A Survey of Visual Preprocessing and Shape Representation Techniques

Bruno A. Olshausen

November 1988

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 88.35

NASA Cooperative Agreement Number NCC 2-408

RIACS
Research Institute for Advanced Computer Science

A Survey of Visual Preprocessing and Shape Representation Techniques

Bruno A. Olshausen

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 88.35
November 1988

Abstract. This survey summarizes many recent theories and methods proposed for visual preprocessing and shape representation. The survey brings together research from the fields of biology, psychology, computer science, electrical engineering, and most recently, neural networks. This report was motivated by the need to preprocess images for a sparse distributed memory (SDM), but the techniques presented herein may also prove useful for applying other associative memories to visual pattern recognition. The material of this survey is divided into three sections 1) an overview of biological visual processing, 2) methods of preprocessing (extracting parts of shape, texture, motion, and depth), and 3) shape representation and recognition (form invariance, primitives and structural descriptions, and theories of attention).

Work reported herein was supported in part by Cooperative Agreement NCC 2-408 between the National Aeronautics and Space Administration (NASA) and the Universities Space Research Association (USRA).

Table of Contents

1. Introduction

Consider for a moment how vision is used by a few living things: The sand wasp (*Philanthus triangulum*) is able to recognize its nest by the pattern of objects lying around it; the kingfisher can target a fish underwater while hovering in the air; and humans can read text in many fonts and sizes. Each of these tasks seems to be performed almost effortlessly, yet any one of them would certainly stump the most powerful of modern digital computers.

How is it that images are processed and understood in biological systems? What is the nature of computation involved in vision, and how might we build machines that "see?" Partial answers to these questions have been offered over the past several decades by researchers in fields of biology, psychology, computer science, and most recently in the burgeoning field of neural networks. This survey is offered as a humble attempt to bring together and summarize many of the recent approaches to visual preprocessing and shape representation that have been proposed.

1.1. The problems of vision

In a sense, vision could be considered the inverse problem of computer graphics. That is, in computer graphics one is given an object with a specified shape, reflectance, illumination, viewing transformation, etc., and then asked to compute the projection onto a 2-D image plane. In vision, one is given only the 2-D image and then asked to compute what created it. This latter problem is underdetermined and hence creates enormous difficulties for machine vision.

Humans are aided to a great extent in visually reconstructing the world by making assumptions about the shape of objects: surfaces are smooth, boundaries are continuous, objects are rigid, etc. When these assumptions fail us we perceive an optical illusion, but most of the time we are unaware of such assumptions and we appear to see everything perfectly. Thus, the apparent ease with which we see can tend to veil the real problems of vision, as Marr (1982) has astutely observed:

> ...in the 1960s almost no one realized that machine vision was difficult. The field had to go through the same experience as the machine translation field did in its fiascoes of the 1950s before it was at last realized that here were some problems that had to be taken seriously. The reason for this misperception is that we humans are ourselves so good at vision. The notion of a feature detector was well established by Barlow and by Hubel and Wiesel, and the idea that extracting edges and lines from images might be at all difficult simply did not occur to those who had not tried to do it. It turned out to be an elusive problem: Edges that are of critical importance from a three-dimensional point of view often cannot be found at all by looking at the intensity changes in an image. Any kind of textured image gives a multitude of noisy edge segments; variations in reflectance and illumination cause no end of trouble; and even if an edge has a clear existence at one point, it is as likely as not to fade out quite soon, appearing only in patches along its length in the image. (p. 16)

In order to begin to understand vision, it is helpful to divide it into two parts: **preprocessing** and **recognition**. The preprocessing part consists of extracting useful features from the image, such as parts of shape, texture, motion, and depth. Such features are used to form a rich description of the visual scene (i.e., something better than simply which pixels are on and which are off) to be fed to the recognition process. Many studies have been done on preprocessing in biological systems - or so-called early vision - and it has been found that this type of processing usually involves many local operations on an image performed almost totally in parallel.

The recognition part consists of the formation of an internal representation of objects and a process for matching or classification based on the description obtained from the preprocessing stage. Somehow, the brain must be capable of representing visual forms independent of such particulars as perspective, illumination, and size. Then, there has to be a process for matching objects from what must be an enormous library of visual forms. Very little is known about how the recognition part is accomplished in biological systems.

The parts of vision may interact, but the problems associated with them may be considered separately, hence simplifying our task in attempting to understand vision.

1.2. Scope of this survey

This survey was motivated by the need to preprocess images for *sparse distributed memory* (SDM). Briefly, SDM provides a simple, massively parallel architecture for an associative memory. Long bit vectors (256-1000 bits, for example) serve as both data and addresses to the memory, and patterns are grouped or classified according to similarity in Hamming distance. (See Kanerva, 1988, for details on SDM, and Keeler, 1988, for a comparison to Hopfield nets.)

In order for SDM to serve as a visual memory, some correspondence must be established between the bits in SDM and the image. Hence, the emphasis of this survey is on preprocessing and representation, with little attention given to classification or matching. In the realm of preprocessing, the emphasis is on extraction of shape, especially 2-D shape, rather than features such as color, motion, 3-D surface orientation, or depth. Also, this survey emphasizes many of the recent applications of neural networks, or biological-type approaches, to visual processing.

Prerequisites. It is assumed that the reader is familiar with neural-networks and also has some background in mathematics. Knowledge of neuroanatomy and neurophysiology is not necessary but would be helpful.

Should the reader be unfamiliar with some of the terminology used in this report, a glossary of technical terms is provided in the Appendix. All terms appearing in *italic font* are defined in the glossary.

Other surveys, books, and collections. Brady (1982) and Binford (1982) have published surveys on the more conventional techniques used in image understanding and machine vision; Horn (1986) and Ballard and Brown (1982) serve as good texts in this area. Arbib and Hanson (1987a) have published a broad overview of theories and techniques used in vision, both in AI and in biological systems, tracing their development from past to present; Fischler and Firschein (1987a) provide a similar perspective. Marr's Vision (Marr, 1982) provides an excellent and insightful analysis of human visual processing from a computational point of view; Pinker (1985a) provides an overview of theories on visual cognition; and a short review by Ballard et al. (1983) discusses some parallel methods for visual computation. Collections edited by Rosenfeld (1986b), Arbib and Hanson (1987b), Fischler and

Firschein (1987b), Schwab and Nusbaum (1986), Pinker (1985b), and Tenenbaum and Barrow (1988) provide a broad assortment of papers covering topics in human and machine vision. As an indication of the vast amount of research going on the computer vision community, Rosenfeld (1988) has compiled a list of over 1400 references to papers on computer vision and image analysis published during 1987.

1.3. Organization of this survey

This survey is organized according to both function and methodology. Section 2, which gives an overview of the current state of knowledge in biological visual processing, is devoted largely to methodology. Section 3 (preprocessing) and section 4 (shape representation and recognition) are organized according to function (edge-detection or form invariance, for example), with particular methods discussed within each functional sub-section.

The bibliography provided at the end of this report is mostly annotated. Some papers are without a summary, which means either that they could not be obtained or that they were not thoroughly read; these papers were included in the bibliography anyway because they may be of interest in the future, or they may be of importance to others.

1.4. Acknowledgments

Pentti Kanerva and Mike Raugh were generous in allotting me the time to thoroughly review the literature necessary for producing this report. Conversations with David Rogers, Jim Keeler, Louis Jaeckel, and David Li were very helpful for providing insight and keeping me going in the right direction. Thanks are also due to Al Ahumada and Mike Raugh for helpful comments on the draft.

2. An Overview of Biological Visual Processing

It is the purpose of this section to summarize the current state of knowledge in biological visual processing. Since many of the techniques and theories discussed in later sections are based on biological models, the terms defined here will become useful later on.

The overall visual processing scheme for most mammals is shown in Figure 1. Visual patterns are captured by the *retina*, and then sent to the *visual cortex* via the *lateral geniculate nucleus*. It should be noted that while much of the knowledge about these areas has been gained from studies on the visual systems in the cat and macaque monkey, most aspects of the organization and function of these areas apply to the human visual system as well.

The material in this section has been extracted largely from Kuffler et al. (1984), Baylor and Shatz (1988), Hubel and Wiesel (1979), Nauta and Feirtag (1979), Van Essen and Maunsell (1983), and Van Essen (1985).

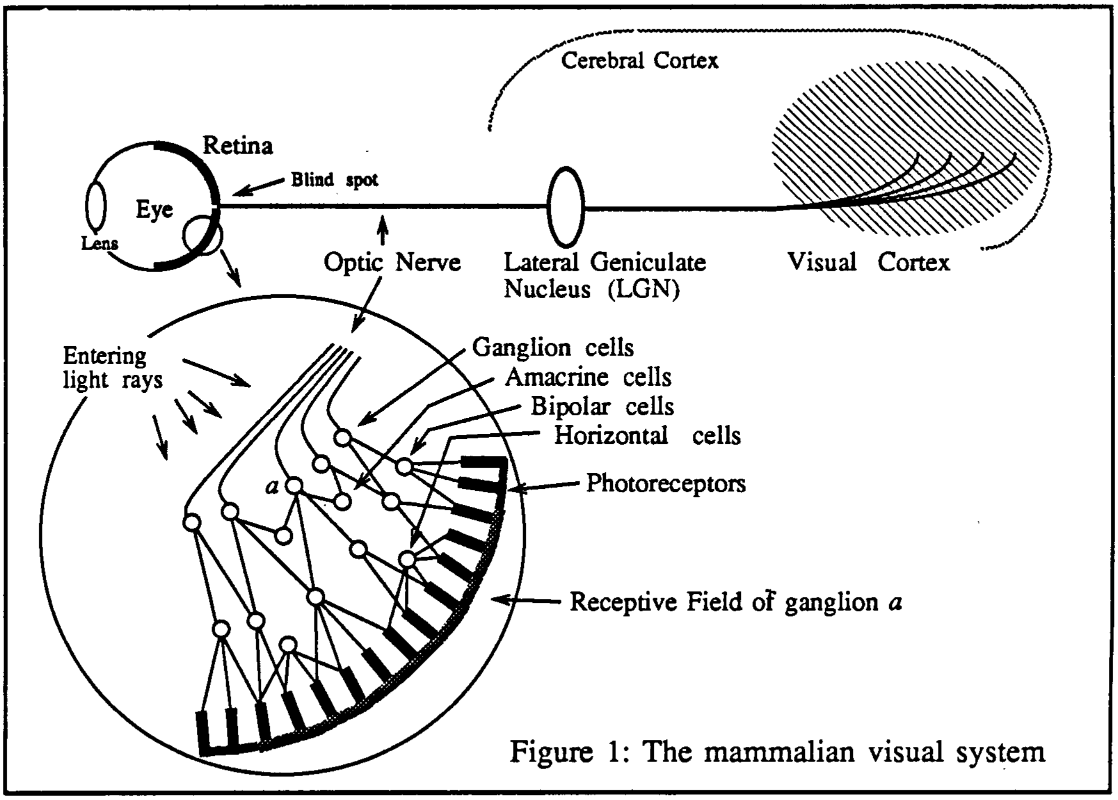

Figure 1: The mammalian visual system

2.1. Retina

Light is focused by the lens of the eye onto the retina, which is a vast array of photoreceptors, interconnecting neurons, and associated wiring (axons and dendrites). Note that light must pass through the wiring and other neurons in order to reach the photoreceptors. In a sense, then, the retina has been wired backwards.

The photoreceptors come in two varieties: *rods* and *cones*. The great majority of photoreceptors are rods, of which there are approximately 120 million packed at an average density of 160,000 per mm^2. Rods are extremely sensitive to light intensity (in fact, one rod can detect a single photon!), but they do not differentiate among the various wavelengths of light (color). Cones, on the other hand, are much fewer in number (6.8 million), require a higher light intensity for activation, and are tuned to respond to different colors. Cones are classified as "red" (561 nm), "green" (531 nm), or "blue" (430 nm) according to the wavelength of light which yields the maximal response. In addition, cones are found mainly at the center of the retina, or *fovea*, packed at an average density of 150,000 per mm^2.

The other neurons in the retina consist of *bipolar cells, horizontal* and *amacrine cells*, and *ganglion cells*. Bipolar cells establish conduction lines from the rods and cones to the ganglion cells, while horizontal and amacrine cells establish inhibitory crosslinks among these conduction lines. The net result of all this interconnectivity is that a ganglion cell effectively collects the outputs from many photoreceptors. (Actually, interactions among neurons in the retina are quite complex; see Loebner, 1987, for example.)

The local group of photoreceptors from which a ganglion cell receives its input is known as a *receptive field.* Ganglion cell receptive fields tend to have *center/surround profiles*, such that uniform illumination on the receptive field elicits no response from a ganglion, but some form of contrast does (see Fig. 2). Receptive fields tend to cover very small areas in the fovea and very large areas in the periphery, making the fovea the area of highest visual acuity.

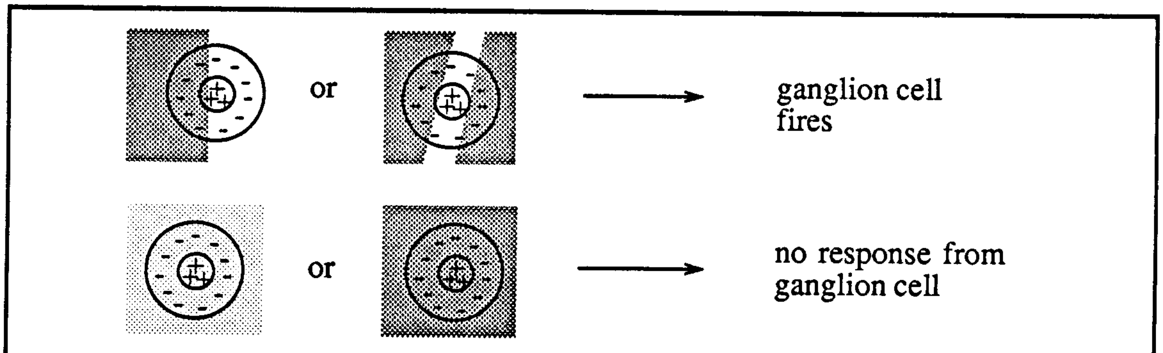

Figure 2: Receptive field profile of a ganglion cell and its response to various spatial patterns. + signs indicate excitory part of the field; - signs indicate inhibitory part of the field.

Retinal ganglion cells may be subdivided into two classes, *X cells* and *Y cells*, on the basis of physical attributes and function. *X* cells are responsive to maintained contrast, have a slow response time, provide high resolution, and they are sensitive to color. *Y* cells, on the other hand, have a transient response to contrast, fast response, but low resolution. It is believed that *X* cells are especially well suited for the analysis of shape, and *Y* cells for the analysis of motion.

The long axons from all the ganglion cells are bundled together to form the *optic nerve*, which exits the eye at the blind spot. (Since there are approximately 1.2 million ganglion cells in the retina, there is an average fan-in of about 100:1 from input to output.) The optic nerve follows two separate paths to the brain. One path leads to the superior colliculus in the midbrain, presumably for the purpose of controlling eye movements (see Sparks and Jay, 1987, for a model of this system). The other path, termed the thalamo-cortical pathway, leads to the visual cortex via the lateral geniculate nucleus (LGN) of the thalamus. The latter path is the one we will follow here.

2.2. Lateral geniculate nucleus (LGN)

Just as there are two eyes, there are also two LGN's, one in each hemisphere of the brain. Each LGN receives axons from the left and right eyes, and the axons terminate in six distinct layers alternating according to left or right eye, as shown in Figure 3. Moreover, the mapping from retina to LGN is topographic, so that neighboring cells in the LGN correspond to neighboring receptive fields in the retina.

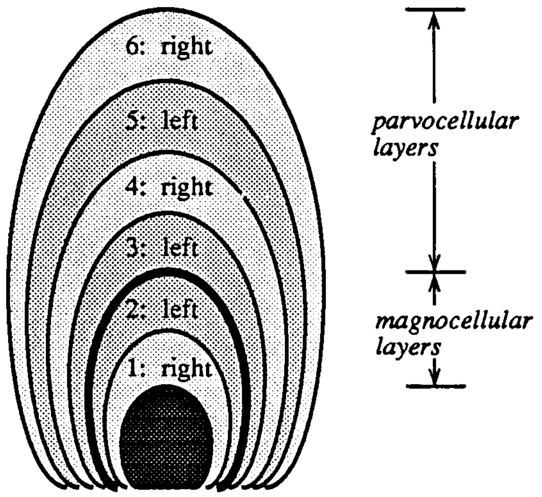

Figure 3: Cross-section of the left LGN. Each layer receives axons from either the left or right eye. Parvocellular layers contain X-type cells; magnocellular layers contain Y-type cells.

The layers of the LGN can be grouped into two parts. Layers 1 and 2, termed the *magnocellular layers,* contain cells that respond like Y cells in the retina. Layers 3-6 are termed the *parvocellular layers* and contain X-like cells.

It is interesting to note that all sensory input (with the exception of olfaction) passes through the thalamus before being processed in the cerebral cortex. Presumably this is done to modify or improve the raw sensory input before being processed by the cerebral cortex. In this case, the LGN (just one part of the thalamus) is bringing together signals from both eyes and grouping signals according to X or Y channels.

2.3. Visual cortex

The visual cortex resides in both hemispheres of the brain at the rear of the cerebral cortex. In humans, it is estimated to occupy 150-250 cm^2, or about 12% of the entire cerebral cortex.

The visual cortex has been subdivided into many different areas, delineated by function and/or neural structure. Each area, as with all of the cerebral cortex, is essentially a two-dimensional layered sheet of neurons. Figure 4 shows how various areas of the visual cortex are interconnected in the macaque monkey. (Many other areas and interconnections are known to exist; this chart is shown here for the sake of simplicity). Interconnections are almost always two-way, such that if area A projects to area B, then

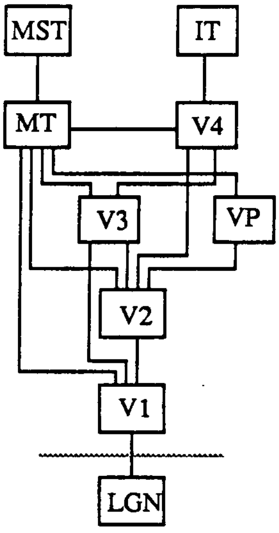

Figure 4: Various areas of the visual cortex and their interconnectivity. Connections between areas are usually two-way. See text for definitions. (Modified from Van Essen and Maunsell, 1983).

area B also projects back to area A. Unnikrishnan et al. (1987) and Miyake and Fukushima (1986) have proposed models to explain this feedback.

Discussed below are several of the more functionally significant areas of the visual cortex.

V1 (area 17, or striate cortex). V1, which receives most of the input from the LGN, comprises approximately 26 cm^2, or 4% of the human cerebral cortex. Like the LGN, V1 contains a more or less topographic map of the visual field. But here, the functional organization is much more complex and the receptive fields are dramatically different.

Essentially three types of cells can be found in V1: *simple cells*, *complex cells*, and *hypercomplex cells*. A simple cell will fire at maximum frequency in response to a small line or edge with a specific angular orientation and position in the visual field. If the orientation is changed by even 20°, the response of the cell will drop by more than 50%; and if the line is not positioned precisely within the receptive field, the response will also drop dramatically. Complex cells are also orientation selective, but exhibit more tolerance to position changes. Hypercomplex cells are responsive to line terminations and corners (also called *endstopped neurons*). It is thought that these three types of cells are connected together in a roughly hierarchical fashion, with simple cells receiving their input from several LGN cells lying along the same line, complex cells receiving their input from several neighboring simple cells of the same orientation preference, and hypercomplex cells receiving excitory input from some complex cells and inhibitory input from others.

The cells of V1 are organized in a columnar fashion according to two parameters: orientation preference and eye preference. That is, cells lying in a column perpendicular to the surface of the cortex are found to respond preferentially to the same orientation and the same eye. All the cells in an *orientation column* are selective to the same angular orientation of a small line or edge in the visual field. Similarly, all the cells in an *ocular dominance column*

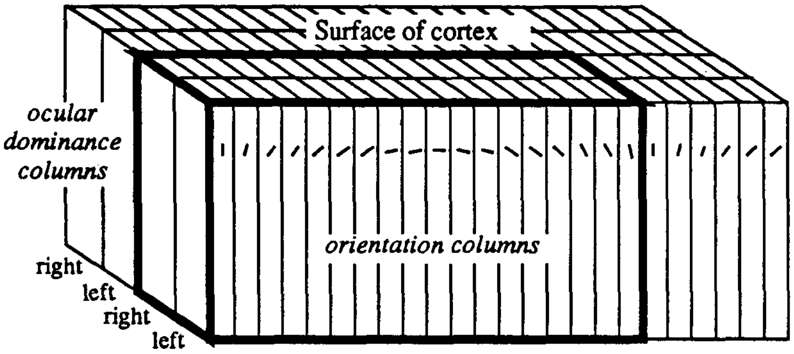

Figure 5: A block of primary visual cortex (V1), showing the columnar organization. Orientation preference changes about 10° every 25-50μm, and eye preference alternates regularly between left and right eye. A hypercolumn is shown in bold outline (horizontal dimension is exaggerated).

are biased toward the same eye. Moving from one column to the next across the surface of the cortex, one finds that orientation preference changes continuously, about 10° every 25-50μm, and eye preference alternates regularly between left and right eye (Fig. 5). About one square millimeter of cortex is needed to contain an entire 180° worth of orientation columns covering both eyes. Hubel and Wiesel have termed each such 1mm^2 portion a *hypercolumn*, of which there are over 4000 in the striate cortex of the monkey.

Although the mapping from retina to V1 is topographic, the map is highly distorted in order to devote more resources to the fovea. For example, close to the fovea, each degree of visual field might have 10 hypercolumns devoted to it, while in the periphery, a degree might be assigned to only a fraction of a hypercolumn.

MT (middle temporal area) and MST (medial superior temporal area). MT neurons are found to be highly selective for direction of motion, speed, binocular disparity, as well as motion in depth. Cells in this area exhibit little or no selectivity to shape or color. Thus, this area seems very well suited for analyzing the three-dimensional trajectories of objects moving in visual space, irrespective of their particular form.

One of the most interesting properties of cells in MST is that some cells respond differently to a moving object when the eyes are stationary than to the equivalent retinal stimulation produced by a stationary object when the eyes are moving. Also, cells in this area tend to have very large receptive fields.

In both areas, the same columnar organization as in V1 is found, except that cells here are grouped according to direction of motion.

V4. V4, along with VP and V2, seems to contain a large fraction of color selective cells. Again, cells are arranged in columns, in this case according to color.

IT (inferotemporal cortex). IT is thought to be heavily involved in visual pattern processing. The cells of IT tend to have very large receptive fields, and so-called *grandmother cells* (cells selective for complex shapes, such as the face of one's grandmother) have been found in this area. Gross (1972) has reported the existence of cells that respond selectively to the silhouette of a monkey's hand, and Perrett (1982) has found neurons in the superior temporal cortex, which receives inputs from IT, that are selectively responsive to faces, or to parts of faces.

3. Preprocessing the Image

The purpose of preprocessing is to extract useful features from the image in order to provide the recognition process with a rich description of a scene. This section discusses how features may be extracted from the raw image - or pixel map - and how they may be used in segmenting the image, or separating figure from ground.

3.1. Edge detection

Since an image is essentially a 2-D distribution of intensity values, one useful operation is to find where, and to what extent, intensity changes occur. From a recognition standpoint, edges are useful because they help define the border or boundary of a shape; and as discussed in Section 2, the retina, LGN, and visual cortex all seem to be actively involved in some form of contrast enhancement or edge detection. Thus, edges are important, and they seem to be detected in biological systems, but what is the computation that underlies the detection of edges?

The Marr-Hildreth edge detector. One of the problems with detecting edges in natural images is that edges generally occur over a wide variety of scales or resolutions. An intensity change may take place over one or two pixels, or it may take place over many pixels. For this reason, Marr and Hildreth (1980) have proposed filtering an image in order to:

1) restrict the range of resolutions over which intensity changes occur in an image,

and

2) maintain the spatial locality of discontinuities (i.e., even though an intensity change may take place over many pixels, you still want to determine its exact location as best as possible).

Marr and Hildreth have found that the filter that best optimizes constraints (1) and (2) is the *Gaussian.*

Once an image has been *Gaussian filtered* at several resolutions (by adjusting σ, the spread of the Gaussian function), the *Laplacian operator* is then applied to find points of maximum intensity change. The Gaussian function and Laplacian operator can be combined into a single filter, called the Laplacian of Gaussian, or *LOG* filter, as illustrated in Figure 6e. Marr and Hildreth have found it useful to plot the zero-crossings of an LOG-filtered image, as these points indicate maxima in the first derivative, which in turn indicate points of maximum intensity change in the image (Fig. 6b-d). When the zero-crossing images from several resolutions are considered together, an edge is indicated by a segment of zero-crossings occurring at the same place for two or more resolutions.

In relating their theory to biological systems, Marr and Hildreth have proposed that the LGN computes the zero-crossings of an LOG filter, and that simple cells compute edge segments (a line of zero crossings). It is interesting to note that Marr and Hildreth's edge detector agrees very nicely with a quantitative model of human spatial vision proposed by Wilson and Bergen (1979). This model predicts that human receptive fields come in four dif-

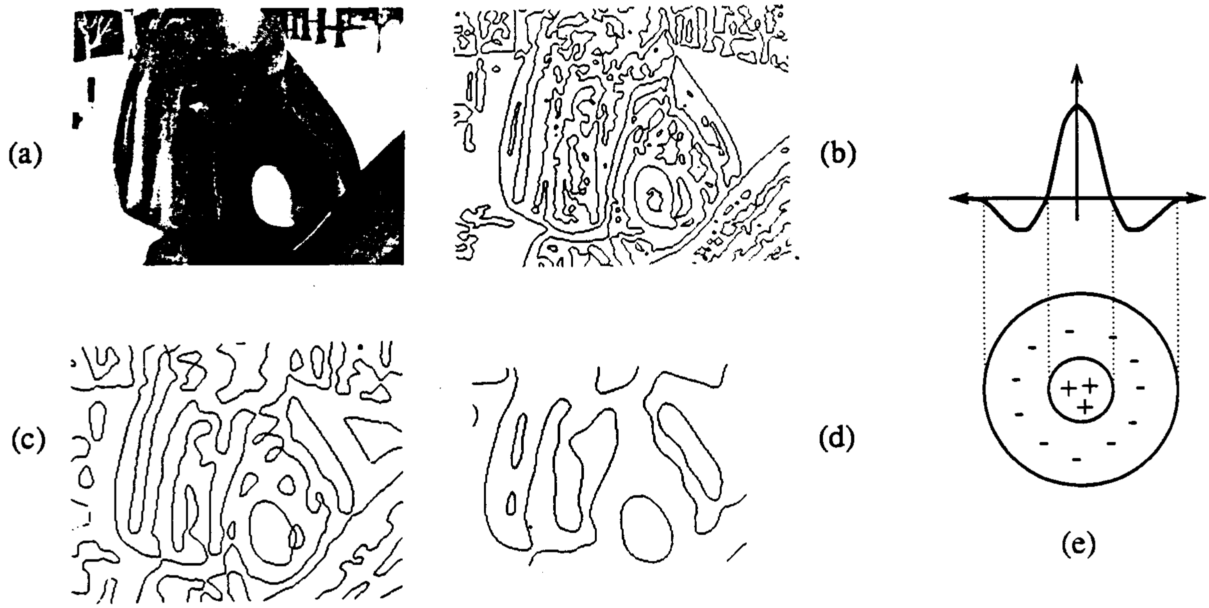

Figure 6: The image (a) is convolved with an LOG filter (e) at several different scales. The zero-crossings of the resulting images are shown in (b-d); the central part of the LOG filter was set at a width of 6 pixels (b), 12 pixels (c), and 24 pixels (d). (a-d copied from Marr and Hildreth, 1980, with permission.)

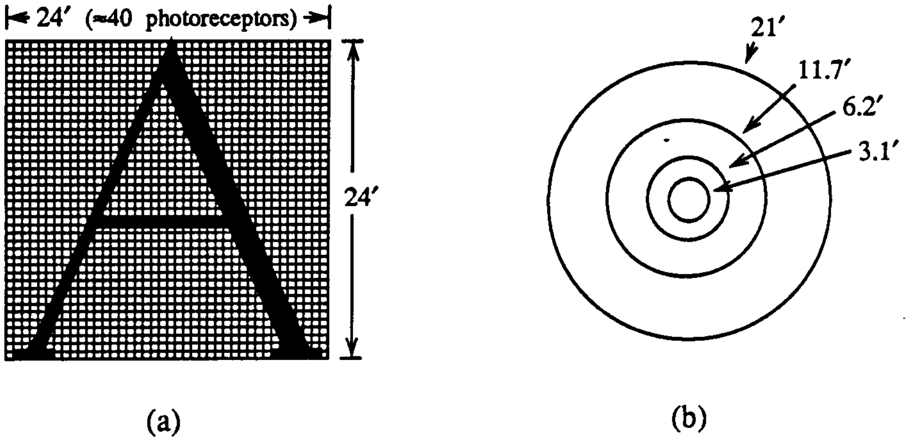

Figure 7: Edge detector resolution at the fovea. (a) The letter A (12 pt. Times font) as it would appear at the fovea when read at a distance of 18 inches from the eye. Each pixel represents a single photoreceptor (on average). (b) The sizes, to scale, of the central parts of the four receptive fields predicted by Wilson and Bergen (1979). (Actually, neuroanatomical evidence points to the existence of a smaller receptive field center measuring approximately 1.3′.)

ferent sizes at any point in the visual field (see Fig. 7) and that each receptive field has a profile similar to the LOG filter. (Wilson and Bergen used the difference of two Gaussian functions, which happens to match the LOG almost perfectly.)

In a similar vein to Marr and Hildreth, Canny (1986) has derived a set of optimal filters for detecting various sorts of intensity changes. For detecting step-edges, Canny has defined a 2D operator based on the gradient of a 2D Gaussian. This operator turns out to be an oriented edge-detector, as opposed to Marr and Hildreth's omnidirectional edge detector.

Nalwa and Binford (1986) have developed a much more complicated method of edge detection based upon fitting a series of surfaces (e.g., cubic spline, tanh) to a window. They are able to obtain better accuracy with this method than with the Marr-Hildreth detector.

The Gabor filter. Watson (1983), Daugman (1985), and Caelli et al. (1987), among others, have used the 2D Gabor function to model the early stages of human visual processing. Basically, the 2D Gabor function is formed from the product of a 2D sine and Gaussian. By changing the phase, frequency, and direction of the sine, as well as the aspect ratio of the Gaussian, a family of 2D Gabor functions can be generated. A few of these are shown in Figure 8. One can readily see how these functions are well suited to serve as image filters for detecting small, linear edges or lines in an image. Daugman (1985) has confirmed this fact, showing that the 2D Gabor function is actually the optimal filter for simultaneously detecting the position, angular orientation, and spatial frequency (or more appropriately, scale) of edges or lines with maximum certainty.

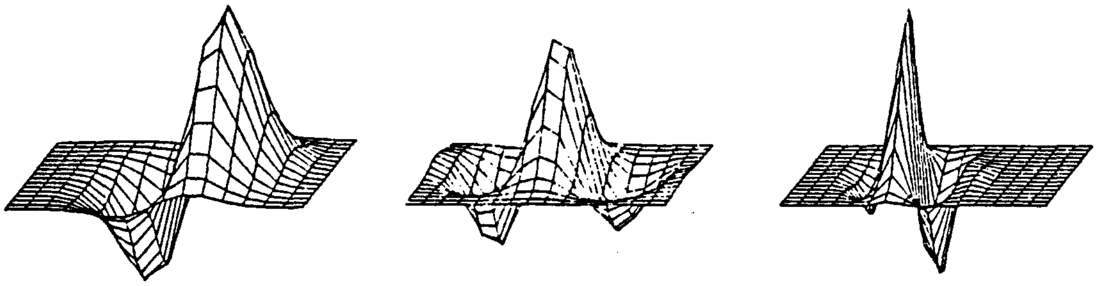

Figure 8: Some 2D Gabor filters, with various settings of phase, spatial frequency, and aspect ratio. Although all of the filters shown here are oriented in the same direction, the orientation selectivity is easily changed by a simple rotation in the plane. (From Daugman, 1985)

Cooperative processes. Edges generally do not occur in isolation. Rather, they usually form part of a global line or boundary. *Cooperative processes* are a way of incorporating such assumptions, much as humans do, in order to aid the edge detection process.

Zucker et al. (1977) have shown how one type of cooperative process called *relaxation labeling* can be used to enhance lines and curves in an image. According this method, a series of filters designed to detect small lines at several angular orientations are convolved with an image in order to generate an array of line-orientation labels. This array is then considered a partially connected graph, with each node connected to its neighbors. Each node in the graph corresponds a location in the image, and each node has an line-orientation label assigned to it. In the relaxation process, each node updates its label to be more compatible with its neighbors, as determined by a set of compatibility weights between line labels. The compatibility weights are chosen such that lines of similar orientation support one another, while lines of perpendicular orientation antagonize one another. "No-line" labels are supported positively by surrounding "no-line" labels and negatively by line labels oriented toward them. The relaxation labeling process converges in only a few iterations so that global lines or curves are enhanced and noisy elements are suppressed. Hummel and Zucker (1983) have refined the relaxation-labeling algorithm and have given conditions and proof of convergence.

Grossberg and Mingolla (1987) have proposed a scheme somewhat similar to Zucker's, called the *boundary contour system.* They use this theory to explain how edges are filled in where part of a boundary is missing, or how an illusory contour is formed from appropriately positioned line-terminations.

Kass et al. (1987) have introduced the concept of "snakes" for finding contours in natural images. A snake is an energy-minimizing spline whose shape is determined by constraint forces. Internal forces act to regularize the spline so that it remains smooth, and image forces pull the snake toward lines and edges in the image. External forces can be used interactively (such as with a mouse) to nudge a snake toward certain image features. This method works remarkably well for latching onto smooth, continuous contours, as well as subjective contours.

Lee and Pavlidis (1987) also use a cooperative process for relaxing a spline along a contour, but their method also allows for important discontinuities such as corners. In this way, corners and vertices are not smoothed over, but are allowed as valid (and important) features. (See also Terzopoulos, 1986.)

Other types of cooperative processes have been proposed by Walters (1987) (see section 3.2), and Canning (1987).

Self-organizing processes. Several researchers have used *self-organizing neural networks* to model the development of simple cells, or edge or bar detectors, in the visual cortex.

Von der Malsburg (1973) devised a model using *Hebbian learning* and on-center/off-surround interactions to show how the orientation selectivity of simple cells in the visual cortex could be developed through experience, rather than being predetermined genetically. In this model, a retina of 19 units is stimulated with lines at 9 different angular orientations. A "cortex" consisting of 338 units is connected to the retina such that each retinal cell excites all the cortical cells through a set of weights. These weights are then modified according to a Hebbian-type rule over repeated presentations of the retinal stimuli. Within the cortex, there is an on-center/off-surround interaction such that the firing of one cell helps to excite its neighbors but inhibits its more distant neighbors. After 100 trials, the cortex exhibits the same type of orientation selectivity found in the visual cortex of mammals, as shown in Figure 9.

Barrow (1987) has taken the work of Von der Malsburg one step further using a slightly different scheme. In Barrow's model, an entire image is used as the test stimulus. Each cell in the "cortex" is connected to a receptive field in the image in a topographic manner, and the weights from retina to cortex are modified according to a *competitive learning rule* devised by Rumelhart and Zipser (1985). After many trials, the cells of the cortex eventually develop to be oriented bar or edge detectors.

An experiment by Wiesel (1982) would seem to lend support to the models of Von der Malsburg and Barrow. In this experiment, a young monkey was exposed exclusively to vertical stimuli for 57 hours. It was then found that cortical cells were much more responsive to vertical stimuli than horizontal stimuli, suggesting that some competition among neurons takes place during early development.

A completely different and enlightening approach to self-organization has been proposed by Linsker (1986). Linsker's model demonstrates that experience is not necessary to develop the edge-detection function of either LGN-type cells or simple cells. In this model, random noise is used as the input to a multilayer network, the architecture of which is shown in Figure 10. The weights of the network are modified according to a Hebbian rule, and after many trials, spatial opponent cells (center/surround type) develop in the third layer. Orientation-selective cells begin to emerge in the seventh layer; and if lateral interactions are allowed at the seventh layer, then a columnar-type organization such as in V1 (primary visual cortex) occurs. Linsker (1988) has shown that the Hebbian learning rule acts to maximize the variance in a layer's response to input patterns, and that each layer in the network actually preserves maximum information about its input from the previous layer.

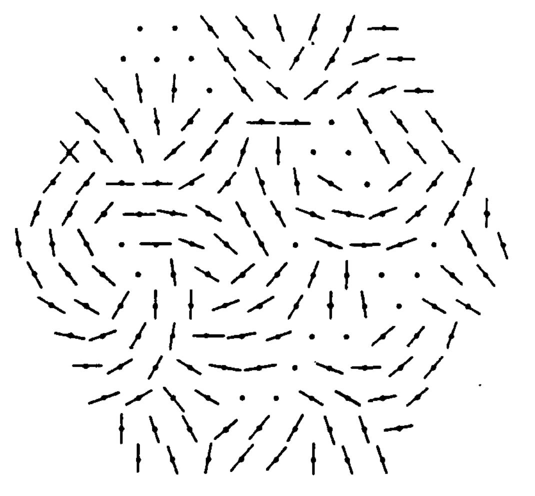

Figure 9: Von der Malsburg's "cortex" after 100 learning trials. The small lines denote the orientation selectivity of neurons in the cortex. (Copied from von der Malsburg, 1973, with permission.)

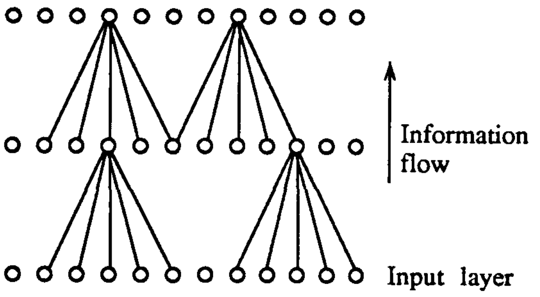

Figure 10: Linsker's network architecture. Each unit within a layer receives inputs from a local area of units (receptive field) in the previous layer (only a few connections are drawn here.

Sanger (1988) has also shown that a Hebbian learning rule can act to maximize information preservation. In particular, he has shown that a single-layer network operating with a Hebb-type rule will learn to compute the Karhunen-Loeve transform (i.e., it will compute the eigenvectors of the input auto-correlation matrix). When trained on 8x8 sub-images from natural scenery, the output units tend to develop center/surround and Gabor-like receptive fields.

Other self-organizing approaches to feature detection have been discussed by Kohonen (1982, 1986) and Grossberg (1976).

3.2. Extracting more complicated shapes

It is not hard to imagine shape features more complicated than edges or lines that would be worthwhile to extract from the image. For instance, Biederman (1986) has demonstrated that such parts as corners and vertices play a critical role in object recognition, as shown in Figure 11. How might such shape features be extracted or inferred from the image?

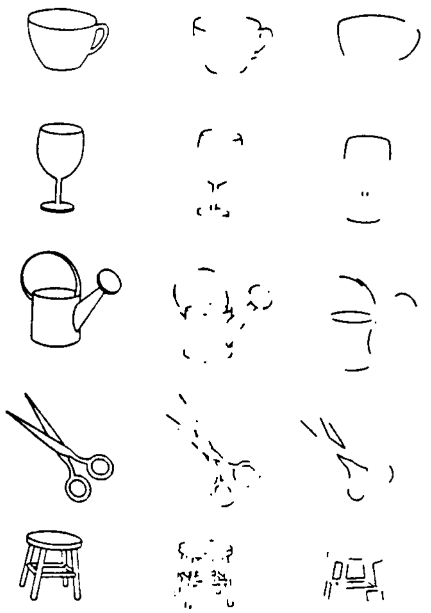

Figure 11: A demonstration of the importance of vertices (or regions of concavity) in recognizing visual objects. Objects in the right column are not recognizable (without first seeing the objects in the left column) because the contour has been deleted or altered at regions of concavity. Objects in the middle column are recognizable because the contour is deleted only at regions of smooth curvature or straight lines. (From Biederman, 1986.)

Curvature detection. Much attention has been paid to the question of how, or if, we detect curvature in an image. Lettvin (1959) has reported finding "net convexity detectors," or neurons that selectively respond to angles of a certain size, in the frog's retina; and Hubel and Wiesel (1965) have reported that hypercomplex and so-called "higher-order hypercomplex cells" appear to respond selectively to line-terminations and corners. Blakemore and Over (1974), Timney and Macdonald (1978), and Wilson (1985) have carried out perceptual experiments to determine how well we detect curvature, but remain inconclusive as to whether we have specific mechanisms for detecting curvature.

Recently, Koenderink and Richards (1988), Koenderink and van Doorn (1987), and Dobbins et al. (1987) have pointed out that endstopped neurons, previously thought to be devoted to detecting line-terminations, may also be used to calculate curvature. Figure 12 shows some 2D operators proposed by Koenderink and Richards for detecting curvature. These operators have receptive fields similar to those of endstopped neurons in the visual cortex. Furthermore, the derivation of these operators can be related to other 1D methods of finding curvature.

Dobbins et al. (1987) hypothesize that an endstopped neuron receives its input from two simple cells: one provides excitory input and has a small receptive field, and the other

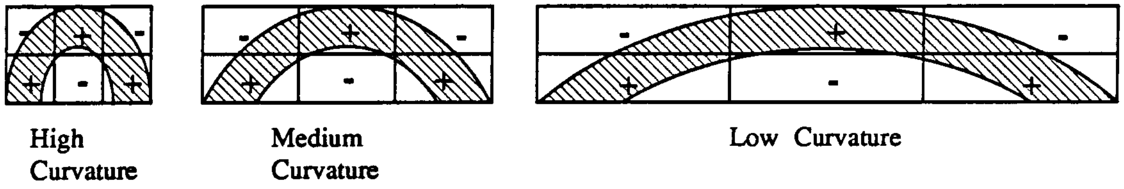

Figure 12: Some 2D curvature operators proposed by Koenderink and Richards (1988). + signs indicate excitory areas; - signs indicate inhibitory areas. These operators have receptive fields similar to endstopped neurons in the visual cortex, but the various aspect ratios allow for selectivity to different radii of curvature.

provides inhibitory input and has a large receptive field. Such an arrangement would provide a measure of curvature, as the radii of curvature in a curve would determine what parts of the excitory or inhibitory fields are activated, and hence the net response of the endstopped neuron. Predictions of this model agree well with the actual response of cortical neurons (in the cat) to semi-circular arcs spanning a wide range of radii.

Schwartz (1980) has suggested that neurons in the inferotemporal cortex (IT) may detect boundary curvature by receiving excitory input from a "line" of neurons in the striate cortex. That is, since orientation preference changes gradually across the surface of striate cortex, a line of active cells across the surface would indicate the presence of a contour of some particular curvature on the retina. Thus, cells in IT that detect oriented lines on the striate cortex would be acting as curvature detectors with respect to the retina.

Hartmann (1985, 1987) has proposed that the visual cortex may employ curve detectors as part of a hierarchical scheme for encoding continuous contours. It is hypothesized that neurons at the lowest level of the hierarchy have receptive fields that are tuned to detect a specific contour element. These elements are then pieced together by neurons at higher levels in order to form a unique representation for an entire contour.

The Neocognitron. Fukushima's Neocognitron (Fukushima, 1980) uses a self-organizing process to develop detectors for more complicated shapes, such as those illustrated in Figure 13. (See section 4.1 for details on the Neocognitron.)

Cooperative Processes. Parent and Zucker (1985) describe a method for inferring the trace of a curve (i.e., the points through which a curve passes) based on a relaxation labeling process for refining tangent and curvature estimates. According to this method, estimated tangents are initially obtained by convolving an image with several oriented line-detectors. The estimated tangents are then constrained by a "co-circularity" relationship between neighboring tangents (two tangents are said to be co-circular if a single circle passes through both tangents). Curvature estimates, obtained from multiple tangents in a local area, are constrained by a consistency relationship among neighbors. The result of the relaxation labeling process is a good estimation of the tangents and local curvatures along a curve, and hence a good recovery of the trace of a curve. This method has been tested with success on both artificial and natural images. Link and Zucker (1988) have performed some perceptual experiments that suggest that such local interactions among tangent and curvature estimates are necessary for detecting corners

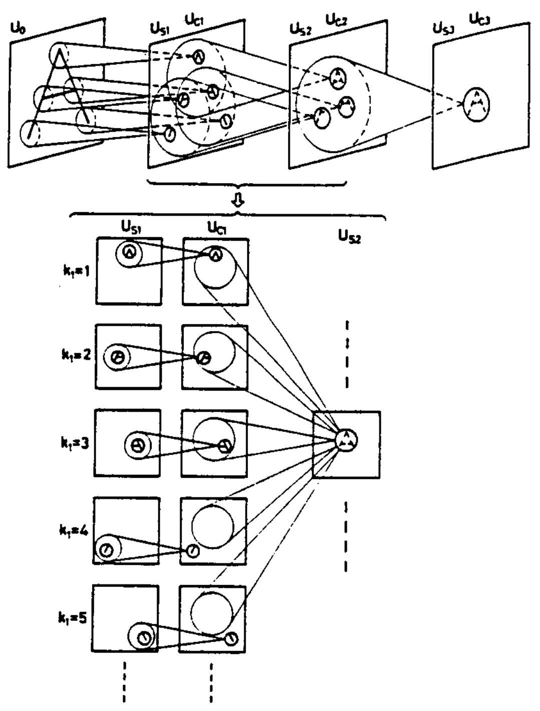

Figure 13: The feature detectors in Fukushima's Neocognitron (see section 4.1 for details). (Copied from Fukushima, 1980, with permission.)

in dotted line drawings.

Walters (1987, 1988a) has done a number of psychological studies to determine what features in an image are perceptually significant; and she has developed a cooperative algorithm for enhancing such features in an image. For example, it is shown that line-terminations become perceptually more significant when placed in certain proximal relationships with other line-terminations. Such features can then be enhanced by interactions among edge-elements in the ρ-*space representation.* Basically, the ρ-space representation is a three-dimensional discretized space, with two dimensions representing spatial position in the image and one dimension representing angular orientation of edges; relationships among neighboring contour elements within this space can act to enhance or suppress certain parts of a contour.

"Non-accidental" patterns. Vistnes (1987) and Lowe and Binford (1982) have discussed methods for detecting "non-accidental" patterns in an image. Basically, non-accidental patterns are those that are more likely to have arisen from underlying physical relationships between constituent features (say, due to the boundary of an object), rather than from some coincidence of viewpoint or location. Vistnes has shown how the principle of non-accidentalness can be applied to the problem of detecting dotted lines and curves amidst a random-dot background.

3.3 Texture

Visual textures are defined as aggregates of many small elements, such as simple spatial patterns or dots of certain colors. The visual world is full of different textures, and indeed, it appears important that all living things be able to discriminate textures: grass, fur, foliage, and water surfaces are just a few. Unlike most visual patterns, textures are not characterized by any one global shape, but rather by some statistical property of the many fine elements that compose them.

Julesz and Bergen (1983) have proposed that so-called *textons* serve as the fundamental elements, or primitives, of texture perception. Textons have specific properties, such as color, angular orientation, width, and length, that enable them to be immediately detected among a group of other textons, such as illustrated in Figure 14. Julesz and Bergen have been able to define quite concisely the properties of textons, but they have not ventured to say how textons

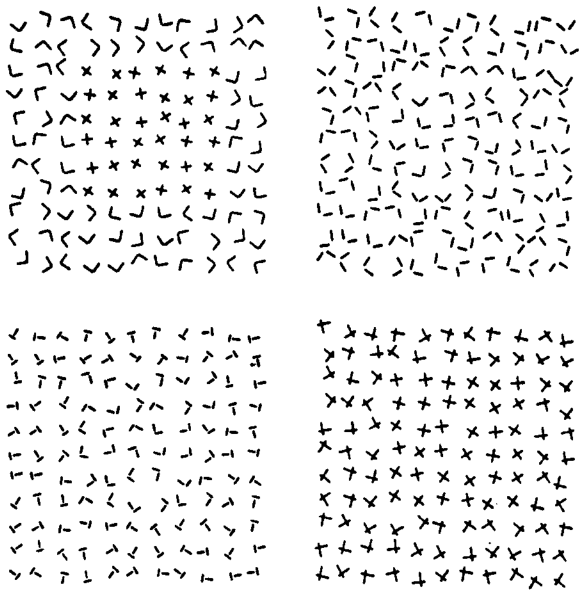

Figure 14: Demonstration that crossing of line segments is a texton. A region of crossing line segments can be immediately differentiated from a background of non-crossing line segments (upper left), but other combinations require much longer serial search. (Copied from Julesz and Bergen, 1983, with permission; Copyright © AT&T Bell Labs.)

may actually be detected in an image.

Voorhees and Poggio (1987) have expanded on the work of Julesz and Bergen by proposing a method for texton detection. They use a Gaussian filter to estimate the amount of background noise in an image, which is then used to determine the threshold for an edge-detection operator (Marr-Hildreth LOG filter). The result is that edges are found along texture boundaries instead of intensity boundaries.

Schwartz (1980b) has proposed a method for texture discrimination in the visual cortex. He has shown that the representation of certain textures in the striate cortex is sufficiently distinct to allow them to be easily differentiated by some higher-level cell.

Other methods of dealing with texture are presented in Lowe and Binford (1982), Kass and Witkin (1985), Zucker (1976, 1986), Vilnrotter et al. (1986), Mesrobian and Skrzypek (1987), and Walters (1988b).

3.4. Motion

Poggio and Koch (1985) have shown how the direction of motion of an object, or *optical flow*, can be computed by using *regularization* techniques. That is, since any one local measurement of motion in an image is incapable of fully specifying the direction of motion of a global object (since a finite aperture can measure only the velocity component perpendicular to the edge moving through it), many local measurements must be combined in order to collectively compute the true direction of motion. By assuming that objects are generally

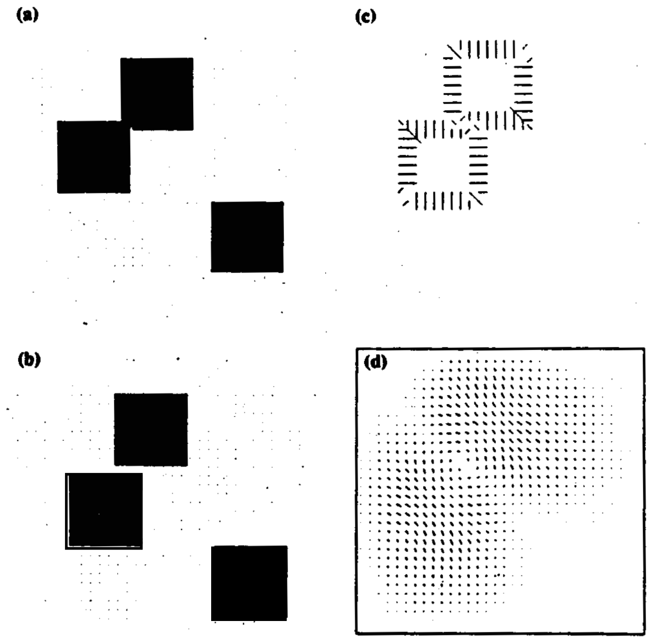

Figure 15: Computing optical flow. The two squares in the upper left of image (a) are displaced in opposite diagonal directions (b) in order to create a motion sequence. The initial computed velocities, based upon locally measured edge movements, are shown in (c) (note that they are perpendicular to the edges). The optical flow field (d) is obtained by applying regularization techniques. (Copied from Hutchinson et al., 1988, with permission.)

smooth and also undergoing smooth movements, it is possible to constrain the direction of motion to vary smoothly across an image. This constraint can be formulated as an *energy functional* that governs the many local motion vectors over an image. Variational principles can then be applied to find the overall direction of motion that minimizes the functional. The result is an optic-flow pattern, such as in Figure 15, that reveals the direction of motion of an object. Hutchinson et al. (1988) have shown how this computation can be implemented in a simple resistive network (using analog VLSI) such that the equations are solved via Kirchoff's current law.

In the visual cortex, Koch (1988) has suggested that local measurements of motion direction are represented in V1, and that MT computes optical flow.

3.5. Depth

It is well known that humans use stereo disparity information to infer depth in a scene. Given the distance between the two eyes and the *disparity*, or relative offset, between the two images produced by an object, the distance to the object can be computed. Julesz (1971) has shown that humans are capable of fusing random-dot stereograms to give the impression of depth, suggesting that the computation of disparity can be based on local comparisons among pixels rather than global comparisons of shape. How, then, are the images being compared at such a fine-grain level to compute disparity?

Marr and Poggio (1976) devised a highly parallel algorithm to compute depth from stereo image pairs. Two constraints are used in the computation: 1) Each pixel may be assigned only one disparity value, and 2) disparity values should vary smoothly almost everywhere, since objects generally have smooth surfaces. After many iterations over all the pixels, a depth map is computed. Marr himself has since criticized this algorithm because the number of iterations required for a solution would make it biologically implausible. He has suggested a second algorithm (Marr, 1982) in which images are first matched at a coarse resolution and then progressively at finer resolutions, thereby reducing the number of iterations required.

In the visual cortex, Schwartz and Yeshurun (1987) have demonstrated that the columnar interlacing of two slightly different images, as provided by the ocular-dominance column system, provides a simple means for extracting disparity.

3.6. Other features

Many other features can be extracted from an image besides those discussed above. Several methods have been proposed for inferring 3-D shape from the image, including shape from stereo, shape from shading, shape from motion, and shape from contour. (Such techniques are generally referred to as "shape from X.") These subjects are covered extensively by Brady (1982) and Horn (1986). Marr (1982) has proposed the *2^1/2-D sketch* as a way of organizing and representing such 3-D information as it is extracted from the 2-D image, as shown in Figure 16.

Recently, Lehky and Sejnowski (1988) applied a neural network to the shape from shading problem. Using the backpropagation algorithm, they were able to train the network

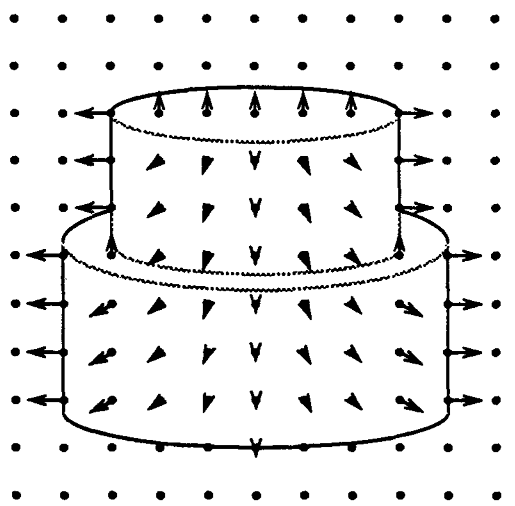

Figure 16: An example of a $2^1/_2$-D sketch. The arrows symbolically represent the 3D orientation of surfaces in the image (a full $2^1/_2$-D sketch would include rough distances to the surfaces as well). Dotted contours show where surface orientations change sharply; solid contours show where depth is discontinuous. The key idea here is that three-dimensional information is indexed relative to the image (i.e., in a viewer centered coordinate frame). Marr hypothesizes that this information is stored relative to an internal reference frame at a later stage. (From Marr, 1982)

to compute the magnitude and orientation of the two principle surface curvatures at the center of an input surface. After 40,000 learning trials, the hidden units in the network happened to develop receptive field profiles much like those of simple cells in the visual cortex. Moreover, the arrangement of weights in the "projective field" of a hidden unit (i.e., the weights from a hidden unit to an output unit) seem to provide information about surface orientation, convexity/concavity, and relative magnitudes of curvature.

3.7. Pyramidal techniques

One of the great difficulties with extracting any kind of features from natural imagery is that they tend to occur at a variety of scales or resolutions. *Pyramidal techniques* offer a method for dealing with this problem by representing an image, or image features, at various resolutions (Fig. 17). At each level in the pyramid the resolution is band-limited, thereby simplifying analysis of the image. (See also section 3.1, The Marr-Hildreth edge detector.)

Witkin (1983) has devised the method of *scale-space filtering* for describing signals over a range of different resolutions. As illustrated in Figure 18, a signal is first filtered with a Gaussian mask at several different widths in order to remove progressive amounts of detail. Then, a "scale-space image" is formed by laying-out the progressively fine-to-coarse filtered signals side by side. In this way, points of maximum change in the signal (zero-crossings of the second derivative) can easily be identified at coarse scales and then traced to finer scales for localization. This provides a convenient way of determining whether changes at the finest scale are due to noise or more global processes. Witkin et al. (1987) have shown the applicability of the scale-space image for matching signals, and Mokhtarian

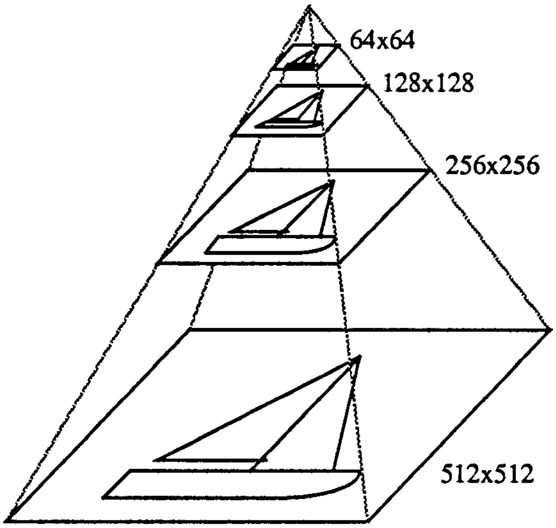

Figure 17: An image *pyramid.* Each level of the pyramid shows progressively less detail than the one below it (the numbers beside each level denote the image resolution). Representing an image in such a way can be of great use to the feature detection process, and also to recognition (see section 4.1).

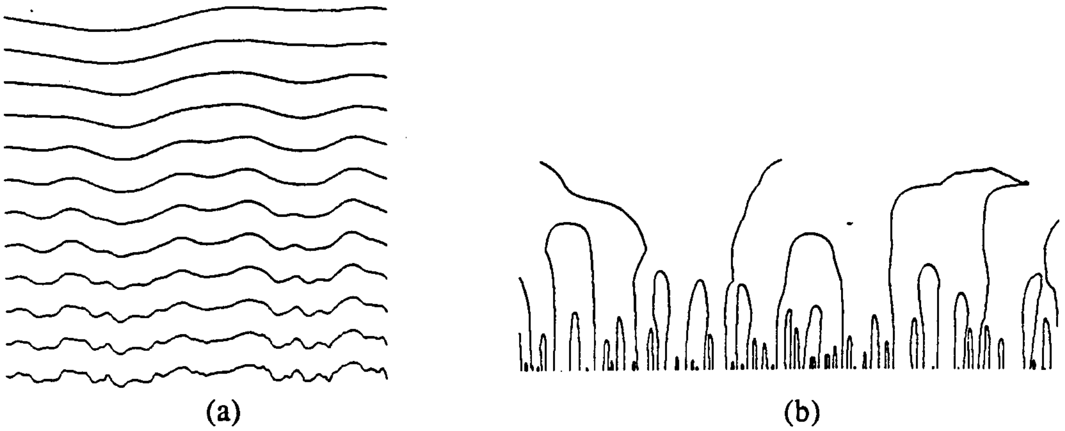

Figure 18: *Scale-space filtering.* (a) A sequence of Gaussian smoothings of a signal $f(x)$, achieved by convolving a Gaussian function with $f(x)$ at various scales. σ, the spread of the Gaussian filter, increases from bottom to top. Each smoothed signal is a constant-σ profile from the scale-space image, which has x and σ as its two spatial dimensions. (b) The contours formed from the zero-crossings of $f''(x)$ in the scale-space image. Again, x is the horizontal dimension, and σ increases from bottom to top. With this representation, points of maximum change in $f(x)$ can be reliably followed from coarse to fine resolutions. If the original signal were two-dimensional (i.e., an image), then the scale-space image would occupy three dimensions (two for spatial position, one for σ). (Copied from Witkin, 1983, with permission.)

and Mackworth (1986) have used the scale-space image to match shapes in an image.

Zucker and Parent (1984) have used a type of pyramidal technique to augment a relaxation labeling process for enhancing lines and curves in an image. In this scheme, edges detected by large-scale operators provide contextual constraint for edges detected by smaller scale operators.

Watson (1987) has used a type of pyramidal technique in modeling the function of simple cells in the striate cortex. By filtering an image at various spatial bandwidths and orientation selectivities, several image pyramids are formed. Images within a pyramid vary according to scale, and orientation varies from one pyramid to the next. This new representation of the image contains just as many pixels as the original image, and the transformation is invertible.

3.8. Segmentation

Segmentation is the process of separating figure from ground, or determining what belongs to the object and what belongs to the background. This process may not necessarily involve recognition; the goal here is to delineate an object within a surrounding field, not to identify it.

Much of the segmentation process has already been discussed in previous sections. For example, some of the cooperative techniques described for detecting lines and curves are essential to delineate the border of an object. Also, segmentation may be readily achieved from textons (Fig. 14) and optical flow (Fig. 15). Discussed below are some methods specifically intended to assemble parts of shape that belong to the same object.

Sejnowski and Hinton (1987) have demonstrated how a neural network may be used to separate figure from ground. In this scheme, each edge extracted from an image is considered to be part of a figure/ground boundary, with one side pointing toward figure and the other toward ground. Then, by interacting with their neighbors lying along the same line, the edges try to find a consistent state so they agree on where the figure is and where the ground is. This method has been shown to successfully segment simple areas, such as rectangles.

Walters (1987) discusses how interactions within the ρ-space representation can be used to segment a boundary contour into sets that have a high probability of depicting a single object.

Other methods of figure/ground separation are discussed in Horn (1986), Ballard and Brown (1982), Rosenfeld (1986), and Weisstein (1986).

4. Shape Representation and Recognition

Assuming that a set of features has been extracted from the image, or that an object has been segmented, how may the object then be recognized? How should shapes be represented, and how is the matching accomplished? This section discusses some theories and methods that address these problems.

4.1. Form invariance

One of the more challenging visual recognition problems in any realistic situation is form invariance. Since the representation of an object on the retina, or image plane, depends critically on the vantage point, there must be some means of re-representing or transforming an object such that its stored representation is independent of the viewing perspective. Otherwise an infinite multitude of object representations would have to be stored.

Presented below are several theories and methods for dealing with the various transformations that can affect an object's representation on the image.

The Hough transform. Ballard (1981) has used the *Hough transform* to detect analytical shapes (e.g., lines and circles) and arbitrary, non-analytical shapes in an image. Figure 19 illustrates a simple application of the Hough transform for detecting lines. Further work by Ballard and others has shown how the Hough transform can be implemented in a neural-network for recognizing both 2-D and 3-D shapes independent of viewing perspective (Ballard and Sabbah, 1983; Hrechanyk and Ballard, 1982; Sabbah, 1982; Ballard, 1984; Ballard and Tanaka, 1985; Ballard, 1986).

One particularly interesting implementation of a Hough transform (actually, a Hough-

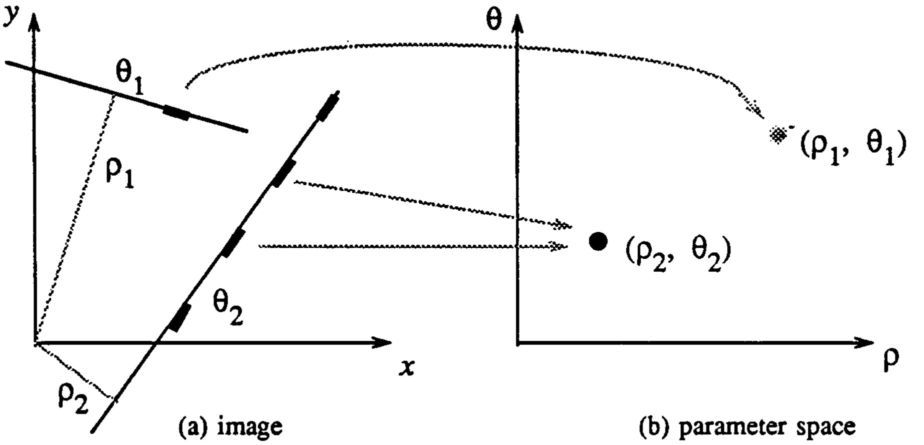

Figure 19: The Hough transform for finding lines. Small line segments in the image (a) are mapped many-to-one to points in parameter space (b). That is, a line segment in the image that falls along a line with slant θ to the x-axis and perpendicular distance ρ from the origin contributes to a sum in parameter-space at coordinates (ρ,θ). Thus, the presence of a line in the image would be indicated by a peak in parameter space.

type transform) in a neural network has been done by Hinton (1981a,b, 1985). As shown in Figure 20, shape features are defined in a model frame (object-based units), and those shape features that comprise an object provide activation to a grandmother node that represents the object. Consequently, an activated grandmother node provides top-down support to those shape features of which it is composed. Shape features, transformation units, and retinal feature units (retina-based units) have a three-way interaction: a retinal feature unit activates a model-frame feature unit gated by a transformation unit, and a transformation unit is activated by simultaneous activity from a retinal feature unit and a model-frame feature unit. The general idea here is that the network eventually settles into a stable state that simultaneously specifies the identification of the object and its transformation from the image to the model frame.

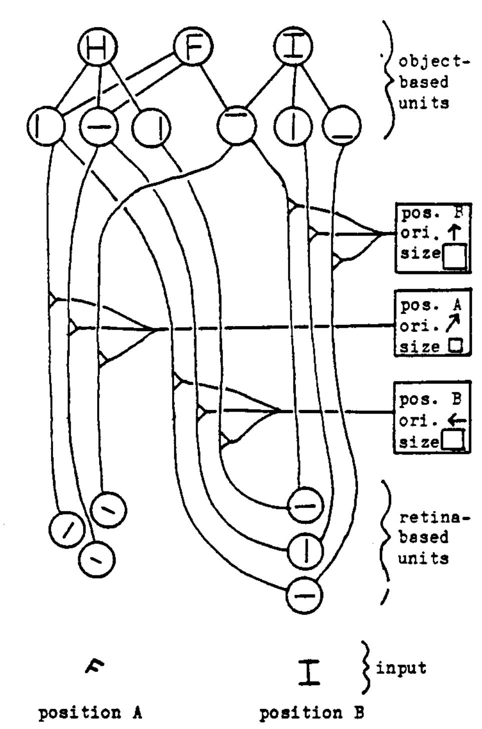

Figure 20: Hinton's implementation of a Hough-type transform in a neural network (see text). (Pos. = Position, Ori. = Orientation.) (Copied from Hinton, 1981, with permission.)

Prazdny (1987a,b) and Tucker (1988) have used a similar technique for recognizing 2D shapes. In their schemes, corners or vertices are extracted from the image and then matched in parallel to all possible model instances through a set of transformation parameters. Each match to a model instance generates a "hypothesis," or a vote for a particular model and a set of viewing-transform parameters. Each hypothesis then projects its model instance back onto the image for verification, and the hypothesis with the highest confidence and the strongest verification yields the identification and transformation of the object. Tucker has programmed this scheme on the Connection Machine, and Prazdny has devised a neural-network implementation.

Fourier/log-polar transform. The Fourier transform, used in combination with the log-polar transform, can be utilized to achieve a representation of an image invariant to shift, scale, and rotation in 2-D. This method is illustrated in Figure 21. First, the 2-D Fourier transform of an image $f(x,y)$ is computed in order to achieve translation invariance. The resulting power spectrum $||F(u,v)||^2$ is then converted to log-polar coordinates and re-represented as $F_{lp}(\log \rho, \theta)$. Thus, rotations of the image will produce shifts on the θ-axis, and expansions or contractions in the image will produce shifts on the (log ρ)-axis (since $\log a{\cdot}\rho = \log a + \log \rho$). These variations can then be eliminated by doing another 2-D Fourier transform on the (log ρ,θ)-plane. Casasent and Psaltis (1976) have applied this technique to the optical correlation of images, except that they used the Fourier-Mellin transform (Altes, 1978) to combine the log and Fourier-transform computations on the ρ-axis. Brousil (1967) used a similar technique for recognition with a single-layer neural network. Carpen-

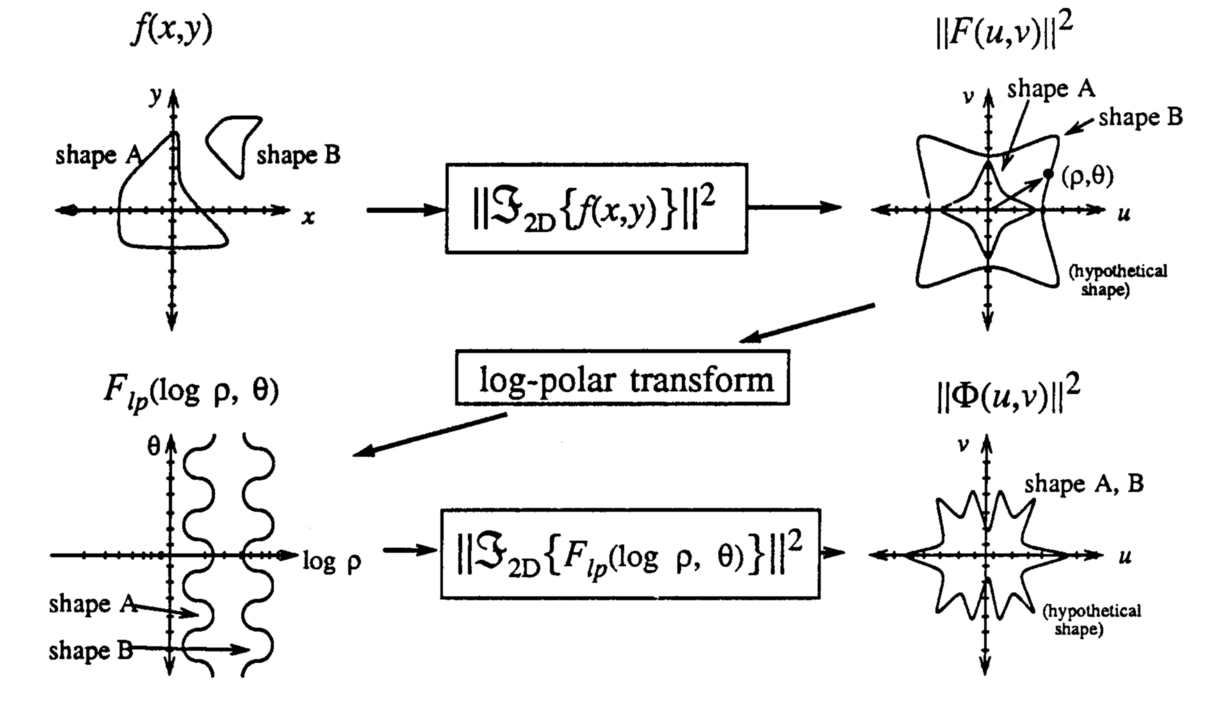

Figure 21: The Fourier/log-polar transform. The shapes *A* and *B* are identical, except that shape *B* is a translated, rotated, and scaled version of shape *A* in the image $f(x,y)$. The power spectrum of the Fourier transform $||F(u,v)||^2$ removes translation; the log-polar transform converts differences in rotation and scale into simple translations in F_{lp}; and a final Fourier transform removes the translations in F_{lp} so that the two shapes *A* and *B* have identical representations.

ter and Grossberg (1987b) and Wechsler and Zimmerman (1988) have used the log-polar transform as the front end to an associative memory.

The Fourier/log-polar transform has also been used in modeling the visual cortex. Schwartz (1980a, 1981, 1985) has proposed that the retinal image is converted to log-polar coordinates on the striate cortex as a by-product of the distorted topographic mapping in that area. Thus, rotations and scale changes on the retina would appear as simple shifts on the surface of the cortex. Cavanagh (1978, 1985) has proposed that a global Fourier/log-polar transform of the retina is formed in inferotemporal cortex from many piecewise Fourier/log-polar transforms in striate cortex. Baron (1987) and Pollen (1971) propose still other schemes.

Fourier descriptors. Zahn and Roskies (1972) invented the technique of using *Fourier descriptors* to form an invariant description of an arbitrary plane closed curve. In this method, a closed curve is represented parametrically as a function of arc length by the accumulated change in direction of the curve along the perimeter. Then, the Fourier coefficients of this function can be used to uniquely describe the curve invariant to changes in rotation, translation, or scale (the perimeter is normalized to 2π).

Schwartz et al. (1983) have suggested that Fourier descriptors may be used to encode shape information in the inferotemporal cortex. In an experiment on inferotemporal neurons in the macaque monkey, it was found that many neurons (54% of 234 visually

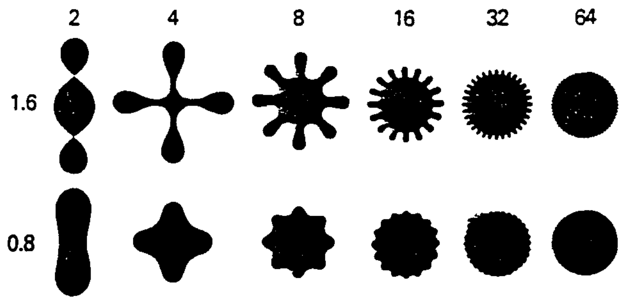

Figure 22: Some examples of Fourier descriptor stimuli, varying in frequency across rows and amplitude down columns. (From Schwartz et al., 1983.)

responsive units) were selective to the frequency of Fourier descriptor stimuli, mostly independent of size and position (Fig. 22). These results suggest that inferotemporal cortex may code for boundary curvature, much as striate cortex codes for local edge orientation.

Representing all transformations. Several methods for achieving form invariance are based on re-representing the visual input at a multitude of possible transformations.

Pitts and McCulloch (1947) have proposed that the brain accounts for size changes by computing magnified and reduced versions of the retinal image, and then performs a sort of averaging operation over all these patterns in order to create an invariant representation.

Crettez and Tanimoto (1985) have proposed using a pyramidal-type scheme for re-representing visual patterns at several different resolutions, and hence several different sizes. As shown in Figure 23, this model assumes that neurons in the visual cortex are organized into layers according to receptive field size. Thus, changes in size on the retina would result in a somewhat invariant pattern simply being shifted up or down the layers. One of the layers would contain a representation of the visual input that matches the size of its internal, stored representation.

Trehub (1987) has proposed a neural-network model that accounts for size and rotation invariance by re-representing the input at various sizes and rotations, and then selecting one that matches.

Copying weights *en masse* within a neural network. Fukushima's Neocognitron (Fukushima, 1980) achieves translation invariance by copying receptive-field weights *en masse* over a "cell plane." As shown in Figure 24, the Neocognitron is composed of several alternating S-layers and C-layers. Each layer is composed of several cell planes, and each cell within a plane receives its input from a receptive field in the previous layer. Upon presentation of a stimulus, the responses of all the S-cells within a layer are compared, and the cell with the highest response adjusts its weights to match the stimulus in the receptive field. This new set of weights is then copied en masse to all the other cells in that cell plane. The C-plane behind each S-plane gathers its input (with fixed weights) from a receptive field in the previous S-plane, so that as patterns are shifted around on the input plane, the responses of C-cells remain somewhat constant. At the output, a unique (grandmother node) representation is formed for each pattern, regardless of translation. The output also exhibits some degree of invariance to changes in size and small distortions in the shape of a

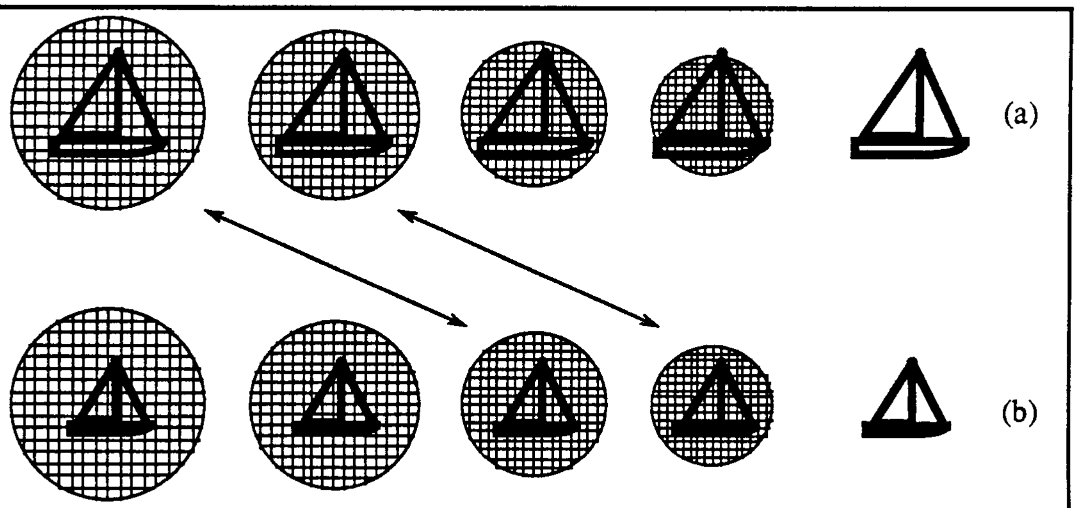

Figure 23: Demonstration of how a pattern shifts through layers as it changes in size, according to the model of Crettez and Tanimoto (1985). Each pixel represents a neuron in the visual cortex with a corresponding receptive field on the retina. Within a layer (i.e., a circle), each neuron has a receptive field of the same size, and the retina is sampled at evenly spaced locations. As the resolution decreases from one layer to the next (i.e., circle to circle from right to left), the receptive fields become larger and the retina is sampled more sparsely. Thus, if one boat is viewed up close (a) and the other at a distance (b), each would still be represented at a variety of scales or resolutions in the visual cortex. Double-arrow lines are drawn between representations of the object at the same scale that could easily be matched to one another on a pixel to pixel basis.

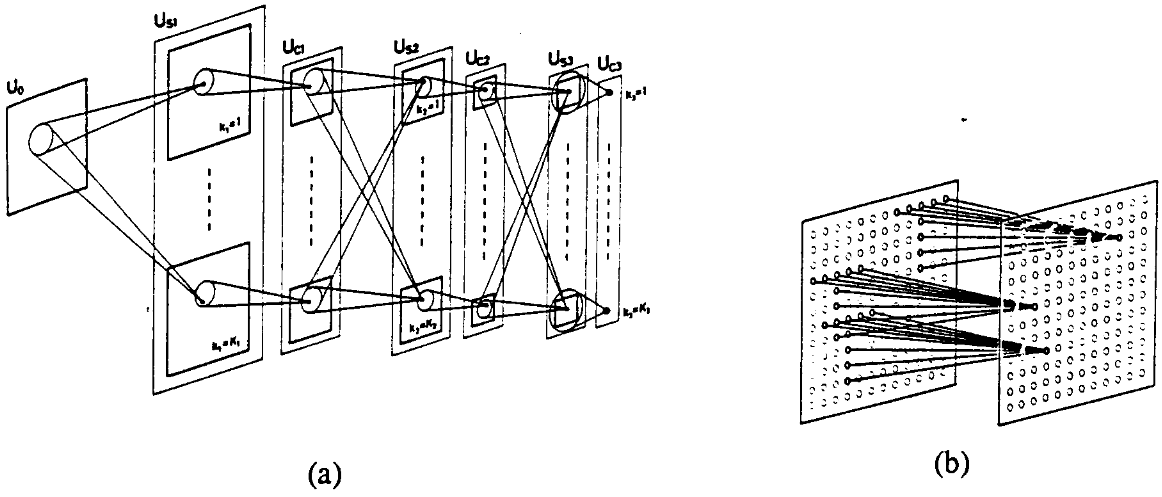

Figure 24: Fukushima's Neocognitron. (a) Interconnections between layers. (b) The learned input interconnections to the cells within a single cell-plane. (Copied from Fukushima, 1980, with permission.)

character.

Rumelhart et al. (1985) have shown how the back-propagation learning algorithm can be applied to the problem of distinguishing a "T" pattern from a "C" pattern in all translations and rotations. In this scheme, a single output unit gathers its input from a layer of hidden units, and each hidden unit gathers its input from a 3x3 receptive field in the input. All the hidden units are constrained to learn the exact same set of weights, so that the whole field of hidden units consists of replications of a single feature detector centered on different regions of the input. This network has arrived at a variety of solutions for the feature detector.

Widrow (1987) has shown how a multilayer network of ADALINE's (ADAptive LInear NEurons) can be constructed as an "invariance net," such that 2-D patterns are transformed into patterns invariant to translation, rotation, and size. The network is composed of a number of "slabs" of neurons, and within each slab, weights are copied (shifted) *en masse* to achieve invariance to up-down and left-right translation. A translation-rotation-scale-invariant net could then be assembled by copying slabs with rotated or scaled weights and putting them together into a single network. This method is currently being tested in simulation experiments.

Training-in associations with a neural network. Yang and Guest (1987) have used the back-propagation algorithm to train a two-layer neural network to recognize 2-D shapes invariant to rotation. They presented four patterns, A, T, H, and R, on a 16x16 array at all rotations in 15° intervals. The number of hidden units was arbitrarily set at 64, and the output consisted of one grandmother node for each pattern. With some modification to the sigmoid threshold function, all four patterns could eventually be recognized at any rotation.

Reference frames. Palmer (1983) has proposed that transformations are dealt with by imposing an *intrinsic frame of reference* on an object. That is, some salient, geometric characteristic of an object's shape (e.g., elongation, symmetry, or motion) is used to define a coordinate system around the object. (Marr, 1982, has proposed a similar scheme, calling it a "natural coordinate system.") Thus, while absolute orientation, position, and size may vary over the retina, relative orientation, position, and size will remain fixed with respect to the intrinsic reference frame. In a sense, then, the intrinsic reference frame factors out the effect of transformations. Palmer suggests that recognition is accomplished by an attentional mechanism that matches the intrinsic reference frame of an object in the image to the reference frame of a stored object.

A number of psychophysical experiments support this hypothesis. For example, Hinton (1979b) has shown that when people are asked to imagine a cube in an "un-natural" reference frame, such as in Figure 25b, they have an extremely difficult time describing the shape accurately. In fact, they usually describe the shape of Figure 25c, which fits much more naturally in the frame. This would seem to indicate that our internal representation of an object is extremely dependent on a reference frame defined with respect to the object. Also, Shepard and Metzler (1971) have demonstrated that the time required to match two objects that have been rotated with respect to one another is linearly proportional to the amount of rotation. Their results seem to indicate that reference frames are being rotated (at 60 degrees per second) in order to obtain a match.

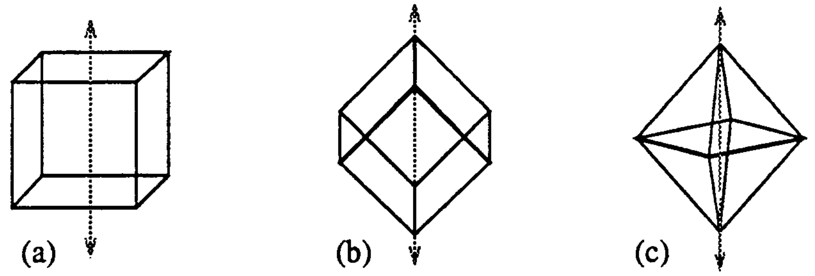

Figure 25: A demonstration of the effect of reference frames in mental imagery (see text).

4.2. Theories of Attention

Attention might be thought of as the executive module in cognition. It is the mechanism that dynamically and serially controls how the various features and parts of a scene are brought together to form an object. This section discusses various theories about the role attention plays in visual perception and cognition.

Feature integration. Treisman and Gelade (1980) have proposed that the visual scene is initially coded along a number of separable dimensions, such as color, orientation, spatial frequency, movement, size, etc., and that these features are registered early, automatically, and in parallel across the visual field. Objects are identified separately at a later stage, which requires focused attention. Treisman and Gelade suggest that attention is directed serially to each stimulus in a display whenever conjunctions of more than one separable feature are needed to characterize or distinguish the possible objects presented. The results of their experiments provide compelling evidence for such a theory, showing a dramatic rise in the time required to identify objects based on a conjunction of features. Their theory would also seem to be supported by the fact that the visual cortex contains parallel paths for processing form, color, and motion (Van Essen, 1985; Barlow, 1980).

Treisman and Schmidt (1982) have shown that when attention is overloaded or diverted, features may be wrongly combined, giving rise to *illusory conjunctions*. For example, brief presentation of a red *T* and a blue *S* may be incorrectly registered as a blue *T* and a red *S*. Hinton (1985) has duplicated such errors in a neural-network model.

Julesz and Bergen (1983) have shown that while differences in textons can be detected immediately, or preattentively, the positional relationship between neighboring textons passes unnoticed. This kind of positional information is extracted only by a time-consuming and spatially restricted process which Julesz calls "focal attention." The aperture of focal attention can be very narrow, and shifting its locus requires about 50ms.

Integrating parts of a scene, accounting for eye movements. Because the fovea covers only 1-2° of the central visual field, some attentional mechanism must be responsible for directing eye movements and integrating the various pieces of a scene captured through the fovea.

Hinton (1981c) has proposed what he calls "spatial working memory" as a means of putting together the various parts of a scene. It is based on three frames of reference: retina-frame, object-frame, and scene-frame. Various shape features and their relationship to each other are represented in the retina-frame. These features then activate a gestalt in the object-frame, which is integrated into some larger whole or scene in the scene-frame.

Other methods for dealing with eye movements are discussed in Ballard (1987b), Baron (1987), Breitmeyer (1986), and Fukushima (1986).

Matching reference frames. Palmer (1983) has suggested that an attentional mechanism functions as the "minds eye," effectively decoupling the internal representation of an object from the stimulus input. The attentional mechanism is capable of scaling, rotating, or translating the input in order to obtain a match between an object's intrinsic frame in the image and its stored, internal reference frame (see section 4.1, Reference frames). Anderson & Van Essen (1987) have proposed a neural shifter circuit that may serve such a purpose (for translation).

4.3. Shape primitives, structural descriptions

A number of researchers have proposed using shape primitives, or some basic set of "shape building blocks," for defining a wide variety of objects.

Asada and Brady (1986) have devised the "curvature primal sketch" for representing the significant changes in curvature along a two-dimensional object boundary. This method assumes that an object's boundary has already been extracted, and that the curvature along its path is represented as a 1-D signal. Then, the zero-crossings of this function are found at various resolutions, or levels of detail, and groups of curvature extrema are partitioned into primitives. An object's boundary can then be described as a composition of such primitives. (See also Fischler and Bolles, 1986, for a discussion of perceptual curve partitioning.)

Richards and Hoffman (1986) have defined a set of six *codons* for describing plane curves, as shown in Figure 26. Because of the strong constraints imposed by the bounding contour projected by 3D objects, only a small set of realistic curves may be generated from any set of codons, thus making the codon representation highly redundant (good for error-correction).

Jaeckel (1988) has proposed that characters could be encoded for a sparse distributed memory (SDM) by decomposing them into lines and arcs of circles. Then, the parameters of location, length, and angular extent (for arcs) for each piece are encoded into bit strings and

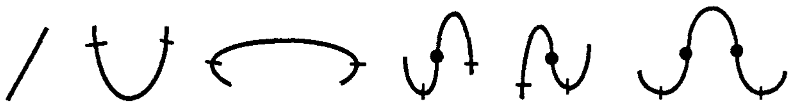

Figure 26: The six codons of Richards and Hoffman (1986). A curve is broken into parts a concave cusps (or minima of negative curvature, when traversing the curve with the figure, or the object delineated by the curve, on the left), and each part is classified as one of six codons according to the number and arrangement of zeroes and maxima of curvature within it. Here, zeros of curvature are indicated by dots, minima by slashes.

fed to an SDM for writing (training) or reading (recognition). Krishnan and Walters (1988) have proposed a similar scheme for recognizing line drawings. Their method is based on decomposing a drawing into "perceptually significant features," such as orientation of edge elements, angular separation of corners, and ratio of chord length to arc length. These features are then encoded into bit strings and fed to an associative memory.

Marr (1982) mentions the use of a hierarchical organization of volumetric primitives, such as generalized cones or cylinders, for defining objects at varying levels of detail. For example, at a coarse scale a hand would be represented as a single cylinder, but at a finer level it would be composed of one cylinder for the palm or wrist part and five cylinders for the fingers and thumb.

In the visual cortex, Schwartz et al. (1983) have proposed that the inferotemporal cortex may use Fourier descriptors to code for shape, as discussed in section 4.1. By combining Fourier descriptors of various frequency and amplitude, it would then be possible to synthesize many closed 2-D shapes with a smooth boundary.

Also, experiments by Perrett et al. (1982) on the superior temporal sulcus in the monkey suggest that the brain may use face primitives, such as the eyes, nose, and mouth, in recognizing faces.

4.4. Other theories of shape representation and recognition

Lowe (1985a,b) describes a framework for preprocessing, representation, and recognition in a vision system for recognizing 3D objects from arbitrary viewpoints. The system is based on three separate mechanisms: 1) a process for finding groupings and structures in image (such as line segments) that are likely to be invariant over a wide range of viewpoints; 2) a process for reducing the size of the search space for object matching, and 3) a method of spatial correspondence for projecting the best-fit model back onto the image for verification and refinement.

Biederman (1986) has proposed using a set of primitive 3D solids, called "geons," for describing 3D objects. He claims that if an arrangement of two or three geons can be recovered from the input, that objects can be quickly recognized even when they are occluded, rotated in depth, novel, or extensively degraded.

Ponce and Chelberg (1987) discuss the use of generalized cylinders for modeling 3D solid shapes. (A generalized cylinder is the solid obtained by sweeping the cross-section of a surface along a curve or the axis of the solid.) They present a fast algorithm for computing set operations (unions, intersections) between different types of generalized cylinders to form compound shapes.

For a good overview of various theories on visual cognition, see Pinker (1985a); Ullman (1986) discusses "visual routines" involved in recognition.

Feldman (1985) presents a general framework for using connectionist networks in visual recognition.

Koffka (1935) and Kohler (1947) are good references for a review of classical gestalt psychology theories of recognition.

Appendix: Glossary of Technical Terms

$2^1/2$-D sketch: A way of organizing and representing 3-D information as it is extracted from the 2-D image. (See Fig. 16)

ADALINE (ADAptive LInear NEuron): A device that can be trained to map a set of input patterns to a set of desired responses. Each bit in the input is multiplied by a weight and summed into the output. On each learning trial, the weights are adjusted according to the LMS rule (see Widrow, 1962) to reduce the mean square error between the desired response and the actual response for a particular pattern presentation. The mean square error is eventually minimized after multiple trials.

Amacrine cell: An interconnecting neuron in the retina; establishes inhibitory crosslinks among the bipolar/ganglion cell connections. (See Fig. 1.)

Attention: The part of cognition that serially controls how the various features or parts of a scene are brought together to form a unified percept. For example, eye movements are an attentional process for constructing an entire scene from the many detailed parts captured in the fovea. (See text p. 28.)

Axon: The fiber extending from a neuron cell body that carries the output signal of the neuron.

Back-propagation learning algorithm: A method for adjusting the weights in a multilayer neural-network so that a desired mapping from input patterns to output patterns may be learned by example. Essentially extends the ADALINE learning rule to a multiple layer network. (See Rumelhart et al, 1985.)

Bipolar cell: An interconnecting neuron in the retina; establishes conduction lines from photoreceptor cells to ganglion cells. (See Fig. 1.)

Boundary contour system: A method developed by Grossberg and Mingolla (1987) for cooperatively finding the edges of an object. Interactions among edge-detectors help to complete missing parts of a boundary. (See text p. 12.)

Center/surround profile: The concentric arrangement of excitory and inhibitory areas within a receptive field. An on-center/off-surround field has an excitory center and inhibitory periphery (see Fig. 2), while an off-center/on-surround field has an inhibitory center and excitory periphery.

Cerebral cortex: The extensive outer layer of gray tissue (densely packed neurons) of the cerebrum, largely responsible for higher nervous functions. (See Fig. 1)

Codon: A 2-D shape primitive used for describing planar curves. (See Fig. 26.)

Competitive learning: A method of unsupervised learning in a neural network (as opposed to learning with a "teacher," or a set of desired responses, as with ADALINE and back-propagation). Units in the network compete with one another for the highest response to an input pattern presented to the network. The unit with the highest response “wins” and adjusts its input weights so that those inputs with the highest activation are given more weight and those inputs with the lowest activation are given less weight. If m patterns are presented to m competing units, then after many trials, each unit will respond optimally to one pattern. (See Rumelhart and Zipser, 1985.)

Complex cell: A neuron in the visual cortex that selectively responds to a bar or edge with a specific angular orientation over a range of positions in the visual field. (See text p. 7.)

Cone: A photoreceptor cell selectively tuned for certain wavelengths of light (color). (See text p. 4.)

Convolution: The process of sliding one function over another function and integrating the product of the two functions at each step. Defined in one dimension as:

$$(f^*g)(x) = \int f(u)g(x-u)du.$$

Cooperative process: The process of dynamically combining many local computations or operations to yield a global result. Each local computation may influence or interact with other computations in accordance with some desired goal. For example, many noisy edge detectors in an image may cooperatively interact with each other so they align in the same direction. (See text p. 11 and 15.)

Dendrites: The fibers extending from a neuron cell body that carry input signals to the neuron.

Disparity: The slight spatial offset that is evident when comparing two images of the same object, where the object taken from slightly different viewing angles. Disparity is an important cue for recovering depth in a scene. (See text p. 18.)

Endstopped neuron: A neuron in the visual cortex that selectively responds to line-terminations; similar to a hypercomplex cell.

Energy functional: A way of describing the "energy" in a dynamical system subject to some cost or constraint. *Variational principles* (in mechanical systems, the *Euler-Lagrange equation*) provide a way to minimize the energy of the system. This technique is useful for solving many of the under-determined problems in vision, where there are fewer equations than there are unknowns. (See text p. 17.)

Excitory connection: A connection from one neuron to another such that activation of the input neuron increases the potential for activation in the receiving neuron.

Filter: In image processing: A mechanism for rejecting or enhancing certain spatial frequencies or features in an image. Filtering is accomplished by convolving a *mask* with an image.

Fourier descriptors: A set of numbers that can be used to uniquely describe a closed curve; these numbers are the Fourier coefficients of the function formed by plotting the accumulated change in direction of a curve along the length of its perimeter. (See text p. 24 and Fig. 22.)

Fourier transform: Transforms a function $f(x)$ in the time or space domain into a function $F(s)$ in the frequency domain. Thus, the function $F(s)$ reveals which temporal or spatial frequencies are dominant or weak in $f(x)$. Defined in one-dimension as:

$$F(s) = \int f(x)e^{-i2\pi xs}dx.$$

Fovea: The central 1-2° of the retina; the area of highest visual acuity. (See text p. 5.)

Gabor function (2D): A function formed from the product of a 2D sine and Gaussian. The aspect ratio of the Gaussian and the angular orientation, frequency, and phase of the 2D sine can take on various values in order to form a family of spatial functions. (See Fig. 8.)

Ganglion cell: A neuron in the retina that effectively collects the outputs of multiple photoreceptors. Ganglion cells form the last processing stage of the retina, and their long axons comprise the *optic nerve*, or output of the retina. (See Fig. 1.)

Gaussian filter: The filter formed by using the Gaussian function as the mask in a convolution.

Gaussian function: In two dimensions: $G(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x^2+y^2)/2\sigma^2}$

Grandmother cell: A neuron that is tuned to respond to one particular pattern. This term stems from the old notion that neurons in the brain are somehow directly related to specific environmental stimuli; for example, one neuron might be responsible for recognizing the face of one's grandmother, and it would fire only when one is looking at, or perhaps imagining, one's grandmother.

Hebbian learning: A form of learning in a neural-network devised by D. O. Hebb (1949). States that the connection strength from neuron *B* to neuron *A* is increased in proportion to the amount of correlated activity between *A* and *B* (i.e., the more the firing of *B* seems to contribute to the firing of *A*, the more weight the connection is given).

Hidden Unit: A unit in the hidden layer of a neural network. The hidden layer in a neural network lies between the input layer and the output layer, hence a unit in the hidden layer is "hidden" from both the input and output. (See Rumelhart et al., 1985.)

Horizontal cell: An interconnecting neuron in the retina; establishes inhibitory crosslinks among photoreceptor cells. (See Fig. 1.)

Hough transform: A parameter-space clustering technique (many-to-one mapping) for finding lines, circles, or arbitrary shapes in a scene. (See Fig. 19.)

Hypercolumn: A small block of visual cortex, approximately 1mm^2, containing 180° worth of orientation columns for both eyes. (See text p. 8 and Fig. 5.)

Hypercomplex cell: A neuron in the visual cortex that selectively responds to line terminations and corners. (See text p. 7.)

Illusory conjunction: A false combination of features that creates an illusory percept; usually caused by attention being overloaded. (See text p. 28.)

Illusory contour: A contour or boundary that is perceived even though it is not explicitly drawn. For example, the entire shape of a triangle may be strongly perceived even though only its vertices are drawn.

Inhibitory connection: A connection from one neuron to another such that activation of the input neuron decreases the potential for activation in the receiving neuron.

Intrinsic reference frame: A reference frame that is defined with respect to an object in an image. The axes of an intrinsic reference frame are aligned with natural features of the object (e.g., elongation, symmetry, or motion). For example, the intrinsic reference frame of a human head might be defined by the major and minor axes of its oval-like shape. (See text p. 27.)

IT (inferotemporal cortex): The part of the visual cortex believed to be involved in shape recognition. (See text p. 8.)

Laplacian operator: In two dimensions, the operator: $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$

Laplacian of Gaussian (LOG) filter: The filter formed by using $\nabla^2 G(x,y)$ as the mask in a convolution (where $G(x,y)$ is the Gaussian function). (See text p. 9 and Fig. 6e.)

Lateral geniculate nucleus (LGN): The part of the thalamus that serves as an intermediate connecting point for signals coming from the retina (optic nerve) on their way to the visual cortex. (See text p. 6 and Figs. 1 and 3.)

Log-polar coordinates: The coordinate system [log ρ, θ], where ρ and θ are related to cartesian coordinates $[x,y]$ by $\rho=\sqrt{x^2+y^2}$ and $\theta=\tan^{-1}(y/x)$. (See text p. 23 and Fig. 21.)

Magnocellular layers: The layers of the LGN containing *Y*-type cells. (See Fig. 3.)

Mask: The function that is to be convolved with an image in order to accomplish a desired filtering operation.

MST (medial superior temporal area): A part of the visual cortex believed to specialize in processing motion. (See text p. 8.)

MT (middle temporal area): A part of the visual cortex believed to specialize in processing motion. (See text p. 8.)

Neocognitron: A multilayer neural network that learns to classify visual patterns using a self-organizing principle. (See text p. 25 and Figs. 13 and 24.)

Neural network: A network formed from neuron-like elements. Each node, or *unit*, in the network sums together the outputs of other units in the network through a set of weights. The weights are usually automatically adjusted over many pattern presentations in order to train the network to perform some desired mapping from input patterns to output patterns. A neural network might be used to model the function of networks of neurons in the brain, or it may be totally unrelated to the brain and devised mainly as a tool for studying learning algorithms or massively parallel computation.

Ocular dominance column: A column of neurons in the visual cortex, perpendicular to the surface of the cortex; each neuron in the column responds preferentially to the same eye (left or right). (See Fig. 5.)

Optic nerve: The bundle of nerve fibers (long axons of the retinal ganglion cells) that serves as the conduction path from the retina to the lateral geniculate nucleus (LGN). (See Fig. 1.)

Optical flow field: A way of representing motion in the visual field. Each point or local area in the image is assigned a direction of motion based upon a local computation (time comparison) in its immediately surrounding neighborhood. (See Fig. 15.)

Orientation column: A column of neurons in the visual cortex, perpendicular to the surface of the cortex; each neuron in the column is selectively tuned to the same angular orientation (of a small line or edge in the visual field). (See Fig. 5.)

Parvocellular layers: The layers of the LGN containing *X*-type cells. (See Fig. 3.)

Photoreceptor: A cell (rod or cone) in the back of the retina that detects light intensity. (See Fig. 1.)

Pyramid: A series of images (usually of the same scene) that span over a range of scales or resolutions. (See Fig. 17.)

ρ-space representation: A three-dimensional discretized space for representing a boundary contour. Two dimensions are for spatial position, and the other is for orientation of edges. Excitory and inhibitory interactions among neighboring points in the space allow for certain parts of a contour in an image to be enhanced or suppressed. (See text p. 16.)

Receptive field: The local group of retinal photoreceptors from which a neuron receives its input (either directly or indirectly). Usually a neuron fires at its maximum frequency in response to a specific pattern of illumination on its receptive field. (See text p. 5 and Figs. 1 and 2.)

Regularization: A way of constraining the solution of a dynamical system to adhere to a sub-space of all possible solutions (see *Energy functional*).

Relaxation labeling: An iterative process for assigning labels to the nodes in a graph. Upon each iteration, the graph is incrementally "relaxed" into a state that best suits the compatibilities between labels at various nodes. (See text p. 11.)

Retina: The array of photoreceptors on the back wall of the eye. Light is focused onto the retina by the lens. (See Fig. 1.)

Rod: A photoreceptor cell that is extremely sensitive to light intensity, but indifferent to the wavelength, or color, of light. (See text p. 4.)

Segmentation: The process of delineating an object from its background. (See text p. 21.)

Self-organizing process: In a neural network: A way of adjusting the weights or synaptic strengths without the aid of a teacher, or a set of known, desired responses. A self-organizing neural network learns to categorize patterns based on the statistics and properties of the input. (See text p. 12.)

Scale-space filtering: A way of representing the changes in a signal or an image over a range of resolutions. (See Fig. 18.)

Simple cell: A neuron in the visual cortex that selectively responds to a small bar or edge with a specific position and angular orientation in the visual field. (See text p. 7.)

Sparse Distributed Memory (SDM): A massively parallel associative memory algorithm/architecture. (See Kanerva, 1988.)

Spline: A curve for data-fitting that is described by a parametric equation. A spline can be reshaped in various ways by adjusting parameters in its equation.

Superior colliculus: The part of the midbrain that receives a portion of the axons from the optic nerve (the rest of the axons go to the LGN). The superior colliculus plays an important role in controlling eye movements. (See text p. 5.)

Superior temporal sulcus (STS): A part of the cerebral cortex that receives inputs from the inferotemporal cortex. The STS has been shown to be involved in face recognition. (See text p. 8 and 30.)

Texton: A primitive texture element, with elementary properties such as color, angular orientation, width, and length. When a group of textons are viewed among a background of other textons, the boundary between the two groups can be perceived immediately. (See Fig. 14.)

Topographic mapping: In cortical neurophysiology: A projection of axons from one cortical area to another such that neighborhood relationships are preserved (i.e., neighboring neurons in one area send their signals to neighboring neurons in another area).

V1 (area 17, or striate cortex): The part of the visual cortex that first receives input from the LGN. (See text p. 7.)

V4: The part of the visual cortex thought to specialize in processing color information. (See text p. 8.)

Variance: A measure of the variation, or spread, of a random variable X about its mean μ. Defined as:

$$\mathrm{VAR}(X) = \mathrm{E}[(X - \mu)^2], \qquad \mu = \mathrm{E}[X]. \quad (\mathrm{E}[] = \text{expected value, or mean})$$

Visual cortex: The part the cerebral cortex devoted to processing visual information. (See text p. 6 and Fig. 1.)

***X* cell:** One of two classes of retinal ganglion cells; well-suited for shape analysis. (See text p. 5)

***Y* cell:** One of two classes of retinal ganglion cells; well-suited for motion analysis. (See text p. 5)

Zero-crossing: The point at which a function crosses zero. The zero-crossings a LOG-filtered image indicate the points of maximum intensity change in the image.

BIBLIOGRAPHY

Altes, R.A., "The Fourier-Mellin transform and mammalian hearing," *Journal of the Acoustical Society of America*, vol. 63, pp. 174-183, 1978.

Demonstrates how the Fourier-Mellin transform can be used to produce a signal representation that is independent of phase (shift) or scale changes in the input. Shows how the Fourier-Mellin transform can be implemented using a bank of proportional bandpass filters.

Anderson, C.H. and D.C. Van Essen, "Shifter circuits: A computational strategy for dynamic aspects of visual processing," *Proceedings of the National Academy of Sciences USA*, vol. 84, pp. 6297-6301, September 1987.

Proposes that "shifter circuits" in the primate visual pathway may be the mechanism responsible for controlling the flow of information from one layer of neurons to the next. Shows how a simple shifter circuit may be constructed, and explains how such circuits could be used for stereo matching, motion compensation, and directed visual attention. Physiological tests for shifter circuits are currently underway.

Arbib, M. and A.R. Hanson, "Vision, brain, and cooperative computation: An overview," in *Vision, Brain, and Cooperative Computation*, ed. M. Arbib and A.R. Hanson, MIT Press, Cambridge, MA, 1987a.

Presents a broad overview of theories and techniques used in vision, both in AI and in biological systems, tracing their development from past to present.

Arbib, M. and A.R. Hanson, eds., *Vision, Brain, and Cooperative Computation*, MIT Press, Cambridge, MA, 1987b. A collection of papers.

Asada, Haruo and M. Brady, "The curvature primal sketch," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 2-14, January 1986.

Presents a novel method for representing the significant changes in curvature along a two-dimensional object boundary. A set of primitive curvature discontinuities are defined, along with their convolutions with the first and second derivatives of a Gaussian. An object's boundary is similarly convolved and then matches are sought between the primitives and critical points along the object's boundary. In this way, an object's boundary can be represented as a composition of curvature primitives.

Attneave, F., "Some informational aspects of visual perception," *Psychol. Rev.*, vol. 61, pp. 183-193, 1954.

Ayache, Nicholas and Olivier D. Faugeras, "HYPER: A new approach for the recognition and positioning of two-dimensional objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 44-54, January 1986.

Babaud, Jean, Andrew P. Witkin, Michel Baudin, and Richard O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 26-33, January 1986.

Ballard, D., "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, pp. 111-122, 1981.

The generalized Hough transform uses parameter space clustering, as in the Hough transform for detecting lines, to detect arbitrary 2-D shapes. A model shape is described in an "R-table" of radii from the center of the shape vs. gradient direction (or tangent direction along the perimeter of the shape). Edge features from the image are then matched in parallel to gradients in the table to determine the correct correspondence between the image shape and model. Changes such as rotation, translation, scale or figure-ground reversals can easily be accounted for by simple transformations to the R-table; but matching in parallel would require increasing the dimensionality of the parameter space. This method would deal well with occlusion or missing information.

Ballard, D. and C.M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

Ballard, D. and D. Sabbah, "Viewer independent shape recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 653-659, 1983.

Extends the generalized Hough transform to 3-D. Also, scale, orientation, and translation are dealt with in sequence rather than in parallel. That is, matching is first done in the parameter space of orientation and scale, and then further according to translation. This reduces the dimensionality of the parameter space.

Ballard, D., G.E. Hinton, and T.J. Sejnowski, "Parallel visual computation," *Nature*, vol. 306, pp. 21-26, 1983.

Review article. Discusses how parallel architectures can be used to compute surface orientation, fuse stereo images, and generate feature maps. Includes a good list of references.

Ballard, D. and Lydia M. Hrechanyk, "Viewframes: A connectionist model of form perception," in *Proceedings, CVPR*, Washington, D.C., June 1983.

Ballard, D., "Parameter networks," *Artificial Intelligence*, vol. 22, pp. 235-267, 1984.

Describes in rather general form a connectionist theory of low-level and intermediate-level vision. The "intrinsic image," something like Marr's primal sketch or 2&1/2-D sketch, provides the low-level description of an image. Features in the intrinsic image are connected many-to-one onto parameter value nodes that represent object features (gestalt parameters). These object features also feedback to the intrinsic image to provide top-down reinforcement. The interconnection between feature space and intrinsic image is termed a "parameter network," as it is an organization of units, each representing the value of a parameter.

Ballard, D. and H. Tanaka, "Transformational form perception in 3-D: Constraints, algorithms, implementation," in *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pp. 964-968, 1985.

Describes a method for using 3-D shape primitives to build a scene description from stereo input. Features in a scene are transformed to match shape primitives via Hough techniques. This is carried out by relaxation in a neural network.

Ballard, D., "Cortical connections and parallel processing: Structure and function," *Behavioral and Brain Sciences*, vol. 9, pp. 67-120, 1986.

An extensive article. Attempts to relate the architecture and function of the visual cortex to the connectionist networks and Hough techniques described in previous papers. Specific examples are given for networks for modeling shape perception and motion perception. Includes an extensive peer commentary and response by Ballard.

Ballard, D., "Modular learning in neural networks," in *Proceedings AAAI* , pp. 279-284, 1987a.

Describes a way of coupling multiple-layer networks into modular hierarchies to increase performance and learning speed.

Ballard, D., "Eye Movements and Spatial Cognition," *University of Rochester Computer Science Department Technical Report*, vol. TR218, November 1987b.

Outlines a theory to explain how eye movements facilitate visual perception. The paper is broad in scope. It is argued that eye movements are more than just a complicating factor in computational vision, but rather play a pivotal role in the computations performed in the process of seeing.

Barlow, H.B., "Critical limiting factors in the design of the eye and visual cortex," *Proceedings of the Royal Society of London*, vol. B212, pp. 1-34, 1981.

Analyzes some of the design aspects of the eye and visual cortex. Barlow points out that there are many more granule cells in layers IVb+c in visual cortex than in corresponding areas in the retina, and he proposes that the granule cells are spatially smoothing and interpolating the sampled image from the retina (similar to Hinton's coarse-coding). This would explain how it is that we can detect spatial offsets much smaller than the sampling width in the retina. A similar theory is proposed for temporal sampling and interpolating. In the analysis of form and the recognition of gestalts, Barlow hypothesizes that the primary

visual cortex extracts "linking features" (such as color, texture, disparity, direction and velocity of motion, and orientation) locally and that these features are non-topographically mapped into other visual areas. Recognition would then be based on clusters of firing units in these areas.

Baron, Robert J., *The Cerebral Computer,* Erlbaum, Hillsdale, NJ, 1987. Chapters 7,8,and 9 on vision

Covers low-level visual processing and proposes theories for higher level representations and recognition as well. The log-polar transform is used to explain size and rotation invariance for patterns in the fovea. It is then proposed that the brain assembles the many patterns obtained from the fovea in order to form an overall image of the world (i.e. a gestalt can be formed by integrating parts of a whole obtained from the fovea). A face recognition experiment is described.

Barrow, H.G. and J.M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision Systems,* ed. A.R. Hanson and E.M. Riseman, Academic Press, New York, 1978.

Barrow, H.G., "Learning Receptive Fields," in *Proceedings of the IEEE First International Conference on Neural Networks,* vol. IV, pp. 115-121, 1987.

A competitive learning scheme, as in Rumelhart & Zipser (1985), is used to show how the transfer function of LGN cells and simple cells in V1 can be developed through experience. The network is exposed to natural images, such as a person's face. The connections from layer to layer are localized, like receptive fields, and the weights adjust their values according to Rumelhart and Zipser's rule.

Baylor, D.A. and C. Shatz, "Retina and anatomy of the visual system," in *Neurobiology 200 Lecture Notes,* pp. 365-413, Stanford University Medical School, Stanford, CA, 1988.

Serves as a good introduction to the human visual system.

Biederman, Irving, "Human Image Understanding: Recent Research and a Theory," in *Human and Machine Vision II,* ed. Azriel Rosenfeld, pp. 13-57, Academic Press, Boston, 1986.

Proposes that a set of 3-D shape primitives (probably no more than 36), called "geons," are used to represent 3-D objects. Objects are classified by the set of geons which compose them, much like phonemes. It is claimed that if an arrangement of two or three geons can be recovered from the input, objects can be quickly recognized even when they are occluded, rotated in depth, novel, or extensively degraded.

Biederman, Irving, "Matching image edges to object memory," in *International Conference on Computer Vision,* pp. 384-392, London, 1987.

More or less a summary of previous work.

Binford, T.O., "Survey of model based image analysis systems," *International Journal of Robotics Research,* vol. 1, pp. 18-64, 1982.

Blakemore, C. and R. Over, "Curvature detectors in human vision?," *Perception,* vol. 3, pp. 3-7, 1974.

Results of this paper suggest that curvature is initially processed by local orientation detectors. There may be a mechanism at a later stage that integrates the output of orientation detectors in order to determine curvature.

Block, H., N.Nilsson, and R. Duda, "Determination and detection of features in patterns," in *Computer and Information Sciences: Collected Papers in Learning, Adaption, and Control,* Spartan Books, Washington, 1964.

Brady, M., "Computational approaches to image understanding," *Computing Surveys,* vol. 14, pp. 3-71, 1982.

Presents a comprehensive survey of techniques used in image understanding, or computer vision. Covers theories of edge detection, segmentation, texture, shape from X (stereo, shading, motion, etc.), and viewpoint-independent representations of objects.

Breitmeyer, Bruno G., "Eye movements and visual pattern perception," in *Pattern Recognition by Humans and Machines: Vision Perception*, ed. Eileen C. Schwab and Howard C. Nusbaum, vol. 2, Academic Press, San Diego, 1986.

Brousil, J.K. and D.R. Smith, "A threshold logic network for shape invariance," *IEEE Transactions on Computers*, vol. EC-16, pp. 818-828, 1967.

A "property filter", made from threshold logic elements, is shown to be suitable for recognizing 2-D shapes independent of translation, rotation, and scale. Basically, the Fourier transform is used in conjunction with a log-polar transform. Patterns are separated by hyperplanes at the output.

Caelli, T., I. Rentschler, and W. Scheidler, "Visual pattern recognition in humans," *Biological Cybernetics*, vol. 57, pp. 233-240, 1987.

Along lines similar to Watson (1983). Uses the output of Gabor filters to form a feature vector. This feature vector is then used to classify visual forms using a least squares minimum distance classifier. Pattern classes are determined adaptively (i.e., dependent on the set of stimuli).

Campbell, F.W. and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology (London)*, vol. 197, pp. 551-566, 1968.

Canning, J., J.J. Kim, and A. Rosenfeld, "Symbolic pixel labeling for curvilinear feature detection," in *Image Understanding Workshop*, pp. 242-256, 1987.

Presents a method for detecting linear features that are approximately one pixel wide in an image. Each pixel in the image has a set of possible masks associated with it (this may be thought of as having a set of hypotheses about the local shape in the region of a pixel). Consistency links between neighboring masks provide a means of constructing connected components and extracting mid-lines.

Canny, J., "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.

Along a similar vein to Marr and Hildreth (1980), describes methods for finding optimal edge detectors (in one-dimensional cross-section) for various sorts of image intensity changes (ridges, roofs, and steps). For step edges, a two-dimensional operator is defined, based upon the gradient of a 2D Gaussian. A method is devised for combining results from operators of different width.

Carpenter, G.A. and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing*, vol. 37, pp. 54-115, 1987a.

A good article on Grossberg's Adaptive Resonance Theory (ART). Describes the basics of the theory, gives some examples, and provides proofs for various aspects of the theory.

Carpenter, G.A. and S. Grossberg, "Invariant pattern recognition and recall by an attentive self-organizing ART architecture in a nonstationary world," in *Proceedings of the IEEE First International Conference on Neural Networks*, 1987b.

Essentially uses the log-polar transform, in conjunction with Adaptive Resonance Theory (ART), to achieve rotation-, scale-, and translation-invariant pattern recognition.

Casasent, D. and D. Psaltis, "Position, rotation, and scale invariant optical correlation," *Applied Optics*, vol. 15, no. 7, pp. 1793-1799, 1976.

Describes a method for optically correlating 2-D patterns with invariance to position, rotation, and scale. The first step is to form the magnitude of the 2-D Fourier transform of the input image. This is then sent through a polar transform which separates the r and theta coordinates. The r coordinate is logarithmicaly scaled, and then another 2-D Fourier transform is formed to get rid of scale and rotation changes (shifts in the log-polar domain). The result is a 2-D function which can be optically correlated with other images, which have been similarly transformed, to determine similarity in form independent of translation, rotation, or size.

Cavanagh, P., "Size and position invariance in the visual system," *Perception*, vol. 7, pp. 167-177, 1978.

Proposes that local log-polar frequency transforms are formed in the striate cortex, with orientation varying along an axis parallel to the surface of the cortex, and log spatial frequency increasing along an axis perpendicular to the surface of the cortex. It is then hypothesized that these local transforms are integrated into a global log-polar frequency transform in inferotemporal cortex. This produces a size and position invariant representation, where size changes appear as shifts of an invariant pattern along the log-frequency axis, and position changes are lost in the amplitude of the Fourier transform.

Cavanagh, P., "Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance," in *Models of the Visual Cortex*, ed. D. Rose and V.G. Dobson, pp. 85-95, John Wiley & Sons, New York, 1985.

Shows how the model of Cavanagh (1978) can be used to achieve near-invariance to rotation as well as size and position. Points out that a form-specific encoding suitable for pattern matching can be formed by taking a Fourier transform of the global log-polar frequency transform in inferotemporal cortex. Provides a good diagram which illustrates the entire scheme, and briefly discusses how this kind of representation can be combined with a structural representation, such as that proposed by Marr.

Cooper, L.N., "Neuron learning to network organization," in *The Sesqucenntenial Symposium*, ed. J.C. Maxwell, Amsterdam, 1984. Cooper vision model

Cowan, W. Maxwell, "The development of the brain," in *The Brain*, pp. 56-69, W.H. Freeman and Company, New York, 1979.

Discusses how connections are formed during fetal development and infancy. Especially interesting is how connections are made in the optic tectum of the frog: axons and dendrites seem to be labeled in some way which tells them where to connect. Gives some idea of what can reasonably be expected to be determined genetically or through experience.

Crettez, Jean-Pierre and Steven L. Tanimoto, "Perceptual constancy and the multi-layer visual model: position and size invariance," *Proceedings of the Seventh International Conference on Pattern Recognition*, pp. 518-520, 1984.

Proposes a multi-layer model (similar to Watson) to explain size and position invariance. Resolution is very fine in the lower layers and gets coarser in higher layers. Distribution of units within a layer is uniform, so each layer displays shift invariance (in the linear systems sense). As a figure changes in size, its representation remains fairly constant, but shifted up or down the layers.

Crick, F. and C Asanuma, "Certain aspects of the anatomy and physiology of the cerebral cortex," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Volume 2: Psychological and Biological Models*, ed. J. L. McClelland and D.E. Rumelhart, pp. 333-371, MIT Press, Cambridge, Mass., 1985.

Provides an overview of the anatomy and physiology of the cerebral cortex. Discusses the neuron, synapse, organization of neurons into layers, cortical areas, inputs, outputs, cortical projections, cell types, behavior of single neurons and groups of neurons, and feature detection.

Daugman, J.G., "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America, Series A*, vol. 2, no. 7, July 1985.

Shows that the family of 2D Gabor functions (product of 2D sine and Gaussian) minimize the joint entropy, or uncertainty, of orientation, spatial frequency, and spatial position. That is, if you want to simultaneously extract angular orientation, spatial freq., and position information from an image, all with maximum specificity, then the Gabor is your function. Elaborates on many of the mathematical properties of Gabor functions, and shows the suitability of Gabor functions as a models of simple cell function.

Deutsch, S., "A simplified version of Kunihiko Fukushima's neocognitron," *Biological Cybernetics*, vol. 42, pp. 17-21, 1981.

Uses a 10 element 1-D visual field for presenting patterns. Shift invariance is demonstrated for simple features.

Dobbins, A., S.W. Zucker, and M.S. Cynader, "Endstopped neurons in the visual cortex as a substrate for calculating curvature," *Nature*, vol. 329, October 1987.

Presents a model for endstopped neurons in the visual cortex, viewing them as curvature detectors. According to this model, two simple cells feed their response into an end-stopped cell. One simple cell has a large receptive field and provides inhibitory input, and the other simple cell has a small receptive field and provides excitory input. This provides a measure for curvature, as the amount of curvature in a curve would determine how much of the excitory or inhibitory region is activated. Predictions of the model agree well with the response of cortical neurons (cat) to semi-circular arcs spanning a wide range of radii.

Duda, R.O. and P.E. Hart, *Pattern classification and scene analysis*, John Wiley, New York, 1973.

Farah, Martha J., "The neurological basis of mental imagery: A componential analysis," in *Visual Cognition*, ed. Steven Pinker, pp. 245-271, MIT Press, Cambridge, Mass., 1985.

Presents a componential model of mental imagery, and attempts to explain how various neurological deficits can be explained in terms of missing or inoperable components in the model.

Farhat, N., S. Miyahara, and K. Lee, "Optical implementation of 2-D neural networks and their application in recognition of radar targets," in *AIP Conference Proceedings 151: Neural Networks for Computing*, ed. J. Denker, pp. 146-152, American Institute of Physics, New York, 1986.

Feldman, J., "A connectionist model of visual memory," in *Parallel Models of Associative Memory*, ed. G. Hinton, Erlbaum Press, Hillsdale, 1981.

A very general, "conversational" paper. Describes in rough outline how such functions as mutual excitation, lateral inhibition, top-down reinforcement, and relaxation can be used to form a model of visual memory.

Feldman, J. and D. Ballard, "Connectionist models and their properties," *Cognitive Science*, vol. 6, pp. 205-254, 1982.

A very general exposition. Defines various types of units (e.g. p units, q units) and connections (e.g. conjunctive, disjunctive, additive), and shows how they can be used in networks which are applicable to problems in cognitive science/AI.

Feldman, J., "Connectionist models and parallelism in high level vision," *Computer Vision, Graphics, and Image Processing*, vol. 31, pp. 178-200, 1985.

Describes in rough outline a connectionist approach to high level vision. Features such as lightness, texture, shape, motion, and size are assumed to be extracted in the intermediate level process. These features are then used to "index" a certain visual object. Objects compete with one another and also provide top-down reinforcement to the features which excite them. The purpose of this paper is only to point out a general approach. Specifics are severely lacking.

Feldman, J., "A functional model of vision and space," in *Vision, Brain, and Cooperative Computation*, ed. M. Arbib and A.R. Hanson, MIT Press, 1987a.

Feldman, J., "Computational constraints on higher neural representations," in *Proceedings of the System Development Foundation Symposium on Computational Neuroscience*, 1987b.

Fischler, M.A. and O. Firschein, *Intelligence: The Eye, The Brain, and The Computer*, Addison-Wesley, Reading, Mass., 1987a.

From the preface: "An intellectual journey into the domain of human and machine intelligence." A large part of the book is devoted to vision, and discusses various interesting facts about human vision as well as methods and problems in computational vision.

Fischler, M.A. and O. Firschein, eds., *Readings in Computer Vision,* Morgan Kaufmann Publishers, Inc., Los Altos, CA, 1987b.

A collection of recent and important papers on computer vision. Covers traditional AI-type approaches as wells as neural-net-type approaches.

Fischler, Martin A. and Robert C. Bolles, "Perceptual organization and curve partitioning," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. PAMI-8, no. 1, pp. 100-105, January 1986.

Fu, K.S., *Syntactic Pattern Recognition and Applications,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

Although the method of pattern recognition discussed here differs greatly from the neural-net approach to pattern recognition, the "applications" part discusses several preprocessing methods that may be useful. The book discusses applications to character recognition, target detection, medical diagnosis, remote sensing, speech recognition, and identification of human faces and fingerprints.

Fukushima, K., "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics,* vol. 36, pp. 193-202, 1980.

The Neocognitron uses self-organizing feature detectors, organized in a hierarchical fashion, to develop a classification code for visual patterns unaffected by shift or small variations in size. The feature detectors in each layer receive their inputs from a pool of units within a certain receptive field in the previous layer. Self-organization is accomplished through a form of lateral inhibition and competitive learning, whereby the feature detectors discover such structures as lines and vertices (in the first layers) and eventually characters (in the final layer). In order to achieve translation invariance, the weight values of receptive fields are copied en masse all over a layer.

Fukushima, K., "A neural network model for selective attention in visual pattern recognition," *Biological Cybernetics,* vol. 55, pp. 5-15, 1986.

Gorman, John W., O. Robert Mitchell, and Frank P. Kuhl, "Partial Shape Recognition Using Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 10, no. 2, pp. 257-266, March 1988.

Grant, P. and J. Sage, "A comparison of neural network and matched filter processing for detecting lines in images," in *AIP Conference Proceedings 151: Neural Networks for Computing,* ed. J. Denker, pp. 194-199, American Institute of Physics, New York, 1986.

Compares the performance of matched filters with associative memories for the detection of lines in images (5x5 or 9x9 windows). The performance of the associative memory is inadequate when patterns are stored according to Hopfield (Hebbian memory matrix). The situation improves considerably when orthonormal basis vectors are used with analog processing on the first pass through the memory. Performance approaches that of matched filters when the nonlinear operator is modified. In all cases, the matched filter was shown to exhibit a reduced computational load.

Grimson, W.E.L. and T. Lozano-Perez, "Recognition and Localization of Overlapping Parts from Sparse Data," in *Three-Dimensional Machine Vision,* pp. 451-510, 1987.

Gross, C.G., C.E. Rocha-Miranda, and D.B. Bender, "Visual properties of neurons in inferotemporal cortex of the Macaque," *Journal of Neurophysiology,* vol. 35, pp. 96-111, 1972.

Investigates the response properties of the receptive fields of neurons in the inferotemporal cortex. The inferotemporal cortex is thought to be a logical place for processing visual patterns, and this paper attempts to pin down what it does. Receptive fields are much larger than those found in LGN or visual cortex, and many cover both left and right fields of vision. Most neurons exhibit the same selectiveness as complex and hypercomplex cells, but some neurons exhibit extreme selectivity, such as responding only to the silhouette of a monkey's hand (grandmother cell??).

Grossberg, S., ''On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems,'' *Biological Cybernetics*, vol. 21, pp. 145-159, 1976.

Describes how feature detectors may be developed by a self-organizing network. Basically uses Hebbian-type learning and competitive learning. Contrast enhancement is used as the competitive learning mechanism.

Grossberg, S. and E. Mingolla, ''Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations,'' *Perception and Psychophysics*, vol. 38, pp. 141-171, 1985.

Grossberg, S. and E. Mingolla, ''The role of illusory contours in visual segmentation,'' in *Proceedings of the International Conference on Illusory Contours*, ed. G. Meyer, Pergamon Press, New York, 1986.

Grossberg, S. and E. Mingolla, ''Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading,'' *Computer Vision Graphics and Image Processing*, vol. 37, pp. 116-165, 1987.

Provides a brief overview of the boundary contour system, in which oriented contrast detectors cooperate interactively to align along object boundaries or lines in an image. A process called "end cutting," in which edge elements perpendicular to a line are activated at line-terminations, is used to show how illusory contours can be formed from a series of properly placed line-terminations. The paper goes on to explain how concepts borrowed from the boundary contour system can be applied to the shape-from-shading problem. Lots of hand waving here.

Hartmann, G., ''Hierarchical contour coding by the visual cortex,'' in *Models of the Visual Cortex*, ed. D. Rose and V.G. Dobson, pp. 137-145, John Wiley & Sons Ltd., New York, 1985.

Attempts to explain how an unlimited number of differently running contours may be completely encoded by a finite number of cortical cells. Encoding is done in a hierarchical fashion: 7 LGN cells drive a cortical cell, 7 cortical cells drive a higher level cell, and so on. At the highest level, the firing of a cell would indicate the rough presence, shape, and location of a continuous contour.

Hartmann, G., ''Recognition of hierarchically encoded images by technical and biological systems,'' *Biological Cybernetics*, vol. 57, pp. 73-84, 1987.

Hebb, D.O., *The Organization of Behavior*, Wiley, New York, 1949.

Hinton, G., ''Relaxation and its role in vision,'' *Ph.D. Dissertation, University of Edinburgh*, 1979a.

Hinton, G., ''Some demonstrations of the effects of structural descriptions in mental imagery,'' *Cognitive Science*, vol. 3, pp. 231-250, 1979b.

Presents some interesting mental imagery tasks and perceptual demonstrations which lend further support to the hypothesis that structural descriptions play an important role of our internal representation of 3-D objects.

Hinton, G., ''A parallel computation that assigns canonical object-based frames of reference,'' *Proceedings IJCAI 1981*, pp. 683-685, 1981a.

Describes a connectionist model for generating a distributed, viewpoint-independent representation for 2-D objects. A set of mapping units transforms retinal features, in this case line segments, into canonical objects (gestalts). Recognition takes place as a competition/relaxation process, where retinal features receive top-down reinforcement from canonical object-based units, and mapping units compete with one another so that only one mapping unit wins (this specifies the viewer transformation for size, rotation, position). Thus, the viewer-transformation and recognition of the object are arrived at simultaneously.

Hinton, G., ''Shape representation in parallel systems,'' in *Proceedings of IJCAI 1981*, pp. 1088-1096, 1981b.

Outlines a general approach for representing shape in a connectionist network (in the framework previously specified, pp. 683-685 of these proceedings). Discusses how this model can be worked into a structural description by forming a hierarchy of feature based units. Also discusses how parts of a scene can be pieced together by "spatial working memory". Shows how coarse coding can be used to reduce the number of units in the network.

Hinton, G., "The role of spatial working memory in shape perception," in *Proceedings of the 3rd Annual Conference of the Cognitive Science Society*, pp. 56-60, 1981c.

Presents some demonstrations of visual perception/cognition tasks which illustrate various aspects of our internal representations of spatial structures. A particular mechanism for spatial representations is proposed. It is based on three frames of reference: retina frame, object frame, and scene frame. Various shape features and their relationship to each other are represented in the retina frame. These features activate some gestalt in the object frame, which is integrated into some larger whole or scene in the scene frame.

Hinton, G. and K. Lang, "Shape recognition and illusory conjunctions," in *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, 1985.

Shows how the phenomenon of illusory conjunctions, as described by Treisman and Schmidt (1982), can be explained with the model described in Hinton's 1981 IJCAI paper. Provides details of the algorithm used for relaxing the network (this was left out in the previous paper).

Horn, B.K.P., *Robot Vision*, MIT Press, Cambridge, Mass., 1986.

This textbook describes conventional techniques used in machine vision: edge detection, regior growing, shape from X, motion, and pattern classification.

Hrechanyk, Lydia M. and D. Ballard, "A connectionist model of form perception," in *Proceedings of the IEEE Workshop on Computer Vision*, 1982.

In addition to using coarse coding to reduce the number of units, Hrechanyk proposes splitting the parameter space, so that rotation and scale are considered apart from translation. Hrechanyk also discusses how different parts of shape might be hierarchically organized.

Hubel, D.H. and T.N. Wiesel, *Journal of Neurophysiology*, vol. 28, pp. 229-289, 1965. discusses "higher-order complex cells"

Hubel, D.H. and T.N. Wiesel, "Brain mechanisms of vision," in *The Brain*, pp. 84-96, W.H. Freeman and Company, New York, 1979.

Provides a clear description of the discoveries of Hubel and Wiesel. Explains the circularly symmetric receptive fields of LGN cells, the orientation selectivity of simple, complex, and hypercomplex cells, ocular dominance, and the columnar organization of the visual cortex.

Hummel, R.A. and S.W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 267-287, 1983.

The purpose of this paper is to present a formal treatment of relaxation labeling processes described earlier (Zucker, 1977). A labeling problem can generally be described as one of assigning labels to nodes in a graph. Given some set of possible labels for each object and a constraint relation over labels at pairs (or n-tuples) of neighboring objects, certain labelings are defined as being consistent. It is shown that a functional exists which can be maximized in the search for consistent labelings. This functional is used to derive an algorithm for "relaxing" the graph into a consistent state.

Hutchinson, James, Christof Koch, Jin Luo, and Carver Mead, "Computing motion using analog and binary resistive networks," *IEEE Computer*, pp. 52-63, March 1988.

Shows how regularization theory can be applied to the computation of motion, as in Poggio et al (1985), and demonstrates how this computation can be implemented in a simple resistive network. The linear equations resulting from the minimization of the cost functional (Euler-Lagrange equations) are solved via Kirchoff's current law. Segmentation of a moving object (separation of figure from ground) can be accounted for by allowing discontinuities in the optic flow. In the network, these discontinuities are formed by putting binary switches between pixel nodes. The network is being implemented using analog VLSI.

Jaeckel, Louis A., *Character recognition using a sparse distributed memory system,* 1988. to be published as a series of RIACS Technical Reports

Describes a method for encoding characters for SDM (Sparse Distributed Memory). Assumes that a character has been properly centered and scaled in the upright position, and that a pre-processor has already broken up the character into line segments or arcs. The parameters of locatioa, length, and angular extent (for arcs) for each piece are encoded into bit strings so that the distance between parameters is related by Hamming distance. Two schemes are presented for addressing SDM with the pieces of a character. One imposes no ordering on the pieces and uses very long bit strings. The other uses shorter bit strings, but requires an ordering of the pieces. Some preliminary results are presented.

Julesz, B., *Foundations of Cyclopean Perception,* University of Chicago Press, Chicago, Ill, 1971.

Julesz, B. and J.R. Bergen, "Textons, the fundamental elements in preattentive vision and perception of textures," *The Bell System Technical Journal,* vol. 62, no. 6, pp. 1619-1645, July-August 1983.

Introduces the notion of "textons," elongated blobs with properties such as color, angular orientation, width, length, etc., as being the fundamental elements detected by the preattentive visual system. Only differences in textons can be preattentively detected; Further processing involving "focal attention" is necessary to determine positional information.

Julesz, B., "Toward an axiomatic theory of preattentive vision," in *Dynamic Aspects of Neocortical Function,* ed. G. Edelman and W.M. Cowan, John Wiley and Sons, New York, 1984.

Kahan, Simon, Theo Pavlidis, and Henry S. Baird, "On the recognition of printed characters of any font and size," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. PAMI-9, no. 2, pp. 274-288, March 1987.

Kandel, E. and J. Schwartz, *Principles of Neuroscience,* Elsevier Publishing, New York, 1985.

Kanerva, P., *Sparse Distributed Memory,* MIT Press, Cambridge, Mass., 1988.

Discusses the theory and principles of sparse distributed memory (SDM). Basically, SDM provides a massively parallel algorithm/architecture for for an associative memory. Long bit vectors serve as both data and addresses to the memory, and patterns grouped or classified according to similarity in Hamming distance.

Kass, Michael and Andrew Witkin, "Analyzing oriented patterns," in *Proceedings of the International Joint Conference on Artificial Intelligence,* pp. 944-952, Los Angeles, 1985.

Oriented patterns produced by natural processes (such as wood grain) are analyzed in terms of flow fields. First, oriented spatial filters are convolved with a pattern in order to determine the direction of flow. The flow field is then used to form a coordinate system in which to view the pattern. Viewing the pattern in flow coordinates can be advantageous for providing preferred directions for edge detection.

Kass, Michael, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," in *International Conference on Computer Vision,* pp. 259-268, London, 1987.

Introduces the concept of "snakes" for finding contours in natural images. A snake is an energy-minimizing spline, whose shape is determined by constraint forces. Internal forces act to regularize the spline so that it remains smooth. Image forces push the snake toward lines, edges, and subjective contours in the image. External forces can be used to interactively (such as with a mouse) nudge a snake toward certain image features. This method works remarkably well for latching onto smooth, continuous contours.

Keeler, J., "Comparison between Kanerva's SDM and Hopfield-type neural network models," *Cognitive Science,* vol. 12, pp. 299-329, 1988.

Develops a mathematical framework for comparing SDM and Hopfield-type neural nets. Limits of capacity and ability to store sequences for both modelsare discussed and compared.

Kersten, D., A. O'Toole, M. Sereno, D. Knill, and J. Anderson, "Associative learning of scene parameters form images," *Applied Optics.* in review

Koch, C., J. Marroquin, and A. Yuille, "Analog "neuronal" networks in early vision," *Proceedings of the National Academy of Sciences USA*, vol. 83, pp. 4263-4267, June 1986.

Koch, C., "Computing optical flow in man and machine," in *Proceedings AAAI Symposium on Physical and Biological Approaches to Computer Vision*, pp. 119-123, 1988.

Koenderink, J.J. and A.J. van Doorn, "Representation of local geometry in the visual system," *Biological Cybernetics*, vol. 55, pp. 367-375, 1987.

Koenderink, J.J. and W. Richards, "Two-dimensional curvature operators," *Journal of the Optical Society of America, Series A*, vol. 5, no. 7, pp. 1136-1141, July 1988.

Presents some methods for detecting planar curvature using two-dimensional operators. One way is to use receptive field profiles similar those of end-stopped line detectors in the visual cortex. The aspect ratio of these receptive fields could be varied in order to detect varying degrees of curvature. Another way is to use co-circularity. According to this method, two tangents (e.g., simple cell responses) at different spatial locations are defined as co-circular if they are both tangent to the same circle. This sort of primitive grouping operation would help to avoid an explosion of receptive fields of increasingly higher order.

Koffka, K., *Principles of Gestalt Psychology*, Harcourt, Brace, and World, New York, 1935.

Kohler, W., *Gestalt Psychology*, Liverright Press, 1947.

Kohonen, T., "Clustering, taxonomy, and topological maps of patterns," in *Proceedings of the 6th International Conference on Pattern Recognition*, pp. 114-125, 1982.

Uses Hebbian learning to form a "topographic" feature map of patterns. Mathematical formulation and proofs of convergence are given. Several examples of topological maps formed in computer simulations are described.

Kohonen, T. and K. Makisara, "Representation of sensory information in self-organizing feature maps," in *AIP Conference Proceedings 151: Neural Networks for Computing*, ed. J. Denker, pp. 271-276, American Institute of Physics, New York, 1986.

Similar to the work of von der Malsburg, except more general. Uses lateral inhibition and Hebbian learning to form a feature map. Associative memory can then be used to deal with incomplete information. An example is given for phonemes.

Krishnan, G. and D. Walters, "Psychologically plausible features for shape recognition in a neural-network," in *Proc. International Conference on Neural Networks*, vol. II, pp. 127-134, San Diego, 1988.

Shows how shapes (line drawings) can be classified by a neural network using psychologically plausible features. Shapes are encoded into bit strings based on properties such as orientation of edge elements, angular separation of corners, and ratio of chord length to arc length. These bit strings are then concatenated to form a one-dimensional feature vector to be stored in an associative memory (with grandmother cells at the output). The associative memory learns shape categories based on a novel learning rule (modified Hebb and Anderson rules).

Kuffler, Stephen W., John G. Nicholls, and A. Robert Martin, *From Neuron to Brain*, Sinauer Associates Inc., Sunderland, Mass., 1984. chapters 2,3, & 20

Covers the anatomy, physiology, and development of the mammalian visual system Good reference for defining terms and illustrations.

Lee, David and Theo Pavlidis, "One-dimensional regularization with discontinuities," in *International Conference on Computer Vision*, pp. 572-577, London, 1987.

Presents a method for regularizing splines (smoothness constraint) which allows for important discontinuities such as corners without smoothing over them.

Lehky, S.R. and T.J. Sejnowski, "Network model of shape-from-shading: neural function arises from both receptive and projective fields," *Nature*, vol. 333, no. 6172, pp. 452-454, June 1988.

Shows how a backpropagation network can learn to compute shape from shading. The input consists of 122 units, each of which calculate a LOG weighted sum of a local area in an image. These units are fully connected to 27 units in the hidden layer; and each unit in the hidden layer is fully connected to 24 units in the output layer. It is desired that each unit in the output layer be maximally responsive for a certain combination of curvature and surface orientation in the input surface. After 40,000 presentations, the network settles on a solution. Interestingly, the hidden units develop receptive field profiles much like those of bar or edge detectors in V1. Even more interesting is the arrangement of weights in the 'projective field' (from the hidden units to the output units), which seem to provide information about surface orientation, convexity/concavity, and relative magnitudes of curvature.

Lettvin, J.Y., H. Maturana, W. McCulloch, and W. Pitts, "What the frog's eye tells the frog's brain," *Proceedings of the IRE*, vol. 47, pp. 1940-1951, 1959.

This was one of the first papers on receptive fields. Reveals four major functions of ganglion cells in the retina of the frog: 1) sustained contrast detectors 2) net convexity detectors 3) moving edge detectors and 4) net dimming detectors.

Link, N.K. and S.W. Zucker, "Corner detection in curvilinear dot grouping," *Biological Cybernetics*, vol. 59, pp. 247-256, 1988.

Attempts to determine how sensitive we (humans) are to corners, and offers a model to explain how corners are detected. Dotted curves are used as the stimuli; and it is shown how sensitivity to orientation discontinuities varies as a function of dot phase (i.e., placement of the dots). Concludes that simple cells must interact to compute curvature.

Linsker, Ralph, "From basic network principles to neural architecture," *Proceedings of the National Academy of Sciences USA*, vol. 83, pp. 7508-7512, 8390-8394, 8779-8783, 1986. three part article series

Demonstrates that a multiple-layer network is capable of developing spatial-opponent cells (center/surround type), orientation selective cells, and hypercolumns as in visual cortex. Connections from layer to layer are localized, like receptive fields, and a Hebb rule is used to modify the synapses. The spatial distribution of synapses for any given cell is gaussian distributed. Spatial-opponent cells develop in the third layer, and orientation selective cells develop in the seventh layer. If lateral connections are added to the seventh layer, then hypercolumns emerge.

Linsker, Ralph, "Self-organization in a perceptual network," *IEEE Computer*, pp. 105-117, March 1988.

Nicely explains the results obtained by Linsker's (1986) multiple-layer network in terms of information theory. Shows how a simple Hebb rule can be used to achieve maximum variance in a cell's output; This corresponds to Principal Component Analysis, a widely used statistical method for feature extraction. Using concepts from information theory, it is then shown that each layer of cells in the network preserves maximum information about its input from the previous layer. In this sense, the cells within a layer become optimal feature analyzers.

Livingstone, Margaret and David Hubel, "Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception," *Science*, May 6, 1988.

Loebner, Egon E., "Concurrency Assurance in vertebrate retinas," in *Proceedings of the IEEE First International Conference on Neural Networks*, vol. IV, pp. 147-159, 1987.

Lowe, D.G. and T.O. Binford, "Segmentation and aggregation: an approach to figure-ground phenomena," in *Proc. DARPA Image Understanding Workshop*, pp. 168-178, 1982.

Describes an approach to low-level vision that is based on measurements that can be computed directly from the image, rather than on prior world knowledge. It is noted that humans are easily capable of detecting patterns in an otherwise random field of dots; how may computer vision systems emulate such an ability? Groupings of image elements are said to be "meaningful" if it is more likely that they arose from underlying physical relationships between constituent features (say, due to the boundary of an ob-

ject), rather than from some coincidence of viewpoint or location. Describes the implementation of a "meaningfulness" algorithm for finding lines.

Lowe, D.G., "Visual recognition from spatial correspondence and perceptual organization," in *IJCAI-85 Proceedings*, pp. 953-959, 1985a.

Describes a vision system called SCERPO (Spatial Correspondence, Evidential Reasoning, and Perceptual Organization) for recognizing 3-D objects from arbitrary viewpoints. The system is based on three separate mechanisms: 1) a process of perceptual organization finds groupings and structures in the image (such as lines segments) that are likely to be invariant over a wide range of viewpoints; 2) evidential reasoning is used to reduce the size of the search space for object matching; and 3) spatial correspondence is used to project the model object (i.e., the previously stored object believed to be present in the image) onto the image for refinement and verification.

Lowe, D.G., *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, Boston, Mass., 1985b.

Provides depth and details of the IJCAI-85 paper.

Marr, D. and T Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283-287, October 1976.

Presents a cooperative algorithm for computing disparity from stereo image pairs. Two constraints are used in the computation: 1) Each pixel may be assigned only one disparity value, and 2) disparity values vary smoothly almost everywhere, since in general objects have smooth surfaces. Illustrations are provided showing results of the computation.

Marr, D. and H.K. Nishihara, "Representation and recognition of the spatial organization of three-dimensional shapes," *Proceedings of the Royal Society of London*, vol. B 200, pp. 269-294, 1978.

Marr, D. and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London*, vol. B 207, pp. 187-217, 1980.

Marr's classic paper on edge detection. Describes a method for 1) detecting intensity changes in images (Laplacian of Gaussian operator) and 2) interpreting intensity changes to form a description of the image, called the raw primal sketch. The method serves as a physiological model of simple cell function.

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

A comprehensive presentation of Marr's work on vision. Discusses the primal sketch, 2&1/2-D sketch, stereopsis, motion, 3-D shape extraction, and 3-D model representation. The primal sketch is a hierarchical representation of the intensity discontinuities in an image, showing raw intensity changes at the lowest level and groupings and alignments at the highest level. Then, the 2&1/2-D sketch incorporates clues from stereo disparity, shading, motion (optical flow), and shape contours to infer depth and surface orientation. Both the primal sketch and 2&1/2-D sketch are done in a viewer-centered coordinate system. For purposes of recognition, 3-D objects are represented in an object-centered coordinate system. The definition of a shape's object-centered coordinate system is based on axes determined by some salient, geometric characteristic of the object (e.g. elongation, symmetry, or motion). Moreover, the representation of shape is modular, so that a shape may be described at varying levels of detail. Objects are thus stored in a "Catalogue of 3-D Models," which is indexed by 3-D shape primitives derived from the 2-D image. Marr considers there to be three levels of understanding for all theories on vision: Computational theory, Representation and algorithm, and Hardware implementation.

Mesrobian, Edmond and Josef Skrzypek, "Discrimination of natural textures: a neural network architecture," in *Proceedings of the IEEE First International Conference on Neural Networks*, vol. IV, pp. 247-258, 1987.

Describes a neural network architecture for discriminating textures. At the lowest level, orientation-specific edge detectors (i.e. simple cells), are used for feature extraction. Higher level units aggregate regions based on similarity, such as complex cells. At a next higher level, boundaries are determined by differentiating aggregated regions.

Miller, B. K. and R. A. Jones, "Reliable formation of feature vectors for 2-D shape representation," in *Proceedings of the SPIE: Computer Vision for Robotics*, vol. 595, pp. 109-118, 1985.

Describes a method for generating a feature vector for describing 2-D shapes with invariance to translation, rotation, and scale. The method is based upon creating a concavity tree from the closed curve which describes a 2-D shape. Invariants of the curve are derived from the tree and they are used as the components of the feature vector. Successful results are demonstrated for three shapes under rotation, translation, scaling, and occlusion.

Minsky, Marvin and Seymour Papert, *Perceptrons (expanded edition)*, MIT Press, Cambridge, Mass., 1988.

Miyake, S. and K. Fukushima, "A neural network for the mechanism of feature-extraction: A self-organizing network with feedback inhibition," in *AIP Conference Proceedings 151: Neural Networks for Computing*, ed. J. Denker, pp. 305-308, American Institute of Physics, New York, 1986.

This model is like the Neocognitron, except that it uses feedback inhibition. Learned features are suppressed in the input layer by feedback inhibition. Novel features remain excited so that they may be learned by self-organization. This method seems to encourage the differentiation of features while preserving stability (i.e. old features don't tend to get wiped out by new features).

Mokhtarian, Farzin and Alan Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 34-43, January 1986.

Presents a method for describing and matching planar curves. A "scale-space image" (see Witkin, 1983) of a curve is formed by finding the zero's of curvature at several levels of detail along its path. The curve is then matched to other curves by comparing their scale space images.

Nagahashi, Hiroshi and Mikio Nakatsuyama, "A pattern description and generation method of structural characters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 112-118, January 1986.

Nalwa, V.S. and T.O. Binford, "On Detecting Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 699-714, Nov. 1986.

Proposes an approach to edge-detection based upon fitting a series of "one-dimensional" surfaces to each window. A least-squares method is used choose the surface that best fits an edge (plane, cubic spline, tanh), and hence best identifies the position and angular orientation of the edge.

Nalwa, V.S., "Edge-detector resolution improvement by image interpolation," in *Image Understanding Workshop*, pp. 981-987, 1987.

Nauta, W.J.H. and M. Feirtag, "The organization of the brain," in *The Brain*, pp. 40-55, W.H. Freeman and Company, New York, 1979.

Provides a very clear description of the organization of the brain. Very useful for getting a feel for what the retina, LGN and visual cortex have to do with other parts of the brain, such as hippocampus, amygdala, and superior colliculus.

Nauta, W.J.H. and M. Feirtag, *Fundamental Neuroanatomy*, pp. 280-315, W.H. Freeman and Company, New York, 1986. Cerebellar cortex, Neocortex, Prospects

Excellent descriptions, diagrams, and electron micrographs of the "circuitry" of cerebellar and cerebral cortex.

Nishihara, H.K., "Practical real-time imaging stereo matcher," *Optical Engineering*, vol. 23, no. 5, pp. 536-545, 1984.

Okajima, K., "A mathematical model of the primary visual cortex and hypercolumn," *Biological Cybernetics*, vol. 54, no. 2, pp. 107-114, 1986.

Proposes a model for visual cortex (not very novel) in which each hypercolumn performs a local spatial Fourier analysis. This is seen as a sort of tomographic representation. Advantages: The hypercolumns are invariant to lateral shift, which is consistent with the Fourier transform, and spatial filtering operations can easily be performed.

Oyster, J.M., F. Vicuna, and W. Broadwell, "Associative network applications to robot vision," *IBM Los Angeles Scientific Center Report No. G320-2777*, Jan. 1986.

Oyster, J.M. and Josef Skrzypek, "Computing shape with neural networks: A proposal," in *Proceedings of the IEEE First International Conference on Neural Networks*, vol. IV, pp. 335-344, 1987.

Proposes a neural structure for the recognition of arbitrary 2-D curves. At the lowest level, local edge, orientation, curvature, corner, and end features are extracted. The features are aggregated by higher levels for detection of more complex features and eventually recognition of overall shape. This paper is greatly lacking in specifics.

Palm, G. and A. Aertsen (eds.), *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory, October 1-5, 1984*, Springer-Verlag, 1986. neural modeling, plus a good collection of classic papers

Palmer, S.E., "The psychology of perceptual organization: A transformational approach," in *Human and Machine Vision*, ed. Jacob Beck, Barbara Hope and Azriel Rosenfeld, pp. 269-339, Academic Press, Orlando, 1983.

A theoretical framework is presented for understanding the phenomena of shape constancy, motion perception, figural goodness, perceptual grouping, and reference frame effects. It is argued that these phenomena can be understood in terms of invariance transformations. For example, shape constancy can be accomplished through a transformation in scale, rotation, and translation from the retinal frame to the model frame.

Parent, P. and S. Zucker, "Curvature, consistency and curve detection," *Technical Report CIM-86-3*, Computer Vision and Robotics Lab, McGill University, Montreal, June 1985.

Describes a method for inferring the trace of a curve based on relaxation labeling. Estimated tangents are constrained by a neighborhood relationship called co-circularity, and curvature estimates are constrained by a curvature consistency relation. The result is an optimal estimation of tangent and curvature information along the path of a curve, and hence a good recovery of the trace.

Pavlidis, T., *Structural Pattern Recognition*, Springer-Verlag, Berlin, 1977.

Perona, Pietro, "Anisotropic diffusion: A scale space technique for edge detection in digital images," in *IEEE Computer Society Workshop on Computational Vision*, Miami, Nov/Dec. 1987. more comprehensive report from this author (at U.C. Berkeley) available by end of summer 1988

Perrett, D.I., E.T. Rolls, and W. Caan, "Visual neurones responsive to faces in the monkey temporal cortex," *Experimental Brain Research*, vol. 47, pp. 329-342, 1982.

Reports that out of a population of 497 neurons recorded in the superior temporal sulcus (STS), at least 48 neurons were selectively responsive to faces. 28 neurons exhibited relatively constant responses despite transformations such as size and rotation (2-D), or other changes such as color or distance. Some neurons showed a bias to particular facial features, such as the mouth or eyes. It is hypothesized that the STS, which receives inputs from the inferior temporal cortex and sends efferents to the amygdala, parietal cortex and frontal cortex, may be specialized to code for faces.

Pinker, Steven, "Visual cognition: An introduction," in *Visual Cognition*, ed. Steven Pinker, pp. 1-64, MIT Press, Cambridge, Mass., 1985a.

Gives an overview of the issues and problems of visual cognition. Discusses theories of shape recognition, such as template matching, feature models, Fourier models, structural descriptions, the Marr-Nishihara theory, reference frames, and massively parallel models (Hinton). Also discusses theories of mental imagery.

Pinker, Steven, ed., *Visual Cognition*, MIT Press, Cambridge, Mass., 1985b. collection of papers

Pitts, W. and W. McCulloch, "How we know universals: The perception of auditory and visual forms," *Bulletin of Mathematical Biophysics*, vol. 9, pp. 127-147, 1947.

Proposes some interesting neural mechanisms for the recognition of invariants. Auditory patterns are scanned sequentially in all translations and visual patterns are scanned in all sizes. Translation in visual patterns is accounted for by an automatic centering mechanism in the superior colliculus which keeps all targets in the fovea centralis.

Plaut, D.C., "Visual recognition of simple objects by a connection network," *University of Rochester Computer Science Dept. Technical Report*, vol. TR143, August 1984.

Plyshyn, Z.W., "What the mind's eye tells the mind's brain: a critique of mental imagery," *Psychological Bulletin*, vol. 80, pp. 1-24, 1973.

Poggio, T., V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, 1985.

Shows how regularization theory can be applied to certain "ill-posed" problems in early vision. For example, to compute the direction of motion from many local measurements, the assumption of smoothness (i.e. uiat the direction of motion probably wiii not change from one pixel to the next, since most moving objects are rigid) constrains the problem in such a way that it is no longer ill-posed. Such a constraint can be mathematically formulated as part of a cost functional. Variational principles can then be applied to find the solution, such as direction of motion (optic flow), which minimizes the the functional.

Poggio, T. and K. Koch, "Ill-posed problems in early vision: From computational theory to analog networks," *Proceedings of the Royal Society of London*, vol. B 226, pp. 303-323, 1985.

Poggio, T., "Making machines (and artificial intelligence) see," *Daedalus*, vol. 117, no. 1, pp. 213-240, 1988.

Pollen, D.A., J.R. Lee, and J.H. Taylor, "How does the striate cortex begin the construction of the visual world?," *Science*, vol. 173, pp. 74-77, 1971.

Proposes that the complex cells in striate cortex compute local Fourier transforms. Such a representation would conserve information and also produce invariant descriptions of visual objects with respect to translation.

Ponce, J. and D. Chelberg, "Localized intersections computation for solid modelling with straight homogeneous generalized cylinders," in *Proc. Image Understanding Workshop*, pp. 933-941, 1987.

Discusses a method for modelling solids based on using generalized cylinders (GC's) as primitives. Presents a fast algorithm for computing set operations (unions, intersections) between different types of GC's to form compound shapes.

Pratt, William K., *Digital Image Processing*, John Wiley & Sons, New York, 1978.

This textbook describes many of the commonly used techniques in digital image processing.

Prazdny, K., "Similitude-invariant pattern recognition using parallel distributed processing," in *Proceedings Sixth National Conference on Artificial Intelligence*, Seattle, WA, 1987a.

Presents a method for position-, rotation-, and scale-invariant pattern recognition of 2-D objects. The approach is similar to Tucker et al. (1988). First, image features such as intersecting lines or vertices are extracted. Every correspondence between an image feature and a model feature is a "vote" for a particular model. Every such correspondence instantiates the model in the image frame for verification; the model broadcasts what features it expects and in what area it expects them ("attention beams"). In the end, the model with the most votes and the best verification score wins.

Prazdny, K., "Position-, rotation-, and scale invariant pattern recognition using parallel distributed processing," in *Proceedings First International Conference on Computer Vision*, London, 1987b.

Same as the 1987 Seattle paper.

Psaltis, D. and J.Hong, "Shift-invariant optical associative memories ," *Optical Engineering*, vol. 26, no. 1, pp. 10-15, 1987.

Reilly, D.E., "A neural model for category learning," *Biological Cybernetics*, vol. 45, pp. 35-41, 1982. Cooper vision model

Richards, Whitman and Donald D. Hoffman, "Codon constraints on closed 2D shapes," in *Human and Machine Vision II*, ed. Azriel Rosenfeld, pp. 207-223, Academic Press, Boston, 1986.

Introduces the idea of "codons" as simple 2-D shape primitives for describing plane curves. A contour or boundary curve can be considered to consist of a string of such codons. Because of the strong constraints imposed by the bounding contour (silhouette) formed from 3D objects, only a small set of "realistic" curves may be generated from any set of codons, thus making the codon representation highly redundant (good for error correction).

Rock, I., *Orientation and Form*, p. Academic, New York, 1973.

Rosenfeld, A., "Some pyramid techniques for image segmentation," in *Pyramidal Systems for Computer Vision*, ed. V. Cantoni and S. Levialdi, Springer-Verlag, Berlin, 1986a.

Discusses the use of pyramidal techniques (representing images at various resolutions) for rapidly extracting global structures from an image. Such techniques are amenable to parallel implementation.

Rosenfeld, A., *Human and Machine Vision II*, Academic Press, Boston, 1986b. A collection of papers.

Rosenfeld, A., "Recognizing unexpected objects: A proposed approach," in *Image Understanding Workshop*, pp. 620-627, 1987.

Rosenfeld, A., "Image analysis and computer vision: 1987," *Computer Vision Graphics and Image Processing*, vol. 42, pp. 234-293, 1988.

A bibliography of over 1400 references (1987 only!) covering research in computer vision and image analysis, arranged by subject matter.

Rumelhart, D. E. and D. Zipser, "Feature discovery by competitive learning," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Volume 1: Foundations*, ed. D.E. Rumelhart and J.L. McClelland, pp. 151-193, MIT Press, Cambridge, Mass., 1985.

Shows how a simple model of competitive learning, involving a form of lateral inhibition and Hebbian learning, can classify "features," or groupings of the input pattern set. Provides a good mathematical analysis of the model and some interesting experimental results on patterns presented on a 2-D grid.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Volume 1: Foundations*, ed. D.E. Rumelhart and J.L. McClelland, pp. 318-362, MIT Press, Cambridge, Mass., 1985.

The definitive paper on back-propagation. Derives the generalized delta rule for gradient descent and shows the results of several interesting experiments. Shows how a back-prop network can be trained for rotation and shift invariance to the letters T and C (weights ar copied en masse to achieve shift invariance).

Sabbah, D., "Design of a highly parallel visual recognition system," *IJCAI*, vol. 7, pp. 722-727, Vancouver, B.C., 1981.

Sabbah, D., "A Connectionist Approach to Visual Recognition," *TR107, Computer Science Department, University of Rochester*, April 1982.

Applies the connectionist theories of Ballard and Feldman to visual recognition in Kanade's Origami World. A conceptual hierarchy is defined, where levels represent the extraction of progressively more complex features. At the lowest level, edge segments are detected. These are coalesced into lines and rays, then L-joints and T-joints, then complex joints and 2-D shapes, and then finally 3-D Origami objects. Between levels, top-down reinforcement is used to further enhance those units which help to comprise something "meaningful" in the level above. Within levels, local lateral inhibition causes units to compete

with one another, which in turn causes noisy, ambiguous or otherwise un-substantiated input to die out. Sabbah demonstrates the operation of the network with a number of examples. Especially evident is the ability of the network to recognize objects which have been partially occluded or which have information missing.

Sabbah, D., "Computing with connections in visual recognition of Origami objects," *Cognitive Science*, pp. 25-50, Winter 1985.

See 1982 tech report.

Sacks, Oliver W., *The Man Who Mistook his Wife for a Hat*, Summit Books, 1985.

Contains an account of a man with visual agnosia. He can tell what features an object is comprised of, but he cannot classify the object as a whole unless some other sensory clue is given (touch, verbal hint).

Sanger, T.D., "Optimal unsupervised learning in a single-layer linear feedforward neural network," in *First Annual INNS Meeting*, Boston, MA, 1988.

Presents a method for unsupervised learning in a neural network that optimizes the amount of information preserved in the output layer (similar to Linsker, 1988). It is shown that a Hebbian learning rule can be used to learn the eigenvectors of the input auto-correlation matrix (that is, the Karhunen-Loeve transform). When the input is an image (actually, many sub-images), the eigenvectors turn out to be very similar to the receptive field profiles of cells in the retina, LGN, and visual cortex. The paper also discusses the significance of these results for texture segmentation and receptive field development.

Schwab, E.C. and H.C. Nusbaum, eds., *Pattern Recognition by Humans and Machines: Vision Perception*, 2, Academic Press, San Diego, 1986. collection of papers

Schwartz, E.L., "Computational geometry and functional architecture of striate cortex," *Vision Research*, vol. 20, pp. 645-669, 1980a.

Shows that the mapping from retina to visual cortex is of the form log(z+c), where z is complex and c is real (in effect, a warped log-polar transform). Demonstrates how scaling and rotation of a pattern on the retina are transformed into shifts of a somewhat invariant pattern on the retina. (Note: this mapping applies to the central 20-30 degrees of visual field. The Fovea Centralis, which covers only the central 1-2 degrees of visual field, would not appear to demonstrate rotation invariance since it lies in the area most warped from the additive constant in the transform.) Gives an explanation for some optical illusions (e.g. MacKay complementary image illusion) based on the local and global architecture of visual cortex. Suggests that neurons infero-temporal cortex may detect boundary curvature on the retina by detecting "lines" on the surface of striate cortex (i.e., because orientation preference changes across the surface of striate cortex, a line of active neurons would most likely indicate a curve).

Schwartz, E.L., "A quantitative model of the functional architecture of human striate cortex with application to visual illusion and cortical texture analysis," *Biological Cybernetics*, vol. 37, pp. 63-76, 1980b.

In addition to the material presented previously (Vision Research, 1980), this article also points out how the cortical representation can be utilized for texture analysis. Schwartz shows how hypercolumns in striate cortex can encode certain textures in such a way that they may be easily segmented.

Schwartz, E.L., "Cortical anatomy, size invariance, and spatial frequency analysis," *Perception*, vol. 10, pp. 455-468, 1981.

Points out several serious flaws with Cavanagh's hypothesis regarding size and position invariance in the human visual system. Namely, he refutes Cavanagh's claim that the visual system is shift invariant and that it does a global Fourier analysis by rejecting the phase from piecewise Fourier transforms. Gives a shortened version of the 1980 paper.

Schwartz, E.L., R. Desimone, T. Albright, and C. Gross, "Shape recognition and inferior-temporal neurons," *Proceedings of the National Academy of Sciences*, vol. 80, pp. 5776-5778, 1983.

Proposes that Fourier Descriptors (Zahn & Roskies 1972) are used to encode shape information in inferotemporal cortex. In an experiment on inferotemporal neurons in the macaque monkey, it was found that many neurons (54% of 234 visually responsive units) were selective to the frequency of the Fourier Descriptor stimuli, mostly independent of size and position. These results suggest that inferotemporal cortex may code for shape on the basis of global boundary curvature, much as striate cortex codes for shape on the basis of local edge orientation.

Schwartz, E.L., "Local and global functional architecture in primate striate cortex: outline of a spatial mapping doctrine for perception," in *Models of the Visual Cortex*, ed. D. Rose and V.G. Dobson, pp. 146-156, John Wiley & Sons Ltd., New York, 1985.

More discussion of the log(z+c) mapping from retina to cortex. Delves more into the local structure of the mapping.

Schwartz, E.L. and Yehezkel Yeshurun, "Towards a non-network approach to "neural modeling": some basic issues of measurement, simulation, and computational significance of brain maps," in *Proceedings of the IEEE First International Conference on Neural Networks*, vol. IV, pp. 225-233, 1987.

Discusses some techniques for 3-D modelling of neuroanatomy. Illustrates a simulation of how a retinal image appears on the visual cortex. Since the mapping from retina to cortex is extremely space-variant (due to the logarithmic spacing of sensors in the retina), a theory is proposed for explaining how the visual system fuses individual "scans" of an object into a single percept.

Sejnowski, T.J., "Open questions about computation in cerebral cortex," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Volume 2: Psychological and Biological Models*, ed. J. L. McClelland and D.E. Rumelhart, pp. 372-389, MIT Press, Cambridge, Mass., 1985.

Poses some interesting questions about computation in cerebral cortex. How is information represented (grandmother cells vs. distributed representations)? How is information processed? Is some sort of iterative relaxation scheme reasonable to expect given the slow switching speed of neurons? What sort of temporal dependencies exist? How can the functional connectivity of the cortex be reconfigured with experience and still make sense?

Sejnowski, T.J. and Geoffrey E. Hinton, "Separating figure from ground with a Bolzmann machine," in *Vision, Brain, and Cooperative Computation*, ed. M. Arbib and A.R. Hanson, MIT Press, Cambridge, Mass., 1987.

A relaxation process is proposed for separating figure from ground. Each edge is considered to be part of the figure/ground boundary, with one side pointing toward figure and the other toward ground. Edges interact with their neighbors so that a consistent state is reached (neighboring edges agree on where the figure or ground is).

Shepard, R.N. and J. Metzler, "Mental rotation of three-dimensional objects," *Science*, vol. 171, pp. 701-703, 1971.

Reports that the time required to correctly match two 3-D objects which have been rotated relative to each other is linearly proportional to the amount of rotation. The objects were displayed as 2-D perspective drawings, and they were rotated either within the picture plane or in depth. The slope of the relationship would imply that "mental rotations" are done at 60 degrees per second.

Sparks, David L. and Martha Jay, "The role of the primate superior colliculus in sensorimotor integration," in *Vision, Brain, and Cooperative Computation*, ed. M. Arbib and A.R. Hanson, MIT Press, Cambridge, Mass., 1987.

The superior colliculus is thought to translate sensory signals from several modalities (visual, auditory, somatosensory) into motor commands for directing eye movements. This paper proposes that the superior colliculus contains a map of motor error and is organized in motor coordinates, rather than sensory coordinates.

Sutherland, N.S., "Object recognition," in *Handbook of Perception, Vol. 3, Biology of Perceptual Systems*, ed. E.D. Carterette, pp. 157-206, Academic Press, New York, 1973.

Tenenbaum, J.M. and A. Witkin, Program chairs, *Symposium: Physical and Biological Approaches to Computational Vision*, AAAI, Menlo Park, CA, 1988.

Terzopoulos, D., ''Regularization of inverse visual problems involving discontinuities,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 413-424, 1986.

Timney, B.N. and C. Macdonald, ''Are curves detected by 'curvature detectors?','' *Perception*, vol. 7, pp. 51-64, 1978.

Describes some experiments aimed at determining whether curves are detected by 'curvature detectors.' This question remains unanswered. Author is inconclusive as to whether curves are detected by a set of linear contour detectors, or detectors designed specifically for curves.

Trehub, Arnold, ''Visual-cognitive neuronal networks,'' in *Vision, Brain, and Cooperative Computation*, ed. M. Arbib and A.R. Hanson, MIT Press, Cambridge, Mass., 1987.

Presents an all-encompassing neural network model of the visual system. Networks are presented for storing and learning patterns, transforming patterns according to rotation or scale, and integrating patterns into objects or scenes.

Treisman, A.M. and G. Gelade, ''A feature-integration theory of attention,'' *Cognitive Psychology*, vol. 12, pp. 97-136, 1980.

It is proposed that the visual scene is initially coded along a number of separable dimensions, such as color, orientation, spatial frequency, brightness, size, etc., and that these features are registered early, automatically, and in parallel across the visual field; Objects are identified separately and only at a later stage, which requires focused attention. The feature-integration theory of attention suggests that attention is directed serially to each stimulus in a display whenever conjunctions of more than one separable feature are needed to characterize or distinguish the possible objects presented. This paper presents compelling evidence for such a theory.

Treisman, A.M. and H. Schmidt, ''Illusory conjunctions in the perception of objects,'' *Cognitive Psychology*, vol. 14, pp. 107-141, 1982.

As a corollary to the feature-integration theory of attention (Treisman 1980), this paper proposes that when attention is diverted or overloaded, features may be wrongly combined, giving rise to "illusory conjunctions." For example, brief presentation of a red T and a blue S may be incorrectly registered as a blue T and a red S. Such experiments suggest that our internal representation contains discrete labels of values on each feature dimension separately, and that a whole object must be resynthesized from a set of these feature labels.

Treisman, A.M., ''The role of attention in object perception,'' in *Physical and Biological Processing of Images*, ed. O. Braddick and A. Sleigh, Springer, London, 1982.

Treisman, A.M. and R. Paterson, ''Emergent features, attention, and object perception,'' *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, no. 1, 1984.

Treisman, A.M., ''Preattentive processing in vision,'' in *Human and Machine Vision II*, ed. Azriel Rosenfeld, pp. 313-334, Academic Press, Boston, 1986.

Presents further research on theories of search and attention. Hypothesizes that search for the presence of a visual primitive is automatic and parallel, whereas search for the absence of the same feature is serial and requires focused attention.

Tucker, Lewis W., Carl R. Feynman, and Donna M. Fritzsche, ''Object Recognition Using the Connection Machine,'' in *Proceedings CVPR-88*, 1988.

Describes a model-based object recognition system and its parallel implementation on the Connection Machine. Similar in approach to Ballard. Local boundary features, in this case corners formed by intersecting line segments, are extracted from the image. These boundary features are matched in parallel to all possible model instances through a set of viewing transform parameters (translation or rotation). Each match to a model instance generates a "hypothesis", that is, a vote for a particular model and a set of viewing transform parameters (translation and rotation). Each hypothesis projects its model instance back onto the image for verification, and that hypothesis which has the highest confidence (most votes) and the strongest verification wins (i.e. object is recognized). Because corner features and hypotheses are assigned one per processor, recognition time is less than linearly proportional to the number of objects in the data base or the complexity of the scene.

Ullman, S., *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA, 1979.

Ullman, S., "Visual routines: Where bottom-up and top-down processing meet," in *Pattern Recognition by Humans and Machines: Visual Perception*, ed. Eileen C. Schwab and Howard C. Nusbaum, vol. 2, Academic Press, San Diego, 1986.

Discusses visual processing in terms of two stages: 1) the creation of base representations (primal sketch, 2&1/2-D sketch), which is a bottom-up, spatially uniform process in a viewer centered frame, and 2) the application of visual routines to base representations, which is a top-down process for defining objects, parts, and spatial relations.

Ungerleider, L.G. and M. Mishkin, in *Analysis of Visual Behavior*, pp. 549-586, MIT Press, Cambridge, MA, 1982.

Unnikrishnan, K.P., A.S. Pandya, and E. Harth, "Role of feedback in visual perception," in *Proceedings of the IEEE First International Conference on Neural Networks*, vol. IV, pp. 259-267, 1987.

Presents a model of visual perception which accounts for the extensive reciprocal connections between lower and higher centers of visual processing (retina, LGN, layers of V1, etc.). When a certain stimulus at one level resembles some pre-defined pattern by the second level, then that stimulus receives top-down enhancement; otherwise, the stimulus is suppressed. The net effect is that extraneous features are suppressed and missing features are completed. (see Sabbah 1982)

Van Essen, D.C. and J.H.R. Maunsell, "Hierarchical organization and functional streams in visual cortex," *Trends in Neuroscience*, vol. 6, pp. 370-375, 1983.

Describes the hierarchical structure, inter-relationships, and function of various areas of cortical visual processing. Briefly discusses the role of MT (middle temporal cortex) and MST (medial superior temporal cortex) in motion analysis, the role of V4 in color perception, and the role of IT (inferotemporal cortex) in form perception (so-called "grandmother cells" responsive to faces (Perrett 1982) or hands (Gross 1972) have been reported in this area). Discusses some of the regularities in the hierarchical structure: Connections between cortical areas tend to be reciprocal, such that if A projects to B then B also projects to A; Also, receptive field size increases at successive stages of the hierarchy. Good list of references.

Van Essen, D.C., "Functional organization of primate visual cortex," in *The Cerebral Cortex*, vol. 3, pp. 259-329, Plenum Press, New York, 1985.

A much more detailed, thorough, and up-to-date version of Van Essen's 1983 TIN paper.

Vilnrotter, F.A., R. Nevatia, and K. Price, "Structural analysis of natural textures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, pp. 76-89, January 1986.

Vistnes, Richard, "Detecting dotted lines and curves in random-dot patterns," in *Image Understanding Workshop*, pp. 849-861, 1987.

Presents a model for explaining human ability to pre-attentively detect dotted lines and curves in images. The model is based on detecting "non-accidental" structures, such as a high density of dots within an elongated region compared with its local surrounding area. Predictions of the model compare well with human performance.

von der Malsburg, Chr., "Self-organization of orientation sensitive cells in the striate cortex," *Kybernetik*, vol. 14, pp. 85-100, 1973.

Uses a model of lateral inhibition and Hebbian learning to show how the orientation selectivity of simple cells in the visual cortex can be developed through experience, rather than being pre-determined genetically. A retina of 19 units is stimulated with 9 different patterns in the form of light bars at all orientations.

A "cortex" consisting of 338 units is connected to the retina such that each retinal cell excites all the cortical cells through a set of weights. These weights are modified according to a Hebbian-type rule. Within the cortex, there is an on-center/off-surround interaction such that the firing of one cell helps excite its neighbors, but inhibits more distant neighbors. After 100 trials, the cortex exhibits much the same behavior as Hubel and Wiesel found in the visual cortex of mammals.

von Seelen, W., H. A. Mallot, and F. Giannakopoulos, "Characteristics of neuronal systems in the visual cortex," *Biological Cybernetics*, vol. 56, pp. 37-49, 1987.

A linear systems approach to modeling the visual cortex: 2-D spatial filtering, feedback, retinotopic mapping, nonlinear cortex couplings. Not too clear, but contains good diagrams of retinotopic mapping and the relation of visual cortex to other areas of the brain.

Voorhees, Harry and Tomaso Poggio, "Detecting textons and texture boundaries in natural images," in *International Conference on Computer Vision*, pp. 250-258, London, 1987.

Proposes a method for extracting textons (Julesz, 1983) from natural images. An image is Gaussian filtered in order to estimate the background noise. The background noise level is then used as the threshold level for an edge detection operator (LOG filter). The result is that edges are found along texture boundaries instead of purely intensity boundaries.

Walters, D., "Selection of image primitives for general-purpose visual processing," *Computer Vision, Graphics, and Image Processing*, vol. 37, pp. 261-298, 1987.

Describes some perceptually significant visual features (namely, relations between line-ends). Shows that all possible relations between line-ends, or end-connections, can be classed as one of just four types of connections. This sort information can be used to enhance contours in an image that have a high probability of being part of an object's contour. Describes the "rho-space representation," a three-dimensional discretized space (2 dim. for spatial pos. in the image and 1 dim. for orientation of edges), with inhibitory and excitory connections among points in the space for enhancing or rejecting certain parts of a contour.

Walters, D., "Orientation based contour descriptors," in *AAAI Symposium on Physical and Biological Approaches to Computational Vision*, pp. 30-32, 1988a.

Contours are represented in the "rho-space representation" (see Walters, 1987). Contour descriptors are extracted from the rho-space representation and provided to the recognition stage for classification. This method was applied to the task of recognizing hand-drawn numerals. Shows that local zero-crossings of curvature can be readily computed from the local neighborhood computations of the rho-space representation.

Walters, D., "Integration of texture properties for texture segmentation," in *AAAI Symposium on Physical and Biological Approaches to Computational Vision*, pp. 33-35, 1988b.

Waltz, D., "Understanding line drawings of scenes with shadows," in *The Psychology of Computer Vision*, ed. P.H. Winston, pp. 19-91, McGraw-Hill, New York, 1975.

Watson, A.B., "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images*, ed. O.J. Braddick and A.C. Slade, pp. 100-114, Springer-Verlag, Berlin, 1983.

Uses the Gabor function (product of 2-d sine and Gaussian) as a model of simple cell function. A feature vector is generated by convolving Gabor functions of various location, size, spatial frequency, and orientation with a simple spatial pattern, such as a grid. Detection or recognition is performed by comparison with a template feature vector (least squares).

Watson, A.B., "The cortex transform: Rapid computation of simulated neural images," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 311-327, September 1987.

Expands somewhat on the work in the 1983 paper. Describes a transform which simulates the representational transformation from retina to visual cortex. Windows of various spatial bandwidth and orientation selectivity are convolved with an image through multiplication in the frequency domain. The result is a series of image pyramids, where resolution varies within a pyramid, and orientation selectivity varies from one pyramid to the next. The transformation is invertible.

Wechsler, H. and G.L. Zimmerman, "2-D invariant object recognition using distributed associative memory," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 811-821, Nov. 1988.

Uses the log-polar transform as the front end to an associative memory. Objects are recognized despite changes in scale and rotation. Some simple examples are given.

Weisstein, Naomi and Eva Wong, "Figure-ground organization and the spatial and temporal responses of the visual system," in *Pattern Recognition by Humans and Machines: Vision Perception*, ed. Eileen C. Schwab and Howard C. Nusbaum, vol. 2, Academic Press, San Diego, 1986.

Widrow, B., "Generalization and information storage in networks of Adaline neurons," in *Self Organizing Systems*, ed. M.C. Yovits, G.T. Jacobi, and G.D. Goldstein, pp. 435-461, Spartan Books, Washington, 1962.

Widrow's classic neural nets paper. Describes the operation of the ADALINE (ADAptive LInear NEuron) and provides a proof of convergence for the LMS algorithm. Demonstrates the applicability of the ADALINE to 2-D pattern recognition tasks, and shows how the ADALINE can be trained to generalize with respect to rotation, size, translation, and noise. For example, to generalize with respect to 90 deg. rotation, the ADALINE forms a matrix of weights which is equal to its transpose. Thus a pattern will produce the same response no matter what its orientation.

Widrow, B., R.G. Winter, and R.A. Baxter, "Learning phenomena in layered neural networks," in *Proceedings of the IEEE First International Conference on Neural Networks*, 1987.

Shows how a multiple layer network can be constructed as an "invariance net," such that 2-D patterns can be transformed into patterns invariant to rotation, size, and translation. This is a kind of "brute force" approach, such as in Fukushima's Neocognitron: Weights are copied and then translated, rotated, or scaled en mass within a "slab" of ADALINEs in order to produce an invariant response. Widrow has a nice, simple rule for training multiple-layer nets called "don't rock the boat."

Wiesel, T.N., "The postnatal development of the visual cortex and the influence of environment," *Nature*, vol. 229, pp. 583-591, 1982.

Describes an experiment where a new-born monkey was deprived of vision in its right eye while the left eye was exposed to vertical stripes for 57 hours; The result was that vertically oriented stimuli became much more effective in driving cortical cells of the left eye than those of the right eye. Horizontally oriented stimuli gave equal responses for both eyes. This suggests that some competition among neurons is taking place during early development. Agrees very nicely with competitive learning models.

Williams, R., "Feature discovery through error-correction learning," *University of California at San Diego, Institute for Cognitive Science Technical Report*, vol. 8501, 1985.

Wilson, H.R. and S.C. Giese, "Threshold visibility of frequency gradient patterns," *Vision Research*, vol. 17, pp. 1177-1190, 1977.

Wilson, H.R. and J.R. Bergen, "A four mechanism model for spatial vision," *Vision Research*, vol. 19, pp. 19-32, 1979.

Proposes that the visual scene is analyzed with four different size-tuned mechanisms (i.e., four different resolutions). Each mechanism is described by the the difference of two 2-D Gaussian functions - essentially a center/surround-type receptive field profile. The four receptive fields have central widths of 3.1', 6.2', 11.7', and 21' at the fovea.

Wilson, H.R., "Discrimination of contour curvature: data and theory," *Journal of the Optical Society of America*, vol. 2, no. 7, pp. 1191-1199, July 1985.

Presents the theory, with experimental results to support it, that curvature discrimination is based upon mechanisms selective for orientation and spatial frequency.

Witkin, A.P. and J.M. Tenenbaum, "On the role of structure in vision," in *Human and Machine Vision*, ed. J. Beck et al., Academic Press, 1982.

Witkin, A.P., "Scale-space filtering," in *Proceedings IJCAI*, pp. 1019-1022, 1983.

Describes a method for richly and compactly describing signals over a variety of scales. Signals are filtered at several scales (several values of sigma are chosen for a Gaussian convolution kernel) and a "scale-space image" is formed by the surface swept out by laying the filtered signals side by side. Extrema can be identified at coarse scales and then traced to finer scales for localization.

Witkin, A.P., Demetri Terzopoulos, and Michael Kass, "Signal matching through scale space," *International Journal of Computer Vision*, vol. 1, no. 2, pp. 133-144, 1987.

Describes a method for matching signals (1-D or 2-D) which have been deformed with respect to each other (such as a motion sequence or stereo pair). The matching process is formulated in terms of energy minimization, involving constraints on smoothness and similarity. An optimal match is first found at a coarse scale and then tracked to a fine scale. Results are presented for a one-dimensional signal, a motion sequence, and a stereo pair.

Wojcik, Zbigniew, "Rough approyimation of shapes in pattern recognition," *Computer Vision, Graphics, and Image Processing*, vol. 40, pp. 228-249, 1987.

Attempts to create a "language" for describing shapes from their contours. A small window is passed over the contour and features are extracted at each point, such as lines, corners, and intersections. These features compose a "sentence" which can be used to universally recognize the object.

Yang, Hedong and Clark C. Guest, "Performance of backpropagation for rotation invariant pattern recognition," in *Proceedings of the IEEE First International Conference on Neural Networks*, vol. IV, pp. 365-370, 1987.

Uses backpropagation to train a 2 layer network to recognize 2-D shapes invariant to rotation. Four patterns, A,T,H, and R, presented on a 16x16 array, are trained in at all rotations in 15 deg. intervals. Thus, rotations between intervals are considered distortions of the trained in patterns. The number of units in the hidden layer is arbitrarily chosen to be 64. Four grandmother cells are used at the output for recognizing each of the four patterns. With some modification in the sigmoid function, all four patterns can be recognized at any rotation.

Zahn, C.T. and R.Z. Roskies, "Fourier descriptors of plane closed curves," *IEEE Transactions on Computers*, vol. C21, pp. 269-281, 1972.

Presents a method for forming an invariant description of plane closed curves. A curve is represented parametrically as a function of arc length by the accumulated change in direction of the curve along the perimeter. The Fourier coefficients of this function can then be used to uniquely describe the curve invariant to changes in rotation, translation, or scale (the perimeter is normalized to 2pi).

Zucker, S.W., "Toward a model of texture," *Computer Graphics and Image Processing*, vol. 5, pp. 190-202, 1976.

Zucker, S.W., R.A. Hummel, and A. Rosenfeld, "An application of relaxation labeling to line and curve enhancement," *IEEE Transactions on Computers*, vol. C-26, no. 4, pp. 394-403, 922-929, April 1977.

Shows how relaxation labeling can be used to enhance lines and curves in images. An image pixel array is considered as a graph of objects with edge orientation labels attached to them (edgels). In the relaxation process, each object updates its label to be more compatible with its neighbors, as determined by a set of compatibility weights between line labels. The compatibility weights are chosen such that edges of similar orientation support one another, while edges of perpendicular orientation antagonize one another; "no-line" labels are supported positively by surrounding "no-line" labels and negatively by line labels oriented toward the point. This process converges, in only a few iterations, so that global lines or curves are enhanced and "noisy" elements are suppressed.

Zucker, S.W. and R.A. Hummel, "Receptive fields and the representation of visual information," *Proceedings of the Seventh International Conference on Pattern Recognition*, pp. 515-517, 1984.

Zucker, S.W. and P. Parent, "Multiple-size operators and optimal curve finding," in *Multiresolution Image Processing and Analysis*, ed. A. Rosenfeld, pp. 200-210, Springer Verlag, Berlin, 1984.

Applies relaxation labeling to line and curve enhancement at multiple resolutions. Edges detected by large scale operators provide contextual constraint for edges detected by smaller scale operators.

Zucker, S.W., "Early orientation selection: Tangent fields and the dimensionality of their support," in *Human and Machine Vision II*, ed. Azriel Rosenfeld, pp. 335-364, Academic Press, Boston, 1986.

Considers two types of orientation structure in images: Type I processes, which are 1-dimensional contours such as boundary curves, and Type II processes, which are 2-dimensional flows such as wood grain or hair. Algorithms are formulated for inferring a vector field of tangents from such patterns.