Visual Recognition of Objects:

Behavioral, Computational, and Neurobiological Aspects

Jack M. H. Beusmans

beusmans@ics.uci.edu
Dept. of Information & Computer Science
University of California
Irvine, CA 92717

August 1987
Technical Report TR-87-21

Abstract. I surveyed work on visual object recognition and perception. In animals, vision has been studied mainly on the behavioral and neurobiological levels. Behavioral data typically show what the visual system, by itself or together with the rest of the organism, is capable of. They show, for example, that humans can recognize objects regardless of size and position, but that rotated objects pose problems. Important insights into the organization of behavior have also been provided by people who suffered localized brain damage. We have learned that the brain is divided into areas subserving different and relatively well-defined behaviors. The visual system itself is also organized in different subsystems; the visual cortex alone contains nearly twenty maps of the visual field. And individual neurons respond selectively to visual stimuli, e.g., the orientation of line segments, color, direction of motion, and, most intriguingly, faces. The question is how the actions of all these neurons produce the behavior we observe. How do neurons represent the shape of objects such that they can be recognized? Before we can answer this question, we have to understand the computational aspect of shape representation, the nature of the problem as it were. Many methods for representing shape have been explored, mainly by computer scientists, but so far no satisfactory answers have been found.

CONTENTS

When one reflects on the computations that must have to be carried out before one can recognize even such an everyday scene as another person crossing the street, one is left with a feeling of amazement that such an extraordinary series of detailed operations can be accomplished so effortlessly in such a short space of time. (Crick 1979)

1. Introduction

Vision is clearly one of the most important senses in animals. Over half the cortical surface of macaque monkeys is devoted to vision (Barlow 1986), and although this proportion is much smaller in humans, their visual area is still quite large in absolute terms. Even though all visual systems, just like the other senses, provide information about the animal's environment, we cannot expect to arrive at a general theory of vision that would apply to all animals. We might find that flies and monkeys both have looming detectors, but that is where the similarity probably would stop. In this survey, I will focus on how vertebrates, particularly mammals such as cats and monkeys, identify objects in their environment using shape information obtained from visual images.

I start with a review of behavioral data, which will give an idea of what the visual system is capable of doing, what it finds problematic, etc. This could be viewed as a description at the system's level. I conclude with an overview of the biological substrate of vision, that is, visual neurons and their organization.

It is clear that there is a large gap between the description of behavior—recognizing objects from a variety of positions, lighting conditions, etc.—and the physiology of neurons and neuronal circuits, the components of the behaving animal's brain. How can we connect these two aspects of visual recognition? How do visual neurons respond to their environment so as to make possible the visual recognition of objects? What aspects of shape are important for recognition? A number of interesting descriptions of shape have been formulated mainly by computer scientists and engineers in an effort to make machines that can recognize objects. I will review the results of this research in the section on computational aspects of object recognition. Understanding shape at the computational or theoretical level will hopefully help us forge a link between brain and behavior.

Interestingly, two thousand years ago the Greeks approached vision in much the same way (Lindberg 1976). As psychologists and philosophers, they were investigating how the shape and color of an object is mediated to the soul of the observer, the all important question being whether the observer sends out rays to the object or whether the object sends rays directly to the observer. As physicians trying to understand and cure eye diseases, they were concerned with the anatomy and physiology of the eye. And as mathematicians they studied perspective or what we today would call optics.[1]

[1] Similarly, it appears that the "two cultures" existed two millennia before C. P. Snow coined the term

Although the representation of shape is the subject of the present survey paper, I will not go into much detail about how shape can be inferred from visual images. For this I refer the reader to Marr (1982) and the following papers. For shape from motion see Ullman (1979, 1984b), Longuet-Higgins and Prazdny (1980), Hoffman and Bennett (1985), Koenderink and van Doorn (1975, 1986a), and Koenderink (1986); for shape from shading see Ikeuchi and Horn (1981), Pentland (1984), and Koenderink and van Doorn (1980); for shape from occluding contours and line drawings see Waltz (1975), Kanizsa (1976), Koenderink (1984d), Hoffman and Richards (1982), Stevens (1981a), Lowe and Binford (1985), Richards et al. (1985), and Malik (1987; in press); for shape from stereo see Julesz (1964), Koenderink and van Doorn (1976a), Marr (1982), Mayhew (1982), Mayhew and Longuet-Higgins (1982), Barnard and Fischler (1982), and Poggio and Poggio (1984); and for shape from texture see Stevens (1981b), Witkin (1981), Julesz (1981) and Aloimonos and Bandyopadhyay (1985). In addition to these visual sources of information, laser range finders are now available that can give a so-called depth map of an object, specifying the distance between observer and object as a function of the direction of the laser beam. For the purpose of this survey I simply assume that shape information has been obtained somehow.

The following works provide surveys of the separate topics discussed in the present paper. For the psychology of visual perception I can recommend Uhr (1966), Zusne (1970), Julesz (1971), Davies et al. (1981), Marr (1982), Howard (1982), Rock (1983), Shepard (1984), Pinker (1984), and Bruce and Green (1985).

Computer vision and pattern recognition is treated comprehensively in Duda and Hart (1973), Requicha (1980), Ballard and Brown (1982), Marr (1982), Binford (1982), Pinker (1984), Besl and Jain (1985), Levine (1985), Kanal and Rosenfeld (1985), Horn (1986), and Chin and Dyer (1986). Each year Rosenfeld publishes a compendium of recent articles on picture processing in *Computer Vision, Graphics, and Image Processing.*

For the neurobiology of the visual system see Rodieck (1979), Van Essen (1979, 1985), Lennie (1980), Mishkin et al. (1983), Rose and Dobson (1985), Levine (1985), and Boothe et al. (1985). Of course, Kandel and Schwartz (1985) contains a wealth of information on the whole field of neuroscience.

"I had intended to omit [it]...since it necessarily involves the theory of geometry and most people pretending to some education not only are ignorant of this but also avoid those who do understand it and are annoyed with them...But afterward I dreamed that I was being censured because I was unjust to the most godlike of the instruments and was behaving impiously toward the Creator in leaving unexplained a great work of his providence" (Galen, second century, quoted in Lindberg, 1976, pp11–12).

2. Behavioral aspects

This paper is about recognizing objects by looking at them, that is, about mapping images of objects onto categories. For humans, perceiving an object's shape is an important intermediary step; everyday experience tells us that we can perceive shape without being able to recognize, but that we cannot recognize without perceiving something. More strikingly, persons whose brain has been damaged in certain locations fail to recognize family members by their faces even though they can perceive the shape of their faces and recognize them by their voice (see sections 2.4 and 2.5). This suggests that the perception of shape provides the substrate for recognition,[1] a substrate commonly referred to as the internal representation of the object's shape. Recognition is the process that associates past experiences with this representation.

Thus, when looking at a face you construct a representation that successfully activates memory and you remember the person's name, past history, etc. After inverting a face you typically have a hard time determining its identity and expression (Fig.2.6c). Since people's names and the meaning of facial expressions are still available to you, we conclude that they are not being accessed, i.e., your internal representation of the inverted face fails to elicit the appropriate memories. We say that the perception of the face stimulus changes as a function of orientation (actually, we cannot talk about "face" if by that we mean the perception that results from looking at an upright face). Translating a face has no such effect: moving it to the left or right does not change its appearance and we have no problem determining its identity and expression. We say that the internal representation of the face, and of objects in general, is translation invariant.

The perception of multistable images such as the Necker cube shown in Fig.2.1 can also be accounted for very elegantly by internal representations. Interpreted as a drawing of a three-dimensional object, this figure yields two three-dimensional interpretations: one interpretation is that of a cube as seen from above, the other of a cube as seen from below. One's perception usually alternates between these two possibilities. We say that each percept corresponds to a particular internal representation or description, and that different descriptions (of the same sensory input) give rise to different percepts. Thus each three-dimensional interpretation of the Necker cube corresponds to a different description. What we "perceive" or "see" is determined by the description of our sensory experience (Sutherland 1968; Rock 1974; Humphreys 1983).

The same is true for other modalities. If you are unfamiliar with a language, you will "hear" an almost continuous stream of sounds whereas you will "hear" pauses between words if you are familiar with the language (Sutherland 1968).

[1] The flow of information is not entirely unidirectional. Memory or verbal descriptions can provide information that influences the perception of ambiguous pictures such as Fig.3-1 in Marr (1982).

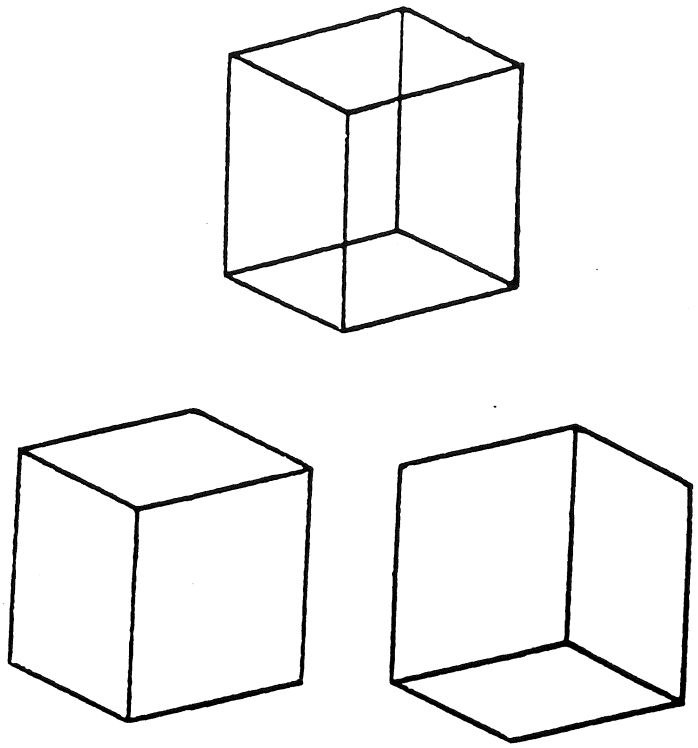


Figure 2.1 **Necker cube and its two 3-D interpretations.**

In this model of perception, commonly referred to as the representational model of mind (see Marr 1982), to perceive is to construct an internal representation. Unfortunately, the term "internal representation" is misleading, at least to my mind. It suggests that an observer re-presents something external that is there independent of its observation, and that this "re-presentation" is to be looked at by ... But to say that I *re-present* the shape, say the roundness or squareness of a rock, is to suggest that it is round or square all by itself. It would be more accurate to talk about our "appreciation" of shape, in this case the roundness or squareness of the rock. It restores the proper subjectivity and it eliminates the need for a homunculus looking at our "re-presentation." Shepard (1984) expressed this in a beautiful metaphor: perceptual systems are stringed like aeolian harps. There is no need for an "intentional hammer", i.e., homunculus, to strike any strings to produce "music," gusts of wind will do, provided the strings are properly tuned.

Koenderink (1980) raised related issues when he noted that the term information is often used ambiguously, sometimes referring to structure in the outside world, sometimes to meaning in the context of an observer. "If you can perceive the solid shape of moving bodies, then it follows that you are receptive to the relevant structures. 'Solid shape' is not present in nature but is a mutual property of perceiver and environment....you [the perceiver] do not "extract" what is already there: what is there depends on *me*" (emphasis in original). Thus

to speak of neurons as feature detectors is misleading: they do not detect features that exist out there. Instead, they *define* what is to be considered a feature. To paraphrase Winograd and Flores (1986), the nervous system is a generator not a filter.[2]

Since the term internal representation is so well established, I will continue to use it but with the above reservations in mind. The present paper thus concerns the internal representation of shape for the purpose of visual recognition. Throughout I will briefly describe the experiments on which conclusions are based because "what an animal *can* do may differ from what it *does* do on any occasion...An animal that fails to perform in a particular way may indeed lack the necessary capacity for it or it may just be that the particular test has failed to activate it" (Herrnstein 1985). And if an animal does solve some perceptual task, it is still a long way from proving that it actually used this or that representation and strategy. And many apparent contradictions in the behavioral literature can be resolved by differences in experimental procedures.

2.1 Parts and wholes

> The Associationists asserted that the experience of complex wholes is built by combining more elementary sensations, while the Gestalt psychologists claimed that the whole precedes its parts, that we initially register unitary objects and relationships, and only later, if necessary, analyze these objects into their component parts or properties (Treisman and Gelade 1980).

As adults we typically perceive objects unitarily: we see telephones, chairs, squares, triangles, etc. We do not see squares as consisting of four line segments at right angles, or as four right angles in a particular configuration. The question is whether this impression is real or not: do our adult visual systems recognize four equally long lines at right angles directly as a square or do they first recognize the lines as such and then their particular arrangement as that of a square.[3] Hebb (1949) argued that our impression of immediacy is illusory. He based his conclusion partly on the way in which adults who had grown up with congenital cataracts learned to identify objects after their vision was restored. Apparently it is not easy to distinguish triangles and squares: After working on it for thirteen days, one person still had to rely on counting the number of corners. Of course, this could merely mean that the perception of identity only becomes immediate after a learning period. It remains to be

[2] Not everybody sees it this way. Rolls (1987) defines the fundamental problem of the senses as the "[reduction of] the redundancy present in the input signals and [the extraction of] a functionally useful information representation."

[3] The right hemisphere seems to be better at perceiving the more global patterns or relationships between components, whereas the left hemisphere seems to see the components only and not their relationships (Bradshaw and Sherlock 1982; Delis et al. 1986). The spatial relationships expressing syntax in sign language, however, are controlled by the language center in the left hemisphere (Bellugi et al. 1983; Poizner and Lane 1979; Poizner et al. 1979, 1984ab; Damasio et al. 1986). Similarly, there is evidence of a left hemispheric superiority in generating mental images (Farah 1985; Kosslyn 1987).

shown that, after identity has been established during learning, perceptual processing s· passes through a stage in which parts are represented as such.

A large number of experiments, in particular those of Treisman and co-workers, provide converging evidence that this is indeed the case. Treisman's feature integration theory of attention holds that the visual image is analyzed and coded in parallel along a number of dimensions or features such as color and movement (Treisman and Gelade 1980). The visual system simultaneously constructs a number of feature maps of the visual image—a color map, a motion map, etc. To combine the features of an object, attention is focused on the corresponding location in the image, thereby linking the feature maps.[4] The evidence for this model mainly consists of reaction time studies in which subjects have to find a target in a display with variable numbers of distractors. For example, the time to decide whether there is a red target in the image does not depend on the number of yellow and green distractors. The red target is said to "pop out." T same holds for simple shapes: an "A" will pop out among "Ts." But conjoining shape and color to form say a"red A" causes decision time to increase linearly with the number of distracting "red Ts", "green As" and "green Ts". In terms of the feature-integration theory, we conclude that there are different feature maps for color and shape. For each feature map, a target "pops out" immediately, whereas conjunctions of two features can only be verified by inspecting each location in the visual field in turn, causing reaction time to increase linearly with the number of distractors. Similarly, the time to verify the presence of an "R" increases with the number of "Ps" and "Qs" as distractors, suggesting that even familiar shapes are composed of separable features.

Nakayama and Silverman (1986) showed that just as conjoining color and shape requires serial search so does conjoining motion and color. However, conjunctions of stereoscopic depth and either color or motion did not require serial search: Each depth plane is searched in parallel for the target.[5] Dick et al. (1987) confirmed that motion could be detected in parallel for small displacements (short-range motion), and added that long-range motion requires serial processing.

The feature-integration model is further supported by illusory conjunctions, errors in conjoining features, causing subjects to report a "red A" in a field of only "red Bs" and "green As." The interpretation is that if features from different locations can be combined to produce an illusion, these features must have existed as such. Thus at approximately the location of the "green A" the visual system registers "green" and "A" separately, and at a "red B" it registers "red" and "B." If the display is shown for say less than 200ms,

[4] However, Prinzmetal et al. (1986) reported a slight influence of attention on feature encoding.

[5] Authors note that single neurons in area MT of visual cortex are tuned to both direction of motion and disparity. Moreover, processing of color and motion is separated within area V2, and results are relayed to different areas: V4 for color and MT for motion. However, in V4 no cells have been found sensitive to both color and disparity (see section 4.2.3).

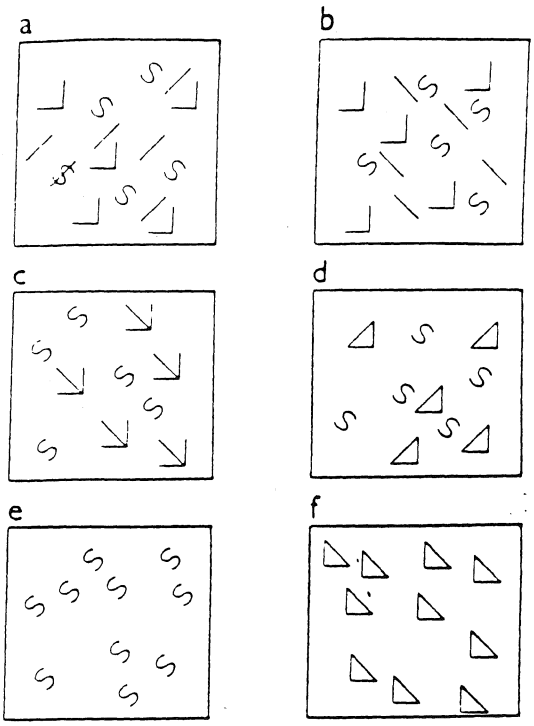


Figure 2.2 **Displays to elicit illusory conjunctions between line segments in different contexts and an "S" to produce a "$" (Treisman and Paterson 1984).**

nearby features "red" and "A"' are occasionally wrongly combined to produce an illusory conjunction. If the theory is correct, conjunction errors can be used to determine what counts as a separate feature and what does not. As Treisman and Paterson (1984) put it: "It conjoins; therefore it is." Using the displays of Fig.2.2, they showed that the sides of triangles can form illusory conjunctions with appropriately oriented "Ss" to produce "$s" (Fig.2.2d) at about the same rate as do isolated lines (Fig.2.2b) and arrows (Fig.2.2c). This means that a "triangle" is not a holistic, unitary percept in the sense of being a fundamental form whose perception precedes that of its components; on the contrary, a triangle is comprised of three lines that can be pried away from it.

Perceptual features need not be as "concrete" as line segments or color. Treisman and Paterson (1984) obtained evidence that the apparently more abstract notion of closure, i.e., space enclosed by a connected line of any shape, might itself be a feature. In a display of angles and lines, conjunction errors led subjects to see triangles. The number of such illusions increased for about half the subjects after adding a circle, presumably having the feature closure, to the display. And if closure were a primitive feature it would result in pop out, as Treisman and Paterson (1984) indeed report. However, Julesz has constructed textures that cannot be segregated in parallel even though their components differ in the closure property (Fig.2.3).

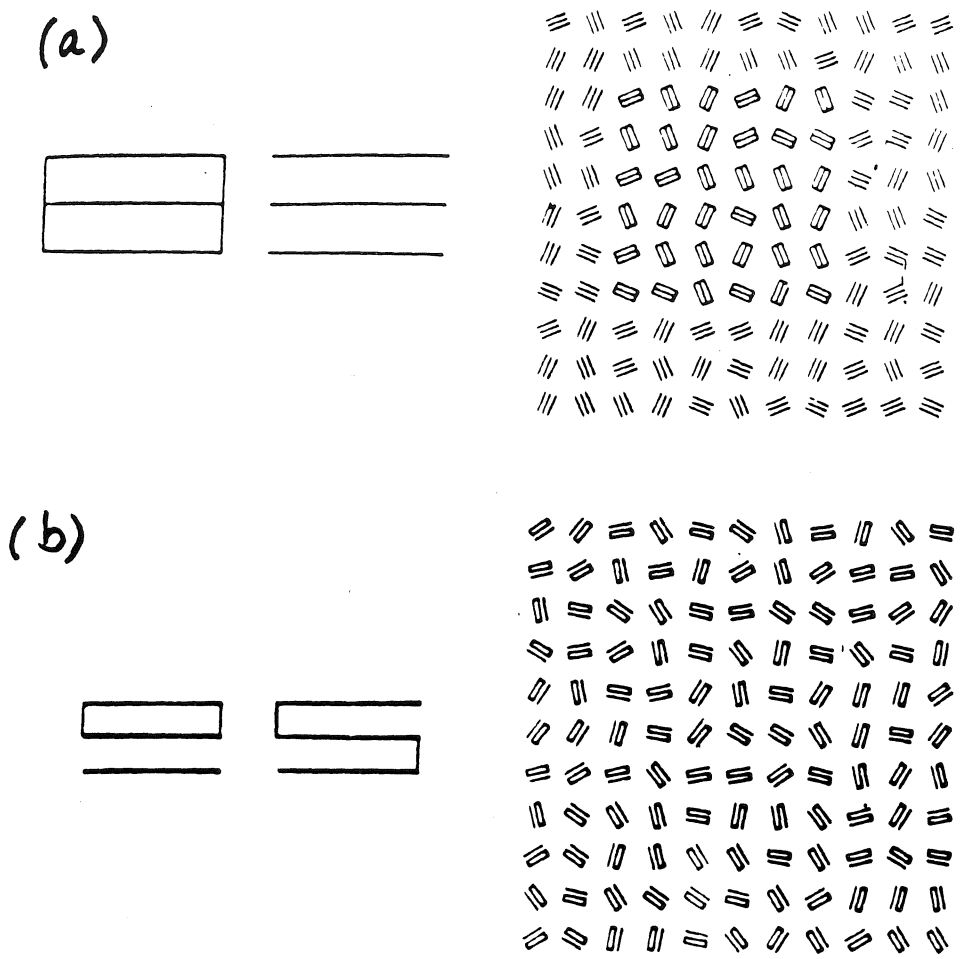


Figure 2.3 **(a) Preattentively discriminable textons differing in the number of terminators resulting in texture segregation. (b) Preattentively indistinguishable texture pairs: number of terminators is equal (Julesz 1984).**

In Julesz' texton theory, closure is not a primitive property, but terminators are (Julesz 1981, 1984); in addition, elongated blobs, crossings of line segments, binocular disparity, motion parallax and flicker also constitute textons. What counts as a texton and what not is an empirical question: If two different textures, shown side by side or one inside the other as in Fig.2.3, are segregated in less than 200ms they are said to consist of different textons. This phase is called preattentive vision; focal attention is needed to distinguish two objects whose textons are the same but differ otherwise (e.g., the precise location of line elements making up a crossing). The obvious suggestion is that these stages correspond to the parallel and serial stages in the feature-integration theory of Treisman. Interestingly, Julesz and Treisman both found that serial search takes about 50ms per item (Julesz (1984) reported 50ms; Treisman and Gelade (1980) reported considerable variation, but their average is about 60ms).

An important aspect of the preattentive stage in both the feature-integration and the texton theories is that features/textons are not glued to a particular location in the visual

field but can float around and cause illusions.[6] A number of theories have been developed to account for this indeterminacy. Wolford (1975; Wolford and Shum 1980) claimed that the positional information associated with a feature degrades or is perturbed over time. Prinzmetal (1981) proposed that the perceptual group (defined roughly by the Gestalt laws of organization) in which a feature is embedded influences the metric of visual space and hence feature perturbations, such that features from the same perceptual group are more likely to be integrated or falsely conjoined than features from different perceptual groups. Following up on this perceptual grouping principle of feature integration, Prinzmetal and Millis-Wright (1984) showed that conjunction errors involving shape and color occur more frequently within words than nonwords.

Assuming that perception indeed goes through a preattentive and attentive stage, which locations are selected for attention and how is attention directed to these locations? Julesz (1984) merely suggested that attention is directed to texton boundaries. Similarly, Koch and Ullman (1984; Ullman 1984) suggested that attention is directed to locations that differ sufficiently from their neighbors. For this purpose, they proposed a retinotopic saliency map which combines information from all feature maps to specify which locations are the most interesting. Focus of attention is not shifted to just any conspicuous location, but is posited to show preference for nearby locations that are similar to the currently attended location. Furthermore, they conjectured that the saliency map may be located within the LGN or area V1 (Koch and Ullman 1984; Sherman and Koch 1985). The LGN is a good candidate because it not only relays information from the retina to the cortex, but also receives extensive feedback from cortical areas.[7] Crick (1984) proposed that the reticular complex of the thalamus, of wich the LGN is a part, implements an internal search light aligning different feature maps.

In the feature-integration theory, perceptual processing proceeds in one direction, from simple features to more complex wholes. However, a number of experiments have shown that local processing can be influenced by context. Letters are recognized faster within a word, and line segments are detected faster within a line drawing having a unitary 3-D interpretation than in an arbitrary context (Weisstein and Harris 1974; Pomerantz et al. 1977; Williams and Weisstein 1978; McClelland and Rumelhart 1981). Fig.2.4a illustrates one such experiment in which the different stimuli were presented in a brief flash, and subjects had to indicate whether a diagonal line segment was present or not. As shown in Fig.2.4b for a variety of experimental conditions, diagonals are detected most accurately in a meaningful

[6] Strangely enough, Sagi and Julesz (1985) argue that, on the contrary, the location of feature gradients is known very accurately already at the preattentive level, and that attention is needed to identify features. However, their experimental procedure differs so much from Treisman's that a direct comparison is hardly possible.

[7] See section 4.2.2 for more details.

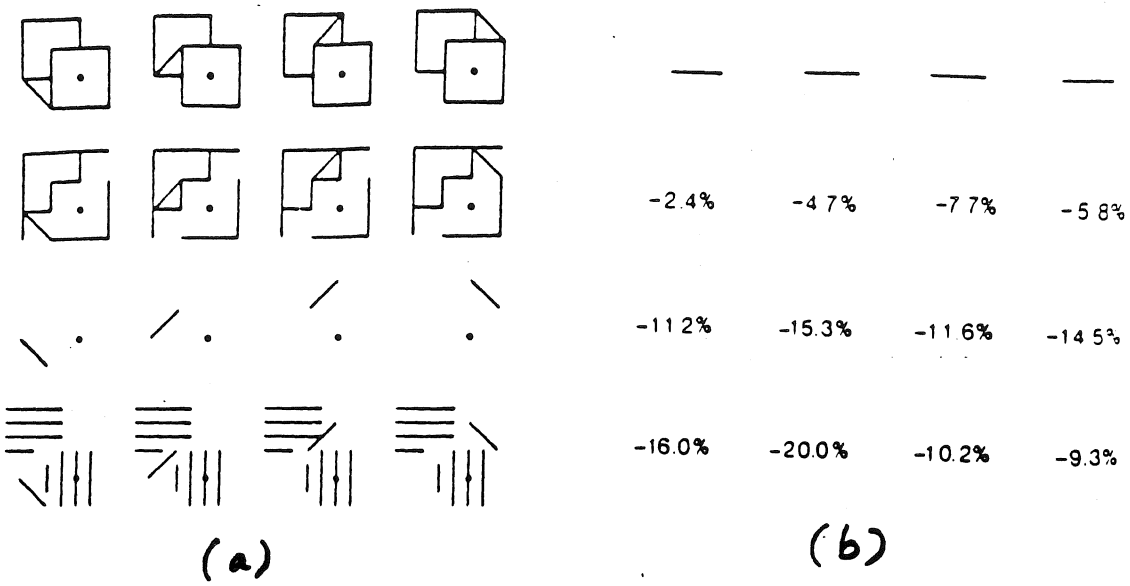


Figure 2.4 (a) The four different contexts in which a diagonal line appeared. (b) Mean difference in accuracy between object context and the three contexts; the four columns correspond to slightly different experimental conditions (Williams and Weisstein 1978).

context.

This effect, known as object superiority, does not necessarily mean that the perception of the whole precedes that of the parts. It might be the result of mutual reinforcement as the following network simulation illustrates.

Rumelhart and McClelland (1981) obtained letter superiority effects with a so-called interactive activation model shown in Fig.2.5a. The model consists of three layers. The lower layer consists of feature detectors that are connected to form letter detectors at the middle level, while the top level combines letter detectors into meaningful word detectors. Higher levels project back to lower ones through excitatory feedback links that reinforce lower level activity. In addition, detectors at the letter and word levels inhibit nodes at the same level, and feature detectors inhibit all letters except for one. Upon presentation of a visual stimulus, some feature detectors increase their activity, at some point activating letter detectors. If the latter have become sufficiently active, they will activate word detectors. Convergent evidence at a high level (i.e., word or letter) is fed back to lower levels reinforcing their activity, which in turn increases activity at the higher levels, etc. Hence the name interactive activation or spreading activation model. It is now clear how context exerts its influence: If a "T" occurs in the context of "TRIP" or "TIME" in the network shown in Fig.2.5a, it will be reinforced more than within "TOVG"; in other words, it will have a lower detection threshold. Similarly, mutual reinforcement serves to resolve ambiguities or fill gaps in the input (Fig.2.5b and c).

As a model of visual letter recognition, McClelland and Rumelhart's model is of course

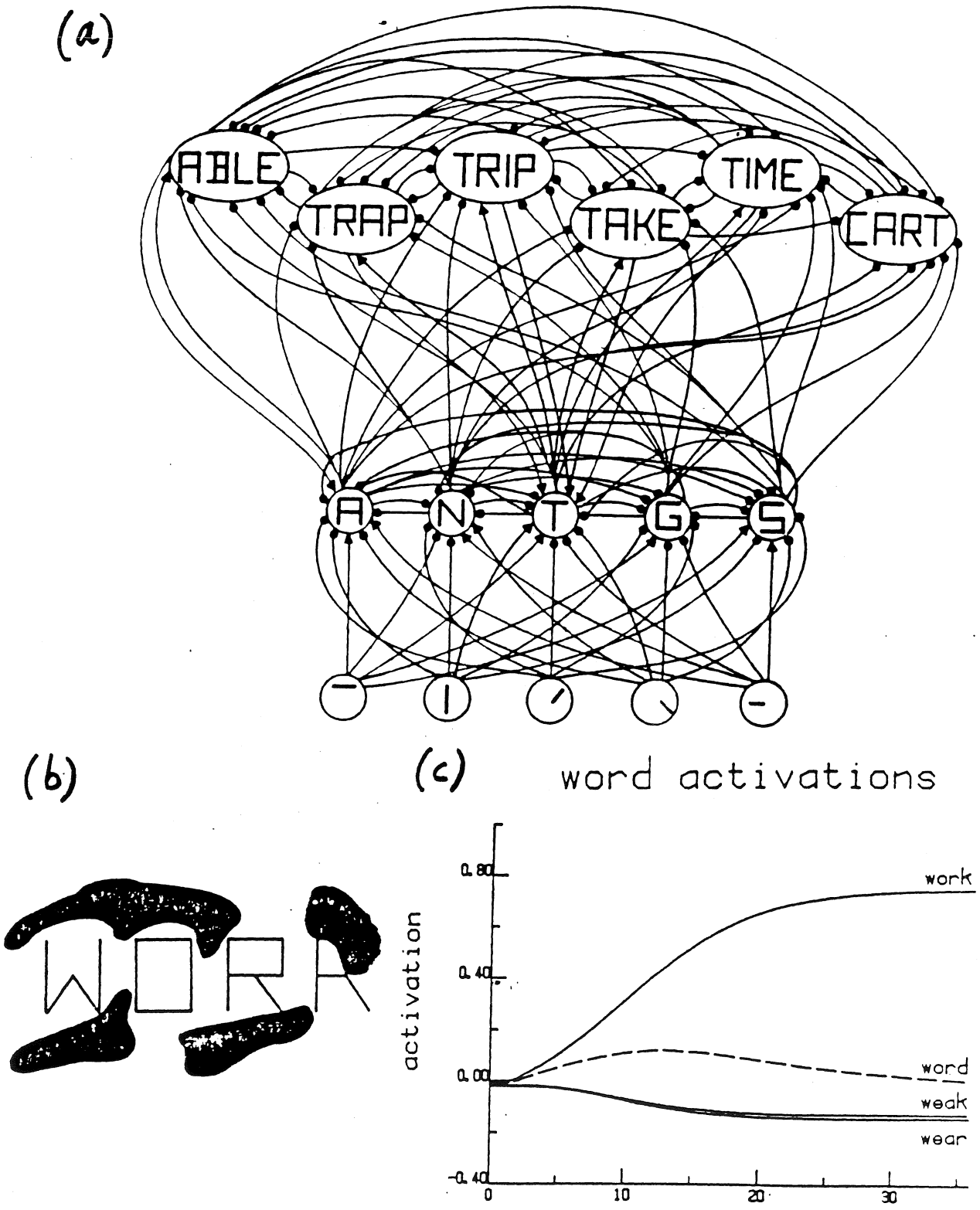


Figure 2.5 (a) Three level network for word and letter recognition; links with arrowhead are excitatory, links with circle are inhibitory. (b) Degraded input. (c) Activation level of several word nodes; note the initial competition between "work" and "word" (McClelland and Rumelhart 1981).

quite crude; their particular feature detectors are not very plausible. It does, however, demonstrate that context superiority effects can result from interactions of local feature detectors, without having to posit holistic processing or that the whole is more than its parts. On the other hand, this demonstration does not exclude the possibility that images can be processed at different scales, and that the results of processing at the larger scales can direct that at smaller scales. In fact, much research is currently directed towards multi-scale representations of images and 3-D objects themselves (Marr 1982; Koenderink 1984a; Koenderink and van Doorn 1978, 1982, 1986b; Witkin 1983; De Graaf et al. 1984; Richards et al. 1986; Mokhtarian and Mackworth 1986; Burton et al. 1986; Pizer et al. 1986, 1987; Kroese 1987).

One of the problems in doing experiments on local and global processing is determining what is local and what is global, and then controlling for it. For this purpose, Navon (1977) introduced so-called compound letters, that is, large letters consisting of smaller ones; for example, a number of small "As" arranged in the form of a "T." These visual stimuli were flashed before subjects who at the same time heard the name of a letter; their task was to report what they had heard as fast as possible. Compared with subjects who receive consistent information, subjects receiving inconsistent information—looking at an "A" and hearing a "T"—take longer to react. Navon turned this around, and used reaction time as a measure of consistency and as a clue to the contents of, in this case, the visual percept. He found that reaction times increased only when the global level of the visual stimulus conflicted with the auditory information, and not when the local level conflicted. This suggests that global processing precedes local processing. However, there are a number of problems. First, do these stimuli really control the local and global levels? It might be that the small letters simply function as place markers, lining up to form two bars say, one horizontal and one vertical together making a "T." Thus, the subjects report a "T" because they saw the local features of a "T", i.e., a horizontal and a vertical bar. Second, what subjects report verbally or otherwise is not necessarily indicative of the sequence of processing, since "reaction times to wholes or to parts may... reflect most directly the speed of access to the final output of perceptual analysis" (Treisman and Paterson 1984, p13).

This last comment about the use of reaction times underlines, appropriately so in my mind, the tentative nature of many of the conclusions reached by the experiments discussed in this section. Treisman and Paterson (1984, p31) readily admit that "[they] use the conclusions from one set of data as the premises for predicting another set of data, and conversely the conclusions from the second set as validating the interpretation of the first," but argue that "the consistency of a fairly large set of results...can strengthen [the theoretical story]." It remains to be seen whether more direct behavioral tests can be designed.

2.2 Form and orientation

In the introduction, I described the task of vision as having to find out what is where. These two aspects of vision are largely independent: the identity of an object does not depend on its position, nor does the identity of an object determine its position.[8] Everyday experience tells us that we can recognize a cat whether it is sitting in a chair or on top of a cabinet, or hanging in the curtains (of course one seldom finds dogs in the latter position), although children might initially link identity and location (Vernon 1966). In our representational model this means that the internal representation of shape is invariant under (certain) movements in the environment.

A good illustration of this principle of shape constancy is provided by size constancy: We do not perceive persons as becoming progressively smaller as they walk away from us, even though their retinal projection decreases in size. Conversely, as illustrated in Fig.2.6a, the same retinal object appears to be larger if other cues such a perspective show it to be farther way. Thus we scale objects by their distance.

The effect of rotation on shape perception is more complicated. Mach's (1902) celebrated example of the diamond and square (Fig.2.6b) illustrates that rotation can drastically change an object's appearance. An even more dramatic illustration is "Thatcher's illusion" (Fig.2.6c). It is very hard to see that the eyes and mouth in the right-hand side face have been inverted, something which is immediately obvious when you turn the page around and look at the faces right side up. In this section, I will discuss the effects of rotation on recognition, effects that might throw some light on the internal representation of shape.

Rock (1973, 1974) was one of the first to systematically analyze the effect of orientation on form. His theory of shape perception takes into account not only the internal configuration of a figure but also its orientation in the environment and with respect to the observer. According to this theory, perceived shape is determined by the viewer's assignment of top, bottom, and side directions, just like perceived size is influenced by perceived distance from the viewer. Information about these directions can be obtained from a number of sources including gravity (through the vestibular system) and the visual environment (Brecher et al. 1972; Schoene 1984).

As we already indicated, there are two obvious coordinate systems in which orientation could be important; one is a viewer-centered retinal coordinate system, and the other is environment-centered. One would expect retinal orientation to be unimportant; and indeed it is easy to demonstrate that tilting one's head does not affect one's perception of figures. A square remains a square, a diamond remains a diamond. Rotating the figure, on the other hand, changes its phenomenal shape because the assignment of top, bottom and sides has

[8] For a review of behavioral evidence, see Leibowitz and Post (1982); for neurobiological evidence, see section 4.1.

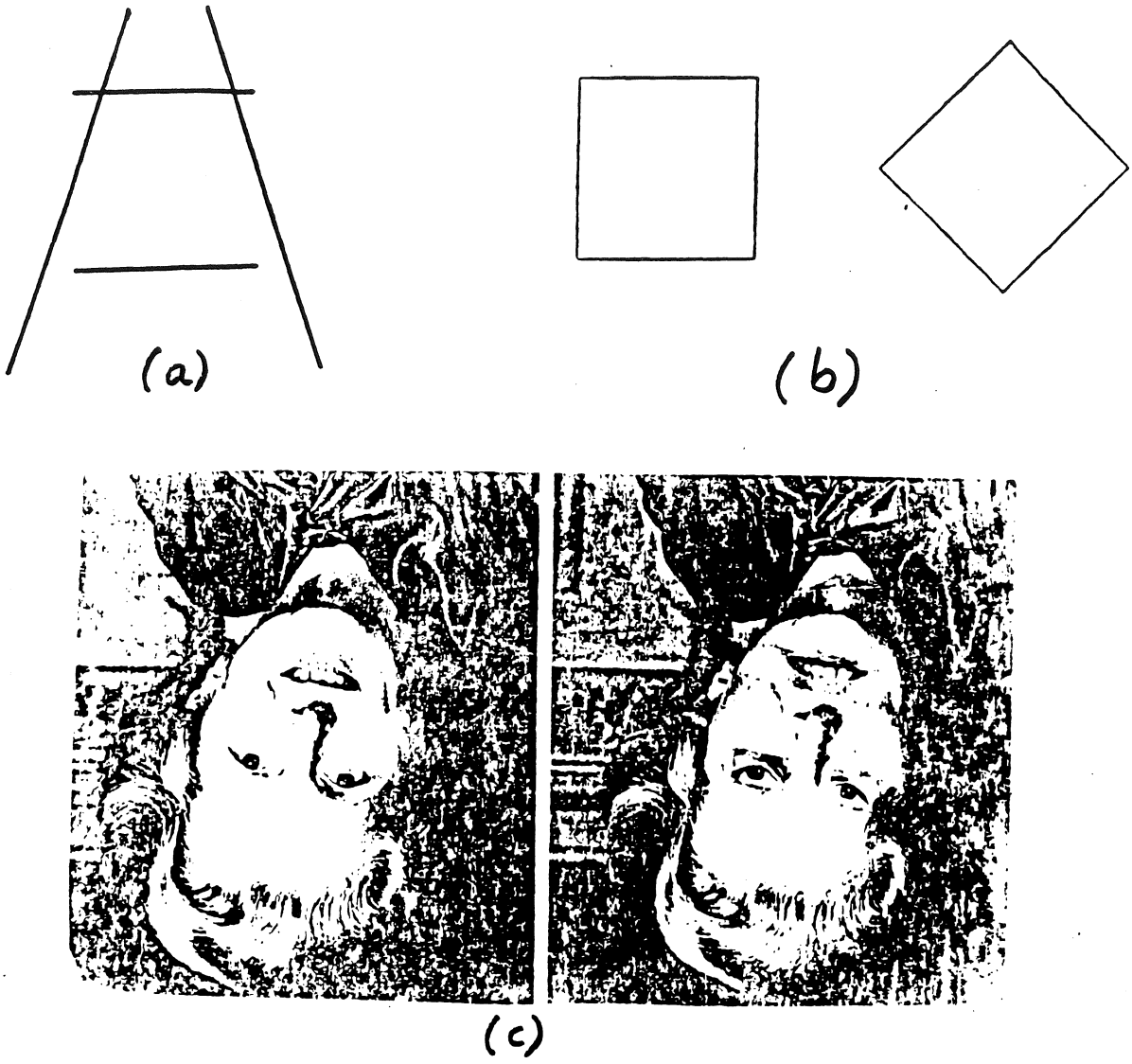


Figure 2.6 (a) size constancy; (b) square and diamond (Mach 1902); (c) Thatcher's illusion illustrates that facial expression is hard to recognize in upside-down faces—the eyes and mouth are inverted in the right-hand side face (Thompson 1980).

changed and with it the internal representation of the figure's shape. Factors influencing this assignment of directions—such as knowledge that the figure has been rotated, pronounced elongation of the figure itself, familiarity—may undo this perceptual change. Corballis and Cullen (1986) showed that for simple 2-D symbols such as capital letters the assignment of top and bottom is largely independent of orientation. In contrast, time to indicate the left or right side increased linearly with rotation from the upright suggesting something like mental rotation (in general, they found that any decision involving mirror image discrimination showed this dependence; we will turn to mental rotation in the next section).

There are circumstances in which retinal orientation does affect perception. Reading print or longhand while one's head is tilted is quite difficult (Kolers and Perkins 1969), it is similarly hard to recognize photographs of faces with one's head upside down. However,

in both cases we can correctly assign directions. Rock (1973, 1974) accounted for these difficulties by postulating that the correction mechanism that allows you to recognize an inverted "E" cannot handle a complete sentence or a face. Correction does not work if there are too many parts all of which have to be corrected simultaneously.

Extending these experiments to 3-D shapes, Rock et al. (1981) found that rotating wireframe figures 90° about the vertical causes the recognition rate to drop from 82% to 43%, whereas rotations of 180° do not affect recognition. The authors explained this by the qualitative change in retinal projection, resulting in an egocentric description very different from the original one, and one which cannot be corrected for. A similar pattern emerged for rotation about the horizontal (of the image plane): A 90° rotation resulted in 21% recognition, while a 180° rotation decreases recognition to 61%. For comparison with the latter, inversion (rotation in the image plane) decreased recognition to 57%. As a result of these and related experiments, Rock and DiVita (1987) concluded that shape description is mainly determined by the image from a particular vantage point, in other words, it is viewer-centered as opposed to object-centered. However interesting these results, I wonder how much they depend on the use of wireframe figures; in particular it is not clear how strong 3-D perception is. It would be important to extend the stimulus set to include solid objects.

Wiser (1981) studied the influence of rotation on figures with an intrinsic, clearly visible, axis of elongation. She used a learning and recognition paradigm in which subjects first learned to recognize 2-D nonsense shapes whose axis of elongation was oriented either vertically, obliquely or horizontally (in a retinal coordinate system, which in this case was the same as the environmental one; Fig.2.7a). That is, the same figure was always presented in the same orientation. In the recognition phase, figures were shown in all three orientations. Interestingly, subjects reacted fastest when figures were oriented vertically, regardless of the orientation in which they had been learned (Fig.2.7b). In fact, with one exception, reaction time increased linearly with the angular departure from the vertical in a manner reminiscent of mental rotation.

Wiser proposed that initially the image is described in the retinal reference frame, next an intrinsic axis is computed (top and bottom directions assigned) after which the so-called perceptual frame of reference is rotated to align with the intrinsic axis. Finally, the shape of the figure is described with reference to the new frame, and this intrinsic description is stored in memory and used for subsequent recognition. Thus she argues for a canonical object-centered description as opposed to a viewer-centered one, at least for figures with salient intrinsic axes. However, this does not explain why inverted faces are so hard to recognize (in the sense of individuation, it is easy to recognize that a picture is of a "face"), even though there is a clear intrinsic axis. Presumably it is hard to keep track of a large number of parts in the perceptual frame defined by the intrinsic axis. This often invoked

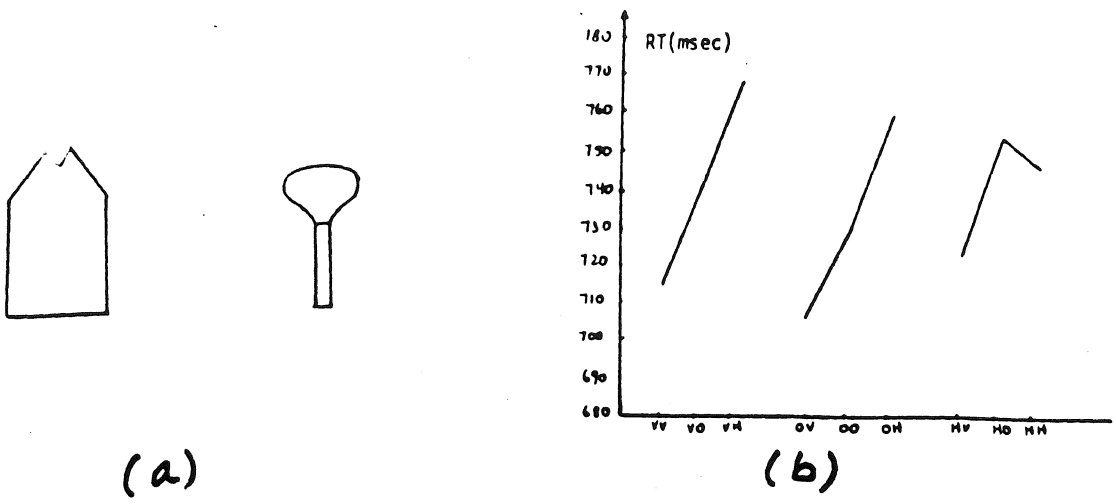


Figure 2.7 (a) Stimuli had clear axes of elongation; (b) Reaction times for "yes responses." Since there were three possible orientations during the learning and three during the recognition phase, there are nine experimental conditions plotted on the abscissa (VV through HH). The first letter indicates orientation during learning, the second during recognition; V=Vertical, O=Oblique, H=Horizontal (Wiser 1981).

hypothesis could be tested by using a stimulus set whose elements can be rank-ordered by complexity. As a measure of complexity one could use the number of parts (object-centered measure) or the number of different projections for 3-D objects (the latter viewer-centered measure is further discussed in section 3.2; see also Fig.3.8).

Pointing out that there are often several intrinsic (object-centered) frames of reference, Hinton (1981ab) suggested that the final choice for the reference frame depends on whether it results in a familiar description. Rather than a strictly serial progression from retinal description to intrinsic axes to intrinsic description to memory we now have a parallel model in which memory actively influences perception. Descriptions in different perceptual frames all vie for attention from memory as it were. Hinton designed a parallel network that simultaneously converges on a particular frame of reference and a shape description in that context. Thus, recognizing an object means recognizing the object and seeing it in a particular orientation. An interesting problem is how several objects or parts of an object that all have different orientations are treated. Presumably, each would give rise to a different canonical frame, and the question is how these different coordinate systems compete. It might be the case that attention can only be focused at one orientation and that this problem simply does not arise in biological systems. Indeed, according to Julesz (1984) relative positions of local features are not computed during preattentive vision.[9] Only with focal vision can we discern these spatial relations and then only, by definition, locally. Shifting focal attention to different locations obliterates all previous information (perceptually speaking).

[9] But see Sagi and Julesz (1985) for a different opinion.

To further probe whether people use viewer- or object-centered coordinate systems, Jolicouer and Kosslyn (1983) constructed 3-D stick figures whose descriptions would be more or less similar depending on the coordinate system used. The stick figures consisted of a main axis to which a small number of limbs were attached. To create two objects having similar object-centered but different viewer-centered descriptions, all one has to do is rotate the object. The converse is more complicated and was presumably accomplished by changing the main axis of an object but not its components. Inspection of Fig.2.8 leaves some doubt about this. In any event, pairwise similarity judgments of subjects could indeed be factored along the viewer- and object-centered dimensions regardless of whether the two objects were shown simultaneously or in sequence. Thus, subjects use viewer- and object-centered information to judge similarity. By instructing subjects to encode the shape of an object such that they could recognize or draw it from their particular point of view, or such that they could recognize or draw it after an arbitrary 3-D orientation, subjects were biased to store a viewer- or object-centered description in memory, respectively. After this learning phase, subjects had to judge 2-D and 3-D similarity. Subjects with a viewer-centered bias judged rotated objects to be less similar than subjects with an object-centered bias, while the similarity judgments for objects with similar viewer-centered descriptions show the opposite trend. The time to judge 2-D similarity was not affected by learning bias (approximately 10.8s in both cases), whereas 3-D similarity was determined faster with an object-centered than with a viewer-centered bias (7.6s as opposed to 9.9s).

These results indicate that subjects can store descriptions with respect to both viewer- and object-centered coordinate systems. However, there does seem to be a slight asymmetry: with an object-centered bias both descriptions are readily available, while this is not the case for a viewer-centered bias. If the situation had been symmetric, 2-D judgements with an object-centered bias would have taken longer, just as 3-D similarity judgments with an viewer-centered bias took longer. This suggests that the object-centered description is built on top of the viewer-centered one. As far as recognition is concerned, the authors conclude that they have not shown that viewer-centered descriptions are actually *used* for recognition, and that the debate over which coordinate system is used has not been resolved.

2.3 Mental rotation

> [T]o draw inferences as to the future...we form for ourselves images or symbols of external objects; and the form which we give them is such that the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured. In order that this requirement may be satisfied, there must be a certain conformity between nature and our thought...The images which we here speak of are our conceptions of things. With the things themselves they are in conformity in *one* important respect, namely, in satisfying the above-mentioned requirement. As a matter of fact, we do not know, nor have we any means of knowing, whether our conception of things are in conformity with them in any other than this *one* fundamental respect. (Hertz 1894, pp1–2)

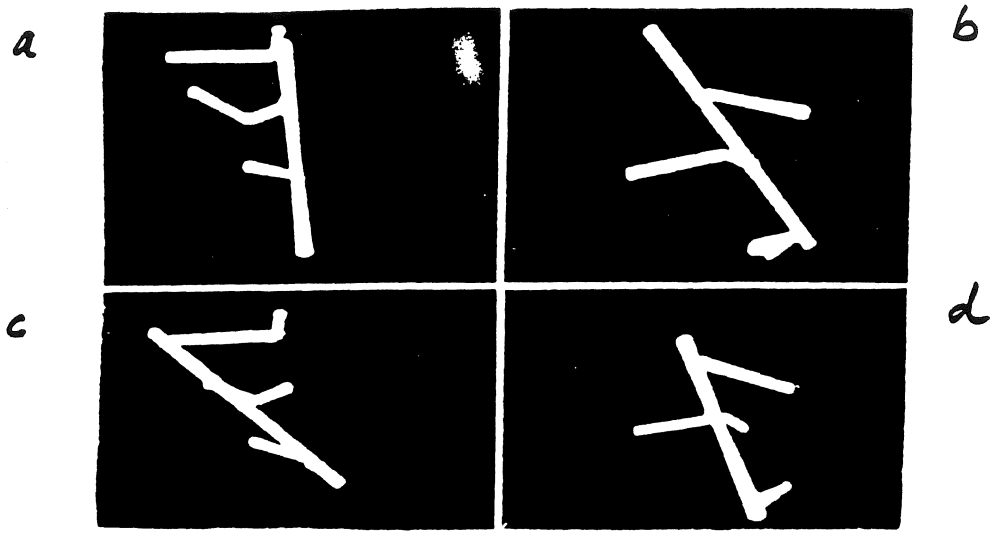


Figure 2.8 **Stick figures with similar viewer-centered or object-centered descriptions. (a) and (b) are rotated versions of the same figure, as are (c) and (d). (c) is obtained from (a) by shifting the main axis; similarly for (b) and (d) (Jolicoeur and Kosslyn 1983).**

Perception is about invariances: even though the retinal image changes continuously,[10] we may perceive constancy, for example, solid shape. One could say that our visual system somehow "decides" whether two images are in fact of the same object. This suggests that one way to study the internal representation of shape is by measuring how long it takes to make this decision.

Shepard and co-workers (Shepard and Metzler 1971) were the first to introduce such experiments. They had their subjects look at two drawings of the same or different 3-D objects and decide as fast as possible whether or not the two drawings derive from the same object. Of interest is how reaction time depends on the difference in viewing angle between the two drawings. However, there is a problem with the difference condition, the condition in which the views are of different objects. If the objects are very different we would expect reaction time to be almost constant because the subject could simply use one feature to distinguish the two objects (for example, one object has a hole and the other does not). This problem was circumvented by using mirror reversed objects in the difference condition (Fig.2.9a). Under these experimental conditions, response time increases linearly with the angular disparity between the two objects shown, irrespective of the plane of rotation (Fig.2.9b).

[10] To be more precise, retinal excitation patterns: to talk about an image is already to interpret, to give meaning, to these patterns. Similarly, Koenderink (1984bc) points out that we have to keep in mind the distinction between the machine—retina, cortex, etc.—and the kind of information it can access, and an observer of the machine. An observer can talk about retinal image or somatotopic maps in the cortex, whereas the machine knows nothing about these things; neurons simply respond to photons, other neurons, etc.

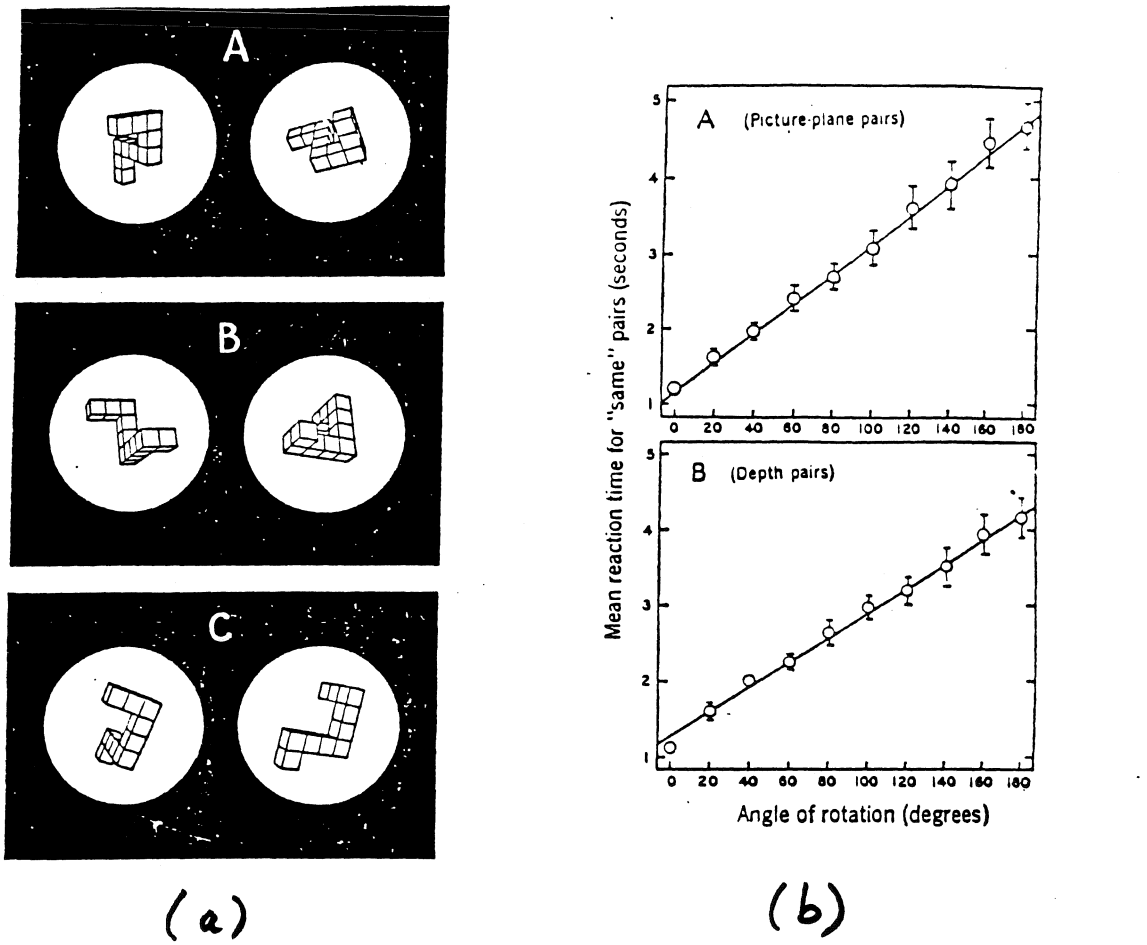


Figure 2.9 **(a) Pairs of perspective drawings: A, "same pair" rotated in picture plane about 80°; B, "same pair" rotated in depth; C, "mirror image pair." (b) Mean reaction times of subjects to indicate that two drawings are indeed of the same object and not of mirror reversed ones: A, for rotation in the picture plane; B, for rotation in depth (Shepard and Metzler 1971).**

In a related experiment, subjects looked at a sequence of views of a 3-D object, much like watching a motion picture. The question is under what conditions subjects experience apparent rigid motion of the 3-D object. Interestingly, the minimum time between successive views increased linearly with their angular difference, again irrespective of the plane of rotation (Shepard and Judd 1976).

These results suggest and subjects indeed report that they "mentally rotate" one object until it maps onto the other (of course, it is the representations of the two objects that are mapped onto each other). And there is evidence that subjects actually image intervening views of the object (Cooper and Shepard 1984). Shepard concluded that "the representation of the possible [spatial] transformations of an object is basic to the representation of the object itself" (Shepard 1982, p51) and that "the representation is not ... of the inherent three-dimensional shape of the object considered purely in itself; it is always of that object

as viewed in a particular orientation" (Shepard 1982, p55).

Referring to Hertz's quote at the beginning of this section, there is a "certain conformity between nature and our thought." When looking at different views of an object in sequence, we connect them and perceive apparent motion, that is, we perceive views of the object that would have been visible had the object actually followed that trajectory.[11] To discuss the nature of the conformity between objects and their representations, Shepard (Shepard and Chipman 1970; Shepard 1982, 1984) introduced the notions of first- and second-order isomorphism. A first-order isomorphism exists between an object and its representation if a property of an object is physically present within its representation. For example, a first-order isomorphic representation of something square is itself square, or the representation of something red is itself red. A second-order isomorphism is said to exist if relations between representations mirror relations between corresponding objects. In that sense, we can call these representations analogous: "To say that an internal process is a mental analog of an external process is, in part, to say that the internal process is similar in important respects to the perceptual process that would take place if a subject were actually to watch the corresponding physical rotation" (Shepard and Judd 1976); in other words, "the intermediate internal states have a natural one-to-one correspondence to appropriate intermediate states in the external world" (Shepard 1978).

Isomorphism goes back to the Gestaltists (Zusne 1970, p109): form is represented in the brain not symbolically but "directly in terms of corresponding points of excitation. The correspondence is not topographical and cannot be topographical. It is topological, that is, order and relationships are preserved, even though distances, angles and curvature are not." And "things look as they do because of the field organization to which the proximal stimulus gives rise" (Koffka 1935 quoted in Zusne 1970, p111). It has been said that because mental rotation mimics real rotation, propositional or symbolic models of mental rotation have to be rejected since they cannot capture this incremental aspect. However, it is straightforward to come up with a propositional theory whose descriptions are related functionally, i.e., analogically. For example, spatial relations among parts of an object are specified with respect to an external coordinate system and two angles can be compared only by repeatedly adjusting one by some constant amount until it equals the other (Minsky 1975; Morgan 1983; Mel 1986).

It is important to note that the phenomenon of mental rotation crucially depends on the use of mirror reverse objects. As mentioned, this was intended to suppress the use of

[11] Shepard (1984) argues that out of the infinitely many possible trajectories we choose the one that is prescribed by Chasles' theorem. This theorem states that two copies of an object at different locations in 3-D space can be mapped onto each other by rotation about a unique axis plus a translation along that axis. Thus, the trajectory is helical or screw-like. For 2-D objects lying in a plane we can dispense with the translation.

distinctive features in discriminating the two objects, and force subjects to mentally rotate images. In fact there is considerable confusion as to what conditions and stimuli produce mental rotation. Hochberg and Gellman (1977) showed that, for 2-D objects rotated in the picture plane, mental rotation rate increases with saliency of landmarks, and that objects having distinctive features are not rotated at all. Cooper and Podgorny (1976), in a related study, failed to find an influence of stimulus complexity on rotation rate when 2-D polygons had to be discriminated from distractors, whose vertices had slightly changed positions. Pylyshyn (1979) attributed this to the fact that distractors were rated equally similar to their respective targets; in other words, the task did not become more complex with increasing number of vertices in the polygones. Shinar and Owen (1973) showed that unfamiliar 2-D objects are mentally rotated, but that this disappears after enough practice; Jolicoeur (1985) demonstrated the same effect for line drawings of 3-D objects. Eley (1982) showed that *identification* of letter-like 2-D symbols is independent of rotation, but that *verification* of whether two symbols are mirror images is proportional to angular difference. Jolicoeur and Landau (1984), on the other hand, did find an effect even for the identification of alphanumeric characters. They argued that reaction time is not a good measure of performance because identification itself is already very fast, causing small increases to be statistically insignificant. Instead, they used the number of errors in brief exposures.

Carpenter and Eisenberg (1978) found that reaction time also depends linearly on angular difference for blind persons who had to rely on touch instead of vision. This raises the possibility that mental rotation acts upon a representation that is more spatial than visual in character.

To test whether other species besides humans use mental rotation for the discrimination of mirror images, Hollard and Delius (1982) trained pigeons to discriminate between identical and mirror images of 2-D forms rotated in the picture plane. Two measures of performance were used: reaction time and percentage error as a function of angular difference. Reaction time for pigeons did not depend on angular difference and was considerably less (between .5 and 1.0s) than that of humans who showed the familiar dependence (1.5s for 0°, 2.7s for 180°). This would suggest that pigeons do not perform anything like mental rotation, a result also reported by Herrnstein (1985). However, the error rates throw doubt upon this conclusion. For pigeons, the error rate profile is symmetric about 90° where, depending on the session, there is a peak of between 10 and 22% errors; for 0 and 180°, the error rate is between 2 and 9%. The results for humans again indicate mental rotation: error rate increases linearly from 2% for 0° to 10% for 180°. Combining Jolicoeur and Landau's reason for choosing error rates over reaction time as a measure of performance when using overlearned stimuli with the fact that the pigeons went through an extensive learning phase, leads me to question Hollard and Delius' interpretation that the absence of any reaction time effect for the pigeons means that their recognition is rotationally invariant.

Mental rotation is clearly an interesting phenomenon; but it would be even more interesting and valuable if we knew more about the conditions in which it occurs. Thus I believe it to be crucial to extend Jolicoeur and Landau's (1984) analysis to patterns other than alphanumeric characters, and to 3-D solids. And if mental rotation is to elucidate the representation of shape, experiments should be designed with particular models in mind, for example, the visual potential (section 3.2).

2.4 Categories

Recognition is the mental process that links current percepts with memory, the repository of past experiences related to the observed objects. But how is memory organized? Does it consist of an unstructured collection of detailed descriptions of objects, or are these descriptions somehow organized, for example, in the form of a taxonomic tree whose nodes describe classes of objects, and whose levels reflect different degrees of abstraction. The structure of memory is clearly a subject of interest in and of itself, but it might also tell us something about the representation of shape.

One way to probe the structure of memory is to measure how fast observers can decide whether a label corresponds to an object. It might be that it takes progressively longer to decide whether a german shepherd is a "german shepherd," a "dog" or an "animal"; in other words, the more abstract a label, the longer it takes to verify the correspondence. Alternatively, verification could be faster with the more abstract names, which could be justified by arguing that it is easier to decide whether something is an animal than a german shepherd (e.g., fewer features are needed). It turns out that human observers are fastest at an intermediate level of abstraction; in the above example, "dog" would have been the first label to be associated with a german shepherd. Rosch et al. (1976) termed this level of abstraction the basic level.

Assuming that objects can be named at different levels of abstraction, are these levels related, and if so how are they related? Jolicoeur et al. (1984) showed that naming an object at a level more abstract than the basic level (the superordinate level) is a two-step process. First the basic-level name is determined on the basis of information derived from the percept, and, second, the basic level provides access to semantic memory and the superordinate name.[12] Fig. 2.10 illustrates that naming is fastest at the basic level, and that naming at the superordinate level does not require extra perceptual processing since reaction time did

[12] Based on the positive correlation of the time it takes to give the superordinate category of a word and a picture; e.g., the word "apple" and the picture of an "apple". When reading the word "apple" one presumably activates the concept "apple" and then the concept "fruit", i.e., one cannot reach the word fruit without passing through apple. If the time it takes to produce the superordinate level of a word correlates positively with the time to name a picture at that level, one has evidence that the superordinate level is accessed through a similar stage, the basic-level category.

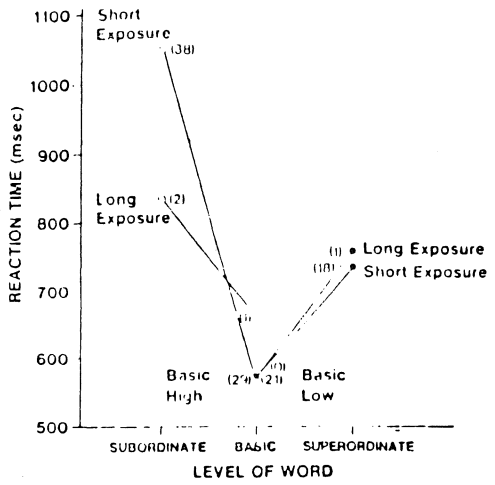


Figure 2.10 **Mean reaction time and percent error (in parentheses) for verifying whether a word is consistent with a picture. Long exposure is 1000ms, short exposure is 75ms. Basic high means that the basic level is at the category level; robin and bird, where bird is the basic level. Basic low means that the basic level is at the exemplar level; apple and fruit, where apple is the basic level (Jolicoeur et al. 1984).**

not depend on exposure time. However, exposure time did affect naming at the subordinate level (a level more concrete than the basic level): if a picture is shown for only 75ms as opposed to 1000ms, reaction time and percent error increase significantly. Thus, decisions at the subordinate level require additional perceptual processing of the image.

However, objects do not always contact memory at the basic level. Jolicoeur et al. (1984) and Murphy and Brownell (1985) found that typicality influences the level at which memory is first contacted: Atypical exemplars such as a penguin for the basic-level category bird are named more often and faster at the subordinate than at the basic level.

Granted there are different levels of categorization, what exactly are the attributes that distinguish the different levels of categorization? Tversky and Hemenway (1984) obtained evidence that the parts of an object play the most important role in its categorization. When asked to describe objects at the subordinate, basic and superordinate levels, subjects used part names most often at the basic level. Categories at the basic level can be accounted for by their part structure: parts are the features shared by members of the same basic level category, and they are the features that distinguish different categories from each other.[13] In contrast, subordinate categories share the same parts and differ in other dimensions, whereas superordinate categories differ in their part structure. For example, "hammer" is a basic-level category, easily described in terms of its gross parts, something which is clearly not the case for its superordinate category "tools." And different kinds of hammers, i.e., different

[13] The tension between similarity and contrast within and between categories lies at the basis of beauty: It is variation on a theme (Humphrey 1973).

subordinate categories, are distinguished by more or less subtle differences in the shape or size of certain parts.

These observations suggest that perception and categorization occur in stages: first, a coarse description in terms of parts and with it categorization at the basic level is obtained; second, more details of the shape of parts, their spatial relationships, and other factors beside shape are taken into account to produce a subordinate categorization.[14]

This two-stage model seems to be corroborated by very selective behavioral impairments due to brain damage (Damasio 1985a). The most dramatic of these impairments is no doubt prosopagnosia, the inability to visually recognize faces:

> I was sitting at the table with my father, my brother and his wife. Lunch had been served. Suddenly.... something funny happened: I found myself unable to recognize anyone around me. They looked unfamiliar. I was aware that they were two men and a woman; I could see the different parts of their faces but I could not associate those faces with known persons...Faces had normal features but I could not identify them. They seemed like strangers, people I had never seen before. (Agnetti et al. 1978)

Note that this person could still recognize a face as a "face," in other words, was able to complete the first stage and determine the basic-level category of the visual stimulus. Accordingly, prosopagnosics are able to indicate and name the parts of a face; they can point out the ears, mouth, nose, and eyes (Damasio 1985ab). However, as the above quotation vividly describes, they are not able to individuate faces visually. This is not due to some memory deficit because persons are recognized as soon as they start talking or they can be recognized by individual peculiarities such as moles or clothes (Meadows 1974). The inability to individuate is not limited to the category of faces; prosopagnosia is often accompanied by agnosias for other objects such as cows (Bornstein et al. 1969), chairs, cars, foods, and clothing (Benton 1980; Damasio et al. 1982; Damasio 1985b; Humphreys and Riddoch 1987).

Since individuation requires processing of detailed visual information, one might surmise that prosopagnosia is caused by deficits in perceptual processing, for example, in the processing of high spatial frequencies. This appears not to be the case. Prosopagnosics are not impaired in face discrimination and can perform demanding perceptual tasks (Benton and van Allen 1968; Benton 1980; Gazzaniga and Smylie 1982). In particular, Rizzo et al. (1986) found that spatial contrast sensitivity for vertical sinusoidal gratings ranging from .39 to 6.27 cycles per degree was normal in one and only slightly impaired in a second patient. Conversely, severe deficits in perceptual processing do not cause prosopagnosias or other agnosias (Damasio 1985a; Rizzo et al. 1986).

Thus prosopagnosics, and agnosics in general, appear to have normal perception and memory. What then is the nature of their deficit? In our discussion of the two-stage model of recognition we only paid attention to the stages themselves. If the neural substrates

[14] This account is only meant to serve as a rough heuristic to understanding categorization. Among other things it ignores the fact that different people can see objects quite differently depending on their familiarity with them (Mervis and Greco 1984).

underlying the two stages are localized in different regions of the brain, a deficit will obtain if their connection is broken (Geschwind 1974, 1979; Kean 1985).

Damasio et al. (1982; Damasio 1985b) suggested that the connection between percept and memory takes the form of a template, "a learned record of fundamental measurements of facial features performed with a given method during previous instances of perception." They liken a template to a lock for the multimodal memory store, the key being a visual stimulus. Now, prosopagnosia results if (1) templates have been destroyed, (2) templates are inaccessible, perhaps because the appropriate key cannot be constructed, or (3) activation produced by templates is blocked. Based on the location of brain damage causing prosopagnosia, templates might be stored in the inferomesial visual association cortices of both hemispheres, their limbic structures effecting the consolidation of these templates.[15]

Interestingly, there is evidence that prosopagnosics do react differentially to familiar faces, albeit at an unconscious level. Familiar faces elicit a significantly larger skin conductance response than unfamiliar faces in normal controls (Tranel et al. 1985) and prosopagnosics (Bauer 1984; Tranel and Damasio 1985). In the context of the proposed template system, it seems that the templates are intact, effecting the increased skin conductance response. Thus, we can conclude that, at least in the two patients tested, activation of memory is somehow blocked.

Summarizing, from both normal and impaired visual perception and recognition we have evidence that we categorize visual stimuli, that these categories are organized hierarchically, and that these processes might be performed by neural structures in specific locations in the brain. Although these data, the neurological ones in particular, are quite suggestive we should be equally cautious in their interpretation and allow for alternative interpretations that do not rely on localization of function (Humphreys and Riddoch 1987). The clinical observations are typically anecdotal, and although clinical test data can be quite extensive they have rarely been designed to test specific hypotheses pertaining to cognitive functioning.

Given that we categorize objects, why do we use the categories we do? Categories can be viewed as entities giving (intended to give) useful information about the world. In this view, basic-level categories are those categories that maximize this information and hence are accessed first. Several measures have been proposed for measuring the usefulness of categories; for instance, by combining the probabilities that a feature will be present given the category and that an object will belong to a particular category given a particular feature (Tversky 1977; Jones 1983; Gluck and Corter 1985). Note that these studies assume that we

[15] As will be discussed in more detail in section 4, the visual system consists of a number of subsystems more or less organized hierarchically. Visual association areas (e.g., areas 18, 19) receive their input from primary visual or striate cortex and project themselves to the pole of the temporal cortices which outputs to the hippocampus. The hippocampus projects to other limbic structures and to the visual (and other) association cortices. In other words, the limbic system seems to perform a pivotal role between perception and multimodal memory.

know what constitutes a feature or a part. This, however, is not the case, even for artificial visual stimuli consisting of simple line segments. The experimenter might unconsciously arrange these line segments so as to produce significant emergent features (Pomerantz et al. 1977), T-junctions or symmetries, for example. In addition, as Herrnstein (1984) pointed out in his review of categorization in non-human animals, features might not be simply necessary or sufficient conditions for category membership but contribute incrementally (see also Hampson and Kibler (1983) and Hampson and Volper (1986)), and instead of being fixed, features might be shaped by the categorization process itself. We clearly need a much better understanding of visual perception and shape representation before we can predict how humans categorize their visual environment.

2.5 Face recognition

Within the field of object recognition, the human face has received special attention for a number of reasons (see Ellis (1975); Davies et al. (1981); Ellis et al. (1986) for reviews). It is a task people perform extremely well even though it means discriminating between objects having the same general shape, and it is an ability indispensable for normal social interactions. Recognition of faces seems to differ from recognition of other objects: inverting the image of a face affects its recognition more than inverting the image of a house. There is evidence that the right hemisphere recognizes faces faster than the left hemisphere, and that the inversion effect is specific to the right hemisphere, long thought to be specialized in the processing of visuo-spatial patterns. These observations led some researchers to suggest that there are special face recognition units in the human brain. In this section, we discuss these findings and illustrate some of the issues raised in previous sections, such as the influence of reference frames on perception and recognition, and whether processing is piecemeal through features or more holistic.

Let us first turn to the question of whether faces are recognized in a manner qualitatively different from other objects. Consider the effect of inverting the image of an object: When asked to decide which face from a pair of faces had been seen in the preceding training session, subjects made 1.25 mistakes on average for upright faces and 4.58 for inverted faces (6 would have been chance performance). Houses, however, elicited a much smaller effect; 2.67 and 3.83 mistakes, respectively (Yin 1970). In addition to recognition, perception of facial expression also deteriorates markedly after inversion, as illustrated by "Thatcher's illusion" (Thompson 1982). In this case, an observer is not fully aware of the strangeness of Thatcher's face brought about by inverting all of its features except the eyes and the mouth (Fig.2.6c). Rock (1973, 1974) suggested that inverted faces are difficult to recognize because the mechanism to correct for disoriented stimuli and thereby obtain a canonical description which can be matched against memory, is overtaxed. It is unlikely that this correction

mechanism resembles mental rotation (Shepard and Metzler 1971), since recognition scores for inverted faces and the ability to perform mental rotation are not correlated (Phillips and Rawles 1979).

The inversion effect depends on age (Carey and Diamond 1977; Diamond and Carey 1977; Carey 1981). Before age 10, children recognize upright and inverted faces equally well; thereafter, they are better at recognizing upright faces, just as adults are. Carey and Diamond (1977) interpreted this developmental change in terms of a switch in the encoding of faces. Before age 10, children recognize faces on the basis of distinguishing features, such as hair or glasses. At this age, they are easily confused if (unfamiliar) faces share the same distinguishing features. After age 10, the child develops the ability to use configurational information which is obtained best from upright faces.

Neuropsychological observations provided further evidence for the special status of faces in recognition and for right hemispheric specialization for faces. Compared with normal subjects and subjects with left cerebral lesions, subjects with right posterior lesions perform worse on upright faces, but slightly better on inverted ones (Yin 1970). No such dissociation between right and left cerebral groups was found for houses. The same right hemisphere advantage was found when normal subjects briefly saw two faces (120 and 150 ms for upright and inverted faces, respectively), one in their left and one in their right visual field (LVF and RVF), and had to indicate which two they had just seen from an array of 12 (Leehey et al. 1978). Upright faces were recognized better when shown in the LVF (processed initially in the right hemisphere) as opposed to the RVF. No difference was found for inverted faces. St. John (1981) also found a significant LVF advantage in reaction time (60 ms) for a task requiring subjects to indicate whether two upright faces, both displayed in the LVF or the RVF, were the same; he found no effect for shoes. Summarizing, there is evidence that the right (posterior) hemisphere plays a special role in the recognition of upright faces; in other words, there is a face recognition system in the right hemisphere that is orientation dependent in addition to an orientation-independent system present in both hemispheres (Kean 1985).

However, a number of problems make these conclusions less than compelling. Sergent and Bindra (1981), for example, argued that the experimental conditions used in tachistoscopic studies are biased in favor of right hemispheric processing: short exposure times to prevent eye movements, peripheral exposure, and highly discriminable images to obtain acceptable error rates. Ellis (1981) noted that the use of artificial stimuli (Bradshaw and Wallace 1971; Davies et al. 1977; Naveh-Benjamin 1982) in which a small number of features is systematically varied, encourages subjects to perform a feature by feature comparison. And, faces are not the only stimuli that exhibit an inversion effect. Letters and words are also hard to recognize upside down (Kolers and Perkins 1969; Rock 1973). In the following, we discuss these problems in some detail.

Ellis et al. (1979) showed that familiar and unfamiliar faces are recognized on the basis of different parts of a face. For famous people, an advantage for internal features was found, while internal and external features proved equally useful for unfamiliar people. Young and Bion (1981) found an inversion effect for familiar people (classmates and colleagues) at ages 7, 11 and beyond. When adults did not know which face to expect there was no effect; but when they were given a list of faces to be used there was a strong effect. In other words, the stimulus material and the task can influence the outcome of these experiments significantly.

Since we are very familiar with upright faces, it is not too surprising that we have difficulty with inverted faces (Ellis 1975). Similarly, whites grown up in a predominantly white environment have difficulty in recognizing faces of orientals, but improve their performance with practice (Elliot et al. 1973). In addition, practice improved discrimination between inverted faces more than discrimination between upright faces (Bradshaw and Wallace 1971). A similar effect was found for photographic negatives which are initially hard to discriminate (Galper 1970). Diamond and Carey (1986) showed that faces are not the only stimuli sensitive to inversion: dog experts showed a comparable decrease in recognition with inverted faces and inverted dogs, whereas novices only showed the inversion effect for faces. They concluded that experts represent visual stimuli differently from novices, namely in terms of "second-order relational" properties rather than distinguishing features. Second-order relational properties, e.g., narrow chin, are used by experts to individuate members of the same class, that is, objects that share the same first-order configuration (e.g., eyes above nose).

Young and Bion (1980) tested children at ages 7, 10 and 13, for their ability to recognize upright and inverted faces. At these three stages in development a right hemisphere advantage was found for upright but not for inverted faces if only four faces were used. With 40 faces, only 13 year old boys showed the effect. In addition to replicating earlier findings of a right hemisphere superiority in processing upright faces, these results show no evidence of a developmental change and once again point to the influence of task difficulty. Flin (1985) further investigated the possibility of a floor effect in which low performance due to the difficulty of the task would mask any potential effect of inversion. Using 10 faces during learning and 10 more as distractors during testing, she found an inversion effect from age 7 to 16. Recognition of upright faces improved markedly with age, whereas recognition of inverted faces remained the same from age 7 to 12, and improved at 16 years. Flin also suspected that a floor effect caused the paraphernalia findings: Children had to discriminate between very similar faces and were thus easily fooled by hats and other paraphernalia. In addition, it is known that high similarity between targets and distractors disrupts adult recognition (Shepherd et al. 1981) To test for this, Flin used similar and dissimilar faces. Four-year olds discriminate dissimilar faces significantly better (40% and 70% errors) than six-year olds (20% and 70%) and 8 year olds (10% and 40%). Thus children rely on paraphernalia when faces are similar and are easily confused by them, but they can use facial information when

the task is made easier, i.e., when dissimilar faces are used.

Summarizing, there is at present no conclusive evidence that faces hold a unique position among the objects that people can recognize. Recognizing inverted words is just as difficult as recognizing inverted faces. In fact, Bruce (1983) proposed a model of face recognition that is adapted from models of word recognition (see next section). However, there is evidence that the right hemisphere has an advantage for encoding visual stimuli, including faces (Levy et al. 1972; Marzi and Berlucchi 1977; Young et al. 1985; Delis et al. 1986).

2.6 A model of object recognition

Our understanding of visual perception and recognition is still very rudimentary and largely confined to functionality. We know something about what is done, but very little about how this accomplished. This is true for almost every level: we can demonstrate that the assignment of figure and ground is important and has to be accomplished, we can demonstrate that inverting faces and words deteriorates their perception and recognition, we know that we can recognize persons from their faces, and that recognition does not necessarily mean knowing a person's name. In other words, we are in the midst of exploring what the visual system can do and how these capabilities are linked. I thus think it appropriate to conclude this section with a brief description of a functional model of human visual recognition, a model about which most researchers seem to agree.

Probably the most important clues about the organization of the human brain come from persons who suffered localized brain damage (Geschwind 1974; Damasio 1985a). Warrington and Taylor (1973) observed that patients with right posterior lesions have difficulty in deciding whether two different views belong to the same object or not, especially with so-called unconventional views of objects. These patients also have problems in naming or identifying objects from unconventional angles: 75% versus 92% correct for right posterior and left anterior groups (Warrington 1982). Similar difficulties arise when objects in a scene are shaded (Benton and Van Allen 1968; Warrington 1982). Conversely, patients with left hemisphere lesions find it difficult to recognize objects even though they can decide whether or not different views are from the same object (Warrington 1975).

On the basis of these data, Warrington and Taylor (1978; Warrington 1982) proposed that visual object recognition consists of two postsensory stages in series (Fig.2.11). In the first stage, percepts are categorized on perceptual grounds, possibly through generalization across viewing position, lighting, etc. Perceptual categorization thus leads to object constancy.[16] After percepts have been assigned an object category, semantic knowledge

[16] The question remaining is: How? Although unconventional views typically show the object in question significantly foreshortened, Warrington and James (1986) found no systematic relationship between the degree of foreshortening and recognition in people with right hemisphere lesions or controls, nor did they

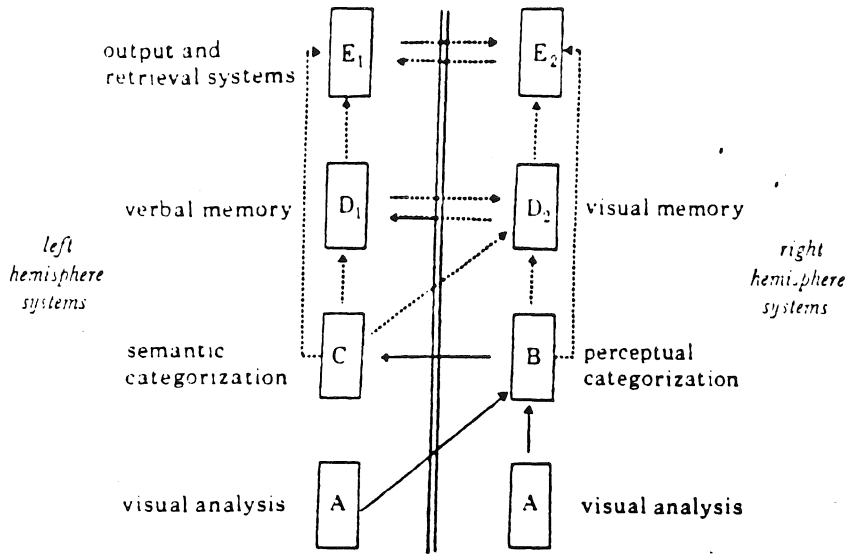


Figure 2.11 **Model for the stages of object recognition (Warrington and Taylor 1978).**

becomes accessible, perhaps by using perceptual categories as indices into semantic memory.

The data also support the idea that perceptual and semantic categorization occur in different regions of the brain or at least have some degree of lateralization. Perceptual categorization is lateralized in the right posterior cortex, semantic categorization in the left posterior cortex (Fig.2.11). Hay and Young (1982) cite studies suggesting that early visual processing does not have any hemispheric preference, but that subsequent processing such as that involved in matching photographs from different angles shows a right hemisphere reaction time advantage (see also Poizner et al. 1984ab).

A number of researchers in face processing have proposed very detailed functional models of face recognition (Ratcliff and Newcombe 1982; Hay and Young 1982; Bruce 1983; Bruce and Young 1986). Evidence for the different stages in these models derives from neurological data, priming effects, and behavioral data from everyday recognition (Warren and Morton 1982; Young et al. 1986abc; Bruce and Valentine 1986). In Fig.2.12, a functional model for face recognition is illustrated; it consists of eight modules, most of which also apply to the recognition of objects besides faces. The second stage in the structural encoding module, producing an "expression-independent description," corresponds to Warrington's perceptual categorization. Subsequent stages effect semantic categorization. Thus the face recognition units, patterned after the logogen model for word recognition (Morton 1969), somehow recognize a particular face providing access to the associated person-specific semantic information. Note that these face recognition units resemble the face templates posited by Damasio et al. (1982, 1985ab) in that they merely provide access to information associated with a particular person and do not themselves contain any of that information. Thus they

find qualitative differences between normal controls and these patients.

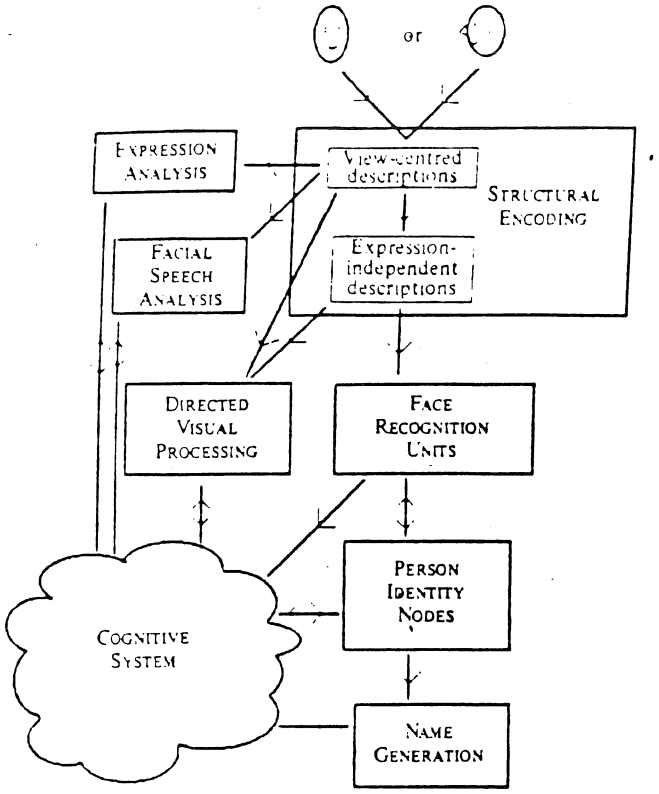


Figure 2.12 A functional model for face recognition (Bruce and Young 1986).

function strictly as indices into memory. In addition to possibly triggering recognition units, the structural code of a face also triggers visual processes that analyze facial expression, sex, age, etc.

3. Computational aspects

In this section, we examine formal theories of visual recognition, i.e., theories that state explicitly what information is used and how that information is represented in order to recognize an object from its visual image(s). Since we restrict ourselves to shape information, these theories specify how shape is represented, and how descriptions of shape can be derived from visual images. Even if they are specified formally as computer programs, these solutions to the problem of object recognition do not necessarily qualify as computational theories in the sense of Marr (1982). This would require, in addition, a grounding in the physical world, which would afford an understanding of why the programs work. For example, in their work on stereopsis, Marr and Poggio (1976) start by making assumptions about the world—that any given point on an object's surface has a unique position in space and time, and that matter is cohesive and surfaces sr oth almost everywhere; this theory then allowed them to derive a cooperative algorithm to match the images from the two eyes and hence compute disparity.

We will start with theories restricted to 2-D objects such as letters, where the 2-D pattern in the image is the object to be recognized. Obviously, the ability to recognize 2-D objects does not suffice for the recognition of 3-D objects whose images can change dramatically under changes in viewing position. Two approaches to the problem of recognizing 3-D objects immediately come to mind. One is to simply collect a number of views and to let this collection be the description of the object's shape. The other is to reconstruct the 3-D shape of the object, say by means of 3-D primitives such as cylinders or spheres.

3.1 Two-dimensional objects

2-D object recognition is commonly formulated as the ability to recognize 2-D objects or patterns consisting of a number of curves and possibly a figure-ground assignment. Usually, objects are to be recognized in all possible positions (translation and rotation invariance) and from all reasonable distances (size invariance). Note that theories striving for rotation invariance cannot, by definition, account for the difficulty humans have in recognizing inverted letters and faces (section 2.2). In the following we assume that the objects have already been separated from the background, i.e., that we know what constitutes the object.

I found it useful to divide the techniques for representing 2-D objects into two classes, one employing shape tem_ ates and the other shape features. The distinguishing property of shape templates is that they treat each part of an object equally: no distinction is made between significant and insignificant parts; indeed, the very notion of part is meaningless. The idea is to find a description or approximation that will afford a categorization into discrete classes. The alternative is to make such distinctions and to view an object as consisting

of features or parts which are distinguishable locally. In so doing, one explicitly assigns structure to an object. Often shape templates derive global descriptors of shape; Fourier descriptors, for instance, indicate to what extent a sinusoidal wave with some frequency is present in the curve as a whole.[1] In general, however, we cannot equate templates with global measures of shape.

As so often, the strength of an approach is at once its weakness. Shape templates avoid the difficulty of defining the features or parts of an object, but, consequently, cannot use "significant" features in situations where objects are partly occluded, miss certain parts, or are deformed. Shape features promise to be more flexible in these situations, a promise that can only be fulfilled, however, if the right set of features can be found. In practice, recognition systems are often somewhere in between these two extremes; for example, by using subtemplates to represent those parts of an object that distinguish it from all others in the data base (this could be called the pragmatic approach to feature definition).

3.1.1 Shape templates

One straightforward way to represent the shape of a curve is to approximate it by a list of line segments that are locally tangent to the curve (Freeman 1974; Ballard and Brown 1982; Scholten and Wilson 1983). Since this list is ordered it is called the **chain code** of the curve (Fig.3.1a).[2] Usually, the line segments are restricted to four or eight different directions, so that each segment can be encoded by a 2- or 3-bit number, and the curve by a string of 2- or 3-bit numbers. A problem with chain codes, especially when used for closed curves, is their dependence on starting position: A different starting position results in a different chain code. We can circumvent this problem by defining a canonical starting position, for example, by choosing the starting position that results in a chain code whose integer value is minimal (interpreting the string of bits as the representation of an integer). This value is referred to as the shape number. Clearly, the shape number does not change if we translate the curve. Although the shape number changes under rotation, we can easily obtain rotational invariance by taking the derivative of the chain code, obtaining a sequence of numbers indicating how much the directions of successive line segments differ, in other words, indicating local curvature. Of course, curvature uniquely specifies the shape of the original curve (Lipschutz 1969).

To compare two curves and determine their similarity, we can compare their chain codes.[3]

---

[1] One could interpret Fourier descriptors as features, indicating the presence or absence of a particular wave length. I chose not to, instead reserving the term feature for a spatially distinct part of an object, as in the "features of a face."

[2] Sometimes referred to as Freeman code.

[3] Tsai and Yu (1985) developed a string matching technique where each symbol in the string represents

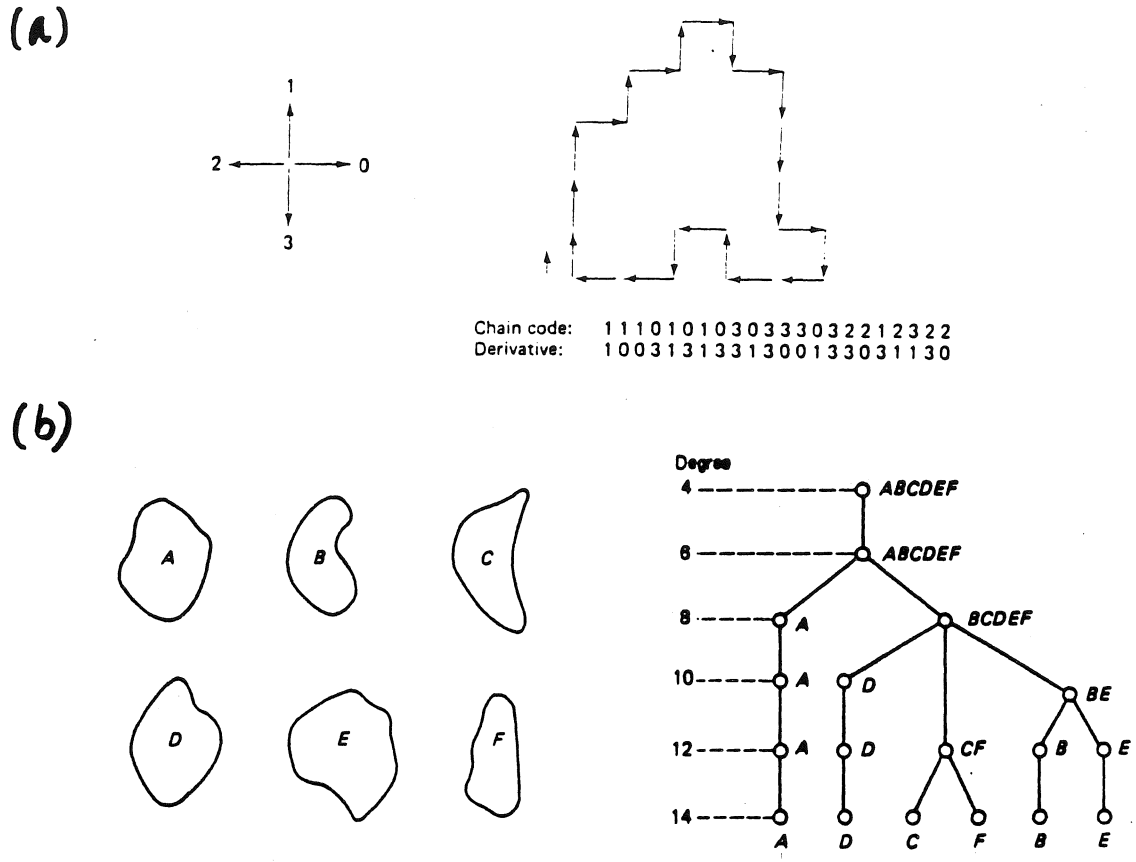


Figure 3.1 (a) Chain code and its derivative; (b) six contours and their similarity tree indicating the degree of similarity (Ballard and Brown 1982).

This allows us to decide whether two curves are exactly the same, but not to measure their similarity quantitatively. The latter can be accomplished by digitizing curves at different resolutions and storing their chain codes as a function of resolution. Comparing two curves then involves comparing their hierarchies of chain codes. The higher the resolution at which their chain codes are the same, the higher the degree of similarity of the curves.[4] As an example, consider the six closed curves shown in Fig.3.1b. The degree of similarity between object C and F is 12, and between C and B it is 8. Human observers would probably judge B and C to be the most similar, not B and E, partly because E is oriented differently from

---

a line segment having a particular length and direction. Interestingly, their procedure is flexible enough to match distorted shapes.

[4] After observing that detection of sinusoidal gratings is largely independent of size but instead depends on the number of periods in the grating, Koenderink and van Doorn (1978, 1982) developed an interesting model for the size distribution of retinal ganglion cells. They assumed that the receptive fields of the ganglion cells differ in size s only, and that the density of ganglion cells varies as $1/s^3$. It then follows that the number of ganglion cells that fire in response to a figure depends on the number of cycles or periods displayed, that is, neural activity level is independent of the size of the display. Burton et al. (1986) implemented this model, and proposed that the elongated receptive fields of cortical neurons are not for detecting oriented line segments but for affording distortion invariance.

B and C. If it were rotated counterclockwise by 120° it would appear more similar to both B and C (see section 2.2 for the influence of orientation on shape perception).

So far we have only defined chain codes for closed curves, an obvious limitation. By combining chain codes with so-called plex grammars (Feder 1971), which specify spatial relationships among shape primitives such as lines, complex patterns can be described. The use of shape grammars is known as the syntactic or structural approach to shape representation (Miller and Shaw 1968; Fu 1982; Lin and Fu 1984, 1986; Shapiro 1985; Bunke and Sanfeliu 1986). We will discuss shape grammars in more detail in section 3.3.1.

A representation reminiscent of the hierarchy of chain codes is the **strip tree** which approximates a curve by so-called strips, rectangles of varying shape and size (Ballard 1981a; Ballard and Brown 1982). At the coarsest level of description, we have one strip simply covering the entire curve; while at the finest level, we have something resembling a chain code (thinking of a line segment as a strip with zero width). A strip tree effectively combines these different levels of resolution: As we go from the root of the tree to its leaves, the representation of the curve becomes progressively more accurate. Using strip trees, we can design efficient algorithms[5] to determine whether two curves intersect, or whether a point lies within a closed curve. Strip trees are related to quad trees (Ballard and Brown 1982), which also represent spatial occupancy hierarchically, but do so by approximating the interior of a curve instead of the curve itself. Using prisms instead of strips, Faugeras and Ponce (1983) generalized this approach to three dimensions.

Instead of approximating a curve with increasing accuracy, one could start with an exact description of a curve, for example by specifying curvature as a function of arc length, then expand this function into a series and retain only the low order coefficients. A well-known example is the Fourier series which describes an arbitrary bounded function $f(t)$ of period 2π as an infinite sum of cosine and sine waves of varying frequency and amplitude:

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nt + b_n \sin nt) = \sum_{n=-\infty}^{\infty} c_n e^{int}, \tag{1}$$

where

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}dt. \tag{2}$$

The set of Fourier coefficients or **Fourier descriptors (FD)** $\{c_n\}$ make up the so-called Fourier spectrum. This technique can be extended to functions of two variables, in particular, $I(x, y)$, the image intensity as a function of spatial coordinates x and y.

[5] Time complexity is O(log n), where n is the number of points describing the curve; this means that the number of computations has an upper bound proportional to log n.

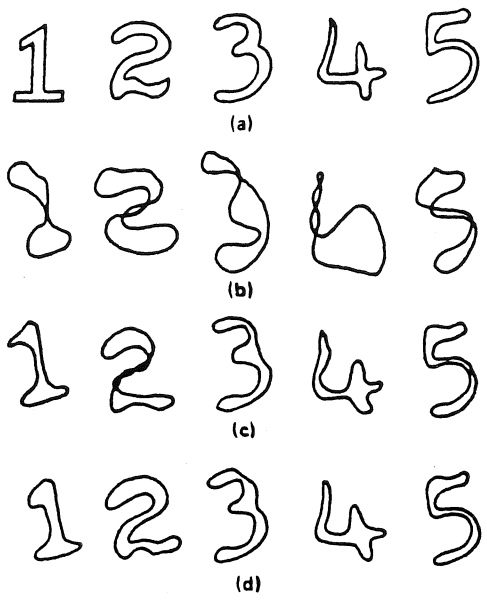


Figure 3.2 **Fourier approximations of the boundary orientation function of some numerals. (a) original; (b) five FDs; (c) 10 FDs; and (d) 15 FDs (Duda and Hart 1973).**

Fourier descriptors can be used in number of ways to describe 2-D objects. One possibility is to compute the two-dimensional Fourier series directly from the image of the object. This is unattractive since the object of interest is usually surrounded or even partially covered by other objects, which of course influences the FDs obtained. One could isolate the object of interest from its surroundings and approximate it by a 2-D Fourier series (Gardenier et al., 1986). Alternatively, one can parametrize the shape of the contour and compute the FDs of the parametrization. This can be done in several ways. One way is to trace the contour and retain the (x, y) coordinates of equidistant points. This generates a complex function $\gamma(l) = x(l) + iy(l)$, with l being arc length. The period of γ is L, the length of the contour (Richard and Hemami 1974; Persoon and Fu 1977; Wallace and Wintz 1980). Wang et al. (1984) used the distance from points on the contour to the center of mass of the figure. Another way is to measure the orientation of the tangent at equidistant points, and generate a so-called boundary orientation function (Persoon and Fu 1977).[6] Fig.3.2 illustrates that lower order Fourier descriptors suffice to reasonably specify the shape of a curve. Note that the lower order FDs capture the global aspects of shape, and the higher order ones the more local variations. A function related to the orientation function is the slope density function, which specifies the reciprocal of the contour's curvature as a function of arc length (Nahin

[6] As discussed in detail in section 4.2.3, Schwartz et al. (1983) found that some neurons of inferior temporal cortex—an area of the cortex involved in object recognition—are selectively tuned to stimuli whose boundary orientation function is composed entirely of the first, second, etc., Fourier coefficients.

1974). The FDs of these different functions can be normalized for position, orientation, scale, and for closed curves, starting position.[7]

FDs have been used to recognize letters, airplane silhouettes, machine parts, etc. To take one specific example, Persoon and Fu (1977) experimented with the so-called Munson numerals "0" to "9," a collection of numbers written by 49 people. To obtain closed curves from the numerals, their outer boundaries were traced and used to compute $\gamma(l)$. After training the system on one third of the available numerals and retaining 8 FDs for each numeral, identification of the remaining two thirds was from 85 to 90% correct, depending on the distance measure used to classify shapes. Surprisingly, the numeral '8' is recognized only in about 55% of the cases, it being taken for a "1" in 20% of the cases. This happens, according to the authors, because the outer boundary of an "8" is similar to that of a "1." Apparently the converse is not true, since a "1" is never classified as an "8" (in fact, "1"s are correctly classified all the time). Thus, although overall performance is reasonable, the system fails to discriminate what seem to be very different shapes. The use of outer boundaries again points to the problem of having to use closed curves, or, at least, curves that do not intersect.

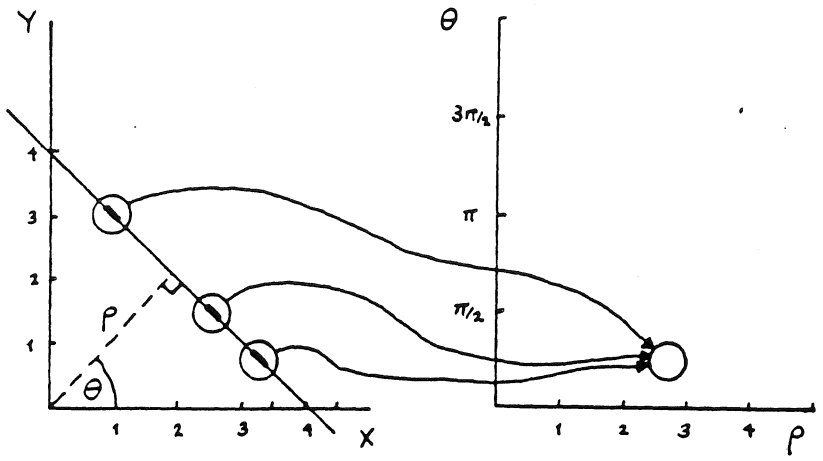
A method originally designed to find lines in noisy images has now been generalized to recognize arbitrary shapes. In the early 1960s, Hough devised a method, now known as the **Hough transform**, to find particle tracks in noisy bubble chamber photographs. His idea was to collect evidence for straight lines in a so-called feature space, which in this case represents all possible lines in the image, for example by their distance from the origin and their slope (Fig.3.3a). Each line segment in the image "votes" for the one point in feature space that represents its slope and location.[8] If, after processing all line segments in the image, some point in feature space has collected more evidence than some threshold value, we conclude that the corresponding line is present in the image.[9] Kushnir et al. (1985) used the Hough transform to represent Hebrew characters printed by hand; that is, the Hough transform was treated as a feature vector, each value in feature space indicating the degree to

[7] Sheng and Arsenault (1986; Sheng and Duvernoy 1986) used circular-Fourier-radial-Mellin descriptors to obtain descriptors of a pattern that are invariant under translation, orientation and scale, that is, descriptors that do not need normalization. Circular Fourier descriptors are based on polar instead of cartesian coordinates, automatically affording rotation invariance. Thus the image is considered to be composed of radial waves. Size invariance results from taking the radial Mellin transform of the circular Fourier descriptors. For a related optical implementation see Casasent and Psaltis (1976).

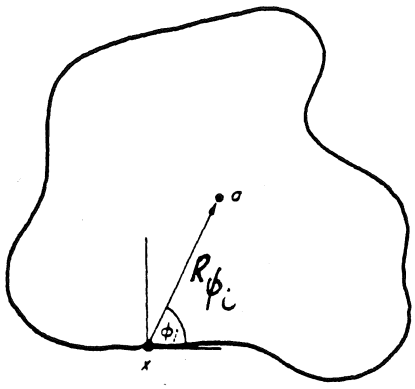
[8] Note that these line segments might come from different "parts" of an object. Strictly speaking, then, points in Hough transform space do not represent features.

[9] Barlow (1986) noted that the brain is faced with a similar problem, namely to bring together related information from widely spaced locations in the retinal image, for example the color of a large object. Since cortical connections are quite local, extending at most 4 mm in either direction (Gilbert 1985), direct communication between neurons representing faraway locations in the cortex is ruled out. This led Barlow to suggest "cortical feature maps" analogous to Hough transforms as a way of associating related information.

(a)



(b)



R-table format

i	ϕ_i	R_{ϕ_i}
0	0	$\{\mathbf{r} \mid \mathbf{a} - \mathbf{r} = \mathbf{x},\ \mathbf{x}$ in $B,\ \phi(\mathbf{x}) = 0\}$
1	$\Delta\phi$	$\{\mathbf{r} \mid \mathbf{a} - \mathbf{r} = \mathbf{x},\ \mathbf{x}$ in $B,\ \phi(\mathbf{x}) = \Delta\phi\}$
2	$2\,\Delta\phi$	$\{\mathbf{r} \mid \mathbf{a} - \mathbf{r} = \mathbf{x},\ \mathbf{x}$ in $B,\ \phi(\mathbf{x}) = 2\,\Delta\phi$
...	...	...

Figure 3.3 (a) Hough transform for lines (Ballard 1984); and (b) R-table for arbitrary planar shapes (Ballard 1981b).

which that feature (line of certain orientation and location) is present in the character. This resulted in an efficient description of the characters, circumventing the problem of defining explicitly what the difference between characters is in terms of features like "a vertical line with a horizontal one at its base pointing to the right." On average, the system recognized 86.9% of the characters, slightly below the human level of 92.4% (characters are shown separately, without the benefit of context).

Ballard generalized the Hough transform to detect arbitrary 2-D objects in an image (1981b; Ballard et al. 1983). The shape of an object is represented by a so-called R-table specifying the location of line segments as a function of their orientation. An R-table is constructed as follows. First, select a reference point, a, in the image (Fig.3.3b). Then, trace the boundary of the 2-D object, and store the difference vector between each boundary point and a as a function of the local gradient direction ϕ (a direction orthogonal to the local

tangent). Note that there might be multiple entries under the same direction ϕ. The feature space is now a set of R-tables representing the objects to be recognized. To find a shape S represented by an R-table R_S we proceed as follows. Let $A(a_x, a_y)$ be a two-dimensional accumulator array initialized to 0. For each edge element $\mathbf{x} = (x, y)$ in the image, compute its gradient direction ϕ_i and use it as an index into R_S. Increment all points $\mathbf{x} + R_S(\phi_i)$ in A. As before, maxima in A indicate the locations of objects with shape S. To account for scale and orientation, two more dimensions can be added to the accumulator array, explicitly representing the object at all scales and orientations.[10] The R-tables of shapes S_i can be combined very easily to describe the resulting composite shape. Letting $\mathbf{y}_i$ be the reference points of S_i, $\mathbf{a}$ the reference point for the new shape S, and $\mathbf{r}_i = \mathbf{a} - \mathbf{a}_i$, then

$$R_S(\phi) = \bigcup_i (R_{S_i}(\phi) + \mathbf{r}_i). \tag{3}$$

Instead of using the orientation of line segments to index the R-table, Turney et al. (1985) used distinguishing features of a curve. Distinguishing features are simply those segments of the curve that distinguish it from the other curves in the data base. The saliency or degree to which a feature determines the identity of a figure can be used to weigh features differentially. The importance of the ability to weigh features was demonstrated by comparing the performance of a system using R-tables based on line segments with that of one based on weighed features. Only the latter was able to correctly locate a partially occluded object.

Neveu et al. (1986) also used R-tables to represent the shape of 2-D objects. In addition, they described certain parts of the object at higher levels of resolution (upto four levels). They did not, however, clearly define what constitutes a part and what does not. That was left to the user who could interactively select regions of the figure for representation at a higher resolution. This hints at a problem of FDs, R-tables and the like, namely the lack of correspondence between components of these representations and identifiable components of the object. Removing part of an object possibly affects all FDs or all entries in the R-table, in a manner that is not obvious (Wang et al. 1984; Pinker 1984; Hoffman and Richards 1984). And although one can easily combine the R-tables of objects to obtain a composite object, the reverse is not true: Given the R-table of the composite object one would be hard put to find its components. A related problem is the insensitivity to small variations, an insensitivity which is not surprising since each Fourier descriptor or point in feature space of the Hough transform indicates the degree to which a particular frequency or line is present in the entire image of the object. Small but significant variations will be lost in the noise.

[10] This can be extended to 3-D objects by assigning one dimension to every degree of freedom (Ballard and Sabbah 1983; Silberberg et al. 1984, 1986).

Thus we are led to explore representations that *explicitly* assign "structure" to an object, for example, by decomposing it into convex parts and specifying the spatial relationships among these parts.

3.1.2 Shape features

To recapitulate, Selfridge and Neisser (1960) noted "that letter patterns cannot be described merely as shapes. It appears that they can be specified only in terms of a preponderance of certain *features*. Thus A tends to be thinner at the top than at the bottom; it is roughly concave at the bottom; it usually has two main strokes more vertical than horizontal, one more horizontal than vertical and so on" [emphasis in original].

But what constitutes a feature? In their letter-recognition program PANDEMONIUM, Selfridge and Neisser (1960) used an *ad hoc* procedure in which the programmer adds as many features as possible, since "there is probably safety in numbers. The designer will do well to include all the features he can think of that might plausibly be useful." Recognizing the shortcomings of their approach and the fundamental nature of the problem they were trying to solve, they concluded that

> **No current program can generate test features of its own. The effectiveness of all of them is forever restricted by the ingenuity or arbitrariness of their programmers. We can barely guess how this restriction might be overcome. Until it is, "artificial intelligence" will remain tainted with artifice.**

And indeed, much work in cognitive psychology, artificial intelligence, taxonomy, and pattern recognition has been directed towards this goal. Here we will briefly discuss what computer vision researchers have learned (see Pavlidis (1980) and Shapiro (1985) for reviews).

It is clear that certain segments of a silhouette carry more information than others. Attneave (1954) demonstrated this by approximating the outline of a sleeping cat with as few as 40 line segments connecting the points of maximum curvature. When asked to segment a silhouette, humans do so at points with highest curvature, a not entirely surprising result (Attneave 1954; Hoffman 1983; Fischler and Bolles 1986). And when considering say a maple leaf, humans divide it into regions, the lobes of the leaf. These intuitions about the way humans appreciate shape led computer vision researchers to explore algorithms that decompose shapes. This work on decomposition can roughly be divided in two classes: in the first, an object is divided into *regions* that satisfy some property, e.g., convexity; in the second, the *outline* of an object is partitioned on the basis of extrema or zeros of curvature, a partitioning which could induce a partitioning of the object itself.

Shapiro and Haralick (1979) sought to partition a figure into clusters, where a cluster is a highly connected part of a figure (points on the perimeter of the figure are connected if they can be connected by a straight line lying completely within the interior of the figure),

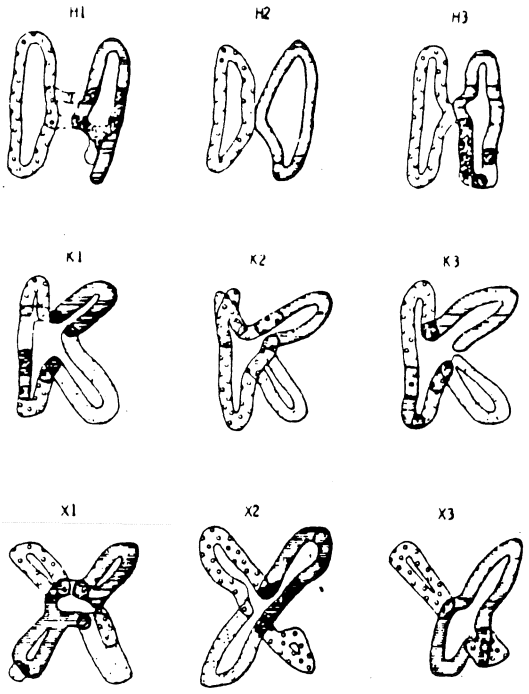


Figure 3.4 **Decomposition of handwritten characters into convex parts (Shapiro and Haralick 1979).**

i.e., maximally convex parts or protrusions.[11] The complement of protrusions, intrusions or concavities, can be found by using lines lying completely outside the shape. We can now decompose the shape into its (almost) convex parts. Note that this decomposition is invariant under translation, rotation, scaling, and skewing. Also note that upon figure-ground reversal, the decomposition of an object changes drastically since concave regions become convex, and vice versa. This decomposition algorithm was tested on the handwritten characters "H," "K," and "X". As can be seen in Fig.3.4, decomposition was quite variable among instances of the same letter. An interesting property of this clustering method is that the resulting clusters or parts are not necessarily disjoint: the horizontal bar in H1 overlaps with the right vertical. A slightly different clustering algorithm developed by Guerra and Pieroni (1982) produced disjoint parts.

To specify the topological relations between parts of the shape, Shapiro (1980) defined two spatial relations, the protrusion and intrusion relation. The protrusion relation (i_1, s, i_2) indicates that the two concavities i_1 and i_2 bound the protrusion or "simple" part s. The intrusion relation does the same, *mutatis mutandis*. If a shape memory consisting of relational descriptions is to be employed, one is naturally interested in quantifying the similarity between two relational descriptions. For this purpose, Shapiro and Haralick (1985) defined a metric on the space of relational descriptions. Given a relational description representative

[11] Maximally convex parts allow some points of a cluster not to be connected.

of some class of objects (prototype), a metric makes it possible to compute the probability that a particular deviation from this prototype would occur. These conditional probabilities can then be used in Bayesian classification schemes (Duda and Hart 1973).

Van der Heydt et al. (1981) developed a related method for decomposing an object into greatest nearly convex regions and used it to separate overlapping chromosomes in metaphase pictures (succeeded in 77 of 88 cases). Bjorklund and Pavlidis (1981) derived a graph from a 2-D object by approximating its constituent curves by line segments. These line segments form the vertices of a graph whose edges specify the appropriate spatial relationships. These relationships might indicate that the two line segments form a convex or a concave corner. The graph is used to decompose the original shape, for example, by searching for cycles arising from nearly convex components. One could also augment the vertices of the graph by associating with each a probability vector indicating the probability of that segment belonging to certain parts of the object. A globally consistent partitioning is then obtained by relaxation (Rutkowski et al. 1981).[12]

In the **symmetric** or **medial axis transform** (Blum 1974; Blum and Nagel 1978), 2-D objects are approximated by a collection of maximal discs, i.e., discs inside the object and tangent to its boundary that are not contained in any other discs (Fig.3.5). The 2-D shape is the union of all such discs, and the symmetric axis is the union of the discs' centers.[13] The symmetric axis can be thought of as the "skeleton" of a shape with "joints" and "branch points," features that can be used to partition the shape. The resulting parts can be further decomposed or simply classified on the basis of their shape, for example, by noting how their boundary curvature with respect to the symmetric axis. The result is a structural description of a 2-D shape. Fairfield (1983) used a different algorithm to decompose the shape. His algorithm does not differentiate between branch points and boundary curvature but only considers boundary curvature. And instead of local curvature, it uses a cumulative measure of orientation change. This measure, called internal concavity, is the net change in spread angle along some part of the symmetric axis; spread angle being the angle formed by the two half lines emanating from a point on the axis and going to the points on the boundary closest to it. Thus the internal concavity of a cylinder is zero since the spread angle is everywhere 180°. Because it is based on angles, internal concavity is scale invariant.

---

[12] Relaxation labeling or cooperative processing are widely used in computer vision (Ballard et al. 1983). It is a technique for satisfying a global constraint by purely local computations. A well-known example is Waltz's (1975) algorithm for interpreting line drawings: In general, each line junction can be interpreted in a number of ways; that is, there is local ambiguity. However, by requiring that neighboring interpretations be consistent, one can arrive at a globally consistent interpretation of the line drawing (see also Rosenfeld et al. 1976; Zucker et al. 1978; Davis and Rosenfeld 1981; Hummel and Zucker 1983). Marr and Poggio (1976) designed a cooperative algorithm to compute disparity from two stereo images.

[13] Lee (1982) developed an efficient, O(log n), algorithm to compute the medial axis of a polygon with n edges.

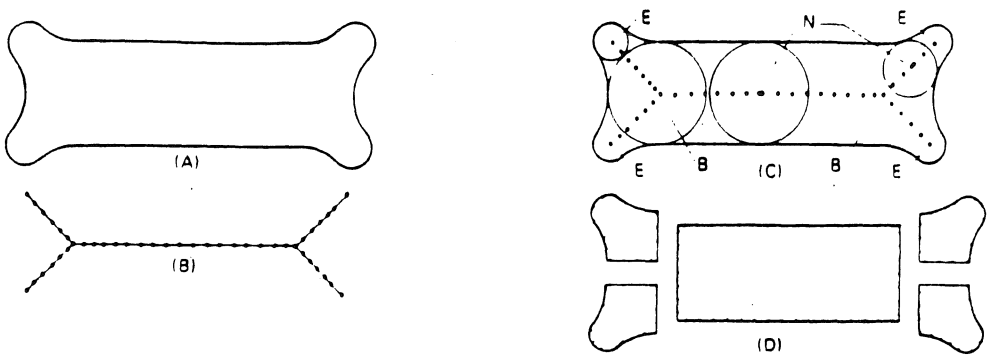


Figure 3.5 Symmetric axis transform. (A) 2-D bone-like shape; (B,C) its symmetric axis; (D) object partitioning at branch points (Blum and Nagel 1978).

This makes the decomposition less sensitive to boundary noise, an oft noted problem of the medial axis transform. Another approach to handling noise is to compute the symmetric axis at different scales (Pizer et al. 1987).

Nackman (1983; Nackman and Pizer 1985) extended the symmetric axis transform to include 3-D objects by replacing discs with spheres and symmetric axes with symmetric surfaces. In related work, Mohr and Bajcsy (1983) used tangential instead of overlapping spheres to approximate 3-D objects.

Just as the various clustering methods, the symmetric axis transform first describes a shape, and then uses the description to decompose the shape. Alternatively, one could first try to decompose the (outline of the) object and then describe the shape of the resulting parts. But what shape properties should one use to decompose an object? Hoffman and Richards (1982; Hoffman 1983; Richards and Hoffman 1985) chose concavities as part boundaries, a choice motivated by the fact that whenever one conjoins two 3-D objects one produces a concavity all along the contour of intersection (except for the nongeneric case in which the objects are partially tangential). This is illustrated in Fig.3.6a. The projection of the contour of intersection yields a concavity in the image. Thus the choice of concave regions, or more generally minima of curvature, as part boundaries is founded on a property or regularity of the physical world (see Brown (1984) for an overview of the various regularities or real world constraints used in computer vision). Dividing the contour at (negative and positive) minima of curvature results in segments that fall into six categories depending on the number of inflection points they contain.[14] Fig.3.6b shows the six contour "codon" types (Richards and Hoffman 1985; Richards et al. 1985; Leyton 1986). Thus, codon 2 indicates the presence of two inflection points in between two part boundaries. Note that a codon description is invariant under translation, rotation and scaling, but changes drastically upon figure–ground

[14] An inflection point has zero curvature, and is thus the transition point between concave and convex parts of the contour.

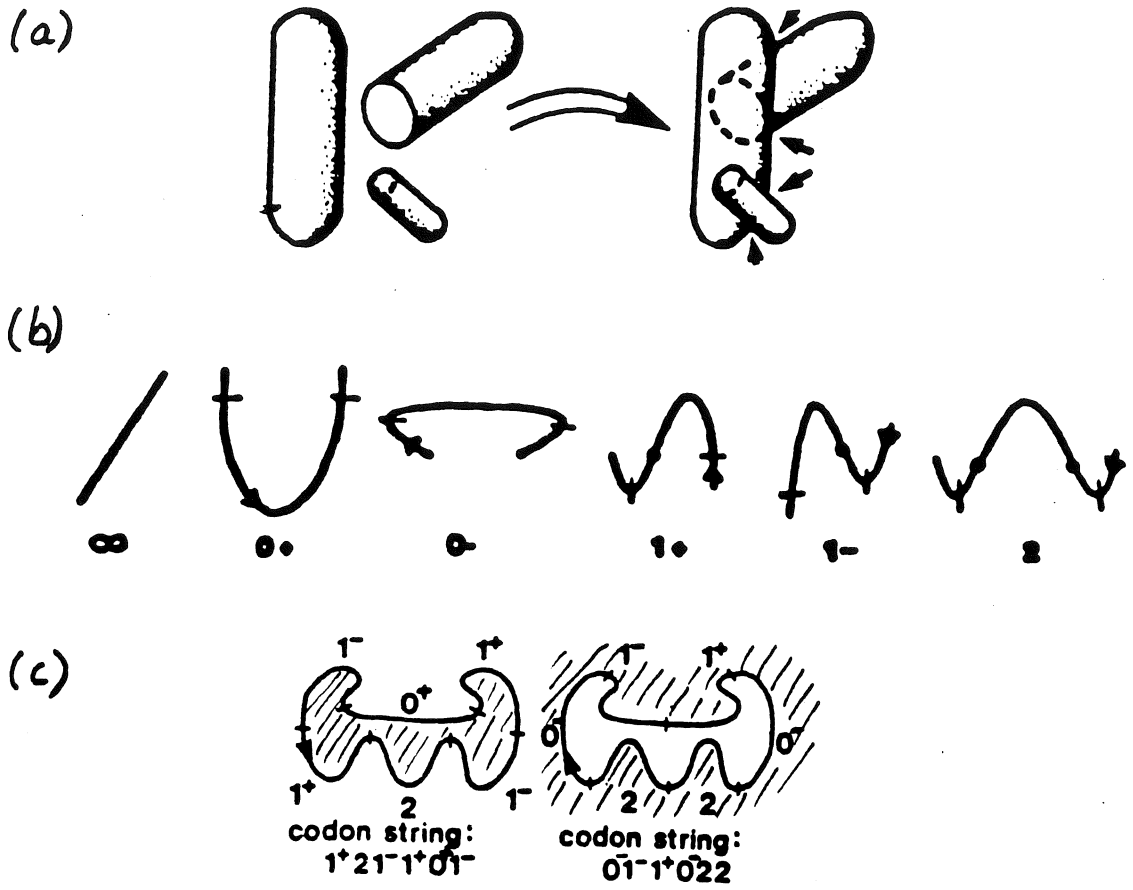


Figure 3.6 (a) Joining of objects gives rise to concavities in the silhouette. (b) The six codon types: ∞, straight line segment; 0^+, no inflections, concave; 0^-, no inflections, convex; 1^+, maximum followed by inflection; 1^-, inflection followed by maximum; and 2, maximum flanked by two inflections. (c) Changing the figure–ground assignment changes the appearance of a curve and its codon string (Hoffman and Richards 1982; Richards and Hoffman 1985).

reversal (Fig.3.6c).

Summarizing, we reviewed a number of representations for 2-D objects, in particular curves and silhouettes. We distinguished between representations that merely describe shape and representations that in addition assign structure to objects. Examples of the former are Fourier descriptors whether they are used to describe the image of the object or the curvature of its outline. Assigning structure to a shape involves decomposing the shape into parts and describing their spatial relationships. Most algorithms decompose an object into (nearly) convex parts, a decomposition which seems to agree with the way human observers appreciate shape.

3.2 Three-dimensional objects represented by multiple views

Recognizing 3-D objects from visual images poses a fundamentally different problem from recognizing 2-D objects. In the latter, all shape information is available from one image, and the only problem is to represent the object's shape appropriately, e.g., with rotational invariance. An image of a 3-D object, on the other hand, is but one of infinitely many views of that object and does, therefore, not contain all shape information. Since we want to recognize objects from all possible vantage points, the question becomes how information derived from a large number of images is to be organized. Researchers have approached this problem in two different ways. In one approach, shape information derived from different vantage points is used to construct a 3-D model. In the other, which we will review in this section, the 3-D shape of an object is represented by a (structured) collection of views from different vantage points.[15]

In the "brute force" method, views from many vantage points are collected without distinguishing between vantage points. Thus, in their airplane recognition system Richard and Hemami (1974) stored the silhouette of a plane at 5° increments in roll (rotation about longitudinal axis) and pitch (rotation about vertical axis) angles, resulting in 666 silhouettes, the so-called reference set.[16] Each silhouette was described by 39 Fourier descriptors, although experiments using 21 FDs gave satisfactory results. Tests were performed with a library of 2664 silhouettes derived from four airplanes: an F-4, Mirage, MIG, and an F-105. The test silhouette was taken from some random vantage point and was therefore generally not explicitly stored in the reference set. Without any noise in the test silhouette, no identification errors were made, while the number of errors increased slightly (to 8 percent) with 20% noise.[17] Orientation was estimated quite accurately: within 2.5° without noise, and within 11° with 20% noise. Doubling the separation of the vantage points, i.e., taking silhouettes at 10° instead of 5° increments in roll and pitch angles, roughly doubled the errors in identification and estimates of orientation.

Wallace and Wintz (1980) used 143 silhouettes to represent the 3-D shape of an airplane. Each silhouette is described by 30 Fourier descriptors. The number of silhouettes per plane

---

[15] Madarass and Thompson (1985) developed a variant in which moving objects are recognized by their so-called feature signature, a time sequence of feature measurements of a moving object—e.g., as it rolls down a slope. Experiments were performed on bars having slightly different cross sections, the difference not being visible from most vantage points. By measuring the cross section as a bar rolls down a plane a signature is obtained. Of course, whether this or any other signature uniquely specifies an object depends on the set of objects to be recognized.

[16] Due to the bilateral symmetry of airplanes and ambiguities in projection, only one eighth the total number of silhouettes taken at 5° intervals, i.e., 72 times 72, need to be stored.

[17] Points comprising the airplane's silhouette were perturbed to simulate Gaussian noise; standard deviation of the noise was expressed as percentage of the distance between the centroid of the silhouette and the point on the silhouette farthest from the centroid.

is relatively small because of an efficient intrapolation procedure between feature vectors. Depending on the resolution of the image, aircraft identification accuracy ranged from 70 to 93% (accuracy is lower than in Richard and Hemami's experiments because of poorer image resolution). In a related experiment, Wallace et al. (1981) described silhouettes by a feature vector consisting of alternating angles and distances, respectively indicating the angular change at curvature extrema in the contour and the distance along the silhouette between preceding and succeeding curvature extrema. As compared with the system using FDs to describe silhouettes, accuracy was slightly lower for low resolution images (which makes sense because local information, i.e., exact location of curvature extrema in the silhouette, becomes less accurate), and slightly higher for the highest resolution.

Noting that this approach to 3-D object recognition invariably results in a large reference set, in turn leading to complex and/or time consuming searches to match test shapes with reference shapes, Wu and Stark (1986) proposed to reduce the entire collection of 2-D views to one 2-D "signature image." The idea is to collapse each 2-D image indexed by the viewing angle ϕ, $I_\phi(x,y)$, into a 1-D function, $P_\phi(x)$, by integrating along one dimension of the 2-D image: $P_\phi(x) = \int I_\phi(x,y)\,dy$. The signature image F combines all collapsed views for instance by setting $F(\rho,\phi) = P_\phi(\rho)$. Ordinary pattern recognition techniques can then be applied to the signature image to extract rotation, scale and brightness invariants suitable for recognition. Wu and Stark demonstrated their approach for a highly restrictive situation in which rotation about the vertical is the only degree of freedom. The question therefore remains how this approach can be extended to the general, unrestricted case. And since the mapping from 3-D object to 2-D signature image is not unique one would have to investigate which objects share the same signature images.

Although a valuable, and in some cases the only source of information, silhouettes provide only part of the shape information one can derive from images. One could also use information about surface orientation or surface type. Using a laser range finder to obtain a depth map,[18] Oshima and Shirai (1983) segmented a scene into regions, each of which could be described by a single surface primitive (planar, ellipsoidal, hyperboloidal, cylindrical, etc.). They described the spatial relationships between regions in terms of connections and angles between regions (fitting a plane to the points of curved regions to compute an angle). The result is a partial model of an object, and "If one view is not enough to describe an object, several typical views are shown and multiple models are made." Unfortunately, the authors did not mention their criteria for including more models in the description of an object, nor did they quantify the results of their experiments. The context of their paper does, however, suggest an interesting learning paradigm in which the system only adds

[18] A depth map indicates the distance to the nearest surface for each direction. It is much like a short-range radar.

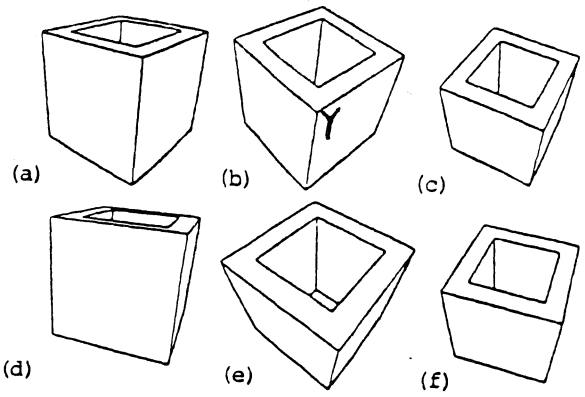


Figure 3.7 **Six views of a cube with a hole. (a) through (d) are topologically equivalent (Chakravarty and Freeman 1982).**

models if the recognition procedure fails to come up with an answer, i.e., whenever input does not warrant a conclusive categorization.

The above methods have one important shortcoming, and that is the absence of structure in the collection of views. Different views of the same object are essentially treated as different objects which happen to have the same name. But, obviously, different views are related by means of their difference in vantage point. In general, views from slightly different vantage points will have only a few differences if any, a fact which could be exploited to reduce the storage requirements of the reference set. These and related ideas have been addressed by Minsky (1975), Koenderink and van Doorn (1979), Chakravarty and Freeman (1982), and Plantinga and Dyer (1987ab).

In his article on frames as a means to structure knowledge, Minsky (1975) proposed that different views (i.e., views that are qualitatively different) of say a cube be represented by different frames, where each frame describes a view of the cube in terms of its faces and their spatial relationships. Different frames are connected by arcs specifying the spatial transformation that brought this change in appearance about. Being not particularly interested in vision, Minsky did not clarify what he meant by qualitatively different views of an object.

Chakravarty and Freeman (1982) expanded Minsky's ideas and implemented a system to recognize polyhedra from different viewing positions. They divided the space of viewing positions into equivalence classes, where two viewing positions are equivalent if their associated images are topologically equivalent. Two images are topologically equivalent if the line segments and line junctions of one can be mapped one-to-one and continuously onto those of the other image. Fig.3.7 shows six views of a cube with a hole in the middle; views (a) through (d) are equivalent, (e) and (f) are different. For the purpose of recognition, each view is described by a 5-tuple specifying the number of junctions of a particular type

(Chakravarty 1979). The junctions are ranked according to the number of visible faces along the lines making up a junction. Junction Y in Fig.3.7b is an example of the highest ranking junction, a junction consisting of three lines each of which has two visible faces. This rank ordering allows the set of views to be organized hierarchically, and therefore to be searched more quickly for the presence of a particular view. An important question, and one that remains to be investigated, is whether the mapping from view to the 5-tuple junction description is one-to-one, in other words, whether a description in terms of number of junctions is unique or not.

Starting with the assertion that an observer's representation of an object is not some passive repository of information relating to the object's shape but is essential to the observer's interaction with that object and the environment in general, Koenderink and van Doorn (1976b, 1979) concluded that it is the purpose of the internal representation of an object to predict how movements of the observer affect the appearance of that object.[19] That is, an object's internal description has to predict how it changes appearance as a result of movements of the observer. This allows the observer to distinguish between changes in the image due to its movements and those of the object itself, something which is essential for veridical perception of the object (Koenderink and van Doorn 1976c). The appearance of an object is simply taken to be the orientation of its visible surfaces, specified by, say, the slant field, a vector field indicating the direction and magnitude of the maximal change in distance between surface and observer at each point in the image.

At the global level, the *aspect* of an object specifies the topological structure of the slant field, i.e., the singularities of the slant field and their spatial ordering. The aspect specifies occluding contours (images of curves on the object that separate visible from invisible parts of he surface), cusps, T-junctions, specular points (Fig.3.8a). For most vantage points, small movements of the observer do not alter the qualitative appearance of the object, that is, leave the aspect of the object invariant. For those vantage points that are unstable in the sense that small changes in viewpoint change the aspect, Koenderink and van Doorn (1976c, 1979) provided an inventory of possible changes in aspect. These changes include the appearance or disappearance of pairs of cusps or T-junctions whenever a convex protrusion appears or disappears. The *visual potential* of an object is a graph whose vertices represent all possible aspects and whose edges specify the spatial relationships between aspects. Thus, the visual potential of a sphere consists of but one aspect as its appearance never changes. Fig.3.8b shows the visual potential of a slightly more complex object, a pyramid; it consists of 14

[19] Held and Hein (1963) showed that development of visually-guided behavior in kittens—examples of which include visually-guided paw placement and discrimination of the high and low side of a visual cliff—depends on visual feedback resulting from self-produced motion. Holst and Mittelstaedt (1950) and Gyr et al. (1979) discuss the role of feedback in the way animals interact with their environment and visual perception.

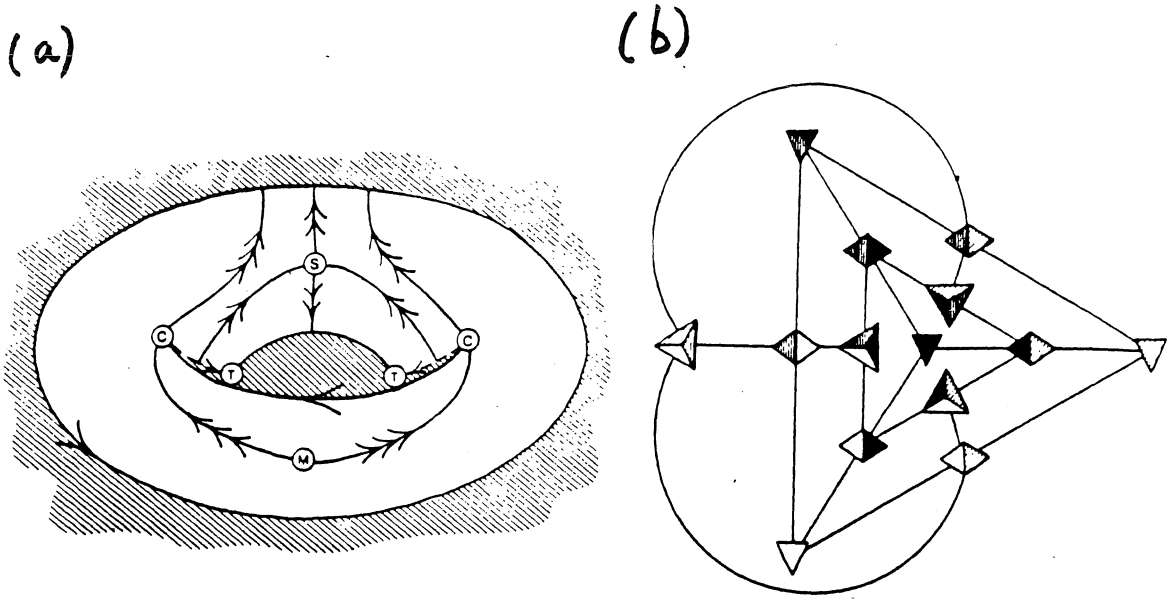


Figure 3.8 **(a) Aspect of a torus. C, cusp; T, T-junction; S, saddle point; M, minimum of distance function. (b) Visual potential of a pyramid (Koenderink and van Doorn 1979).**

aspects which fall into three classes, namely aspects with one, two, or three visible faces.

The visual potential allowed Koenderink and van Doorn to quantify the notion of complexity of shape. The distance between two aspects, A and B, is simply the smallest number of edges one has to traverse in going from A to B. This measure of distance, call it d, is a true metric since $d(A,B) = 0$ iff $A = B$, $d(A,B) = d(B,A)$ and $d(A,B) + d(B,C) \geq d(A,C)$. The complexity of a shape is just the largest distance between any two aspects in its visual potential. The intuitively simplest shape, the sphere or more generally the ovoid, has complexity zero, because its visual potential has only one aspect.

So far the visual potential or characteristic views of an object have only been analyzed for rigid objects. It will be necessary to expand the analysis to include nonrigid objects. For example, what is the visual potential of the human hand? It might be profitable to decompose the visual potential into regions that can change independently.

Koenderink and van Doorn's choice of the visual potential as the internal representation of an object was primarily motivated by its ability to predict changes in the object's appearance quite easily. This does not mean that other representations cannot do the same; see, for example, section 3.3.2 on Gaussian images.

Plantinga and Dyer (1987ab) generalized the visual potential by representing aspects from all viewing directions. In particular, they propose that objects be represented in a four-dimensional space which is the 2-D image cross the 2-D space of all viewpoints. Korn and Dyer (1987) describe algorithms to efficiently manipulate and search through a collection of 2-D views.

3.3 Three-dimensional objects represented by 3-D models

In the previous section, we discussed the representation of 3-D objects by means of collections of viewer-centered aspects. We now explore object-centered descriptions, that is, descriptions that do not depend on viewing position. I have divided these descriptions into four categories. In the first, 3-D objects are described as combinations of surface patches, for example quadratic surface patches. The spatial relations between these patches are specified by a shape grammar, which is not unlike a grammar for languages. This is known as the syntactic or structural approach. The second method, the extended Gaussian image, also represents the surface of a solid, but does so without dividing the surface into patches. The Gaussian image of an object specifies the distribution of surface normals on the object's surface. Of course, the Gaussian image could be used to divide an object's surface into patches yielding a structural description; that is, these two approaches do not necessarily exclude each other, but might form different stages in some vision system.

The two remaining categories explicitly represent the volume taken up by the object under consideration. In the bounding volume approximation we specify as well as possible the volume an object occupies; the resulting representation is not unlike a "volume occupancy array" (Ballard and Brown 1982). In the fourth and last approach we will discuss, a solid is decomposed and described in terms of volumetric primitives.

3.3.1 Shape grammars of surface patches

Not surprisingly, surface-based descriptions of solids are used extensively in computer graphics. We will not review all the possible ways in which surface patches can be described (see Requicha (1980), Ballard and Brown (1982), Foley and van Dam (1982) for comprehensive overviews), but focus on methods for combining surface patches to form a 3-D object. In particular, we will look at the syntactic approach of Fu (1982) and Lin and Fu (1984).

We assume that surface patches are described somehow, for example by bicubic splines. To connect these patches we introduce the idea of a surface patch to which other surface patches can be attached along "attaching curves." A primitive which can be connected with n others is called an n attaching curve entity or NACE, a generalization of Feder's (1971) n attaching point entity for specifying 2-D interconnections. A 3-D-plex-grammar specifies how NACEs can be combined:

$$G_p = (N, \Sigma, P, S, I, i_o), \tag{4}$$

where N (respectively Σ) is a finite, nonempty set of NACEs called the nonterminals (respectively terminals), S is the start NACE, I is a finite set of symbols identifying the attaching curves on each NACE, i_o is the null identifier, and, finally, P is a set of production or rewrite

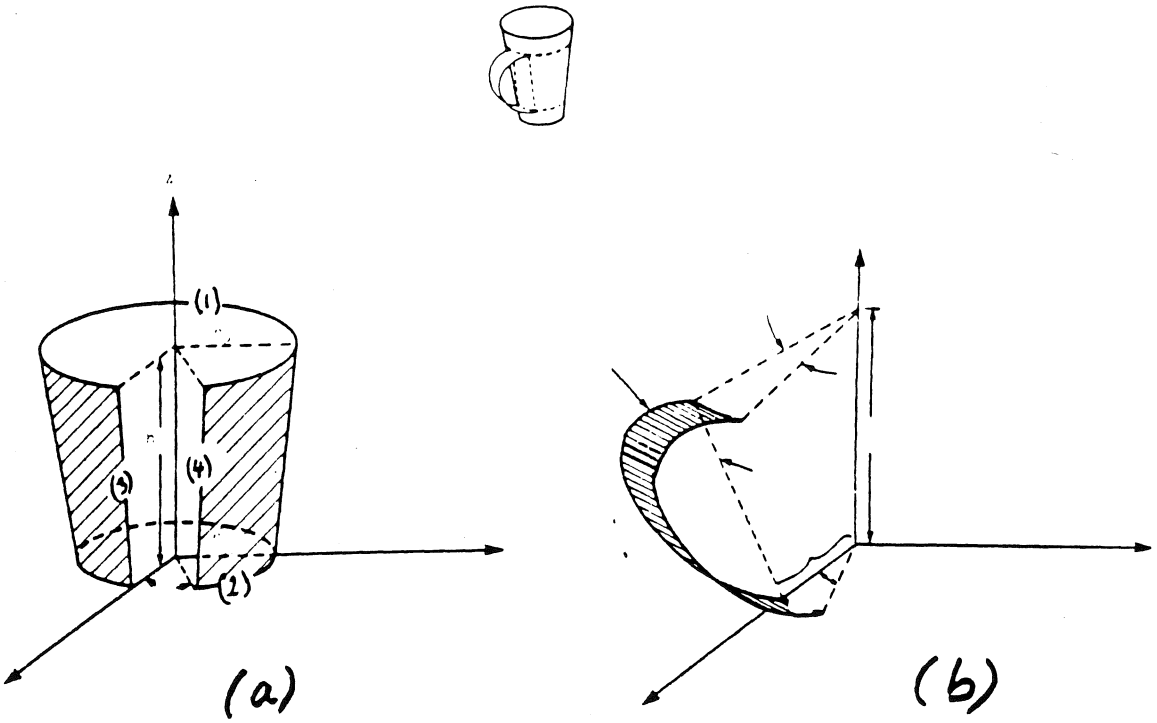


Figure 3.9 Syntactic description of a cup. The "middle" surface of the cup consists of part of a cylindrical surface (a) and a handle (b) (Lin and Fu 1984).

rules. Interconnections of NACEs are specified by mapping the appropriate identifiers onto each other.

For a context-free grammar, productions are of the form

$$A\Delta_A \rightarrow \chi\Gamma_\chi\Delta_\chi, \tag{5}$$

specifying how the NACE A is replaced by the plex structure $\chi\Gamma_\chi$, where χ a nonempty list of NACEs and Γ_χ specifies how the NACEs in χ are interconnected. Furthermore, Δ_A and Δ_χ specify how the new plex structure $\chi\Gamma_\chi$ is embedded in the already existing structure of which A was a part.

As an example, consider the cup shown in Fig.3.9. Its plex grammar is

$$\begin{aligned}
G_p &= (N, \Sigma, P, S, I, i_o), \\
N &= \{\langle\text{cup}\rangle, \langle\text{top}\rangle, \langle\text{middle}\rangle, \langle\text{handle}\rangle, \langle\text{bottom}\rangle\}, \\
\Sigma &= \{\langle\text{a}\rangle, \langle\text{b}\rangle, \langle\text{c}\rangle, \langle\text{d}\rangle, \langle\text{e}\rangle, \langle\text{f}\rangle\}, \\
S &= \langle\text{cup}\rangle, \\
I &= \{0, 1, 2, 3, 4\}, \\
i_o &= 0,
\end{aligned}$$

and P, the set of production rules, is

$$
\begin{aligned}
&(i)\ \langle \text{cup} \rangle \{\} \rightarrow \langle \text{top} \rangle \langle \text{middle} \rangle \langle \text{bottom} \rangle \{210; 021\} \{\} \\
&(ii)\ \langle \text{top} \rangle \{1\} \rightarrow \langle \text{a} \rangle \{\} \{2\} \\
&(iii)\ \langle \text{middle} \rangle \{2\} \rightarrow \langle \text{b} \rangle \langle \text{handle} \rangle \{34; 43\} \{11; 22\} \\
&(iv)\ \langle \text{handle} \rangle \{4\} \rightarrow \langle \text{c} \rangle \langle \text{e} \rangle \{11; 22\} \{11; 22; 30; 40\} \\
&(v)\ \langle \text{bottom} \rangle \{1\} \rightarrow \langle \text{d} \rangle \langle \text{f} \rangle \{21\} \{10\}.
\end{aligned}
$$

The terminals ⟨a⟩ through ⟨f⟩ refer to the surface patches in which the surface is divided. Fig.3.9a shows ⟨b⟩, a segment of a cylinder derived via rule (iii), in the process producing a ⟨handle⟩, which according to (iv) consists of a segment ⟨c⟩ and a curved surface ⟨e⟩, the actual handle (Fig.3.9b). It is clear, as Lin and Fu (1984) themselves point out, that the decomposition of a complex object can be nonintuitive, and the associated structural description rather complex. Lin and Fu (1986) developed an algorithm to determine whether a scene contains an object included in the class of objects defined by some plex grammar G_p. In case of recognition, the algorithm outputs a sentence describing the exact structure of the object. The scope of their system was limited to objects whose surface contains at least some polygons.

3.3.2 Extended Gaussian images

Instead of describing the surface of a solid by *a priori* surface patches, we could use the normals to the solid's surface. The idea is to specify how much of the surface is oriented in each direction. For example, all surface elements of a plane have the same orientation, i.e., their normals all point in the same direction. For a sphere the opposite holds: all normals point in different directions. Intuitively, then, there is a potentially interesting relationship between an object's shape and the distribution of its normals.

To investigate this relationship, Gauss (see Hilbert and Cohn-Vossen 1952) introduced a mapping, now named after him, in which each unit surface normal **N** is translated to the origin of a unit or Gaussian sphere S as shown in Fig.3.10a. This is also called a mapping by parallel normals. After translation, each normal touches the Gaussian sphere at some point P; hence P is called the Gaussian image of the point(s) on the solid whose surface normal is **N**. The normals of some small area δO on the object will be mapped onto a patch δS of the Gaussian sphere (Fig.3.10b). The area of δS depends on the curvature of δO: if its curvature is small, say approximately zero in an almost planar area, then all normals will map to approximately the same point on S; in other words, δS will be small. Conversely, if δO is curved considerably, its normals point in quite different directions and will map to widely spaced points on S; in other words, δS will be large. It seems then that we could

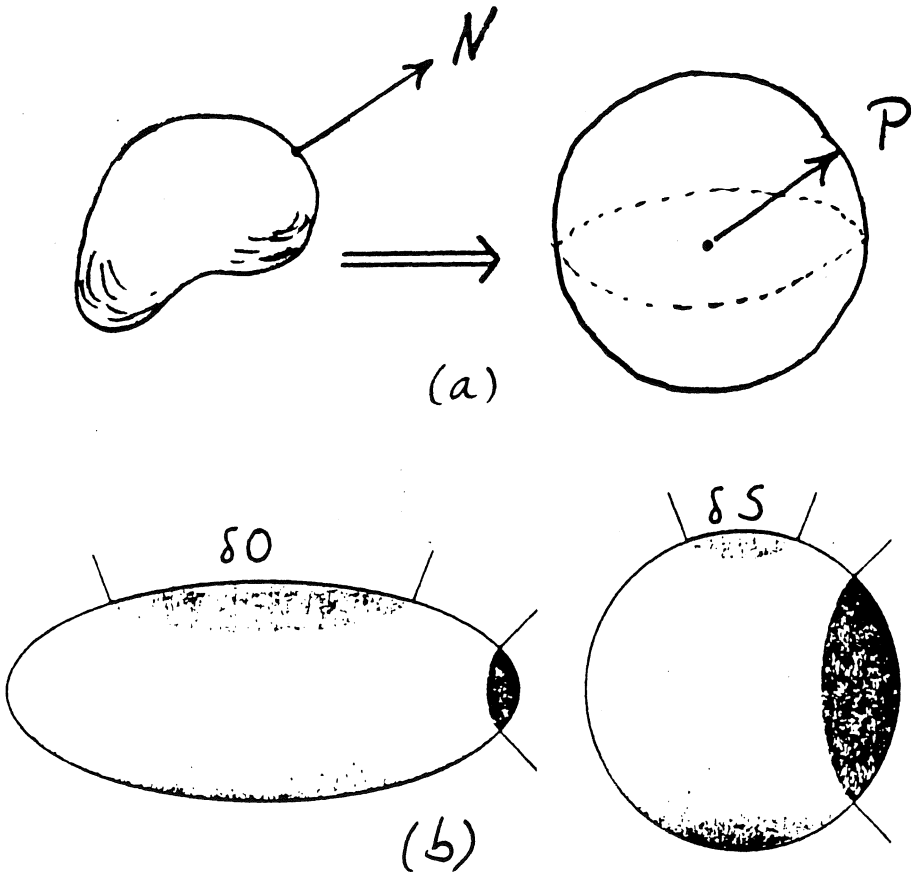


Figure 3.10 **(a) Mapping by parallel normals. (b) Gaussian curvature is the ratio of δO and δS (Horn and Ikeuchi 1984).**

use the relationship between the area of δO and δS to specify the shape of δO. In fact, the Gaussian curvature K of a surface is defined as the limit of $\delta S/\delta O$, as δO approaches zero (Hilbert and Cohn-Vossen 1952; Horn 1984).

By specifying, for each point on the Gaussian sphere, the Gaussian curvature of its pre-image we obtain an extended Gaussian image (EGI) of the solid. Another way to extend the Gaussian image is to specify the inverse of Gaussian curvature, $1/K$. Integrating over the Gaussian sphere we get

$$\int_S \frac{1}{K} dS = \int_S \frac{dO}{dS} dS = \int_O dO = O, \tag{6}$$

that is, the surface area of the object (Horn 1984). Van Hove and Verly (1985) defined yet another extension of the Gaussian sphere. They retained the second fundamental coefficients which completely specify curvature (in a Monge patch representation), making it possible to invert the EGI, i.e., to recover the surface shape from corresponding points on the EGI.

A Gaussian sphere thus extended also allows one to recover the occluding contour of the object as seen from any direction. The occluding contour is the image of those points on the object's surface whose surface normal is perpendicular to the viewing direction. It follows that the Gaussian image of these points forms a great circle[20] perpendicular to the viewing direction. Thus to compute the silhouette or occluding contour of the object as seen from any direction, one simply takes the great circle on the EGI perpendicular to that direction and computes local curvature of the contour from the fundamental coefficients. This shows that one does not have to retain something like a visual potential in order to be able to predict how changes in viewing position will change the object's appearance.

Now we have defined a representation of an object, namely an extended Gaussian image, the usual questions come to mind. Are EGIs invariant under translation, rotation, and scaling? Is the mapping between EGI and object one-to-one? It is clear that EGIs are invariant under translation because translation does not affect the orientation of surface normals. Although EGIs are not invariant under rotation, the relationship between the EGI of an object in its standard orientation and its EGI after rotation is straightforward. Rotating an object about a certain axis causes its EGI to be rotated about the same axis to the same extent. Thus, the EGI can be used to recover the orientation of a (convex) object (Ikeuchi 1983; Horn and Ikeuchi 1984; Brou 1984; Little 1985b). Similarly, EGIs depend on scale, but can easily be made scale invariant by dividing the values on the EGI e.g, K or $1/K$) by the surface area of the object, which is easily obtained through equation 6. In other words, we normalize the EGI with respect to surface area.

The mapping between EGI and object is one-to-one only for convex objects whose Gaussian curvature is everywhere greater than zero. For this class of objects, Minkowski showed that there is a unique object for any given EGI (Pogorelov 1973).[21] On the basis of this proof, Little (1983, 1985a) developed an iterative procedure to reconstruct a convex polyhedron from its EGI. Thus EGIs can be used to recognize convex objects (Smith 1979; Ikeuchi 1981). To recognize non-convex objects we could, as Smith (1979) suggests, first decompose an object into convex parts and then represent each part by means of an EGI. This will only work if one can formulate an algorithm that decomposes similar objects in roughly the same way. We already discussed decompositions into convex parts for 2-D objects, and will turn to the 3-D case in section 3.3.4.

---

[20] A great circle on a sphere is a circle whose center is the origin of the sphere. The equator is a great circle of the earth and so are the meridians, but the tropics of Cancer and Capricorn are not.

[21] Minkowski showed that there is a unique convex object for any positive, continuous function $K(n)$ defined on the Gaussian sphere. This result can be generalized to any twice differentiable strictly increasing function f of the two principal curvatures R_1 and R_2. Minkowski's problem is the special case $f(R_1, R_2) = R_1 R_2$; Christoffel's problem is $f(R_1, R_2) = R_1 + R_2$.

3.3.3 Boundary volume approximation

Wang et al. (1984; Martin and Aggarwal 1983) used a two-stage method to recognize 3-D objects from their silhouettes. First, they reconstruct as well as possible an object's 3-D shape or bounding volume from a series of silhouettes. Note that the different vantage points of the observer have to be known quite accurately in order to reconstruct the 3-D shape of an object in this manner, and that this method can only be applied to rigid objects. In the second stage, they derive object invariants from the bounding volume approximation. These invariants include the so-called principal silhouettes, that is, the three silhouettes one obtains by projecting the bounding volume along each of its three principal axes.[22] The shape of the three silhouettes can be described by Fourier descriptors (FDs) as discussed in section 3.2. If an object has any holes, each will be described by a separate set of FDs and its position and orientation with respect to the silhouette. The principal silhouettes and the principal moments[23] serve as indices into shape memory.

Two experiments were performed with a shape memory consisting of 10 different kinds of trucks and cars. In one case, three silhouettes sufficed to recognize the object correctly, and in the other case, five were sufficient. In addition to the drawbacks mentioned above, Wang et al. (1984) noted that their method will fail whenever parts of an object are missing since the correct principal axes (which form the basis for the object invariants) cannot be recovered. This points towards the need for a more structured representation, one which can take such deviations from the prototype into account.

3.3.4 Decomposition into volumetric primitives

As has become clear in the previous discussions, the ability to decompose a complex object is almost a prerequisite for any successful representation of shape. If one refrains from decomposing an object, as in the principal silhouettes method of the previous section, the resulting representation is sensitive to noise and missing parts. Of course any representation has to deal with imperfect samples or deviating members of a category; the important point is that with principal silhouettes local deviations from the prototype affect the representation globally. Imagine, for example, that one of the wheels of a car is missing. Since this influences the computation of the principal axes, this local change has a global influence—it changes the description of all other parts of the car. We will know look at some methods for decomposing

[22] Principal axes of an object are the eigenvectors of its inertia matrix. These directions describe the distribution of the object in space: In one direction, the object is spread out most, roughly its axis of elongation; and in another, orthogonal direction, it is spread out least. The third principal direction is simply orthogonal to the first and second directions.

[23] The moments of inertia with respect to the principal axes; see Sadjadi and Hall (1980) and Bamieh and DeFigueiredo (1984) for other examples of the use of moment invariants in recognition.

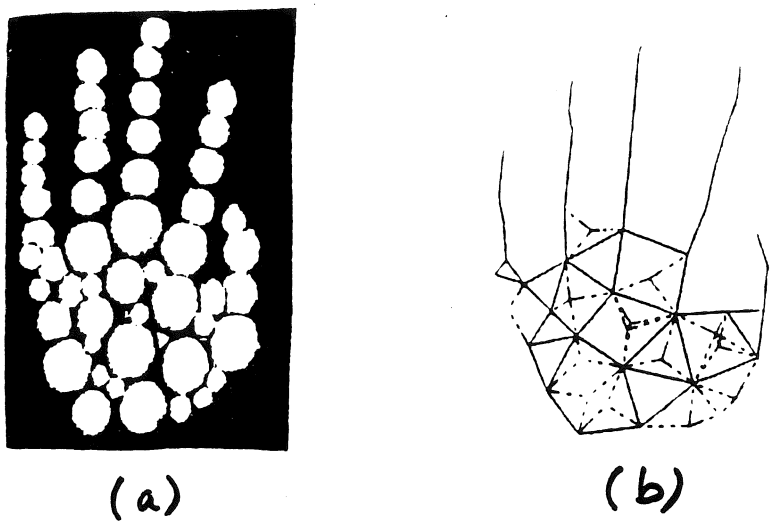


Figure 3.11 (a) Approximation of a hand by tangential spheres. (b) Graph of spherical approximation, showing elongated and flat parts (Mohr and Bajcsy 1983).

solids.

O'Rourke and Badler (1979; extended by Mohr 1982) developed an algorithm to approximate 3-D objects by a collection of overlapping spheres, analogous to the way in which 2-D objects are approximated by overlapping disks in the symmetric axis transform. Mohr and Bajcsy's (1983) approach differed slightly in that they used tangential instead of overlapping spheres. In both cases, the decomposition of an object into spheres is considered to be but an intermediate step in shape representation, one that would allow properties such as elongation or flatness to be easily computed. For example, in Fig.3.11a we see how a hand can be approximated by 54 tangential spheres. This approximation naturally leads to a graph whose vertices represent the spheres and whose edges specify the spatial relationships between the spheres (Fig.3.11b). This graph allows higher-level descriptors such as "elongated thin part" or "flat part" to be extracted. Thus the hand is described as consisting of five elongated parts (fingers), a flat part (palm), and their interconnections.

A high-level primitive which captures local properties directly and can be derived directly from an image, is the **generalized cylinder** (Agin and Binford 1976; Agin 1981) or **generalized cone** (Nevatia and Binford 1977; Brooks 1981, 1983). A generalized cylinder is defined by its axis, which can be any curve in 3-D (hence the generalized), and planar cross-section. The cross-section "sweeps out" some 3-D volume as it follows the axis. The generalized cone is more general since the cross-section is allowed to change as it moves along the axis. The idea is to approximate the shape of an object by these primitives, and to assign each the role of "part of the object." Now, in general, one can describe a solid in a number of ways using generalized cones, and, consequently, decompose it in a number of ways. This number can be limited by applying heuristics that require parts to be generalized cones having a smooth axis and smoothly changing cross-section, and by preferring elongated

and cylindrical cones. Of course, this problem can also be minimized by choosing the right domain, e.g., airplanes in the vision system ACRONYM (Brooks 1981, 1983).

Marr and Nishihara (1978; for discussion see Sutherland 1979; Nishihara 1981; Marr 1982) formulated three criteria by which to judge shape representations designed for recognition. First, descriptions in terms of the proposed representation should be reliably computable, in other words, they should be accessible from an image and degrade gracefully with increasing noise levels. Second, all objects of interest should have a unique description. Third, descriptions should be stable under small changes in shape, yet sensitive enough to capture small differences in shape. These considerations led them to choose an object-centered representation using volumetric primitives over a viewer-centered one like the visual potential. The decisive property in favor of the object-centered representation was the ease with which it allows objects to be described at different scales. This is illustrated in Fig.3.12: Depending upon the scale, an arm consists of one or two cones.

While Marr and Nishihara's criteria are hard to argue with, it is not clear whether their object-centered representation meets them in practice; nor is it clear that a representation like the visual potential fails to meet them. As we saw already, it is hard to recover generalized cones from an image and, so far, *ad hoc* heuristics have proved indispensable. Given this difficulty, it is not hard to imagine that descriptions might not always be unique, in particular for objects that are not obviously built from generalized cones. A representation like the visual potential simply circumvents these problems—assuming that the topological structure of the slant field, the basis for the visual potential, can be recovered reliably. And to describe an object at different scales we can simply retain the visual potential at a number of different resolutions.[24] One major difficulty with the visual potential, and one that remains to be investigated, is how it changes under bendings of the object, that is, changes in the spatial relationships among parts of the object. The object-centered representation naturally accommodates these changes because spatial relations are represented explicitly as predicates connecting the volumetric primitives.

Instead of having the approximation of an object dictate its decomposition, one could decouple the two by first partitioning an object and then describing the resulting parts. Hoffman (1983; Hoffman and Richards 1984) proposed that objects be partitioned along curves having a local minimum of curvature. This choice of part boundaries was motivated by the observation that whenever one combines two solids one creates a contour of intersection along which curvature has a local minimum (Bennett and Hoffman 1985; see Fig.3.6a and section 3.1.2). Beusmans et al. (1987) extended this analysis and derived rules to decompose objects without holes into parts that are usually convex. This partitioning reveals the object's

[24] For related work on the representation of the retinal image at different scales or resolutions, see Wilson and Bergen (1979), Koenderink and van Doorn (1978; 1982), Witkin (1983) and Koenderink (1984a), Pizer et al. (1986, 1987).

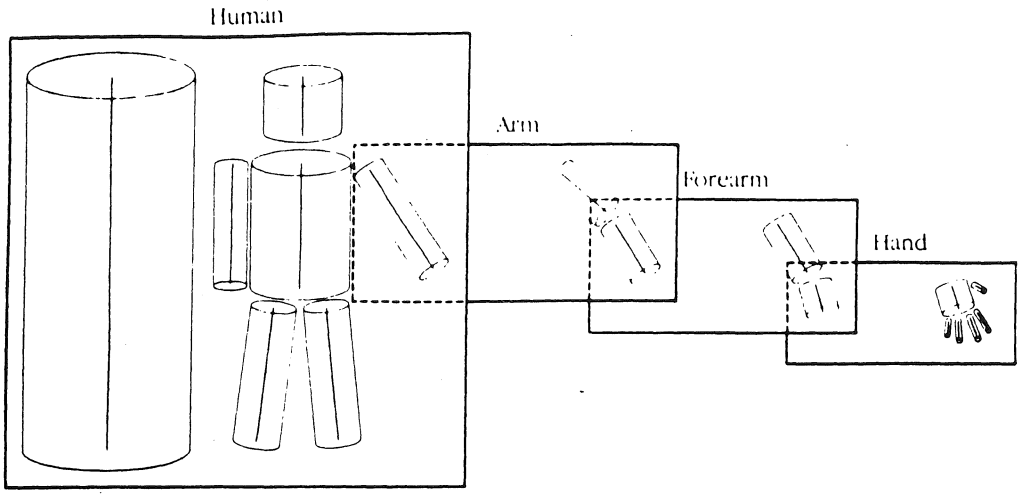


Figure 3.12 **A 3-D model description of a human being. Note that a part can be described at different levels, and that spatial relationships are specified in distributed coordinate systems. The fingers are described with respect to a coordinate system at the hand level instead of some global coordinate system (Marr and Nishihara 1978).**

"deep structure," i.e., its parts and their spatial relationships. Spatial relations could be described by the relational model for 3-D objects developed by Shapiro et al. (1984). This model assumes that 3-D objects consist of sticks, plates and blobs, having one, two, and three significant dimensions, respectively. The spatial predicates are qualitative, specifying, for example, that two sticks are connected, but not exactly how.

Koenderink and van Doorn (1986b) noted that in academic art theory shape is viewed as a "hierarchically ordered structure" suggesting that "the perceptual approach is *dynamic*, not in a temporal sense but in the sense that a partial order is apparent that relies on a *hypothetical evolution* or morphogenesis that is an integral part of the shape description: the shape is thought of as has been formed from a primeval, shapeless, ovoid blob that was articulated in first rough then finer steps, finally leading to the present object [emphasis in original]." One way of retracing this morphogenesis is by a blurring process described in terms of the diffusion equation (Koenderink 1984a). Interestingly, the blurring of two sufficiently close blobs until they form one object results in a contour along which there is a minimum of curvature, i.e., the part boundary proposed by Hoffman and Richards (1982, 1984). This relationship between object decomposition and blurring is quite intriguing and deserves further study.

3.4 Discussion

We have discussed representations of shape proposed by computer vision researchers. Although programs have been written to recognize characters, airplane silhouettes, etc., more or less successfully, it is not clear how these methods generalize. And one wonders whether Brousil and Smith (1967) would still maintain that "although much effort has been invested in pattern recognition in recent years, the problem of classifying character classes having substantial variation lies essentially unsolved." Their diagnosis, "a lack of an underlying theory," has been taken to heart by a number of researchers, most notably in my view, Marr, Koenderink and van Doorn.

Marr (1982) championed what he called the computational approach, which emphasizes the functional aspect of vision; that is, it considers the question "What is the purpose of this visual information processing task?" to be of paramount interest. The answer to this question constitutes the underlying theory. Chomsky (1965) had already made a similar distinction between the *what* and *how* aspects of language understanding. His theory was one of competence, indicating what the syntax of English should be like, while disregarding how an English sentence should be parsed to produce the correct interpretation. While one can argue about the specifics of proposed theories, the important fact is that there is something substantial to argue about. That is, I agree wholly with Sutherland (1979) when he concludes his review of Marr and Nishihara's 3-D model by observing that "the decisions...taken in theory construction are always motivated: There is none of the 'I put that in because it seems to work, goodness knows why' attitude that characterizes the *ad hocery* of some practitioners of artificial intelligence. Marr and Nishihara's paper provides a starting point, not a finished theory."

Another good starting point is available in the work of Koenderink and van Doorn, who also analyze what the visual system is for, and proceed with exploring what information the visual image can possibly provide to satisfy this goal.

Before closing this section, I would like to briefly discuss an approach to shape representation that goes back to the forties, when Pitts and McCulloch (1947) and Wiener (1961) came to realize that "Gestalten" or figures are in fact equivalence classes of certain group actions.[25] Thus the "square" is an equivalence class under the group of planar translations, that is, we perceive a "square" no matter where it is located in the visual field. Similarly, under the group of dilations a "square" remains a "square;" this group captures what we call size constancy. However, the same is not true for the group of rotations: Rotating a "square" by 45° results in a "diamond." Hoffman (1966) expressed the usual perceptual constancies in terms of a Lie group of transformations, i.e., the group of "infinitesimal trans-

[25] A group is a set G with a binary, associative operation $\cdot$, such that G is closed under $\cdot$, G contains an identity element and its own inverses.

formations." More recently, this work has been reiterated (Marko 1973; Foster 1977; Dodwell 1983). Related work concerns auto-correlation in which a figure is represented in terms of transformations that map it onto itself (Uttal 1975; Gluender 1986; Kroese 1987). Thus a circle is mapped onto itself by the group of rotations; the square by the subgroup of rotations about 90°.

An important contribution of the work on transformations is that it shows that a representation does not have to copy, to re-present. In that sense, many of the representations we reviewed probably do not escape Pitts and McCulloch's (1947) denunciation of "Gestalt psychologists who will not conceive a figure being known save by depicting it topographically on neuronal mosaics, and [of] the neurologists...who must have it fed to some specialized neuron whose business is, say, the reading of squares."

This emphasis on the transformations an object can be subjected to in the real world is reminiscent of Shepard's insistence on the analogical character of representations. He believes that the essence of a representation is to reflect certain important, in this case spatial, relations among real world objects: "For the system of constraints that governs the projections and transformation of such bodies in space must long ago have become internalized as a powerful, though largely unconscious, part of our innate perceptual machinery" (Shepard 1978).

4. Neurobiological aspects

> Sensations are set by the encoding functions of the sensory nerve endings and by the integrated neural mechanisms of the central nervous system. Afferent fibers are not high fidelity recorders, for they accentuate certain stimulus features, neglect others. The central neuron is a story-teller with regard to the nerve fibers, and it is never completely trustworthy, allowing distortions of quality and measure... *Sensation is an abstraction, not a replication of the real world....*Each of us lives within the universe—the prison—of his own brain. Projecting from it are millions of fragile sensory nerve fibers, in groups uniquely adapted to sample the energetic states of the world around us: heat, light, force, and chemical composition. *That is all we ever know of it directly; all else is logical inference.* (Mountcastle 1975)

In this section, we will briefly review what is known about the visual system of certain animals, mainly cats and monkeys. Of course it is impossible to do justice to the complexity of this system in only a few pages, but I do hope I will leave you with an idea of its overall organization and functioning. We start with an overview of the visual system and identify its main components. Then we continue with a more detailed discussion of the pathway devoted to object recognition, giving special attention to the retina, lateral geniculate nucleus, and the various cortical visual areas.

4.1 Overview of the mammalian visual system

In mammals, an image of (part of) the outside world is projected onto the retina of each eye. The neurons in the retina somehow respond to the pattern of illumination that is the image and send the result to a number of other brain structures via the optic nerve. It appears that with few exceptions, the optic nerve in vertebrates projects to the following six regions in the brain (Rodieck 1979):

(1) Suprachiasmatic nuclei of the hypothalamus (above the optic chiasm, the crossroads of the optic nerves from the left and right eye). The suprachiasmatic nuclei are essential for the maintenance of a number of circadian rhythms, but do not depend on visual input for their rhythmic activity. It is thought that visual input keeps the rhythm synchronized.

(2) Accessory optic nuclei in the midbrain. Project to the cerebellum and prefrontal cortex, possibly to help regulate the coordination of head and compensating eye movements (Cooper and Magnin 1986).

(3) Pretectum (rostral of superior colliculus) which mediates pupillary light reflexes. The pretectal area also receives input from the superior colliculus, visual cortex, ventral lateral geniculate nuclei, and the frontal eye fields.

(4) Ventral lateral geniculate nucleus, vLGN (part of thalamus; the pregeniculate nucleus in primates). vLGN also receives input from the visual cortex, superior colliculus, pretectum and cerebellum, and projects to the superior colliculus, pretectum, accessory

optic nuclei, and the su achiasmatic nuclei. Its function is not clear, but it seems to be involved in the pupillary reflex in rats, and eye movements in monkeys and cats.

(5) Dorsal lateral geniculate nucleus, dLGN (part of the thalamus). Main target of the optic tract in mammals with well developed vision. In monkeys, dLGN contains six registered retinotopic maps which project to the visual cortex, and it also receives extensive feedback from these cortical areas.

(6) Superior colliculus (part of brainstem; analogous to optic tectum in lower animals). It also receives projections from cortical areas 17, 18, 19 and the lateral suprasylvian area, projections which are in topographical register with the projections from the retina. Additionally, there are topographic maps derived from the auditory, somatosensory, electrosensory (in electric fish), or infrared visual systems (Sparks and Nelson 1987). In rattle snakes, for instance, the infrared-derived map is in register with the retinotopic map (Newman and Hartline 1982). This alignment of different maps is obviously important for combining information about the location of objects, but it can also serve to calibrate one map by another map. In owls, the auditory map is calibrated by the visual map, but not vice versa (Knudsen and Knudsen 1985; Harris 1986). The superior colliculus itself projects to the pulvinar and lateral posterior nucleus of the thalamus, both of which project mainly to extrastriate areas in visual cortex (Bender 1981). In monkeys, the superior colliculus helps control the direction of gaze, and in frogs it presumably controls capturing of flies. In cats, two types of collicular neurons have been identified, one of which is sensitive to the relative motion between a small spot and a moving background grating, whereas the other type is tuned to absolute motion of a spot (Mandl 1985).

It is important to keep in mind that even among mammals, these patterns of interconnections vary considerably. For example, in primates the dLGN projects mainly to striate cortex and not to extrastriate cortex, whereas in cats both pathways exist (unlike a monkey, a cat without striate cortex is not blind).

At the cortical level, a large number of visual areas has been identified on the basis of function, cytoarchitecture, interconnections, and topographic organization (Zeki 1978; Van Essen 1979, 1985; Mishkin et al. 1983; Kaas 1987). The primary visual or striate cortex (Fig.4.1), for example, has a very distinct cytoarchitecture which clearly sets it apart from bordering cortex; in fact, its name derives from the stripes visible in cross sections (Fig.4.7). In addition, the striate cortex has a complete map of the visual field (Fig.4.2). Thus far almost twenty visual areas have been identified in the macaque monkey (Van Essen 1985); and, indeed, a major portion of its cortex is devoted to vision. Fig.4.2 illustrates the layout of the currently known visual areas of the macaque monkey, an Old World primate. These cortical areas have extensive interconnections, but also project to a number of subcortical

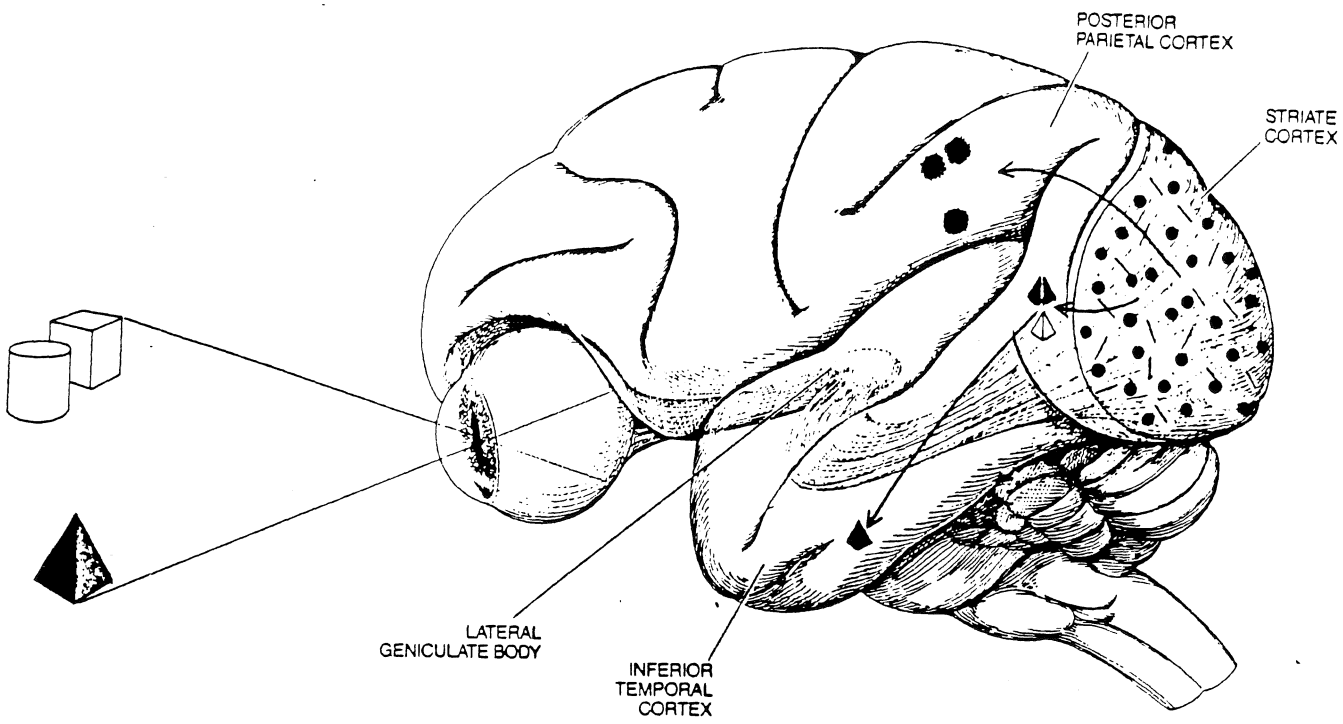


Figure 4.1 Visual system of the macaque monkey has two main pathways, one devoted to objects themselves, the other to the relation of objects with their environment, i.e., their spatial location and movement. After reaching the striate cortex, each pathway follows a different course: the object-oriented pathway passes through a number of stations until it reaches the pole of the temporal lobe, and the spatial pathway continues to the posterior parietal cortex (Mishkin and Appenzeller 1987).

structures (Graham et al. 1979; Mishkin and Appenzeller 1987).

An important question is whether there is some order in this multitude of visual areas and interconnections. Van Essen and Maunsell (1983) formulated principles that allow these areas to be organized in a hierarchy. They noted that, in general, connections between areas are reciprocal in the sense that if neurons of area A synapse with neurons in area B then the reverse is also true, i.e., neurons of area B synapse with neurons in area A. However, the two interconnections usually are between different cortical layers: The "forward" connections arise in superficial layers of the cortex and terminate mainly in layer IV, whereas "feedback" connections originate in both superficial and deep layers and terminate mainly outside layer IV. The first type is called "forward" because the dLGN projects mainly to layer IV of the striate cortex. We now use the "forward" connections to assign the visual areas to different levels of a hierarchy. Each area is assigned to the level immediately above that of all the areas having a "forward" connection to it, and areas having no "forward" connections are assigned to the first level. Connections that are not clearly "forward" or "feedback" are assumed to be between areas at the same level. Applying these rules to the visual areas of

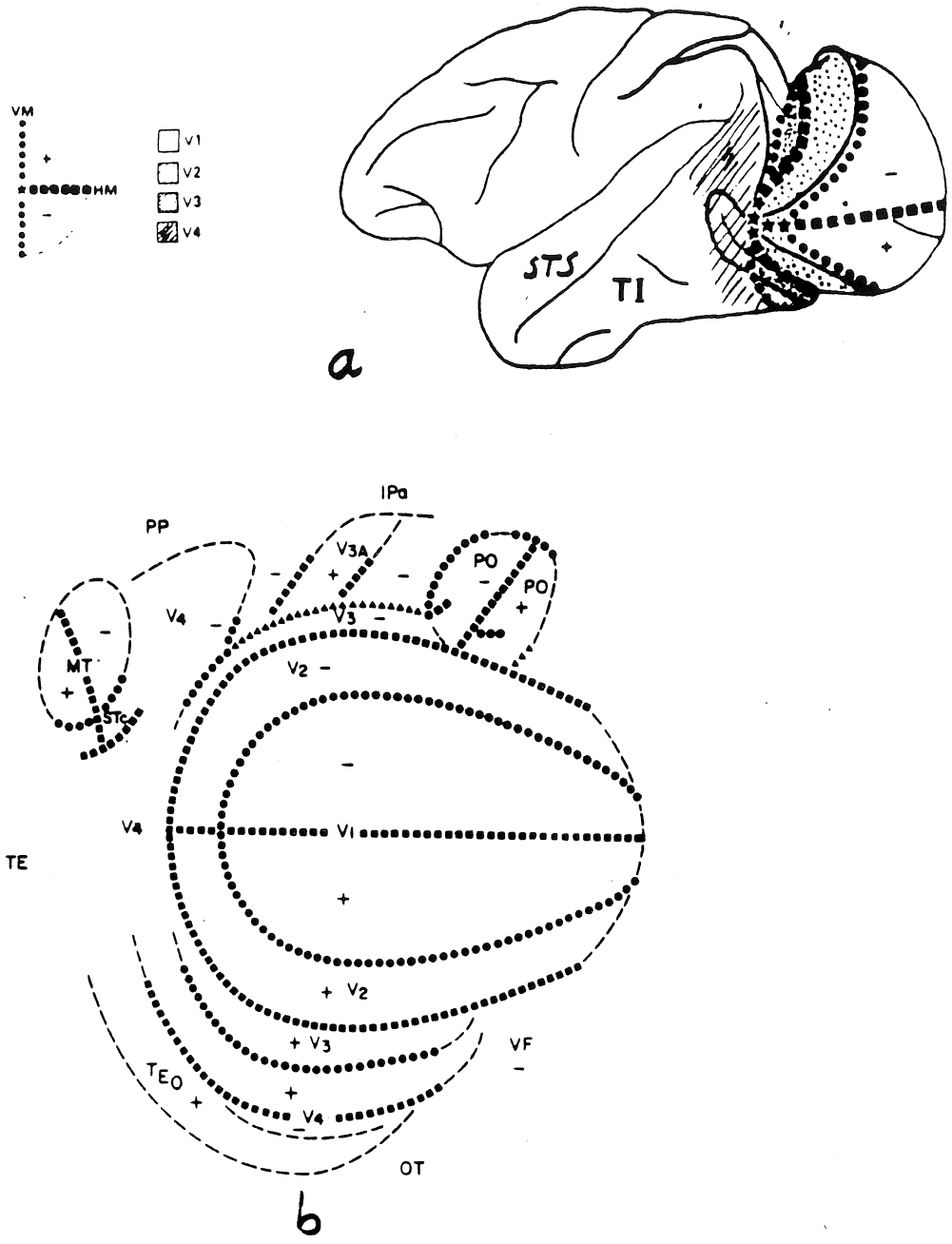


Figure 4.2 **(a) Side view of the** macaque left hemisphere **indicating the layout of visual areas and the projections of the vertical and** horizontal meridian of the **right visual hemifield. Note that the fovea is represented in adjacent areas in V1, V2, V3,** and V4. TI, **inferior temporal cortex; STS, superior temporal sulcus; +, upper visual hemifield; –,** lower visual hemifield (Desimone **et al. 1985). (b) Flattened representation of the visual areas of the** macaque left hemisphere (Gattass **et al. 1985).**

the macaque monkey results in a six-level hierarchy (Fig.4.3).

If this hierarchy is indeed meaningful one would expect higher levels to correspond to

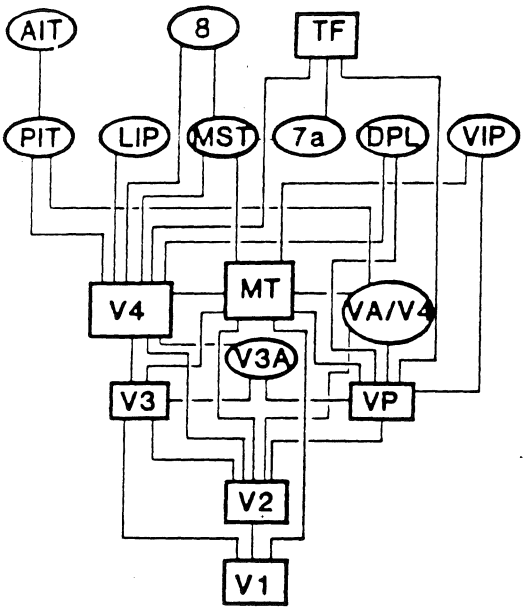


Figure 4.3 Hierarchy of the visual areas in the macaque cortex. The pathways that have been tested were found to be reciprocal. V1, primary visual or striate cortex; VP, ventral posterior (considered to be part of V3 by some authors); MT (V5), middle temporal; PIT, posterior inferotemporal; LIP, lateral intraparietal; MST, medial superior temporal; VIP, ventral intraparietal; AIT, anterior inferotemporal (Van Essen 1985).

higher levels of visual processing. And indeed, one measure of complexity of visual processing, the size of receptive fields, increases in going from bottom to top in the hierarchy. One would also expect that visual areas at the same level would process different kinds of visual information. And again there is evidence that there are at least two functional streams; one devoted to spatial vision, the other to object recognition and color (Schneider 1969; Macko et al. 1982; Van Essen and Maunsell 1983; Mishkin et al. 1983; Shipp and Zeki 1985; Desimone et al. 1985). Note that different authors emphasize different aspects of spatial vision. Van Essen and Maunsell (1983) and Sakata et al. (1985) refer primarily to motion analysis, whereas Macko et al. (1982), Mishkin et al. (1983), and Andersen et al. (1985) emphasize spatial localization. Of course these two aspects are closely related, since motion often leads to the localization of an object.

The pathway specialized in motion analysis proceeds from V1 to MT (middle temporal), and from there to areas MST (medial superior temporal) and VIP (ventral intraparietal) (Van Essen and Maunsell 1983). The majority of cells in MT, MST and VIP are sensitive to motion direction, motion speed and binocular disparity, ignoring stimulus shape and color. Although neurons of area V1 show similar properties, their receptive fields are about two orders of

magnitude smaller, indicative of a less advanced stage of information processing. That processing becomes more sophisticated as one goes up in the hierarchy is further illustrated by cells in MST and area 7a which distinguish between optic flow produced by movement in the scene and optic flow resulting from eye movements, a distinction not made at the V1 or MT level (Sakata et al. 1985).

The pathway subserving object recognition includes areas V1, V2, VP, V4, and IT. Of those, V2, VP, and V4 have many cells involved in color analysis, while remaining selective for orientation and disparity. As for object recognition, cells have been found in temporal areas that respond to specific stimuli such as hands and faces (Rolls 1984; Perrett et al. 1982, 1984, 1985; Desimone et al. 1984, 1985; Baylis et al. 1985; Kendrick and Baldwin 1987). And as in the spatial pathway, these cells have very large receptive fields, suggesting convergence of information. However, it is not at all clear what their triggering properties are and what (if any) are the intermediate stages of image analysis, stages between edge detection and recognition.

Thus, in solving the "what is where" problem, it seems that the visual system devotes one subsystem to the "what" and another to the "where" question. To solve the original problem, the "what" and "where" have to be combined. This can be done by having the subsystems project to the same location in the brain, i.e., by combining their results explicitly. A candidate location would be the hippocampus (Mishkin et al. 1983), a subcortical structure of the temporal lobe. The hippocampus has so-called place-field cells which fire only if the animal is at a particular location relative to some set of landmarks (O'Keefe 1979; Zipser 1985). Assuming that landmarks are recognized visually, one could say that place-field cells combine the what and where. Alternatively, the place-field cells fire only if some object is seen from a particular vantage point. In this case, spatial information is represented implicitly by means of the view or aspect of a known object.

Vision is only one of the senses that informs an organism about its environment. Other important sensory systems include touch and hearing. Since these systems refer to different aspects of the same object—stroking a cat, one can hear it purr and see it roll on the floor—one might expect them to converge on certain regions of the brain. This is indeed the case. All three sensory systems, vision, touch, and hearing, have overlapping projections to the premotor and prefrontal regions of the frontal lobe, and the parahippocampal gyrus of the temporal lobe (Pandya and Seltzer 1982). The premotor cortex projects to the motor cortex, suggesting that it translates sensory "input" into "motor" output. Lesions in premotor cortex confirm this interpretation: the animal is unable to respond to contralateral sensory stimulation and cannot learn motor tasks involving crossmodal integration (Pandya and Seltzer 1982). The amygdala, another structure that receives extensive input from all sensory systems, is thought to mediate crossmodal associations. Removal of the amygdala leads to the Klüver-Bucy syndrome in which monkeys repeatedly and indiscriminately investigate

inedible objects, seemingly unable to remember the results of tasting the object when looking at it or touching it. And together with the hippocampus, the amygdala seems to be involved in memory formation by way of the basal forebrain's cholinergic system which innervates the final stages of the visual systems, that is, areas where "perceptions" are though to be consolidated (Murray and Mishkin 1985; Bachevalier et al. 1985; Mishkin and Appenzeller 1987).

4.2 Visual pathways

As we saw in the previous section, the visual pathway subserving object recognition passes through three main structures: the retina, the dorsal lateral geniculate nucleus in the thalamus, and the cortex. We will now discuss each in some detail, look at its structure and the properties of its cells, and from that try to understand its function.

4.2.1 Retina

The eye functions much like a camera, projecting an image of the world onto a photo-sensitive surface, in this case a more or less hemispherical surface called the retina (Fig.4.4a). In functional terms, the input to the retina is a pattern of light and its output is a pattern of activity in the retinal ganglion cells, whose fibers converge on the blind spot and continue as the optic nerve. However, the activity in the optic nerve does not merely reproduce the input pattern of light and dark; that is, the output of a ganglion cell at position (x, y) is not proportional to the light intensity at (x, y) as sensed by the photoreceptors cells. Instead, as we will see, ganglion cells signal differences in light intensity, not absolute values (we will also see that this is not true for low light levels: in starlight, ganglion cells do signal absolute light intensity). Thus the analogy between eye and camera breaks down immediately after the photoreceptor layer, at least for daylight conditions.

The retina consists of a number of different cells, commonly divided into five classes. In addition to the **photoreceptors** and **ganglion cells**, there are **bipolar cells, horizontal cells,** and **amacrine cells**, which together form a complex network connecting photoreceptors and ganglion cells. Bipolar cells directly connect photoreceptors and ganglion cells; horizontal cells interconnect photoreceptors and also connect photoreceptors and bipolars; and amacrine cells receive input from bipolars and direct output to other amacrine cells, bipolars and ganglion cells (Fig.4.4b). These five basic cell types can be further subdivided on the basis of morphology, physiology, and neurotransmitters, reaching a total of 60 for the cat retina.[1] I will not discuss all retinal cell types, but restrict the discussion to photorecep-

[1] For reviews see Barlow and Mollon (1982), Sperling (1983), Kandel and Schwartz (1985), and Masland (1986). The May 1986 issue of *Trends in Neurosciences* is devoted to information processing in the

tors and ganglion cells, and some of their interconnections.

Retinal photoreceptors come in two types: rods and cones. Rods in the cat retina are extremely sensitive, especially in the blue-green part of the spectrum (500 nm) where absorption of one quantum results in observable reactions at the ganglion level (Stryer 1987; Schnapf and Baylor 1987). Cones are less sensitive, having a threshold which is about three log units higher than that of rods, and come in three varieties, each tuned to a slightly different range of wavelengths, having peak sensitivities at 450, 500, and 550 nm. As shown in Fig.4.4b, both rods and cones respond to light absorption by membrane hyperpolarization.[2]

There are on the order of 10^8 rods and 5×10^6 cones, arranged in a regular mosaic (Fig.4.4c), whose composition depends on position in the retina: there are more cones in the fovea than in the periphery, and more rods in the periphery than in the fovea (Sterling 1983; Curcio et al. 1987; Lia et al. 1987). The fact that the optic nerve has only 10^6 axons[3] suggests a high degree of convergence in the photoreceptor–ganglion pathway. And Sterling et al. (1986) do indeed report that approximately 1500 rods converge on a single ganglion cell. They also report a diverging pathway which starts with 1 rod, passes through 2 rod bipolars, 5 amacrine cells, and 8 cone bipolars before it reconverges on 2 ganglion cells.

Given their different sensitivities, one might have expected the two photoreceptors, the rods and cones, to be the first stages of two parallel pathways. It turned out, however, that the rod and cone system already interact at the receptor level: cones receive input from rods via gap junctions, but not vice versa. (Sterling 1983; D'Zmura and Lennie 1986). And although rods and cones synapse to separate populations of bipolar cells, the rod and cone bipolars, their pathways converge again since rod bipolars synapse to amacrine cells which in their turn are connected to cone bipolars. Finally, cone bipolars are connected to ganglion cells. Thus there seem to be two pathways, one rod–cone–cone bipolar–ganglion cell, and the other rod–rod bipolar–amacrine cell–cone bipolar–ganglion cell. Sterling (1983) suggests that during dark adaptation the visual system switches from the first to the second pathway, a switch possibly regulated by certain types of horizontal and amacrine cells.

As mentioned before, the output of ganglion cells is not simply proportional to local light intensity, at least beyond certain light levels. Instead, they typically respond to a small spot of light in a dark field but not to a uniform field, even though the local light intensity is similar. Kuffler (1953) was the first to characterize the receptive fields of cat ganglion cells and distinguish between **ON-** and **OFF-center** cells. The receptive field of an

---

retina.

[2] That is, the membrane potential increases from -25mV to about -60mV. Note that photoreceptors are exceptional in this respect: most sensory receptors depolarize in response to a stimulus.

[3] Perry and Cowey (1985) estimated that there are between 1.4 and 1.8×10^6 ganglion cells, and 3.2×10^6 cones in a monkey retina; Masland (1986) reports that there are 3.5×10^5 ganglion cells in a rabbit retina.

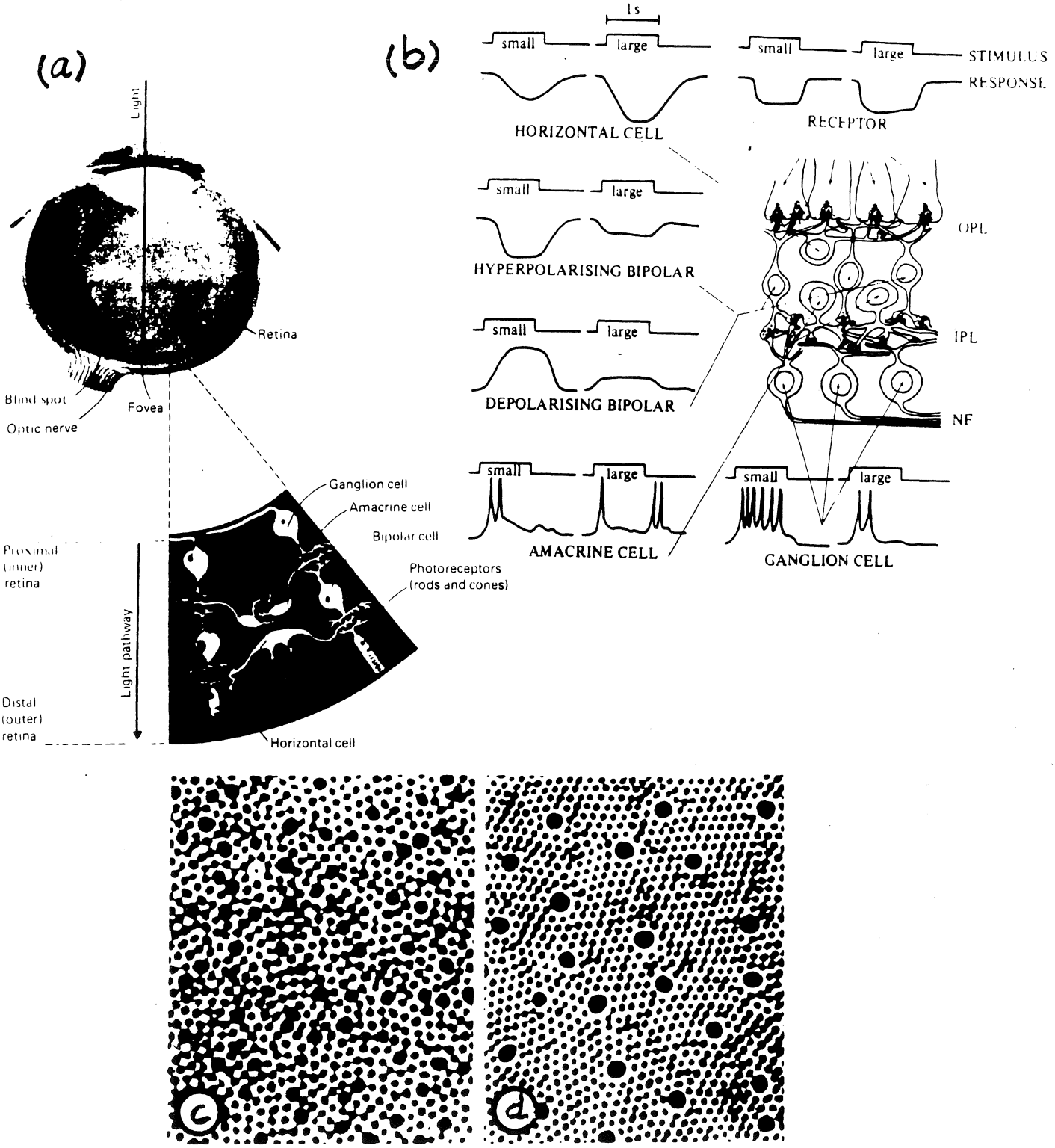


Figure 4.4 (a) Cross section of eye and retina (Kandel and Schwartz 1985). (b) Response of each of the five main retinal cell types to a small and large stimulus. OPL, outer plexiform layer; IPL, inner plexiform layer; NF, nerve fibers forming optic nerve. (Barlow and Mollon 1982). (c) Rod and cone mosaics at level of inner segments in cat retina. Central area; density for cones, $32\times10^3/\text{mm}^2$; for rods, $338\times 10^3/\text{mm}^2$. (d) About 1 mm from central area; density for cones, $11\times 10^3/\text{mm}^2$; for rods, $537\times 10^3/\text{mm}^2$ (Sterling 1983).

ON-center ganglion cell consists of a central region and an annulus surrounding the center; illumination of the center increases output, whereas illumination of the surround decreases output. Off-center cells show the opposite behavior: Illumination of the center decreases output, illumination of the surround increases it.[4] The ON-center ganglion cells are slightly less numerous (Sterling 1983).

In the mid-sixties, a further distinction was drawn between cells having linear spatial summation—X-cells—and cells having nonlinear summation—Y-cells (Enroth-Cugell and Robson 1966). This distinction became less clear-cut when it was shown that X-cells responded linearly only to low contrast stimuli (Enroth-Cugell et al. 1983), and that Y-cell have linear as well as nonlinear components (Shapley and Lennie 1985). A few years later, X- and Y-cells were also shown to differ in time course of their response, the response of X-cells being "sustained" and that of Y-cells "transient." Lennie (1980) criticized this characterization because time course of response depends critically on such factors as light adaptation (the more light-adapted, the more transient the response) and stimulus contrast (transient response is only elicited by high contrast). Fleet et al. (1984) also questioned the distinction between transient and sustained responses, and noted that both X- and Y-cells respond to some extent "transiently."

The functional differences between X- and Y-cells are correlated with morphological differences. X-cells (called beta cells by anatomists) have a medium sized cell body and axon, and a narrow (20–300μm diameter), densely branched, dendritic tree. Y- or alpha cells, on the other hand, have a large cell body and axon, and a wide (180–1000μm diameter), sparsely branched, dendritic tree (Sterling 1983).

The X- and Y-type ganglion cells occur in very different proportions: In the cat, 55% of all ganglion cells are of the X-type, 4% of the Y-type, and the remaining 41% of the W-type. This last type does not seem to have the distinct center-surround organization and appears to be specialized for detecting moving stimuli (Kandel and Schwartz 1985). As always, note that these proportions are species dependent. In the monkey, there are only a few W-like cells, whereas rodents do not have any X-like ganglion cells (Stone and Dreher 1982). The targets of the three types also differ: X-cells project to the LGN, W-cells to the superior colliculus, and Y-cells to both. In other words, already at the level of the retina, there is a large degree of parallelism in the processing of the visual image.

To pursue the similarities and differences between species a little further, what are called X (Y)-cells in cats are called P(M)-cells in monkeys, because of their differing projections to the LGN: P-cells project to the parvocellular layers and M-cells to the magnocellular layers of the dLGN (see next section for more details). Table I summarizes the properties of these

[4] However, according to Allman et al. (1985), a moving spot as far away as 90°, that is, outside the surround, can still decrease a cell's response.

Table I. Properties of Ganglion Cells in the Cat and Monkey (Shapley and Perry 1986).

	Cat X	Cat Y	Monkey P	Monkey M
Relative				
Relative receptive field center size	smaller	larger	smaller	larger
Relative axonal conduction velocity	slower	faster	slower	faster
Dendritic field diameter	smaller	larger	smaller	larger
Absolute				
Non-linear subunit input	No	Yes	No	No M_X Yes M_Y
Axonal conduction velocity	18 m s^{-1}	50 m s^{-1}	13 m s^{-1}	21 m s^{-1}
Colour opponency	No	No	Yes	No
Central receptive field size	0.1°	0.3°	0.01°	0.06°
Projection pattern to brain	Unbranched	Branched	Unbranched	Unbranched
Peak cell density mm^{-2}	4500	200	29 600	3700
Contrast gain	High	High	Low	High

classes of ganglion cells. Although there are obvious similarities between X and P-cells, and Y and M-cells, the differences between these classes are just as obvious. On the basis of this and related evidence, Shapley and Perry (1986) suggest that M-cells can be divided into M_X and M_Y classes. In addition to the data shown in Table I, Kaplan and Shapley (1986) showed that M-cells are more sensitive to contrast than P-cells. They conjecture that the M-cell pathway operates at low contrast and low to intermediate spatial frequencies (large dendritic and receptive field), and that the P-cell pathway takes over when M-cells become saturated.

As a summary of the above discussion, let us consider the results of a detailed electronmicroscopic study of the circuitry of the beta (or X-type) ganglion cell (Sterling 1983; Sterling et al. 1986). As illustrated in Fig.4.5, both the ON- and OFF-center beta cells receive input from two bipolars, which presumably already have a center-surround organization. Each pair of bipolars seems to function in a "push-pull" manner, one of the bipolars exciting the beta cell and the other inhibiting it. This might account for the wide range of spike frequencies observed in beta cells (from 0 to 700 spikes/s). Also shown in Fig.4.5 is the alternative pathway used after dark adaptation, in which beta-cells are excited via the rod–rod bipolar–amacrine–cone bipolar–beta pathway. This switch in pathways causes the receptive fields of the beta-cells to lose their antagonistic center-surround organization since the cone bipolars are now indirectly excited or inhibited at the axonal level instead of at the dendritic level, the supposed origin of the antagonism between center and surround.

Given the many different types of ganglion cells, one wonders about their role in visual processing. Some suggest that neighboring, parallel rows of ON- and OFF-center cells form

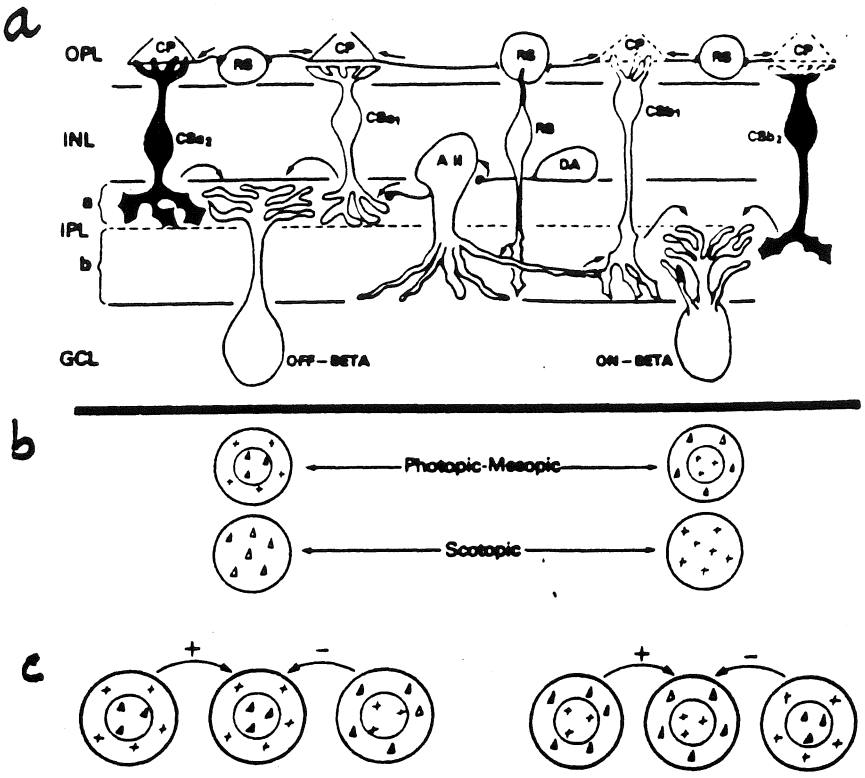


Figure 4.5 (a) Schema of microcircuitry of retinal beta cells. (b) Receptive fields of beta cell under light (phototopic-mesotopic) and dark (scotopic) conditions. (c) Receptive field of beta cells is generated by the "pushing" and "pulling" of an excitatory and an inhibitory bipolar. OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer; CP, cone pedicle; RS, rod spherule; CB, cone bipolar; RB, rod bipolar; AII, AII amacrine; DA, dopamine amacrine (Sterling 1983).

the basis for orientation detection (Marr 1982; Heggelund and Moors 1983). And indeed ON- and OFF-center cells converge at the cortical level (Swindale 1986). If both cell types are thus combined, one would expect that blockage of say the ON-center pathway would impair vision dramatically. This appears not to be the case. Of all the visual functions tested, only contrast sensitivity and the detection of light increment are significantly affected after chemically[5] blocking the ON-center pathway (Schiller 1982; Schiller et al. 1986). The ability to discriminate between gratings with different orientations is not impaired. This suggests that orientation can be computed within the OFF system itself, without input from the ON system. Poggio (see Ullman 1986) proposed just such a structure: two parallel, nonoverlapping rows of OFF-cells converging on a cortical cell. If an edge falls just in between these two rows, one will be active (dark side of the edge), the other inactive (light side of the edge). And the cortical cell combining the first row with an "AND" and the second with a

[5] by means of the glutamic acid analogue 2-amino-4-phosphonobutyric acid.

"NOT-AND," will thus signal the presence of an edge whose orientation coincides with that of the two rows. Of course, a complementary edge detector can be built with ON-center cells. Note that the displacement between the two rows could be used to encode spatial frequency: the larger the displacement the larger the spatial frequency it is designed to detect. But, of course, the results of blocking the ON-center pathway do not exclude the possibility that orientation is detected by converging rows of ON- and OFF-cells, they merely show that there are other means.

Sherman (1979) suggested that Y-cells are primary involved in basic shape analysis while X-cells serve to add more details. This suggestion was motivated by three observations. First, blurring or defocusing, which eliminates high spatial frequencies from the image, does not affect recognition whereas the opposite, elimination of low spatial frequencies, does. Second, at spatial frequencies below .5 cycles/degree, X-cells are more sensitive to contrast than Y-cells; their contrast sensitivity for higher frequencies being similar. And third, lesions in the striate cortex of cats (which block the X pathway but not the Y pathway since the latter proceeds through extrastriate and striate areas) do not impair shape recognition.

Noting that cortical cells are very quiet until presented with appropriate stimuli, Lennie (1980) proposed that Y-like cells regulate the sensitivity of cortical cells. Without any stimulus, Y-like cells inhibit cortical cells, but when a stimulus is present the inhibition is removed. Lennie calculated that, in the cat, there are approximately 950 hypercolumns[6] for each eye, and that each column is served by no more than 4 Y-like cells. Interestingly, the axons of Y-cells conduct spikes faster than those of X-cells, 30–40 m/s versus 18-25 m/s (Stone and Dreher 1982). Since the same is true for the X-like and Y-like cells of the dLGN, signals traveling along fast-conducting axons take about 6ms to go from retina to cortex as opposed to 10ms for slow-conducting axons (Lennie 1980). In other words, Y-like cells could disinhibit cortical cells before the X-like signals arrive. In fact, Maffei (1985) suggests that cortical complex cells which receive Y-like input control the activity and function of the simple cells. This functional role of the complex cells also explains the observation that complex cells have spontaneous activity whereas simple cells do not: The spontaneous spiking of complex cells inhibits simple cells.

4.2.2 Dorsal lateral geniculate nucleus

The dorsal lateral geniculate nucleus (dLGN) forms part of the thalamus, a collection of nuclei in the middle of the brain. In Old World monkeys, apes, and man, dLGN consists of layers (laminae) of cell bodies, each layer containing a retinotopic representation of the contralateral visual hemifield. The central 15–20° of the visual field are mapped onto six

[6] Small patches in striate cortex thought to encode all possible orientations at particular retinal positions; see section 4.2.3 and Fig.4.10.

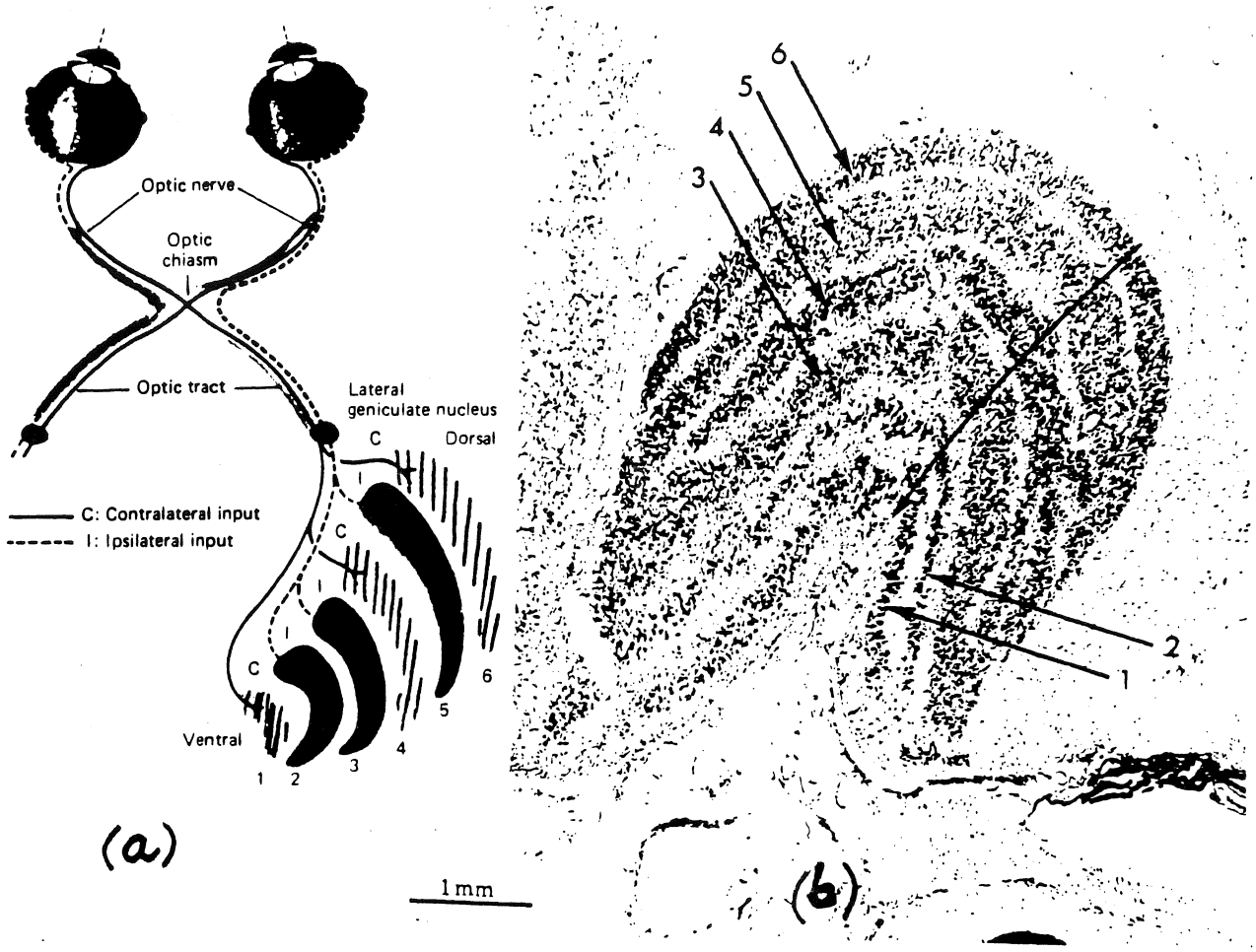


Figure 4.6 (a) Projection of the retinas onto the LGN. Layers 1 and 2 are magnocellular, the remaining parvocellular (Kandel and Schwartz 1985). (b) Section of the right LGN of an adult macaque, stained for cell bodies. Each of the six layers contains a map of the left visual hemifield. The six maps are in register such that the neurons along the arrow all respond to the same point in the visual field (Hubel and Wiesel 1979).

layers, the remaining part onto four layers. As shown in Fig.4.6, half the laminae receive their input from the ipsilateral eye, the other half from the contralateral one. Interestingly, this lamination is only prevalent in animals having substantial binocular visual fields, other animals (e.g., rat, squirrel) have fewer and less clearly separated laminae (Lennie 1980).

The six layers found in certain primates are divided into four dorsal (or parvocellular for small cells) and two ventral (or magnocellular for large cells) layers. The magnocellular and parvocellular layers also differ functionally, the differences resulting from their being targets of different populations of retinal ganglion cells, X- (Y-) cells projecting to the parvocellular (magnocellular) layers. Thus cells in parvocellular layers are color-sensitive

and have low contrast sensitivity, whereas cells in magnocellular layers are broadly tuned and have a contrast gain approximately ten times higher than that of cells in parvocellular layers. Kaplan and Shapley (1986) suggest that these two classes are adapted to function at different light intensities: cells with high contrast gain operate at low light levels, and as they become saturated with increasing light intensity, cells with low contrast sensitivity take over. This is reminiscent of the switching between the pathways subserving scotopic (low light levels) and phototopic (high light levels) vision in the retina; it suggests that these pathways remain parallel at least until the dLGN.

Besides the differences between parvo- and magnocellular layers, there are differences within the four parvocellular layers. Schiller and Malpeli (1978) found that the two dorsal layers (5 and 6) have predominantly ON-center cells (93.7% with red, green, and white test spots), while the two ventral layers (3 and 4) have mainly OFF-center cells (81.6%). Curiously, when using blue spots as test stimuli, ON-center cells are found in layers 3 and 4, but rarely in 5 and 6, where the other ON-center cells are found (OFF-center blue cells are already absent in the retina). The presence of OFF-center cells in layer 3 is further corroborated by the fact that blocking layer 3 has the same effect on cortical cells as blocking the OFF-center pathway at the retinal level (Schiller 1982). In both cases, the dark edge response of cells in the striate cortex is reduced or even eliminated.

As mentioned before, cortical projections of the dLGN are species dependent. For instance, in cats dLGN projects to both striate (17) and extrastriate (18,19) cortex, whereas its projection is limited to striate cortex in macaque monkeys (Rodieck 1979; Lennie 1980). Within these pathways exist subpathways, going from particular geniculate layers (and therefore presumably particular cell types) to particular cortical layers. Consider the simplest example, that of the monkey. Parvocellular layers (X-like) terminate mainly in layer IVcβ but also in IVa and VI. Magnocellular layers (Y-like) project to IVcα, IVcβ, and VI (Fig.4.13). The situation in the cat is more complex, but if we consider only the projections to striate cortex, we find a similar pattern: X-like cells terminate in layers IVc and VI, Y-like cells in IVab and VI.

Receptive fields of geniculate cells do not differ substantially from those of retinal ganglion cells, i.e., they are of the ON/OFF antagonistic center-surround type (Hubel and Wiesel 1977; Lennie 1980; Kandel and Schwartz 1985; Shapley and Lennie 1985). This does not mean that the dLGN merely relays signals from retina to cortex. If this were the case one would expect the majority of the synapses in the dLGN to have originated from retinal ganglion cells, which, it turns out, is not the case (Sherman and Koch 1985). Only between 10 and 20% of the synapses belong to retinal ganglion cells, and, in fact, around 50% of all synapses come from layer VI in the visual cortex. The remaining are thought to have a local origin and to be inhibitory. In addition, biophysical studies have shown that the behavior of geniculate neurons can be changed such that it does not relate to incoming retinal signals

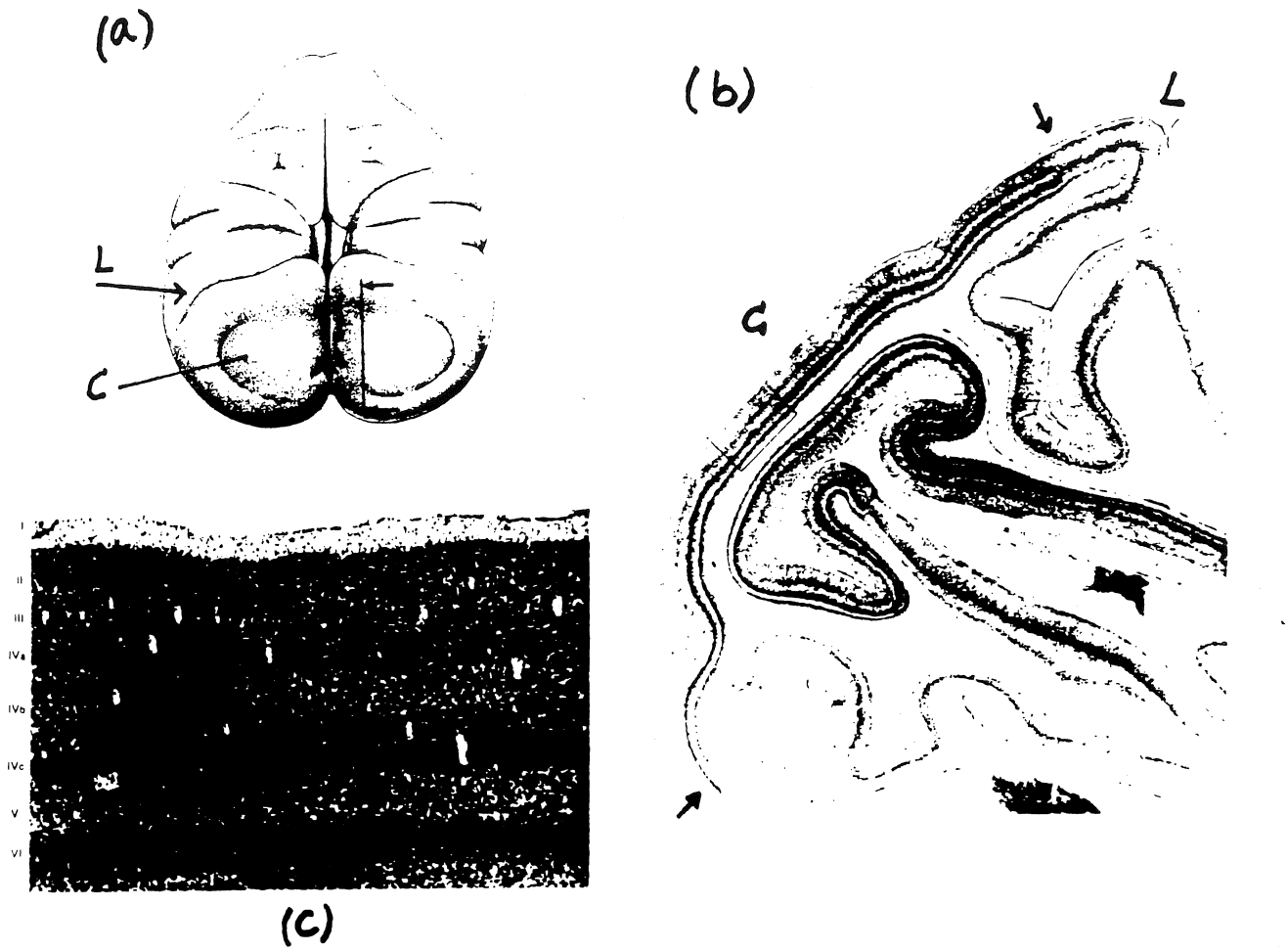


Figure 4.7 (a) Occipital lobe of macaque seen from behind and above, demarcated in front by the lunate sulcus (L). It consists mainly of striate cortex (shaded). If followed medially, the striate cortex curves around and continues underneath the convexity (C) of the occipital lobe. (b) Cross section of the striate cortex along vertical line in (a). Note the convexity and the fold underneath it. Arrows indicate the striate–extrastriate border. (c) Outer 2mm of striate cortex, magnification of rectangle in b (Hubel and Wiesel 1979).

but assumes a more autonomous character. This led Sherman and Koch (1985) to propose that dLGN modulates signals from retina to cortex as a function of the behavioral state of the animal.

4.2.3 Visual areas in the cortex

The **primary visual cortex**, also known as striate cortex, area 17, or V1, occupies most of the occipital lobes in primates (Fig.4.7a). Fig.4.7b illustrates the structure of the cortex which is about 2mm thick and consists of alternating layers of cells and processes. It also indicates that there is a histological difference between striate and extrastriate cortex, the

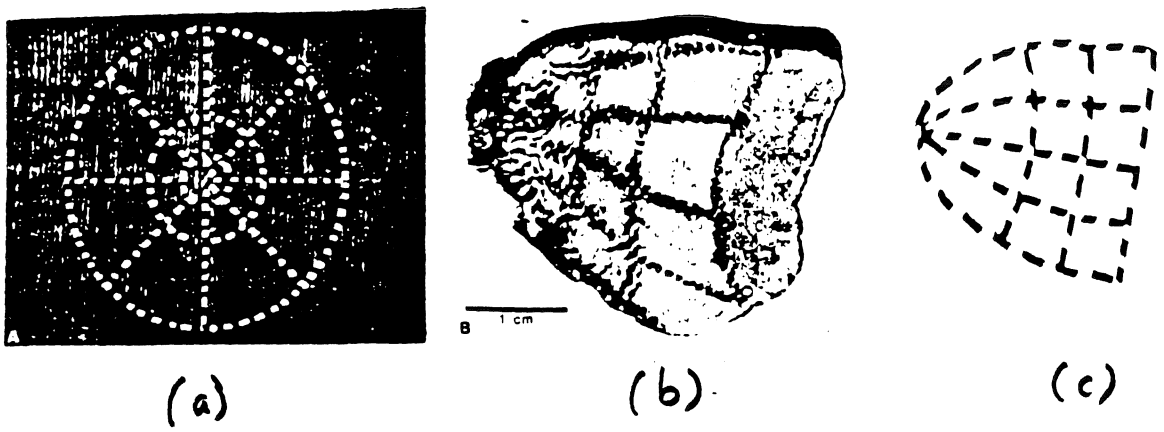


Figure 4.8 (a) Visual stimulus consisting of flickering squares. (b) Pattern of activity in striate cortex while watching stimulus of (a). Activity is measured by radioactive deoxyglucose uptake (Tootell et al. 1982). (c) Visual stimulus of (a) transformed by $\log(z+0.3)$. Except for the periphery, this map corresponds well with the experimental data (Schwartz 1984).

former being more clearly striped when staining cell bodies using the Nissl method. The most common numbering of the layers is shown in Fig.4.7c.

The striate cortex of each hemisphere contains a retinotopic map of the complete contralateral hemifield plus about 2° of the ipsilateral hemifield along the vertical meridian. The mapping between visual field and striate cortex in a macaque monkey is shown in Fig.4.8 (Tootell et al. 1982). At the top we see the visual stimulus: three concentric rings and eight rays emanating from the locus to be foveated by the monkey. The rings and rays are composed of small squares that flicker at 3 Hz. At the bottom we see the resulting pattern of activation in the striate cortex at the level of layer IVb,c.[7] One way to describe this mapping is by its "magnification factor": the ratio of the distance between two nearby points on the cortex and the distance between the two corresponding points on the retina. In other words, the magnification factor is the derivative of the function that maps the retina onto the cortex. The value of this derivative has been found to be inversely proportional to eccentricity, i.e., it decreases as one moves away from the fovea, suggesting that the map itself is logarithmic. Schwartz (1977ab, 1980, 1984; see also Mallot 1985) proposed the complex logarithm $\log(z+a)$, where points z on the retina are given by their polar coordinates $z=(r,\phi)$, and a is a small number, typically less than five. For small z, $\log(z+a) \approx \ln a + z/a$, a linear map. For large z, $\log(z+a)$ behaves like $\log z$, which maps polar coordinates (r,ϕ) into cartesian coordinates $(\log r, \phi)$. Thus, concentric rings (constant r) are mapped onto vertical lines, and rays (constant ϕ) are mapped onto horizontal lines. Fig.4.8 juxtaposes the function $\log(z+.3)$ and the cortical activity map obtained by Tootell et al.; the two patterns match

[7] Measured by glucose utilization in the 2-[^{14}C]deoxy-D-glucose method.

reasonably well except for the peripheral visual field.

Cells in striate cortex differ fundamentally from retinal ganglion and geniculate cells. They often respond to stimulation from both eyes instead of from one eye only, and they are most responsive to line segments as opposed to circular spots. Even though many cells in striate cortex respond to visual stimuli from both eyes and are tuned to horizontal disparity (Poggio and Poggio 1984; Poggio et al. 1985), they still respond to monocular stimulation. Interestingly, this response is not symmetric: in many cases, one eye will elicit a much stronger response than the other (Hubel and Wiesel 1968, 1970, 1977). Accordingly, cells have been characterized by their **ocular dominance**, some being dominated by the left eye, others by the right eye. In some species, cells with the same ocular dominance are grouped in parallel bands, between 350 and 500μm wide, running perpendicular to the striate-extrastriate border (Fig.4.9).[8] This pattern is clearly visible in Old Word monkeys, apes and humans, but conspicuously absent in New World monkeys (Hendrickson 1985). The ocular dominance columns are found in layers IVa,c, while layers II and III show rows of blobs, the spacing between blobs suggesting that they are the continuation of the layer IV bands. A cytochrome oxydase stain, another measure of cellular activity, reveals the same rows of blobs in layers II and III, but also in layers IVa and IVc (Hendrickson 1985). Since this is the case for both Old and New World monkeys, it seems that the rows of blobs and the ocular dominance columns are not related.[9]

In addition to combining inputs from both eyes, cells in striate cortex analyze orientation in the visual field, at least that is suggested by their response to lines and edges at different orientations.[10] In their studies of receptive field properties, Hubel and Wiesel (1968, 1977) distinguished between **simple**, **complex**, and **hypercomplex** cells. Simple cells respond quite selectively to the position and orientation of an edge. Deviations as small as 10° from the preferred orientation can silence a simple cell. Compared with simple cells, complex cells have a larger receptive field, are insensitive to the position of the stimulus within that field, and are slightly less selective for orientation. Hypercomplex cells differ from complex cells in that they are sensitive to the length of the edge: extending the edge in one or both directions can completely inhibit a hypercomplex cell (Although in widespread use, Wiesel and Gilbert (1986) suggest that the term "hypercomplex" be abandoned since simple cells also show end inhibition). This "end-inhibition" in hypercomplex cells can be selectively

[8] In prefrontal cortex, ipsilateral projections from the parietal lobe alternate with callosal projections from the contralateral frontal lobe (Goldman-Rakic and Schwartz 1982).

[9] Interestingly, antiserum to calbindin, an calcium-binding protein, stains cortex except at the cytochrome oxydase rich blobs (Celio et al. 1986), and intracellular calcium decreases the affinity of receptors for GABA (Inoue et al. 1986), an inhibitory neurotransmitter thought to be involved in generating orientation selectivity (Wolf et al. 1986).

[10] However, cells in areas 17 and 18 also respond to auditory stimuli (Fishman and Michael 1973), and proprioceptive signals from extraocular muscles (Buisseret and Maffei 1977; Buisseret and Singer 1983).

Figure 4.9 (a) Ocular dominance columns at layer IVC in striate cortex of macaque monkey. Dashed line indicates the striate–extrastriate border. (b) Fingerprint of human index finger at same scale (Hubel and Wiesel 1977).

suppressed by inactivating layer VI with the inhibitory transmitter γ-aminobutyric acid or GABA (Bolz and Gilbert 1986).

Although it is satisfying to neatly order cortical cells and put them in discrete categories, it is not clear to what extent these categories are discrete as opposed to being the two extremes along a continuum of receptive field properties (Shapley and Lennie 1985). Consider, for example, the response to a drifting sine grating. The response of simple cells typically is modulated: its mean spike rate increases and decreases as the light and dark bars of the grating pass over its receptive field. Complex cells respond with an increased spike rate, a response which becomes modulated only for very low spatial frequencies. If we map cortical cells as a function of the modulation of their response to drifting sine gratings, we obtain a continuous distribution with simple and complex cells overlapping. While this is true for the cat, the distribution in the monkey is bimodal, one peak corresponding to simple cells, the other to complex cells. "Whether this difference between species is real remains to be investigated. The question of the discreteness of the simple and complex categories therefore remains troubling, and requires further investigation with quantitative techniques" (Shapley and Lennie 1985).

As their response to moving gratings indicates, some cortical cells are sensitive to temporal frequency as well as spatial frequency (Tolhurst and Movshon 1975), and their receptive fields should thus be characterized in space-time instead of in space only. On the basis of psychophysical experiments, Burr and Ross (1986) constructed the spatio-temporal receptive field of human motion detectors. They found that the temporal component of receptive fields is very similar, having a summation period of 100ms and an optimal response at a temporal frequency of 10 Hz. Emerson et al. (1985) reported similar receptive fields for motion detectors in the cat's striate cortex.

Given that cells in the striate cortex are quite sensitive and selective for orientation and that the striate cortex contains a retinotopic map of the visual field, the question arises how the two are fitted together, that is, how are the orientation-selective cells ordered on the cortical surface? One could, for example, surmise that different cortical layers are selective for different orientations, and that moving along the cortical surface, one encounters cells tuned to slightly different locations in the visual field. This simple arrangement turns out not to be true. Instead, one finds that cells underneath a certain cortical position share the same orientation preference (Hubel and Wiesel 1968, 1977; Mountcastle 1978). In other words, if one records from an electrode as it proceeds through the cortex in a direction perpendicular to the cortical surface, then one will find that all cells respond maximally to the same orientation (except for cells in layers IVB and IVC which are broadly tuned). However, recently Bauer and coworkers (Bauer 1982; Bauer et al. 1983) found an abrupt shift in orientation preference at the IV–V border. On average, this shift was 55°, with 70% of the electrode tracks showing a shift between 45 and 90°. If one inserts the electrode at a small angle, almost parallel to the surface, one finds that optimal orientation changes slowly, on average about 10° for every 25–50μm; less often it does not change or it changes abruptly.

Hubel and Wiesel (1977) proposed that the ocular dominance and the orientation systems are independent and locally orthogonal: ocular dominance alternates in one direction, orientation preference alternates in the perpendicular direction. Combining this with the retinotopic organization of V1 leads to the concept of a hypercolumn (Fig.4.10a): V1 is divided into hypercolumns, with each hypercolumn representing all orientations and ocular dominance values at a certain location in the visual field. A hypercolumn is about 2mm deep, i.e., it spans the cortex, and has a square cross-section with sides of 1mm. In one direction, orientation changes continuously over 180°, and in the other direction, ocular dominance changes from dominance by one eye to dominance by the other. Braitenberg and Braitenberg (1979; modified by Dow and Bauer (1984) and Goetz (1987)) proposed an alternative organization for the orientation columns (Fig.4.10b). Unlike the hypercolumns, their centric organization can account for sudden shifts in orientation selectivity (halfway along the oblique line in Fig.4.10b there is a sudden shift).

Several attempts have been made to test whether striate cortex is divided into something resembling hypercolumns. In one experiment, activity was assessed by radioactive deoxyglucose uptake while the animal was watching a vertical grating. As is clear from Fig.4.10b, the pattern of cortical activity is very intricate and not as regular as one would expect on the basis of a hypercolumn organization. Unfortunately it is not quite clear what distinguishes areas labeled by the radioactive deoxyglucose from the unlabeled ones. Noting the discrepancy between their electrophysiological results which show a 55° shift in orientation tuning as the electrode crosses the IV–V border, and the deoxyglucose uptake experiments which show no signs of such a shift since cortical columns are labeled in all six layers (Fig.4.10c), Bauer and coworkers (1983) point out that deoxyglucose experiments do not discriminate between inhibitory and excitatory activity, and that one should therefore be cautious when inferring orientation preference from deoxyglucose patterns (Shapley and Lennie 1985). And in fact there is increasing evidence that orientation selectivity in simple cells results from cortical inhibitory activity (Maffei 1985; Wiesel and Gilbert 1986).

Using voltage-sensitive dyes, Blasdel and Salama (1986) extended the results obtained with deoxyglucose labeling. In this technique, the cortex is stained with a dye which emits fluorescence signals as a function of membrane potentials, including action potentials. One can literally see (after some computer processing of the cortical image) which parts of the cortex are excited. This makes it possible to study the same piece of cortex while the animal watches a vertical grating, and then repeat the procedure with a horizontal grating. Blasdel and Salama (1986) found that the cortical surface consists of small modules (0.5–1mm wide) separated by so-called fractures.[11] Within a module, orientation selectivity changes smoothly in one direction but hardly in the others (Fig.4.11a). Fractures, in contrast, are

[11] But see Grinvald et al. (1986) for a critique.

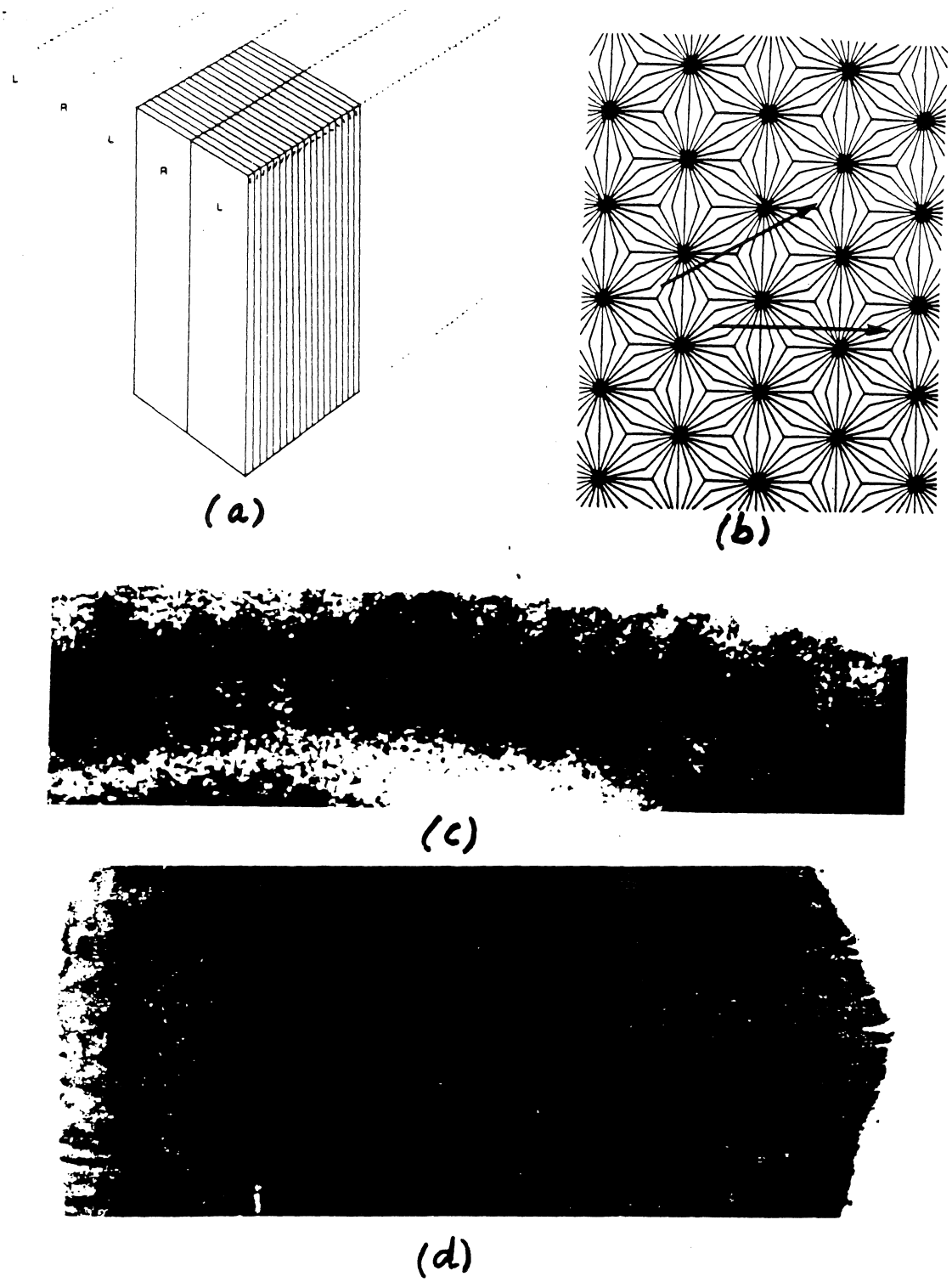


Figure 4.10 (a) Hypercolumn, a proposed building block of striate cortex. In one direction, ocular dominance changes, in the other direction, orientation selectivity changes. (b) Centric organization of orientation selectivity (Braitenberg and Braitenberg 1979). (c) Cross section of striate cortex showing dark areas with high radioactive deoxyglucose uptake as a result of the monkey's watching of a vertical grating. Layer IV is stained throughout. (d) Topview of the same piece of cortex. Dark band represents continuously labeled layer IV (Hubel and Wiesel 1979).

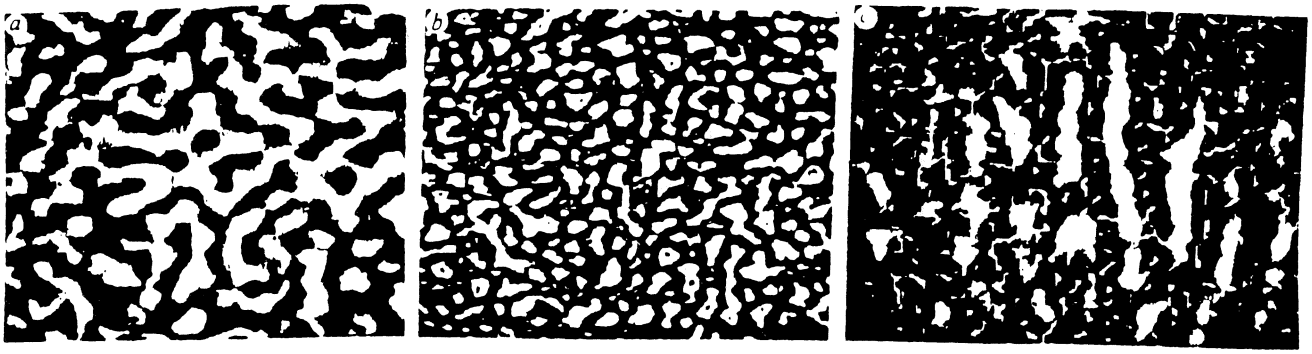


Figure 4.11 (a) Top view of striate cortex indicating orientation preferences. Each shade of gray corresponds to selectivity for a particular orientation. (b) Gradient of orientation preference. In this photocopy of the original color-coded image, darker areas indicate steeper gradients. (c) Ocular dominance columns overlayed with fractures (shifts in orientation preference of more than 45°). Note that the fractures tend to run along the ocular dominance columns. Horizontal width of each image corresponds to 8mm on the cortical surface (Blasdel and Salama 1986).

characterized by sudden shifts (more than 45° in 40μm) in orientation selectivity (Fig.4.11b). And as can be seen in Fig.4.11c, fractures tend to run along ocular dominance columns or cross them at right angles. Thus, the fractures lie in the same location as the cytochrome oxydase rich blobs whose cells show no preference for orientation (Hendrickson 1985), but are sensitive to color. An attractive interpretation would be that each module represents a particular point in visual space, analyzing it in terms of depth and orientation of edges. In other words, a module corresponds to a hypercolumn. In general, however, modules do not represent all possible orientations or ocular dominance stripes. But it is not clear, a priori, that this would be necessary, as it is quite easy to envision a system that encodes orientation by means of the relative response of populations of cells broadly tuned for horizontal and vertical direction (or some other subset of all possible directions). And indeed, in the fovea, the horizontal and vertical orientations are represented disproportionately (Mansfield 1974). A more fundamental question is whether the striate cortex is really designed to encode orientations of line segments. Perhaps orientation selectivity is an epiphenomenon, a consequence of the encoding of shape by measuring the deformation component of optic flow (Koenderink 1986).

Area V2 (part of area 18) forms a ring around striate cortex (V1) and is organized retinotopically. In monkeys, V2 receives its input from V1, whereas, in cats, it also receives input from dLGN. Receptive fields of cells in V2 are similar to the ones in V1, although there seem to be fewer simple cells (Hubel and Wiesel 1970; Van Essen 1979, 1985). In addition, there are "binocular depth cells" tuned to binocular disparity (Hubel and Wiesel 1970), and

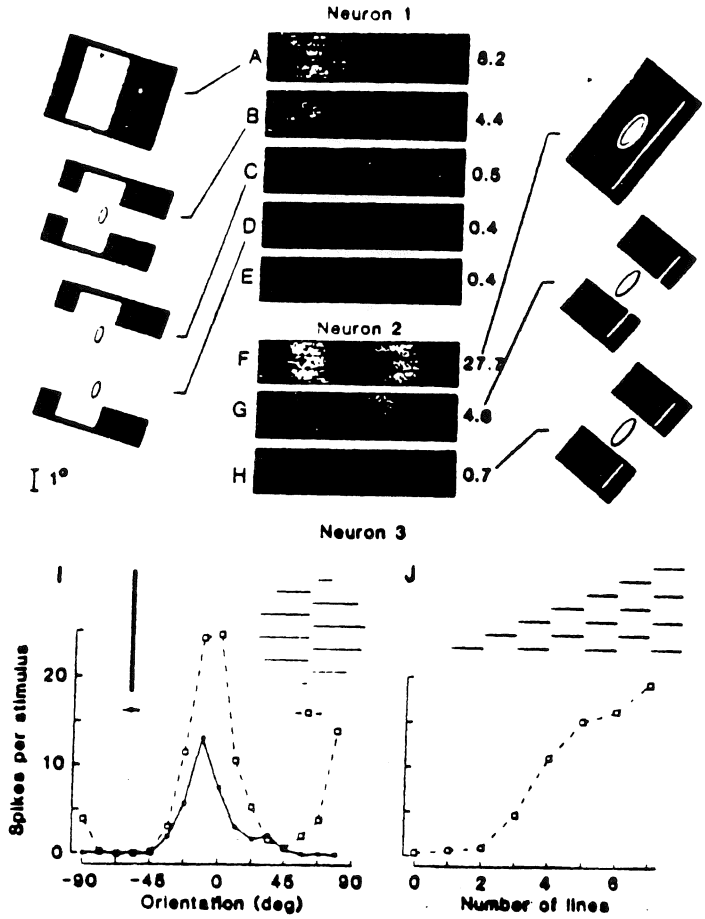


Figure 4.12 **Response to edges, bars, and stimuli producing illusory contours, of area 18 neurons of rhesus monkey. The stimuli were moved back and forth across the receptive fields (neuron 1, 1° at 1 Hz; neurons 2 and 3, 2° at 1 Hz). Each was presented 8 (I), 16 (J), or 24 (A through H) times. Ellipses indicate the receptive fields obtained with a simple edge or bar; the cross in A and F indicates the fixation point; numbers on the right are mean number of spikes per stimulus cycle. Neuron 1 responds to a light edge (A) and to a subjective edge (B), but not to either half (C and D). Neuron 2 responds to a bar of light (F) and to a subjective continuation of two half bars (G), but not when the half bars are slightly changed such as to preclude their subjective continuation (H). Neuron 3 responds to a bar but also, and even better, to a line defined by abutting gratings (I); and its response is a function of the number of abutting lines (J) (Von der Heydt et al. 1984).**

cells tuned to different directions of motion in three-dimensional space (Cynader and Regan 1978). Von der Heydt et al. (1984) found cells responsive to so-called illusory or subjective contours (Kanizsa 1976). The responses of some of these cells (Fig.4.12, neurons 1 and 2) could be explained by local excitatory connections between cells with the same optimal orientation (Gilbert 1983, 1985), filling in the gaps as it were. But for other illusions, such

as virtual lines defined by endpoints of lines (Fig.4.12, neuron 3), this explanation obviously does not apply, and although local connections between cells whose preferred directions are orthogonal have been found (Matsubara et al. 1985), these connections are thought to be inhibitory (Gilbert 1985).

To learn more about the functional organization of V2, Tootell et al. (1983) stained that region for cytochrome oxydase, a mitochondrial enzyme indicative of high metabolic activity. The resulting pattern is shown in Fig.4.13b. As mentioned before, V1 contains blobs with high metabolic activity; V2 contains alternating wide and narrow stripes, most clearly visible in layers IV and V, and less so in the other layers, suggesting a columnar organization. In squirrel monkeys, the stripes are about 400-700μm wide and approximately one millimeter apart. Part of the stripes have a high degree of myelination, perhaps due to an afferent input. Shipp and Zeki (1985) reported that the narrow stripes project to area V4, and the wide stripes to V5 (also known as MT). In accordance with the known functions of these two areas, orientation selective cells are found exclusively in the wide stripes and the interstripes, whereas color selective cells are encountered in the narrow stripes (Hubel and Livingstone 1985). Fig.4.13 summarizes the interconnections of cortical areas V1, V2, V4 and V5.

Area V3 is located around V2, and split into two regions. The dorsal region receives input from V2 and V1, although the latter is restricted to the lower visual field. The more ventrally located region receives input from V2 and is restricted to the upper hemifield; this region is also referred to as VP, and considered by some to be a separate area (Desimone et al. 1985; Gattass et al. 1985). Cells in VP are selective for color, orientation, and disparity (Van Essen and Maunsell 1983). V3 projects to V4.

Area V4 receives input from V2 and V3 and projects to IT, the inferior temporal cortex. Its topographical organization is not clear: While some authors report multiple representations of the visual field (Zeki 1978), others show evidence of a single but crude representation (Desimone et al. 1985; Gattass et al. 1985). Depending on eccentricity, receptive fields in V4 are between 4 (at 1°) and 6 (at 3°) times larger than in V1. V4 cells are selective for orientation and spatial frequency, having a sensitivity comparable to that of V1 cells. And just as in V1, cells in V4 can be characterized as being more like simple or like complex cells. The majority of cells in V4 are sensitive to color, having spectral bandwiths similar to color-opponent cells in retina and dLGN. Interestingly, V4 cells also respond to white light; the average response to white light being 60% of the maximal response, quite unlike the color-sensitive cells in striate cortex. It may be that the computations in V4 result in color constancy, the ability to perceive the same color regardless of illumination (Desimone et al. 1985; Wiesel and Gilbert 1986).

Area IT, for inferior temporal cortex (coincides with architectonic area TE; Desimone and Gross 1979), is involved in object recognition as is clear from lesion experiments (Dean

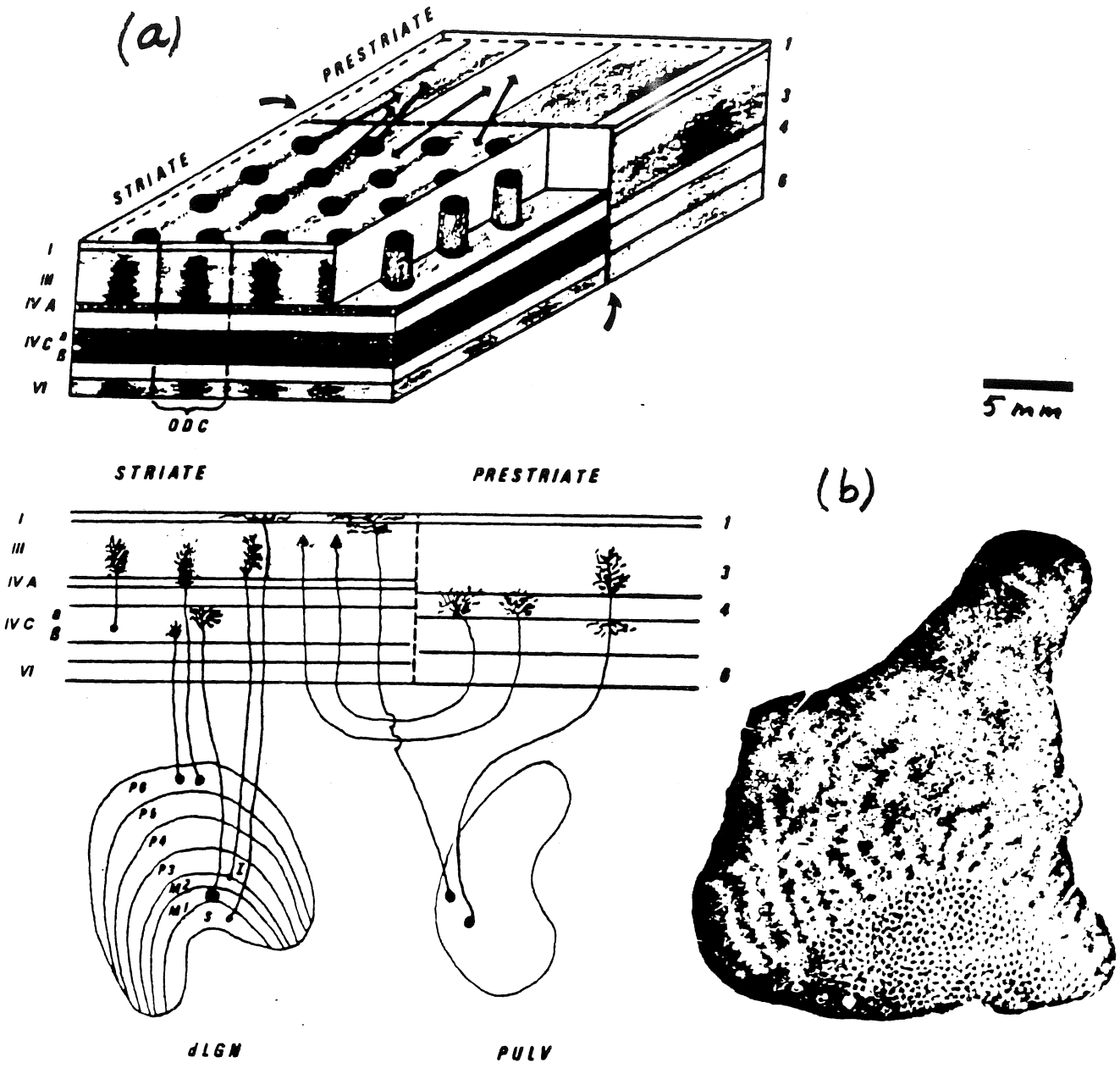


Figure 4.13 (a) Schema of interconnections between dLGN, pulvinar, V1 (striate), and V2 (prestriate). Blobs in V1 are cytochrome oxydase rich. Parvocellular layers of dLGN project to IVCα (and branch to III) and IVCβ (synapse to spiny stellate cells which project to III). The relationship between the pulvinar projections to striate layers I and II and the cytochrome oxydase rich blobs is not known. Prestriate cortex has alternating narrow and wide stripes of high cytochrome oxydase content. The striate blobs connect reciprocally with the narrow stripes; while the striate interblob system is connected reciprocally to the bands with low cytochrome oxydase content. (Hendrickson 1985). (b) Flat-mounted section of layer III from the lateral surface of squirrel monkey cortex stained for cytochrome oxydase. Anterior in the brain is toward the top, dorsal is toward the right. The dotted region is central V1; the adjoining ring marked by stripes is central V2 (Tootell et al. 1983).

1982).[12] If IT is removed after an animal has learned some discrimination task, its performance deteriorates considerably on the same task and it takes much longer to learn a new task (Gross 1978). This deficit is not due to a loss of acuity or related visual functions, and is restricted to tasks requiring visual discrimination. IT seems to be the last stage in the object recognition pathway that is purely visual; the areas it projects to—e.g., premotor, prefrontal, and parahippocampal cortex, and amygdala—integrate vision with the other senses, for example, with hearing or touch.

Receptive field properties of IT cells differ qualitatively from those of cells from areas V1, V2, V3 and V4. First, receptive fields are much larger having a median size of $26° \times 26°$,[13] and almost always include the center of gaze and extend into both visual hemifields (Gross et al. 1969; Desimone and Gross 1979; Richmond et al. 1983; Desimone et al. 1984; Moran and Desimone 1985). The distribution of receptive field size is not completely random: the most anterior part and the dorsal part (in particular the floor of the superior temporal sulcus) of IT have larger fields, upto $60° \times 60°$. And nearby cells tend to have similar receptive field sizes (Desimone and Gross 1979). Having large receptive fields is at once advantageous and problematic. It is advantageous since the shape of an object is now separated from its location, thus laying the foundation for position invariance. It is problematic because now a large number of objects can be present within a single receptive field. Fuster and Jervey (1981) and Moran and Desimone (1985) showed that this problem is solved through selective attention: If the monkey attended to one stimulus and ignored another one (both inside the receptive field), the response to the ignored stimulus decreased. They also found this effect in area V4, but not in V1 (striate cortex).

The second difference between receptive fields of IT neurons and neurons from areas preceding IT, concerns the stimuli that trigger maximal response. Instead of being tuned to simple stimuli such as bars and edges at particular orientations and lengths, IT neurons require complex stimuli and even real 3-D objects (Desimone and Gross 1979; Desimone et al., 1984, 1985). IT neurons are also less selective: A large fraction (41%, or 44 out of 110) responded well to all stimuli tested (Desimone et al., 1984), and 61% of the more selective neurons still responded weakly to every stimulus tested. Only very few neurons (5 out of 151) were selective for particular objects; 2 were apparently tuned to hands, 3 to faces. A large proportion (37%) of face selective cells were found in the STS, and most cells were either tuned to the full face or to a profile.

The face selective cells in the STS have been the focus of much recent work (Perrett et al. 1982, 1984, 1985; Rolls 1984, 1987; Baylis et al. 1985; Kendrick and Baldwin 1987).

[12] See Bachevalier et al. (1985) and Mishkin and Appenzeller (1987) for the effects of lesions in subcortical systems of the temporal lobe.

[13] Compare with $1° \times 1°$ for excitatory receptive fields including the center of gaze in V4 (Desimone et al. 1985).

Perrett et al. (1985) found that 8–9% of neurons in the fundus of the anterior STS of the macaque monkey increased their firing rate in response to faces but not to other visual stimuli. 63% of neurons (115 out of 182) were sensitive to head orientation. Of these, 48 were most responsive to full faces and 39 to faces in profile. Note first, that these neurons were broadly tuned to orientation, and, second, that no neurons were found that responded most to faces seen at a 45° angle. The remaining 37% responded equally well to frontal and side views of faces. Although no significant differences in response latency were found, the orientation-insensitive cells might receive input from several orientation-sensitive cells, thereby progressing from a description that depends on viewpoint to one that does not. The large majority of cells responds to faces per se, i.e., does not distinguish between different individuals. This led Rolls (1985) to suggest that it is the different patterns of activity within a set of face specific neurons that individuate faces. Interestingly, inversion of the faces did not affect the firing rate, but did increase response latency 10-60 ms in 15 out of 26 cells, the remainder not being affected, perhaps explaining why monkeys trained on upright faces take longer to recognize inverted faces.

Having found that some IT neurons respond to certain stimuli and not to others, the question remains as to exactly what stimulus properties trigger these neurons. Rolls et al. (1985) band-pass filtered faces, and found that the majority of IT neurons respond well to high-pass and low-pass filtered faces, a finding in agreement with psychophysical data (Harmon 1973; Harmon and Julesz 1973; Fiorentini et al. 1983). Given that neurons in areas projecting to IT are sensitive to orientations of line segments, IT neurons might collect the response of a number of such cells and thus be tuned to particular silhouettes. Schwartz et al. (1983; see also Schwartz 1984) described the boundary orientation of a silhouette by means of Fourier descriptors (FD; see section 2.1) and tested neurons for their sensitivity to silhouettes with different frequencies and amplitudes (Fig.4.14a). 54% (out of 234) of visually responsive neurons in IT cortex respond selectively to frequency, their response being largely independent of the size, location and contrast reversal of the stimulus (Fig.4.14b). And, generally, response was proportional to the amplitude of the stimulus. Of course, FDs cannot describe complex 2-D shapes such as faces let alone the shape of 3-D objects, and can, therefore, not be used exclusively by IT neurons (a point readily acknowledged by Schwartz et al.).

We saw that most IT neurons increase their activity when stimulated by a complex visual stimulus, regardless of its shape. Using evoked potentials, Srebro (1985) reported that recognition of faces and triangles is correlated with activity in large areas of the temporal cortex. Face-related activity covered a larger area than triangle-related activity, suggesting that different regions of temporal cortex serve different memories. John et al. (1986) estimated that on the order of 10^7 neurons increased their activity (at the $P < 0.05$ significance level) during simultaneous discrimination of two concentric circles and a star (excluding "normal" visual

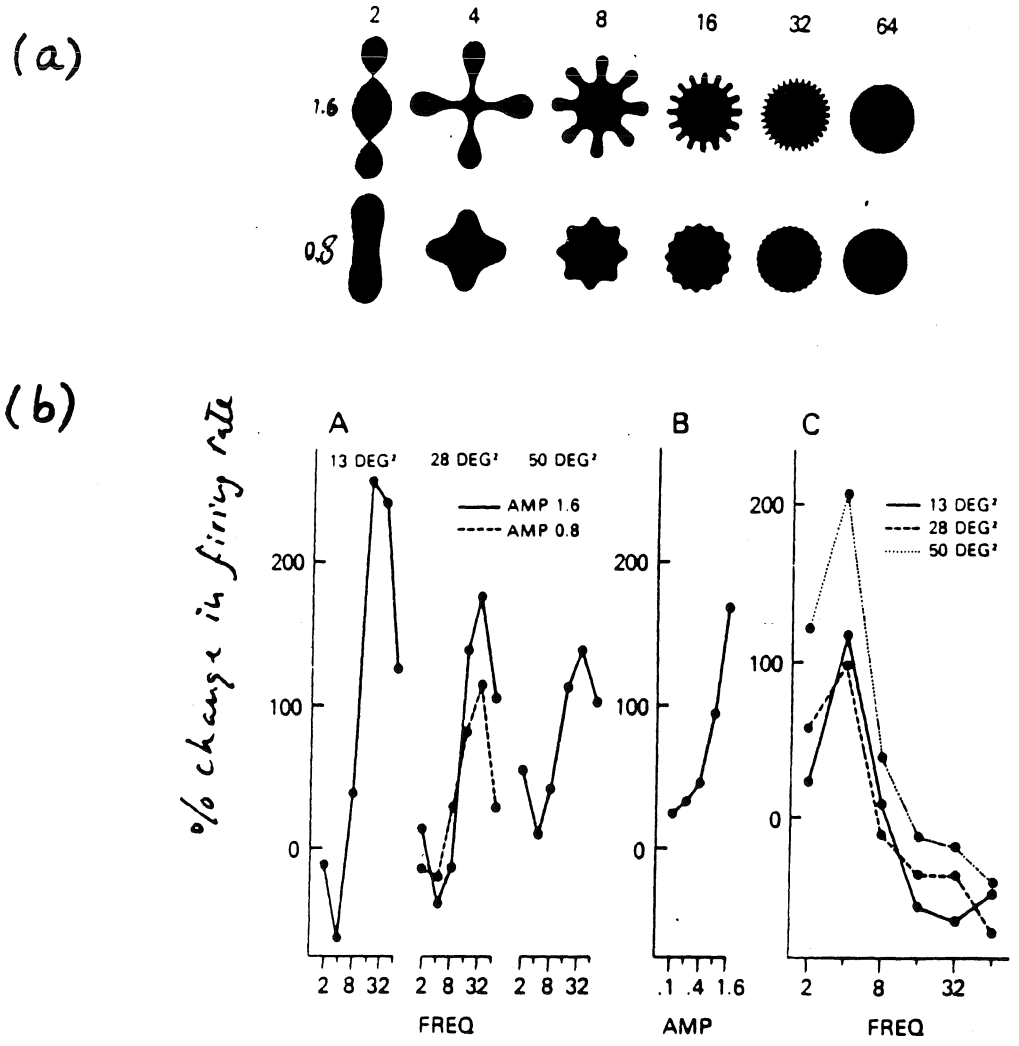


Figure 4.14 (a) Examples of FD stimuli varying in frequency (2–64 cycles/perimeter, cp) and amplitude (0.8 and 1.6). (b) Responses of IT neurons to FD stimuli. Responses are plotted as percent change in firing rate over the mean spontaneous rate (mean of 10 presentations). Different neurons are tuned to different frequencies; in A, they are tuned to 16–32 cp, and in C to 4 cp. Note that the tuning curve remains similar over changes in stimulus size and amplitude. Response is proportional to amplitude (B) (Schwartz et al. 1983).

processing). In view of this large number, these authors question the validity of current feature extraction models that rely on discrete pathways and convergence on specific percept or memory detectors. They conclude that "[i]n view of the large number of neurons involved, the question of how the information represented in these neurons can be evaluated and appreciated by the brain becomes of critical theoretical interest. No conceivable neuron or set of neurons, no matter how diffuse its synaptic inputs, can evaluate the enormous amount of neural activity here shown to be involved in retrieval of even a simple form discrimination. Memory and awareness in complex neural systems may depend upon presently unrecognized

properties of the system as a whole, and not upon any of the elements that constitute the system."

4.3 Summary and discussion

We have traced the pathway subserving visual object recognition as it starts in the retina, continues to the dLGN and ends in a hierarchy of visual areas in the cortex. Summarizing, the retina converts a pattern of light and dark into a pattern of pulses along the optic nerve, where each pulse signals (something like) local contrast. In the optic chiasm the axons of the two optic nerves are partitioned into those serving the right and those serving the left visual field. The first continue to the left dLGN, the second to the right dLGN. In the monkey, the central visual field is mapped onto six layers in the dLGN, each layer receiving input from a particular class of ganglion cells. The mapping from retina to each layer is retinotopic, and the layers are in register with respect to each other. The dLGN appears to function as a modulator of retinal input to the cortex; in terms of responses to visual stimuli, its cells resemble retinal ganglion cells in their center-surround organization. The dLGN projects retinotopically to the striate cortex (and depending on the species, also to extrastriate areas). Here neurons are selective for stimuli such as bars and edges at particular orientations and having a particular size. Moreover, input from the two eyes is combined to obtain stereopsis. The striate area is only the first in a complex hierarchy of almost twenty cortical areas. Since many of these areas have only been identified recently, much remains to be learned about their contributions to perception and recognition.

Thus although much has been learned over the past two decades, this knowledge is quite fragmentary and, one can only agree with Wiesel and Gilbert (1986) that it is still a major mystery how the activities of different cells are integrated to produce a single percept. Even worse, it is often not clear what the function of a cell is; it is one thing to find cells tuned to spatial frequency, and another to conclude that the corresponding tissue computes a Fourier transform; see same holds, mutatis mutandis, for bar detectors (Graham 1979; De Valois and De Valois 1980; Albrecht et al. 1980; MacKay 1981; Braddick 1981; Shapley and Lennie 1985; Bossomaier and Snyder 1986).[14] Or consider neurons selective for disparity. We noted that there are cortical neurons tuned for binocular disparity, that is, they respond maximally to a particular (horizontal) difference in the position of a bar projected on the left retina, and another bar projected onto the right retina. This finding does not tell us much about what this neuron computes in a more complex image. Granted it signals a particular horizontal disparity, several questions remain unanswered: Disparity between what and what? How do the different disparity tuned neurons interact and cooperate? Contrast, for example, the

[14] Thus one could describe traffic noise by tuba detectors, each detector tuned to a different frequency and intensity range.

following answers to the first question. Gillam et al. (1984) and Mitchison and McKee (1985) suggest that matching occurs between the edges in an image and that areas in between are assigned depth by a process of interpolation. Marr and Poggio (1979) propose an algorithm that first performs a feature-by-feature match on a coarse copy of the image, and continues with copies that are progressively more detailed.

Although we might not know the exact function of neurons, we are able to say something about how they encode information. Assuming that spiking rate is a valid measure of a neuron's functional output, one can say that neurons are broadly tuned, meaning that they do have a preference for say a particular wave length, orientation or disparity but also respond to nearby wave lengths, orientations and disparities (Hubel and Wiesel 1977; Kandel and Schwartz 1985). This does not necessarily imply inaccuracy in the representation of a stimulus property since the activities of a number of broadly tuned neurons can be combined (Erickson 1982; Ballard et al. 1983; Hinton et al. 1986). Georgopoulos et al. (1986), for example, studied the relationship between arm movement and the activity of motor neurons in the arm area of the motor cortex in rhesus monkeys. They found that motor neurons are broadly tuned, maintaining a considerable spiking rate with movement in any direction. In other words, individual neurons are inaccurate predictors of the direction in which the arm is going to be moved or is being moved. This suggests that the direction of movement is encoded by the combined activity of large numbers of neurons. Indeed, by viewing neurons as vectors—each vector pointing in the preferred direction of the neuron it represents and having a length proportional to the increase in the neuron's spiking rate—and combining the contributions of 224 neurons by means of a vector sum, the resulting vector pointed in the direction of the arm movement with a high degree of accuracy.

Accordingly, a very interesting topic of research is the behavior of large collections of neurons or neural nets (Amari 1977; Grossberg 1976, 1980; von der Malsburg and Willshaw 1981; Hopfield 1982; Cooper et al. 1985; Hopfield and Tank 1986; Ballard 1986; von der Malsburg and Schneider 1986; Bear et al. 1987). Especially intriguing is work by Linsker (1986abc), showing that antagonistic center-surround organization and orientation selectivity develop naturally in a hierarchical network of neuronal layers.

This discussion about encoding raises the question of how the shape of objects is represented such that they can be recognized by their shape as inferred from visual images. The failure to find any cells in area IT that are tuned to objects other than hands and faces (Desimone et al. 1984), together with the finding that between 5 and 100 million neurons in a cat's brain increased their response as a result of watching a simple, familiar stimulus (John et al. 1986), point towards a distributed representation. And even though some neurons are selective for hands or faces, they are still broadly tuned, for example, for head orientation (Perrett et al. 1985). Additionally, most neurons respond to any face, perhaps implying that facial identity is somehow encoded by the combined activity of all face-selective

neurons. Contrary to the face-selective neurons, the hand-selective neurons have not been tested with real hands, stimuli that do not necessarily show the characteristic five pronged figure. It might thus be the case that the so-called hand-selective neurons are not selective for hands, but are selective for the shape of silhouettes, some of which look like hands.

Whether the identity of objects is encoded by single neurons or the combined activity of a large number of neurons, it is fair to say that it is still a mystery on the basis of what characteristics objects are recognized. And it is not clear that studies that rely on image filtering, 2-D silhouettes, etc., can lead to the answer. They basically ignore what I believe to be the central problem in object recognition, namely, how to find identity in a multitude of differing appearances.

5. Discussion and conclusions

During the fifties, it was demonstrated for the first time that the behavior of animals could be directly traced to the activity of certain neurons. It was found that moving a small, dark object in front of a frog caused some of its retinal ganglion cells to respond vigorously and it caused the frog to snap at the object as if it were a fly. This suggested that the retinal ganglion cells signal the presence of a fly-like object at a particular location in space, and that their firing elicits the snapping response (Lettvin et al. 1959). It also became clear that the optic nerve does more than simply route image intensities to the brain; already at the retinal level, neurons perform complex computations "rejecting unwanted information and passing useful information" (Barlow 1953).

This clearly successful line of research was continued in the sixties and seventies, and in fact still continues. It was discovered that neurons in the postretinal stages of visual processing also respond best to particular visual stimuli and not to others, and that the response of these cells could be explained by assuming that they are connected in some appropriate manner to certain other neurons. Thus, the orientation selectivity of a cell in the primary visual area could arise from its connections to rows of center-surround LGN cells, themselves not selective for orientation. The picture that began to emerge was that of a hierarchically organized visual system. At the bottom of the hierarchy neurons signal local changes in contrast; at the next level, neurons respond to oriented edges, motion in certain directions, etc.; and at the top, cells indicate the presence of very specific patterns, for example that of a hand or a face.

Sutherland (1968) outlined a theory of visual recognition which clearly reflects this view: "The visual input is analysed by a processor that extracts local features (mainly bars, edges and ends) simultaneously at all points on the input picture. When a picture is memorized, a rule is written into a store describing the output from the processor in a highly abstract language...The language used for the descriptive rules contains hierarchical elements and this allows for pictures to be segmented in different ways. When a picture is 'recognized' the output from the processor is matched to a stored description. What we see depends upon the rule to which the picture is matched." Besides incorporating the physiological know-how of that time, this processor was also designed to take into account a number of behavioral observations, including the well-known constancies (size, translational, and brightness), the transfer of learned discriminations, the problems with rotated figures, and perceptual learning. And at the same time, computer programs had been written to analyze and interpret line drawings in terms of three-dimensional objects. The programs showed that it is possible to combine local information from an image, integrate it, and arrive at conclusions about the outside world. An excellent review of these three aspects of visual recognition—behavior, neurophysiology, and computer vision—is given by Barlow et al. (1972).

These findings had of course profound implications for the relationship between behavior and the nervous system, implications that were perhaps most clearly expressed by Barlow (1972): "The central proposition is that our perceptions are caused by the activity of a rather small number of neurons selected from a very large population of predominantly silent cells." The fourth dogma of his neuron doctrine for perceptual psychology states that perception and neurons are related quite simply: for every element of perception there is one cortical neuron. The converse is not true; not every cortical neuron signals a perceptual event. However, Barlow is not exactly clear on what he considers a perceptual element, but it seems to be reserved to something one can be conscious of.

Since that time our knowledge has increased considerably in all three aspects of visual recognition. As far as behavior is concerned, we have learned about mental rotation, categorical perception, and preattentive vision. Neuroscientists have found many new cortical and subcortical maps of the visual field, and discovered that the visual environment profoundly influences the developing visual system. They also found that, even for such a simple percept as a square, large numbers of neurons are selectively involved. In computer vision, many ways of representing three-dimensional shapes, especially in the context of computer-aided-design, have been invented. However, the problem of describing real objects in terms of these representations and deriving the description from images of natural scenes remains.

The most important advance, I believe, has been in the way we think about the visual system. Instead of viewing it as a filter of the "real world," it is more appropriately considered the creator of its own world.[1] The visual observer is not a passive entity, but interacts and measures the environment. Thus, paraphrasing Koenderink's remarks at a recent conference,[2] what we call the shape of an object is the visual system's means of predicting the results of possible interactions with that object. There is no such thing as shape *an sich* only shape in the context of interactions between observers and their environment.[3]

---

[1] Or as Ramón y Cajal put it ca. 1898: "As long as our brain is a mystery, the universe—the reflection of the structure of the brain—will also be a mystery.

[2] Mathematical Problems of Computational Vision, Univ. of California Summer School in Nonlinear Science, Berkeley, California, June 22–29, 1987.

[3] It is rather amusing that this resembles the ancient extramission theory of vision, at least superficially.

REFERENCES

Agin GJ and Binford TO, "Computer Description of Curved Objects," IEEE T. Computers **25**: 439–449 (1976).

Agin GJ, "Hierarchical Representation of Three-Dimensional Objects Using Verbal Models," IEEE Trans. Pattern Analysis and Machine Intell. **3** (2): 197–449 (1981).

Agnetti V, Carreras M, Pinna L, and Rosati G, "Ictal Prosopagnosia and Epileptogenic Damage of the Dominant Hemisphere," Cortex **14**: 50–57 (1978).

Albrecht DG, De Valois RL, and Thorell LG, "Visual Cortical Neurons: Are Bars or Gratings the Optimal Stimuli?" Science **207**: 88–90 (1980).

Allman J, Miezin F, and McGuinness E, "Stimulus Specific Responses from Beyond the Classical Receptive Field: Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons," Ann. Rev. Neurosci. **8**: 407–430 (1985).

Aloimonos J and Chou PB, "Detection of Surface Orientation and Motion from Texture: I. The Case of Planes," Dept. Computer Sci., Univ. of Rochester, Tech. Rep. 161 (1985).

Amari SI, "Field Theory of Self-Organizing Neural Nets," IEEE Trans. Systems, Man, and Cybernetics **13**: 741–748 (1983).

Andersen RA, Essick GK, and Siegel RM, "Encoding of Spatial Location by Posterior Parietal Neurons," Science **230**: 456–458 (1985).

Attneave F, "Some Informational Aspects of Visual Perception," Psychol. Rev. **61**: 183–193 (1954).

Bachevalier J, Parkinson JK, and Mishkin M, "Visual Recognition in Monkeys: Effects of Separate vs. Combined Transection of Fornix and Amygdalofugal Pathways," Exp. Brain Res. **57**: 554–561 (1985).

Ballard DH, "Strip Trees: A Hierarchical Representation for Curves," Comm. Assoc. Comp. Mach. **24**: 310–321 (1981a).

Ballard DH, "Generalizing the Hough Transform to Detect Arbitrary Shapes," Pattern Recogn. **13**: 111–122 (1981b).

Ballard DH and Brown CM, *Computer Vision*, Prentice Hall, Englewood Cliffs, New Jersey (1982).

Ballard DH and Sabbah D, "Viewer Independent Shape Recognition," IEEE Trans. Pattern Analysis and Machine Intell. **5**: 653–660 (1983).

Ballard DH, Hinton GE, and Sejnowski TJ, "Parallel Visual Computation," Nature **306**: 21–26 (1983).

Ballard DH, "Parameter Nets," Artif. Intell. **22**: 235–267 (1984).

Ballard DH, "Cortical Connections and Parallel Processing: Structure and Function," Brain and Behav. Sci. **9**: 67-120 (1986).

Bamieh BA and de Figueiredo RJP, "A Framework and Algorithms for Identification and Attitude Determination of Space Objects From Camera Data," Dept. Electr. and Comp. Eng., Rice Univ., Tech. Rep. EE 8407 (1984).

Barlow HB, "Summation and Inhibition in the Frog's Retina," J. Physiol. **119**: 69–88 (1953).

Barlow HB, "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology?" Perception **1**: 371–394 (1972).

Barlow HB, Narasimhan R, and Rosenfeld A, "Visual Pattern Analysis in Machines and Animals," Science **177**: 567–575 (1972).

Barlow HB and Mollon JD (eds.), *The Senses*, Cambridge Univ. Press, Cambridge, England (1982).

Barlow HB, "Why Have Multiple Cortical Areas?" Vision Res. **26**: 81–90 (1986).

Barnard ST and Fischler MA, "Computational Stereo," Computing Surveys **14**: 553–572 (1982).

Bauer R, "A High Probability of an Orientation Shift Between Layers 4 and 5 in Central Parts of the Cat Striate Cortex," Exp. Brain Res. **48**: 245–255 (1982).

Bauer R, Dow BM, Snyder AZ, and Vautin R, "Orientation Shift Between Upper and Lower Layers in Monkey Visual Cortex," Exp. Brain Res. **50**: 133–145 (1983).

Bauer RM, "Autonomic Recognition of Names and Faces in Prosopagnosia: A Neuropsychological Application of the Guilty Knowledge Test," Neuropsychol. **22**: 457–469 (1984).

Baylis GC, Rolls ET, and Leonard M, "Selectivity Between Faces in the Responses of a Population of Neurons in the Cortex in the Superior Temporal Sulcus of the Monkey," Brain Res. **342**: 91–102 (1985).

Bear MF, Cooper LN, and Ebner FF, "A Physiological Basis for a Theory of Synapse Modification," Science **237**: 42–48 (1987).

Bellugi U, Poizner H, and Klima ES, "Brain Organization for Language: Clues from Sign Aphasia," Human Neurobiol. **2**: 155–170 (1983).

Bender DB, "Retinotopic Organization of Macaque Pulvinar," J. Neurophysiol. **46**: 672–693 (1981).

Bennett BM and Hoffman DD, "Shape Decompositions for Visual Shape Recognition: The Role of Transversality," In: Richards WA (ed.), *Image Understanding II*, Ablex, Norwood, New Jersey (1985).
Benton AL and Van Allen MW, "Impairment in Facial Recognition in Patients with Cerebral Disease," Cortex **4**: 344–358 (1968).
Benton AL, "The Neuropsychology of Facial Recognition," Am. Psychol. **35**: 176–186 (1980).
Besl PJ and Jain RC, "Three-Dimensional Object Recognition," Computing Surveys **17**: 75–145 (1985).
Beusmans JMH, Hoffman DD, and Bennett BM, "Description of Solid Shape and Its Inference from Occluding Contours," J. Opt. Soc. Am. A **4**: 1155–1167 (1987).
Binford TO, "Survey of Model-Based Image Analysis Systems," Int. J. Robotics Res. **1**: 18–63 (1982).
Bjorklund CM and Pavlidis T, "Global Shape Analysis by k-Syntactic Similarity," IEEE Trans. Pattern Analysis and Machine Intell. **3**: 144–154 (1981).
Blasdel GG and Salama G, "Voltage-Sensitive Dyes Reveal a Modular Organization in Monkey Striate Cortex," Nature **321**: 579–585 (1986).
Blum H, "A Geometry for Biology," Annals New York Acad. Sci. **231**: 19–30 (1974).
Blum H and Nagel RN, "Shape Description Using Weighted Symmetric Axis Features," Pattern Recognition **10**: 167–180 (1977).
Bols J and Gilbert CD, "Generation of End-Inhibition in the Visual Cortex Via Interlaminar Connections," Nature **320**: 362–365 (1986).
Boothe RG, Dobson V, and Teller DY, "Postnatal Development of Vision in Human and Nonhuman Primates," Ann. Rev. Neurosci. **8**: 495–545 (1985).
Bornstein B, Sroka H, and Munitz H, "Prosopagnosia with Animal Face Agnosia," Cortex **5**: 164–169 (1969).
Bossomaier T and Snyder AW, "Why Spatial Frequency Processing in the Visual Cortex?" Vision Res. **26**: 1307–1309 (1986).
Bower GH and Glass A, "Structural Units and the Redintegrative Power of Picture Segments," J. Exp. Psychol.: Human Learning and Memory **2**: 456–466 (1976).
Braddick O, "Spatial Frequency Analysis in Vision," Nature **291**: 9–11 (1981).
Bradshaw JL and Wallace G, "Models for the Processing and Identification of Faces," Perception & Psychophysics **9** (5): 443–448 (1971).
Bradshaw JL and Sherlock D, "Bugs and Faces in the Two Visual Fields: The Analytic/Holistic Processing Dichotomy and Task Sequencing," Cortex **18**: 211–226 (1982).
Braitenberg V and Braitenberg C, "Geometry of Orientation Columns in the Visual Cortex," Biol. Cybern. **33**: 179–186 (1979).
Brecher GA, Brecher MH, Kommerell G, Sauter FA, and Sellerbeck J, "Relation of Optical and Labyrinthean Orientation," Optica Acta **19**: 467–471 (1972).
Brooks RA, "Symbolic Reasoning Among 3-D Models and 2-D Images," Artif. Intell. **17**: 285–348 (1981).
Brooks RA, "Model-Based Three-Dimensional Interpretations of Two-Dimensional Images," IEEE Trans Pattern Analysis and Machine Intell. **5**: 140–150 (1983).
Brou P, "Using the Gaussian Image to Find the Orientation of Objects," Int. J. Robotics Res. **3** (4): 89–125 (1984).
Brousil JK and Smith DR, "A Threshold Logic Network for Shape Invariance," IEEE Trans. Electr. Computers **16**: 818–828 (1967).
Brown CM, "Computer Vision and Natural Constraints," Science **224**: 1299–1305 (1984).
Bruce V, "Recognising Faces," Phil. Trans. R. Soc. London **224**: 423–436 (1983).
Bruce V and Green PR, *Visual Perception: Physiology, Psychology and Ecology*, Lawrence Erlbaum Assoc., Publishers, London (1985).
Bruce V and Young A, "Understanding Face Recognition," Br. J. Psychol. **77**: 305–327 (1986).
Bruce V and Valentine T, "Semantic Priming of Familiar Faces," Q. J. Exp. Psychol. **38**: 125–150 (1986).
Buisseret P and Maffei L, "Extraocular Proprioceptive Projections to the Visual Cortex," Exp. Brain Res. **28**: 421–425 (1977).
Buisseret P and Singer W, "Proprioceptive Signals from Extraocular Muscles Gate Experience-Dependent Modifications of Receptive Fields in the Kitten Visual Cortex," Exp. Brain Res. **51**: 443–450 (1983).
Bunke H and Sanfeliu A (eds.), "Advances in Syntactic Pattern Recognition. A Special Issue," Pattern Recogn. **19** (4) (1986).
Burr D and Ross J, "Visual Processing of Motion," Trends in Neurosci. **9**: 304–407 (1986).
Burton GJ, Haig ND, and Moorhead IR, "A Self-Similar Stack Model for Human and Machine Vision," Biol. Cybern. **53**: 397–403 (1986).

Carey S and Diamond R, "From Piecemeal to Configurational Representation of Faces," Science **53**: 312–314 (1977).

Carey S, "The Development of Face Perception," In: Davies G, Ellis H, and Shepherd J (eds.), *Perceiving and Remembering Faces*, Academic Press, London (1981).

Carpenter PA and Eisenberg P, "Mental Rotation and the Frame of Reference in Blind and Sighted Individuals," Perception & Psychophysics **23**: 117–124 (1978).

Casasent D and Psaltis D, "Position, Rotation, and Scale Invariant Optical Correlation," Appl. Opt. **15**: 1795–1799 (1976).

Celio MR, Schaerer L, Morrison JH, Norman AW, and Bloom FE, "Calbindin Immunoreactivity Alternates with Cytochrome c-Oxydase-Rich Zones in Some Layers of the Primate Visual Cortex," Nature **323**: 715–717 (1986).

Chagas C, Gattass R, and Gross C (eds.), *Pattern Recognition Mechanisms*, Exp. Brain Res. Suppl. **11** (1985).

Chakravarty I, "A Generalized Line and Junction Labeling Scheme with Applications to Scene Analysis," IEEE Trans. Pattern Anal. Machine Intell. **1**: 202–205 (1979).

Chakravarty I and Freeman H, "Characteristic Views as a Basis for Three-Dimensional Object Recognition," Proc. Society for Photo-Optical Instrumentation Engineers **336**: 37–45 (1982).

Chin RT and Dyer CR, "Model-Based Recognition in Robot Vision," Computing Rev. **18**: 67–108 (1986).

Chomsky N, *Aspects of the Theory of Syntax*, The MIT Press, Cambridge, Mass. (1965).

Cooper HM and Magnin M, "A Common Mammalian Plan of Accessory Optic System Organization Revealed in all Primates," Nature **324**: 457–459 (1986).

Cooper LA and Podgorny P, "Mental Transformations and Visual Comparison Processes: Effects of Complexity and Similarity," J. of Exp. Psychol.: Human Learning and Memory **2**: 503–514 (1976).

Cooper LA and Shepard RN, "Turning Something Over in the Mind," Sci. Am. **251**: 106–114 (1984).

Cooper LN, Munto P, and Scofield C, "Neuron Selectivity: Single Neuron and Neuron Networks," In: Levy WB, Anderson JA, and Lehmkuhle S (eds.), *Synaptic Modification, Neuron Selectivity, and Nervous System Organization*, Lawrence Erlbaum Assoc., Publishers, Hillsdale, New Jersey (1985).

Corballis MC and Cullen S, "Decisions about the Axes of Disoriented Shapes," Memory & Cognition **14** (1): 27–38 (1986).

Crick F, "Thinking About the Brain," Sci. Am. **241** (3): 181–188 (1979).

Crick F, "Function of the Thalamic Reticular Complex: The Searchlight Hypothesis," Proc. Natl. Acad. Sci. USA **81**: 4586–4590 (1984).

Curcio CA, Sloan KR Jr., Packer O, Hendrickson AE, and Kalina RE, "Distribution of Cones in Human and Monkey Retina: Individual Variability and Radial Symmetry," Science **236**: 579–582 (1987).

Cynader M and Regan D, "Neurones in Cat Parastriate Cortex Sensitive to the Direction of Motion in Three-Dimensional Space," J. Physiol. **274**: 549–569 (1978).

Damasio AR, Damasio H, and Van Hoesen GW, "Prosopagnosia: Anatomic Basis and Behavioral Mechanisms," Neurology **32**: 331–341 (1982).

Damasio AR, "Disorders of Complex Visual Processing: Agnosias, Achromatopsia, Balint's Syndrome, and Related Difficulties of Orientation and Construction," In: Mesulam MM (ed.), *Principles of Behavioral Neurology*, F. A. Davis Co., Philadelphia, pp259–288 (1985a).

Damasio AR, "Prosopagnosia," Trends in Neurosci. **8**: 132–135 (1985b).

Damasio AR, Bellugi U, Damasio H, Poizner H, and Van Gilder J, "Sign Language Aphasia During Left-Hemisphere Amytal Injection," Nature **322**: 363–364 (1986).

Davies G, Ellis H, and Shepherd J, "Cue Saliency in Faces as Assessed by the 'Photofit' Technique," Perception **6**: 263–269 (1977).

Davies G, Ellis H, and Shepherd P (eds.), *Perceiving and Remembering Faces*, Academic Press, London (1981).

Davis LS and Rosenfeld A, "Cooperating Processes for Low-Level Vision: A Survey," Artif. Intell. **17**: 245–263 (1981).

Dean P, "Visual Behavior in Monkeys with Inferotemporal Lesions," In: Ingle DJ, Goodale MA, and Mansfield RJW (eds.), *Analysis of Visual Behavior*, The MIT Press, Boston, Mass. (1982).

Delis DC, Robertson LC, and Efron R, "Hemispheric Specialization of Memory for Visual Hierarchical Stimuli," Neuropsychologia **24**: 205–214 (1986).

Desimone R and Gross CG, "Visual Areas in the Temporal Cortex of the Macaque," Brain Res. **178**: 363–380 (1979).

Desimone R, Albright TD, Gross CG, and Bruce C, "Stimulus-Selective Properties of Inferior Temporal Neurons in the Macaque," J. Neurosci. 4: 2051–2062 (1984).

Desimone R, Schein SJ, Moran J, and Ungerleider LG, "Contour, Color and Shape Analysis Beyond the Striate Cortex," Vision Res. **25**: 441–452 (1985).

De Valois RL and De Valois KK, "Spatial Vision," Ann. Rev. Psychol. **31**: 309–341 (1980).

Diamond R and Carey S, "Developmental Changes in Representation of Faces," J. Exp. Child Psychol. **23**: 1–22 (1977).

Diamond R and Carey S, "Why Faces Are and Are Not Special: An Effect of Expertise," J. Exp. Psychol.: General **115**: 107–117 (1986).

Dick M, Ullman S, and Sagi D, "Parallel and Serial Processes in Motion Detection," Science **237**: 400–402 (1987).

Dodwell PC, "The Lie Transformation Group Model of Visual Perception," Perception & Psychophysics **34**: 1–16 (1983).

Dow BM and Bauer R, "Retinotopy and Orientation Columns in the Monkey: A New Model," Biol. Cybern. **49**: 189–200 (1984).

Duda RO and Hart PE, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York (1973).

D'Zmura M and Lennie P, "Shared Pathways for Rod and Cone Vision," Vision Res. **26**: 1273–1280 (1986).

Eley MG, "Identifying Rotated Letter-Like Symbols," Memory & Cognition **10**: 25–32 (1982).

Elliott ES, Wills EJ, and Goldstein AG, "The Effects of Discrimination Training on the Recognition of White and Oriental Faces," Bull. Psychon. Soc. **2** (2): 71–73 (1973).

Ellis HD, "Recognising Faces," Br. J. Psychol. **66**: 409–426 (1975).

Ellis HD, Shepherd JW, and Davies GM, "Identification of Familiar and Unfamiliar Faces From Internal and External Features: Some Implications for Theories of Facial Recognition," Perception **8**: 431–439 (1979).

Ellis HD, "Theoretical Aspects of Face Recognition," In: Davies G, Ellis H, and Shepherd P (eds.), *Perceiving and Remembering Faces*, Academic Press, London (1981).

Ellis HD, Jeeves MA, Newcombe F, and Young A (eds.), *Aspects of Face Processing*, Nijhoff Publ., Dordrecht, the Netherlands (1986).

Emerson RC, Citron MC, and Vaughn WJ, "Cortical Responses to Motion Depend on Nonlinear Spatiotemporal Interactions," Invest. Ophthalm. Visual Sci. (suppl) **26** (7): 7 (1985).

Enroth-Cugell C and Robson JG, "The Contrast Sensitivity of Retinal Ganglion Cells of the Cat," J. Physiol. **187**: 517–552 (1966).

Enroth-Cugell C, Robson JG, Schweitzer-Tong DE, and Watson AB, "Spatio-Temporal Interactions in Cat Retinal Ganglion Cells Showing Linear Spatial Summation," J. Physiol. **341**: 279–307 (1983).

Erickson RP, "The Across-Fiber Pattern Theory: An Organizing Principle for Molar Neural Function," Contributions to Sensory Physiol. **6**: 70–110 (1982).

Fairfield J, "Segmenting Dot Patterns by Voronoi Diagram Concavity," IEEE Trans. Pattern Analysis and Machine Intell. **5**: 104–110 (1983).

Farah MJ, "The Neural Locus of Mental Image Generation: Converging Evidence from Brain-Damaged and Normal Subjects," Proc. 7th Ann. Conf. Cogn. Soc. pp19–25 (1985).

Faugeras OD and Ponce J, "Prism Trees: A Hierarchical Representation for 3-D Objects," Proc. 8th Int. Joint Conf. Artif. Intell., Karlsruhe, West Germany, pp982–988 (1983).

Feder J, "Plex Languages," Information Sciences **3**: 225–241 (1971).

Fiorentini A, Maffei L, and Sandini G, "The Role of High Spatial Frequencies in Face Perception," Perception **12**: 195–201 (1983).

Fischler MA and Bolles RC, "Perceptual Organization and Curve Partitioning," IEEE Trans. Pattern Analysis and Machine Intell. **8**: 100–105 (1986).

Fishman MC and Michael CR, "Integration of Auditory Information in the Cat's Visual Cortex," Vision Res. **13**: 1415–1419 (1973).

Fleet DJ, Jepson AD, and Hallett PE, "A Spatio-Temporal Model for Early Visual Processing," Dept. Comp. Sci., Univ. of Toronto, Tech. Rep. RCBV-TR-84-1 (1984).

Flin RH, "Development of Face Recognition: An Encoding Switch?" Br. J. Psychol. **76**: 123–134 (1985).

Foley JD and van Dam A, *Fundamentals of Interactive Computer Graphics*, Addison-Wesley Publ. Co., Reading, Mass. (1982).

Foster DH, "Visual Pattern Recognition by Assignment of Invariant Features and Feature-Relations," Optica Acta **24**: 147–157 (1977).

Freeman H, "Computer Processing of Line-Drawing Images," Comp. Surveys **6**: 57–98 (1974).

Fu KS, *Syntactic Pattern Recognition and Applications*, Prentice Hall, Englewood Cliffs, New Jersey (1982).

Fuster JM and Jervey JP, "Inferotemporal Neurons Distinguish and Retain Behaviorally Relevant Features of Visual Stimuli," Science **212**: 952–955 (1981).

Galper RE, "Recognition of Faces in Photographic Negative," Psychon. Sci. **19** (4): 207–208 (1970).

Gardenier PH, McCallum BC, and Bates RHT, "Fourier Transform Magnitudes Are Unique Pattern Recognition Templates," Biol. Cybern. **54**: 385–391 (1986).

Gattass R, Sousa APB, and Covey E, "Cortical Visual Areas of the Macaque: Possible Substrates for pattern recognition mechanisms," Exp. Brain Res. Suppl. **11**: 1–20 (1985).

Gazzaniga MS and Smylie CS, "Facial Recognition and Brain Asymmetries: Clues to Underlying Mechanisms," Ann. Neurol. **13**: 536–540 (1983).

Georgopoulos AP, Schwartz AB, and Kettner RE, "Neuronal Population Coding of Movement Direction," Science **233**: 1416–1419 (1986).

Geschwind N, *Selected Papers on Language and the Brain*, D. Reidel Publ. Co., Dordrecht–Holland (1974).

Geschwind N, "Specializations of the Human Brain," Sci. Am. **241** (3): 158–171 (1979).

Gilbert CD, "Microcircuitry of the Visual Cortex," Ann. Rev. Neurosci. **6**: 217–247 (1983).

Gilbert CD, "Horizontal Integration in the neocortex," Trends in Neurosci. **8**: 160–165 (1985).

Gillam B, Flagg T, and Finlay D, "Evidence for Disparity Change as the Primary Stimulus for Stereoscopic Processing," Perception & Psychophysics **36**: 559–564 (1984).

Gluck MA and Corter JE, "Information, Uncertainty, and the Utility of Categories," 7th Ann. Conf. Cogn. Sci. Soc. pp133–138 (1985).

Gluender H, "Neural Computation of Inner Geometric Pattern Relations," Biol. Cybern. **55**: 239–251 (1986).

Goetz KG, "Do "d-Blob" and "I-Blob" Hypercolumns Tessellate the Monkey Visual Cortex?" Biol. Cybern. **56**: 107–109 (1987).

Goldman-Rakic PS and Schwartz ML, "Interdigitation of Contralateral and Ipsilateral Columnar Projections to Frontal Association Cortex in Primates," Science **216**: 755–7577 (1982).

Graaf CN de, Toet A, Koenderink JJ, Zuidema P, and van Rijk PP, "Some Applications of Hierarchical Image Processing Algortihms," In: Deconinck F (ed.), *Information Processing and Medical Imaging*, Nijhoff Publ., Dordrecht, pp343–369 (1984).

Graham J, Lin CS, and Kaas JH, "Subcortical Projections of Six Visual Cortical Areas in the Owl Monkey, Aotus Trivirgatus," J. Comp. Neurol. **187**: 557–580 (1979).

Graham N, "Does the Brain Perform a Fourier Analysis of the Visual Scene?" Trends in Neurosci. **2**: 207–208 (1979).

Grinvald A, Lieke E, Frostig RD, Gilbert CD, and Wiesel TN, "Functional Architecture of Cortex Revealed by Optical Imaging of Intrinsic Signals," Nature **324**: 361–364 (1986).

Gross CG, Bender DB, and Rocha-Miranda CE, "Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey," Science **166**: 1303–1306 (1969).

Gross CG, "Inferior Temporal Lesions Do Not Impair Discrimination of Rotated Patterns in Monkeys," J. of Comp. and Physiol. Psychol. **92**: 1095–1109 (1978).

Grossberg S, "On the Development of Feature Detectors in the Visual Cortex with Applications to Learning and Reaction-Diffusion Systems," Biol. Cybern. **21**: 145–159 (1976).

Grossberg S, "How Does the Brain Build a Cognitive Code?" Psychol. Rev. **87**: 1–51 (1980).

Guerra C and Pieroni GG, "A Graph-Theoretic Method for Decomposing Two-Dimensional Polygonal Shapes into Meaningful Parts," IEEE Trans. Pattern Analysis and Machine Intell. **4**: 405–190 (1982).

Gyr J, Willey R, and Henry A, "Motor-Sensory Feedback and Geometry of Visual Space: An Attempted Replication," Behav. and Brain Sciences **2**: 59–94 (1979).

Hampson S and Kibler D, "A Boolean Complete Neural Model of Adaptive Behavior," Biol. Cybern. **49**: 9–19 (1983).

Hampson S and Volper DJ, "Linear Function Neurons: Structure and Training," Biol. Cybern. **53**: 203–217 (1986).

Harmon LD, "The Recognition of Faces," Sci. Am. **229**: 70–83 (1973).

Harmon LD and Julesz B, "Masking in Visual Recognition: Effects of Two-Dimensional Filtered Noise," Science **180**: 1194–1197 (1973).

Harris WA, "Learned Topography: The Eye Instructs the Ear," Trends in Neurosci. **9**: 97–99 (1986).

Hay DC and Young AW, "The Human Face," In: Ellis AW (ed.), *Normality and Pathology in Cognitive Functions*, Academic Press, London, pp173–202 (1982).

Hebb DO, *Organization of Behavior*, John Wiley & Sons, London (1949).

Heggelund P and Moors J, "Orientation Selectivity and the Spatial Distribution of Enhancement and Suppression in Receptive Fields of Cat Striate Cortex Cells," Exp. Brain Res. **52**: 235–247 (1983).
Held R and Hein A, "Movement-Produced Stimulation in the Development of Visually Guided Behavior," J. of Comp. and Physiol. Psychol. **56**: 872–876 (1963).
Hendrickson AE, "Dots, Stripes and Columns in Monkey Visual Cortex," Trends in Neurosci. **8**: 406–410 (1985).
Herrnstein RJ, "Objects, Categories, and Discriminative Stimuli," In: Roitblatt HL, Bever TG, and Terrace HS (eds.), *Animal Cognition*, Lawrence Erlbaum Ass., Hillsdale, New Jersey (1984).
Herrnstein RJ, "Riddles of Natural Categorization," Phil. Trans. R. Soc. London **8**: 129–144 (1985).
Herts H, *The Principles of Mechanics*, Dover Publications, New York, 1956, translated from 1st German edition (1894).
Heydt L van der, Dom F, and van den Berghe H, "Two-Dimensional Shape Decomposition Using Fuzzy Subset Theory Applied to Automated Chromosome Analysis," Pattern Recognition **13** (2): 147–157 (1981).
Heydt R von der, Peterhans E, and Baumgartner G, "Illusory Contours and Cortical Neuron Responses," Science **224**: 1260–1262 (1984).
Hilbert D and Cohn-Vossen S, *Geometry and the Imagination*, Chelsea, New York (1952).
Hinton GE, "The Role of Spatial Working Memory in Shape Perception," Proc. 3rd Ann. Conf. Cogn. Sci. Soc., Berkeley, Calif., pp56–60 (1981a).
Hinton GE, "A Parallel Computation that Assigns Canonical Object-Based Frames of Reference," Proc. 7th Int. Joint Conf. Artif. Intell., Vancouver, Canada, pp683–685 (1981b).
Hinton GE, McClelland JL, and Rumelhart DE, "Distributed Representations," In: Rumelhart DE, McClelland JL and the PDP Research Group, *Parallel Distributed Processing. Vol.1: Foundations*, The MIT Press, Cambridge, Mass. (1986).
Hochberg J and Gellman L, "The Effect of Landmark Features on Mental Rotation Times," Memory & Cognition **5**: 23–26 (1977).
Hoffman DD and Richards WA, "Representing Smooth Plane Curves for Recognition: Implications for Figure-Ground Reversal," Proc. Natl. Conf. Artif. Intell., Pittsburgh, Penn., pp5–8 (1982).
Hoffman DD, "The Interpretation of Visual Illusions," Sci. Am. **249** (6): 154–162 (1983).
Hoffman DD and Richards WA, "Parts of Recognition," Cognition **18**: 65–96 (1984).
Hoffman DD and Bennett BM, "Inferring the Relative Three-Dimensional Positions of Two Moving Points," J. Opt. Soc. Am. A **2**: 350–353 (1985).
Hoffman WC, "The Lie Algebra of Visual Perception," J. Math. Psychol. **3**: 65–98 (1966).
Hollard VD and Delius JD, "Rotational Invariance in Visual Pattern Recognition by Pigeons and Humans," Science **218**: 804–806 (1982).
Holst E von, and Mittelstaedt H, "Das Reafferenzprinzip," Naturwissenschaften **20**: 464–476 (1950).
Hopfield JJ, "Neural Networks and Physical Systems with Emergent Collective Computational Properties," Proc. Natl. Acad. Sci. USA **79**: 2554–2558 (1982).
Hopfield JJ and Tank DW, "Computing with Neural Circuits: A Model," Science **233**: 625–633 (1986).
Horn BKP, "Extended Gaussian Images," Proc. IEEE **72**: 1671–1686 (1984).
Horn BKP and Ikeuchi K, "The Mechanical Manipulation of Randomly Oriented Parts," Sci. Am. **251** (8): 100–111 (1984).
Horn BKP, *Robot Vision*, The MIT Press, Cambridge, Mass. (1986).
Howard IP, *Human Visual Orientation*, John Wiley & Sons, New York (1982).
Hubel DH and Wiesel TN, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," J. Physiol. **195**: 215–243 (1968).
Hubel DH and Wiesel TN, "Cells Sensitive to Binocular Depth in Area 18 of the Macaque Monkey Cortex," Nature **225**: 215–243 (1970).
Hubel DH and Wiesel TN, "Functional Architecture of Macaque Monkey Visual Cortex," Proc. R. Soc. London B **198**: 1–59 (1977).
Hubel DH and Wiesel TN, "Brain Mechanisms of Vision," Sci. Am. **241** (3): 130–144 (1979).
Hubel DH and Livingstone MS, "Complex-Unoriented Cells in a Subregion of Primate Area 18," Nature **315**: 325–327 (1985).
Hummel RA and Zucker SW, "On the Fou: ıtions of Relaxation Labeling Processes," IEEE Trans. Pattern Analysis and Machine Intell. 5: 267–287 .)83).
Humphrey NK, "The Illusion of Beauty," Perception **2**: 429–439 (1973).
Humphreys GW, "Reference Frames and Shape Perception," Cogn. Psychol. **15**: 151–196 (1983).

Humphreys GW and Riddoch MJ, "On Telling Your Fruit From Your Vegetables: A Consideration of Category -Specific Deficits After Brain Damage," Trends in Neurosci. **10**: 145–148 (1987).
Ikeuchi K, " Recognition of 3-D Objects Using the Extended Gaussian Image," Proc. 7th Int. Joint Conf. Artif. Intell., Vancouver, Canada, pp595–600 (1981).
Ikeuchi K and Horn BKP, "Numerical Shape from Shading and Occluding Boundaries," Artif. Intell. **17**: 141–184 (1981).
Ikeuchi K, "Determining Attitude of Object From Needle Map Using Extended Gaussian Image," MIT AI Memo No. 714 (1983).
Inoue M, Oomura Y, Yakushiji T, and Akaika N, "Intracellular Calcium Ions Decrease the Affinity of the GABA Receptor," Nature **324**: 156–158 (1986).
John ER, Tang Y, Brill AB, Young R, and Ono K, "Double-Labeled Metabolic Maps of Memory," Science **233**: 1167–1175 (1986).
Jolicoeur P and Kosslyn SM, "Coordinate Systems in the Long-Term Memory Representation of Three-Dimensional Shapes," Cogn. Psychol. **15**: 301–345 (1983).
Jolicoeur P, Gluck MA, and Kosslyn SM, "Pictures and Names: Making the Connection," Cogn. Psychol. **16**: 243–275 (1984).
Jolicoeur P and Landau MJ, "Effects of Orientation on the Identification of Simple Visual Patterns," Can. J. Psychol. **38**: 80–93 (1984).
Jolicoeur P, "The Time To Name Disoriented Natural Objects," Memory & Cognition **13**: 289–303 (1985).
Jones GV, "Identifying Basic Categories," Psychol. Bull. **94**: 423–428 (1983).
Julesz B, "Binocular Depth Perception without Familiarity Cues," Science **145**: 356–362 (1964).
Julesz B, *Foundations of Cyclopean Perception*, Univ. of Chicago Press, Chicago (1971).
Julesz B, "Textons, the Elements of Texture Perception, and Their Interactions," Nature **290**: 91–97 (1981).
Julesz B, "A Brief Outline of the Texton Theory of Human Vision," Trends in Neurosci. **7**: 41–45 (1984).
Kaas JH, "The Organization of Neocortex in Mammals: Implications for Theories of Brain Function," Ann. Rev. Psychol. **38**: 129–151 (1987).
Kanal LH and Rosenfeld A, *Progress in Pattern Recognition*, Elsevier Science Publ., North Holland (1985).
Kandel ER and Schwartz JH (eds.), *Principles of Neural Science*, Elsevier, New York (1985).
Kanizsa G, "Subjective Contours," Sci. Am. **234**: 48–52 (1976).
Kaplan E and Shapley RM, "The Primate Retina Contains Two Types of Ganglion Cells, With High and Low Contrast Sensitivity," Proc. Natl. Acad. Sci. USA **83**: 2755–2757 (1986).
Kean ML, "Disconnected Memories," In: Weinberger NM, McGaugh JL and Lynch G (eds.), *Memory Systems of the Brain*, The Guilford Press, New York (1985).
Kendrick KM and Baldwin BA, "Cells in Temporal Cortex of Conscious Sheep Can Respond Preferentially to the Sight of Faces," Science **236**: 448–450 (1987).
Knudsen EI and Knudsen PF, "Vision Guides the Adjustment of Auditory Localization in Young Barn Owls," Science **230**: 545–548 (1985).
Koch C and Ullman S, "Selecting One Among the Many: A Simple Network Implementing Shifts in Selective Visual Attention," MIT AI Memo 770 (1984).
Koenderink JJ and van Doorn AJ, "Invariant Properties of the Motion Parallax Field Due to the Movement of Rigid Bodies Relative to an Observer," Optica Acta **22**: 773–791 (1975).
Koenderink JJ and van Doorn AJ, "Geometry of Binocular Vision and a Model for Stereopsis," Biol. Cybern. **21**: 29–35 (1976a).
Koenderink JJ and van Doorn AJ, "Visual Perception of Rigidity of Solid Shape," J. Math. Biol. **3**: 79–85 (1976b).
Koenderink JJ and van Doorn AJ, "The Singularities of the Visual Mapping," Biol. Cybern. **24**: 51–59 (1976c).
Koenderink JJ and van Doorn AJ, "Visual Detection of Spatial Contrast; Influence of Location in the Visual Field, Target Extent and Illuminance Level," Biol. Cybern. **30**: 157–167 (1978).
Koenderink JJ and van Doorn AJ, "The Internal Representation of Solid Shape with Respect to Vision," Biol. Cybern. **32**: 211–216 (1979).
Koenderink JJ, "Why Argue About Direct Perception?" Behav. and Brain Sci. **3**: 390–391 (1980).
Koenderink JJ and van Doorn AJ, "Photometric Invariants Related to Solid Shape," Optica Acta **27**: 981–996 (1980).
Koenderink JJ and van Doorn AJ, "Invariant Features of Contrast Detection: An Explanation in Terms of Self-Similar Detection Arrays," J. Opt. Soc. Am. **72**: 83–87 (1982).
Koenderink JJ, "The Structure of Images," Biol. Cybern. **50**: 363–370 (1984a).

Koenderink JJ, "Geometrical Structures Determined by the Functional Order in Nervous Nets," Biol. Cybern. **50**: 43–50 (1984b).
Koenderink JJ, "Simultaneous Order in Nervous Nets from a Functional Standpoint," Biol. Cybern. **50**: 35–41 (1984c).
Koenderink JJ, "What Does the Occluding Contour Tell Us About Solid Shape?" Perception **13**: 321–330 (1984d).
Koenderink JJ, "Optic Flow," Vision Res. **26**: 161–180 (1986).
Koenderink JJ and van Doorn AJ, "Depth and Shape from Differential Perspective in the Presence of Bending Deformations," J. Opt. Soc. Am. A **3**: 242–249 (1986a).
Koenderink JJ and van Doorn AJ, "Dynamic Shape," Biol. Cybern. **53**: 383–396 (1986b).
Kolers PA and Perkins DN, "Orientation of Letters and Their Speed of Recognition," Q. J. Exp. Psychol. **5**: 275–280 (1969).
Korn MR and Dyer C, "3-D Multiview Object Representations for Model-Based Object Recognition," Pattern Recogn. **20**: 91–103 (1987).
Kosslyn SM, "Seeing and Imagining in the Cerebral Hemispheres: A Computational Approach," Psychol. Rev. **94**: 148–175 (1987).
Kroese BJA, "Local Structure Analyzers as Determinants of Preattentive Pattern Discrimination," Biol. Cybern. **55**: 289–298 (1987).
Kuffler SW, "Discharge Patterns and Functional Organization of Mammalian Retina," J. Neurophysiol. **16**: 37–68 (1953).
Kushnir M, Abe K, and Matsumoto K, "Recognition of Handprinted Hebrew Characters Using Features Selected in the Hough Transform Space," Pattern Recogn. **18**: 103–114 (1985).
Lee DT, "Medial Axis Transformation of a Planar Shape," IEEE Trans. Pattern Analysis and Machine Intell. **4**: 363–114 (1982).
Leehey S, Carey S, Diamond R, and Cahn A, "Upright and Inverted Faces: The Right Hemisphere Knows the Difference," Cortex **14**: 411–419 (1978).
Leibowitz HW and Post RB, "The Two Modes of Processing Concept and Some Implications," In: Beck J (ed.), *Organization and Representation in Perception*, Lawrence Erlbaum Ass., Publ., Hillsdale, New Jersey, pp343–363 (1982).
Lennie P, "Parallel Visual Pathways: A Review," Vision Res. **20**: 561–594 (1980).
Lettvin JY, Maturana HR, McCulloch MS, and Pitts WH, "What the Frog's Eye Tells the Frog's Brain," Proc. Inst. Radio Engineers **47**: 1940–1951 (1959).
Levine MD, *Vision in Man and Machine*, McGraw-Hill, New York (1985).
Levy J, Trevarthen C, and Sperry RW, "Perception of Bilateral Chimeric Figures Following Hemispheric Deconnexion," Brain **95**: 61–78 (1972).
Leyton M, "Constraint-Theorems on the Prototypification of Shape," Proc. 5th Natl. Conf. Artif. Intell., Philadelphia, pp141–153 (1986).
Lia B, Williams RW, and Chalupa LM, "Formation of Retinal Ganglion Cell Topography During Prenatal Development," Science **236**: 848–851 (1987).
Lin WC and Fu KS, "A Syntactic Approach to 3-D Object Representation," IEEE Trans. Pattern Analysis and Machine Intell. **6**: 351–364 (1984).
Lin WC and Fu KS, "A Syntactic Approach to 3-D Object Recognition," IEEE Trans. Systems, Man, and Cybernetics **16**: 405–422 (1986).
Lindberg DC, *Theories of Vision from al-Kindi to Kepler*, Univ. of Chicago Press, Chicago (1976).
Linsker R, "From Basic Network Principles to Neural Architecture: Emergence of Spatial-Opponent Cells," Proc. Natl. Acad. Sci. USA **83**: 7508–7512 (1986a).
Linsker R, "From Basic Network Principles to Neural Architecture: Emergence of Orientation-Selective Cells," Proc. Natl. Acad. Sci. USA **83**: 8390–8394 (1986b).
Linsker R, "From Basic Network Principles to Neural Architecture: Emergence of Orientation-Columns," Proc. Natl. Acad. Sci. USA **83**: 8779–8783 (1986c).
Lipschutz MM, *Differential Geometry*, McGraw-Hill, New York (1969).
Little JJ, "An Iterative Method for Reconstructing Convex Polyhedra from Extended Gaussian Images," Proc. Natl. Conf. Artif. Intell., Washington, D.C., pp247–250 (1983).
Little JJ, "Recovering Shape and Determining Attitude from Extended Gaussian Images," Dept. Comp. Sci., Univ. of British Columbia, Tech. Rep. TN 85-2 (1985a).

Little JJ, "Determining Object Attitude From Extended Gaussian Images," Proc. 9th Int. Joint Conf. Artif. Intell. pp960–963 (1985b).
Longuet-Higgins HC and Prazdny K, "The Interpretation of a Moving Retinal Image," Proc. R. Soc. London B **208**: 385–397 (1980).
Lowe DG and Binford TO, "The Recovery of Three-Dimensional Structure from Image Curves," IEEE Trans Pattern Analysis and Machine Intell. **7**: 320–326 (1985).
Mach E, *The Analysis of Sensations*, Dover Publ., New York 1959, translated from 1st German edition (1902).
MacKay DM, "Strife over Visual Cortical Function," Nature **289**: 117–118 (1981).
Macko KA, Jarvis CD, Kennedy C, Miyaoka M, Shinohara M, Sokoloff K, and Mishkin M, "Mapping the Primate Visual System with [2-14C]Deoxyglucose," Science **218**: 394–397 (1982).
Madarass RL and Thompson WB, "Recognition of Moving Objects Using Feature Signatures," IEEE Trans. Pattern Analysis and Machine Intell. **7**: 713–717 (1985).
Maffei L, "Encoding and Processing of Visual Information in Cortical Neurons," Exp. Brain Res. Suppl. **11**: 97–116 (1985).
Malik J, "Recovering Three-Dimensional Shape from a Single Image of Curved Objects," Comp. Sci. Div., Univ. of Calif., Berkeley, Tech. Rep. UCB/CSD 87/340 (1987).
Malik J, "Interpreting Line Drawings of Curved Objects," Int. J. Comp. Vision (in press).
Mallot HA, "An Overall Description of Retinotopic Mapping in the Cat's Visual Cortex Areas 17, 18, and 19," Biol. Cybern. **52**: 45–51 (1985).
Malsburg C von der, and Willshaw D, "Co-operativity and Brain Organization," Trends in Neurosci. **4**: 80–83 (1981).
Malsburg C von der, and Schneider W, "A Neural Cocktail-Party Processor," Biol. Cybern. **54**: 29–40 (1986).
Mandl G, "Responses of Visual Cells in Cat Superior Colliculus to Relative Pattern Movement," Vision Res. **25**: 267–281 (1985).
Mansfield RJW, "Neural Basis of Orientation Perception in Primate Vision," Science **186**: 1133–1135 (1974).
Marko H, "Space Distortion and Decomposition Theory," Kybernetik **13**: 132–143 (1973).
Marr D and Poggio T, "Cooperative Computation of Stereo Disparity," Science **194**: 283–287 (1976).
Marr D and Nishihara HK, "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," Proc. R. Soc. London B. **200**: 269–294 (1978).
Marr D and Poggio T, "A Computational Theory of Human Stereo Vision," Proc. R. Soc. London B **204**: 301–328 (1979).
Marr D, *Vision*, W. H. Freeman, San Francisco (1982).
Martin WN and Aggarwal JK, "Volumetric Descriptions of Objects from Multiple Views," IEEE Trans. Pattern Analysis and Machine Intell. **5**: 150–158 (1983).
Marzi CA and Berlucchi G, "Right Visual Field Superiority for Accuracy of Recognition of Famous Faces in Normals," Neuropsychologia **15**: 751–756 (1977).
Masland RH, "The Functional Architecture of the Retina," Sci. Am. **255** (6): 102–111 (1986).
Matsubara J, Cynader M, Swindale NV, and Stryker MP, "Intrinsic Projections within Visual Cortex: Evidence for Orientation Specific Local Connections," Proc. Natl. Acad. Sci. USA **82**: 935–939 (1985).
Mayhew JEW, "The Interpretation of Stereo-Disparity Information: The Computation of Surface Orientation and Depth," Perception **11**: 387–403 (1982).
Mayhew JEW and Longuet-Higgins HC, "A Computational Model of Binocular Depth Perception," Nature **297**: 376–378 (1982).
McClelland JL and Rumelhart DE, "An Interactive Activation Model of Context Effects in Letter Perception: Part I, an Account of Basic Findings," Psychol. Rev. **88**: 375–407 (1981).
Meadows JC, "The Anatomical Basis of Prosopagnosia," J. of Neurol, Neurosurgery, and Psychiatry **37**: 489–501 (1974).
Mel BW, "A Connectionistic Learning Model for Three-Dimensional Mental Rotation, Zoom, and Pan," 8th Ann. Conf. Cogn. Sci. Soc., Amherst, Mass., pp562–571 (1986).
Mervis CB and Greco C, "Parts and Early Conceptual Development: Comment on Tversky and Hemenway," J. Exp. Psychol.: General **113** (2): 194–197 (1984).
Miller WF and Shaw AC, "Linguistic Methods in Picture Processing–A Survey," AFIPS Fall Joint Comp. Conf. **33**: 279–290 (1968).
Minsky M, "A Framework for Representing Knowledge," In: Winston P (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, pp211–277 (1975).

Mishkin M, Ungerleider LG, and Macko KA, "Object Vision and Spatial Vision: Two Cortical Pathways," Trends in Neurosci. **6**: 414–417 (1983).
Mishkin M and Appenzeller T, "The Anatomy of Memory," Sci. Am. **256** (6): 80–89 (1987).
Mitchison GJ and McKee SP, "Interpolation in Stereoscopic Matching," Nature **315**: 402–404 (1985).
Mohr R, "A Refinement of a Spherical Decomposition Algorithm," IEEE Trans. Pattern Analysis and Machine Intell. **4**: 51–51 (1982).
Mohr R and Bajcsy R, "Packing Volumes by Spheres," IEEE Trans. Pattern Analysis and Machine Intell. **5**: 111–116 (1983).
Mokhtarian F and Mackworth A, "Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes," IEEE Trans. Pattern Analysis and Machine Intell. **8**: 34–43 (1986).
Moran J and Desimone R, "Selective Attention Gates Visual Processing in the Extrastriate Cortex," Science **229**: 782–784 (1985).
Morgan MJ, "Mental Rotation: A Computationally Plausible Account of Transformation Through Intermediate Steps," Perception **12**: 203–211 (1983).
Morton J, "The Interaction of Information in Word Recognition," Psychol. Rev. **76**: 165–178 (1969).
Mountcastle VB, "The View from Within: Pathways to the Study of Perception," Johns Hopkins Med. J. **136**: 109–131 (1975).
Mountcastle VB, "An Organising Principle for Cerebral Function: The Unit Module and the Distributed System," In: Edelman GM and Mountcastle VB, *The Mindful Brain*, The MIT Press, Cambridge, Mass. (1978).
Murphy GL and Brownell HH, "Category Differentiation in Object Recognition: Typicality Constraints on the Basic Category Advantage," J. Exp. Psychol.: Learning, Memory and Cognition **11**: 70–84 (1985).
Murray EA and Mishkin M, "Amygdalectomy Impairs Crossmodal Association in Monkeys," Science **228**: 604–606 (1985).
Nackman LR, "Curvature Relations in Three-Dimensional Symmetric Axes," Computer Graphics and Image Processing **20**: 43–57 (1982).
Nackman LR and Pizer SM, "Three-Dimensional Shape Description Using the Symmetric Axis Transform I: Theory," IEEE Trans. Pattern Analysis and Machine Intell. **7**: 187–202 (1985).
Nahin PJ, "The Theory and Measurement of a Silhouette Descriptor for Image Pre-Processing and Recognition," Pattern Recogn. **6**: 85–95 (1974).
Nakayama K and Silverman GH, "Serial and Parallel Processing of Visual Feature Conjunctions," Nature **320**: 264–265 (1986).
Naveh-Benjamin M, "The Effect of Complexity on Interpreting 'Chernoff' Faces," Human Factors **24**: 11–18 (1982).
Navon D, "Forest Before Trees: The Perception of Global Features in Visual Perception," Cogn. Psychol. **9**: 353–383 (1977).
Nevatia R and Binford TO, "Description and Recognition of Curved Objects," Artif. Intell. **8**: 77–98 (1977).
Neveu CF, Dyer CR, and Chin RT, "Two-Dimensional Object Recognition Using Multiresolution Models," Computer Vision, Graphics, and Image Processing **34**: 52–65 (1986).
Newman EA and Hartline PH, "The Infrared 'Vision' of Snakes," Sci. Am. **246** (3): 116–127 (1982).
Nishihara HK, "Intensity, Visible-Surface, and Volumetric Representations," Artif. Intell. **17**: 265–284 (1981).
O'Keefe J, "A Review of the Hippocampal Place Cells," Progress in Neurobiology **13**: 419–434 (1979).
O'Rourke J and Badler N, "Decomposition of Three-Dimensional Objects into Spheres," IEEE Trans. Pattern Analysis and Machine Intell. **1**: 295–305 (1979).
Oshima M and Shirai Y, "Object Recognition Using Three-Dimensional Information," IEEE Trans. Pattern Analysis and Machine Intell. **5**: 353–361 (1983).
Pandya DN and Seltzer B, "Association Areas of the Cerebral Cortex," Trends in Neurosci. **5**: 386–390 (1982).
Pavlidis T, "Algorithms for Shape-Analysis of Contours and Wave Forms," IEEE Trans. Pattern Analysis and Machine Intell. **2**: 301–312 (1980).
Pentland AP, "Local Shading Analysis," IEEE Trans. Pattern Analysis and Machine Intell. **6**: 170–187 (1984).
Perrett DI, Rolls ET, and Caan W, "Visual Neurones Responsive to Faces in the Monkey Temporal Cortex," Exp. Brain Res. **47**: 329–342 (1982).
Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, and Jeeves MJ, "Neurones Responsive to Faces in the Temporal Cortex: Studies of Functional Organization, Sensitivity to Identity and Relation to Perception," Human Neurobiol. **3**: 197–208 (1984).
Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, and Jeeves MA, "Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction," Proc. R. Soc. Lond. B **223**: 293–317 (1985).

Perry VH and Cowey A, "The Ganglion Cell and Cone Distributions in the Monkey's Retina: Implications for Central Magnification Factors," Vision Res. **25**: 1795–1810 (1985).

Persoon E and Fu KS, "Shape Discrimination Using Fourier Descriptors," IEEE Trans. Systems, Man, and Cybernetics. **7**: 170–179 (1977).

Phillips RJ and Rawles RE, "Recognition of Upright and Inverted Faces: A Correlational Study," Perception **8**: 577–583 (1979).

Pinker S, "Visual Cognition: An Introduction," Cognition **18**: 1–63 (1984).

Pitts W and McCulloch WS, "How We Know Universals. The Perception of Auditory and Visual Forms," Bull. Math. Biophysics **9**: 127–147 (1947).

Pizer SM, Koenderink JJ, Lifshitz LM, Helmink L, and Kaasjager ADJ, "An Image Description for Object Definition, Based on Extremal Regions in the Stack," In: Bacharach SL (ed.), *Information Processing in Medical Imaging*, Nijhoff Publ., Dordrecht, pp24–37 (1986).

Pizer SM, Oliver WR, and Bloomberg SH, "Hierarchical Shape Description Via the Multiresolution Symmetric Axis Transform," IEEE Trans. Pattern Analysis and Machine Intell. **9**: 505–511 (1987).

Plantinga H and Dyer C, "The Asp: A Continuous Viewer-Centered Representation for 3D Object Recognition," Computer Sci. Dept., Univ. of Wisconsin, Madison, Report 682 (1987a).

Plantinga H and Dyer C, "The Aspect Representation," Computer Sci. Dept., Univ. of Wisconsin, Madison, Report 683 (1987b).

Poggio GF and Poggio T, "The Analysis of Stereopsis," Ann. Rev. Neurosci. **7**: 379–412 (1984).

Poggio GF, Motter BC, Squatrito S, and Trotter Y, "Responses of Neurons in Visual Cortex (V1 and V2) of the Alert Macaque to Dynamic Random-Dot Stereograms," Vision Res. **25**: 397–406 (1985).

Pogorelov AV, *Extrinsic Geometry of Convex Surfaces*, Am. Math. Soc., Providence, Rhode Island (1973).

Poizner H and Lane H, "Cerebral Assymmetry in the Perception of American Sign Language," Brain and Language **7**: 210–222 (1979).

Poizner H, Battison R, and Lane H, "Cerebral Asymmetry for Perception of American Sign Language: The Effects of Moving Stimuli," Brain and Language **7**: 351–362 (1979).

Poizner H, Bellugi U, and Iragui V, "Apraxia and Aphasia for a Visual-Gestural Language," Am. J. Physiol. **246**: 868–883 (1984a).

Poizner H, Kaplan E, and Bellugi U, "Visual Spatial Processing in Deaf Brain-Damaged Signers," Brain and Cognition **3**: 281–306 (1984b).

Pomerantz JR, Sager LC, and Stoever RJ, "Perception of Wholes and of Their Component Parts: Some Configural Superiority Effects," J. Exp. Psychol: Human Percept. and Perform. **3**: 422–435 (1977).

Prinzmetal W, "Principles of Feature Integration in Visual Perception," Perception & Psychophysics **30**: 330–340 (1981).

Prinzmetal W and Millis-Wright M, "Cognitive and Linguistic Factors Affect Visual Feature Integration," Cogn. Psychol. **16**: 305–340 (1984).

Prinzmetal W, Presti DE, and Posner ME, "Does Attention Affect Visual Feature Integration?" J. Exp. Psychol.: Human Perception and Performance **12**: 361–369 (1986).

Pylyshyn ZW, "The Rate of "Mental Rotation" of Images: A Test of a Holistic Analogue Hypothesis," Memory & Cognition **7**: 19–28 (1979).

Ratcliff G and Newcombe F, "Object Recognition: Some Deductions from the Clinical Evidence," In: Ellis AW (ed.), *Normality and Pathology in Cognitive Functions*, Academic Press, London, pp147–171 (1982).

Requicha AAG, "Representations for Rigid Solids: Theory, Methods, and Systems," Computing Surveys **12**: 437–464 (1980).

Richard CW and Hemami H, "Identification of Three-Dimensional Objects Using Fourier Descriptors of the Boundary Curve," IEEE Trans. Systems, Man, and Cybernetics **4**: 371–378 (1974).

Richards WA and Hoffman DD, "Codon Constraints on Closed 2D Shapes," Computer Vision, Graphics, and Image Processing **31**: 265–281 (1985).

Richards WA, Koenderink JJ, and Hoffman DD, "Inferring 3D Shapes from 2D Codons," MIT AI Memo 840 (1985).

Richards WA, Dawson B, and Whittington D, "Encoding Contour Shape by Curvature Extrema," J. Opt. Soc. Am. A **3**: 1483–1492 (1986).

Richmond BJ, Wurtz RH, and Soto T, "Visual Responses of Inferior Temporal Neurons in Awake Rhesus Monkey," J. Neurophysiol. 50: 1415–1432 (1983).

Rizzo M, Corbett JJ, Thompson HS, and Damasio AR, "Spatial Contrast Sensitivity in Facial Recognition," Neurology **36**: 1254–1256 (1986).

Rock I, *Orientation and Form*, Academic Press, New York (1973).
Rock I, "The Perception of Disoriented Figures," Sci. Am. **230** (1): 78–85 (1974).
Rock I, DiVita J, and Barbeito R, "The Effect on Form Perception of Change of Orientation in the Third Dimension," J. of Exp. Psychol.: Human Perception and Performance **7**: 719–732 (1981).
Rock I, *The Logic of Perception*, The MIT Press, Cambridge, Mass. (1983).
Rock I and DiVita J, "A Case of Viewer-Centered Object Perception," Cogn. Psychol. **19**: 280–293 (1987).
Rodieck RW, "Visual Pathways," Ann. Rev. Neurosci. **2**: 193–225 (1979).
Rolls ET, "Neurons in the Cortex of the Temporal Lobe and in the Amygdala of the Monkey with Responses Selective for Faces," Human Neurobiol. **3**: 209–222 (1984).
Rolls ET, Baylis GC, and Leonard CM, "Role of Low and High Spatial Frequencies in the Face-Selective Responses of Neurons in the Cortex in the Superior Temporal Sulcus in the Monkey," Vision Res. **25**: 1021–1035 (1985).
Rolls ET, "Information Representation, Processing, and Storage in the Brain: Analysis at the Single Neuron Level," In: Changeux JP and Konishi M (eds.), *The Neural and Molecular Bases of Learning*, John Wiley & Sons, Cambridge, Mass., pp503–540 (1987).
Rosch E, "Basic Objects in Natural Categories," Cogn. Psychol. **8**: 382–439 (1976).
Rose D and Dobson VG (eds.), *Models of the Visual Cortex*, John Wiley & Sons, New York (1985).
Rosenfeld A, Hummel RA, and Zucker SW, "Scene Labeling by Relaxation Operations," IEEE Trans. Systems, Man, and Cybernetics **6**: 420–433 (1976).
Rutkowski WS, Peleg S, and Rosenfeld A, "Shape Segmentation Using Relaxation," IEEE Trans. Pattern Analysis and Machine Intell. **3**: 368–439 (1981).
Sadjadi FA and Hall EL, "Three-Dimensional Moment Invariants," IEEE Trans. Pattern Anal. and Machine Intell. **2**: 127–439 (1980).
Sagi D and Julesz B, " 'Where' and 'What' in Vision," Science **228**: 1217–1219 (1985).
Sakata H, Shibutani H, KawanoK, and Harrington TL, "Neural Mechanisms of Space Vision in the Parietal Association Cortex of the Monkey," Vision Res. **25**: 453–463 (1985).
Schiller PH and Malpeli JG, "Functional Specificity of Lateral Geniculate Nucleus Laminae of the Rhesus Monkey," J. Neurophysiol. **41**: 788–797 (1978).
Schiller PH, "Central Connections of the Retinal ON and OFF Pathways," Nature **297**: 580–583 (1982).
Schiller PH, Sandell JH, and Maunsell JHR, "Functions of the ON and OFF Channels of the Visual System," Nature **322**: 824–825 (1986).
Schnapf JL and Baylor DA, "How Photoreceptor Cells Respond to Light," Sci. Am. **256** (4): 40–47 (1987).
Schneider GE, "Two Visual Systems," Science **163**: 895–902 (1969).
Schoene H, *Spatial Orientation*, Princeton Univ. Press, Princeton, New Jersey (1984).
Scholten DK and Wilson SG, "Chain Coding with a Hexagonal Grid," IEEE Trans. Pattern Analysis and Machine Intell. **5**: 526–533 (1983).
Schwartz EL, "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception," Biol. Cybern. **25**: 181–194 (1977a).
Schwartz EL, "Afferent Geometry in the Primate Visual Cortex and the Generation of Neuronal Trigger Features," Biol. Cybern. **28**: 1–14 (1977b).
Schwartz EL, "Computational Anatomy and Functional Architecture of Striate Cortex: A Spatial Mapping Approach to Perceptual Coding," Vision Res. **20**: 645–669 (1980).
Schwartz EL, Desimone R, Albright TD, and Gross CG, "Shape Recognition and Inferior Temporal Neurons," Proc. Natl. Acad. Sci. USA **80**: 5776–5778 (1983).
Schwartz EL, "Anatomical and Physiological Correlates of Visual Computation from Striate to Infero-Temporal Cortex," IEEE Trans. Systems, Man, and Cybernetics **14**: 257–271 (1984).
Selfridge OG and Neisser U, "Pattern Recognition by Machine," Sci. Am. **203**: 60–68 (1960).
Sergent J and Bindra D, "Differential Hemispheric Processing of Faces: Methodological Considerations and Reinterpretation," Psychol. Bull. 89: 541–554 (1981).
Shapiro LG and Haralick RM, "Decomposition of Two-Dimensional Shape by Graph-Theoretic Clustering," IEEE Trans. Pattern Analysis and Machine Intell. **1**: 10–20 (1979).
Shapiro LG, "A Structural Model of Shape," IEEE Trans. Pattern Analysis and Machine Intell. **2**: 111–126 (1980).
Shapiro LG, Moriarty JD, Haralick RM, and Mulgaonkar PG, "Matching Three-Dimensional Objects Using a Relational Paradigm," Pattern Recogn. **17**: 385–405 (1984).

Shapiro LG, "Recent Progress in Shape Decomposition and Analysis," In: Kanal LN and Rosenfeld A (eds.), *Progress in Pattern Recognition 2*, Elsevier Science Publ., pp113–123 (1985).
Shapiro LG and Haralick RM, "A Metric for Comparing Relational Descriptions," IEEE Trans. Pattern Analysis and Machine Intell. **7**: 90–94 (1985).
Shapley R and Lennie P, "Spatial Frequency Analysis in the Visual System," Ann. Rev. Neurosci. **8**: 547–583 (1985).
Shapley R and Perry VH, "Cat and Monkey Retinal Ganglion Cells and their Visual Functional Roles," Trends in Neurosci. **9**: 229–235 (1986).
Sheng Y and Arsenault HH, "Experiments on Pattern Recognition Using Invariant Fourier-Mellin Descriptors," J. Opt. Soc. Am. A **3**: 771–776 (1986).
Sheng Y and Arsenault HH, "Circular-Fourier-Radial-Mellin Transform Descriptors For Pattern Recognition," J. Opt. Soc. Am. A **3**: 885–888 (1986).
Shepard RN and Chipman S, "Second-Order Isomorphism of Internal Representations: Shapes of States," Cogn. Psychol. **1**: 1–17 (1970).
Shepard RN and Metzler J, "Mental Rotation of Three-Dimensional Objects," Science **171**: 701–703 (1971).
Shepard RN and Judd SA, "Perceptual Illusion of Rotation of Three-Dimensional Objects," Science **191**: 952–954 (1976).
Shepard RN, "The Mental Image," Am. Psychol. **33**: 125–137 (1978).
Shepard RN, "Perceptual and Analogical Bases of Cognition," In: Mehler J, Walker ECT and Garrett M (eds.), *Perspectives on Mental Representation*, Lawrence Erlbaum Assoc., Hillsdale, New Jersey, p49–67 (1982).
Shepard RN, "Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imagining, Thinking, and Dreaming," Psychol. Rev. **91**: 417–447 (1984).
Shepherd J, Davies G, and Ellis H, "Studies of Cue Saliency," In: Davies G, Ellis H, and Shepherd J (eds.), *Perceiving and Remembering Faces*, Academic Press, London (1981).
Sherman SM, "The Functional Significance of X and Y cells in Normal and Visually Deprived Cats," Trends in Neurosci. **2**: 192–195 (1979).
Sherman SM and Koch C, "The Anatomy and Physiology of Gating Retinal Signals in the Mammalian Lateral Geniculate Nucleus," MIT AI Memo 825 (1985).
Shinar D and Owen DH, "Effects of Form Rotation on the Speed of Classification: The Development of Shape Constancy," Perception & Psychophysics **14** (1): 149–154 (1973).
Shipp S and Zeki S, "Segregation of Pathways Leading from Area V2 to areas V4 and V5 of Macaque Monkey Visual Cortex," Nature **315**: 322–325 (1985).
Silberberg TM, Davis L, and Harwood D, "An Iterative Hough Procedure for Three-Dimensional Object Recognition," Pattern Recogn. **17**: 621–629 (1984).
Silberberg TM, Harwood DA, and Davis LS, "Object Recognition Using Model Points," Computer Vision, Graphics, and Image Processing **35**: 47–71 (1986).
Smith DA, "Using Enhanced Spherical Images for Object Representation," MIT AI Memo No. 530 (1979).
Sparks DL and Nelson JS, "Sensory and Motor Maps in the Mammalian Superior Colliculus," Trends in Neurosci. **10**: 312–317 (1987).
Srebro R, "Localization of Cortical Activity Associated with Visual Recognition in Humans," J. Physiol. **360**: 247–259 (1985).
St.John RC, "Lateral Asymmetry in Face Perception," Can. J. Psychol. **35**: 213–223 (1981).
Sterling P, "Microcircuitry of the Cat Retina," Ann. Rev. Neurosci. **6**: 149–185 (1983).
Sterling P, Freed M, and Smith RG, "Microcircuitry and Functional Architecture of the Cat Retina," Trends in Neurosci. **9**: 186–192 (1986).
Stevens KA, "The Visual Interpretation of Surface Contours," Artif. Intell. **17**: 47–73 (1981a).
Stevens KA, "The Information Content of Texture Gradients," Biol. Cybern. **42**: 95–105 (1981b).
Stone J and Dreher B, "Parallel Processing of Information in the Visual Pathways. A General Principle of Sensory Encoding?" Trends in Neurosci. **5**: 441–446 (1982).
Stryer L, "The Molecules of Visual Excitation," Sci. Am. **257** (1): 42–50 (1987).
Sutherland NS, "Outlines of a Theory of Visual Pattern Recognition in Animals and Man," Proc. R. Soc. London B **171**: 297–317 (1968).
Sutherland NS, "The Representation of Three-Dimensional Objects," Nature **278**: 395–398 (1979).
Swindale NV, "Parallel Channels and Redundant Mechanisms in Visual Cortex," Nature **322**: 775–776 (1986).
Thompson P, "Margaret Thatcher: A New Illusion," Perception **9**: 483–484 (1980).

Tolhurst DJ and Movshon JA, "Spatial and Temporal Contrast Sensitivity of Striate Cortical Neurons," Nature **257**: 674–675 (1975).
Tootell RBH, Silverman MS, Switkes E, and De Valois RL, "Deoxyglucose Analysis of Retinotopic Organization in Primate Striate Cortex," Science **218**: 902–904 (1982).
Tootell RBH, Silverman MS, De Valois RL, and Jacobs GH, "Functional Organization of the Second Cortical Visual Area in Primates," Science **220**: 737–739 (1983).
Tranel D, Fowles DC, and Damasio AR, "Electrodermal Discrimination of Familiar and Unfamiliar Faces: A Methodology," Psychophysiology **22**: 403–408 (1985).
Tranel D and Damasio AR, "Knowledge Without Awareness: An Autonomic Index of Facial Recognition by Prosopagnosics," Science **228**: 1454–1455 (1985).
Treisman AM and Gelade G, "A Feature-Integration Theory of Attention," Cogn. Psychol. **12**: 87–136 (1980).
Treisman AM and Paterson R, "Emergent Features, Attention, and Object Perception," J. Exp. Psychol.: Human Perception and Performance **10**: 12–31 (1984).
Tsai WH and Yu SS, "Attributed String Matching with Merging for Shape Recognition," IEEE Trans. Pattern Analysis and Machine Intell. **7**: 453–462 (1985).
Turney JL, Mudge TN, and Volz RA, "Recognizing Partially Occluded Parts," IEEE Trans. Pattern Analysis and Machine Intell. **7**: 410–421 (1985).
Tversky A, "Features of Similarity," Psychol. Rev. **84**: 327–352 (1977).
Tversky B and Hemenway K, "Objects, Parts, and Categories," J. Exp. Psychol.: General **113**: 169–193 (1984).
Uhr L (ed.), *Pattern Recognition*, John Wiley & Sons, New York, (1966).
Ullman S, *The Interpretation of Visual Motion*, The MIT Press, Cambridge, Mass. (1979).
Ullman S, "Visual Routines," Cognition **18**: 373–415 (1984a).
Ullman S, "Maximizing Rigidity: The Incremental Recovery of 3-D Structure from Rigid and Nonrigid Motion," Perception **13**: 255–274 (1984b).
Ullman S, "Artificial Intelligence and the Brain," Ann. Rev. Neurosci. **9**: 1–26 (1986).
Uttal WR, *An Autocorrelation Theory of Form Detection*, Lawrence Erlbaum Ass., Hillsdale, New Jersey (1975).
Van Essen DC, "Visual Areas of the Mammalian Cerebral Cortex," Ann. Rev. Neurosci. **2**: 227–263 (1979).
Van Essen DC and Maunsell JHR, "Hierarchical Organization and Functional Streams in the Visual Cortex," Trends in Neurosci. **6**: 370–375 (1983).
Van Essen DC, "Functional Organization of Primate Visual Cortex," In: Peters A and Jones EG (eds.), *Cerebral Cortex. Vol. 3: Visual Cortex*, Plenum Press, New York, pp259–329 (1985).
Van Hove PL and Verly JG, "A Silhouette-Slice Theorem for Opaque 3-D Objects," IEEE Proc. **221**: 933–936 (1985).
Vernon MD, "The Nature of Perception and the Fundamental Stages in the Process of Perceiving," In: Uhr L (ed.), *Pattern Recognition*, John Wiley & Sons, New York, pp61–83 (1966).
Wallace TP and Wintz PA, "An Efficient Three-Dimensional Aircraft Recognition Algorithm Using Normalized Fourier Descriptors," Computer Graphics and Image Processing **13**: 99–126 (1980).
Wallace TP, Mitchell OR, and Fukunaga K, "Three-Dimensional Shape Analysis Using Local Shape Descriptors ," IEEE Trans. Pattern Analysis and Machine Intell. **3**: 310–323 (1981).
Waltz D, "Understanding Line Drawings of Scenes with Shadows," In: Winston P (ed.), *The Psychology of Computer Vision*, McGrawe-Hill, New York (1975).
Wang YF, Magee MJ, and Aggarwal JK, "Matching Three-Dimensional Objects Using Silhouettes," IEEE Trans. Pattern Analysis and Machine Intell. **6**: 513–518 (1984).
Warren C and Morton J, "The Effects of Priming on Picture Processing," Br. J. Psychol. **73**: 117–129 (1982).
Warrington EK and Taylor AM, "The Contribution of the Right Parietal Lobe to Object Recognition," Cortex **9**: 152–164 (1973).
Warrington EK, "The Selective Impairment of Semantic Memory," Q. J. of Exp. Psychol. **27**: 635–657 (1975).
Warrington EK and Taylor AM, "Two Categorical Stages of Object Recognition," Perception **7**: 695–705 (1978).
Warrington EK, "Neuropsychological Studies of Object Recognition," Phil. Trans. R. Soc. London B **298**: 15–33 (1982).
Warrington EK and James M, "Visual Object Recognition in Patients With Right-Hemisphere Lesions: Axes or Features?" Perception **15**: 355–366 (1986).
Weisstein N and Harris CS, "Visual Detection of Line Segments: An Object-Superiority Effect," Science **186**: 752–755 (1974).
Wiener N, *Cybernetics*, The MIT Press, Cambridge, Mass. (1961, 2nd ed.).
Wiesel TN and Gilbert CD, "Visual Cortex," Trends in Neurosci. **9**: 509–512 (1986).

Williams A and Weisstein N, "Line Segments are Perceived Better in a Coherent Context than Alone: An Object-Line Effect in Visual Perception," Memory & Cognition **6** (2): 85–90 (1978).
Wilson HR and Bergen JR, "A Four Mechanism Model for Threshold Spatial Vision," Vision Res. **19**: 19–32 (1979).
Winograd T and Flores F, *Understanding Computers and Cognition*, Ablex Publ. Corp., Norwood, New Jersey (1986).
Wiser M, "The Role of Intrinsic Axes in Shape Recognition," Proc. 3rd Ann. Conf. Cogn. Sci. Soc., Berkeley, Calif., pp184–186 (1981).
Witkin AP, "Recovering Surface Shape and Orientation from Texture," Artif. Intell. **17**: 17–45 (1981).
Witkin AP, "Scale-Space Filtering," Proc. 8th Int. Joint Conf. Artif. Intell., Karlsruhe, West Germany, pp1019–1022 (1983).
Wolf W, Hicks TP, and Albus K, "The Contribution of GABA-Mediated Inhibitory Mechanisms to Visual Response Properties of Neurons in the Kitten's Striate Cortex," J. Neurosci. **6**: 2779–2795 (1986).
Wolford G, "Perturbation Model for Letter Identification," Psychol. Rev. **82**: 184–199 (1975).
Wolford G and Shum KT, "Evidence for Feature Perturbations," Perception & Psychophysics **27**: 409–420 (1980).
Wu R and Stark H, "Three-Dimensional Object Recognition From Multiple Views," J. Opt. Soc. Am. A **3**: 1543–1557 (1986).
Yin RK, "Face Recognition by Brain-Injured Patients: A Dissociable Ability?" Neuropsychol. **8**: 395–402 (1970).
Young AW and Bion PJ, "Absence of Any Developmental Trend in Right Hemisphere Superiority for Face Recognition," Cortex **16**: 213–221 (1980).
Young AW and Bion PJ, "Accuracy of Naming Laterally Presented Known Faces by Children and Adults," Cortex **17**: 97–106 (1981).
Young AW, Hay DC, and McWeeny KH, "Right Cerebral Hemisphere Superiority for Constructing Facial Representations," Neuropsychol. **23**: 195–202 (1985).
Young AW, McWeeny KH, Hay DC, and Ellis AW, "Access to Identity-Specific Semantic Codes from Familiar Faces," Q. J. Exp. Psychol. **38**: 271–295 (1986a).
Young AW, McWeeny KH, Ellis AW, and Hay DC, "Naming and Categorizing Faces and Written Names," Q. J. Exp. Psychol. **38**: 297–318 (1986b).
Young AW, McWeeny KH, Hay DC, and Ellis AW, "Matching Familiar and Unfamiliar Faces on Identity and Expression," Psychol. Res. **48**: 63–68 (1986c).
Zeki S, "Functional Specialization in the Visual Cortex of the Rhesus Monkey," Nature **274**: 423–428 (1978).
Zipser D, "A Computational Model of Hippocampal Place Fields," Behav. Neurosci. **99**: 1006–1018 (1985).
Zucker SW, Krishnamurthy EV, and Haar RL, "Relaxation Processes for Scene Labeling: Convergence, Speed, and Stability," IEEE Trans. Systems, Man, and Cybernetics **8**: 41–48 (1978).
Zusne L, *Visual Perception of Form*, Academic Press, New York (1970).