VERSION SPACES. See Concept learning; Learning.

VISION, EARLY

The study of biological perception has had an enormous influence on the development of computational vision. Unfortunately, the most common observation about biological—in particular, human—vision is its immediacy: simply open your eyes and a percept of the world appears. This apparently effortless speed implies to many that vision is relatively simple and that general-purpose vision systems should not be too difficult to construct. However, this effortless speed results not from the simplicity of vision but rather from the immense amount of specialized wetware, the biological equivalent of hardware, dedicated to it. In humans the visual system occupies a major portion of the cortex. Vision is, in fact, immensely complex, and it has turned out to be immensely difficult to construct successful vision systems. Endless applications in areas as diverse as robotics (qv), biomedicine, and remote sensing all support this conclusion.

Two basic problems confront the designers of complex early vision systems: What are the fundamental pieces, or the individual tasks comprising vision, and how should they be solved. The problems are clearly related; either the designer has a problem for which he can foresee a solution or he has a solution "in search of a problem." Although much of the insight into how to decompose vision has come from mathematics, physics, and computer science/engineering, perhaps the most powerful influence to date has been from the study of biological vision systems. This entry is a historical survey of the modern development of early computational vision. The goal is to illustrate how diverse the influences on computational vision have been and to argue that such diversity is necessary.

The entry proceeds as follows. It begins about 100 years ago with the two great vision scientists von Helmholtz (1821–1894) and Mach (1838–1916), see General References. Two themes emerge from their different views of vision. In Helmholtz a clear separation can be seen between low-level and high-level processing, or what is now sometimes called early and later processing; and in Mach there is a separation between the analysis of a task and the mechanism proposed to accomplish it. (Early processing does not imply a temporal dimension; rather, the term "early" denotes processing from the retina back into the cortex, and "later" denotes the latter stages of cortical processing). These two themes were present in the first attempt at a complete computer vision system—the one by L. Roberts—and they persist to the present. In Roberts's system there was a clear separation between low-level processing, or the extraction of a cartoonlike line drawing out of an image, and high-level processing, or the recognition of objects. And the mechanisms applied at these levels depended on the tasks; the low-level mechanism being one of so-called edge detection (qv) and the higher level one of object matching into a database. Modern computational theories, such as the one proposed by Marr (1), postulate more elaborate interfaces: "primal sketches." Although this thread is common, the main evolution of computational vision has been an appreciation of the immense complexity involved in both of these stages, with one paradigm after another attempting to grapple with it. Different paradigms have arisen for low- and high-level processing, and some have even emerged for intermediate stages. Strong forms of so-called inverse optics, pieced together with little or no interaction, have now given way to an increased appreciation of abstract structure. That is, it has now become clear that it is essentially impossible to exactly invert the scene projection process; rather, the search is on for discovering which aspects of the structure of the world can—and should—be recovered.

The detailed evolution of the field can be thought of in terms of two pendula, one representing the tension between low-level and high-level vision, and the other between the formulation of the task and the techniques chosen to solve it. This is very much a personal view, as are the examples that I have chosen to illustrate how these pendula swing back and forth in time. Interestingly, early on in the development of the field they were assumed to be rather separate from one another, but as the field began to mature, their interrelationships became more clear as well. The tension between low and high level vision developed into a concern for the type of knowledge to be applied, the specifics of which are clearly related to both task formulation and technique employed.

The vision problem can be summarized as follows. Three-dimensional physical structure in the scene projects into two-dimensional structure in the image. This process must be inverted; i.e., somehow, physical structures must be inferred from image structures. For each class of related physical and image structures a microinverse problem can be formulated, and many such problems exist, as we shall describe. Early (low level) vision consists of those problems for which the solution is driven by general-purpose assumptions and special-purpose hardware, whereas later (high level) vision consists of those problems for which the solution is driven by special-purpose assumptions and general-purpose hardware. Or stated differently, in early vision, if something is understood about structure (of the world), something can be inferred about function in the visual system; whereas in later vision it appears that function must be understood before structure.

The focus of this entry is on the evolution of ideas rather than on algorithms. There is more concentration on the classical foundations of the field than on current approaches. Several other articles in this *Encyclopedia* address these different aspects of vision in detail, and cross-references to them are indicated whenever possible. Furthermore, given space limitations the entry needs to be somewhat selective about material. Many book length treatments of computational vision (2–8), image processing (9,10), and visual perception (11–16) are available and should be consulted along with this entry. It is also worthwhile to consult the annual list of publications in computer vision and image processing compiled every year and published by Rosenfeld in the journal *Computer Vision, Graphics, and Image Processing*. Also, several recent collections have emphasized the relationships between biological and computational vision (17,18).

First Paradigm: Segmentation in Low Level and High Level Vision

The study of vision begins with the grand decomposition suggested by Hermann von Helmholtz (19).

Helmholtz: Physiological Optics and Unconscious Inference. In his treatise on physiological optics (19) Helmholtz sketched a theory of vision in which the eye acted as a transducer of light into the nervous system, which then performed "unconscious inferences" in order to compose internal versions of percepts. That is, he asserted that there was a low-level component to vision, dominated by physics and physical models, and a high level component in which the inferences (qv) took place. Unfortunately, the only language that he had for talking about inferences was the rather loose one of what he took to be "conscious inferences," or the logic of premises and conclusions. High level vision is, he therefore asserted, the same sort of activity as is normally involved in cognition and thinking, although one is unaware of it.

Although Helmholtz was (unfortunately) rather vague about unconscious inferences, his studies of early vision are still remarkably fresh and insightful. To illustrate, consider his study of the transduction properties of the eye. Perhaps inspired by his work in physics, he countered a rather widespread belief that the eye was a "perfect" optical instrument by actually measuring its optical properties. He observed, as is commonly known today, that the eye is far from perfect (20). It exhibits the many different forms of aberration and distortion to which all physically realized systems are susceptible.

The result of such optical imperfections in the eye is that images do not fall on the retina in perfect focus but are blurred regardless of how well the lens is functioning. Helmholtz looked for perceptual consequences of such blurring and found many, one of which he believed to be the Mueller–Lyer illusion (see Fig. 1). His reasoning was as follows. On a figure such as the Mueller–Lyer, the areas between the lines forming the acute angles will be blurred more than the areas within the obtuse ones, thereby stretching the lines into the acute angles more than the obtuse ones. Such a distortion is precisely in the direction of the illusion and was, for Helmholtz, its causal explanation.

Such is visual theorizing of the best sort. A task is posed (what are the optical properties of the eye?) and solved in a theoretical fashion that is consistent with empirical data (the spherical abberation was actually measured). Finally, the theory was applied to explain observed phenomena (such as the Mueller–Lyer illusion).

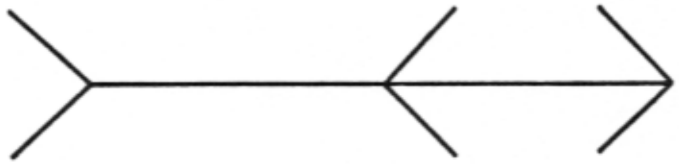

Figure 1. Mueller–Lyer illusion. Although the arrowheads delimit two line segments of equal length, the one enclosed in the convex region appears shorter than the one in the concave region. Helmholtz believed that this is because people's visual systems somehow fill in" the convex portion more than the concave portion, thereby effecting the length judgments. Gregory (12), on the other hand, believed that interpretations of these line segments as projections of 3-D structures lies at the foundation of the illusion.Whatever the mechanism responsible, however, such illusions indicate that everyone's perception of structure in the world is not veridical but rather depends on contextual influences from many possible sources.

Helmholtz was correct in observing that the eye is an imperfect optical instrument. But he was mostly wrong in that his explanation of the Mueller–Lyer illusion cannot account for the entire effect. This has been determined only recently using an elaborate optical technique to project a highly focused image onto the retina (21). Such techniques indicate that optical blurring can account for at most 15% of the illusory effect. Nevertheless, considerations of the physics inherent in the imaging process will emerge in different ways as a major theme in computational vision.

Roberts's System: Segmentation and Matching. Inherent in Helmholtz's theory is a distinction between the types of processing that take place early on and then later in the visual process. Such a distinction lies at the basis of modern computational theories as well, beginning with the first real attempt to design a full system. In a seminal thesis Roberts (22) described what is probably the first computer-vision system. Although it is not clear that he was directly influenced by Helmholtz, he, too, decomposed processing into a low level stage, in which a line drawing was abstracted out of an image, and a high level stage, in which the line drawing was matched against a universe of prototypes. Thus the physically motivated processing was concentrated on the extraction of the line drawing, and "unconscious inferences" were used to match it into a database of objects.

In order to effect this matching, it was necessary for Roberts to restrict the possible universe of objects that his system could encounter. He worked in a miniworld of polyhedral objects composed entirely of blocks—the so-called blocks world—a class of assumptions that influenced computational vision for more than a decade. Even though the universe of objects was simple, however, it did not follow that the matching would be trivial; in the process various transformation parameters such as complex object decomposition, depth, and rotation had to be computed.

The early portion of Roberts's system was concerned with the problem of edge detection (qv), or the identification of those positions in images that indicate interesting physical events. The locus of these positions then comprised a line drawing. The motivation behind "line drawings" can be seen intuitively in cartoons, or drawings which are, in some sense, equivalent to full images. That is, both convey sufficient information to satisfy one's high-level, inferential processes. More specifically, the line drawing was taken to represent a segmentation of the image into meaningful pieces, each of which was taken to be the projection of a meaningful portion of a physical object. The outlines of these pieces then comprised the line drawing.

Roberts's approach to edge detection was based on the observation that distinct physical events (say, the sides of a cube) give rise to distinct image events (intensities). And the image locations at which these events meet have special significance: they are the edges of the cube. Thus, if the points of intensity change could be located and joined, the result would be a perfect line drawing of a projected cube.

To locate the edges of the cube, observe further that intensity changes rapidly there. Calculus tells one that, for smooth functions, rapid changes in the value of the function at a point are accompanied by large values in the derivative of the function at that point. Hence Roberts derived a discrete approximation to a gradient operator elegantly simple in computational form. His plan was to convolve this operator against the image, and then to select the strongest of these convolution values by thresholding. Finally, a linking process would select high gradient "edge" points to be fit by straight lines. Note how a distinct higher level constraint enters Roberts's formulation at this point: the straight sides of his scene polyhedra project into straight "edges" in the image. Helmholtz made similar observations about straight and converging lines. There is further discussion about the influence of such "high-level" or domain-specific knowledge (qv) on early processing later in the entry.

However, as described below, there is a lot more to low-level, early processing than was thought at this time. Roberts was never actually able to get a perfect line drawing from this early processor, and there is a basic sense in which the perfect line drawing is still elusive. But this is a fascinating story in itself since what began as the pursuit of the perfect line drawing has evolved into the full study of early vision. Again the entry starts historically, this time with Helmholtz's contemporary E. Mach.

Similarities versus Differences. Before beginning the discussion of edge detection, however, consider a tangential point. Segmentation (qv) can be approached in two different ways: either by searching for differences, or points along segmentation boundaries or by searching for similarities, or points, within segmentation regions. Edge detection is the approach for determining which points are different, and it is the approach adopted by Roberts. Region growing (see Region-based segmentation) is another approach to segmentation designed for determining which points are the same (8). The rationale for region growing was that, since differentiation emphasizes noise, segmentations might be found more reliably by smoothing "within" regions rather than differentiating between them.

However, it should be stressed that the duality of edge detection and region growing does not imply that one or the other is unnecessary. Rather, they are complementary and almost always work together. Consider Roberts's system again; note that after edge detection (differentiation), similar points (i.e., the locations at which the differential convolution survived thresholding) are linked into lines; this linking process is a kind of one-dimensional region-growing process. That is, points are defined to be similar if they are associated with a common straight-line segment in the least-mean-square sense. Thus, following linking, it is as if all of the boundary points had been "smoothed" into a bounding contour since their (prethresholding) individual differences are now gone. The key information provided by the discontinuity measurement—the edge operator—has now been summarized into a more global, abstract form. This complementarity emerges again and again throughout the evolution of early vision; eventually it emerges as grouping. Consider now the edge-detection route, however, since this was by far the most prominent of the two.

Laplacians, Mach, and Edges. The modern study of edges has its roots in the studies of E. Mach. To contrast him with his contemporary Helmholtz, Mach was interested in image sharpening rather than blurring, and in explanations couched in terms of neural networks rather than in physiological optics. The phenomenon of sharpening is known as Mach bands, or the addition of subjective bright and dark lines (bands) on either side of an intensity change (see Fig. 2). Such bands indicate that the "eye" (i.e. the visual system) is sensitive not only to image intensities but also to their (first and second) derivatives.

Mach bands give a clear indication that the subjective impression of brightness and of contrast is highly dependent on spatial context. That is, one's impressions of brightness and of contrast are not isomorphic with the intensity of light impinging on the retinas but rather are derived—or computed—from it.

How can these computations be understood? Mach believed that psychophysical laws, such as the ones underlying brightness and contrast phenomena, had their proper explanation in terms of properties of neural networks, not in terms of pure physics or purely "physical events." The particulars of Mach's explanation were posed mathematically in terms of "a reciprocal interaction of neighboring areas of the retina" (23). He formulated mathematical relationships involving the Laplacian operator, a symmetric second differential of the image intensities (see below). He cited (then) current neuroanatomical data by Ritter (23) that postulated a regular arrangement of cells on the retina and characterized the function of these cells mathematically. And he postulated that the result of the neural interactions between these cells was a "sensation surface" on which the brightness effects were present. Thus, Mach, in discussing such surfaces, was talking directly about representations (read: *re*-presentations); he was concerned with possible constraints from the "wetware."

Lateral Inhibition: From Operators to Cooperative Computation. Although Mach was able to infer the nature of processing taking place immediately after the retina, it was not until a revolutionary innovation in neurophysiology—the development of microelectrodes for single-cell recording—that his inferences could be verified experimentally. This was first done in the eye of the horseshoe crab *limulus* and has led to much more accurate mathematical models. Such models are said to exhibit lateral inhibition, or a regular structure in which the response at a particular retinal point is derived from excit-

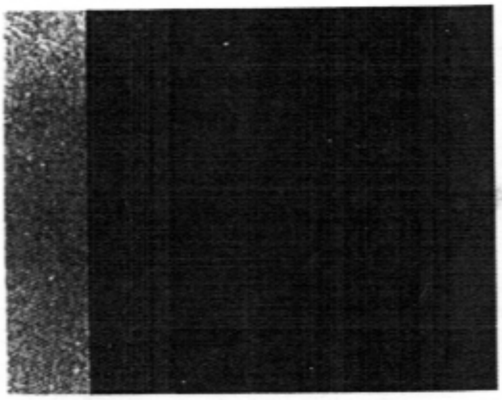

Figure 2. Illustration of Mach bands (11). The image consists of a sequence of rectangular regions of constant intensity, the "edges" between regions are therefore perfect "step edges." However, the intensity does not appear constant to a viewer. On either side of each edge are two Mach bands, a darker one (on the dark side of the edge) and a lighter one on the other side. Again, these bands indicate that what one sees is not "what's out there" but rather is a context-dependent computation driven by it together with additional constraints.

atory contributions at that point together with inhibitory interactions from neighboring points (11,24) (see Fig. 3). Notice, in particular, the regular neural architecture for implementing lateral inhibition, in which the same local structure is repeated across the spatial array. Viewed spatially, the lateral inhibitory structure looks circularly symmetric, with an excitatory central area enclosed within a negative, or inhibitory, surround. Or, in other words, the response at a retinal point is a function of the context around that point.

An essential aspect of this context is the presence of intensity changes in the visual array. As said in the discussion of Roberts's system, such changes are important because they often indicate the presence of physical object contours. In fact, the functional significance of Mach bands has often been attributed to their edge-enhancement effect: if one is to navigate through the physical world on the basis of sensory information, one certainly needs to locate object contours.

Lateral excitatory and inhibitory networks are without doubt one of the most ubiquitous mechanisms in biological vision systems. Lateral inhibitory networks play a clear role in regulating the dynamic range of the eye (25) and otherwise performing a sort of local sharpening, or maxima selection, at the neural level (11). But more generally they have led to a general view of the kind of computational architecture that should be employed in early vision, an architecture of regularly interconnected networks of rather simple processors. But before these networks are developed, consider the local view of lateral-inhibitory-based edge operators.

Discontinuity and Edge Detection. The classical approach to edge detection is differentiation. This is typically accomplished in two stages: the convolution of an operator against the image and some process for interpretation of the operator's responses. Or stated in more general terms, the stages consist of a measurement process followed by a detection process. It is noted above that there is a basic sense in which the two are complementary: if edge detection is a differential process, interpretation must be an integrative one (recall Region growing). Note that there are two convergent approaches to the design of measurement operators, either as numerical approximations to derivatives of different order or as inferences about how primates might do it.

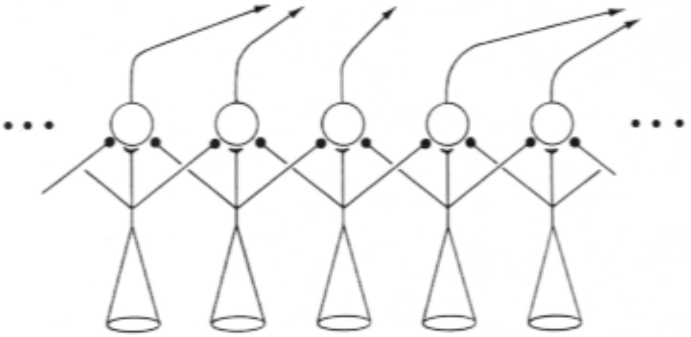

Figure 3. Example of the computational structure of lateral inhibition. In this diagram the cones represent light receptors, and the large circles represent "summation" devices. Note how the level of activity in each receptor contributes positively to the result (excites the summation device), whereas the level of activity in neighboring receptors inhibits it (filled circles). In general, the architecture is one of arrays of units with near-neighbor interactions, both excitatory and inhibitory, with the local computation a simple one.

Edge detection as Differentiation. If the surfaces of physical objects project different image intensities, it would seem that the locations of the physical object "edges" could be inferred from the places where intensity changes rapidly. These rapid intensity changes can be detected, under the assumption that the world projects smooth image-intensity functions to which differential calculus applies, by locating those positions at which the first derivative (spatial gradient) is high; or where the second derivative crosses zero (this assumption, and the approach that it implies, is questioned below). However, there are important numerical issues to be confronted as well, so that the estimates of the derivatives are as accurate as they can be. Trade-offs between these two issues—differentiation and numerical stability—are classical. They led to better approximations to the gradient than the Roberts operator [see the Sobel operator in Duda and Hart (3), as well as Kirsch (26); the numerical issues are discussed in Hildebrandt (27) and implications for edge detection are in many textbooks (2,4,8). Before proceeding, it should be noted that in spite of the predominant identification of edge detection with differentiation, other approaches emerged. Chief among these were a formulation of edge detection as hypothesis testing, so that both the differences in edge profiles and their inherent noisy variation could be taken into account (5,28,29), and an observation that fitting "surfaces" to intensity distributions was a more numerically stable way of finding step discontinuities (30–32). But they still did not work sufficiently well (see Fig. 4).

Shown below is the evidence for the second major influence: early primate vision.

Shape of Visual Receptive Fields. A virtual revolution in the understanding of early visual physiology took place from single-cell recordings in cat and monkey visual systems. The receptive field of a cell indicates how arrangements of light stimuli will effect its activity; in effect, the receptive field characterizes aspects of what the neuron is doing. Hubel and Wiesel (33) discovered a striking arrangement in receptive field structure. Their discovery can be appreciated as follows. Suppose an electrode is indicating the level of activity (firing rate) of a neuron in the visual system. If a spot of light is shone somewhere on the retina, processing may percolate back to influence the firing of that cell. This will be true for some locations in the retinal array, and it may be either excitatory—leading to an increase in the firing rate—or inhibitory, leading to a decrease (below some "resting" or spontaneous level). The shapes of these receptive fields in retinal ganglion cells resemble a circular-surround organization that has been modeled as a difference of two Gaussians (34,35) (see Fig. 5). And they come in two flavors: excitatory center with inhibitory surround, and inhibitory center with excitatory surround.

In the cortex, however, the structure of receptive fields changes dramatically. Here they exhibit the additional property of being orientation selective. That is, individual cells response better to lines and intensity edges than to isolated points. And their response varies as a function of the orientation of the lines and points. The receptive fields are elongated (see Fig. 6). The temptation to identify them with operators for edge and line detection is overwhelming, and this is normally done. However, at this time identification was almost purely local, with little consideration of the network interaction necessarily taking place between these local pieces. The only interactions considered were those required for constructing hierarchies of cells as building blocks to more complex functionality.

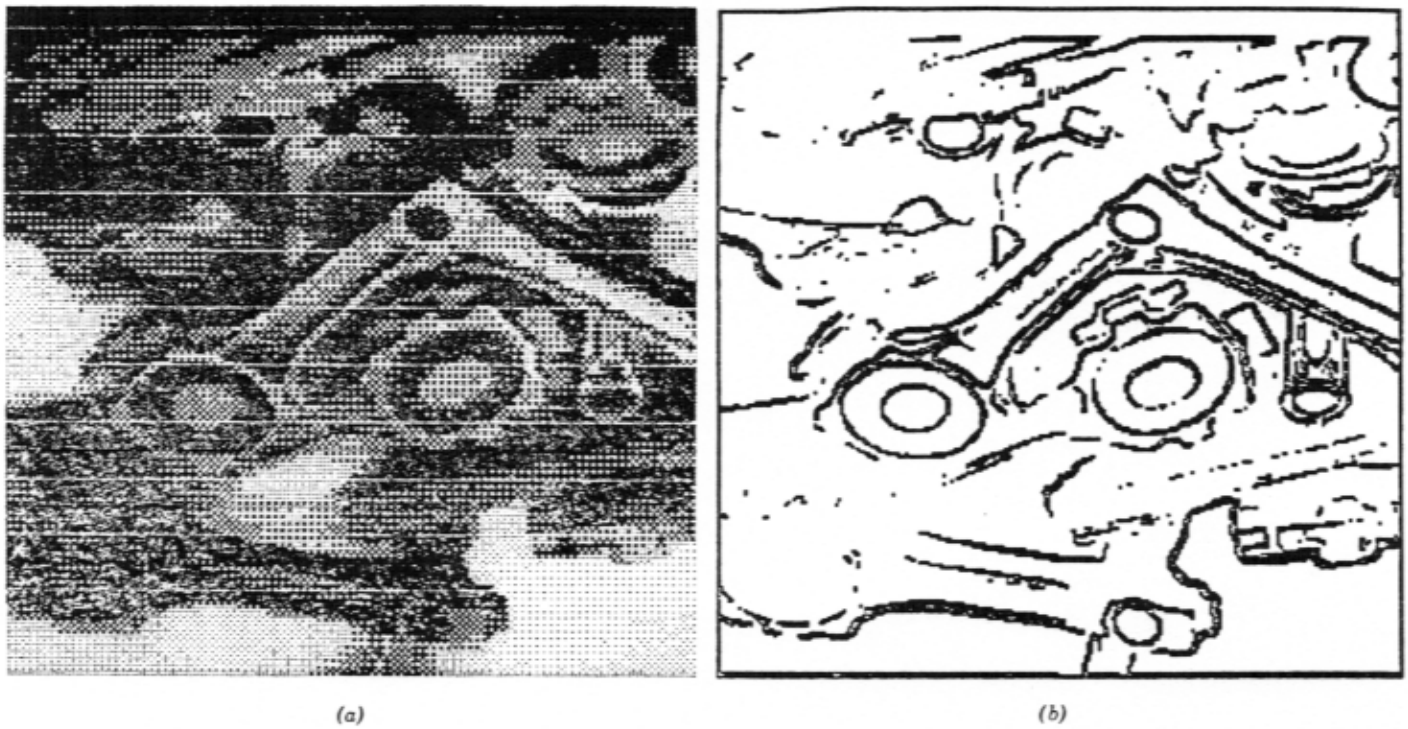

Figure 4. Illustration of the Sobel edge detector. (*a*) Original image of automotive parts (256 × 256) pixel resolution. For display this image has been quantized to six gray levels. (*b*) Binary image indicating the positions at which the maximal Sobel responses were located. Thresholds were specified by a locally adaptive algorithm. Note that although some of the prominent edges have been found, it is certainly not the case that all have been found. Problems clearly arise in image areas that do not contain steplike edge changes.

Edge Detection and Scale. Visual receptive fields span a range of sizes, a point that is loosely consistent with quite a bit of psychophysics regarding threshold perception. Consider, e.g., a display consisting of a sinusoidal grating. The minimal contrast necessary to see the grating is a function of its spatial frequency (11). Wilson and Bergen (36) have empirically determined that these psychophysical data are consistent with four separate channels of processing. Although these channels have not yet been related quantitatively back to the physiology, they seem to indicate (at least abstractly) a number of parallel functional streams. But this point has always been puzzling to computational modelers. If the receptive fields of these cells participate in edge detection, why the variability in scale? What role does scale play in edge detection? Several suggestions have emerged. First, given the noise problems inherent in early vision, from quantization, occlusion, and receptor processes, some sort of averaging would seem necessary to reduce it. For example, if one wished to measure a local feature reliably, say an edge configuration, increasing the size of the operators could lead to increased performance (equivalent detectability with decreasing signal-to-noise ratio) (37). However, there must be more to it than this because larger operators will cover more of the image; hence they may cover more structure than, say, one edge.

In addition to numerical issues, observe that structure in the world arises at different scales as well. Observe, in particular, that certain physical events are highly localized in space (say, the locus of points along which the faces of a cube meet), whereas others are much less localized; they span more space (say, the locus of points defining an animal's limb). These observations indicate the scale at which different physical events are taking place. Land (38) observed, e.g., that changes in physical objects are usually highly localized (say, at the occluding edge between them), whereas changes in lighting are typically much more diffuse. The scale of intensity events thus purportedly "decomposes" lighting from reflectance. Witkin (39) has suggested a scale space (qv) for studying events at these different scales (see also General-Purpose Models Revisited). Could it be that the variation in receptive field size is tuned to events of different "scales" in the world?

The Elusive Edge Operator. Marr and Hildreth (40) tried to link the above notion of scale with specific ideas for edge detection. Selecting from the above facts, they observed that the circular surround operators could be approximated mathemat-

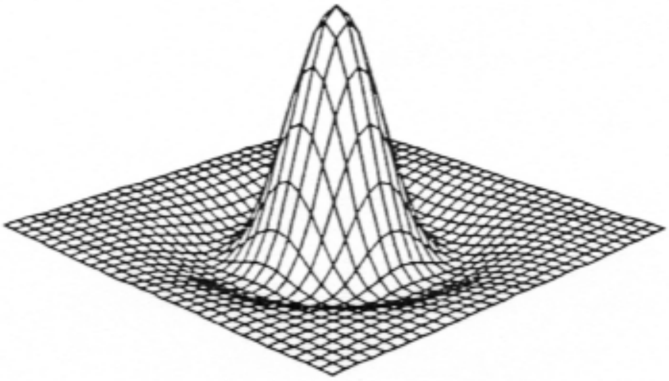

Figure 5. Illustration of Laplacian of a Gaussian edge operator. The illustration also approximates the difference-of-Gaussian receptive fields typical of those found in primate retina.

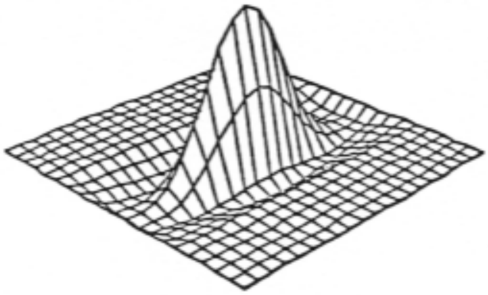

Figure 6. Type of simple cell receptive field of the sort that could be found in early primate visual cortex. Note that it resembles the circular-surround retinal receptive field in Figure 5, although now it is elongated. Such elongation illustrates an orientation preference. In computer vision such operators are known as "line detectors."

ically as $\nabla^2 G$, the Laplacian of a Gaussian; that since the Laplacian is a second-derivative operator (recall Mach (23)), step changes in intensity can be localized by its zero crossings; and that the different sizes of operators would be sensitive to events (i.e., "edges") at different spatial scales (see Edge detection).

Such is a wonderful confluence of events. The problem of edge detection, with which researchers in computational vision have been preoccupied for two decades, could now be solved in a way that is consistent with psychophysical and neurophysiological data and, moreover, provides a functional explanation for it. Unfortunately, however, the scheme cannot work in general for two reasons. First, implicit in it (and in the design of most other edge operators) is an assumption that the intensity structure is a step function across the edge and that the curvature of the edge is zero; i.e., that the edge is straight. It can be shown that these are precisely the conditions under which this operator works successfully but that few edges in natural scenes are of this form (see Fig. 7). Rather, there is a plethora of different physical configurations that can give rise to edges, and these must be taken into account (41) (see also From Structure into Function, below). Moreover, consequent psychophysical predictions from the Marr–Hildreth operator have not been supported (42). The perfect edge operator remains elusive.

Image Representation and Communication. If the circular-surround receptive fields—especially those in the retina and the lateral geniculate nucleus (LGN)—are not involved in edge detection, what other function might they be accomplishing? Since the retina is "an outgrowth of the brain," the problem of communicating image information from the retina to the cortex stably and reliably arises. Laughlin et al. (37) have shown that linear predictive coding theory leads to a model that fits these receptive fields strikingly well (at least for certain animals), and considerations of numerical stability lead to the separation of opposite contrast data (43). It thus would seem that edge detection is likely to begin in the cortex, which opens up the door for much more complex processing. The classical idea of a single edge-detection operator seems unlikely; it remains necessary to discover the mappings between physical scene structure and images and between image structure and visual function.

Naive Physiology: Hierarchies of Feature Detectors. Although lurking in the background, the influence of (then) current notions in physiology on computer vision has always been quite strong. The basic model was feature detection (see Fea-

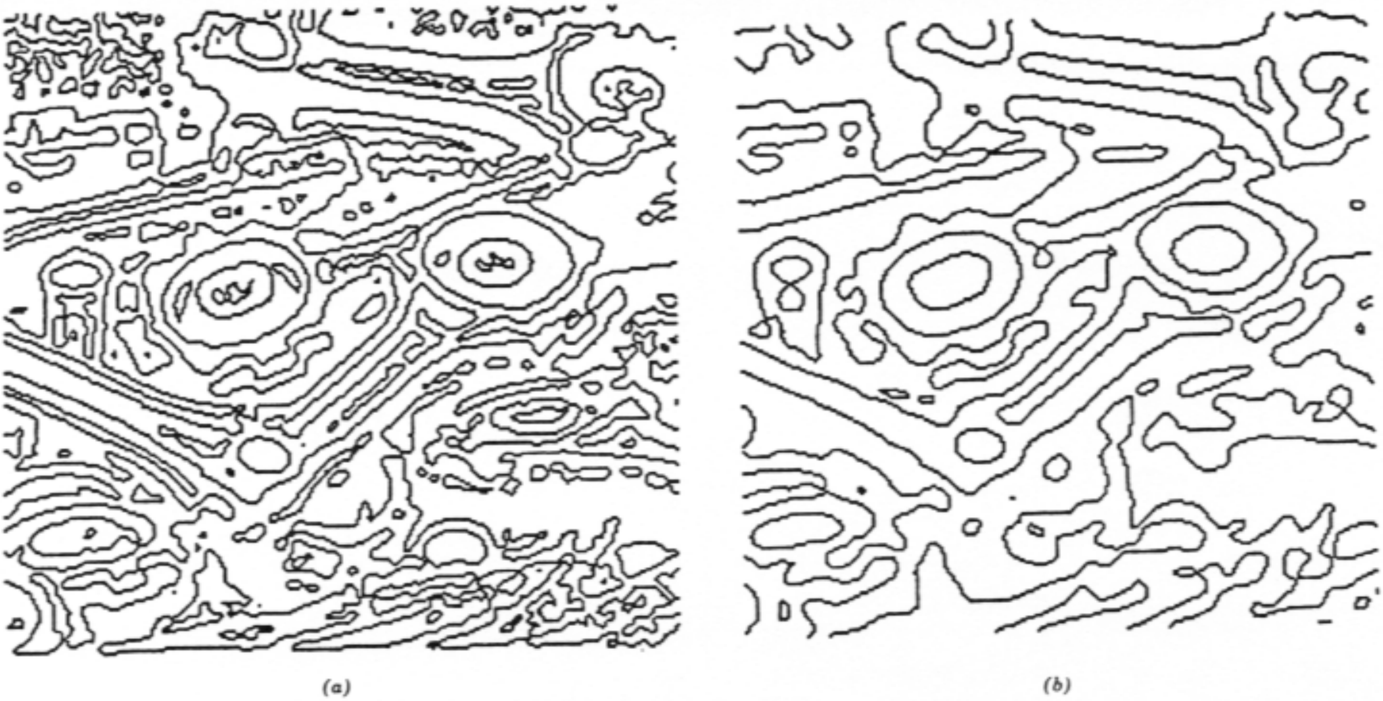

Figure 7. Edge locations (zero crossings) obtained with the operator in Figure 5. (*b*) Compare with Figure 4*b*. Note that the zero crossings form closed contours, although these sometimes have little connection with the physical objects comprising the scene. Two size operators are shown [a small one in (*a*) and a large one in (*b*)] in order to illustrate that the problem is not simply one of "scale." Neither one is completely satisfactory for locating edges.

ture extraction), a two-stage procedure in which operators were first convolved against the image, and then the best match (i.e., the strongest convolution) was selected by a process of thresholding. The operators were somehow matched to the image projections of certain stimuli in the world, following the ideas presented in the classic paper by Lettvin, Maturana, McCullough, and Pitts (44) in which circular-surround receptive fields most sensitive to movement of spots of a particular size and velocity were interpreted as bug detectors. In addition to this general perspective, the model of physiology that was most strongly influencing researchers in computer vision was the hierarchical one put forth by Hubel and Wiesel beginning in 1962 [see the review in the paper by Hubel and Wiesel (33)] in which visual neurons exhibited three types of receptive field structures: simple, complex, and hypercomplex. Simple cells were defined as those in which the subdomains were linear and separable (hence, simple to characterize); complex cells as those in which the subdomains were nonlinear and overlapping (and hence complex to characterize); and hypercomplex cells, which were the most difficult of all to capture. Hubel and Wiesel further hypothesized a hierarchical relationship between cells: retinal ganglion cells fed into the LGN, maintaining the circular-surround receptive fields discussed above. These LGN cells are then combined into elongated simple cells, which are then combined successively into complex and hypercomplex cells. It was loosely asserted that simple cells are involved in edge and curve detection and hypercomplex cells in detection of "higher order" properties such as "corners" or "end points," although even at this time problems were surfacing. How, e.g., could such cells detect dashed curves in noise (45)? More precisely, the receptive fields were modeled as operators, and the question became: how could operator convolutions followed by thresholding detect dashed curves in noise? No clear function was proposed for complex cells, and (as shown below, the wonderful simplicity of hierarchical arrangements of feature detectors gives way to more realistic computations. There is more to "bug" detection in the frog than circular-surround receptive fields, and there is more to "edge" detection in humans than individual simple cells.

Knowledge in the Edge-Detection Process. As the measurements of image intensity changes evolved (recall the "edge-detection" operators), so have the processes for interpreting, or selecting, a "truest" one from among them. Such selection is necessary because the value of the convolution is nonzero almost everywhere due to microstructure and noise. Somehow the significant responses must be separated from the insignificant ones.

Thresholding, Local Maxima Selection, and Hough Transforms. In the simplest case normal thresholding suffices: Simply select the strongest (highest value) convolutions; all others are discounted. The idea behind thresholding is that true edges will project to real intensity differences and hence will lead to high convolution values. This holds for nonoriented edge operators when the decision is, essentially, whether a particular image location is part of an edge. The case of oriented operators is just slightly more complicated since not only the presence of edges must be separated from their absence (the noise responses must be eliminated) but the correct orientation of the edge must be chosen as well. Postulating the orientation at a point to be the same as the orientation of the operator mask with the highest convolution value is, in a sense, the best match, and thresholding is one way to select it.

Thresholding can therefore be interpreted as a selection based on maxima in some first-order statistic, and the idea has evolved to include statistics of more general or less local features. Perhaps the first example in computer vision is the Hough transform (qv) (3), in which long straight lines are found by histogramming local estimates of their orientation and intercept. The Hough transform has been generalized and applied by Ballard (46), Davis (47), and others.

Thresholding can also be viewed as a mapping that takes an image (in which entries are the value of a particular convolution at a point) into another, binary image (in which the locations with values that survived thresholding have value 1 and all others have value 0). The recovery of more global structures thus requires further processing, say tracking along the 1s to find lines and contours.

Vision as Controlled Hallucination. Although the above arguments may seem persuasive at first, unfortunately thresholding's only virtue is its simplicity; it rarely works in practice. How can threshold values be selected? More than two decades of research have shown that, for any but the simplest of images and tasks, the knowledge required cannot be represented as threshold values (8). More sophisticated processing techniques must be employed, and a basic question arises regarding how to structure them. Two schools of thought have been prominent. Originally, it was believed that edge descriptions could somehow be extracted directly out of arbitrary images. This led to elaborate thresholding algorithms, thinning algorithms, etc. (3,8). But it is not at all clear that these mechanisms are sufficiently powerful to utilize the knowledge required for edge detection. It is very difficult to evaluate the results of individual algorithms out of any system's context, and early limitations in these algorithms, or apparent failures in larger contexts, led to the consideration of drastically different techniques. The suggestion that vision is, to a real extent, a process of "controlled hallucination" (48) emerged, the thrust of which implied that hypotheses about (or knowledge of) the object being imaged should directly influence the edge-extraction process. That is, knowledge from the highest, most abstract levels should influence the earliest, most primitive ones (e.g., Ref. 49). Key to this change in view is the observation that not only are intensity edges difficult to locate but many of the edges that define objects also have no image-intensity counterparts. It is possible to see edges where there is no change in intensity (50) (see Fig. 8).

Top-Down versus Bottom-Up. The question of which knowledge should be applied in the edge-detection process is a special case of a more general one: should image analysis be top-down or should it be bottom-up (see Processing, bottom-up and top-down)? Should it be data-driven or hypothesis-driven? Different schools emerged, with Rosenfeld and the image-processing community being associated with the bottom-up idea and much of the more traditional AI community associated with top-down approach. Perhaps the most prominent attempt at top-down edge detection was that of Shirai (51), in which a knowledge-based edge detector was designed to work only for images of cubes.

Flow of Control in Knowledge-Based Edge Finding. Shirai's system worked as follows. Standard image-differential techniques were used to locate a prominent intensity change. Because the universe of possible objects was limited, as in Roberts's case, to the blocks world, it could then be assumed that such a prominent edge point would be part of the bounding contour around the cube. After finding the orientation and

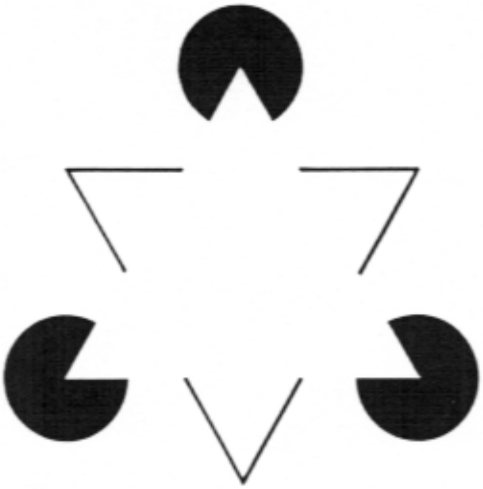

Figure 8. Kanisza subjective edge. Note how the apparent (bright) triangle is indicated by the missing corners in the dark circles and by the line terminations. Edges seem to be present even with no intensity changes.

length of this edge, it was then possible to hypothesize a putative cube model with certain size, etc., parameters instantiated to particular values. This model could then be used to predict the location of other putative edges, which could then be verified by looking at the image intensities in detail.

Complexity of Edge Detection. The use of knowledge all the way down to the lowest levels of the edge-finding process thus becomes quite complex, necessitating elaborate mechanisms for its control. Although this can lead to very high performance levels in restricted universes (such as the blocks microworld), it also leads to brittle, highly specialized systems with little generality. Extending them to ever-so-slightly larger domains became arbitrarily difficult. In terms of object models, there is a formidable gap between the blocks world and the real world. This forced another look at complexity trade-offs and the flow of processing in vision systems. In retrospect it seems that too much was hoped for, with regard to the performance of local edge operators, a lesson that has still not clearly penetrated computational vision.

Introduction of Surface Constraints. Intermediate between knowledge about the exact object and about its edges is knowledge about how they fit together. In the blocks world, knowledge about edges is intimately linked to knowledge about surfaces, and other researchers began to introduce surface-intersection constraints directly into their programs as well. Mackworth (52), e.g., used the idea of gradient space (53–55)., a representation of object-surface-normal properties (not image-intensity gradients) to determine which edge segments could physically belong together for polyhedral objects, i.e., gradient space makes explicit relations between the coordinates of polyhedral surface gradients and lines in an orthographically projected image (see also Refs. 56 and 57). Another intermediate step involved shape estimation, see Refs. 58 and 59.

Rigidity of Early Systems. The problem with these early systems was rigidity; The knowledge on which they were based—e.g., the Shirai constraints or the gradient-space constraints—were "hardwired" into the programs. They could work only for the idealized-object classes within which the constraints held. Generalization was difficult, if not impossible. Historically it was time to back off from the detailed problems and to have a broader look, which is exactly what happened. As shown below, the network parallelism that was so obvious in early visual physiology now begins to play a much more prominent role. It suggests in particular a framework, a point of view toward vision, of how to derive and use general-purpose constraints from abstract assumptions about images and the world, assumptions that are far more realistic than those within the blocks world. The importance of intermediate levels of knowledge, such as that first suggested by gradient space, increases greatly, but its form is drastically different so that its use can become much more fluid and adaptable. For a further, in-depth discussion of the use of high-level knowledge in vision systems, see Tsotsos (60).

Constraints and Assumptions. There is often a minor point of confusion in the study of vision between the terms constraint and assumption. Assumptions about the universe, e.g., that it consists of flat surfaces, allow the derivation of constraints in algorithms, e.g., the equations for fitting planes rather than for fitting fifth-order polynomials. Clearly, the assumption of the blocks world leads to many such constraints, some of which have been described.

Organization and Complexity

Two observations speak against the top-down, rigid analysis of images, and they are both related to complexity in a particular way (61). First, if one sees only what one expects to see, how can the visual system be responsive to unexpected events in the world? In the limit one would not even need to open his/her eyes! Second, on the basis of what trigger features—or database keys—are models selected? Somehow they must be derived from the image, and if the keys are just intensities, there is an immense complexity barrier to be overcome: there are an infinite number of different scenes that could project into any particular image. How can the correct one be selected in a reasonable amount of time?

The antidote to complexity is organization, and the answer to the complexity dilemma is classical. To illustrate, consider the Dewey Decimal Classification (Melvil Dewey, Lake Placid Education Foundation, 1922). If the books in a library were organized randomly, and there were n of them, it would take on average $n/2$ examinations to find any particular one. But if the books were organized, say according to hierarchical categories as in the Dewey decimal system, the savings in search time could be enormous (under certain schemes search time can be shown to grow with $log(n)$ which, for large n is much slower than $n/2$). Analogously, one needs to organize the knowledge in vision systems. The issue is not whether knowledge should be used, but how, what kind, and when it should be used. Intermediate, abstract (with respect to the models) knowledge must somehow be incorporated that captures the regularities of objects in the world. Of course, the earlier observation that physical edges project into image-intensity changes is one kind of knowledge, but this must be further generalized.

Abstraction and General-Purpose Models. Another demonstration can be invoked in support of intermediate levels of organization. Suppose you were looking at a totally unfamiliar scene, e.g., one from a scanning electron microscope or from an ultrasound scanner. Although it is unlikely that one would be able to recognize the object—the scene can even be chosen so that it does not contain any real objects—one will still be able to describe what he/he sees. The description will be in general terms, perhaps involving geometric forms, apparent contours or corners, and will be of the sort that can be derived from any image. Such intermediate descriptions, and the knowledge incorporated into making (and interpreting) them, are what should be sought. What is needed are not assumptions about the entire universe of objects; as in the blocks world, by the time these assumptions give rise to useful constraints, they are too constraining. Rather, more abstract assumptions are needed about the intermediate kinds of structure that can arise in a wide class of natural scenes. These assumptions and the constraints they give rise to will provide the backbone for early processing.

Ubiquity of Uncertainty. Although the metaphor of libraries is instructive regarding the role of organization, it is misleading in its directness. A better example would be a library in which the titles of some of the books were partly obscured. The reason for this was illustrated by the elusive edge operator. The term "operator" as it is usually used in computational vision denotes, in mathematical terms, a linear operator with local spatial support. Although the design of nonlinear operators was also attempted (62), their success was little better. In general, it is impossible to design an operator that responds iff a particular feature is present. Rather, they respond partly to whether a feature is present (the mathematical problem here is related to noninvertibility of operators and L_2 matching theory). Somehow these responses must be interpreted, and it is in these interpretation processes that the general-purpose constraints are embodied. It has been discovered that much of this interpretation can be carried out in mechanisms suitable for parallel implementation, thereby dealing with spatial complexity in a distributed fashion. Before doing so, however, it is essential to stress the change in viewpoint from segmentation to description.

From Segmentation to Description. The corners and occluding boundaries that arise in the blocks world are only a small subset of the diversity of physical events of interest in the natural world. Some of these are primitive events, like the change in orientation at the corner of a cube, and others are compound, like the texture of a forest. Thus, it becomes essential to bring out the many levels of structure if there is to be any hope of eventually agglomerating them. The more explicit they are, the easier (in a computational sense) they will be to use. Borrowing insights from another area that was attempting to grapple with complexity—structured programming—Marr (63) proposed three principles to underlie the development of vision systems. The first of these was a principle of least committment, or the postponing of limiting decisions as long as possible. This is another way to pose the blocks-world criticism. The second principle was one of explicit naming, in which the distinct entities of importance to vision are named so that their description is easily referenced. It, too, deals with complexity. Finally, there is the principle of modularity, the reasons for which were discussed above in the context of organization. In early vision modularity is realized most often by decomposing operations in space as well as into levels.

Rise of Parallelism. In the discussion of lateral inhibition it was pointed out that there were two ways to proceed. The first was toward local (differential or edge) operators, the chosen one. The second direction, or the observation that much of the wetware in early vision appears parallel in structure, is now explored. To understand the advantages of parallel computational machinery, it is helpful to compare it with sequential machinery.

Suppose one wanted to perform a convolution of a particular operator—say, an edge-detection operator—against an image. To obtain numbers, suppose further that it takes η operations to evaluate the convolution at each location. Specifically, suppose the operator consists of a mask whose spatial support is $n \times n$), or n^2, locations. Then, by the convolution formula, there are $\eta = n^2$ (multiplications at each point) + n^2 (additions) to add them up. Then, for a 100×100 image it would take $10^4\eta$ operations to perform the entire convolution of the operator against the image. For biological "machinery" this is quite a long time. But rather than spend this amount of time performing all of the individual convolutions one after another, biology seems to have evolved a faster solution: Perform all of them in parallel. This trades the cost in time for one in space (special-purpose hardware) but provides the answer in the time required for only η operations. In almost all situations this is very fast indeed.

Researchers in computational vision were anxious to understand and, if possible, to capitalize on the parallelism constraint. There were two problems to face: first, what classes of algorithms could be decomposed into parallel ones and, second, how could knowledge be imbedded (represented) within them? It is rather straightforward to show that parallel convolution machines can be built: Take a large number of processors, or "units," and arrange them so that the units are connected to their neighbors. Then weight the interaction connections with the coefficients that define the convolution operator and simply have the units perform the required multiplications and additions. Hierarchies or layered collections of such units could then be conceived, with interconnections between units established both across and between layers. This is the standard view of parallelism, and commercial hardware to accomplish such convolutions is now widely available (64).

Neural Modeling. But the promise of parallel networks is much more than just measurements and convolutions. An abstract characterization of neurons into binary {0, 1} units led Pitts and McCullough (65) to discover relationships between neural networks and a particular logical calculus. Other early researchers attempted to introduce some of the uncertainty apparent in neurons (66) and to deal with fault tolerance (67,68). Perceptrons (qv), or linear-threshold devices were introduced in the 1950s as devices actually capable of substantial pattern recognition, and Hebb (69) suggested, in an exciting book, how neural assemblies could underlie much of behavior and learning. But, unfortunately, most of these earlier claims about perceptrons were based on technical notions that have turned out to be inadequate (49), and although much of the earlier enthusiasm remains, new conceptual approaches became mandatory.

Cooperative Processes and Energy Minimization. Another conceptual approach to parallel processing emerged from two sources, one in psychology and the other in AI. The psychological contribution is considered first.

Stereopsis (see Stereo vision) is the process by which information about the depth of objects in the 3-D scene is extracted from relationships between the two retinal images. By trigonometry, a point in depth will project to different positions on each retina, and this difference in retinal disparity is proportional to depth. Thus, depth information could be inferred if retinal disparities could be computed, but this requires the establishment of correspondence relations between structures in one image and those in the other that derive from the same physical event.

Two insights into what parallel processing could do were provided by two (seemingly) different approaches to the stereo-correspondence problem. The first of these was obtained by Julesz (70), who envisioned the entities in each eye as different kinds of abstract "magnetic dipoles." This then implied that the dipoles could cooperate with one another as they relaxed into an equilibrium configuration. This is, of course, analogous to the Ising model of ferromagnetism (71) in which the local-interaction terms are given by the equations of magnetism restricted to nearest neighbors. Such magnetic-interaction terms model an abstract "affinity" or "compatibility" between local pieces of the retinal image.

The second approach is based on another metaphor from physics. Consider a mountain of sand and a billiard ball, and think of the mountain as representing an "energy landscape." If the ball were rolled down the mountain, intuitively it would seek a minimal-energy position (72). The Gestalt psychologists observed early on that such notions could be applied to model vision. Consider the way that Sperling (73) first applied minimization to the stereo problem. He derived his "energy landscape" from considerations about similarity between retinal images. Correspondence can then be viewed as a process of finding the disparity matches (or correspondences) between images that maximizes a measure of their total similarity.

Since stereo correspondence can be formulated in both ways, it is reasonable to conclude that they are two ways of expressing the same thing. Although it may not be clear what the exact relationship between Julesz's and Sperling's approaches is without writing the equations in full, one can see that Julesz's approach concentrated on local interactions whereas Sperling's approach concentrated on their global "sum." (Some literary freedom is taken here, since Sperling also reduces his model to an interactive network. This paper is especially interesting to read for the early view of connectionism that it proposes.) But an essential ingredient was still missing: How could optimization problems of the sort in which Sperling was interested be solved by the kind of local interactions in which Julesz was interested? The search for the answer goes back to the emergence of parallelism within computer vision. Fischler and Elschlager (74) indicated that the direction should be toward abstract structural matching.

Constraint Satisfaction and Discrete Relaxation Labeling. To return to computational vision, allow the pendulum to swing from technique back to task. Building on the work of Guzman (75), Clowes (48), and Huffman (76), consider an observation made by Waltz (77) about using knowledge in vision systems. Waltz and the others above were interested in the high-level side of a vision system designed to function in the blocks world or in aspects of what might be thought of loosely as the high-level side of Roberts's system. They were concerned, in particular, with what is called line labeling, or assigning semantically meaningful labels to the lines in a line drawing. Consider, for a moment, how such lines could arise physically. The outside edge of a cube occludes the background, and the edges between visible faces represent surface orientation changes. It would seem, then, that at least in the blocks world, it is possible to enumerate all of the ways in which physical edges (or their line representations) could arise. Waltz's task was to assign such representations to the lines in a line drawing so that these could then be synthesized into objects and their interrelationships. The synthesis was to be accomplished by search (qv): try all combinations, say in a depth-first (see Search, depth, first) manner, until a match is found. Unfortunately, the simplicity of the search is overwhelmed by the combinatorics.

Necessity therefore motivated Waltz to shift his attention from the global task (which was now precisely formulated) to its local constituents. Waltz observed in particular that much of the search was unnecessary; it went into examining combinations that were in principle impossible. In fact, many of these impossible combinations could be detected locally, and then pruned, before the full graph was searched. All that was needed were rules, say, about how lines could combine at junctions, to detect many physically impossible situations. Thus, in order to complete his search, Waltz implemented a sequential process that wandered around the graph of combinatorial possibilities, deleting impossible ones. The efficiency was thereby improved to the point that the global searches could be completed after all locally impossible combinations were removed.

The next step was to show that the sequential elimination of labels could be done in parallel (78). This step involved a shift in concentration from the task—labeling line drawings—to the technique. It yielded an algorithm in which, intuitively speaking, each label looked around at all of its neighbors and determined whether it was compatible with each of them; i.e., whether there was (at least) one interpretation (of the possibly many) that could be associated with each line meeting at a junction such that the pair formed a local combination that was realizable in a physical object. If not, the (inconsistent) label as discarded. Of course, this inconsistent label may have been the only one supporting another label on one of its neighboring lines, so that the check for local consistency had to be iterated. In this way information about inconsistencies propagates throughout the entire graph. Such a process was first known as discrete relaxation and has now developed into the study of constraint satisfaction (qv) (see also Waltz filtering).

Continuous Relaxation Labeling. The above algorithms are symbolic in the sense that they deal with explicit symbols (say, occluding edge) associated with explicit image structures (say, a line). The image structures are easily abstracted to a graph, and the problem then becomes one of labeling a graph; or, more precisely, selection from among a set of labels (symbols) associated with nodes in the graph of a particular subset of labels that is consistent according to relations defined over pairs (or triples, etc.) of labels associated with neighboring nodes. The selection need not be done solely on the basis of discrete relations, however; and the labels need not simply be an unordered set. Rather, continuous measures can be distributed over the label sets at each node, and the label-to-label relationships can be continuous rather than all-or-none functions. This essentially establishes a connection back to the subsection above on optimization and replaces the idea of discrete updating with a continuous, analog-type process. That is, the selection can now be done by maximizing a global criterion function of appropriate form or by solving a variational problem with a particular structure (79). Such processes have been called relaxation-labeling processes, in analogy with relaxation techniques for solving systems of differential equations, and their analysis and application in early vision has been widespread (80,81).

Tasks, Tools, and Techniques. At this point in the discussion it is worthwhile to clarify a distinction that is often confused in computational vision: that between tasks—e.g. determining what is an edge—and techniques—how can edges be detected. Marr (1) put it slightly differently: He claimed that one needed to make a distinction between the problem and the algorithm for solving the problem. He further distinguished between the algorithm and the implementation of the algorithm. The discussion in the preceding subsections evolved into one of tools and techniques and can be summarized by general queries of the form: What class of computations can be implemented on parallel, distributed hardware? Such investigations should lead to algorithms and their analysis. Two distinct kinds of analysis are, in fact, necessary. The first kind relates classes of algorithms to abstract characterizations of techniques; for example, there are many algorithms for solving linear-programming problems or for solving optimization problems. The second kind of analysis relates to properties of a particular algorithm (or class of algorithms): will they converge; are they sequential or parallel; are they numerically stable, etc.

Although there is more on Gestalt psychologists later in this entry, it is worth noting at this point that it can be devastating to jump to particular conclusions regarding how algorithms can be implemented. The Gestalt psychologists took the electromagnetic metaphor quite literally, and when electrical potentials were discovered in the brain, they assumed that their minimization metaphor had been substantiated biologically. They thus took it quite literally for many aspects of brain and behavioral function, and the movement suffered substantially as a result (82).

Although Marr advocated treating problem, algorithm, and implementation separately, there are clearly important relationships between them. The study of tasks interfaces with the study of algorithms through the abstract characterization of what they can do. For example, if edge detection could be formulated cleanly as an optimization problem, appropriate optimization algorithms could be chosen. However, if it is not formulated cleanly and does not work, it is impossible to uniquely ascribe blame to either the task specification or to the technique proposed to solve it. Often both are at fault.

In general, it has been the case that there is a sort of pendulum of activity swinging between a concentration on tasks and on tools and techniques. Of course, to actually accomplish anything, one must pay attention to both issues: what is the task and what techniques can be applied to solve it. It is just this duality that keeps the pendulum swinging. Whenever a particular approach fails, either the formulation or the fabrication (the task or the technique) must be blamed, so the researcher then swings over to the other side. Of course, leaving out highly engineered situations, it is almost always the case that both the task and the technique have been inadequately formulated, which is why the pendulum keeps swinging! Therefore, to proceed, allow the pendulum to swing back to the task side.

Vision as "Inverse Optics." Perhaps because they were physicists, both Mach and Helmholtz considered how images were formed, fields known today as photometry, optics, and physiological optics. This perspective has had an immense influence on shaping the second major paradigm for computational vision, or what might be viewed as second-generation—or second-paradigm—vision systems. [The first generation, of course, is typified by the Roberts system (22).] The particular—backward, or inverse—way in which photometry entered is illustrated first; then the second paradigm is developed. For consideration of other optical phenomena in computer vision, see Ref. 54.

Shading from Shape (and Light Source). Light is emitted from a source, reflects off the surfaces of objects, and, if it is not obscured or absorbed by some intermediate object, is captured by the photoreceptors in our eyes. The standard formulation for matte reflection without highlights is well known to physicists, and Mach used it in his research in the latter part of the nineteenth century. The image intensity I (at each image point) is given by

$$I = \rho(\mathbf{N} \cdot \mathbf{L})$$

where $\mathbf{N}$ is the normal vector to the surface at the point, $\mathbf{L}$ is the light-source vector, and ρ is a scalar coefficient of surface reflectance. I is a function of two variables (say, x and y retinotopic or image coordinates), whereas $\mathbf{N}$ and $\mathbf{L}$ are vectors in 3-D space. Clearly, if one knows the scene (or, more particularly for this special case, $\mathbf{N}$, $\mathbf{L}$, and ρ), one can calculate the image. In words, the shading in the image of an object will vary in an appropriate way with the object, the viewing conditions, and the lighting conditions. But vision is concerned with running the above calculations backward.

Shape from Shading. The inversion of the image-formation process is underdetermined. There are always essentially an infinite number of scenes that could have given rise to a particular image. Somehow 3-D variables must be inferred from 2-D ones. Many different sources of constraint are possible, but those that lead to an estimate of where the light source is (with respect to the surface, say), where the surface is and how it is oriented (with respect to the viewer, say), and what the surface's reflectance properties are would seem to be among the most useful. Recently Horn (83), Woodham (55), Pentland (84,84a), and others have tried to recover information about surface shape from changes in illumination (or shading), developing an activity initiated by Mach. Recently, in fact, an entire industry of approaches known as "shape-from-X," where X can be texture, contour, or motion (85,86), has emerged (see Shape analysis).

Beck (87), among others, has studied such phenomena psychophysically. But since image formation is a complex of processes, it follows that segmentation is not just a matter of image differences but can also be a matter of inferred physical object differences (segmentation is the decomposition of the image into pieces that arise as the projection of distinct physical events whose recovery would enable or support object inferences). Waltz's research was an important case in point; see also Mackworth (52). The next decade (1975–1985) of computational vision research was, to a large extent, an attempt to verify this observation; to calculate and to make explicit these intermediate inferences from images back into the scene domain.

Second Paradigm: From Segmentation to Surfaces

Image differences can arise from differences in lighting (say, cast shadows), in surface orientation (as at the corner of a cube), in surface composition (as when one object obscures another), etc. The next major focus of computational vision research was exploring these differences both individually and together. Although edge detection seems impossible solely on the basis of image intensities, each of the above individual properties, if it could be computed, could then be differentiated. The result would be a description not only of where intensities changed but also of which (estimated) scene properties were changing as well.

The key difference between this second paradigm and the first one is how the line is drawn between low- and high-level vision. In the first paradigm the goal of low-level processing was a segmentation, or the recovery of a line drawing of (intensity) outlines. Somehow these outlines are to be matched against a database of prototypical object outlines. Such matching is impossible for natural objects, however, since intensity discontinuities do not always correspond with object structure. For example, intraobject texture differences may obscure interobject edge differences. In the second paradigm the attempt is to recover a richer, more abstract intermediate description, namely, surfaces. The difference between the paradigms is that now, in the second paradigm, segmentation is to be carried out with respect to abstract, inferred properties, not only with respect to image-intensity properties. Then the matching can be guided by (abstract) object as well as a priori properties. And this general-purpose, intermediate level of description will incorporate information from many different sources, including stereo, motion, shape from shading, as well as many different kinds of constraints (such as object smoothness).

The influence from psychology is again clear here. Gibson (88), who, like Helmholtz, had very limited ideas about the need for computation early on in the visual process, nevertheless had a clear idea of the importance of inferring abstract "surface" properties from images. To this end he and his students studied motion, texture, stereopsis, and shading or many of the modes through which information about surfaces could be obtained (to be more precise, the statement here is not that earlier researchers, including Helmholtz, e.g., were unaware of these sources of information but rather that Gibson illustrates the enthusiastic revival of interest in them). Each of these appeared to be such a rich source of information to him that one could almost understand why he thought the visual system could resonate to them. It is striking, in fact, to compare illustrations from his highly influential 1950 book and Marr's (1) much more recent attempt to lay out a computational viewpoint. They are remarkably similar! Curiously, the enthusiasm surrounding Gibsonian "resonances" continues to the present; see the commentaries of Ullman (89).

How, then, is it possible to structure the surface-finding processes, i.e., those processes that actually perform the inference of surfaces? The influence from physiology is again strong, as are inputs from mathematics and computation.

Interacting Modules for Surface Interpolation. Two approaches emerged within this second paradigm almost simultaneously, one based on a working assumption of independence and the other on one of dependence.

Primal Sketches. The first of the frameworks around which second-paradigm computational vision developed was the primal sketch idea of Marr (63). The primal sketch is an explicit representation of the "important information about the two-dimensional image, primarily the intensity changes there and their geometrical organization" (90). The primal sketch is therefore a data structure, and with development it actually evolved into several data structures: the raw primal sketch, which held the results of "edge detection" (qv); the full primal sketch, in which geometric relationships were first made explicit; and the $2\frac{1}{2}$-D sketch, in which certain depth properties were first computed. Model descriptions were then inferred from these data.

Marr took the principle of modularity very seriously and proposed a research program in which each of these different stages, or the processing that actually comprised them, could be studied independently. Separate projects in stereo, motion, and edge detection were therefore begun. Although modularity or independence is arguably just a first approximation, the competing framework stressed interrelationships.

It is interesting to examine the progress from Roberts's line drawings to Marr's primal sketch. Primal sketches are simply an elaboration of the data structure interfaces common to all computer-vision systems. The question was—and is—precisely where they should be placed within the system.

Intrinsic Images. The other dominant framework for second-paradigm vision was founded on the idea that since the surface–light source–viewer arrangements were all intrinsically coded within images, the role of early vision was to explicate them. Hence Barrow and Tenenbaum (91) proposed the idea of intrinsic images, or a collection of arrays aligned in retinotopic (image) coordinates. Each array made one of the local intrinsic properties explicit, including image intensity, surface normal, distance (or disparity), and lighting. Since none of these intrinsic images could be computed by itself, functions were further defined between them which embodied, say, the photometric relationships that must hold between them. Each image further made discontinuities explicit, so that this information could be propagated between images as well (see Fig. 9).

Slicing Up The World For Constraints. The approaches to vision so far have indicated several different ways in which assumptions about the world can be obtained. First, it is possible to have a complete but artificial world. This is the case for the blocks world, from which we can conclude that although some quantitative analyses thereby become possible (say, Horn's image-intensity calculations), the results do not extend to more general universes. A second way to slice up the world is by introducing intermediate constraints that hold for some aspects of the general world. Examples of this are smoothness of surfaces (92,93), the continuity of edges (94,95) and the rigidity of objects (96–98). But beware, when selecting assumptions, that they do not conflict with the quantitative requirement above: There is almost a sensitivity principle operating here in which the consequences of violating an assumption should vary in proportion to the universe over which it holds.

Symbolic Side of Early Vision. In addition to making intrinsic properties explicit, the above framework began to confront the symbolic content of early vision as well. Implicit in the previous generation was an assumption that each stage was interpreted—in the logical sense of the term—by a subsequent one. Now researchers began to worry about the semantics of these symbols; i.e., about what they meant (99,100). Early vision was shown to involve inferences about "what's out there." Such concerns developed from the concern in high-level vision with symbols and appeared explicitly in mecha-

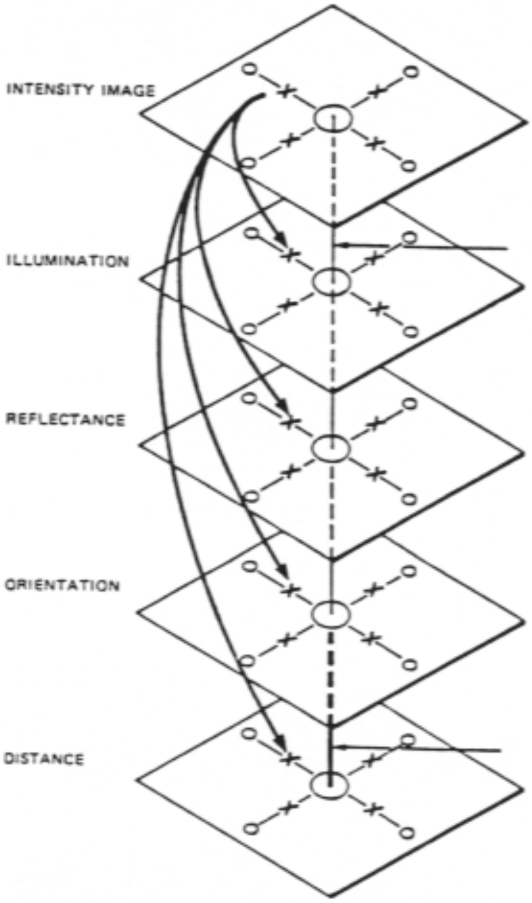

Figure 9. Typical second-paradigm computational model, in this case Barrow and Tenenbaum's (91) intrinsic images. The planes are intended to represent various intrinsic features, such as surface reflectance, arranged retinotopically, and the arrows indicate quantitative constraints between them (equations connecting them). But such purely quantitative, global models now appear to suffer from intractable stability and conditioning problems.

nisms such as relaxation labeling. There is an essential difference between relaxation labeling as it emerged in the middle 1970s and neural modeling, which is much older; in relaxation labeling the labels are symbols. The importance of meaningful symbols, and the inferential processes that manipulated them, increased from this time on in all of vision.

Realization of Second Paradigm. The realization of the second paradigm has been attempted for the past decade, and hence much of the research is covered in detail in more specialized chapters in this *Encyclopedia;* see, in particular, Color vision; Dot-pattern analysis; Motion analysis; Scale-space methods; Template matching; Edge detection; Generalized cylinder representation; Image analysis; Optical flow; Scene analysis; Shape analysis; Stereo vision; Texture analysis; Visual depth map.

Organization of Primate Visual Systems. Decomposition of function arose in physiology as well as in computer vision. One certainly has the feeling that such anatomically apparent modularity lies behind the strong form of modularity advocated by Marr (1). Consider the macroscopic organization of the primate visual system. The first level of organization is rather coarse and can be characterized in terms of anatomically distinct areas to which the visual input is mapped. Recognizing that perhaps a few new visual areas will still be discovered, there are something in the range of 15–20 that have been at least partly characterized (101). It is tempting to relate a visual function to each of these areas, say, with one solving stereo, another solving edge detection, etc. It is also the case that different neural layers within each of these (anatomical) areas may be involved in different functions, which adds up to another factor of 3 or 4. However, this relating is dangerous since each visual area may well be involved in more than one function; recall, e.g., that simple cells are both orientation- and velocity-tuned. An even more telling example has been found by Regan and Cynader (102) in cells whose receptive fields are both motion- and disparity-sensitive. That is, receptive fields exist whose stimulus dependencies involve both velocity and disparity simultaneously; they cannot be decomposed. Even keeping such compositions in mind, however, a first-order estimate of functional complexity is possible. The number of distinct visual areas is significantly larger than 2 (for separating low- and high-level vision) and less than, say, 100, even taking the different layers into account.

The detailed organization of each of these visual areas exhibits structure of its own. The simple, complex, and hypercomplex cells of Hubel and Wiesel (33) are found in the first few cortical visual areas. Fascinatingly, at least in the first of these visual areas (V1 in primates, area 17 in cats) these cells are not arranged randomly but rather form striking columnar patterns according to a number of criteria. First, there are ocular-dominance columns in which cells take their dominant input from the left or the right visual field. Then there are orientation columns in which cells have a preferred orientation that changes sequentially along tangential (electrode) penetration. Finally, the receptive fields vary in size over several orders of magnitude, with some less than 10′ (minutes of visual angle) and others more than 10°. In general, complex cells have larger receptive fields than simple cells. It would definitely seem that this organization provides strong support for parallel processing across spatial (and other) dimensions, although the mapping from anatomical area to visual function is still very obscure.

Parallel functional streams. Although only a few of the properties of cortical receptive fields have been considered, the story is undoubtedly more complex than this. Returning to the retina, the original view that ganglion cells only exhibit spatial circular-surround receptive fields has been replaced with a more modern one in which their temporal characteristics matter as well. Even in the classical view there is evidence of two parallel-processing streams with respect to contrast: one which is ON center/OFF surround, and the other which is OFF center/ON surround. It is now widely believed that there are many different classes of retinal ganglion cells, X and Y being among the most prominent. X cells are roughly linear in their response and have relatively slow connections back to cortex. Y cells, on the other hand, are nonlinear in their response, have fast connections, and have receptive fields roughly twice the size of X cells. Both scale in size with retinal eccentricity. Also, there is a third class of so-called W cells that is beyond the scope of this entry.

Although functional ideas regarding the X and Y systems are still hypothetical, it is held that the X system appears to be more concerned with spatial vision, whereas the Y system may be involved in motion (103). However, the significant point is that the X and Y pathways proceed in separate, parallel channels from the retina through the LGN to the cortex. There is even evidence that they give rise to separate populations of simple and complex cells as well. Thus, in addition to the hierarchical, layered organization (see Naive Physiology (above)), visual processing would appear to be accomplishing several functions in parallel as well. Furthermore, physiological evidence points to separate enervation to (at least some) simple and complex cells, which brings the strict hierarchy into question. And more recently the processing of color has been relegated to a (partially) separate system (104). The feeding forward and the feeding back of visual information does not follow the simple branches of a tree but rather flows around a graph. Complexity rises again.

What kinds of functions could the neurons along all of these parallel, layered pathways be computing? Although the separation of color and luminance information makes evolutionary sense, and the separation of contrast makes mathematical sense, how can other visual functions be decomposed into parallel streams? It is fascinating that this is precisely the question that began to dominate researchers in computational vision, for different reasons, during this second paradigm. Decomposition and organization are the only ways to deal with complexity.

Summary of the Second Paradigm. In summary of this section, the second major paradigm in computational vision, it is interesting to stress that the partition of early and later processing, which began with Helmholtz (19) and Roberts (22), has continued. The primal sketch and intrinsic images are far more complex interfaces than Roberts's line drawing but nevertheless serve the same putative function. They certainly contain much more information, which underlines the complexity of early vision, a point that has emerged many times. But the metaphor of photometry on which this style of analysis is based has become questionable. Just how far quantitative methods can be pushed remains an open question. The missing ingredient is qualitative structure.

General-Purpose Models Revisited

Although the modular view has many attractions (recall the argument above regarding the Dewey decimal classification), the formulation it assumed within the second paradigm has serious problems. A decade of concentrated effort has not led to generally functional systems; rather, there have been only "local" successes, or successes only under many explicit and highly restrictive assumptions. The situation is reminiscent of the blocks-world experience, in which knowledge of (or assumptions about) the physical situation were so restrictive as to be suffocating.

The problem with second-paradigm vision systems can be seen from the discussion of "inverse optics"; for the program to succeed, all of the details of surface normals, surface-reflectance functions, spatial arrangements, etc., need to be recovered exactly. There is no place for approximation, noise, or uncertainty. But approximation, noise, and uncertainty are present in any real physical vision system. The strong form of the second paradigm cannot succeed.

The examples from perception presented above also argue strongly against the second paradigm. The Mueller–Lyer illusion (Fig. 1) and Mach bands (Fig. 2) indicated that what is seen is not exactly what is there; rather, it is the result of a context-dependent computation. And the subjective Kanizsa edge (50) (Fig. 7) suggests that the computation involves a form of inference.

If the exact structure of the scene cannot be recovered everywhere, which parts can be? Which aspects of structure are necessary (sufficient?) for visual inferences? Thus, the question of what kinds of knowledge are employed, and to what ends, rises again. There are two major issues in particular: whether the knowledge is truly quantitative and how to slice up the world so that appropriate general assumptions can be made.

Qualitative versus Quantitative Knowledge. The fact that image-intensity formation is a complex process led to an early observation that intensity profiles are not always shaped like a step function but also arise as roofs or slopes (28). The structure of these intensity functions carries important information about the physical scene that gave rise to it (53), an observation that was instrumental, in large part, for the concentrated effort on "inverse optics." But much of this research was quantitative in motivation (if not in fact), with the goal of precisely formulating the systems of differential equations and solving them exactly. Unfortunately, as Mach indicated, this is impossible in general, and those restrictive situations in which it is may not be all that relevant to the solution of the general vision problem. Although photometers must be quantitative in their interpretation of, say, shading profiles on the moon as seen through telescopes, they are functioning as physicists and their goal is to recover the surface topography of the moon as accurately as possible from that source of information. It would appear that vision systems cannot accomplish this in general for all scenes.

Qualitative Shape from Shading. Because of its central position, the methodological presupposition within the second paradigm that one's visual system functions as an inverse photometer must be questioned. The evidence, in fact, is against it from a biological perspective. For example, if computer-graphics techniques were used to render images of cylinders with different cross-sectional profiles, ranging, say, from circular to triangular, one would expect the percept of the figure to change as well. This does not happen, however (105); rather, a range of different surface shapes are all perceived as roughly circular. The impression is that shading cues are more qualitative than quantitative. Cavanagh (106), in a recent psychophysical study, has shown that depth cues such as occlusion and shading interact only in certain ways, depending on whether they derive from color, motion, stereo, etc.

The next example raises another kind of problem with strict quantitative constraints in early vision.

Rigidity of Rigidity Assumption in Structure from Motion. Humans have the remarkable ability to see structure from moving-dot patterns (96,97). There is, of course, much more to motion than is considered in this entry; consult the other articles in this *Encyclopedia* for a more complete treatment. Imagine, e.g., a clear cylinder covered with tiny specks of dust. Statically, the dust pattern would appear to just be random dots arranged on a plane. If the cylinder were rotated, however, its full 3-D structure would be apparent. Ullman (98) has referred to this ability as inferring structure from motion, and a good deal of research, both computational and psychophysical, has been focused on it (96,97,107, and 108). In particular, the apparent perceptual preference to see rigid configurations has inspired researchers to build the rigidity constraint directly into the computation, leading to models in which systems of algebraic equations need to be solved exactly.

However, the computations arising from such formulations are difficult, if not impossible, to realize biologically or perceptually because of a numerical problem called ill-conditioning. A system of equations is ill-conditioned if a small change in input (independent variable) leads to an arbitrary change in output (dependent variable). Such small changes in input arise from errors in measurement or lack of high numerical precision, and hence such quantitatively sensitive algorithms seem ideally unsuited for real vision. Visual systems must be designed to function in the presence of noise and uncertainty. Global rigidity assumptions are too rigid; they fracture under the slightest pressure from uncertainty.

General Algorithms and General Position. The above point about rigidity can be made in different terms by the notion of general position. Imagine viewing a line drawing of a cube. From all positions except one this line drawing will be topologically similar; there is one, however, at which it looks different: when the corner nearest the viewer and the corner farthest from the viewer align along the viewing axis. In this singular configuration that line drawing no longer looks 3-D; rather, it looks like a flat triangulated hexagon (see Fig. 10). Note, however, that such singular configurations are destroyed if one moves ever so slightly off the axis connecting the two corners. In all but this one configuration the cube is viewed from a general position. By definition, then, a general-positional view of an object is one in which the image does not change qualitatively. Other examples of singular arrangements could be obtained if one were to view scenes from singular viewpoints; recall the Ames room demonstrations (109). They did not look the same from any other viewpoint. Vision algorithms must be general in precisely this sense; the rigidity assumption is analogous to a singular view.

From Structure to Function

In the beginning of the entry, two pendula are described, one of which indicates the tension between low and high-level knowledge and the other between task and technique. Although the field has become much more experienced, the pendula still remain. Ten years ago there was great concern about whether processing was top-down or bottom-up; now it is what kinds of knowledge must be applied; what are legitimate assumptions to make and what constraints do they imply. However, the assumptions and constraints are usually ones of principle; e.g., the assumption of rigidity in motion (97,98). How should they be made discrete in practice? The need for modularity has arisen within both tensions; the question has become which structures in the scene project into which structures in the image (informational modularity), and how might these be recovered reliably and consistently (processing modularity)? Spatial parallelism is clearly indicated in the early stages; how high it goes is still an open question (110,111).

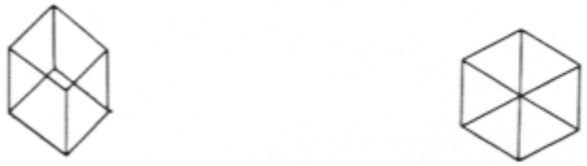

Figure 10. Two line drawings of cubes. The one on the left illustrates a cube from an arbitrary viewpoint; although the details change, the topological features remain the same. On the right is an illustration from a singular viewpoint, in which the three-dimensionality may or may not be present. Any slight change in viewpoint destroys the triangulated hexagon and results in a drawing like the one on the left.

The idea here is more than that certain scene structures are preserved under projection. Helmholtz (19) and Gibson (88) were both aware of this (the classic examples from Helmholtz concern how straight lines in space project into straight lines in the image; Gibson, of course, discovered texture gradients and optical flow). The idea is that vision consists of a fabric of inferences, some or many of which are interrelated. Not all of the scene information is recoverable directly; rather, only some kinds of structure in the scene domain give rise to image structures from which the "projection" transformation can be "inverted." It is this network of local (in every sense of the term) inferences on which the global scene recovery is anchored. Somehow the global structure must be interpolated from the local anchors. It is the discovery of such structures that provides the keys into visual function.

On the Mechanisms For Early Inferences. Three major lessons have been learned about how to approach early vision: the mechanisms for structuring the early inferences must incorporate assumptions that hold abstractly over significant subclasses of real-world structure; they must be as insensitive to noise and uncertainty as possible; and they must be utilizable within distributed "hardware" (either networks of neurons or VLSI circuits) to deal with the complexity of size. It has been slowly realized that such inferences can be carried out by distributed optimization, relaxation, or related machinery. The constraints are more satisfiable, when posed within minimization- or hypothesis-testing frameworks than as rigid systems of equations (112–115). And since constraints on these inferences exist from many different sources, the mechanisms within which they are utilized must permit their interaction as well.

Local Structure of Intensity Discontinuities. Both issues—how qualitative and how encompassing assumptions are—emerge in the search for the elusive edge-detection process (recall The Elusive Edge Operator). Concentrate on the first lesson listed above: that there should be a significant but "local" problem, i.e., a significant amount of knowledge about the local structure of images that provides a solid foundation for inferring a local piece of the scene that gave rise to it. Since the range of physical events is large that can project into what one should like to call edges in the image, it follows that the description of the edge must be rich enough to support decisions about the physical cause (edge events can arise from surface-reflectance changes, surface-orientation changes, occlusions, and lighting changes taken individually or in combination). Simple image differentials are unlikely to work, since they only localize changes in intensity (e.g., Ref. 116). More complex processing is needed for determining not only the locations of the image-intensity discontinuities but also the shape of the intensity function in a local neighborhood around it.

The importance of the local structure of image intensities in the neighborhood of an edge has been realized since the blocks world. Binford and Horn (29) matched intensity profiles along an edge; however, their matching relied heavily on assumptions about the blocks world (e.g., planarity of surfaces). More recently Quam (117) used the intensity structure across linear features to track roads, and Witkin (118) correlated intensity structure across putative edges to verify their presence. Most important, however, Witkin tried to use these across-edge correlations to infer something about the physical event that gave rise to the edge, observing that when objects stand in an occlusion relationship, the pattern of intensities on either side of the occlusion should be similar, but not necessarily across the occlusion. Such information clearly supports the mapping from the image back to the scene under certain circumstances.

One can go further. Not only should intensity structures be similar on either side of an edge, but their detailed structure can be used to support inferences about the edge (41). To illustrate, recall the above discussion of shape from shading. Consider a matte surface smoothly varying in orientation. By the image-irradiance equation it follows that the projected intensity function will be smooth as well. Now, suppose that a shadow were cast on the surface. This would add a step to the (log) intensity distribution, resulting in an image structure in which the slope of the intensity function would be the same (and almost certainly not flat!) on either side of the discontinuity. Therefore, to locate this type of edge, and to distinguish it from others, it is necessary to describe both the intensity function and its slope on either side of a discontinuity. Other schemes that simply locate discontinuities at the cost of irretrievably modifying the local intensities cannot support inferences back to the world.

Leclerc and Zucker (41) have developed a nonlinear scheme for accomplishing this estimation of local intensity structure concurrently with detection of discontinuities. It involves spatial interactions between convolutions of similar sizes as well as "level" or scale interactions between different sizes. Such intra- and interscale interactions are necessary because of the trade-off between neighborhood size and noise immunity: the larger the neighborhood, the better the performance at detecting and describing an edge. But the performance will degrade if more local edge events are smoothed over in the process. The network interactions are necessary to guarantee that this does not take place.

Edge detection is therefore not just a matter of finding the right edge operator but rather requires understanding the interactions between measurements as well. This leads to inferences. The structure across edges has been stressed—the local structure of intensities in the neighborhood of discontinuities; now a different aspect of the structure is considered—that along curves rather than across edges. This is dealt with by differential geometry.

Analysis of Orientation: Curve and flow recovery. As another example, consider the (related) problem of curve detection. Curves are an important subclass of structure in the world because they satisfy the above criteria: They occur in almost all images (as occluding contours, surface creases, hair and other surface coverings, etc.), and they provide information on which subsequent surface inferences can be based (115, 119). How can they be recovered, or inferred, from images?

Early approaches to curve detection were barely distinguishable from early edge detection. Simple cell operators were convolved against images and maximal values selected by thresholding. This is because they respond maximally when centered exactly on a line and oriented similarly to it; hence they have been called line detectors. Unfortunately, they run into the same problems as "edge" detectors.

Lines versus Tangents. What relationship do so-called line detectors have to the actual detection of a curve? A logical place to start is by asking what a curve really is: Mathematically, a curve is a function that maps an interval of the line $\mathfrak{R}^1$ into an embedding space (say, the plane $\mathfrak{R}^2$ or space $\mathfrak{R}^3$). What is given in an image is not a curve, but rather a discrete sampling of the trace of a curve, or a set of points through which it passes. Curve detection is the process of inferring a curve from its trace subject to additional constraints.

It is important to view curves abstractly and mathematically rather than pragmatically as image structures because it suggests that the intermediate structures should be abstract as well; in this case lines are not what need be detected but rather tangents to curves. Curvature can then also be shown to play a role, so that tangent fields can be recovered by minimizing a functional of curvature variation (115). Since the tangent is the first derivative of the curve with respect to arc length, the global curve can be easily recovered from such local representations of it by a process of integration. Tangent fields are an arrangement of discretized and quantized tangents in retinotopic coordinates.

If the problem were viewed as one of line detection rather than tangent detection, subsequent stages (smooth global curve inferences, placement of corners and discontinuities, etc.) would be difficult to formulate in a mathematically consistent manner.

Viewing curves abstractly also raises additional possibilities for interpreting scale as well. Recall the traditional view that larger operators detect larger events (recall Edge Detection and Scale). However, since curvature is a relationship over neighboring tangents along a curve, it suggests that perhaps the role of the larger operators is related to these higher order properties (115). One should certainly expect this to be the case mathematically, and computationally it works as well. Indications are that the biological equivalent to curvature measurements are embodied in the hypercomplex cells (see Naive Physiology, above) or, more precisely, in their "end-stopping" property. This is a case in which the explanation of biological data together with basic mathematics has yielded a successful computational-vision algorithm (115); but see also Ref. 120.

Parallel Surface Contours. Consider a surface covered in parallel pinstripes. These will project into "parallel" curves in the image. Stevens (119) studied the inverse situation, or the inference of a surface from a collection of "parallel" curves. Such displays give strong impressions of surfaces when viewed (see Fig. 11).

Flow Patterns. It is rarely the case that the surfaces of objects in natural scenes are covered with regularly arranged contours, however. The more natural case is that the curves are arranged so densely that they cover the surface in a physical sense, often winding in and out of occlusion relationships. This is the case for hair and fur patterns consisting of only roughly parallel arrangements of curves (hairs), or in the case of motion, such flow patterns arise in waterfalls, or when the projected image of a complex physical arrangement changes rapidly, as when one runs through a forest. Note that hairs are different from the pinstripe patterns discussed above in that pinstripes almost never touch and are continued for a long distance, whereas hairs almost always touch and are rarely

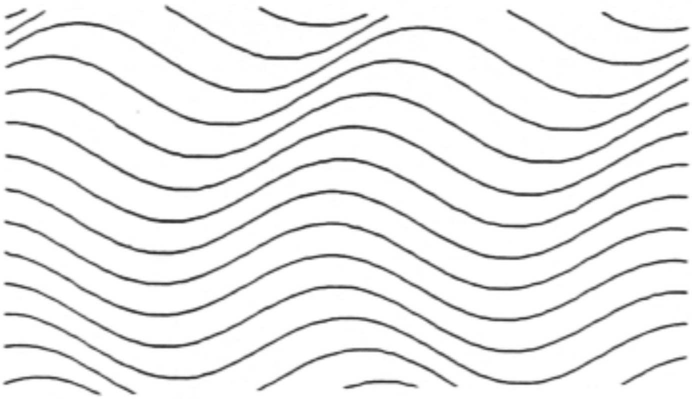

Figure 11. Illustration of how parallel surface contours (pinstripes) provide a strong cue for surface orientation.

visible for long distances. Somehow the surface covering (the hair pattern) must be recovered from the (relatively) sparse information available in the image as highlights.

The issue here is another abstract mathematical one, and the way in which it connects image events (patterns of intensities) to events in the world (patterns of hairs). Topologically, curves are 1-D constructs; surface coverings are, like surfaces, 2-D constructs. One should therefore expect the recovery of surface-covering descriptions to require processing in addition to that required for curve inferencing. In particular, the recovery of orientation information for flow patterns involves a direct form of interpolation, or the spreading of curve information, to "fill in" the areas between "highlights" of projected image structure (see Fig. 12). The result is a tangent field, or description of the orientation information, over an entire 2-D.

Texture. There are other regularities in addition to those in image structure that are connected to orientation. Such arrangements are called textures (88), and their structural geometric regularities have been modeled by Zucker (121). Beck (122) has shown that most of the perceptually relevant structure of (static) textures is contained in intensity, orientation, and color distributions, and Haralick (123) has reviewed the different approaches to computing texture descriptions.

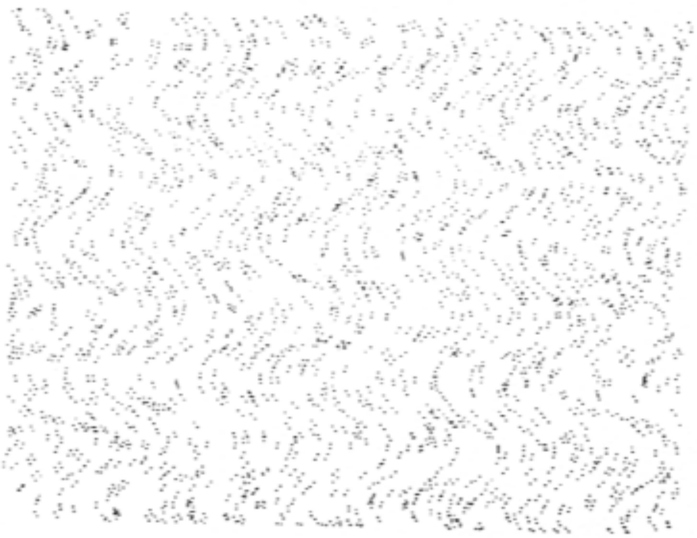

Figure 12. Flow pattern covering à surface with the same kind of undulations as the one in Figure 11. Now, however, the covering is more like a hair covering. Local indication of orientation is provided by pairs of random dots; global 2-D organization is inferred by one's visual system. Note that, unlike Figure 11, there are no long contours and that they are not evenly spaced but rather appear to cover the surface densely.

If textures are that class of image structures that indicates regularity, at least in some sense, the other end of the extreme are those image structures that indicate (and arise from) chaos; consider, e.g., clouds, trees covered with leaves, and the surf. Pentland (84) has studied how such fractals can arise in images and how their structure can be inferred.

Surfaces. Once information about surface coverings, contours, and edges is available, the question arises regarding how to infer surfaces (or other abstract structures) from them. Many other examples of such inferences abound; recall the discussion of shape-from-X methods above (Tasks, Tools, and Techniques).

To illustrate the technique, recall the observation of Helmholtz that straight lines in space project to straight lines in images. Now, if it were known that the lines in space were regularly arranged, say as a square grid or as (in the case of contours) circles, their projection onto a surface would give rise to a regular geometric distortion; the square grid would go into a trapezoidal one, and the circles would go into ellipses. Hence, given these figures, the projection equation could be inverted for them (85,124). Again note that this is not a case of doing inverse optics in general but rather doing it for a local problem (a specific grid). Such ideas underlie much of the 3-D vision in robotics (qv), in which structured light arrays are projected onto objects using either patterned sources or lasers (125).

Kass and Witkin (126) present a more radical attempt at structural inference; they would like to characterize the process by which objects are formed, as, e.g., the flow of tree bark around a knot.

Corners and Parts. At a level of abstraction up from curves and edges, even sparsely supported, are structures that are derivable from them, say their discontinuities, inflexions, and extrema. For a discussion of how to detect corners, see Brady and Asada (120), Davis (58), and Zucker (115). In a collection of parallel contours, e.g., corners that align along smooth contours provide a curve inferencing problem one level higher (than the curve inferencing), and similarly these corner contours could be grouped. The study of such grouping processes is related to the way in which complex objects are decomposed into parts, an area currently under investigation (95,127). Intersections between objects almost always map to singularities in the bounding contour, for example.

Figure/Ground and Structural Grouping. The Gestalt psychologists emphasized the role of structure and organization in early vision nearly 65 years ago (e.g., see Ref. 128). They identified a phenomenological level of processing in which figure was separated from ground according to a number of principles, including proximity, good continuation, common fate, etc. But their proposals have defied quantification until now (16). Perhaps the reason is because of the complexity issues that have been discussed here. There are, e.g., many different ways in which curves can arise in the physical world, as is the case for edges, and their subsequent uses are likely different. Bounding contours arise from inferences that link discontinuities with similar local intensity structures, and surface contours may have similar intensities along them. Discussed above is how curves such as surface contours can be detected, but this requires an assumption that the given trace of the curve is dense enough to properly stimulate the initial convolutions. Curves that are sparse with respect to this metric will certainly require different techniques.

Grouping is a term usually taken to denote the agglomeration of local structures into more global, and perhaps more abstract, wholes. The standard example is Wertheimer's (129) observation that collections of points are not seen as points but rather as figures. As discussed above, the Gestalt psychologists believed that the (physical) minimization of real quantities was the mechanism by which this grouping was accomplished, in much the same way that soap bubbles minimize surface area. Although it is now known that minimization can be accomplished by more computational methods (e.g., see the section Relaxation Labeling), the original idea that grouping could be accomplished by a single mechanism is extant [e.g., Marr (1)]. Given the arguments about the diversity of inference mechanisms for curves and edges, however, this seems unlikely, in much the same way that a single-edge mechanism now seems unlikely. Rather, one might guess that there are a diversity of grouping processes, each tailored to different (classes of) function (115). Curve and edge processes are discussed above; a few other examples follow. This argument has now been generalized by Witkin and Tenenbaum (130), Lowe and Binford (131), and Lowe (132). It holds that, in general, structure is unlikely to arise randomly in images; rather the structure of the world is such that, should events somehow "line up," a common cause should be attributed to them. General-purpose models capture this common cause at early, but still abstract, levels.

Differential Equation Metaphor. The way in which, say, occluding contours and hair patterns relate to the physical world leads to another difference between them, a difference that can be generalized to include many other classes of structural information. Occluding contours arise when surfaces intersect projectively; they give rise to intersurface constraints. Hair patterns, on the other hand, provide information about the particular surface on which they lie; they give rise to intrasurface constraints. This difference is fundamental to the processes that must put various sources of information together. Again one is led to the metaphor provided by differential equations, in which the solution is governed by two distinct classes of constraints: the differential operator, which constrains how the solution varies over its domain, and the boundary condition, which constrains the domain and the value at the edges of this domain. Note that, for 2-D differential equations, such as Laplace's equation, the differential operator is an infinitesimal function of two variables, whereas the boundary condition is given by, say, the values along a 1-D contour. Infinitesimally such differential operators represent the kind of constraint available from flow patterns, whereas the boundary conditions resemble 1-D contours. In terms of the previous examples, the Laplacian corresponds to the orientation information provided by an infinitesimal "piece" of a waterfall, and the boundary condition corresponds to its bounding contour.

To be more specific, recall that Laplace's equation is

$$\nabla^2 u(x, y) = 0,$$

where ∇^2 denotes the Laplacian (differential) operator $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$. If Ω is an open neighborhood, a well-defined problem would be to find u in Ω from prescribed values of $\nabla^2 u$ in Ω and of u on $\partial\Omega$, the boundary of Ω. Such problems are known as Dirichlet problems. Within Ω, u is completely determined by the constraints provided by Laplace's equation and by the value of u along the boundary $\partial\Omega$.

The above example is, of course, metaphorical. This is not to say that intrasurface constraints are Laplacians. Rather, it is the abstract mathematical form that is of concern, and it is not limited to waterfalls. Other sources of static intrasurface constraint come from monocular shape cues, such as shape from shading (150, 151), from binocular stereo disparities (70,93,133), and so on as has been discussed. Intrasurface cues only hold for particular surfaces; they take abrupt jumps as the projected image from one surface undergoes a transition to that from another. Similar arguments hold for the transitions in lighting, which has also been discussed, say from an illuminated area to one in a cast shadow (41). Topologically all of these transitions are 1-D contour boundary conditions that constrain the area over which the other, 2-D intrasurface constraints can be integrated.

Free Boundary-Value Problems. Clearly there are many sources of both inter- and intrasurface constraint, including motion, stereo, shading, texture, color, and their differences. I have already argued that their inference will involve interpolation and that "edges" provide the boundary conditions for limiting them. How can these different sources of information be put together? Clearly situations will arise in which the intrasurface information will be ambiguous, as will the intersurface information, and both inferences must occur simultaneously so that they can mutually constrain one another. Intuitively the situation is like computing the shape of a soap bubble over a flexible ring; clearly, the final shape will depend both on the ring and on the soap. Such problems are called free boundary-value problems, and they arise in many areas of mathematical physics (134).

An early attempt to implement these ideas in a simplified vision context involved dot clusters, in which the problem was to label the dots defining the edge of clusters simultaneously with labeling the dots interior to the clusters (135). Separate processes for intracluster and intercluster (edge) labeling were specified, as were their interactions. Briefly, one might speculate that, in the vision context, the intracluster processes would be integrative, region-growing-type algorithms over intensity (concretely) or shape and reflectance (abstractly) constraints; the intercluster processes would delimit the various type of edges between them. More recent attempts at integrating information from different sources in vision systems are discussed by (136). Much remains to be done along these lines.

Generalization of the Framework. The framework provided by inter- and intrasurface information, or, in different terms, by differential equations, holds not only for the features described here but also for abstractions over them. Contours arising from abrupt changes, in, say, a flow or hair pattern could provide the boundary constraint to a higher level process. This could correspond to a physical situation in which the underlying surface changed orientation abruptly but the surface markings smoothed it over somewhat. Thus, issues of how to differentiate flows become as important as the flows themselves.

High Level Vision

The beginning of this entry differentiates between low-level and high level vision by asserting that low level vision is the study of general constraints on special-purpose hardware whereas high level vision is that of special constraints on general-purpose hardware. In the evolution of this entry many different low level constraints are uncovered with nontrivial roots in higher level vision. The earliest work on computational grouping may be Guzman's (75) "matched *T*s" for joining pieces of contour occluded by a common block together. As these different constraints were refined and "moved down" in visual systems, the need emerged for intermediate-level structures to support them. Perhaps the earliest tenable idea in this direction is Binford's (137) notion of a generalized cylinder (qv); see also Brady and Asada (120). More specialized constraints are bound to emerge here, in the sense that they are more global. Rather than searching for local structures that can be inverted, here one is searching for a vocabulary of intermediate objects; a kind of modeling language for prototypes. The most recent contribution in this direction is a proposal by Pentland (138) for "superquadrics," an extension of quadric surfaces done for solid modeling in computer graphics (139). But it is not yet clear how these superquadric constructs relate to more traditional graphics modeling primitives (140).

The requirements for intermediate structures are influenced both by what is coming from "below" and what remains to follow from "above." Lowe (132) makes the point that the result of grouping operations should provide indices into model databases, and attention (141), may well provide a connection in the opposite direction. Robotics imposes its own special constraints (e.g., see Ref. 142).

High level vision has a very different structure than what has been described throughout most of this entry. Although constraints still play a fundamental role (143), now they relate properties of objects (in object databases) to image structure. "Analog," continuous methods, of the sort that have been described throughout, get replaced by more symbolic programming tools (4,144–146,152); for a review, see Ref. 147. It is these symbolic tools that provide the general "inference engine" for interpreting the specific, high level constraints. And these high level constraints are more symbolic than those in early vision; see, e.g., the complex frames and other data structures described in the references above. The mixture of the two can often provide nice solutions to constrained engineering problems (e.g., see Refs. 148 and 149; see also Image understanding).

Conclusions

Light reflected from physical objects gives rise to images. Vision is the inverse of this process: the recovery of descriptions of objects in the world from images of them. It is clearly an underconstrained problem: somehow a description of 3-D scenes must be recovered from 2-D images. Yet it is possible, as the human visual system demonstrates. But where does the trick lie? How is the structure of the world reflected in the structure of one's visual system? Which aspects of the structure of the world are important, and how are they—should they be—organized? Is it in the gross organization, or in the details of neural interconnections? How can the processing be described so that it could be understood and tested? How does one's internal percepts relate back to the world? And how can principles be uncovered that allow observations about biological perception to be related to machine perception? These are the kinds of questions that computational-vision research would like to be able to answer.

As shown, there are many ways to approach these questions. Psychology, physiology, anatomy, evolutionary biology, mathematics, computer science, engineering, physics, philosophy, and psychoanalysis all have something to contribute. The diversity of these fields gives some indication of the diversity of constraints that are active in vision, and the goal of this entry has been to illustrate how they can work together. A metaphorical example may help, in retrospect, to put the pieces of this entry together.

Suppose that you were an extraterrestrial who happened to land on this planet in the midst of a museum collection of clocks. That some relationship existed between the objects might be inferred from their proximity, but how might you go about discovering it. Physical and anatomical observations would reveal differences in their microstructure, with some objects composed of sand and others of wood or metal. However, the introduction of the abstract concept of energy would greatly unify the investigation and might even point out some macroscopic principles of organization; namely, that clocks contain (or depend on) a source of energy and have internal structure that acts reliably on the external world. For example, the internal gears might move hands or the transistors control LED displays. To take the next step toward understanding how these devices relate to one another requires another abstract concept: the mathematical notion of periodicity and modular arithmetic. Now consider concepts of time and eventually connect them with the many tasks for which timekeeping is important, ranging from agriculture to social customs.

Of course, the above vignette is grossly oversimplified. The difficulties inherent in making the many leaps are immense. But the point is clear: theoretical ideas from many different levels lead to constraints that percolate via reductionism and constructivism to other levels.

Investigations of the problems of vision rarely yield complete theories. Rather, their contribution results in the formulation of constraints for shaping any theory. Such constraints stand no matter whether the parent theoretical framework changes. The evolution of one's understanding of these constraints has been the principal theme running through this entry; this is probably what is considered to be progress in understanding vision.

BIBLIOGRAPHY

1. D. Marr, *Vision*, W. H. Freeman, San Francisco, CA, 1982.
2. D. Ballard and C. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.
3. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
4. M. Levine, *Vision in Man and Machine*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
5. T. Pavlidis, *Structural Pattern Recognition*, Springer, New York, 1977.
6. T. Pavlidis, *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, MD, 1982.

7. R. Nevatia, *Machine Perception*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

8. A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.

9. W. Pratt, *Digital Image Processing*, Wiley-Interscience, New York, 1978.

10. R. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1977.

11. T. Cornsweet, *Visual Perception*, Academic Press, New York, 1970.

12. R. Gregory, *The Intelligent Eye*, McGraw-Hill, New York, 1970.

13. R. Haber and M. Herschenson, *The Psychology of Visual Perception*, Holt, Rhinehart, and Winston, New York, 1973.

14. L. Kaufman, *Sight and Mind*, Oxford University Press, New York, 1974.

15. I. Rock, *The Logic of Perception*, MIT Press, Cambridge, MA, 1984.

16. W. Uttal, *A Taxonomy of Visual Processes*, Erlbaum, Hillsdale, NJ, 1981.

17. J. Beck, B. Hope, and A. Rosenfeld (eds.), *Human and Machine Vision*, Academic Press, Orlando, FL, 1983.

18. O. Braddick and A. Sleigh (eds.), *Physical and Biological Processing of Images*, Springer, New York, 1983.

19. H. von Helmholtz, in J. P. C. Southall (ed.), *Treatise on Physiological Optics*, Dover (reprint), Mineola, NY, 1962.

20. G. Fry, *Blur of the Retinal Image*, Ohio State University Press, Columbus, 1955.

21. S. Coren, L. Ward, C. Porac, and R. Fraser, "The effect of optical blur of visual-geometric illusions," *Bull. Psychon. Soc.* **11**(6), 390–392 (1978).

22. L. Roberts, "Machine perception of 3-dimensional solids," in J. Tippett (ed.), *Optical and Electro-Optical Information Processing*, MIT Press, Cambridge, MA, 1965.

23. Reference 22, p. 267.

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

25. F. S. Werblin, "Functional organization of a vertebrate retina: Sharpening up in space and intensity," *Ann. NY. Acad. Sci.* **193** (1972).

26. R. Kirsch, "Computer determination of the constituent structure of biological images." *Comput. Biomed. Res.* **4**, 315–328 (1971).

27. A. Hildebrandt, *Introduction to Numerical Analysis*, Wiley, New York, 1956.

28. A. Herskovitz and T. Binford, On Boundary Detection. Al Memo 183, MIT, Cambridge, MA, 1970.

29. B. Horn, The Binford-Horn Line Finder, Al Memo 285, MIT, Cambridge, MA, 1973.

30. J. M. S. Prewitt, Object Enhancement and Abstraction, in A. Rosenfeld and J. Prewitt (eds.). *Picture Processing and Psychopictorics*, Academic Press, New York, 1970.

31. R. Haralick, and L. Watson, "A facet model for image data." *Comput. Vis. Graph. Im. Proc. 15*, 113–129 (1984); R. Haralick, "Digital step edges from zero crossing of second directional derivative, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-6**, 58–68 (1984).

32. M. Heuckel, "An operator which locates edges in digital pictures," *JACM* **18**, 113–125 (1971).

33. D. Hubel and T. Wiesel, "Functional architecture of macaque monkey visual cortex. *Proc. Roy. Soc. London B* **198**, 1–59, (1977)—*a review.*

34. R. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," *Vis. Res.* **5**, 583–601 (1965).

35. C. Enroth-Cugell and J. Robson, "The contrast sensitivity of retinal ganglion cells of the cat." *J. Physiol.* (Lond.) **187**, 517–552 (1966).

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

37. S. Laughlin, M. Srinivasan, and A. Dubs, "Predictive coding: A fresh view of inhibition in the retina," *Proc. Roy. Soc. London B*, 427–459 (1982).

38. E. H. Land, "The retinex theory of color vision," *Sci. Am.* **237**(6), 108–128 (1977).

39. A. Witkin, Scale Space Filtering, in A. Pentland (ed.), *From Pixels to Predicates*, Ablex, Norwood, NJ, 1986, pp. 5–19.

40. D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy Soc. London B* **207**, 187–217 (1980).

41. Y. Leclerc and S. W. Zucker, "The local structure of intensity changes in images," *IEEE Trans.* **PAMI 9**, (1987).

42. R. Watt and M. Morgan, "Mechanisms responsible for the assessment of visual location: Theory and evidence," *Vis. Res.* **23**, 97–109 (1983).

43. S. W. Zucker and R. Hummel, Receptive Fields and the Representation of Visual Information, *Human Neurobiology* **5**, 121–128 (1986).

44. J. Lettvin, H. Maturana, W. McCulloch, and W. Pitts, "What the frog's eye tells the frog's brain," *Proc. IRE* **47**, 1940–1951 (1959).

45. H. Barlow, R. Narasimhan, and A. Rosenfeld, "Visual pattern recognition in machines and animals," *Science* **177**, 567–575 (1972).

46. D. Ballard, "Parameter networks," *Artif. Intell.* **22**, 235–267 (1984).

47. L. Davis, "Hierarchical generalized Hough transform and line-segment based generalized Hough transforms," *Patt. Recog.* **15**, 277–285 (1982).

48. M. B. Clowes, "On seeing things," *Artif. Intell.* **2**, 79–116 (1971).

49. M. Minsky and S. Papert, *Perceptions*, MIT Press, Cambridge, MA, 1969.

50. G. Kanisza, *Organization in Vision*, Praeger, New York, 1979.

51. Y. Shirai, Analyzing Intensity Arrays Using Knowledge about Scenes, in P. Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975, pp. 93–114.

52. A. K. Mackworth, "Interpreting pictures of polyhedral scenes," *Artif. Intell.* **4**, 121–137 (1973).

53. B. Horn, "Understanding image intensities," *Artif. Intell.* **8**, 201–231 (1977).

54. S. Shafer, *Shadows and Silhouettes in Computer Vision*, Kluwer Academic, Boston, MA, 1985.

55. R. Woodham, "Analyzing images of curved surfaces," *Artif. Intell.* **17**, 17–45 (1981).

56. G. Falk, "Interpretation of important line data as a three-dimensional scene," *Artif. Intell.* **3**, 77–100 (1972).

57. K. Turner, *Computer Perception of Carved Objects Using a Television Camera*, Ph.D. Thesis, University of Edinburgh, 1974.

58. L. Davis, "Shape matching using relaxation techniques," *IEEE Trans. PAMI* **PAMI-1**, 60–72 (1979).

59. H. Freeman, "Computer processing of line drawing images," *Comput. Surv.* **5**, 57–97 (1974).

60. J. Tsotsos, "Knowledge of the visual process: Content, form and use," *Patt. Recog.* **17**, 13–28 (1984).

61. S. Zucker, A. Rosenfeld, and L. Davis, "General Purpose Models: Expectations about the Unexpected," *Proceedings of the Fourth International Joint Conference Artificial Intelligence*, Tblisi, Georgia, 1975, pp. 716–720.

62. A. Rosenfeld, "A nonlinear edge detection technique," *Proc. IEEE* **58**, 814–816 (1970).

63. D. Marr, "Early processing of visual information," *Proc. Roy. Soc. (London)* **B275,** 483–534 (1976).

64. S. Tanimoto and L. Uhr, *Structured Computer Vision*, Academic Press, New York, 1983.

65. W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.* **5,** 115–133 (1943).

66. J. McCarthy and C. Shannon, *Automata Studies*, Princeton University Press, Princeton, NJ, 1956.

67. S. Winograd and J. Cowan, *Reliable Computation in the Presence of Noise*, MIT Press, Cambridge, MA, 1963.

68. M. Arbib, *The Metaphorical Brain*, Wiley, New York, 1972.

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

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

71. E. Ising, "Contribution to the theory of ferromagnetism," *Z. Physik.* **31,** 253–258 (1925).

72. D. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1969.

73. G. Sperling, "Binocular vision: A physical and a neural theory," *Am. J. Psych.* **83,** 461–534 (1970).

74. M. Fischler and R. Elschlager, "The representation and matching of pictorial structures," *IEEE Trans. Comput.* **22,** 67–92 (1973).

75. A. Guzman, Decomposition of a Visual Scene into Three-Dimensional Bodies, in A. Grasselli (ed.), *Automatic Interpretation and Classification of Images*, Academic Press, New York, 1969.

76. D. Huffman, Impossible Objects as Nonsense Sentences, in Meltzer and Michie (eds.), *Machine Intelligence*, Vol. 6, Edinburgh University Press, Edinburgh, 1971.

77. D. Waltz, Understanding Line Drawings of Scenes with Shadows, in P. Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975.

78. A. Rosenfeld, R. Hummel, and S. W. Zucker, "Scene labelling by relaxation operations," *IEEE Trans. Sys. Man Cybernet.* **SMC-6,** 420–433 (1976).

79. R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Patt. Anal. Mach. Intell.* **PAMI-5,** 267–287 (1983).

80. D. Ballard, G. Hinton, and T. Sejnowski, "Parallel visual computation," *Nature* **306,** 21–26 (1983).

81. L. Davis and A. Rosenfeld, "Cooperating processes for low-level vision: A survey," *Artif. Intell.* **17,** 245–264 (1981).

82. W. Kohler, *The Task of Gestalt Psychology*, Princeton University Press, Princeton, NJ, 1969.

83. B. Horn, Obtaining Shape from Shading Information, in P. Winston (ed.), *The Psychology of Computer Vision*, McGraw Hill, New York, 1975, pp. 115–156.

84. A. Pentland, "Local shading analysis," *IEEE Trans. PAMI* **PAMI-6,** 170–187 (1984).

84a. A. Pentland, "Fractal based descriptions of natural scenes," *IEEE Trans. PAMI* **PAMI-6,** 661–675 (1984).

85. J. Kender, Shape from Texture, Technical Report, Computer Science Department Carnegie-Mellon University, Pittsburgh, PA, 1980.

86. A. Witkin, "Recovering surface shape and orientation from texture," *Artif. Intell.* **17,** 17–47 (1981).

87. J. Beck, *Surface Color Perception*, Cornell University Press, Ithaca, NY, 1972.

88. J. J. Gibson, *The Perception Of The Visible World*, Houghton Mifflin, Boston, MA, 1950.

89. S. Ullman, "Against direct perception," *Behav. Br. Sci.* **3,** 373–415 (1980).

90. Reference 1, p. 37.

91. H. Barrow and J. M. Tenenbaum, Recovering Intrinsic Scene Characteristics from Images, in A. Hanson and E. Riseman (eds.), *Computer Vision Systems*, Academic Press, New York, 3–26, 1978.

92. W. E. L. Grimson, *From Images to Surfaces*, MIT Press, Cambridge, MA, 1983.

93. D. Terzopoulos, Multilevel Computational Processes for Visible Surface Reconstruction, Ph.D. Thesis, MIT, Cambridge, MA, 1984.

94. J. J. Koenderinck and A. van Doorn, "Photometric invariants related to solid shape," *Opt. Acta* **27,** 981–996 (1980).

95. J. J. Koenderinck and A. van Doorn, "The shape of smooth objects and the way contours end," *Perception* **11,** 129–137 (1982).

96. G. Johanssen, *Configurations in Event Perception*, Almquist and Wiksells, Uppsala, Sweden, 1950.

97. H. Wallach and D. O'Connell, "The kinetic depth effect," *J. Exp. Psych.* **45,** 205–217 (1953).

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

99. J. Fodor, *The Modularity of Mind*, MIT Press, Cambridge, MA, 1984.

100. D. Dennett, *Brainstorms*, Bradford Books/MIT Press, Cambridge, MA, 1978.

101. D. van Essen and J. Maunsell, "Two-dimensional maps of the cerebral cortex," *J. Comp. Neurol.* **191,** 255–281 (1980).

102. M. Regan and M. Cynader, "Motion-in-depth neurons; effects and speed and disparity," *Invest. Ophth. Vis. Sci.* **20,** 148 (1981).

103. G. Orban, *Neuronal Operations in the Visual Cortex*, Springer, New York, 1984.

104. S. Zeki and S. Shipp, "Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey cortex," *Nature* **315,** 322–324 (1985).

105. H. Barrow and J. M. Tenenbaum, "Computational vision," *Proc. IEEE* **69,** 572–595 (1981).

106. P. Cavanagh, "Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape, *Comput. Vis. Graph. Im. Proc.,* **37** (1987).

107. D. Hoffman and B. Flinchbaugh, "The interpretation of biological motion," *Biol. Cybernet.* **42,** 195–204 (1982).

108. J. Cutting, "Coding theory adapted to gait perception," *J. Exp. Psych. Hum. Perc. Perf.* **7,** 71–87 (1981).

109. W. Ittleson, *Visual Space Perception*, Springer, New York, 1960.

110. J. Feldman and D. Ballard, "Connectionist models and their properties," *Cog. Sci.* **6,** 205–254 (1982).

111. S. Ullman, "Visual Routines," *Cognition* **18,** 97–159 (1984).

112. M. Brady and A. Yuille, An Extremum Principle for Shape from Contour, *Proc. of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, FRG, 1983.

113. B. Horn and B. Schunk, "Determining optical flow," *Artif. Intell.* **17,** 185–204 (1981).

114. K. Ikeuchi and B. Horn, "Numerical shape from shading and occluding boundaries," *Artif. Intell.* **17,** 141–184 (1981).

115. S. W. Zucker, "Early orientation selection: Tangent fields and the dimensionality of their support," *Comput. Vis. Graph. Im. Proc.* **32,** 74–103 (1985).

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

117. L. Quam, Road Tracking and Anomaly Detection, *Proceedings of the DARPA Image Understanding Workshop*, pp. 51–55, 1978.

118. A. Witkin, Intensity Based Edge Classification, *Proceedings of the Second AAAI Conference*, Pittsburgh, PA, pp. 36–41, 1982.

119. K. Stevens, "The visual interpretation of surface contours," *Artif. Intell.* **17,** 47–74 (1981)..

120. M. Brady and Asada, "Smoothed local symmetries and their implementation," *Int. J. Robot. Res.* **3,** 36–61 (1984).

121. S. W. Zucker, "On the structure of texture," *Perception* **5,** 419–436 (1976).

122. J. Beck, Texture Segregation, in J. Beck (ed.), *Organization and Representation in Perception,* Erlbaum, Hillsdale, NJ, 1982.

123. R. Haralick, Statistical and Structural Approaches to Texture, *Proceedings of the Fourth International Joint Conference on Pattern Recognition,* Kyoto, Japan, 1978, pp. 45–69.

124. T. Kanade, "Recovery of the three-dimensional shape of an object from a single view." *Artif. Intell.* **17,** 409–460 (1981).

125. G. Dodd and L. Rossol, *Computer Vision and Sensor-Based Robots,* Plenum, New York, 1979.

126. M. Kass and A. Witkin, Analyzing Oriented Patterns, *Proceedings of the Ninth International Joint Conference on Aritificial Intelligence,* Los Angeles, 1985, pp. 944–952.

127. D. Hoffman and W. Richards, Parts of Recognition, in A. Pentland (ed.), *From Pixels to Predicates,* Ablex, Norwood, NJ, 1986, pp. 268–294.

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

129. M. Wertheimer, "Laws of organization in perceptual forms," *Psych. Forsch.* **4,** 301–350 (1923); translated in W. Ellis, *A Source Book of Gestalt Psychology,* Routledge and Kegan Paul, London, pp. 71–88, 1938.

130. A. Witkin and J. M. Tenenbaum, On the Role of Structure in Vision, in J. Beck, B. Hope, and A. Rosenfeld (eds.), *Human and Machine Vision,* Academic Press, New York, 1983.

131. D. Lowe and T. Binford, Segregation and Aggregation: An Approach to Figure/Ground Phenomena, *Proceedings of the DARPA Image Understanding Workshop, 1982,* pp. 168–178.

132. D. Lowe, Perceptual Organization and Visual Recognition, Ph.D. Thesis, Stanford University, 1984.

133. J. Mayhew and J. Frisby, "Psychophysical and computational studies towards a theory of human stereopsis," *Artif. Intell.* **17,** 349–385 (1981).

134. D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and their Applications,* Academic Press, New York, 1980.

135. S. W. Zucker and R. A. Hummel, "Toward a low-level description of dot clusters: Labelling edge, interior, and noise points," *Comput. Graph. Im. Proc.* **9,** 213–233 (1979).

136. D. Terzopoulos, Integrating Visual Information from Multiple Sources, in A. Pentland (ed.), *From Pixels to Predicates,* Ablex, Norwood, NJ, 1986, pp. 111–142.

137. T. Binford, Visual Perception by Computer, *Proceedings of the IEEE Conference on Systems and Control,* Miami, 1971.

138. A. Pentland, "Perceptual organization and the representation of natural form," *Artif. Intell.* **28,** 293–331 (1986).

139. A. Barr, "Superquadrics and Angle-Preserving Transformations," *IEEE Computer Graphics and Applications,* 11–23 (1981).

140. J. Foley and A. van Dam, *Fundamentals of Interactive Computer Graphics,* Addison-Wesley, Reading, MA, 1982.

141. A. Triesman, "Preattentive processing in vision," *Comput. Vis. Graph. Im. Proc.* 1–22 (1985).

142. O. Faugeras, Steps toward a Flexible 3-D Vision System for Robotics, in H. Hanufusa and H. Inoue (eds.), *Robotics Research: The Second International Symposium,* MIT Press, Cambridge, MA, 1985.

143. R. Brooks, "Symbolic reasoning among 3-D models and 2-D images," *Artif. Intell.* **17,** 285–348 (1981).

144. T. Garvey, Perceptual Strategies for Purposive Vision, Technical Note 117, SRI International, Menlo Park, CA, 1976.

145. A. Hanson and E. Riseman, *Computer Vision Systems,* Academic Press, New York, 1978.

146. J. K. Tsotsos, J. Mylopolous, D. Covvey, and S. W. Zucker, "A framework for visual motion understanding," *IEEE Trans. Patt. Anal. Mach. Intell.* **PAMI-2,** 563–573 (1980).

147. T. Binford, "Survey of model based image analysis systems," *Int. J. Robot. Res.* **1,** 18–64 (1982).

148. F. Ferrie, M. D. Levine, and S. W. Zucker, "Cell tracking: A modeling and minimization approach," *IEEE Trans. Patt. Anal. Mach. Intell.* **4,** 277–291 (1982).

149. M. Levine and S. Shaheen, "A modular computer vision system," *IEEE Trans. PAMI* **PAMI-3,** 540–556 (1981).

150. B. K. P. Horn, Obtaining shape from shading information, in P. Winston (ed.), *The Psychology of Computer Vision,* McGraw-Hill, New York, 1975.

151. A. Pentland, Local shading analysis, *IEEE Trans. Patt. Analysis and Machine Inst.* **PAMI-6,** 170–187 (1984).

152. M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs,* Plenum, New York, 1980.

S. W. Zucker
McGill University

Thanks to the Natural Sciences and Engineering Research Council (Canada), grant A4470, and the Canadian Institute for Advanced Research for their support.

VISUAL DEPTH MAP

Early vision (qv) is often characterized as a process that reconstructs the three-dimensional properties of scenes from their 2-D images. The dimension lacked by images is the distance along lines of sight from the viewer to points on physical surfaces in the environment. This distance is known as depth.

Humans effortlessly gain a strong sense of depth as they navigate the natural world with both eyes open, and a weaker though definite depth percept results from viewing a monocular image of a 3-D scene. Furthermore, synthetic visual stimuli such as random-dot stereograms and kinetic displays are known to elicit compelling percepts of coherent surfaces in depth.

The recovery of depth from images may be formulated as an inverse problem, converse to the direct problem in computer graphics of rendering images of 3-D geometric models. Evidently, the human visual system is adept at solving this inverse problem at an early processing stage, and it is reasonable to hypothesize the existence of an internal representation of three-dimensionality, putatively in terms of perceived depth z over the visual field. This representation, naturally expressed as a single valued depth function $z = u(x, y)$, where x and y are retinotopic or image coordinates, is commonly known as the visual depth map.

The visual depth map has attracted substantial attention in the study of human and machine vision. The first section briefly examines techniques for acquiring initial 3-D information to construct depth maps. These include range sensors engineered to actively acquire depth data from the environment and computational processes of early vision that perform a 3-D analysis of images. In computational vision the visual depth