COMPUTER VISION AND HUMAN PERCEPTION

an essay on the discovery of constraints

Steven W. Zucker
Computer Vision and Graphics Laboratory
Department of Electrical Engineering
McGill University
Montreal, Quebec, Canada

ABSTRACT

The study of vision, in both man and machine, is viewed as the discovery of constraints. Computational constraints often imply assumptions necessary for achieving a problem's solution, while psychological and neurophysiological ones restrict the manner in which such solutions can be achieved. These ideas are illustrated by several examples of research related to the early processing of visual information. The development of the paper takes place historically, starting with Helmholtz and Mach, as well as comceptually, from the concrete to the abstract, and anatomically, from the eye to the brain.

I INTRODUCTION

Computer vision and human perception — two realizations of the process of seeing, one embedded in computers and the other in people. Clearly there is a metaphorical level in which these two activities have much in common. But is it only a metaphorical level, with fundamental differences always keeping them separate? Or is there real substance to the metaphor, so that each side could benefit from interacting with the other. We shall argue, in this paper, for the latter. Our position is that, since the process of vision is an immensely complex one, theories at many different levels of abstraction must be utilized. As Marr and Poggio (40) have stated: "The CNS needs to be understood at four nearly independent levels of description: (1) that at which the nature of the computation is expressed; (2) that at which the algorithms that implement a computation are characterized; (3) that at which an algorithm is committed to particular mechanisms; and (4) that at which the mechanisms are realized in hardware." Traditionally, computer vision operates at the computational level of description, while the study of human perception has been more concerned with input/output or neurophysiological descriptions.

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 whether or not the parent theoretical framework changes. The evolution of our understanding of these constraints is the principle theme running through this paper; this is what we take to be progress in understanding vision. As we shall illustrate, constraints have been discovered that fall into three main categories: computational, behavioural, and implementational.

The preparation of this paper was supported by the National Sciences and Engineering Research Council. Harold Hubschman, Peter Sander, and Demetri Terzopoulos provided constructive, critical, and ocassionally complimentary comments.

Computational constraints are the most abstract. Given a statement of a visual problem, these are the constraints that must be in effect for a particular solution of that problem to be correct. Mathematically they are required to transform underdetermined situations into determined onec/. In the broadest sense, the need for constraints can be seen from the image formation process. A view of a three-dimensional scene is projected onto our two-dimensional retinas; to recover a description of the scene, somehow the loss in this degree of freedom must be overcome. This requires the introduction of constraints. Discovering what these constraints can, and should,be, is a subtle process; instances of it will occupy much of this essay. For example, each ray of light impinging on our retinas is obtained from a certain product of illumination and surface reflectance. When this relationship is expressed mathematically, there are clearly infinite combinations that could satisfy it. But, if illumination is assumed to be constant and distant, then the pattern of perceived illumination becomes proportional to surface reflectance. And, if the surface is further assumed to be uniformly reflective, then it becomes proportional to surface orientation. As each of these assumptions is understood as a constraint on the solution, a unit of progress is made toward understanding which constraints could be active for the general vision problem.

The other two classes of constraints manifest themselves less as assumptions and more as restrictions. They specify what the visual system has available for implementing solutions, as well as intermediate states encountered while achieving them. They may be characterized in terms of the available "machinery", as in the case of neurophysiology, or they may be characterized behaviourly, as in the case of psychology.

Because of the complexity of vision, it is our position that each of these different kinds of constraints is needed, or the likelihood of discovering the correct explanation is seriously diminished. Without the computational theories and constraints, one is faced with the problem of inferring what staggering numbers of neurons are doing, without a suitable language for describing either them or their scope. The problem is perhaps even more difficult than inferring what a digital computer is doing in terms of the electronics. Imagine, for example, trying to infer the scheduling algorithm, or even the need for a scheduler or operating system, without our present computational background. As another example, recall Chomsky's classical critique of Skinner's behaviourism: try to express the notion of attack underlying a particular chess strategem in terms of conditioned collections of neurones. To appreciate the need for the other, behavioural constraints, just recall how many different techniques there are for solving systems of partial differential equations, or optimization problems. Such constraints could participate in the decision to use a simplex or a gradient algorithm, running on a parallel or a sequential machine.

While constraints shape theories, they rarely do so to the point of uniqueness. Such underdetermined theories specify competency [8] or sufficency of a system; they state what could be happening, not necessarily what is happening. As additional constraints are added, however, the theories become sharper and more focused, particularly when the constraints span several descriptive levels. In the limit, we believe, enough constraints will become known at each level so that a complex of sufficent theories will become, or inspire, the correct one.

In summary, the other general points of the paper are that

1. Computer vision can provide a language for posing theories of visual information processing, and such languages are essential;

2. computer vision can provide a capability for carrying out experiments that are essentially impossible to perform without confounding within the human visual system;

3. evidence about human perception can provide clues for computer vision that would not be obvious otherwise, and vice versa;

4. theories at different descriptive levels are instructive, if not necessary, to restrict experimental and theoretical scope at all levels of explanation, whether one is concerned with computer or human perception or both. In this paper, however, we shall primarily be concerned with human perception.

While some of the above points have taken on new importance given the current development of computation, the most basic theme -- the necessity for multiple-levels of description — is a classical one. This theme is evident when one looks across the writings of the great vision scientists, and we shall illustrate it with brief (and perhaps overly simplified) views of Hermann von Helmholtz and Ernst Mach. The progression that we shall follow will be both historical and conceptual, with Helmholtz portrayed as a physicist and Mach as a neural modeler. We shall then return to Helmoltz, because of his strong position on the role of "unconscious inferencing" in perceptual processing. Conceptually we shall progress from the eye to the brain, and the concrete to the abstract. The examples will be chosen from early visual information processing. The current paradigm for vision we take to be a loose, but logical, development from the earlier positions, although it does substantial refinement of them. Helmholtz and Mach were both serious philosophers, physicists, and mathematicians, as well as psychologists. Thus their views about vision spanned many of the descriptive levels to which we have referred.

2. THE EARLIEST CONSTRAINT
PHYSICAL IMPERFECTIONS IN THE EYE

Parmenides (ca. 500 B.C.) explained the existence of visual illusions by observing; "The eyes and ears are bad witnesses when they are at the service of minds that do not understand their language". We shall begin our discussion of vision with a discussion of the earliest possible "language" in the visual understanding process — the optics of the eye. As a medium, we shall use the Mueller-Lyer illusion, one of the most extensively-studied (but still not completely understood) geometric illusions. And H. von Helmholtz will provide the conceptual viewpoint for our investigation.

In his treatise on Physiological Optics [20], 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. Such unconscious inferences we shall take to mean computations, a notion that Helmholtz was (unfortunately) rather vague about. The only language that he had for talking about them was that of "conscious inferences", or the logic of premises and conclusions. Other portions of his investigations were incredibly concrete and clear, however, such as his study of the transduction properties of the eye, and it is this with which we shall now be concerned. 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. It exhibits the many different forms of spherical aberration and distortion to which physically-realized systems are susceptable.

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 filled in (i.e., 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 problem 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 issusion).

Helmholtz was correct in observing that the eye is an imperfect optical instrument. But he was wrong, in part, in that his explanation of the Mueller-Lyer illusion cannot account for the entire effect. This has been determined very recently using an elaborate optical technique, an artificial pupil, to project a highly focused image onto the retina (10]. Such techniques indicate that optical blurring accounts for roughly 15% of the illusory effect. Nonetheless, the constraints stand as contributors. (We shall discuss other contributions to the Mueller-Lyer later in the paper.)

3. LATERAL INHIBITION AND NEURAL MODELS

We now turn from a phenomenon of blurring to one of sharpening, from explanations in terms of optical mechanisms to ones embodied in neural networks, and from Helmholtz to Ernst Mach. The phenomenon of sharpening that we shall discuss is commonly known as Mach bands — it is the addition of subjective bright and dark lines (bands) on either side of an intensity edge (see Fig. 2). Such bands indicate that the eye responds 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, our impressions of brightness and of contrast are not isomorphic with the intensity of light impinging on our retinas, but rather are derived — or computed — from it.

Mach's theoretical position was based on a belief 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 'psychical events'. "The psychophysical law holds...for the relation of the primary stimulus and the last nerve excitation with which the conscious sensation goes. Indeed, this is precisely because the excitations in the sense organs are filtered through a complicated web of nerves." [45,299-300]

The particulars of Mach's explanation were posed mathematically in terms of "a reciprocal interaction of neighboring areas of the retina" (45,267]. He cited (then) current neuro-anatomical data by Ritter that postulated a regular arrangement of cells on the retina, and characterized the function of these cells mathematically. Thus he was concerned with possible constraints from the •wetware*. 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.

While 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 excitatory contributions at that point together with inhibitory interactions from neighboring points [11]— 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 sturcture looks circularly symmetric, with an excitatory central area surrounded by a negative, or inhibitory, area. 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. Such changes are important because they often indicate the presence of physical object contours, one of the most fundamental constraining relationships between the physical and the visual worlds. In fact, the functional significance of Mach bands has often been attributed to their edge-enhancement effect -- if we are to navigate through the physical world on the basis of sensory information, we certainly need to locate object contours. But this kind of explanation is pure teleology. It almost implies that there should be a little "homonculus" inside our heads whose job was to look at the visual (retinal) image to locate edges. Enhancement would then make his job easier.

Lateral inhibition may be one of the most ubiquitous mechanisms in biological vision systems. It plays a clear role in regulating the dynamic range of the eye [52] and otherwise performing a sort of local sharpening, or maxima selection, at the neural level [11]. But these are all very low-level functions; whether it actually helps the human visual system to find edges still remains an open question.

4. EDGES AND FEATURES: CAN THEY BE DETECTED?

Neurophysiology, in addition to verifying lateral inhibition in certain animals, also inspired a revolutionary theory of how the early visual system functions. In a striking series of observations, Hubel and Wiesel [27] measured the receptive fields of different cells in the lateral geniculate nucleus and the visual cortex of the cat and monkey. (The receptive field is the arrangement of retinal cells — rods and cones — which, when stimulated with a pattern of light, influence the activity of the cell under measurement. The lateral geniculate nucleus is the first major processing station between the retinal ganglia and the visual cortex.) The structure of these receptive fields (with respect to certain of their defining characteristics) was striking; "Roughly four classes of cells can be distinguished, in a series of ascending complexity... These are termed 'circularly symmetric', 'simple', 'complex', and 'hypercomplex*. We assume that cells at each stage receive their major input from cells at the previous stage, with the circularly symmetric cells receiving their inputs predominantly from geniculate cells. Circularly symmetric cells, as their name implies, show no preference to any orientation of lines, and indeed, seem similar in their properties to geniculate cells. Simple cells are the first in the hierarchy to show orientation specificity, so that the rearrangements responsible for orientation specificity are presumed to take place between the circularly symmetric and the simple cells. A simple cell responds to an optimally orientated line in some narrowly defined position: even a slight displacement of the line to a new position, without change in orientation, renders the line ineffective. A complex cell, on the contrary, is probably just as specific in its orientation requirements as the simple cell, but is far less particular about the exact positioning of the line___ Hypercomplex cells, finally, resemble complex cells in all respects but one: extending the line beyond the region from which responses are envoked produces a marked reduction or complete abolition of the response." [24, p. 8). The structure of these receptive fields is as shown in Fig. 3 , and the interpretation of the simple cells by the psychological community was immediate: "It takes little imagination to describe these simple cortical fields as edge detectors and line detectors." HB,p. 54).

While such neurophysiological observation is striking, and certainly introduces strong constraints on what the visual system is doing, as well as how it is doing it, is the jump *from* observation to a theory of edge and line detection correct? That is, do the simple cells detect lines and edges? We pose this conjecture to highlight one of the main contributions of computer vision to the understanding of human perception — in addition to providing constraints, it provides us with a means of testing them. In this case, computer vision can test a version of the above conjecture: Are simple cells a sufficient mechanism for detecting lines and edges? The answer, it turns out, is no, at least for the manner in which our intuitions first indicated. While simple cells can detect lines and edges in certain clearcut situations, they are not sufficent to accomplish the task in arbitrary ones. See Fig. 5. Such computational experiments also reveal the problem with simple cells and other such "feature-detector" theories — their response is ambiguous. They respond not only to lines and edges, but to other 'noise' patterns as well. As we shall show, however, it does look as if they are involved in the edge finding process, but are not the whole story.

1L INTENSITY EDGES AND PHYSICAL CONTOURS

To more properly appreciate the complexity of the edge-finding problem, consider how the underlying physical events constrain the image intensities. A physical edge can be said to exist if there is a change in surface reflectance, orientation, or illumination. The resultant image profiles have been examined by Binford [21], who found, for edge-like patterns, that there are three primary classes: step, roof, and spike. Thus edges come in many different guises, or functional forms, certain properties of which can be related back to physical configurations. Horn [22] found, e.g., that step edges are likely to correspond to occluding surface boundaries, but that these inverse constraints are rather weak ones. Thus the search for a single, perfect, one-step edge detector, like the simple cells, begins to feel elusive. Furthermore, edge "events", such as surface reflectance or orientation changes, can take place at many different scales [39]. Highlights are usually sharp, and shadows blurry. An intensity gradient arising from a curved surface and a single light source will span a much wider physical distance than the abrupt shift caused when one surface occludes another. Any general purpose edge-finding scheme must take such functional and scale dependence into account.

To make matters worse, there are still other confounding constraints. If we view the circularly-symmetric center surround cell as a discrete approximation to a Laplacian (i.e., second spatial derivative) operator, then numerical analysis tells us that the more smoothing incorporated into the operator, the more stable its performance. Such smoothing is necessary to counteract many forms of 'noise', such as that which is introduced by the sampling process. But as more smoothing is incorporated, the likelihood of evaluating the operator across an image of distinct, but small, physical edges increases. Such events make the response of the operator even less reliable.

One solution to these conflicting constraints, developed for computer vision systems, is the use of hierarchies of operators at different sizes [32], Incredibly, it now seems that a form of this solution is used by the human visual system as well. A rather large body of psychophysical evidence, beginning with the work of Campbell and Robson [6], and summarized into a clean descriptive theory recently by Wilsen and Bergen [53]*1* indicates that visual information is decomposed, very early, into a number (perhaps 4 or 5) of separate channels, each of a different spatial resolution. The finest such channel carries information limited by the actual placement of receptor cells in the fovea, while the broadest carries highly smoothed information. Marr and Hildreth [39] have used this observation, together with the observed properties of the first of the Hubel/ Wiesel cells (the circulary symmetric ones), to propose a new theory of how the human visual system actually begins to find edges. Based on the assumption that the first stage of edge finding should be directionally insensitive, they proposed a first stage of the edge finding process based on operators that model the channel blurring followed by the Laplacian; an image of such an operator is shown in Fig. 6. Note that there is a hierarchy of five such operators. The response of these operators (since they correspond to second derivatives, the responses across edges are zero crossings) is shown in Fig. 7.

ii. FURTHER CONSEQUENCES OF "NEURAL" BLURRING

The Marr/Hildreth zero-crossing scheme makes explicit assumptions about the existence of distinct channels, each of which implements a certain degree of blurring. Our belief in the existence of these channels would certainly be strenghtened if we could find other perceptual effects that were also consistent with such blurring. The original conjecture was in fact based on a wealth of such effects — contrast sensitivity in the perception of gratings. These, however, are highly technical, and our present goal is to search for more readily observable ones.

The first example is one that we are already familiar with — the Mueller-Lyer illusion. We have discussed Helmholtz's observations about blurring and have noted that they can only account for a small fraction of the illusory effect. We now have another possible source of blurring — the spatial-frequency limited channels in the early visual system. Experiments in our laboratory indicate that the smallest amount of channel blurring possible, that from a hypothetical channel for high visual acquity, would correspond to approximately 13% of an illusory effect for a perfect image [55] — see Fig. 8. The larger channels would, of course, imply more of a distortion in the direction of the illusion. During normal viewing of the illusion, it seems clear that the optical and channel blurrings should be additive.

The next example is also visually striking. It can be obtained by taking a normal checkerboard pattern and shifting every other row a fraction of a cycle — see Fig. 9. Note that the lines of the checkerboard no longer appear straight; rather, they are skewed slightly off the horizontal. This phenomenon was first studied by Munsterburg [43], after it was brought to his attention by a weaving instructor who could not understand why his students could not weave a straight linel More recently, a variant was discovered by Gregory [16).

Related phenomena were known to Helmholtz, who referred to their cause with the term "irradiation". Can our smoothing and differentiation constraints be held accountable again? The answer is yes — see Fig. 10. The zero-crossing contours describe trapezoids, not rectangles; smoothing followed by differentiation does introduce distortion of the right kind. Conversely, if we were to avoid the smoothing and differentiation constraints, e.g., by presenting the cafe wall image as a random dot stereogram [27), then we would expect the illusion not to be present. Experiments in our laboratory, with a (512 x 512) digital stereogram, do indicate this to be the case.

As a final example, we can pose a converse conjecture. If edge events are signaled by intensity arrays whose profile changes, then, if the profile were changing slowly enough, the differential operator should miss it. In particular, the channel theory allows an estimate of the largest operators, so that we can derive a lower bound for perceivable edge profiles. The constraint that edge profiles change more slowly than this may underlie some of the classical lightness illusions, such as the Cornsweet-0*Brien [11]. We have begun to investigate this conjecture in our laboratory, and preliminary tests seem to hold. In any case, evidence seems to be accumulating in the right direction.

But all is not done. While the zero crossings give a strong indication of where some edges lie, and separate some scale effects, they do not respond always, and only, to edges. And the responses from the different size operators have to be unified into a single coherent edge description, rather than the series of decomposed estimates. (This is not to say that there is no information in the different channels, or that this decomposed information is not useful. In fact, we shall see examples of where this is precisely the case.) To state matters another way, we cannot take the response of the local operators as definitive; the responses still need to be interpreted. Furthermore, even if we could find perfect local indications of where edges lie, they still have to be grouped (i.e., joined up) into longer contours. In the case of the cafe wall illusion, for example, the sides of the trapezoids have to be grouped into long lines. We shall examine possible approaches to grouping following a discussion of mterpretation.

7. INTERPRETING THE ZERO CROSSINGS

While the Marr/Hildreth operators answer some of the problems associated with the location of intensity changes, they are only a first step toward the solution- They do carry information useful in this task, but only in the form of a signal or measurement — they must still be interpreted just like the simple cells discussed previously. Computer vision has developed a wide range of such interpretation schemes, two of which we shall now consider through the use of simple cells. The difference with the earlier discussion of simple cells is that now they will be used to interpret the zero-crossings, rather than interpreting the raw intensities. In a sense, then, the edge finding problem has become one of line finding.

We shall take the problem of interpreting the zero-crossings to be one of asserting the underlying edge segment signaled by the zero crossings. There may, of course, actually be none. Perhaps the simplest such interpretation scheme begins with the evaluation of the responses of a number of simple cells at different orientations centered on a given image point, and selects the local orientation at that point to be the one difined by the simple cell with the strongest response. Such an interpretation scheme is purely local, and amounts to thresholding, or maxima selection. It is the kind of computation that can be accomplished, e.g., by a lateral inhibitory network. Because the available information is limited in scope, however, if the strongest response is uncertain, then it is not clear how much faith we should place in the result of local maxima selection.

A stronger interpretation strategy would be one in which an external criterion were imposed onto the process: select, for example, the interpretation that yields the minimal total curvature for the edge segments. Except for universes such as bubble chamber photographs, however, this constraint is too artificial. Human, and most machine, perception must be much more general purpose [1,60]. Another criterion can be obtained by examining the manner in which zero-crossings are imaged through the simple cells. Small ones give a response that is indicative of the presence of lines, but not definitive, because their spatial extent is too limited. The larger channels, on the other hand, have more context but are a bit too coarse for accurate fine measurements. Putting these two sources of information together, however, yields the capability for specifying context, with the larger masks, and details, with the smaller ones. It gives rise, in optimization terms, to a criterion in which the small mask responses are interpreted to make them as consistent as possible with the responses of the larger masks. Or, more precisely, the goal of the optimization process can be stated as follows: find the underlying pattern which is most likely to give the measured responses, both small and large [59]. It is interesting to note that the development of such a criterion suggests that the channels should increase in size by a factor of two, which is in fact the observed relation between them. Also, the development of such computational interpretation strategies suggests novel roles for the simple cells. For example, it has long been known from the neurophysiology that termination points, of lines and edges, are important and probably explicit [24]. if they are also to be found from the zero-crossings, as would seem logical, then we would like to suggest that this may be accomplished by using the "edge" mask, or directional first derivative simple cells, to signal zero-crossings that stop. Within the context of the above strategy for interpreting the zero-crossings, this would simply amount to adding a few more masks to the same process. Stepping back, the role of the masks can thus be said to add empirical constraints to the edge-finging process, particularly to the zero-crossing interpretations.

8. BACK TO UNCONSCIOUS INFERENCE AND HYPOTHESIS FORMATION

At this point in our discussion we would like to stress a change in our orientation. Rather than looking at effects whose explanation lies in a clear mechanism, such as the optical imperfections of the eye or neural networks implementing lateral inhibition, now we are considering much more abstract questions: how, for example, can the zero-crossings be interpreted? Interpretation is a symbolic act, whos'e explanation is most likely to be found in computational terms. Once these are understood, then the likelihood of finding the underlying neural implementation would seem much higher.

Such symbolic acts of interpretation are an example of what we believe Helmholtz had in mind when he spoke about unconscious inference, although they are probably at a much lower level than he believed necessary. He was concerned that there was a "false assumption that the mental operations we are discussing take place in an undefined, obscure, halfconscious fashion; that they are, so to speak, mechanical operations, and thus subordinate to conscious thought, which can be expressed in language. I do not believe that any difference in kind between the two functions can be proved" [19, p. 181]. A more modern champion of this position is Richard Gregory, who has regarded perception "as a matter of building up and testing hypotheses" [15, p.162), or the induction of hypotheses from information that is often highly underconstraining.

If we compare Helmholtz (and Gregory) with Mach, the difference in their positions (at least so far as they have been portrayed here) becomes apparent: Helmholtz was striving towards constraints on the algorithmic solution level, while Mach was striving towards constraints on the implementation level. This immediately raises the question of reducibility between levels: are theories expressed in the language of computation uniquely expressible in the language of hardware, such as the physics of transducers. This is sometimes possible, as for example the reduction of thermodynamics to statistical mechanics [44], or the computation of the smoothing and differentiation operations discussed in Sec. 5 directly in terms of models for the X and Y cells [46]. This latter example is expecially exciting, because it directly spans the bridge between two levels, one addressed by computer vision and one addressed by neurophysiology. But in general it is not possible; at some point, information processing becomes intentional [12], That is, it depends on the abstract state that the machine is in. Precisely where this occurs in perception is still an open question, perhaps made more decidable by the possibility of computational models.

Computational models of perception have two essential components — representational languages for describing information, and mechanisms that manipulate those representations. One important mechanism is clearly the creation of descriptions, and the inferential side of perception makes the need for this explicit. Constraints may be active, as we shall see, on both what is represented, and on how it is manipulated; i.e., constraints are active on both representations and mechanisms.

One of the strongest arguments for having explicit abstract representations is the fact that they provide explanatory terms for otherwise difficult (if not impossible) notions. Perhaps the clearest example of this is a subjective figure, or a structure constructed purely from context.

9. EDGES CAN BE SUBJECTIVE

The edges that we have been talking about up to now have all had manifestations in the image intensities. It is possible, however, to create the impression of edges in contexts where there are no actual intensity differences; see Fig. 11 for examples due to Kanisza [29]. Such subjective edges (and other figures) are so compelling that apparent intensity differences across them can actually be measured psychophysically. More importantly, however, these subjective edges appear to be associated with depth changes, as if they were the result of inferred surface discontinuities [9].

Should subjective edges, once formed, be considered as the same kind of abstract entity as intensity edges? That is, can subjective edges behave in a manner similar to intensity edges. The answer is yes, and it can be illustrated by another geometric illusion — the Poggendorf. It is present whether or not the defining edges are subjective (see Fig. 12). In representational terms, then, it would seem that edges ought to be considered as symbolic descriptive entities, whether or not they are subjective. Such a position is further consistent with our previous discussion about interpreting the zero-corssings; the result of the interpretation process is the establishment of these symbolic entities.

Once symbolic entities have been created, they can serve as input to later processes. Some of the clearest illustrations of this occur in perceptual grouping.

I0. GROUPING AND THE CONSTRUCTION OF ABSTRACT ENTITIES

Grouping is a generic name for a class of processes that take local entities of one kind and join, or combine, or "group" them into another one. It is illustrated nicely with the following example. Consider an array of random dots. If a copy of this dot array is first rotated and then superimposed on the original, the resultant pattern is not only one with twice the density of dots; it also exhibits clear structure [14] (see Fig. 13). Such patterns are called random dot Moire patterns.

How shall we model grouping phenomena? If we view the local entities as dots, then one possibility would be to specify a relation over the dots that describes which ones participate in the agglomerative structure. Technically such a relation can be viewed as a "virtual line" [47]; i.e., a relation that indicates which pairs of dots define the apparent circles. The process of grouping, in this case, thus results in the establishment of virtual lines.

While grouping in the case of random dot Moire patterns is clearly constructive, that is, it imposes structure not present in the intensities, it is important to realize that the agglomeration of local edge and line segments into long lines and curves is no less so. The earliest stages of visual processing decompose the retinal array into discrete, local pieces, at some of which line (or edge) segments are signaled. Nor is it necessary for the entities participating in the grouping to be explicit in the intensities -- subjective figures can be grouped as well (Fig. 14). The picture of low level visual processing thus beginning to emerge is one of many levels of descriptive entity construction and grouping, with interaction taking place between processes when warrented by the individual constraints undergoing satisfaction. Precisely what these constraints might be, as well as how they may be satisfied, are considered with the next topics.

11. GROUPING IN SPACE AND GROUPING IN TIME

The examples of grouping that we have considered up until now have all been grouping in space, such as the linking of spatially neighboring dots. An analagous form of grouping takes place in time, and can be used for motion computations. Such grouping establishes a correspondence between descriptive entities at different times (e.g., derived from temporal image sequences) that, presumably, denote the successive representations of the same physical event. Once such a correspondence has been established, it becomes possible to obtain the structure of rigid objects undergoing Euclidean motions [50].

Consider, again, a random dot array, this time with the dots moving. Since the motion of each dot in the array does not normally influence the motions of the others, Ullman [50] has argued that a viable assumption at the base of the human correspondence process in that all motions be considered independently. He has further argued, given this assumption, that the appropriate correspondence relations between tokens can be found by minimizing a functional (i.e., a sum) of "affinities" between tokens. Such affinities are proportional, e.g., to the length and orientation differences between line segments, or to the distance between dots. The motion correspondence process, then, requires machinery for minimizing functional criteria, a process to which we shall return.

12. FURTHER CONSTRAINTS ON GROUPING

What are the appropriate models for guiding grouping? Since the goal of vision is to produce descriptions of the three-dimensional structure of the world, it would seem that constraints about this 3-D structure should enter the grouping process. Curves in the image are, after all, the projection of a space curve denoting a physical edge contour. Constraints about the differential geometry of space curves therefore matter, and have been studied by Barrow and Tenenbaum [1], Huffman [25], Stevens [47], and Witkin [54]. But the full nature of grouping processes is still largely unknown. Recall that, while discussing the cafe wall illusion (Fig. 9), we produced trapezoids that eventually formed, we claimed, essential data for generating the 'tilted* lines. The mechanism responsible for grouping these local segments into global lines is unclear, but behaviourly it would appear to be the same as the one responsible for the Fraser spiral (see Fig.). Undoubtedly, there is a smoothness constraint lurking somewhere.

While curves carry clues about physical contours, or the joins between surfaces, information about surface orientation internal to these boundaries is carried by texture. Texture may be viewed as a summary of the descriptions computed by the various grouping processes, some of which may agglomerate tokens across surfaces [3]. It was first actively studied by a psychologist -- J.J. Gibson [13] — as a source of depth information, and is only now producing algorithms for inferring local surface orientation from textural cues [30]. For the scope of descriptive structures possibly underlying texture, see [28,37,57].

A third source of information about surfaces comes from stereopsis, our built-in range finder. Several models have been proposed to describe the human stereopsis model, one cooperative (see Sec. 15) [27,41),and the other sequential [42]. This sequential model, by the way, makes use of the zero-crossing information from the separate channels individually, with the course channels achieving a match before the fine ones attempt to; for computational experiments with this matcher, see [17]. The fact that both of these modeling classes work correctly, to a large extent, raises an importont question: how can two computational models, so different in structure, both solve the same problem. The answer can be found directly in terms of the levels discussed in the Introduction. An essential part of the stereo problem is determining which tokens in each eye's description correspond to the same physical event. Once this is determined, trigonometry can be used to obtain depth. The two algorithms perform this match differently. But, at a more abstract level, one can hypothesize the existence of a criterion or suitable function, underlying the match. The two algorithms are then just different ways of optimizing this criterion. The specification of this criterion still remains, as does the refinement of neurophysiological and psychophysical constraints that will permit us to decide between these two algorithmic implementations, or to derive another one (cf. [31]).

Stereo is not, however, the only way to obtain direct information about surfaces. When objects move, or an observer moves, optical flow provides another rich source.

13. OPTICAL FLOW

The discussion so far has moved from intensities to abstract descriptions derived from them, such as edges and subjective figures. However, intensities carry much more information than we have exploited so far, such as information coded into their temporal changes. Such changes give rise to a vector field, which, if could be computed exactly, would be sufficent for inferring a great deal about the motion of surfaces [35]. This, by the way, is another of Gibson's early contributions. Consider, for example, moving toward a uniform field of dots. The vector field associated with differences in position between the dots across time would point radially outward.

To get a feel for optical flow, consider the Mueller-Lyer illusion again. We have already established blurring as a causal factor in the illusion, and have noted that increases in blurring increase the subjective strength of the illusion. Now, suppose the blurring takes place in real time; that is, suppose you were to watch the image go successively in and out of focus. What kinds of object motions would you see? We have performed the experiment, and have discovered two [55]. The first of these is a motion of the insides of the convex arrows moving toward one another. It rapidly attracts one's attention, and appears to be a grouping phenomenon of the sort discussed in Sec. 10. Interestingly, the identical motion of the concave arrows is not perceived.

The second perceived motion is completely different. It is of an object moving in depth. As the figure becomes more blurred, it appears to approach the observer, and, as the figure becomes focused, it receedes. Insight into this percept can be obtained from the optical flow vector field.

Horn and Schunck [23] have derived an algorithm for computing optical flow by assuming that the brightness of each point in the image is constant within a world of smooth surfaces, constant illumination, and constant shading. This gives rise to a differential equation (the rate of change of brightness at a point is 0), which is satisfied approximately. The algorithm attempts to satisfy this equation everywhere by minimizing *an* expression for the deviation from zero in the image-based estimates. The results of applying this algorithm to successively blurred versions of the Mueller-Lyer illusion are shown in Fig. 16.

14. REPRESENTATIONS AND DATA STRUCTURES

A quick glance at the optical flow pattern in Fig. 16 immediately suggests a data structure for representing it — generalized cylinders. Indeed, this has been one of the most widely used three-dimensional representational structures since they were first introduced by Binford [4]. More recently, Marr [38] gave them additional credibility by discovering that, under certain smoothness assumptions, generalized cylinders embodied the information contained in topological contours; i.e., the properties invariant under projective transformations. It remains an open problem, however, as to whether they are the best such representation, (for heuristic arguments, see [ ]).

Part of the appeal of generalized cylinders is that they provide a coarse 3-D representation intermediate to the requisite indexing into richer, and more detailed, families of models. One of the principle lessons learned by computational models of perception so far is that such intermediate structures should proliferate throughout low-level vision as well. The idea is a classical one; Dewey understood the use of organization to deal with complexity when he designed the library's decimal system. And researchers in computer vision have stressed the importance of explicitly computing (or maintaining) representations of intensities, illumination, depth, surface orientation, etc.[1,37] Since each of these properties can be defined locally, it makes sense from a computational point of view to organize them as arrays indexed by a retinal-centered coordinate system, as Mach did. Marr has termed one such representation the "primal sketch". Whether this is a dense enough representation to encompass all of grouping still remains to be seen, as does the manner in which such abstract descriptions are represented by the wetware of the brain. Undoubtedly, many new constraints, still to be discovered, are required to decide the issue.

15, COOPERATIVE COMPUTATION AND OPTIMIZATION

At several points in our discussion we have been faced with decisions in the face of ambiguous situations; we had to decide which tokens correspond to identical physical events, we had to decide which dots correspond to the subjective circles in random dot Moire patterns, we had to decide whether our zero crossings were indicating local edge elements, and what the orientation of these edge elements might be. Most of these dec* isions have already been characterized in terms of constraints represented as criteria to be minimized, so we shall now turn our discussion to how such minima might be found. That is, having discussed representations, we shall now discuss algorithms and mechanisms.

As is well known, the field of minimization and optimization algorithms is a large, well-developed one [36], Most of these algorithms are not of immediate relevance for us, however, because, with our stated interest in vision, we must always be concerned with possible constraints from other levels. Some of the tightest of these constraints come from the implementation level; whatever algorithm we develop must be implementable on the hardware available. For the early visual system, this amounts to rather regular arrangements of sparsely interconnected units, such as neurones, each of which can perform a simple computation. Such arrangements are attractive evolutionaryly, because they make the construction of complex systems possible from simple components [48]. And they are attractive for computer vision, because they are one of the few design methodologies currently available for VLSI technology [34].

The most convenient form of sparsely-interconnected computational networks for our purposes is one in which processors are arranged spatially so that each one interacts only with its spatial neighbors. The class of computations performable in such networks may seem, at first, to be quite limited. If each node can only use data available to itself and to its neighbors, then there is no way that data from larger distances can exert any influence on the outcome of the computation. But, if we permit iteration in the network, then data can, in effect, propagate its influence over larger areas. Metaphorically, myopia is conquered by permitting neighbors to glimpse neighbors by iteration, and so on. For certain computations, the units can all operate in a lock step, parallel fashion, iterating toward a common result. This is what we shall refer to as cooperative computation.

Is it possible to perform minimization in cooperative networks? The answer is yes, as we shall now demonstrate by introducing one of the most widely studied cooperative networks — relaxation labeling processes.

16. RELAXATION LABELING PROCESSES

Consider a graph in which the nodes represent entities, and the edges indicate which entities constrain each other. Now, let a set of labels be attached to each entity, each of which represents possible interpretations for that entity. Finally, let a measure of confidence be associated with each label, indicating how likely that label is for the associated node.

Given this initial structure, the problem is to select a labeling for the graph which is most likely given a model of how the entities fit together. Perhaps the earliest such example in computer vision occured in the blocks world of convex polyhedra, in which programs attempted to label the sides of line drawings with the physical edge configuration that they were representing. Sharp local constraints existed between pairs of lines meeting at a junction within this universe, because physical edges can fit together only in certain ways [25]; for a recent review, see [2]. For example, within this universe, a line denoting two surfaces meeting to form a convex fold could not join a line denoting two surfaces forming a concave one, but it could meet one denoting an occluding surface (consider a triangular flap of cardboard bent up; the bend is the convex junction, and on either of its sides is an occluding one). The constraints in this world, then, were tables of which pairs (or triples, etc.) could form physically meaningful configurations.

Within the blocks world, this constraint information was used to restrict the search for legal labelings. The basic idea, as developed by Waltz [5U , was that labels need not be considered in the global search if they were not consistent, according to the constraint tables, with their neighbors. He thus filtered the possible labels according to the following rule: discard all labels that did not have at least one label on each of their neighbors with which they were consistent. The rule could be applied in parallel to each node-label pair, and, since the label sets change after each application, it could be iterated until no further changes took place. It is, therefore, a cooperative computation.

Although Waltz filtering was developed in this discrete, symbolic manner, it can be reformulated in optimization terms. If we view labels as either being present or absent, and constraints as either being true or false (i.e., 0 or 1, for both labels and constraints), then the effect of a label's context is to add support to the label. Since each neighboring node must have at least one consistent label, we can define the support from a node as one when the condition is satisfied, otherwise zero. The goal of Waltz filtering, then, is to maximize support for each label; otherwise, they will be discarded.

Discrete Waltz filtering can be generalized to continuous optimization by generalizing the certainty measure attached to each label from {PRESENT, ABSENT} to the continuum [0,1] , and similarly for the logical constraints. Thus Ullman's affinities, e.g., become the constraints. The problem, then, becomes one of how to extend the definition of support to this continuous case, and finally of how to maximize it in a cooperative fashion. These problems have been solved by Hummel and Zucker [26], using the tools of variational calculus. In short, their algorithm is one of gradient ascent; it operates by computing a direction for the iteration to, say, maximize an increase in the value of a functional, and then taking a step in this direction. It is interesting to note that, although this algorithm was derived for optimization purposes, it solves a much richer class of problems. It is intimately related to the algorithms for solving systems of (partial) differential equations, such as the one discussed previously for computing optical flow.

17. LABELING LINES AND LINKS

As a demonstration of the relaxation labeling process, recall the problem of interpreting the responses of oriented simple cells. In particular, let us suppose that the simple cells are trying to interpret the zero-crossing contours into locally-straight segments, called EDGES, with an associated orientation. How shall we constrain the interpretation process? Our general goal, as previously stated, is to find the labels, i.e., the EDGE segments, which were most likely to give the observed response for (at least two levels of) the observed simple cell responses. While this can be done [59], it would require more space than we have here to develop it. Rather, let us consider a simpler constraint: minimize a measure of curvature, so that curves are continued as smoothly as possible across intersections. Such a constraint could be said to implement the Gestalt law of good continuation [33], a summary of observational experience (but hardly a theory of constraint).

Given that we wish to minimize curvature, how might, the visual system obtain suitable affinities, or constraints? One approach is to use the simple cells' receptive fields not only as operators for signaling EDGES, but also as explicit representational descriptions of them. Then, constraints for orientation good continuation can be derived by using the information implicit in the arrangement of the receptive fields. That is, by overlaying receptive fields of different orientations, and then counting the amount of overlap between them, constraints between pairs of EDGES can be derived. For example, for a vertical orientation of one mask, the affinity to the other mask would drop off (roughly) exponentially as it were rotated away from the vertical. These constraints are actually proportional to the likelihood of a long vertical zero-crossing segment represented by two fine masks, in the context of a larger mask indicating a strong vertical segment in the same area (see Sec. 7) See Fig. 17. The full system of constraints, of course, must take both the orientation and the response of the larger masks into account.

18. MULTI-LEVEL COOPERATIVE SYSTEMS

As we have indicated, early visual processing seems to involve many different representations, from zero crossings to subjective figures. The issue of satisfying constraints between representational levels, as well as across them, therefore arises. And, given the underdetermined nature of many of these problems, it seems most logical that as many constraints as possible should be considered concurrently. Multi-level relaxation systems thus seem a natural solution, and in this section we shall discuss one for both labeling and linking EDGES. As before, the system presented is simplified, but, we believe, of the right sort for many forms of grouping.

The first level of our system is identical to the one just presented; it labels the response of simple-cell-like operators with assertions about oriented EDGE segments. The second level groups these segments into longer curves by labeling a relation over spatially neighboring segments as either connected or not-connected. The connected relation joins segments analogously to the way in which virtual lines joined the dots in random Moire patterns. The constraints for this second level come from the intensities; the likelihood of EDGES being connected is proportional to the similarity between their intensity profiles. The results are shown in Fig. 18 [56].

The results of these two relaxation examples should be viewed as computational experiments. They permit one to develop both the practical feel of a particular approach, and the possibility of performing mathematical analyses of it. Most importantly for the study of human perception, however, they expand the modeling vocabulary of. the visual theorist substantially. Previously, neurophysiologists have spoken of lateral inhibitory interactions between orientation detectors to overcome ambiguities in the null firing rate of neurons [5]; such mechanisms implement a limited form of enhancement. Now we have the capability of discussing implementations that achieve optimal constraint satisfaction. When one looks at the details, the required machinery is not significantly different. There is even a partial conceptual similarity as well. When the structure is clear enough everywhere, cooperative algorithms can become equivalent to local maxima selection [58]. That is, choosing local maxima everywhere may result in a global one. But in general, this is not the case.

19. SUMMARY AND CONCLUSIONS

In this essay we have trie"d to follow three paths simultaneously: an historical one, from Helmholtz and Mach to the present; an anatomical one, from the eye to the brain; and a conceptual one, from the concrete to the abstract. Because the process of vision is so complex, explanations can be put forward at many different levels, from assumptions necessary to solve abstract vision problems, to restrictions on the machinery that will implement the solution. We have illustrated constraints at each of these levels, because, when they are taken together, we believe that they demonstrate the way in which progress can be made in understanding vision in general. If constraints at any of these levels were missing, we argued, the remaining ones would be too underconstrained for viable theorizing. This was illustrated in the particular instance of cooperative algorithms, the need for which came from computational theories, but the form of which came from implementation constraints.

The essay concentrated on early perception, in part because this is the most well understood component, but more to illustrate the need for computer vision in the understanding of human perception. For it is here that the explanations would presumably be most concrete. While this is true for the physics of the eye, it does not seem to be so for as fundamental a process as the location of edges; this requires, we argued, an interpretative component to decipher the transducers' signals. By the time the visual system begins to hypothesize surfaces and volumes, the language of computations and representations seems even more necessary. While the computational level may not be strictly necessary to understand the reflex-like mechanisms in lower organisms like the fly, it would appear that both Helmholtz and Mach were right about the human visual system — inferences take place, and they are realized by mechanisms implemented in neurones. They were simply theorizing at different descriptive or explanatory, levels. We argued this point by developing the need for optimal interpretation strategies, and then showing how they could be realized, at least in principle.

Although computational theories are necessary, in the sense that we have argued, they may never be unique. There are often several different, but equivalent, ways in which the same phenomenon can be explained; soap films, for example, can be described physically as the surface with minimal area, a global characterization, or locally in terms of their differential geometry. In this essay we saw several examples from vision: lateral inhibition can be viewed as an enhancement process, or as implementing the receptive fields of center-surround cells. They may even play a role in the spatial-frequency limited channels. And simple cells were viewed teleologically as edge and line finders, and as approximations to first and second (spatial) derivatives. Finally, we encountered two separate stereo algorithms. But each of these alternative theories was instructive, illustrating, once again, how important different explanatory points of view can be.

REFERENCES

{1] Barrow, H., and Tenenbauro, J.M., Recovering intrinsic scene characteristics from images, in A. Hanson and E. Riseman, eds., Computer Vision Systems, Academic Press, New York, 1978.

[2] Barrow, H., and Tenenbaum, J.M., Computational vision, PROC. IEEE, 1981, 69, 572-595.

(3J Beck, J., Texture segregation, TR 971, Computer Science Center, University of Maryland, 1980.

[4] Binford, T., Visual perception by computer, IEEE Conf. Systems and Control, Miami, 1971.

[5] Blakemore, C., Carpenter, R., and Georgeson, M., Lateral inhibition between orientation detectors in the human visual system, NATURE, 1970, 228, 37-39.

[6] Campbell, F.W., and Robson, J.G., Applications of Fourier analysis to the visibility of gratings, J. PHYSIOL. (LONDON), 1968, 197, 551-556.

[7] Chomsky, N., Review of Skinner, LANGUAGE, 1957, 35, 26-58.

[8] Chomsky, N., Aspects of the theory of syntax, MIT Press, Cambridge, 1965.

[9] Coren, S., Subjective contour and apparent depth, PSYCH. REV., 1972, 79, 359.

[10] Coren, S., Ward, L., Porac, C, and Fraser, R., The effect of optical blur of visual-geometric illusions, BULLETIN OF THE PSYCHONOMIC SOC., 1978, 11(6), 390-392.

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

[12] Dennett, D., Brainstorms, Bradford Books, Vermont, 1978.

[13] Gibson, J., The perception of the visible world, Houghton Mifflin, Boston, 1950.

[14] Glass, L., Moire effect from random dots, NATURE, 1969, 243, 578-580.

[15] Gregory, R., The Intelligent Eye, McGraw-Hill, New York, 1970.

[16] Gregory, R., and Heard, P., Border locking and the cafe wall illusion, PERCEPTION, 1979, 8, 365-380.

[17] Crimson, W.E.L., A computer implementation of a theory of human stereo vision, AI Memo 565, MIT, 1980.

[18] Haber, R., and Herchenson, M., The Psychology of Visual Perception, Holt, Rhinehart, and Winston, New York, 1973.

[19] Helmholtz, H., von, Popular Scientific Lectures, Dover (reprint), New York.

[20] Helmholtz, H., von. Treatise on Physiological Optics, J.P.C. Southall, ed., Dover (reprint), 1962.

[21] Herskovitz, A., and Binford, T., On boundary detection, AI Memo 183, MIT, 1970.

[22] Horn, B., Understanding image intensities, ARTIFICIAL INTELLIGENCE, 8, 1977, 201-231.

[23] Horn, B., and Schunck, B., Determining optical flow, AI Memo 572, MIT, 1980.

[24] Hubel, D., and Wiesel, T., Functional architecture of macaque monkey visual cortex, PROC. ROY. SOC. (LONDON), B, 1977, 198, 1-59.

[25] Huffman, D., Impossible objects as nonsense sentences, in Machine Intelligence 6, Meltzer and Michie (eds.), Edinburgh U. P., 1971.

[26] Hummel, R. and Zucker S., On the foundations of relaxation labeling processes, TR 80-7, McGill University, Montreal, 1980.

[27] Julesz, B., Foundations of Cyclopean Perception, University of Chicago Press, 1971.

[28) Julesz, B., Textons, the elements of texture perception, and their interactions, NATURE, 1980, 290, 91-97.

[29] Kanizsa, G., Margini quasi-percettivi in campi con stimolazione omogenea, RIV. di PSICOLOGIA, 1955, 49, 7-30.

[30] Kender, J., Shape from texture, Technical Report, Computer Science, Carnegie-Mellon University, 1900.

[31] Kidd, A., Frisby- J., and Mayhew, J., Texture contours can facilitate stereopsis by initiating vergence eye movements, NATURE, 1979, 280, 829-832.

[32] Klinger, A., and Tanimoto, S. (eds.), Structured Computer Vision, Academic Press, New York, 1980.

[33) Koffka, K., Gestalt Psychology, Harcourt, Brace and World, New York, 1935.

[34) Kung, H., Systolic YLSI, Technical Report, Computer Science, Carnegie-Melion University, 1979,

[35] Longuet-Higgins, H., and Prazdny, K-, The interpretation of a moving retinal image, PROC. ROY. SOC. (LONDON), B, 1980, 385-397.

[36] Luenberger, D., Optimization by Vector Space Methods, Wiley, New York, 1969.

[37] Marr, D., Early processing of visual information, PROC. ROY. SOC, B, 1976, 275, 483-534.

[38] Marr, D., Analysis of occluding contour, PROC. ROY. SOC. (LONDON), B, 1977, 197, 441-475.

[39] Marr, D., and Hildreth, E., Theory of edge detection, PROC. ROY. SOC. (LONDON), B, 1980, 207, 187-217.

[40) Marr, D. and Poggio, T., From understanding computation to understanding neural circuitry, NEUROSCIENCE RESEARCH PROGRAM BULLETIN, 15, 1977, 470-488.

[41] Marr, D., and Poggio, T., Cooperative computation of stereo disparity, SCIENCE, 1976, 194, 283-287.

[42] Marr, D., and Poggio, T., A computational theory of human stereo vision, PROC. ROY. SOC. (LONDON), B, 1979, 204, 301-328.

[43] Moulden, B., and Renshaw, J., The Munsterberg illusion and 'irradiation[1], PERCEPTION, 1979, 8, 275-301.

[44] Nagel, E., The Structure of Science, Harcourt, Brace, and World, New York, 1951.

[45] Ratliff, F., Mach bands: Quantitative studies on neural networks in the retina, Holden Day, San Francisco, 1965.

[46] Richter, J., and Ullman, S., A model for the spatio-temporal organization of X and Y-type ganglion cells in the primate retina, AI Nemo 573, MIT, 1980.

[47] Stevens, K., Computation of locally parallel structure, BIOL., CYBER., 29, 19-26.

[48] Szentagothai, J., The neuron network of the cerebral cortex: A functional interpretation, PROC. ROY. SOC. (LONDON), B, 1978, 201, 219-248.

[49] Stevens, K., Surface perception from local analysis of texture and contour, TR 512, AI Lab, MIT, 1980.

[50] Ullman, S., The Interpretation of Visual Motion, MIT Press, Cambridge, 1979.

[51] Waltz, D., Understanding line drawings of scenes with shadows, in P. Winston (ed.), The Psychology of Computer Vision, McGraw-Hill, New York, 1975.

[52] Werblin, F.S., Functional organization of a vertebrate retina: Sharpening up in space and intensity, ANN. N. Y. ACAD. SCI., 193, 1972.

[53] Wilson, H., and Bergen, J., A four mechanism model for threshold spatial vision, VISION RESEARCH, 1979, 19, 19-32.

[54] Witkin, A., Shape from contour, Ph.D. ^he^is, Psychology, MIT, 1980.

[55] Zucker, S., Motion and the Mueller-Lyer illusion, TR 80-2, McGill University, 1980.

[56] Zucker, S., Labeling lines and links: An experiment in cooperative computation, in R. Haralick, (ed.), Consistent Labeling Problems in Pattern Recognition, Plenum, New York, 1980.

[57] Zucker, S., and Cavanagh, P., Constructive texture perception: Orientation anisotropics in discrimination, TR 80-8, McGill University, Montreal.

[58] Zucker, S., Leclerc, Y., and Mohammed, J., Relaxation and local maxima selection: Conditions for equivalence, IEEE TRANS. PAMI, 1981, 3, 117-127.

[59] Zucker, S., Parent, P., and Sander, P., Empirical compatibilities for finding lines, Technical Report, McGill University, in preparation.

[60] Zucker, S., Rosenfeld, A., and Davis, L., General purpose models: Expectations about the unexpected, PROC. IJCAI 4, Tiblisi, 1975.

Fig 1. The Mueller-Lyer illusion.

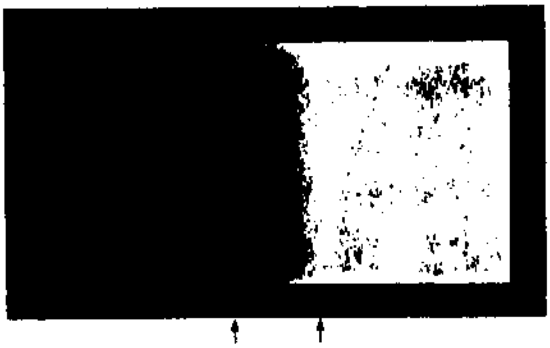

Fig 2 Mach bands; note the bright and dark lines on either side of the edge.

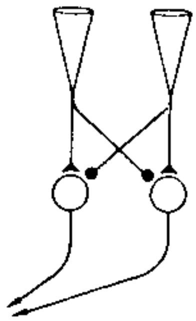

Fig 3 A mechanism for implementing lateral inhibition.

Fig 4 The structure of the simple cells' receptive fields

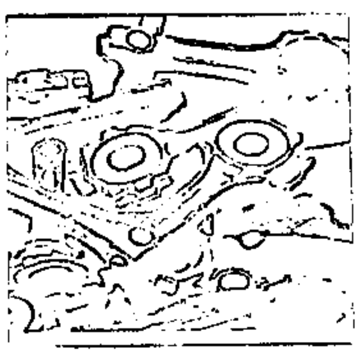

Fig 5 Illustration of the ambiguous response of a simple 'edge detector'. Note especially the weak response in the area of overlap between the parts

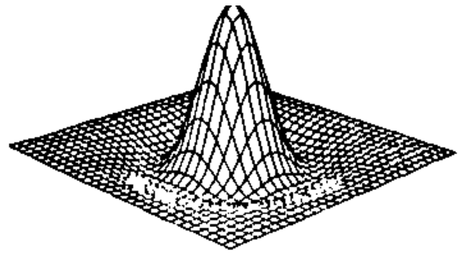

Fig 6 The Marr/Hildreth operator

Fig 7 The zero-crossings of the Marr/Hildreth operator

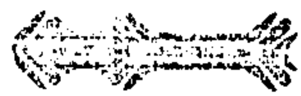

Fig 8 The zero-crossings around a blurred Mueller-Lyer figure

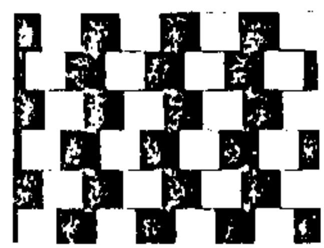

Fig 9 The cafe-wall illusion

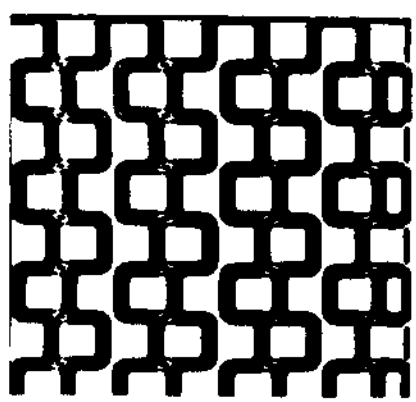

Fig 10 The zero-crossings on the cafe-wall illusion.

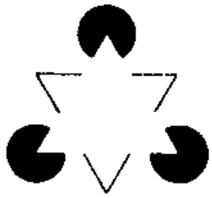

Fig 11 A Kanizsa triangle

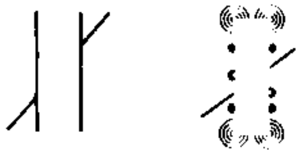

Fig 12 The Poggendorf illusion with subjective edges. Note how the line appears skewed running through the subjective rectangle.

Fig 13 A random dot Moire pattern

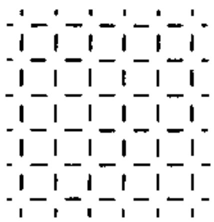

Fig 14 The Ehrenstein illusion Note that the subjective dots form swaths when it is rotated 45 degrees.

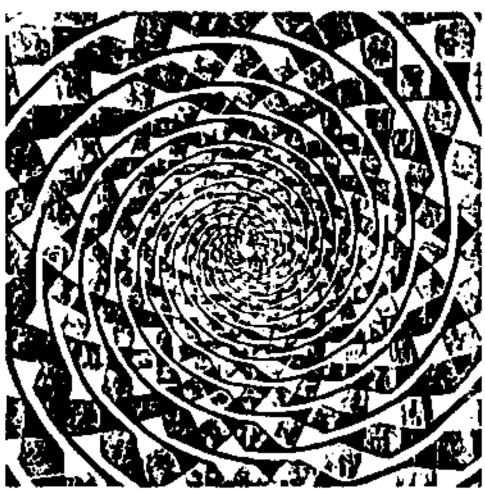

Fig 15 A Fraser spiral, which is really concentric circles

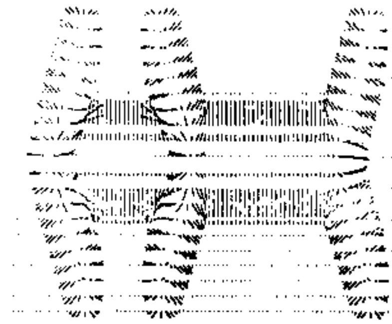

Fig 16 The optical flow vector field around the blurred Mueller-Lyer.

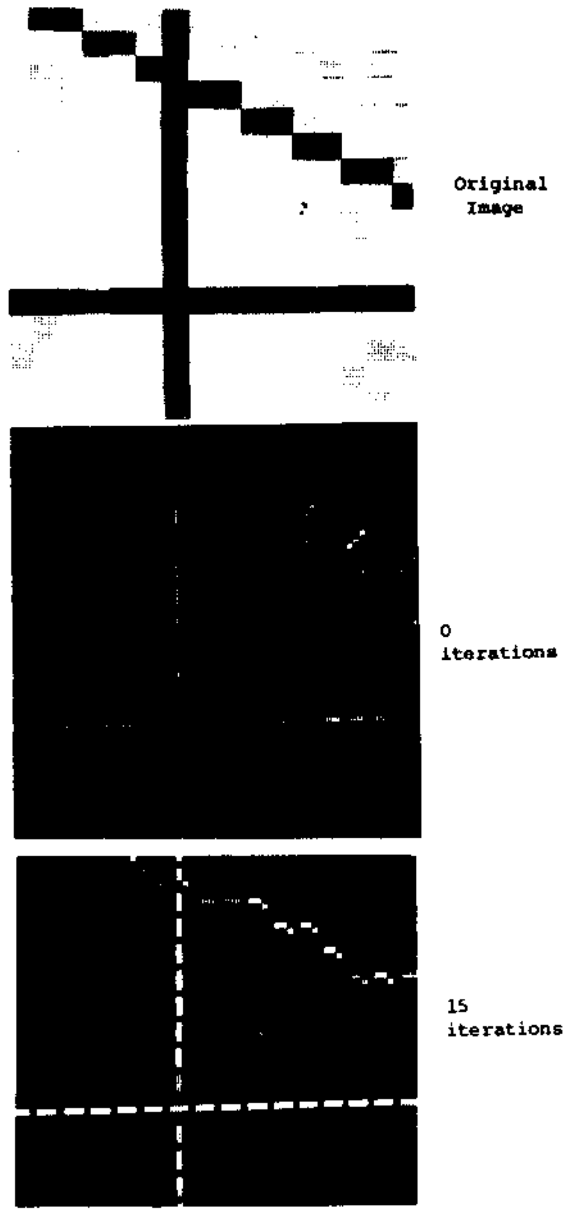

Fig 17 The results of a relaxation process for interpreting the response of simple "line" cells

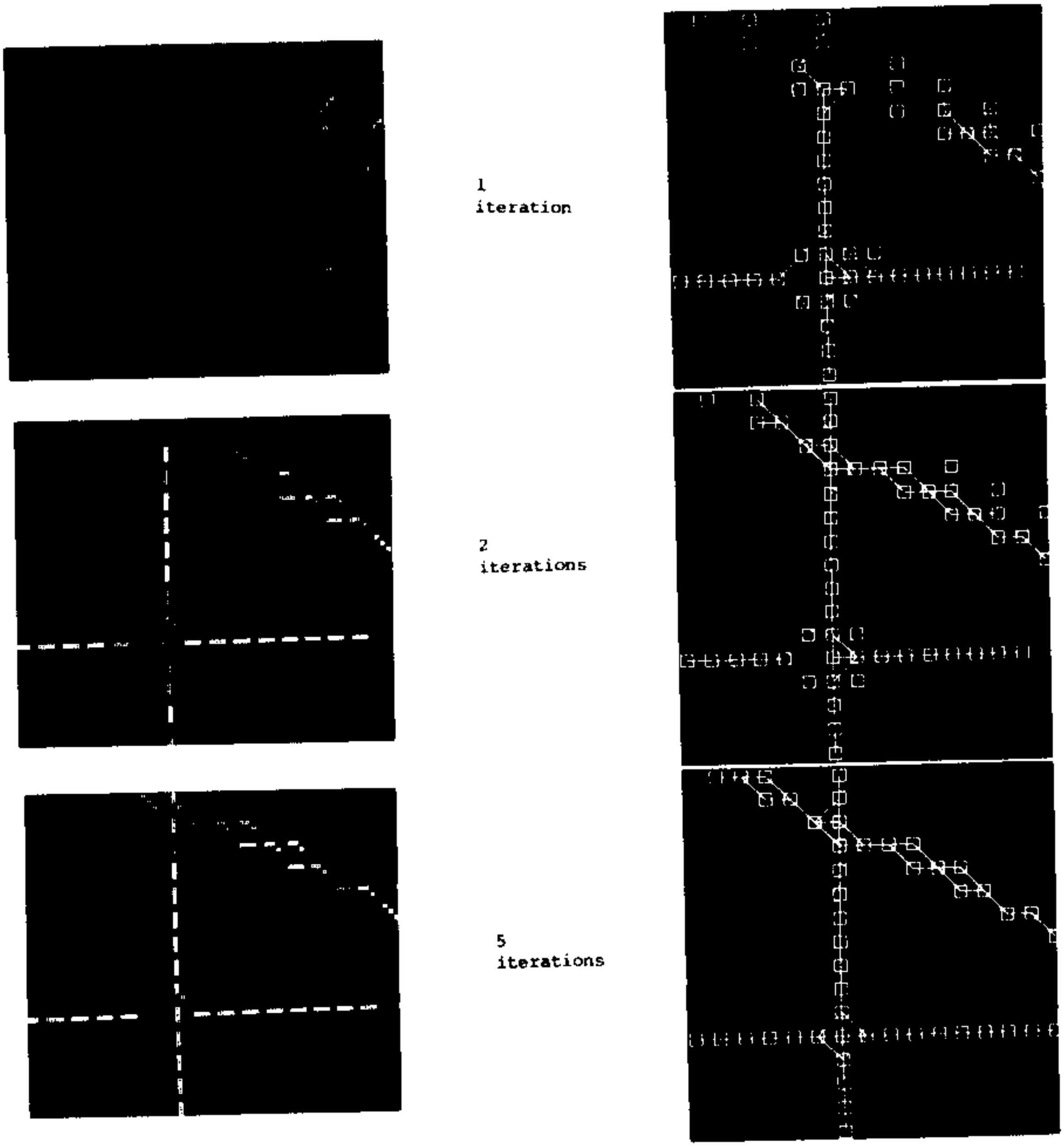

Fig 10. The results of a 2-level relaxation process for labeling and linking the simple cells' interpretations. The EDGE labels are displayed on the left, and the LINK labels on the right.

AUTHOR INDEX

*** indicates that the paper was not received in time for publication

Volume Two begins on page 395