MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 919 December, 1986
C.B.I.P. Memo No. 21

The Analysis of Visual Motion: From Computational Theory to Neuronal Mechanisms

Ellen C. Hildreth and Christof Koch

ABSTRACT: This paper reviews a number of aspects of visual motion analysis in biological systems, from a computational perspective. We illustrate the kinds of insights that have been gained through computational studies and how these observations can be integrated with experimental studies from psychology and the neurosciences, to understand the particular computations used by biological systems to analyze motion. The particular areas of motion analysis that we discuss include early motion detection and measurement, the optical flow computation, motion correspondence, the detection of motion discontinuities, and the recovery of three–dimensional structure from motion.

© Massachusetts Institute of Technology 1986

Acknowledgments. This report describes research done within the Artificial Intelligence Laboratory and the Center for Biological Information Processing (Whitaker College) at the Massachusetts Institute of Technology. Support for this research is provided by grants from the Sloan Foundation and from the Office of Naval Research, Engineering Psychology Division.

This empty page was substituted for a blank page in the original document.

INTRODUCTION

The measurement and use of visual motion is one of the most fundamental abilities of biological vision systems, serving many essential functions. For example, a sudden movement in the scene might indicate an approaching predator or a desirable prey. The rapid expansion of features in the visual field can signal an object about to collide with the observer. Discontinuities in motion often occur at the locations of object boundaries and can be used to carve up the scene into distinct objects. Motion signals provide input to centers controlling eye movements, allowing objects of interest to be tracked through the scene. Relative movement can be used to infer the three–dimensional (3–D) structure and motion of object surfaces, and the movement of the observer relative to the scene, allowing biological systems to navigate quickly and efficiently through the environment. More generally, the analysis of visual motion helps us to maintain continuity of our perception of the constantly changing environment around us.

This article reviews our current understanding of a number of aspects of visual motion analysis in biological systems, from a computational perspective. We illustrate the kinds of insights that have been gained through computational studies and how they can be integrated with experimental studies from psychology and the neurosciences, to understand the particular computations used by biological systems to analyze motion. In the remainder of this introduction, we briefly describe the computational approach to the study of vision and discuss the areas of motion analysis that are addressed in this review.

The Computational Study of Vision

One of the most important tenets underlying a *computational* approach to the study of biological vision is the belief that the brain, like a computer, can be thought of as a machine that processes information extracted from the environment, resulting in some sort of action. Like Aristotle, Galen, and Descartes before us, we often think of the brain in terms of our most successful machines, which today happen to be digital computers. We must be careful in making such an analogy, however. The electrochemical environment of neurons, their means of transmitting information, and their overall architecture is very different from that of the wires and etched crystals of semiconducting material that comprise computers. The Turing machine, a core concept of computer science, works in a discrete mode in a world determined by classical physics. Such a machine can only approximate the truly analog operations of biological hardware in a world governed by the laws of quantum physics.

Although their hardware differs greatly, both biological systems and machines can perform similar functions that rely on the same mathematical and physical principles. Thus, there exists a level of description of the tasks performed by these two systems that is independent of the underlying hardware. In order to understand how natural or artificial systems can solve problems like sensing motion or depth or manipulating the environment,

we must understand the nature of the problem — for example, whether it can be solved at all and what constraints the physical world imposes on the solution — before we can fully understand the detailed procedures used to find a solution.

A computational approach to the study of biological systems, based on the founding principles of the field of Artificial Intelligence, was elucidated by Marr and Poggio (1977; Marr, 1982). Marr was attracted to the field of Artificial Intelligence after experiencing certain limitations of other theoretical approaches to brain research in his early work on the cerebellum (Marr, 1969). Although his model for learning in the cerebellum has led to important experimental work (for example, Ito, 1984), Marr abandoned this line of research after realizing that it did not shed light on how complex motor behavior can actually be achieved.

In his later work in computational vision, Marr elucidated three distinct levels of analysis that are necessary for understanding an information processing task:

- A *computational theory* analyzes what problem is being solved and why, and investigates the natural constraints that the physical world imposes on the solution to the problem.
- An *algorithm* is a detailed step–by–step procedure that represents one method for yielding the solution indicated by the theory.
- An *implementation* is a physical realization of the algorithm by some mechanism or hardware.

These levels could suggest a prescription for conducting research on complex problems; that is, one first formulates a theory, then derives an algorithm, and lastly designs a mechanism that implements the algorithm:

$$\text{theory} \Rightarrow \text{algorithm} \Rightarrow \text{mechanism.}$$

Despite the initial success of this approach, research over the past few years has shown that computational theories, even if complemented by psychophysical experiments revealing how humans perform visual tasks, have inherent limitations in understanding the brain. In particular, the nature of the hardware can profoundly influence the type of algorithm needed to solve a particular problem. Thus, while the computational theory and properties of the hardware can often be studied independently, the algorithmic level is influenced by both. A given computation, such as the computation of stereo depth or motion, usually can be performed by several different algorithms. These algorithms depend not only on the nature of the computation itself, but also on the properties and limitations of the hardware in which the algorithm is implemented. Thus, in order to explain the functions of a visual system at its different levels, not only must the abstract, computational nature of a task be understood, but also the properties of the underlying hardware. The flow of information is therefore in both directions:

$$\text{theory} \Rightarrow \text{algorithm} \Leftarrow \text{mechanism.}$$

These observations stress the importance of integrating the results of computational studies with those of experimental studies of biological vision systems.

Other introductions to the computational approach described here can be found, for example, in Poggio (1984), Morgan (1985), Ullman (1986), and Hildreth and Hollerbach (1985). The latter review also addresses the limitations and successes of the computational approach in the area of motor control.

Other "Computational" Approaches to the Study of Biological Systems

The term *computational* is often used within the neurosciences to denote very different concepts. For example, certain neural modeling approaches that study how neuronal networks can operate and how these operations can be extrapolated to explain higher brain functions frequently are termed "computational." Examples of this include the seminal work by McCulloch and Pitts (1943) on neuronal networks, the work on perceptrons (Minsky and Papert, 1969) and parallel "connectionist" networks (Ballard, 1986), as well as Marr's original work on the cerebellum. The word "computational" in this case refers to the detailed working of specialized hardware, such as linear threshold automata, rather than to an analysis of information processing at a level independent of the underlying hardware. Similarly, connectionist theories refer directly to neuronal hardware and therefore lack the characteristics of Marr's notion of a computational theory (Koch, 1986). Although they have made important contributions to automata theory and theoretical cybernetics, we want to emphasize a distinction between these approaches and that described by Marr and Poggio (1977; Marr, 1982). It is of course essential to understand the properties of the biological hardware — neurons, dendrites, synapses, channels, etc. — in order to understand what algorithms the brain uses to analyze its environment, and a substantial fraction of this article is devoted to aspects of neuronal hardware. We believe, however, that to fully understand a complex information processing system, it is necessary first to understand the nature of the tasks the system is required to perform.

Finally, *computational* is used in yet another sense, as in *computational chemistry* or *computational biophysics*. This term generally refers to the extensive use of computers to simulate a given chemical or biophysical system, such as the reconstruction of the tertiary structure of simple proteins by using the principles of quantum physics and chemistry (Clementi, 1985) or the simulation of the electrical properties of an array of pyramidal cells in the hippocampus (Traub *et al.*, 1984). In the following pages we refer frequently to such simulations of biophysical circuits.

Overview of Visual Motion Analysis

The pattern of movement in a changing image is not given to the visual system directly, but must be inferred from the changing intensities that reach the eye. The 3–D shape of object surfaces, the locations of object boundaries, and the movement of the observer relative to the scene can in turn be inferred from the pattern of image motion. Typically, the overall analysis of motion is divided into two stages: first, the measurement of movement in the changing two–dimensional (2–D) image, and second, the use of motion measurements, for example to recover the 3–D layout of the environment. It is not clear whether motion analysis in biological systems is necessarily performed in two distinct stages, but this division has served to facilitate theoretical studies of motion analysis and to focus empirical questions for perceptual and physiological studies.

The measurement of movement can itself be divided into multiple stages and may be performed in different ways in biological systems. In the human visual system alone, motion may be measured by at least two processes, termed *short–range* and *long–range* processes (for example, Braddick, 1974, 1980). The short–range process analyzes continuous motion, or motion presented discretely but with small spatial and temporal displacements from one moment to the next. The long–range process may then analyze motion over larger spatial and temporal displacements, as in apparent motion. Evidence indicates that these two processes interact at some stage (Clatworthy and Frisby, 1973; Green and von Grünau, 1983), but initially they may be somewhat independent.

The subsequent uses of motion measurements impose different requirements on the precision and completeness with which image motion must be represented. The localization of object boundaries requires the detection of sharp changes in direction or speed of movement, but may not need a precise representation of absolute velocities everywhere. Object tracking requires knowledge of the gross translation of an object, but not information about the detailed relative movements that take place within the object. The recovery of the accurate 3–D shape of a moving object, on the other hand, appears to require a more precise and complete estimate of the local variations of motion across object surfaces. Motion analysis in the human visual system may ultimately involve the interaction of many processes, some fast but rough, others slow but more accurate, and still others that are specialized for specific tasks such as detecting object boundaries or looming motion. These processes must work together in a way that provides a versatile and robust motion analysis system.

In this review, we first address the earliest stage of motion measurement. We discuss two important theoretical models of motion detection, correlation and gradient models, and present relevant psychophysical and physiological data regarding biological motion detectors. We then discuss at length possible biophysical mechanisms that implement the computations underlying motion discrimination in retinal and cortical neurons. Later stages of motion measurement are then discussed in a subsequent section, which addresses the computation of an instantaneous 2–D velocity field, long–range motion correspondence, and

the detection of motion discontinuities. Finally, we discuss the recovery of 3–D structure from relative motion. This article is not intended as an exhaustive overview of work on motion analysis. Rather, we highlight some of the areas that exhibit fruitful interactions between computational and experimental studies. Two recent reviews of motion analysis include the surveys by Barron (1984), focusing on computational methods for deriving and interpreting optical flow, and by Nakayama (1985), focusing primarily on the psychophysics and physiology of motion.

EARLY MOTION DETECTION AND MEASUREMENT

Detecting Motion: Theory

Before motion can be used to reconstruct the 3–D structure of objects, the visual system must first reliably detect and measure relative motion in the 2–D image. What types of schemes have been proposed for this initial detection, how are these schemes related and what are their computational properties? The most general property of any motion discrimination system is that the underlying operation must be nonlinear. As first noted by Poggio and Reichardt (1973), no linear operation can extract the direction of motion of a moving stimulus. The schemes proposed for motion detection fall broadly into two classes: (1) correlation–like schemes (Hassenstein and Reichardt, 1956; Poggio and Reichardt, 1973; van Santen and Sperling, 1984) and (2) gradient schemes (Fennema and Thompson, 1979; Horn and Schunck, 1981; Marr and Ullman, 1981). As we shall see, most biological motion detection schemes cannot reliably measure velocity even for one–dimensional motions, because their output typically depends on contrast and on a mixture of velocity and spatial structure of the moving pattern (Reichardt, Poggio and Hausen, 1983).

CORRELATION MODELS The best known motion detection scheme is based on research done over the last thirty years on movement perception in insects. On the basis of open– and closed–loop experiments performed first on the beetle, *Chlorophanus*, and later on the fruitfly, *Drosophila*, and the housefly, *Musca Domestica*, a number of researchers, most notably W. Reichardt, were led to the following conclusions regarding motion discrimination in insects (Hassenstein and Reichardt, 1956; Varju and Reichardt, 1967; Götz, 1968, 1972; Reichardt, 1969; Poggio and Reichardt, 1976; Reichardt and Guo, 1986):

i) A sequence of two light stimuli impinging on adjacent receptors is the elementary event that evokes an optomotor response.

ii) The relation between the stimulus input to these two receptors and the optomotor output follows the rule of algebraic sign multiplication. For instance, stimulating receptor 1 with alternating dark to light changes and receptor 2 with light to dark transitions leads to a turning response of the insect opposite to the direction of stimulus successions, while dark to light transitions presented to both receptors elicits a turning reponse in the direction of the stimulus succession.

iii) The strength of the optomotor response is proportional to the product of the two stimuli.

On the basis of these experimental conclusions, a minimum mathematical model of motion perception in insects was formulated. Figure 1a shows a modified version of this *correlation* model. The image is sampled by a receptor with a point–like receptive field. The input to the receptor can thus be described by $I(t)$. The output of the receptor is subsequently passed through a linear high–pass filter, removing steady–state components of

the output of the receptor, before being multiplied with a low– or band–pass filtered signal from a neighboring receptor. Thus, at this stage the signal strength is given by:

$$R(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W(t_1, t_2) I(t - t_1) I(t - t_2) dt_1 dt_2$$

where $W(t_1, t_2)$ represents the lumped transfer–function for the different filters. Subsequently, the output of the multiplication operation is integrated over time. A little analysis will show that the output of this stage is equivalent to the autocorrelation of the input function $I(t)$. Let us assume that the low–pass filter actually corresponds to a fixed delay $\delta t > 0$. We are then essentially multiplying a linearly transformed version of $I(t)$ with itself, but shifted by the total amount $\Delta t = \delta t + \Delta x / v$ (where $\Delta x > 0$ is the spacing between the receptors and v the velocity of the stimulus), and integrating the resulting function over time. For a range of negative velocities, i.e. movement from the right to the left, Δt will be very small and the final output of this subunit will be large. For positive velocities, that is for movements in the opposite direction, the two functions $I(t)$ and $I(t + \Delta t)$ are out of synchrony and their product, integrated over time, will be small. The output of this subunit is then subtracted from the output of the complementary subunit to yield the total detector response. It follows that if the output of the right subunit exceeds the output of the left subunit, the detector response is positive, indicating rightward motion; likewise, if the output of the left subunit exceeds the output of the right subunit, detector response is negative, indicating leftward motion. This theoretical model has a number of properties that can be tested experimentally. Two of the most interesting are *phase invariance* and *spatial aliasing* (for an overview see Reichardt, 1969).

Imagine a light pattern consisting of a number of superimposed sinusoidal gratings of different spatial frequencies. Because the process of autocorrelation, i.e. multiplication and subsequent integration, destroys all of the information that is inherent to the specification of the phases of the gratings, the output of the motion detectors is invariant to any changes in the phase relations of the sinusoidal gratings. Because any pattern $I(t)$ can be decomposed into its Fourier components, it follows that this class of motion detectors does not sense the relative position of the Fourier components. This important result has been tested and confirmed in experiments with the beetle, *Chlorophanus*, the fruitfly, *Drosophila*, and with *Musca* by evaluation of the time–averaged optomotor reactions to the angular motion of a fixed pattern painted on the inside of a drum. Moreover, the total time–averaged response is simply given by the sum of the time–averaged response to the individual Fourier components (Poggio and Reichardt, 1973). Figure 2a shows the angular distribution of the brightness of two distinct patterns, obtained by superposition of the different Fourier components. These patterns only differ with respect to their phase relations. Yet the fruitfly reacts equally to motions of the two patterns (Götz, 1972).

For any particular sinewave grating, the temporal phase difference between the two inputs to the multiplication will depend on the distance between its input channels, Δx,

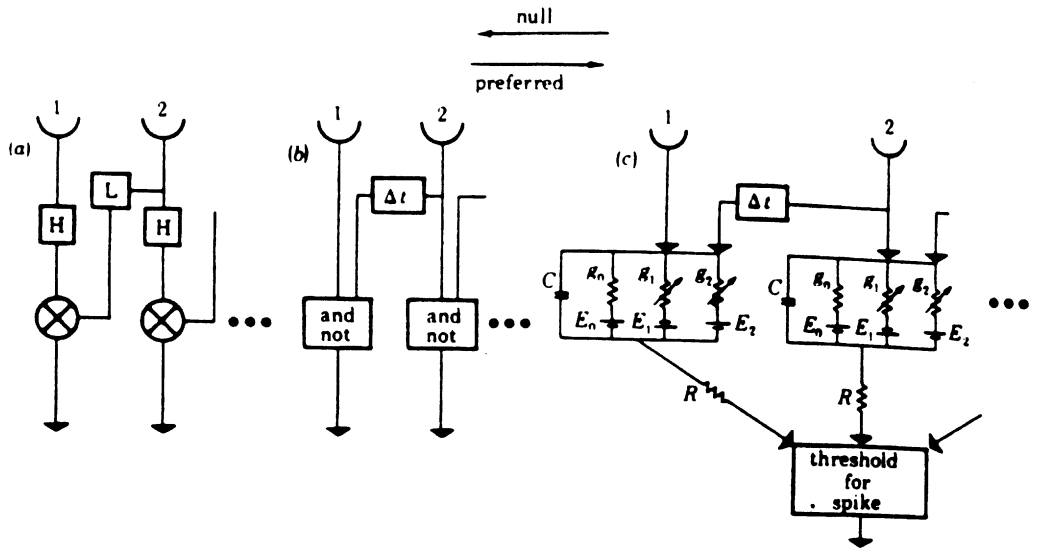

Figure 1. (a) A direction selective subunit of the correlation model of Hassenstein and Reichardt (1956) as modified by Kirschfeld (1972). The two inputs are multiplied after low pass filtering with different time constants. If an average operation is made on the output, the overall operation is equivalent to cross–correlation of the two inputs. Subsequently, the time–averaged response of this subunit is subtracted from the response of a similar but mirror–symmetric subunit to yield the final movement sensitive response. (b) The functional scheme proposed by Barlow and Levick (1965) to account for direction selectivity in the rabbit retina. A pure delay Δt is not necessary: a low pass filtering operation is sufficient. (c) The equivalent electrical circuit of the synaptic interaction assumed to underlie direction selectivity as proposed by Torre and Poggio (1978). The interaction implemented by the circuit is of the type $g_1 - \alpha g_1 g_2$, where g_1 and g_2 represent the excitatory and inhibitory synaptic inputs. From Torre and Poggio (1978).

and on the spatial wavelength λ of the sinewave grating used. The original correlation model displays spatial aliasing: if one changes the spatial period of the grating, but not its direction of motion, the sign of the detector response reverses, indicating an incorrect motion. Within the wavelength region $\lambda > 2\Delta x$, the moving sinusoidal pattern is resolved by the receptor system as the number of samples received per period λ at any time is greater than or equal to two. If, however, $\lambda < 2\Delta x$, optimal resolution of the periodic pattern breaks down, because less than two samples per wave length of the pattern are observed (see also Shannon's sampling theorem) and the detector signals the incorrect direction for $\Delta x < \lambda < 2\Delta x$ (Figure 2b). This inversion of apparent motion does occur in various insects and has been used to determine the grating constant of the receptor spacing (Reichardt, 1969).

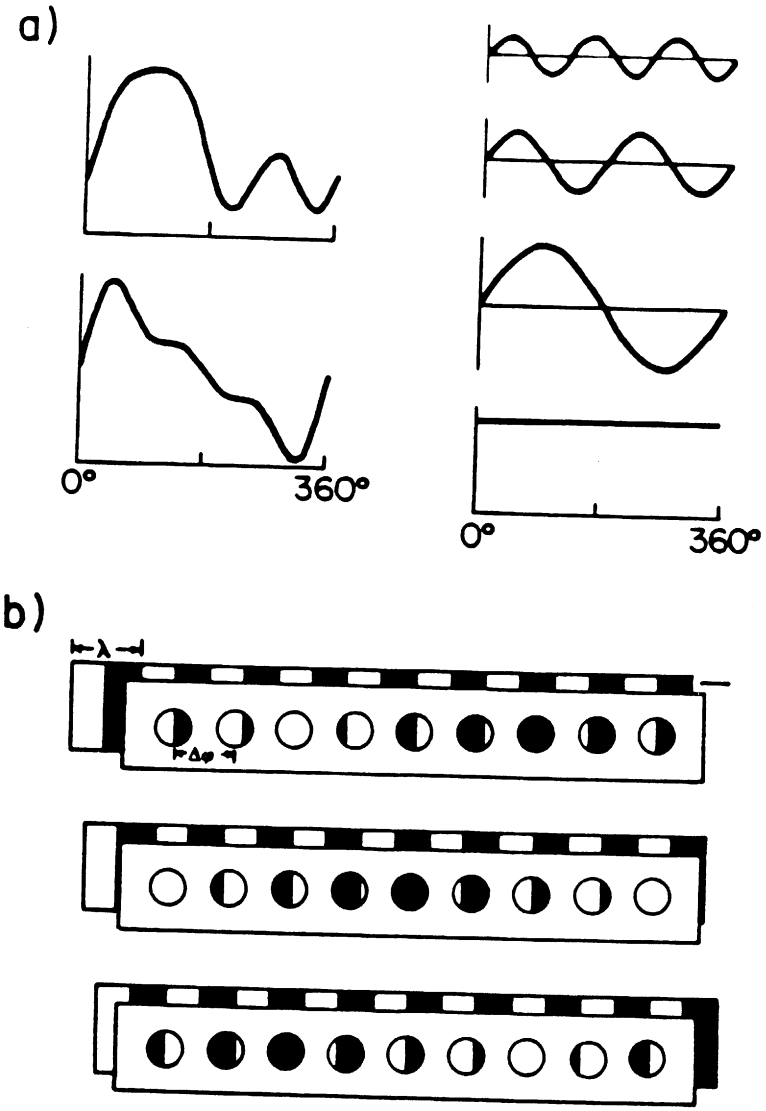

Figure 2. Two experimental predictions of the correlation model. (a) Phase invariance: The left part of the figure shows two different light patterns received by a photoreceptor at different angular positions of the environment. Both distributions contain the same set of Fourier components shown in the right part of the figure, but with different phases. However, insects like the housefly, the fruitfly or the beetle respond with the same optomotor reaction to both patterns. At the moment, it is not known whether direction selective cells in the mammalian visual system show phase invariance. (b) Inverse motion perception: interference phenomena in the insect eye elicited by a moving pattern with a comparatively small spatial wavelength. When the distance $\Delta\phi$ between input channels in the insect's eye is between one–half and one spatial period λ of the pattern of excitation, the correlation model signals the incorrect direction of motion. The insect is compelled to follow this apparent motion in the direction opposite to the "true" direction of motion. Redrawn from Götz (1972).

This property of the original correlation model can be avoided by replacing the point–shaped receptive field of the receptor in the original Reichardt model with a spatial–dependent receptive field of finite extent (Fermi and Reichardt, 1963; Götz, 1965; Reichardt, Poggio and Hausen, 1983; van Santen and Sperling, 1984, 1985). Van Santen and Sperling show how to choose the receptive field in their *elaborated Reichardt detector* so that the sign of the detector output is correct for any drifting sinewave grating. van Santen and Sperling (1985) showed that the elaborated Reichardt model is fully equivalent to two recently proposed models of human motion detection: an elaborated version of the motion detector of Watson and Ahumada (1985) and the "spatiotemporal energy" motion detector of Adelson and Bergen (1985). These and similar models characterized by a multiplication–like nonlinearity are all equivalent to the correlation model (Poggio and Reichardt, 1973).

GRADIENT MODELS Gradient schemes rely on the relationship between the spatial and temporal gradients of image intensity. In the case of the one–dimensional movement of an intensity profile $I(x,t)$ over a small displacement dx in time dt, the temporal derivative of image intensity $I_t \approx (I(x,t+dt) - I(x,t))/dt$ and the spatial derivative of the intensity $I_x \approx (I(x+dx,t) - I(x,t))/dx$ are related by

$$v = \frac{dx}{dt} = -\frac{I_t}{I_x}$$

where v is the velocity of the pattern. This method was originally proposed by Limb and Murphy (1975) and later extended by Fennema and Thompson (1979). The approach carries over to the 2–D case (Horn and Schunck, 1981). Here, however, due to a fundamental limitation in the measurement process, termed the *aperture problem* (discussed later), only the component of the velocity in the direction of the brightness gradient can be measured. If we assume that the motion measurement process occurs along an edge, only the velocity component at right angles to the edge can be recovered. It is given by

$$v = -\frac{I_t}{\sqrt{I_x^2 + I_y^2}}$$

where I_x, I_y are the spatial derivatives in the x and y directions. This equation is strictly only correct for rigid, translating patterns, with no rotation, seen under orthographic projection (Schunck, 1984). For sufficiently small temporal and spatial displacements dx, dy, and dt, however, the equation approximates the correct one. Gradient schemes suffer from the disadvantage that they require computation of the derivatives of the intensity values, an operation that is extremely sensitive to noise.

A quantized version of the gradient scheme was proposed by Marr and Ullman (1981). This model operates on locations in the images where the light intensity changes significantly. Marr and Hildreth's analysis (1980) showed that zero–crossings, that is locations where the Laplacian of the image is zero, correspond closely to intensity edges in the original

image. Marr and Ullman track the motion of zero–crossings in the following way. An edge detector S of the Marr and Hildreth type signals the absence or presence of a zero–crossing at location x. This detector has two variants, one for transitions from dark to light (termed a light–on edge) and one for light to dark transitions (light–off edge). A second type of detector, termed a T unit, samples the temporal derivative of the intensity in approximately the same patch of the visual field as the edge–detecting unit. One version of this unit, T^+, only signals when the temporal derivative is positive, that is when a light–on edge has moved to the left or a light–off edge moves to the right, whereas T^- only responds to a reduction in light intensity. Combining the output of an S and a T unit conjunctively yields a set of detectors signaling the left (or rightward) motion of light–on or light–off edges. Marr and Ullman tentatively identify the edge detecting S units with sustained X–like On– or Off–center cells and the T^+ and T^- units with transient Y–like On– or Off–center cells. Computer experiments on some images have shown that this gradient scheme can recover motion information from image sequences. Note that their model, different from other gradient schemes, does not provide an estimate of the local velocity, but only its sign, that is the direction of motion, although some measure of velocity could be extracted.[1]

MOTION PRIMITIVES What are the primitives used to detect and measure motion, and at what stage in the analysis of the image does the detection of motion take place? For instance, are the initial measurements of the light intensity in the photoreceptors taken as primitives, or are the measurements extracted after the filtering and smoothing of the visual input at the stage of the retinal ganglion cells or even cortical cells? Finally, more symbolic primitives such as zero–crossings, edges, and line segments or even endpoints, corners, breaks, local deformities of objects, or discontinuities in line orientation could also be used. The advantage of matching more symbolic tokens, such as zero–crossings, across the image is that these tokens mark interesting points in an image, for instance locations where the image intensity changes most. Tokens are generally far more stable to changes and noise in the illumination than the original intensities or some filtered version of them. Moreover, because tokens presumably are sparsely distributed in the image, far fewer points must be matched and ambiguities can be avoided. If, however, large areas of the image contain no tokens, for instance if the light intensity changes little, these areas will not have any motion measurements assigned initially (these areas could be filled in later on). A further disadvantage of symbolic primitives is that they must be unambigously identified before they can be matched, thus preventing an early computation of motion.

For the visual system of the fly, the experimental evidence suggests that the primitive is simply some measure of local intensity flux (Reichardt *et al.*, 1983). For the short–range motion system, Hildreth (1984) discusses the evidence that motion measurement may rely on the detection of the movement of features such as zero–crossings, or some similar measure

[1] It can be shown formally that for small contrast amplitudes, the correlation–model and the gradient scheme are equivalent (T. Poggio, personal communication).

operating on the smoothed intensity values, and that the limits on spatial and temporal displacements observed empirically in the short–range motion system are the consequence of the limited spatial and temporal extent of the initial filtering (see also Marr and Ullman, 1981). Much more work needs be done, however, before the question of the primitives used by the motion system can be answered.

Detecting Motion: Psychophysics

Both gradient and correlation schemes are local, involving only limited parts of the visual scene, and are therefore likely to provide a dominant input to the short–range process, which appears to operate on motion restricted to a spatial range of up to $10' - 15'$ minutes of visual arc and an interstimulus interval less than $80 - 100$ *msec.* (Braddick, 1974, 1980). Because these separations in space and time are small, establishing correspondence between items in consecutive images is considerably easier than in the long–range process (see next section). Finally, the short–range process is assumed to operate directly on the light intensities, filtered intensities or on edges or zero–crossings. Interestingly, color seems to provide little if any input to the short–range process (Ramachandran and Gregory, 1978). In the following, we discuss the (limited) human perception evidence that has been used to discriminate between the various models of motion computation discussed above.

One of the main properties of the Reichardt correlation model is that its output responds not only to pattern velocity but also to structural properties of the pattern contrast. This property allows the motion detector to be used as pattern discriminator, at least in flies (Reichardt *et al.*, 1983; Reichardt and Guo, 1986). Specifically, it can be shown (for instance, in Poggio and Reichardt, 1973, 1976) that the time–averaged response of the correlation subunit depends on the ratio, for each spatial Fourier component, of the pattern velocity v and the spatial wavelength λ of the stimulus used. Thus wavelength and velocity trade off against each other and, as a consequence, the correlation model cannot reliably measure the speed of movement. This property, first confirmed with behavioral experiments for the fly, *Musca Domestica* (Eckert, 1973), also seems to extend to the human visual system. If subjects fixate a point while square or sinusoidal gratings of variable spatial wavelength are moved past the fixation point at various speeds, their perception of velocity depends linearly on both the speed and spatial frequency of the gratings (Diener *et al.*, 1976; Burr and Ross, 1982). These experiments seem consistent with a multiplicative–like second–order correlation model.

A striking prediction of the original Reichardt model is motion inversion: if the wavelength of the stimulus pattern is less than twice the separation between input channels, the insect will perceive motion in the direction opposite to the true direction of motion (Reichardt, 1969; Götz, 1972). Because humans, in contrast to insects, generally do not seem to show spatial aliasing, the point–like receptive field assumption of the original correlation model must be abandoned in favor of extended receptive fields (Fermi and Reichardt, 1963).

It can then be shown that motion reversal can be prevented (see, for instance, van Santen and Sperling, 1984). Van Santen and Sperling (1984, 1985) test this "elaborated Reichardt" model with a number of psychophysical experiments. In particular, by varying the contrast of neighboring vertically oriented bars moving in a horizontal direction, they show that the total response of the subject depends on the product of the amplitudes of the two bars, a finding that offers support for the multiplication principle.

Psychophysical evidence in favor of the gradient scheme is presented by Moulden and Begg (1984). In one particularly ingenious experiment, they show polarity and direction–specific effects on motion discrimination in response to adaptation to a non–moving, spatially homogenous stimulus, and provide evidence for channels tuned to detect an increase or decrease in the light intensity (Marr's and Ullman's (1981) T^+ and T^- units).

Thus, the current psychophysical evidence does not decisively favor one particular theory.

Detecting Motion: Circuitry and Biophysics

Having described some of the algorithms proposed to underlie motion detection, we now discuss in more detail the biophysical mechanisms that may be used for motion detection. Numerous nerve cells in the visual system of both invertebrates and vertebrates respond differentially to motion. Moving a visual stimulus, say a dark bar on a light background, in the *preferred* direction elicits a vigorous response from the cell, whereas movement in the opposite direction, termed the *null direction*, yields no significant response. *Direction selective* cells, first described in the frog's retina in a classical paper by Maturana *et al.* (1960), have subsequently been identified in the third optic ganglion of the house fly (for a review of the extensive literature see Hausen, 1982a,b), in the retina of pigeons (Maturana and Frenk, 1963; Holden, 1977), rabbits (Barlow, Hill and Levick, 1964; Barlow and Levick, 1965), ground squirrels (Michael, 1966), and cats (Stone and Fabian, 1966; Cleland and Levick, 1974), and in the visual cortex of both cats and monkeys (Hubel and Wiesel, 1959, 1962; Schiller, Finlay and Volman, 1976; Orban, Kennedy and Maes, 1981). Analyzing these cells afford us the opportunity to study the elementary biophysical events underlying a well characterized but nonlinear (that is, nontrivial) operation in single nerve cells.

In most mammals, except cats and primates, the first cells that seem to discriminate the direction of motion are the retinal ganglion cells. Thus, in the rabbit's retina approximately one–quarter of the ganglion cells can be described as direction selective. In the cat retina, however, less than 1% of the physiological identified ganglion cells are direction selective (Rodieck, 1979) while no such cells have been reported in the monkey's retina.[2] Because neither cells in the A and A1 layers of the lateral geniculate nucleus (LGN) of the cat nor cells in the magno– and parvo–cellular layers in the monkey are strongly direction selective,

[2]Due to the inevitable electrode bias, this does not necessarily imply such cells do not exist in the primate retina.

the appearance of substantial numbers of direction selective neurons in the primary cortex of both animals strongly suggests that this property arises first in the cortex.

COMPUTING THE DIRECTION OF MOTION IN THE RETINA: *Early Experiments* Barlow and Levick (1965) systematically explored directional selectivity in the retina of the rabbit by using extracellular recordings. About 20% of the ganglion cells in the visual streak give both On and Off responses to stationary, flashed stimuli and are direction selective for moving stimuli. These cells therefore compute the direction of motion independent of the contrast of the stimulus (i.e. dark stimulus on a light background or vice versa). A smaller proportion of ganglion cells ($\approx$ 7%) are direction selective and of the On–type, that is, they respond only to light–on edges. These cells project to the accessory optic system in the midbrain and are believed to be crucial for the control of the optokinetic nystagmus (Oyster, Takahashi and Collewijn, 1972) and image stabilization (Simpson, 1984). Off–type direction selective cells have neither been reported in the rabbit or cat, although they are found in the turtle. Two important conclusions can be drawn from Barlow and Levick's (1965) report. First, inhibition is crucial for direction selectivity. On the basis of this evidence Barlow and Levick proposed that sequence discrimination is based upon a scheme whereby the response to the null direction is vetoed by appropriate neighboring inputs (the AND NOT gate in Figure 1b). Directionality is achieved by an asymmetric delay — or by a low pass filter — between excitatory and inhibitory channels from the photoreceptors to the ganglion cell. This model can be considered as an instance of the Reichardt correlation model. Second, this veto operation must occur within small independent subunits distributed throughout the receptive field of the cell, because movement of a bar over 0.25° to 0.5° elicits a direction selective response (whereas the whole receptive field subtends 4.5°; Barlow and Levick, 1965). Thus, the site of the veto operation is extensively replicated throughout the receptive field of the direction selective cell. Confirming evidence for the critical role of inhibition comes from experiments in which inhibition is blocked with pharmacological agents (Caldwell, Daw and Wyatt, 1978; Ariel and Daw, 1982; Ariel and Adolph, 1985), a situation that results in an equal response for both preferred and null directions (see below).

A Biophysical Model We can now ask how this operation is implemented at the level of the hardware, i.e. at the level of retinal cells. Torre and Poggio (1978) proposed a specific biophysical mechanism implementing the neural equivalent of a veto operation.

When two neighboring regions of a dendritic tree experience simultaneous conductance changes, induced by synaptic inputs, the resulting postsynaptic potential is generally not the sum of the potentials generated by each synapse alone; that is, synaptic inputs may interact in a highly nonlinear fashion. This is particularly true for an inhibitory synaptic input that increases the membrane conductance with an associated ionic battery that reverses at, or very near, the resting potential E_{rest} of the cell. Activating this type of

inhibition, called *silent* or *shunting* inhibition, is similar to opening a hole in the membrane: its effect is only noticed if the intracellular potential is substantially different from E_{rest}. Torre and Poggio (1978) showed in a lumped electrical model of the membrane of the cell that silent inhibition can cancel effectively the excitatory postsynaptic potential (EPSP) induced by an excitatory synapse without hyperpolarizing the membrane. Moreover, for small synaptic conductance inputs the interaction between excitation and silent inhibition is multiplication–like, thereby approximating the nonlinear operation underlying the correlation–scheme (see legend to Figure 1c). Pairs of excitatory and inhibitory synapses distributed throughout the dendritic tree may compute the direction of motion at many independent sites throughout the receptive field of the cell, in agreement with the physiological data. Because nonlinearity of the interaction is an essential requirement of this scheme, Torre and Poggio suggest that the optimal location for excitation and inhibition are fine distal dendrites or spines of the direction selective ganglion cell.

Because this analysis left out the precise conditions required to produce effective and specific nonlinear interactions in a dendritic tree, Koch, Poggio and Torre (1982, 1983) used one–dimensional cable theory to analyze the interaction between time–varying excitatory and inhibitory synaptic inputs in a morphologically characterized cat retinal ganglion cell (of the δ type; see Boycott and Wässle, 1974). They were able to prove rigorously in the case of steady state synaptic conductance inputs, that in a passive and branched dendritic tree the most effective location for silent inhibition (most effective in terms of reducing an EPSP) must always be on the *direct path* between the location of the excitatory synapse and the soma.

Detailed biophysical simulations of highly branched and passive neurons show that this on–the–path condition can be quite specific. If the amplitude of the inhibitory conductance change is above a critical value, inhibition can reduce excitation by as much as a factor of 10, as long as inhibition is located between the excitatory synapse and the soma. Inhibition more than about $10\mu m$ behind excitation or on a neighboring branch 10 or $20\mu m$ off the direct path is ineffective in reducing excitation significantly. This specificity in terms of spatial positioning of excitatory and inhibitory synapses carries over into the temporal domain. For maximal effect, inhibition must last at least as long as excitation and the inhibitory and excitatory conductance changes must occur nearly synchronously (Koch *et al.*, 1983; Segev and Parnas, 1983). Finally, the on–the–path condition is also valid in the presence of action potentials: in order for silent inhibition to block the propagation of a spike past a branching point, it must be located at most $5\mu m$ from the branch point (O'Donnell, Koch and Poggio, 1985). Because such a precise mapping imposes stringent conditions on the specificity of the positioning of synapses during development of the retinal circuitry, one simple developmental rule to follow is that a pair of excitatory and inhibitory inputs originating from interacting photoreceptors should contact the ganglion cell dendrite close to one another.

The specificity of silent inhibition contrasts with the action of a hyperpolarizing synap-

tic input (i.e. a conductance change with an associated battery below E_{rest}). In this case, the interaction between excitation and inhibition will be much more linear, that is, the inhibitory synapse will reduce the EPSP generated by the excitatory synapse by an amount roughly proportional to the inhibitory conductance change with less regard to the relative spatial positioning of excitatory and inhibitory synapses (Koch *et al.*, 1982; O'Donnell, Koch and Poggio, 1985; Koch and Poggio, 1986).

Critical Predictions of the Model How does the model fare against experimental evidence? The following lists some of the most important predictions:

- On–Off direction selective cells receive distinct excitatory and inhibitory synaptic inputs. The reversal potential of the inhibitory input is close to the resting potential of the cell (probably acting via a $GABA_A$ receptor).
- Bicucculin should abolish direction selectivity.
- Inhibitory synapses are not more distal to the soma than excitatory synapses.
- Direction selectivity is computed at many independent sites in the dendritic tree before spike initiation at the axonal hillock.
- The direction selective cell should show a δ–like morphology, with a highly branched, bistratified dendritic tree with small diameter dendrites or possibly spines.
- On–Off direction selective cells are expected to show little interaction between a dark bar/spot and light bar/spot moving in opposite directions within the receptive field.

Currently, the main support for this hypothesis derives from intracellular recordings in retinal ganglion cells from the turtle (Marchiafava, 1979) and the bullfrog (Watanabe and Murakami, 1984). Moving a spot or bar in the preferred direction gives rise to a somatic EPSP with superimposed action potentials whereas null direction stimulation results in a smaller EPSP without a hyperpolarization. The reduced somatic EPSP in the null direction appears to be caused by an inhibitory process that increases the membrane conductance with an associated reversal potential at or very near the resting potential of the cell. This silent inhibition is revealed by injecting a steady–state depolarizing current into the soma, giving rise to a hyperpolarization (see Figure 3). Preliminary evidence from rabbit ganglion cells indicates the presence of a similar inhibitory input (F. Amthor, personal communication).

Within the last few years, two groups have determined the morphological structure of On–Off direction selective ganglion cells. Using a fluorescent stain, Jensen and DeVoe (1983) visualized these cells in the turtle retina, and Amthor, Oyster and Takahashi (1984), used horseradish peroxidase (HRP) in the rabbit. The overall morphology of these cells is similar in the two species. Rabbit direction selective ganglion cells have several distinct features that allow visual identification on purely morphological grounds (Figure 4a). (1)

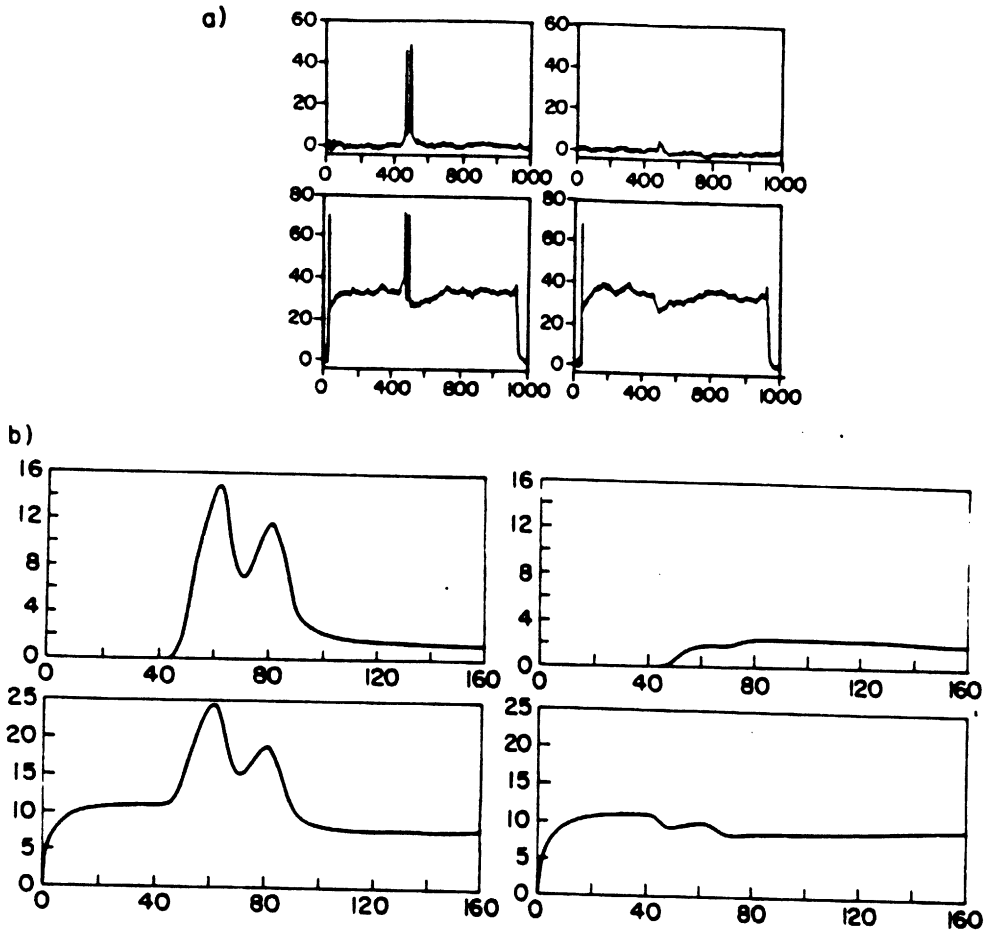

Figure 3. (a) The effect of intracellular current injection upon the photoresponse in an intracellular recorded direction selective turtle ganglion cell. The response in the preferred and null directions are shown in in the left and right part of (a). The lower record shows the photoresponse while $0.23nA$ current was being injected into the soma. Adapted from Marchifava (1979). (b) Simulated intracellular potential at the soma of the reconstructed rabbit On–Off direction selective ganglion cell shown in Figure 4, assuming a purely passive membrane. The two distinct peaks correspond to the leading edge, receiving On input, and the trailing edge, receiving Off input. In the bottom half, a step current of $0.091nA$ was being injected into the soma. Preferred direction is left and null direction right. From Koch *et al.* (1986b).

These cells have two levels of dendritic ramification. This observation is consistent with studies that have divided the inner plexiform layer into On and Off laminae (Famiglietti and Kolb, 1976). (2) The dendritic branches of the direction selective cells are of very small

diameter relative to other rabbit ganglion cells. Moreover, the dendrites carry spines or spine–like structures. (3) The dendritic branching pattern is quite complex, with dendrites forming apparent loops. Note that although the cell drawn in Figure 4a has an asymmetric placement of the soma with respect to the dendritic tree, preferred and null direction do not appear to be predictable from the gross dendritic morphology of these cells. Thus, the morphology of direction selective cells agrees well with previous predictions (Koch *et al.*, 1982).

In order to model massive synaptic input to a direction selective ganglion cell, the passive electrical properties of the anatomically reconstructed cell shown in Figure 4a was simulated on the basis of one–dimensional cable theory (O'Donnell, Koch and Poggio, 1986; Koch, Torre and Poggio, 1986). The computation of the voltages is carried out by a circuit simulation program, SPICE, first applied to biophysical circuit modeling by Segev *et al.* (1985). Figure 3 shows the resulting somatic depolarization in the absence and in the presence of a depolarizing current step injected at the soma, in comparison with experimental records obtained from turtle ganglion cells (Marchiafava, 1979). The intracellular potential can also be displayed in color throughout the entire cell (O'Donnell, Koch and Poggio, 1986; Koch *et al.*, 1986).

Presynaptic Circuitry How much do we know about the origin and properties of the excitatory and inhibitory inputs to direction selective cells? Considerable evidence implicates acetylcholine (ACh) as the excitatory neurotransmitter underlying direction selectivity in the rabbit retina (Ariel and Adolph, 1985). If all synaptic transmission in the perfused retina is blocked by pharmacological manipulation of the bathing medium, On–Off direction selective cells can be driven by direct application of ACh, thus implying that these cells are the postsynaptic target for cholinergic synapses. Ariel and Daw (1982) found that upon application of physostigmine, a drug that inhibits the hydrolysis of ACh after it has bound to the postsynaptic membrane, ganglion cells lose their ability to discriminate motion. Other properties like speed and size specificity and radial grating inhibition do not seem to be affected. This result may at first seem paradoxical, because physostigmine increases the effectiveness of ACh. One simple explanation is that this increased effectiveness during null direction serves to overcome the inhibition and to initiate action potentials at the soma. In turtle retina, simliar experiments yield similar results (Ariel and Adolph, 1985).

Recently, Masland and colleagues (Masland, Mills and Cassidy, 1984; Tauchi and Masland, 1984) identified two unique populations of cholinergic amacrine cells. In the rabbit retina, the only cells synthesizing and releasing ACh are two groups of amacrine cells distributed in the On and Off layers. Using radioactive labeled ACh, Masland, Mills and Cassidy demonstrated that these two subtypes of amacrine cells release ACh transiently either at the onset (cells in the On layer) or at the offset of light (cells in the Off layer). Because the cells have a unique morphology reminiscent of fireworks, they are called *starburst*

amacrine cells. These cells appear to be presynaptic to bistratified ganglion cells, with the morphological attributes of the direction selective cells of Amthor *et al.*, (1984).

The inhibitory input for motion discrimination is believed to be mediated by the neurotransmitter, γ–aminobutyric acid (GABA). Caldwell *et al.* (1978) and Ariel and Daw (1982) infused picrotoxin, a potent antagonist of GABA, into the rabbit retina. Within minutes after the start of drug infusion, the response of direction selective cells in the null direction increased dramatically, so that the cell became equally responsive to movement in both directions. A few minutes after drug infusion was discontinued, the cell again became direction selective. In the turtle retina, direct application of ACh leads to spontaneous firing in direction selective cells, during blockage of synaptic transmission via a low calcium concentration and EGTA (Ariel and Adolph, 1985). This ACh–induced spike activity can be suppressed by GABA, thus indicating that both ACh and GABA receptors must coexist on the membrane of turtle direction selective ganglion cells. In the rat retina, the only cells staining for glutamic acid decarboxylase (GAD; the rate–limiting enzyme for the synthesis of GABA) are amacrine cells (Vaughn *et al.*, 1981). These cells make synapses onto processes of bipolar, amacrine and ganglion cells in descending order of frequency.

Thus, at least in the turtle and rabbit retina, the excitatory and the inhibitory inputs to direction selective ganglion cells appear to derive from cholinergic and GABAergic amacrine cells. This finding does not exclude, however, direct input from bipolar cells that may be responsible, for instance, for the center–surround organization of direction selective cells.

Alternative Models What are the alternative models for the neuronal operations underlying motion discrimination? If one assumes that direction selectivity is first expressed at the level of the ganglion cells then the experimental evidence of Barlow and Levick (1965) and the intracellular recordings of Marchiafava (1979) and Watanabe and Murakami (1984) in conjunction with the pharmacology (Ariel and Adolph, 1985) argue in favor of our postsynaptic, silent inhibition scheme. Although both Werblin (1970) and Marchiafava (1979) have failed to record direction selective responses in bipolar or amacrine cells, the possibility that the critical computations occur presynaptic to the ganglion cell cannot be excluded. Indeed, DeVoe and his collaborators (DeVoe, Guy and Criswell, 1985) have recorded from direction selective amacrine and bipolar cells in the retina of the turtle. Their evidence points toward an alternative or coexistent presynaptic site for the critical computation underlying direction selectivity in the turtle. A second piece of evidence favoring a presynaptic arrangement is the influence of GABA on ACh. GABA inhibits the light evoked release of ACh in the rabbit retina (Massey and Neal, 1979; see Figure 4).

Other classes of presynaptic models for motion discrimination have been proposed (Dowling, 1979; Koch and Poggio, 1986; Koch *et al.*, 1986): Because GABAergic processes synapse onto bipolar, amacrine, and ganglion cells, the site of the critical computation underlying direction selectivity could either be a bipolar cell exciting the starburst amacrine cell or the starburst amacrine cell itself. Starburst amacrine cells have dendrites that are

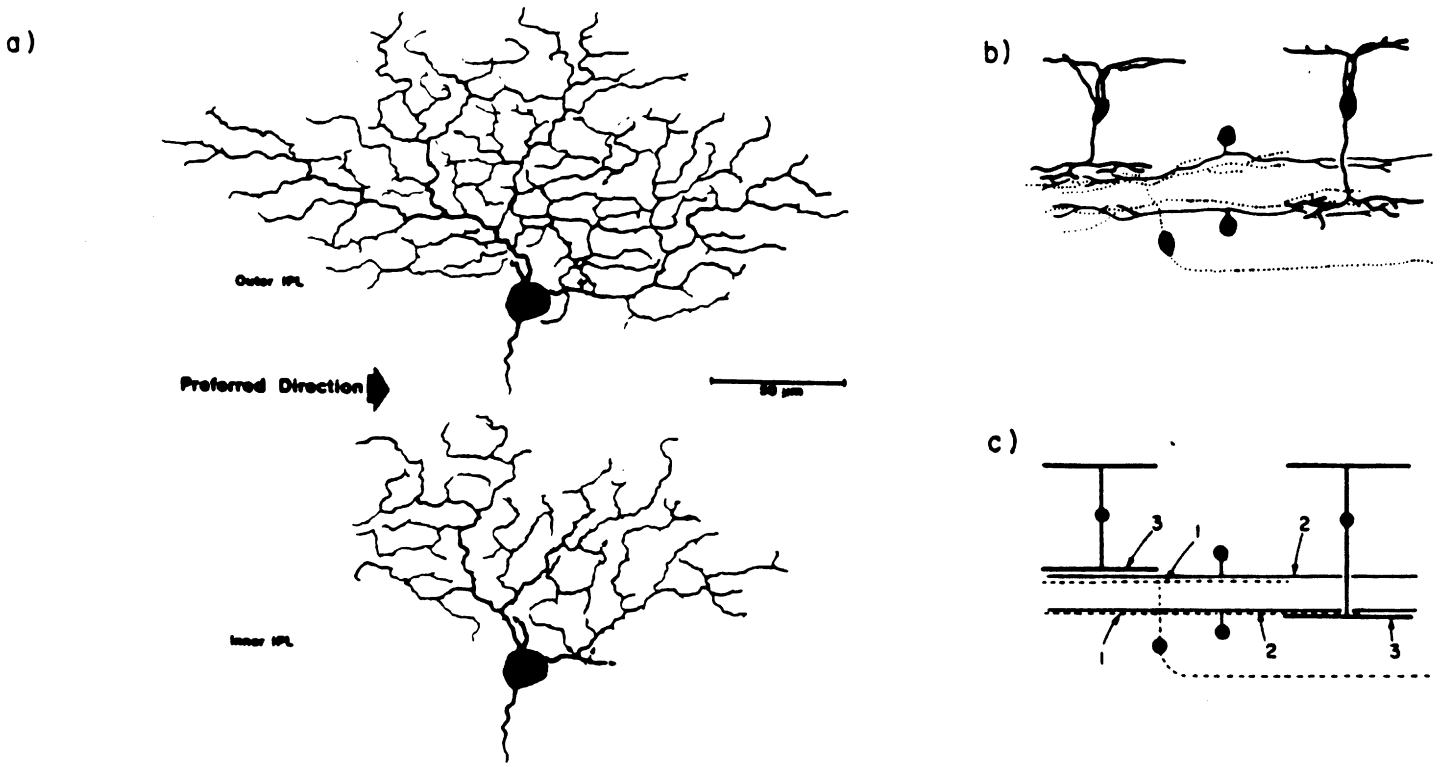

Figure 4. (a) Camera lucida drawing of an HRP–injected On–Off direction selective cell in the visual streak of the rabbit retina. The dendritic fields have been drawn in two parts: "outer" refers to the part of the inner plexiform layer (IPL) closest to the inner nuclear layer, where the cells of the Off pathway make synaptic connections, while "inner" is the layer closest to the ganglion cell layer where the the On pathway is connected. There are no obvious asymmetries in the cell that are correlated with the preferred direction. Adapted from Amthor *et al.* (1984). (b) A simplified schematic of the excitatory pathway from the outer plexiform layer (OPL) to the On–Off direction selective ganglion cell in the rabbit. Depolarizing (On) and Hyperpolarizing (Off) bipolar cells convey the visual information from the OPL to the On or Off part of the IPL. Here they most likely synapse either directly, possibly using glutamate or aspartate as excitatory neurotransmitter, or indirectly, via other amacrine cells, onto the cholinergic starburst amacrine cells. These amacrine cells feed in turn directly onto the bistratified On–Off ganglion cells. (c) Possible sites for the computations underlying motion discrimination. GABAergic amacrine cells can veto the excitatory pathway either at the level of the ganglion cell (1), at the starburst amacrine cells (2) or bipolar cells (3). Current evidence seems to favor site (1). The On and Off pathways are segregated up to the cell body of the On–Off direction selective cell. From Koch *et al.* (1986b).

probably decoupled from each other and the soma (Miller and Bloomfield, 1983). Only the distal–most portion of the dendrites give rise to conventional chemical synaptic output, whereas the bipolar and amacrine cell input is distributed throughout the cell (Famiglietti, 1983). Thus, each dendrite may behave from an electrical point of view as an independent subunit, acting as the morphological basis of Barlow and Levick's subunits (1965). At least two biophysical mechanisms could underlie direction selectivity: 1) the AND NOT veto scheme, now implemented at the level of bipolar or amacrine cells, or 2) a linear interaction between an excitatory synapse and a hyperpolarizing synapse followed by synaptic rectification (Koch and Poggio, 1986). In this case, the nonlinearity essential for direction selectivity (Poggio and Reichardt, 1973) would be implemented by a synaptic transduction mechanism that only allows transmission of depolarizing events. For these presynaptic models, the release of neurotransmitter, whether from the bipolar onto the amacrine cell or from the amacrine onto the ganglion cell, would in itself be direction selective.

We would like to point out that both pre– and postsynaptic models may turn out to be correct. For instance, the direction selective bipolar and amacrine cells recorded by DeVoe *et al.* (1985) have a smaller velocity range than direction selective ganglion cells. Thus, a rough estimate of the direction of a moving stimuli could be computed at the level of bipolar/amacrine cells while ganglion cells would perform similar but finer measurements.

COMPUTING MOTION IN THE VISUAL CORTEX Much more work has been done on the biophysical mechanisms underlying direction selectivity in the retina than in the cortex. Therefore, our discussion of cortical mechanisms will necessarily be brief. As mentioned above, cells in the primary visual cortex of cats and primates are likely to compute the direction of motion, because the geniculate input shows no evidence of direction selectivity. Moreover, if the inhibition mediated by local interneurons is removed by application of bicuculline, an antagonist of GABA (Sillito, 1977; Sillito *et al.*, 1980), direction selectivity of cortical cells is severly reduced or abolished.[3]. This experiment, similar to Ariel and Daw's experiment in the retina (1982), underscores the importance of inhibition for direction discrimination.

An extension of the veto mechanism outlined above has been proposed to underlie direction selectivity in the visual cortex (Poggio, 1982; Koch and Poggio, 1985). The basic idea is as follows: A single LGN On–center neuron (or a row of such cells) excites a cell in area V1 whenever a light–on stimulus falls within its receptive field center. A neighboring On–center LGN cell reduces the activity of the cortical neuron by a delayed silent inhibition. Because it is unlikely that LGN cells have an inhibitory effect on their postsynaptic targets, the second geniculate cell excites an interneuron, possibly in layer 4c, which in turn inhibits the direction selective cell. This seems plausible in light of the fact that direction selective

[3]The crucial nature of inhibition for motion discrimination seems to be well preserved across species. Injecting picrotoxin, a GABA antagonist, into the third optic ganglion of the blowfly, *Calliphora Erythrocephala*, abolishes motion discrimination at both the cellular and the behavioral level (Bülthoff and Bülthoff, 1986).

cells in the primate cortex first occur one synapse beyond layer 4c, i.e. in layer 4b (Dow, 1974). If the silent inhibition is located either on the direct path between excitation and the soma or very near the excitatory synapse, it will effectively veto excitation in the null direction. Adding a similar but inverted circuit constructed of geniculate Off–center neurons endows our cortical neuron with direction selectivity for both light–on and light–off edges moving in the same direction — the most common type of direction selective cell (the S2 cell of Schiller *et al.*, 1976). These Off–center neurons, whose receptive field overlap with the fields of their On–center counterparts, map onto a different part of the dendritic tree of the direction selective cortical cell. This prediction, i.e. that direction selectivity for light–on and light–off edges results from the independent convergence from the geniculate, is supported by experiments done by Schiller (1982) in the monkey and by Sherk and Horton (1984) in the cat, using the pharmacological agent APB. APB infusion into the retina reversibly blocks the On pathway at the level of the retinal outer plexiform layer and eliminates the response of the cortical direction selective cell to light edges while leaving the response to dark edges intact.

One intriguing possibility is that dendritic spines might be the specialized sites for the synaptic veto operation to take place. $5 - 20\%$ of spines on cortical cells have been reported to carry symmetrical and asymmetrical synaptic profiles on the same spine (see, for instance, Jones and Powell, 1969; Sloper and Powell, 1979). Such an arrangement can be used to perform a highly tuned temporal discrimination operation, essentially without influencing the rest of the neuron (Koch and Poggio, 1983). With a fast excitatory and a much slower inhibitory conductance change simultaneously occuring on the same spine, inhibition will effectively veto excitation if it sets in before the start of excitation (null direction). Activating the inhibition some fraction of a millisecond after the start of excitation will not influence excitation to any significant degree (preferred direction).

Very recently, Saito *et al.* (1986) have proposed that a more complex type of motion discrimination, namely cells in the superior temporal sulcus of the macaque monkey that respond only to either expanding or contracting size change of patterns or to rotation of patterns in one direction, is based on local synaptic veto operations occurring at numerous independent sites in the dendritic tree of these cells. Finally, it has recently been proposed that the synaptic veto mechanism underlies the direction selective response to cells in the somatosensory cortex of awake monkeys when wheels with surface gratings are rolled over their skin (Warren *et al.*, 1986).

Open Questions

Evidence still seems inadequate to present a clear–cut case for either the correlation or the gradient scheme for human motion discrimination. In fact, both schemes may be used by the human visual system. Because the physiological and behavioral data seems to indicate the validity of the correlation model for invertebrates and a large class of vertebrates

it may be hypothesized that the Reichardt correlation model, possibly implemented via the synaptic veto mechanisms of Torre and Poggio (1978), is used in the primate retina to endow some cells with direction selectivity. These cells, which cannot exist in very large numbers, project to the superior colliculus and from there possibly to the cortex. Motion discrimination in the cortex could be computed *de nouveau* within simple cells in the striate cortex by use of a different scheme, for instance the gradient scheme of Marr and Ullman (1981) or the implementation of the correlation model based on AND–NOT type of synaptic logic (Poggio, 1982; Koch and Poggio, 1985). Psychophysical experiments may thus be unable to separate these two models. Clearly, what is needed are physiological experiments, e.g. single cell recordings using some of the psychophysical paradigms, to identify unambiguously the algorithm used to detect motion.

In the section on the biophysical mechanisms possibly underlying direction selectivity, we discussed the strengths and limitations of simulating biophysical hardware, that is neurons. Modeling the events underlying a particular computation at the cellular level can give us valuable insights into the elementary operations underlying information processing at the single cell level, operations that cannot be resolved by present experimental techniques because of the small distances and the brief times involved. Thus, the major justification of this approach is its *predictive power.* Computer simulations should provide a number of detailed predictions that can be evaluated experimentally. Ideally, these predictions should be nontrivial and should rule out alternative explanations.

The major drawback of this approach is that any model is only as good as its fundamental assumptions. For instance, most of the studies addressing properties of the synaptic veto operation assume the absence of any significant electrical nonlinearity, such as dendritic spikes. This proviso must be taken into account when comparing experiments with the theoretical predictions, and the effect of this simplifying assumption on the mechanism in question must be carefully assessed (see O'Donnell, Koch and Poggio, 1985). Biophysical models of the electrical properties of neurons depend on a host of parameters and assumptions, most of which are poorly characterized. Thus, the foremost requirement of any detailed model of cellular properties must be *robustness*: varying some parameter, such as the membrane resistance, by a given amount should not lead to drastically changed properties in the circuit except if some critical, and specified, value has been crossed. Ideally, one would like to show that some particular behavior occurs for a broad range of parameters and is not overly sensitive to any one of them. If the model's behavior varies dramatically by changing a parameter, for instance the location of inhibition with respect to excitation, this dependency should be studied carefully, because it may lead to interesting predictions. Any model that overly constrains a parameter seems biologically unreasonable.

THE INTEGRATION OF EARLY MOTION MEASUREMENTS

Solving the Aperture Problem

The motion detection mechanisms described in the preceding section provide only partial information about the 2–D pattern of movement in the changing image, due to a problem often referred to as the *aperture problem* (Wallach, 1976; Fennema and Thompson, 1979; Burt and Sperling, 1981; Horn and Schunck, 1981; Marr and Ullman, 1981; Adelson and Movshon, 1982). Consider the computation of the projected 2–D velocity field for the rotating wireframe object illustrated in Figure 5a. Suppose that the movement of features on the object were first detected by using operations that examine only a limited area of the image, such as those performed by neural mechanisms with spatially limited receptive fields. The information provided by such mechanisms is illustrated in Figure 5b. The extended edge **E** moves across the image, and its movement is observed through a window defined by the circular aperture **A**. Through this window, it is only possible to observe the movement of the edge in the direction perpendicular to its orientation. The component of motion along the orientation of the edge is invisible through this limited aperture. Thus it is not possible to distinguish between motions in the directions **b**, **c** and **d**. This failure to distinguish between motions when the object is viewed through a small window has been referred to as the aperture problem, and is inherent in any motion detection operation that examines only a limited area of the image.

As a consequence of the aperture problem, the measurement of motion in the changing image requires two stages of analysis: the first stage measures components of motion in the direction perpendicular to image features; the second combines these components of motion to compute the full 2–D pattern of movement in the image. In Figure 5c, a circle undergoes pure translation to the right. The arrows along the contour represent the perpendicular components of velocity that can be measured directly from the changing image. These component measurements each provide some constraint on the possible motion of the circle, as illustrated in Figure 5d. The bold vector **v** represents the local perpendicular component of motion at a particular location in the image. The possible true motions at that location are given by the set of velocity vectors whose endpoint lies along the line l oriented perpendicular to the vector **v**. Examples of possible true velocities are indicated by the dotted vectors. The movement of image features such as corners or small spots can be measured directly. In general, however, the first measurements of movement provide only partial information about the true movement of features in the image, and must be combined to compute the full pattern of 2–D motion.

The measurement of movement is difficult because in theory, there are infinitely many patterns of motion that are consistent with a given changing image. For example, in Figure 5e, the contour **C** rotates, translates and deforms to yield the contour **C′** at some later time. The true motion of the point **p** is ambiguous. Additional constraint is required to identify a

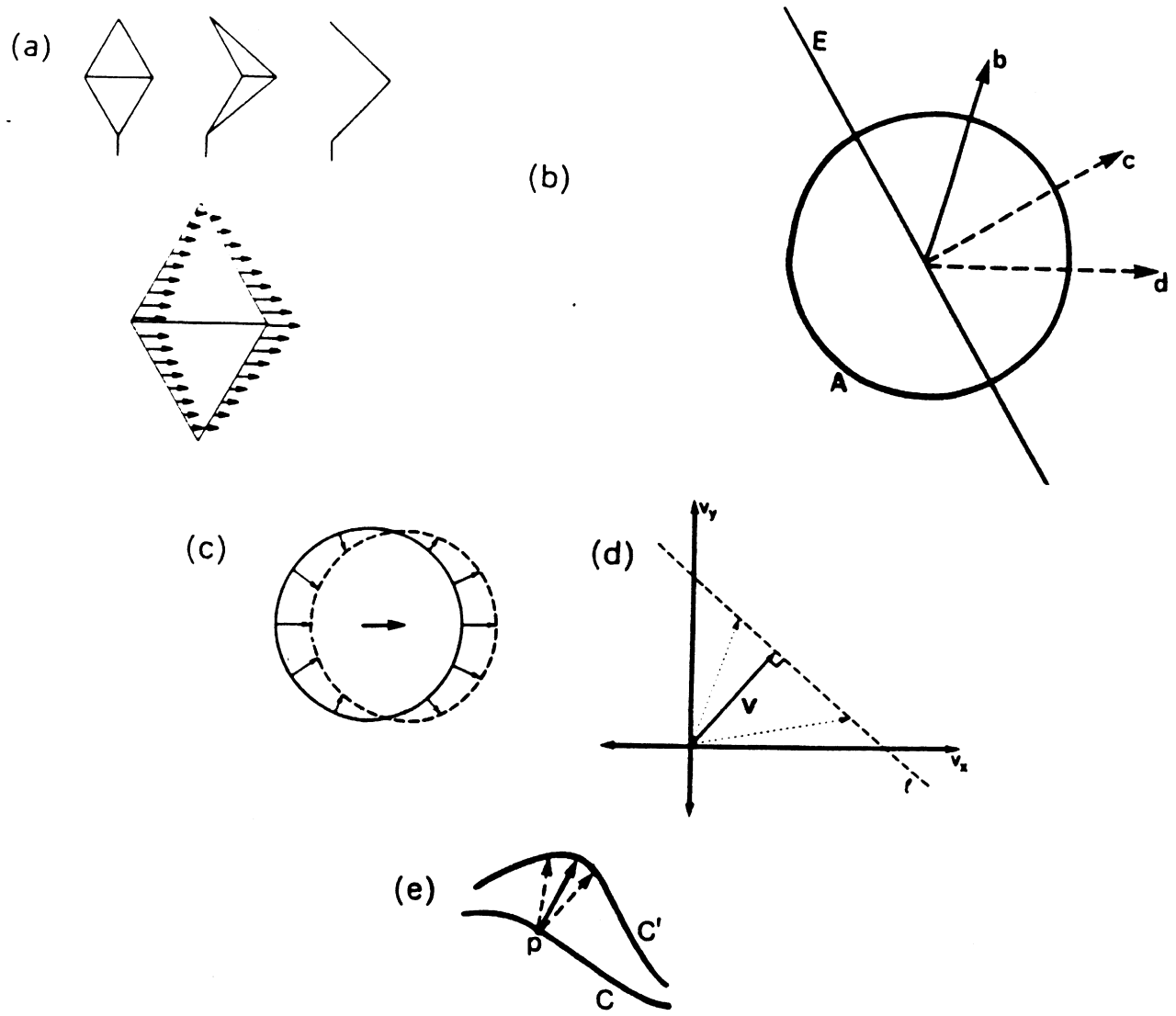

Figure 5. The aperture problem in motion measurement. (a) On the top are three views of a wireframe object undergoing rotation around a central vertical axis. On the bottom, the arrows along the contours of the object represent the instantaneous velocity field at one position in the object's trajectory. For simplicity, an orthographic projection is used. (b) An operation that views the moving edge **E** through the local aperture **A** can compute only the component of motion **c** in the direction perpendicular to the orientation of the edge. The true motion of the edge is ambiguous. (c) The circle undergoes pure translation to the right; the arrows represent the perpendicular components of velocity that can be measured from the changing image. (d) The curve **C** rotates, translates, and deforms over time to yield the curve **C'**. The velocity of the point **p** is ambiguous. (e) The vector **v** represents the perpendicular component of velocity at some location in the image. The true velocity at that location must project to the line *l* perpendicular to **v**; examples are shown with dotted arrows.

unique solution.[4] It should also be noted that in general, it may not be possible to recover the 2–D projection of the true 3–D field of motions of points in space, from the changing image intensities. Factors such as changing illumination, specularities, and shadows can generate patterns of optical flow in the image that do not correspond to the real movement of surface features. The additional constraint used to measure image motion can yield at best a solution that is most plausible from a physical standpoint.

Many physical assumptions could provide the additional constraint needed to compute a unique pattern of image motion. One possibility is the assumption of pure translation. That is, it is assumed that velocity is constant over small areas of the image. This assumption has been used both in computer vision studies and in biological models of motion measurement (for example, Lappin and Bell, 1976; Pantle and Picciano, 1976; Fennema and Thompson, 1979; Anstis, 1980; Marr and Ullman, 1981; Thompson and Barnard, 1981; Adelson and Movshon, 1982). Methods that assume pure translation may be used to detect sudden movements or to track objects across the visual field. These tasks may require only a rough estimate of the overall translation of objects across the image. Tasks such as the recovery of 3–D structure from motion require a more detailed measurement of relative motion in the image. The analysis of variations in motion such as those illustrated in Figure 5a requires the use of a more general physical assumption.

Davis, Wu and Sun (1983) proposed a computational method for solving the aperture problem that assumes that the pattern of image motion can be approximated locally by rigid motion in the image plane. In more recent studies, the local image motions have been modeled by second–order polynomials in the image coordinates (Wohn, 1984; Waxman and Wohn, 1985; Wohn and Waxman, 1985; Waxman, 1986). This approach implicitly assumes that the image locally represents the projection of a quadric surface patch in motion.

Other computational studies have assumed that velocity varies smoothly across the image (Horn and Schunck, 1981; Hildreth, 1984; Nagel, 1984; Nagel and Enkelmann, 1984, 1986; Anandan and Weiss, 1985; Scott, 1986). The assumption rests on the principle that physical surfaces are generally smooth; that is, variations in the structure of a surface are usually small, compared with the distance of the surface from the viewer. When surfaces move, nearby points tend to move with similar velocities. There exist discontinuities in movement at object boundaries, but most of the image is the projection of relatively smooth surfaces. Thus, it is natural to assume that image velocities vary smoothly over most of the visual field. A unique pattern of movement can be obtained by computing a velocity field that is consistent with the changing image and has the least amount of variation possible. In other words, a pattern of movement is derived for which nearby points in the image move with velocities that are as similar as possible.

The use of the smoothness assumption for motion measurement has several important attributes from a computational perspective. First, it allows general motion to be analyzed.

[4]Like many early vision problems, the measurement of motion is an *ill–posed problem*, as formalized by Hadamard (Poggio, Torre and Koch, 1985). A body of mathematics known as *regularization theory* may serve to unify the solution to many ill–posed problems in vision.

Surfaces can be rigid or nonrigid, undergoing any movement in space. It is always possible to compute a projected velocity field that preserves the variation in the local pattern of movement. Second, the smoothness assumption can be embodied in the motion measurement computation in a way that guarantees a unique solution (Hildreth, 1984). Third, the velocity field of least variation can be computed straightforwardly, using standard computer algorithms (Horn and Schunck, 1981; Hildreth, 1984; Nagel and Enkelmann, 1984, 1986; Anandan and Weiss, 1985), as well as simple analog resistive networks (Poggio, Torre and Koch, 1985; Poggio and Koch, 1985).

From the perspective of perceptual psychology, one can ask whether the human visual system derives patterns of movement that are consistent with those predicted by a computation that uses the smoothness assumption. In particular, one can ask whether an incorrect pattern of motion is perceived in situations in which a computer algorithm also fails. The method for computing the velocity field suggested by Hildreth (1984) is guaranteed to yield the correct solution for at least two classes of motion: (1) pure translation, and (2) general motion (translation and rotation) of rigid 3–D objects whose edges are essentially straight. For example, the computation yields the correct velocity field for the moving object of Figure 5a. For smooth curves undergoing rotation, this computation sometimes yields a solution that differs from the correct projected velocity field. The human visual system also appears to derive an incorrect perception of motion in these situations (Hildreth, 1984). Comparisons between the results of computational models and perceptual behavior have so far been only qualitative, however. Open questions remain regarding whether the human visual system maintains a local representation of the pattern of image motions, and whether perceived motion is quantitatively consistent with that expected from a computation that uses the smoothness constraint. Perceptual studies indicate that when visual patterns undergo uniform translation, human observers can match velocity directions to a resolution of about 1° (Levinson and Sekuler, 1976; Nakayama and Silverman, 1983). It is not yet known, however, whether such precision of velocity direction is also obtained when the velocity field varies continuously across the visual field.

A second issue that arises regarding the solution to the aperture problem is the question of whether the early motion measurements are integrated over 2–D areas of the image or along connected contours such as edges. Models such as that suggested by Horn and Schunck (1981) integrate these measurements over areas, while the model proposed by Hildreth (1984) integrates motion measurements along connected contours. This issue was addressed in a recent perceptual study by Nakayama and Silverman (1984a). Their study used a simple distorted line, oscillating up and down. When viewed alone, a central diagonal section of the line appeared to move in an oblique direction, so that the entire figure appeared nonrigid. The figure could be made to appear to move rigidly up and down by the introduction of additional features that were unambiguously moving up and down. Nakayama and Silverman introduced both breaks on the contour and short segments off the contour. They found that both the breaks on the line and the segments off the line could

cause the central part of the line to appear to move up and down, but the features on the contour had a much stronger effect, in that their distance from the center could be very large. The segments had to be very close to the line in order to exert any influence on the perception of its motion. These phenomena suggest that the integration of motion constraints along contours may play a stronger role in the human visual system, an observation that is also supported by perceptual demonstrations presented by Hildreth (1984).

The local perpendicular components of motion are not always combined by the human visual system. The conditions governing whether or not these measurements are combined were studied by Adelson and Movshon (1982) and by Nakayama and Silverman (1983). In the Adelson and Movshon study, the stimulus patterns consisted of two superimposed sinewave gratings at different orientations, moving in the direction perpendicular to their orientations. Together, the two gratings formed a single rigid pattern, moving in a direction consistent with the constraints imposed by the two components. Under some conditions, the gratings did not form a single coherent pattern perceptually; rather, the two components appeared to split and move independently of one another. The coherence of the combined pattern was found to decrease with an increase in any of the following factors: (1) the difference in contrast between the two gratings, (2) the angle between the primary directions of the gratings, (3) the difference between the two spatial frequencies and (4) the speed of movement of the overall pattern. In a later study by Adelson (1984), it was shown that the two components of motion would also appear to split if they were presented on different depth planes. This observation suggests that stereo disparity enters into the solution to the aperture problem in motion. Nakayama and Silverman (1983), by using stimuli consisting of sinewave lines, demonstrated that two components of motion tend not to be combined if their orientations are very similar (i.e. they differ by at most about 30°). These perceptual studies suggest that early measurements of the perpendicular components of motion are not always combined by the human visual system. Under some conditions, they will remain separate, resulting in a perception of motion that corresponds directly to the pattern of components. More generally, these studies provide implicit support for the notion that motion measurement takes place in two stages, with the first stage providing the perpendicular components of motion and the second stage combining these components into a single coherent pattern of motion. More explicit psychophysical support for a two–stage motion measurement computation is presented in Movshon *et al.* (1985).

The motion measurement problem can also be examined from a physiological perspective. Early movement detectors in biological systems have spatially limited receptive fields and therefore face the aperture problem. Stimulated by a theoretical analysis of the aperture problem, Movshon *et al.* (1985) sought and found direct physiological evidence for a two–stage motion measurement computation in the primate visual system. Two visual areas that include an abundance of motion–sensitive neurons are cortical areas V1 and the middle temporal area of extrastriate cortex (MT), located in the posterior bank of the superior temporal sulcus (for example, see Maunsell and Van Essen, 1983; Van Essen and

Maunsell, 1983; Allman, Miezin and McGuinness, 1985; Saito *et al.*, 1986). The explicit role of area MT in the cortical analysis of visual motion was confirmed recently by Newsome *et al.* (1985), who showed that small restricted chemical lesions in area MT of the macaque monkey led to a behavioral deficit in the monkey's ability to match the velocity of smooth pursuit eye movements with the velocity of visual targets. Moreover, lesions in the cat's Claire–Bishop area, which is assumed to correspond to area MT in the macaque anatomically, led to a much reduced ability of behaving cats to distinguish small moving figures from both moving and stationary surrounds (Strauss and von Seelen, 1986). Movshon *et al.* (1985) explored the type of motion analysis taking place in the primate's MT, by using the same stimulus with superimposed sinewave gratings used by Adelson and Movshon (1982). The results of these experiments indicate that the selectivity of neurons in area V1 for direction of movement is such that they could provide only the component of motion in the direction perpendicular to the orientation of image features. These neurons essentially only respond to a single component of the combined grating pattern, independent of the presence of the second grating. Area MT, however, contains a subpopulation of cells, referred to as *pattern* cells, that appear to respond to the 2–D direction of motion of the combined grating pattern. For example, imagine a sinewave grating moving diagonally up this page (bottom left to top right) and a second pattern superimposed on the first, moving diagonally down the page (top left to bottom right). A neuron in V1 whose best direction id diagonally upward would respond to the superimposed pattern, as though the downward moving diagonal were not even present. A pattern cell in MT, however, would respond to the superimposed patterns as though they were moving directly across the page from left to right. Thus, these pattern cells may serve to combine motion components to compute the real 2–D direction of velocity of a moving pattern. These experiments do not yet distinguish between the use of the simple assumption of pure translation, as suggested in the study by Movshon *et al.*, 1985, and a more general assumption such as smoothness. Stimulus patterns undergoing more complicated motions are required to make such a distinction. If the pattern cells in area MT employ the assumption of smoothness in their computation of motion, one would expect to find direct interaction between pattern cells that analyze nearby areas of the visual field.

Poggio and Koch (1985; Poggio *et al.*, 1985) presented hypothetical neural implementations of regularization algorithms in terms of very simple linear, electrical or chemical, analog networks. In particular, they proposed an implementation for the computation of the smoothest velocity field as suggested by Hildreth (1984). From these networks, a neural circuit is then designed that behaves in a similar way. Examples of the electrical and neural networks are shown in Figure 6. In the network of Figure 6a, the currents $\mathbf{I}_i$ and conductances $\mathbf{g}$ and $\mathbf{g}_i$ represent measurements of the perpendicular components of velocity and other properties of a moving contour obtained directly from the image. The voltages $\mathbf{V}_i$ represent the tangential component of velocity (i.e. the component of velocity in the direction parallel to the orientation of features in the image) that is recovered by the computation

of the full 2–D velocity field. These analog resistive networks allow a fast computation of the smoothest velocity field and are guaranteed to converge to the correct solution (Poggio and Koch, 1985). In the corresponding neural implementation of Figure 6b, the tangential component of the velocity field is represented by the voltages $\mathbf{V}_i$ along a dendrite, which are sampled by dendro–dendritic synapses. Measurements from the image are represented by synaptically mediated current injections $\mathbf{I}_i$ and other synaptic inputs $\mathbf{R}_i$ (for instance, a silent $GABA_A$ type inhibitory synapse) that control the membrane resistance. The full 2–D velocity field is represented implicitly by the combination of the currents $\mathbf{I}_i$ and the voltages $\mathbf{V}_i$. This hypothetical neural implementation was not intended as a specific model for the measurement of motion in area MT. Rather, its intent was to show that it is possible for neural hardware to exploit a model of this computation that incorporates a general assumption such as smoothness of the velocity field. Models such as this can help to focus experimental questions regarding the actual neural circuitry in areas such as MT.

Long Range Motion Correspondence

The preceding section addressed computational models that might underlie the short–range process. The computation of a velocity field requires that motion in the image be roughly continuous. The perception of motion by the human visual system does not, however, require that objects move continuously across the visual field. Motion can be inferred when features are presented discretely at positions separated by up to several degrees of visual angle and with long temporal intervals between presentations. There are many visual patterns that yield qualitatively different perceptions of motion, depending on the size of the spatial and temporal displacements between frames (for example, Ternus, 1926; Anstis, 1970, 1980; Braddick, 1974, 1980; Anstis and Rogers, 1975; Pantle and Picciano, 1976; Petersik and Pantle, 1979; Shepard and Judd, 1976; Burt and Sperling, 1981; Green and von Grünau, 1983; Hildreth, 1984; Anstis and Mather, 1985). Although the short–range and long–range motion processes may interact at some stage (Clatworthy and Frisby, 1973; Green and von Grünau, 1983), there is evidence that they are initially distinct processes (Mather, Cavanagh and Anstis, 1985; Gregory, 1985; Anstis and Mather, 1985).

The long–range motion phenomena illustrate the ability of the human visual system to derive a correspondence between elements in the changing image, over considerable distances and temporal intervals. Under these conditions, there is no continuous motion of elements across the image to be measured directly. A correspondence computation is therefore likely to underlie the long–range motion process. Two issues arise regarding this computation: first, what features in the image are matched from one moment to the next, and second, how is a unique correspondence of features established? Similar to the velocity field computation, many possible matchings between features in two images exist, and additional constraints must be imposed to compute a single correspondence that is most plausible from a physical standpoint.

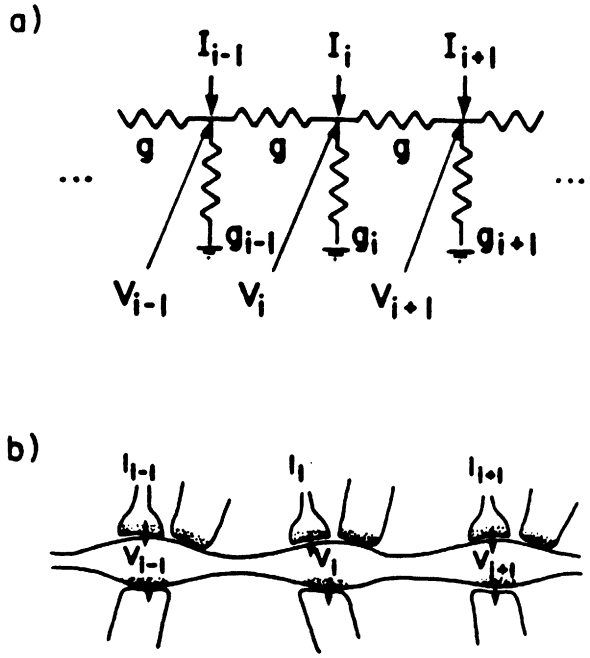

Figure 6. Analog models of the velocity field computation. (a) A simple resistive network that computes the smoothest velocity field. The conductances $\mathbf{g}$ and $\mathbf{g_i}$, and the currents $\mathbf{I_i}$ represent properties of a moving contour that are measured directly from the image. In particular, $\mathbf{g_i}$ is proportional to the square of the contrast of the contour at location i. The 2-D velocity field along the contour is represented implicitly by the combination of these inputs and the resulting voltages $\mathbf{V_i}$. (b) A hypothetical neural implementation of the circuit shown in (a). Synaptic mediated currents $\mathbf{I_i}$, and additional inputs $\mathbf{R_i}$ (possibly a $GABA_A$ type of synapse) represent properties of a moving contour. The resulting voltages $\mathbf{V_i}$, sampled by dendro–dendritic synapses, together with the input currents, represent local velocities along the contour. Redrawn from Poggio and Koch (1985).

The possible image features that could form the matching elements span a wide range, from simple edge and line segments, points, and blobs, to texture boundaries, subjective contours and groups of primitive features, and even to structured forms or entire objects. Motion measurement schemes used in computer vision, reviewed for example in Thompson and Barnard (1981), Ullman (1981a) and Barron (1984), have considered most of these possible matching elements. In general, the earlier tokens such as edge and line segments are easier to compute, but there is greater ambiguity in the matching of these tokens from one moment to the next. The use of primitive tokens also allows the correspondence process to operate on arbitrary objects undergoing complex shape changes. More complex tokens

such as structured forms can simplify the correspondence process, but more computation is required to extract these features from the image, and there is less flexibility in the types of motion that can be analyzed.

Perceptual studies suggest that many long–range motion phenomena can be explained in terms of a correspondence of elements such as edges, bars, line terminations, points, etc. (Ullman, 1979). The human visual system can also establish a correspondence between groups of primitive elements even when the constituents of the groups are not the same (Riley, 1981), subjective contours and texture boundaries (Ramachandran, Rao and Vidyasagar, 1973; Riley, 1981) and subjective surfaces (Ramachandran, 1985). Properties of primitive elements such as orientation, contrast and size can influence the correspondence computation (for example, Frisby, 1972; Kolers, 1972; Ullman, 1979, 1981b), although it is possible to establish a correspondence between objects that differ significantly in their components (Navon, 1976; Anstis and Mather, 1985). Chen (1985) has suggested that topological features such as connectivity, closure and the presence of holes can play a role in motion correspondence, but it is not clear whether these properties are made explicit in the description of the matching elements, or whether they are reflected in the constraints that are used to establish a unique correspondence of elements between frames.

The rules or constraints that are used by the human visual system to establish a correspondence of elements between frames have also been explored in many studies. Early perceptual studies focused on the role of the time and distance between elements in successive frames (for example, Ternus, 1926; Kolers, 1972; Burt and Sperling, 1981). When the elements in motion are isolated dots, each dot in general 'prefers' to match its nearest neighbor in the subsequent frame, although this constraint sometimes can be violated locally when a field of dots in motion interacts (Ullman, 1979; Burt and Sperling, 1981). The distance metric that is used in the correspondence process appears to be based on 2–D distances between elements rather than 3–D distances (Ullman, 1979; Mutch, Smith and Yonas, 1983; Tarr and Pinker, 1985). Ramachandran and Anstis (1983, 1985) showed that 'inertia' can influence correspondence; that is, in ambiguous situations, moving elements will tend to maintain the same direction of motion over time.

A computational model of correspondence presented by Ullman (1979) assumes independence of the matching elements. Subsequent studies have revealed situations in which the independence assumption appears not to hold. For example, the perceived motion of a feature can be influenced by the motion of other features connected to it along a contour (Hildreth, 1984; Chen, 1985). Ramachandran and Anstis (1985) created a display in which a local pattern of dots whose motion was two–way ambiguous was repeated in a large array. Each local subpattern could in principle be perceived as moving in either of two directions, but observers always perceived the array of patterns as all moving in the same direction. The correspondence established within one subpattern of the display could influence the correspondence of dots in neighboring subpatterns.

To summarize, much is known about the matching elements used in long range corre-

spondence, and the rules or constraints used to match elements. Many recent perceptual studies were motivated by computational models of the correspondence process. At present, however, there are no computational models that adequately account for all of the long–range motion phenomena observed in perceptual studies. Recent physiological studies that explored the response of MT neurons to apparent movement stimuli (Newsome, Mikami and Wurtz, 1982, 1986; Mikama, Newsome and Wurtz, 1986) suggest that area MT might provide some of the neural substrate for the interpretation of long–range motion.

The Detection of Motion Discontinuities

If two adjacent surfaces undergo different motions, a discontinuity generally occurs in the optical flow or velocity field along their boundary. The explicit detection of motion discontinuities allows the detection and localization of object boundaries in the scene. Other cues to the presence of boundaries often occur as well, such as sharp changes in stereo disparity or texture, but perceptual studies suggest that it is possible to detect object boundaries on the basis of motion information alone (Anstis, 1970; Regan and Spekreijse, 1970; Julesz, 1971) and to use the relative motions in the vicinity of these boundaries to infer the relative locations of surfaces in depth (Kaplan, 1969; Nakayama and Loomis, 1974; Mutch and Thompson, 1985).

It is advantageous to detect motion discontinuities as early as possible, for two reasons. First, the fast detection of a sudden relative movement in the environment can serve as an early warning system, alerting the observer to a possible prey or predator, or to the sudden movement of an object toward the viewer. It is essential not only to detect the presence of movement, but also to identify the outline of the object. A second reason for detecting motion discontinuities early is that they facilitate the subsequent measurement of 2–D motion in the image. It was noted ealier that the computation of a velocity field requires the integration of local measurements of the perpendicular components of motion. Motion measurements should only be combined within single surfaces, as the combination of measurements across object boundaries will generally yield errors in the velocity field. If detected early, the motion discontinuities can define regions of the image within which the local motion measurements should be combined.

With regard to computational schemes, one issue that arises is the question of what stage in the analysis of the image should discontinuities first be detected. Three alternatives present themselves. First, motion discontinuities could be localized prior to the computation of the full velocity field, just after the initial measurements of the perpendicular components of motion in the image (for example, Schunck and Horn, 1981; Hildreth, 1984). Schunck and Horn used simple heuristics to avoid combining motion measurements that are likely to occur on surfaces undergoing different motions. Hildreth presented a scheme to detect sudden changes in the perpendicular components of motion, which uses techniques that were previously used for edge detection (Marr and Hildreth, 1980). H. Bülthoff and T.

Poggio (1986, personal communication) use the binary output of simple correlation–like detectors, signaling motion to the left or right, to localize discontinuities in dense random–dot patterns. Surprisingly, such a simple measure gives a fairly accurate assessment of discontinuities, at least for random–dot stimuli.

A second possible stage at which boundaries can be detected is after the velocity field has been computed explicitly everywhere. For example, Nakayama and Loomis (1974) proposed a local center–surround operator to detect boundaries in optical flow fields. Similar ideas are incorporated in models suggested by Clocksin (1980), and Thompson, Mutch and Berzins (1982, 1985; Mutch and Thompson, 1985), which use a Laplacian operator applied to components of the optical flow field. In other schemes explored, for example, by Potter (1977) and Fennema and Thompson (1979; Thompson, 1980), region–growing techniques are used to group together elements of similar velocities.

Finally, the velocity field and its discontinuities could be computed simultaneously. In a scheme suggested by Wohn (1984) and Waxman (1986), the motion segmentation problem is approached by detecting "boundaries of analyticity" at which an approximation of the local image flow by second order polynomials breaks down. The boundaries are located within the process that models the local motion field. Koch, Marroquin and Yuille (1986a) have proposed that binary line processes, first introduced in the solution of vision problems by Geman and Geman (1984), can successfully demarcate motion boundaries. At locations where this line process is set, an unobservable line or edge is postulated to interrupt the otherwise smooth velocity field, segmenting the image into its natural components. The appropriate algorithm can be formulated as an energy minimization problem that maps naturally into simple analog networks (Koch, Marroquin and Yuille, 1986).

A detailed neural circuitry for the detection of motion discontinuities by the housefly was proposed by Reichardt, Poggio and Hausen (1983 Reichardt and Poggio, 1979). Large field binocular "pool" cells summate the output of a retinotopic array of small field elementary movement detectors (EMD) over a large part of the visual field of the two compound eyes. The EMD signal movement in one of two directions: progressive, i.e. movement from front to back, and regressive, i.e. movement from back to front. The pool cells inhibit in turn, via a silent or shunting inhibition (see the section on circuitry and biophysics), the signals provided by the EMD, irrespective of their preferred direction. After inhibition of each channel, all signals from the EMD feed into a large field output cell. This circuit shows two important properties: it detects relative motion of a moving figure superimposed on a stationary background of the same texture as the figure, and its output, the optomotor response, is independent of the size of moving figure. Motion discontiuities are signaled by significant activity in the output cells. The model agrees well with behavioral data from the fly. Moreover, elements of the proposed circuitry can be identified with anatomically and physiologically characterized cells in the visual system of the fly (Egelhaaf, 1985).

Physiological studies have revealed center–surround mechanisms that are organized antagonistically for direction of motion in many vertebrate species (for example, Sterling

and Wickelgren, 1969; Collett, 1972; Bridgeman, 1972; Frost, 1978; Frost, Scilley and Wong, 1981; Frost and Nakayama, 1983). Motion–sensitive cells with this organization have been found recently in area MT of the Owl monkey (Miezin, McGuinness and Allman, 1982; Allman, Miezin and McGuinness, 1985) and in striate cortex of the cat (Orban *et al.*, 1986). The existence of center–surround relative motion detection mechanisms across such a range of species suggests that a similar strategy may be utilized in the underlying computations. Richards and Lieberman (1982) show in psychophysical studies that some viewers are "blind" to shearing motions, and suggest that the neural substrate for detecting such discontinuous motions may be independent from mechanisms detecting other motion boundaries.

Psychophysical studies of motion discontinuities have mainly used dynamic random dot patterns, in which only motion cues signal the presence of boundaries. Braddick's (1974, 1980) studies revealed a limit on the spatial and temporal displacements required to perceive coherent motion in dense random dot patterns, and showed that it was possible to detect a boundary between coherent and incoherent fields of motion. Experiments by Baker and Braddick (1982a) and van Doorn and Koenderink (1982, 1983) suggest that the detection of discontinuities is not based on a computation that explicitly measures only relative movement; rather, an absolute measurement of motion takes place first, followed by a process that compares nearby motions to locate discontinuities. Baker and Braddick (1982b) showed that the ability to discriminate the orientation of a patch that moves against an uncorrelated background varies little with dot density and increases with the patch size (see also Chang and Julesz, 1983). In general, the size of a patch of moving dots that can be discriminated against a differentially moving background increases with larger displacements of the dots between frames (Hildreth, 1984). This phenomenon may reflect the limitations of multiple spatial frequency channels involved in the early detection of motion. Other perceptual studies have shown that spatial frequency plays a role in determining the maximum displacements that allow the perception of coherent motion in random dot patterns (Chang and Julesz, 1983; Nakayama and Silverman, 1984b).

It is important to draw a distinction between the ability to detect differences in motion, and the ability to localize a boundary between surfaces undergoing different motions. For example, if two adjacent fields are undergoing motion in the same direction, a 5% difference in speed is sufficient to detect relative movement (McKee, 1981; Nakayama, 1981). To localize a boundary, however, requires much larger differences in speed, between $50 - 60\%$ (van Doorn and Koenderink, 1982, 1983; Hildreth, 1984). If two adjacent surfaces undergo motions with similar speeds but different directions, then an angular change in direction of at least $20°$ is required to localize the position of the boundary (Hildreth, 1984).

Experimental studies have provided much insight into the nature of the mechanisms that underlie the detection of motion discontinuities in biological systems. Many fundamental questions still remain, however. Perhaps the most basic open question concerns at what stage in the analysis of motion the discontinuities are first detected. It is not known,

for example, what representation of motion forms the input to the center–surround mechanisms observed in area MT by Allman, Miezin and McGuinness (1985). These mechanisms may operate directly on the perpendicular components of motion, or they may operate on the real 2–D directions of image motion. Psychophysical studies have not yet addressed this issue directly. Furthermore, while physiological studies reveal that some sort of center–surround mechanisms are involved in the detection of relative movement, little is known about what these mechanisms really compute and how they compute this information. Further computational studies are needed to examine possible algorithms for detecting motion boundaries that may utilize these center–surround mechanisms.

THE RECOVERY OF THREE–DIMENSIONAL STRUCTURE FROM MOTION

The Computational Problem and Related Perceptual Studies

When an object moves in space, the motions of individual points on the object differ in a way that conveys information about its 3–D structure, as illustrated in Figure 5a. The directions of motion in this case are all horizontal, but the speed of movement varies in a way that depends on the structure of the object. Using wireframe objects such as that shown in Figure 5a, Wallach and O'Connell (1953) showed that the human visual system can derive the correct 3–D structure of moving objects from their changing 2–D projection alone. Other perceptual studies also demonstrated this remarkable ability (for example, Green, 1961; Braunstein, 1962, 1976; Johansson, 1973, 1975; Rogers and Graham, 1979; Ullman, 1979; Cutting, 1982; Cutting and Proffitt, 1982). Relative motion in the image is also created by movement of the observer relative to the environment, and can be used to infer observer motion from the changing image (Gibson, 1950; Lee and Aronson, 1974; Johansson, 1971; Lee, 1980).

Theoretically, the two problems of (1) recovering the 3–D structure and movement of objects in the environment and (2) recovering the 3–D motion of the observer from the changing image, are closely related. The main difficulty faced by both is that infinitely many combinations of 3–D structure and motion could give rise to any particular 2–D image. To resolve this inherent ambiguity, it is necessary to impose additional constraint that allows most 3–D interpretations to be ruled out, leaving one that is most plausible from a physical standpoint. Computational studies have used the *rigidity* assumption to derive a unique 3–D structure and motion; they assume that if it is possible to interpret the changing 2–D image as the projection of a rigid 3–D object in motion, then such an interpretation should be chosen (for example, Ullman, 1979, 1983; Clocksin, 1980; Prazdny, 1980, 1983; Longuet–Higgins, 1981; Longuet–Higgins and Prazdny, 1981; Tsai and Huang, 1981; Hoffman and Flinchbaugh, 1982; Bobick, 1983; Mitiche, 1984, 1986; Mitiche, Seida and Aggarwal, 1985; Waxman and Ullman, 1985). When the rigidity assumption is used in this way, the recovery of structure from motion requires the computation of the rigid 3–D object that would project onto a given 2–D image. The rigidity assumption was suggested by perceptual studies that described a tendency for the human visual system to choose a rigid interpretation of moving elements (Wallach and O'Connell, 1953; Gibson and Gibson, 1957; Green, 1961; Jansson and Johansson, 1973; Johansson, 1975, 1977).

Computational studies have shown that the rigidity assumption can be used to derive a unique 3–D structure from the changing 2–D image. Furthermore, this unique 3–D interpretation can be derived by integrating image information only over a limited extent in space and in time. For example, suppose that a rigid object in motion is projected onto the image plane by using orthographic projection. Three distinct views of four points on

the moving object are sufficient to compute a unique rigid 3–D structure for the points (Ullman, 1979). In general, if only two views of the moving points are considered or fewer points are observed, there are multiple rigid 3–D structures consistent with the changing 2–D projection. If a perspective projection of objects onto the image is used instead, then two distinct views of seven or eight points in motion are usually sufficient to compute a unique 3–D structure for the points (Longuet–Higgins, 1981; Tsai and Huang, 1981). If the instantaneous velocity of movement in the image is known at discrete points, then under perspective projection, the position and velocity at five points may be sufficient to derive a unique structure (Prazdny, 1980; Roach and Aggarwal, 1980). Longuet–Higgins and Prazdny (1981) originally showed that if the continuous velocity field is known everywhere within a region of the image, then the velocity field together with its first and second spatial derivatives at a point is consistent with at most three possible surface orientations at that point. Waxman, Kamgar–Parsi and Subbarao (see Waxman, 1986) have recently shown that a unique solution can usually be determined in this case. Finally, for the case of orthographic projection, 3–D structure can be recovered uniquely if both the velocity and acceleration fields are known within a region (Hoffman, 1982). Additional theoretical results have been obtained for classes of restricted motion, such as planar surfaces in motion (Hay, 1966; Koenderink and van Doorn, 1976; Buxton *et al.*, 1984; Longuet–Higgins, 1984; Murray and Buxton, 1984; Kanatani, 1985; Waxman and Ullman, 1985; Ullman, 1985; Negahdaripour and Horn, 1985; Subbarao and Waxman, 1985), pure translatory motion of the observer (Clocksin, 1980; Lawton, 1983; Jerian and Jain, 1984), planar or fixed axis rotation (Hoffman and Flinchbaugh, 1982; Webb and Aggarwal, 1981; Bobick, 1983; Bennett and Hoffman, 1985; Sugie and Inagaki, 1984), translation perpendicular to the rotation axis (Longuet–Higgins, 1983), and motion of quadratic surfaces (Waxman and Ullman, 1985; Waxman and Wohn, 1985). A review of early theoretical results regarding the recovery of structure from motion can be found in Ullman (1983).

The theoretical results summarized above are important for the study of the recovery of structure from motion in biological vision systems, for at least two reasons. First, they show that by using the rigidity assumption, a unique structure can be recovered from motion information alone. It is not necessary to make further physical assumptions, in order to obtain a unique solution. Second, these results show that it is possible to recover 3–D structure by integrating image information over a small extent in space and in time. This second observation could bear on the neural mechanisms that compute structure from motion; in principle, they need only integrate motion information over a limited area of the visual field and a limited extent in time.

The above computational studies of the recovery of structure from motion also provide algorithms for deriving the structure of moving objects. Typically, measurements of the positions or velocities of image features give rise to a set of mathematical equations whose solution represents the desired 3–D structure. The algorithms generally derive this structure from motion information extracted over a limited area of the image and a limited extent in

time. Testing of these algorithms reveals that although this strategy is possible in theory, it is not reliable in practice. A small amount of error in the image measurements can lead to very different (and often incorrect) 3–D structures. This behavior is due in part to the observation that over a small extent in space and time, very different objects can induce almost identical patterns of motion in the image (Ullman, 1983, 1984).

This sensitivity to error inherent in algorithms that integrate motion information only over a small extent in space and time suggests that a robust scheme for deriving structure should use image information that is more extended in space or time or both. This conclusion is supported in recent computational studies (Bruss and Horn, 1983; Lawton, 1983; Ullman, 1984; Adiv, 1985; Negahdaripour and Horn, 1985, Waxman and Wohn, 1985; Wohn and Waxman, 1985). Lawton (1983) showed that recovery of the translatory motion of an observer could be coupled with the solution to the motion correspondence problem over an extended region of the image, to yield a robust solution. Adiv (1985) presented an algorithm for recovering the motion parameters for several moving objects, which assumes that object surfaces are piecewise planar. The extraction of the motion parameters uses a least–squares approach that minimizes the deviation between the measured flow field (at a large number of points) and that predicted from the estimated motion and structure (Bruss and Horn, 1983). Negahdaripour and Horn (1985) also addressed the recovery of the motion of an observer relative to a stationary planar surface, and showed that a robust recovery of the observer motion and the orientation of the plane is possible when dense measurements of the spatial and temporal derivatives of image brightness are integrated over a large region of the changing image. Thus, consideration of motion information that is more extended in space can lead to a stable recovery of structure. The study by Ullman (1984), elaborated below, demonstrated that a robust recovery of structure is also possible when motion information is integrated over an extended period of time. The extension in time can be achieved, for example, by considering a large number of discrete frames or by observing continuous motion over a significant temporal extent.

With regard to the human visual system, the dependence of perceived structure on the spatial and temporal extent of the viewed motion has not yet been studied systematically, but the following informal observations have been made. Regarding spatial extent, two or three points undergoing relative motion are sufficient to elicit a perception of 3–D structure (Borjesson and von Hofsten, 1973; Johansson, 1975), although theoretically the recovery of structure is less constrained for two points in motion, and perceptually the sensation of structure is weaker. An increase in the number of moving elements in view appears to have little effect on the quality of perceived structure (for example, Petersik, 1980). Regarding the temporal extent of viewed motion, Johansson (1975) showed that a brief observation of patterns of moving lights generated by human figures moving in the dark (commonly referred to as biological motion displays) can lead to a perception of the 3–D motion and structure of the figures. Other perceptual studies indicate that the human visual system requires an extended time period to reach an accurate perception of 3–D structure (Wallach and

O'Connell, 1953; White and Mueser, 1960; Green, 1961; Doner, Lappin and Perfetto, 1984; Inada *et al.*, 1986). A brief observation of a moving pattern sometimes yields an impression of structure that is "flatter" than the true structure of the moving object. Thus, the human visual system is capable of deriving some sense of structure from motion information that is integrated over a small extent in space and time. An accurate perception of structure may, however, require a more extended viewing period.

Most methods compute a 3–D structure from motion only when the changing image can be interpreted as the projection of a rigid object in motion. They otherwise yield no interpretation of structure or yield a solution that is incorrect or unstable. Algorithms that are exceptions to this can interpret only restricted classes of nonrigid motions (Bennett and Hoffman, 1985; Hoffman and Flinchbaugh, 1982; Koenderink and van Doorn, 1986). The human visual system, however, can derive some sense of structure for a wide range of nonrigid motions, including stretching, bending and more complex types of deformation (Johansson, 1964; Jansson and Johansson, 1973; Todd, 1982, 1984). Furthermore, displays of rigid objects in motion sometimes give rise to the perception of somewhat distorting objects (Wallach, Weisz and Adams, 1956; White and Mueser, 1960; Green, 1961; Braunstein, 1962; Sperling *et al.*, 1983; Braunstein and Andersen, 1984; Hildreth, 1984; Adelson, 1985). These observations suggest that while the human visual system tends to choose rigid interpretations of a changing image, it probably does not use the rigidity assumption in the strict way that previous computational studies have suggested.

Ullman (1984) proposed a more flexible method for deriving structure from motion that interprets both rigid and nonrigid motion. Referred to as the *incremental rigidity scheme*, this algorithm uses the rigidity assumption in a different way from previous studies. It maintains an internal model of the structure of a moving object that consists of the estimated 3–D coordinates of points on the object. The model is continually updated as new positions of image features are considered. Initially, the object is assumed to be flat, if no other cues to 3–D structure are present. Otherwise, its initial structure may be determined by other cues available, from stereopsis, shading, texture, or perspective. As each new view of the moving object appears, the algorithm computes a new set of 3–D coordinates for points on the object that maximizes the rigidity in the transformation from the current model to the new positions. This is achieved by minimizimg the change in the 3–D distances between points in the model. Thus the algorithm interprets the changing 2–D image as the projection of a moving 3–D object that changes as little as possible from one moment to the next. Through a process of repeatedly considering new views of objects in motion and updating the current model of their structure, the algorithm builds up and maintains a 3–D model of the objects. If objects deform over time, the 3–D model computed by the algorithm also changes over time. Other models have been proposed that impose rigidity by requiring that the 3–D distances between points in space change very little from one moment to the next (for example, Mitiche, 1984, 1986; Mitiche, Seida and Aggarwal, 1985), although these models do not build up a 3–D model incrementally as in Ullman's proposed

scheme.

The method proposed by Ullman (1984) was motivated in part by the limitations of previous computer algorithms and in part by knowledge of the human visual system. The method has overcome limitations of previous computational studies in two ways. First, it provides a reliable recovery of structure in the presence of error in the image measurements, by integrating image information over an extended time period. Second, it allows the interpretation of nonrigid motions. These are essential qualities for any method that is proposed as a viable model for the recovery of structure from motion by the human visual system. This method also has other attributes that are consistent with human perceptual behavior: (1) it sometimes yields a nonrigid interpretation of rigid structures in motion, (2) a brief viewing time results in a structure that is "flatter" than the true structure of the object, (3) it allows a 3–D interpretation of scenes containing as few as two points in motion (Borjesson and von Hofsten, 1973; Johansson, 1975), and (4) it provides a natural means for integrating multiple sources of 3–D information.

A recent computational study by Grzywacz and Hildreth (1985) has extended Ullman's incremental rigidity scheme, presenting a formulation of the algorithm that makes direct use of instantaneous velocity information over an extended time, and showing how the algorithm can be modified to use perspective projection of the scene onto the image. With regard to the use of velocities, previous studies had suggested that the recovery of 3–D structure from velocity information at a single moment is inherently unstable (Prazdny, 1980; Ullman, 1983). Through computer simulations and a theoretical analysis, Grzywacz and Hildreth showed that the integration of velocity information over an extended time does not overcome this problem of instability. The velocity based formulation of the incremental rigidity scheme does not yield a robust computation of structure over an extended time; rather, the solution oscillates between good and poor estimates of the 3–D structure of a moving object. More generally, if discrete views of moving elements are used instead, the incremental rigidity scheme performs best when the spatial changes between views are large. For example, if an object is rotating, the algorithm computes a better 3–D structure for the object if larger angular rotations between discrete frames are considered.

With regard to the human visual system, it is unlikely that discrete movie–like "snapshots" form a direct input to the recovery of 3–D structure from motion. Second, if a short–range motion measurement system exists and provides essentially instantaneous measurements of movement in the changing image, these measurements should be used in some way to interpret the 3–D structure of the scene. These short–range measurements may, however, form the input to a longer–range tracking operation that integrates image motion information over a more extended time for the accurate recovery of 3–D structure. In any case, the short–range measurements can also be used to identify motion discontinuities, which are likely to indicate the locations of object boundaries in the scene. Knowledge of object boundaries can improve the overall recovery of structure from motion.

This discussion of the structure–from–motion problem illustrates a number of impor-

tant points that often arise in the computational study of other problems in the early stages of vision. First, a single solution to the problem cannot be obtained from information in the image alone; additional constraint is required. Second, theoretical studies can be used to show that a general physical assumption such as rigidity is sufficient to solve the structure–from–motion problem uniquely. Third, an assumption such as rigidity can be incorporated in many ways into an algorithm to recover structure. The development of a reliable algorithm requires a cycling between computer implementation, testing, and refinement. Finally, perceptual studies can suggest and test particular assumptions and reveal aspects of the algorithm used by the human visual system for solving a given problem. It is typical of computational studies that the initial methods proposed for solving a problem only loosely consider the detailed observations of biological systems. These first studies uncover useful aspects of the problems, however. Later studies then combine this knowledge of the problem with observations of biological systems to derive models that more closely reflect the computations carried out in biological systems.

Physiological Studies of the Recovery of Structure from Motion

Physiological studies have uncovered neurons in higher cortical areas that are sensitive to properties of the motion field that may be relevant to the recovery of the 3–D structure and motion of surfaces in the environment, or to the recovery of the motion of the observer relative to the scene. Many studies have revealed neurons sensitive to uniform expansion or contraction of the visual field, a property that is correlated either with translation of the observer forward or backward, or equivalently, motion of an object toward or away from the observer. Such neurons have been found, for example, in the posterior parietal cortex of the monkey (Motter and Mountcastle, 1981; Andersen, 1986). Other neurons have been found that are sensitive to global rotations in the visual field (Andersen, 1986; Sakata *et al.*, 1985). All of these neurons have large receptive fields, so they probably lack the spatial sensitivity required to derive the detailed shape of an object surface from relative motion. In the human visual system, the accurate recovery of object shape from motion may be an ability that is restricted to the central region of the eye; the ability to interpret 2–D structure–from–motion displays appears to degrade rapidly as one moves away from the fovea (S. Ullman, personal communication). Siegel and Andersen (1986) showed that motion processing in area MT is critical to the recovery of structure from motion.

The neurons sensitive to relative movement that were discussed in the context of motion discontinuities may also contribute to the recovery of 3–D structure. Certainly the detection and localization of object boundaries is essential to the construction of a 3–D representation of surfaces in the scene. Mechanisms such as the "convexity" detector suggested by Nakayama and Loomis (1974) may also derive information about the relative depths of surfaces on either side of a motion boundary. The computational study by Mutch and Thompson (1985) also addressed this issue.

Regan and Beverley (1979, 1983) have hypothesized the existence of 'changing–size' detectors (analogous to detectors of uniform expansion or contraction in the visual field) based on psychophysical evidence from adaptation studies. They also suggested that the changing–size detectors may be distinct from neural mechanisms signaling motion in depth (Beverley and Regan, 1979). Neurons exist in area 18 of the cat visual cortex (for example, Cynader and Regan, 1978, 1982) and area V1 of the primate visual cortex (Poggio and Talbot, 1981) that appear to be selective for direction of movement in depth. These studies of cells responsive to movement in depth used binocularly viewed moving bars, however, so they may address the interaction between binocular stereopsis and motion measurement for the recovery of movement in space, rather than the recovery of structure from motion alone.

CONCLUDING REMARKS

In this review we have tried to integrate studies from computation, psychophysics, physiology and biophysics into a computational framework. The interaction between these different approaches promises to be fruitful in furthering our understanding of motion analysis in biological vision systems, because the various perspectives each provide valuable and different insight into how vision systems analyze motion information.

Perceptual studies, for example, help to define the problems in motion analysis that are solved, and reveal the quantitative ability with which the human visual system can solve these problems. We have seen that many problems in motion analysis do not have a unique solution, and additional constraint must be imposed to solve them. There are often different choices for the assumptions that could be embodied in the underlying computations, which critical perceptual experiments can attempt to distinguish. There are also many algorithms that could solve a given problem, and different algorithms might fail in different ways. Again, critical perceptual experiments can be designed to determine whether the human visual system fails in the same way. It is often the case that perceptual studies provide initial hints about the strategies used in the underlying computations.

Studies from physiology and biophysics can reveal what parts of the visual system are involved in a particular computation, and what the elementary operations are that neurons use in processing motion information. Properties of the underlying hardware also constrain the nature of the algorithms and representations that are used in motion computations. Detailed computer models of neuronal networks subserving motion measurement have helped to focus further experimental questions regarding physiological and biophysical behavior. Finally, physiological methods can help eliminate ambiguities in perceptual studies. Because the primate visual system may have evolved a variety of different algorithms to cope with a particular problem, a psychophysical paradigm may be unable to distinguish between these different algorithms, while single–cell recordings may do so.

Computational studies help to focus questions for perceptual studies about the assumptions, representations, and algorithms used by the human visual system to analyze motion. Implementations of proposed algorithms have provided powerful predictive tools for making hypotheses about what the behavior of the system ought to be if it is performing motion computations in particular ways. In the case of physiological studies, by elucidating the problems that need to be solved in motion analysis, computational studies can aid the initial exploration of the function of neurons in motion–sensitive areas in the visual pathway. By elucidating possible methods by which computations can be performed, computational studies can help to refine our understanding of how neurons function and by what mechanisms.

Acknowledgments: We thank Tomaso Poggio and Shimon Ullman for useful comments on a draft of this manuscript.

REFERENCES

Adelson, E. H. 1984. Binocular disparity and the computation of two–dimensional motion. *J. Opt. Soc. Am. A* 1:1266.

Adelson, E. H. 1985. Rigid objects that appear highly non–rigid. *Invest. Ophthalmol. Vision Sci. Suppl.* 26:56.

Adelson, E. H., Bergen, J. R. 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2:284–299.

Adelson, E. H., Movshon, J. A. 1982. Phenomenal coherence of moving visual patterns. *Nature* 300:523–525.

Adiv, G. 1985. Determining three–dimensional motino and structure from optical flow generated by several moving objects. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI–7:384–401.

Allman, J., Miezin, F., McGuinness, E. 1985. Direction– and velocity–specific responses from beyond the classical receptive field in the middle temporal area (MT). *Perception* 14:105–126.

Amthor, F.R., Oyster, C. W., Takahashi, E. S. 1984. Morphology of on-off direction-selective ganglion cells in the rabbit retina. *Brain Res.*, 298: 187–190.

Anandan, P., Weiss, R. 1985. Introducing a smoothness constraint in a matching approach for the computation of optical flow fields. *Proc. IEEE Workshop on Computer Vision: Representation and Control*, Bellaire, MI, October, pp. 186–194.

Andersen, R. A. 1986. The anatomy and physiology of the inferior parietal lobule. In *Development of Spatial Relations*, ed. U. Belugi, Chicago: Univ. Chicago Press, *in press.*

Anstis, S. M. 1970. Phi movement as a subtraction process. *Vision Res.* 10:1411–1430.

Anstis, S. M. 1980. The perception of apparent motion. *Philos. Trans. R. Soc. London Ser. B* 290:153–168.

Anstis, S. M., Mather, G. 1985. Effects of luminance and contrast on direction of ambiguous apparent motion. *Perception* 14:167–179.

Anstis, S. M., Rogers, B. J. 1975. Illusory reversal of visual depth and movement during changes of contrast. *Vision Res.* 15:957–961.

Ariel, M., Daw, N. W. 1982. Pharmacological analysis of directionally sensitive rabbit retinal ganglion cells. *J. Physiol.* 324: 161–185.

Ariel, M., Adolph, A. R. 1985. Neurotranmitter inputs to directionally sensitive turtle retinal ganglion cells. *J. Neurophysiol.* 54:1123–1143.

Baker, C. L., Braddick, O. J. 1982a. Does segregation of differently moving areas depend on relative or absolute displacement. *Vision Res.* 7:851–856.

Baker, C. L., Braddick, O. J. 1982b. The basis of area and dot number effects in random dot motion perception. *Vision Res.* 10:1253–1260.

Ballard, D. 1986. Cortical connections and parallel processing: structure and function. *Behav. Brain Sci.* 9:67–120.

Barlow, H. B., Hill, R. M., Levick, R. E. 1964. Retinal ganglion cells responding selectively to direction and speed of image motion in the retina. *J. Physiol.*, 173:377–407.

Barlow, H. B., Levick, R. W. 1965. The mechanism of directional selectivity in the rabbit's retina. *J. Physiol.* 173:477–504.

Barron, J. 1984. A survey of approaches for determining optic flow, environmental layout and egomotion. *Univ. Toronto Tech. Rep. Res. Biol. Comp. Vision*, RBCV–TR–84–5.

Bennett, B. M., Hoffman, D. D. 1985. The computation of structure from fixed axis motion: nonrigid structures. *Biol. Cybern.* 51:293–300.

Beverley, K. I., Regan, D. 1979. Separable aftereffects of changing–size and motion–in–depth: different neural mechanisms? *Vision Res.* 19, 727–732.

Bobick, A. 1983. A hybrid approach to structure–from–motion. *Proc. ACM Interdisc. Workshop on Motion: Representation and Perception*, Toronto, Canada, pp. 91–109.

Borjesson, E., von Hofsten, C. 1973. Visual perception of motion in depth: application of a vector model to three–dot motion patterns. *Percept. Psychophys.* 13:169–179.

Boycott, B. B., Wässle, H. 1974. The morphological types of ganglion cells of the domestic cat's retina, *J. Physiol.* 240:397–419.

Braddick, O. J. 1974. A short–range process in apparent motion. *Vision Res.* 14:519–527.

Braddick, O. J. 1980. Low–level and high–level processes in apparent motion. *Philos. Trans. R. Soc. London Ser. B* 290:137–151.

Braunstein, M. L. 1962. Depth perception in rotation dot patterns: effects of numerosity and perspective. *J. Exp. Psychol.* 6:41–420.

Braunstein, M. L. 1976. *Depth Perception Through Motion.* New York: Academic Press.

Braunstein, M. L., Andersen, G. J. 1984. A counterexample to the rigidity assumption in the visual perception of structure from motion. *Perception* 13:213–217.

Bridgeman, B. 1972. Visual receptive fields sensitive to absolute and relative motion during tracking. *Science* 178:1106–1108.

Bruss, A., Horn, B. K. P. 1983. Passive navigation. *Comput. Vision Graph. Image Proc.* 21:3–20.

Bülthoff, H., Bülthoff, I. 1986. GABA–Antagonist inverts movement and object detection in flies. Submitted for publication.

Burr, D. C., Ross, J. 1982. Contrast sensitivity at high velocities. *Vision Res.* 22:479–484.

Burt, P., Sperling, G. 1981. Time, distance, and feature trade–offs in visual apparent motion. *Psych. Rev.* 88:171–195.

Buxton, B. F., Buxton, H., Murray, D. W., Williams, N. S. 1984. 3D solutions to the aperture problem. In *Advances in Artificial Intelligence*, ed. T. O'Shea, Amsterdam: Elsevier.

Caldwell, J. H., Daw, N. W., Wyatt, H. J. 1978. Effects of picrotoxin and strychnine on rabbit retinal ganglion cells: lateral interactions for cells with more complex receptive fields. *J. Physiol.* 276:277–298.

Chang, J. J., Julesz, B. 1983. Displacement limits, directional anisotropy and direction versus form discrimination in random–dot cinematograms. *Vision Res.* 23:639–646.

Chen, L. 1985. Topological structure in the perception of apparent motion. *Perception* 14:137–208.

Clatworthy, J. L., Frisby, J. P. 1973. Real and apparent movement: evidence for unitary mechanism. *Perception* 2:161–164.

Cleland, B. G., Levick, W. R. 1974. Properties of rarely encountered types of ganglion cells in the cat's retinaand an overall classification. *J. Physiol.*, 240:457–492.

Clementi, E. 1985. *Ab initio* computational chemistry. *J. Phys. Chem.*, 89:4426–4436.

Clocksin, W. F. 1980. Perception of surface slant and edge labels from optical flow: a computational approach. *Perception* 9:253–269.

Collett, T. 1972. Visual neurons in the anterior optic tract of the privet hawk moth. *J. Comp. Physiol.* 78:396–433.

Cutting, J. E. 1982. Blowing in the wind: Perceiving structure in trees and bushes. *Cognition* 12:25–44.

Cutting, J. E., Proffitt, D. R. 1982. The minimum principle and the perception of absolute, common, and relative motions. *Cognit. Psychol.* 14.

Cynader, M., Regan, D. 1978. Neurons in the cat parastriate cortex sensitive to the direction of motion in three–dimensional space. *J. Physiol.* 274:549–569.

Cynader, M., Regan, D. 1982. Neurons in cat visual cortex tuned to the direction of motion in depth: effect of positional disparity. *Vision Res.* 22:967–982.

Davis, L., Wu, Z., Sun, H. 1983. Contour–based motion estimation. *Comput. Vision Graphics Image Proc.* 23:313–326.

DeVoe, R. D., Guy, R. G., Criswell, M. H. 1985. Directionally selective cells of the inner nuclear layer in the turtle retina. *Invest. Ophthalmol. Vis. Sci. Suppl.* 26:311.

Diener, H. C., Wist, E. R., Dichgans, J., Brandt, Th. 1976. The spatial frequency effect on perceived velocity. *Vision Res.* 16:169–176.

Doner, J., Lappin, J. S., Perfetto, G. 1984. Detection of three–dimensional structure in moving optical patterns. *J. Exp. Psychol.: Human Percept. Perform.* 10:1–11.

Doorn, A. J. van, Koenderink, J. J. 1982. Visibility of movement gradients. *Biol. Cybern.* 44:167–175.

Doorn, A. J. van, Koenderink, J. J. 1983. Detectability of velocity gradients in moving random–dot patterns. *Vision Res.* 23:799–804.

Dow, B. M. 1974. Functional classes of cells and their laminar distribution in monkey visual cortex. *J. Neurophysiol.* 37:927–946.

Dowling, J.E. 1979. Information processing by local circuits: the vertebrate retina as a model system. In *The Neurosciences: Fourth Study Program*, Schmitt, F.O., and F.G. Worden, eds., Cambridge: MIT Press, pp. 163–181.

Eckert, H. E. 1973. Optomotorische Untersuchungen am visuellen System der Stubenfliege *Musca Domestica L.*. *Kybernetik* 14:1–23.

Egelhaaf, M. 1985. On the neuronal basis of figure–ground discrimination by relative motion in the visual system of the fly: III. Possible input circuitries and behavioural significance of the FD–cells. *Biol. Cybern.*, 52:267–280.

Famiglietti, E. V., 1983. On and Off pathways through amacrine cells in mammalian retina: the synaptic connections of "starburst" amacrine cells. *Vision Res.*, 23:1265–1279.

Famiglietti, E. V., Kolb, H. 1976. Structural basis for on– and off-centre reponses in retinal ganglion cells. *Science*, 194:193–195.

Fennema, C. L., Thompson, W. B. 1979. Velocity determination in scenes containing several moving objects. *Comput. Graph. Image Proc.* 9:301–315.

Fermi, G., Reichardt, W. 1963. Optomotorische Reaktionen der Fliege *Musca Domestica* L. *Kybernetik*, 2:15–28.

Frisby, J. P. 1972. The effect of stimulus orientation on the phi phenomenon. *Vision Res.* 12:1145–1166.

Frost, B. J. 1978. Moving background patterns after directionally specific responses of pigeon tectal neurons. *Brain Res.* 151:599–603.

Frost, B. J., Nakayama, K. 1983. Single visual neurons code opposing motion independent of direction. *Science* 220:744–745.

Frost, B. J., Scilley, P. L., Wong, S. C. P. 1981. Moving background patterns reveal double opponency of directionally specific pigeon tectal neurons. *Exp. Brain Res.* 43:173–185.

Geman, S., Geman, D. 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian Restoration of Images. *IEEE Trans. Patt. Anal. Machine Intell.*, 6:721–741.

Gibson, J. J. 1950. *The Perception of the Visual World.* Boston: Houghton Mifflin.

Gibson, J. J., Gibson, E. J. 1957. Continuous perceptive transformations and the perception of rigid motion. *J. Exp. Psychol.* 54:129–138.

Götz, K. G. 1965. Die optischen Übertragungseigenschaften der Komplexaugen von *Drosophila*. *Kybernetik*, 2:215–221.

Götz, K. G. 1968. Flight control in *Drosophila* by visual perception of motion. *Kybernetik*, 4:199–208.

Götz, K. G. 1972. Principles of optomotor reactions in insects. *Bibl. Ophthal.* 82:251–259.

Green, B. F. 1961. Figure coherence in the kinetic depth effect. *J. Exp. Psychol.* 62:272–282.

Green, M., von Grünau, M. 1983. Real and apparent motion: one mechanism or two? *Proc. ACM Interdisc. Workshop on Motion: Representation and Perception*, Toronto, Canada, pp. 17–22.

Gregory, R. L. 1985. Movement nulling: for heterochromatic photometry and isolating channels for 'real' and 'apparent' motion. *Perception* 14:193–196.

Grzywacz, N. M., Hildreth, E. C. 1985. The incremental rigidity scheme for recovering structure from motion: positions vs. velocity based formulations. *MIT Artif. Intell. Memo* 845. Also to appear in *J. Opt. Soc. Amer. A.*

Grzywacz, N. M., Yuille, A. 1986. Motion correspondence and analog networks. *MIT Artif. Intell. Memo 845.*

Hassenstein, B., Reichardt, W. E. 1956. Functional structure of a mechanism of perception of optical movement. In *Proc. First Intl. Cong. Cybernetics*, Namar, pp. 797–801.

Hausen, K. 1982a. Movement sensitive interneurons in the optomotor system of the fly. I. The horizontal cells: structure and signals. *Biol. Cybern.*, 45:143–156.

Hausen, K. 1982b. Movement sensitive interneurons in the optomotor system of the fly. II. The horizontal cells: receptive field organization and response characteristics *Biol. Cybern.*, 46:67–79.

Hay, C. J. 1966. Optical motions and space perception — an extension of Gibson's analysis. *Psychol. Rev.* 73:550–565.

Hildreth, E. C. 1984. *The Measurement of Visual Motion.* Cambridge: MIT Press.

Hildreth, E. C., Hollerbach, J. M. 1985. The computational approach to vision and motor control. *MIT Artif. Intell. Memo* 846. Also in *Handbook of Physiology*, Vol. 5, Sect. 1, ed. F. Plum, American Physiological Society, *in press.*

Hoffman, D. D. 1982. Inferring local surface orientation from motion fields. *J. Opt. Soc. Am.* 72:888–892.

Hoffman, D. D., Flinchbaugh, B. E. 1982. The interpretation of biological motion. *Biol. Cybern.* 42:195–204.

Holden, A. L. 1977. Responses of directional ganglion cells in the pigeon retina. *J. Physiol.*, 270:253–269.

Horn, B. K. P., Schunck, B. G. 1981. Determining optical flow. *Artif. Intell.* 17:185–203.

Hubel, D. H., Wiesel, T. N. 1959. Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.*, 148:574–591.

Hubel, D. H., Wiesel, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154.

Inada, V. K., Hildreth, E. C., Grzywacz, N. M., Adelson, E. H. 1986. The perceptual buildup of three-dimensional structure from motion. *Invest. Ophthal. Visual Sci. Suppl.* 27:142.

Ito, M. 1984. *The cerebellum and neural control.* New York: Raven Press.

Jansson, G., Johansson, G. 1973. Visual perception of bending motion. *Perception* 2:321–326.

Jensen, R. J., DeVoe, R. D. 1983. Comparisons of directionally selective with other ganglion cells of the turtle retina: intracellular recording and staining. *J. Comp. Neurol.* 217:271–287.

Jerian, C., Jain, R. 1984. Determining motion parameters for schemes with translation and rotation. *IEEE Trans. Patt. Anal. Machine Intell.* PAMI–6:523–530.

Johansson, G. 1964. Perception of motion and changing form. *Scand. J. Psychol.* 5:181–208.

Johansson, G. 1971. Studies on visual perception of locomotion. *Perception* 6:365–376.

Johansson, G. 1973. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14:201–211.

Johansson, G. 1975. Visual motion perception. *Sci. Am.* 232:76–88.

Johansson, G. 1977. Spatial constancy and motion in visual perception. In *Stability and Constancy in Visual Perception*, ed. W. Epstein, New York: Wiley.

Jones, E. G., Powell, T. P. S. 1969. Morphological variations in the dendritic spines of the neocrotex. *J. Cell Sci.*, 5:509–529.

Julesz, B. 1971. *Foundations of Cyclopean Perception.* Chicago: Univ. Chicago Press.

Kanatani, K. 1985. Structure from motion without correspondence: general principle. *Proc. Image Understanding Workshop*, Miami, FL, pp. 107–116.

Kaplan, G. A. 1969. Kinetic disruption of optical texture: The perception of depth at an edge. *Percept. Psychophys.* 6:193–198.

Kirschfeld, K. 1972. The visual system of *Musca*: studies on optics, structure and function. In *Information Processing in the Visual System of Arthropods*, R. Welner, ed., Berlin: Springer–Verlag, pp. 61–74.

Koch, C., 1986. What's in the term connectionist? *Behav. Brain Sci.* 9:100–101.

Koch, C., Marroquin, J., Yuille, A. 1986a. Analog "neuronal" networks in early vision. *Proc. Natl. Acad. Sci. USA* 83:4263–4267.

Koch, C., Poggio, T. 1983. A theoretical analysis of electrical properties of spines, *Proc. R. Soc. Lond. B.* 218:455–477.

Koch, C., Poggio, T. 1985. The synaptic veto mechanism: does it underlie direction and orientation selectivity in the visual cortex? In *Models of the Visual Cortex*, D. Rose and V. Dobson, eds., Sussex: John Wiley, pp. 408–419.

Koch, C., Poggio, T. 1986. Biophysics of Computational Systems: Neurons, Synapses and membranes. In: *New Insights into Synaptic Function*, G. M. Edelman, W. E. Gall and W. M. Cowan, editors, Neurosciences Research Foundation and John Wiley and Sons, *in press.*

Koch, C., Poggio, T., Torre, V. 1982. Retinal ganglion cells: A functional interpretation of dendritic morphology. *Philos. Trans. R. Soc. B.* 298: 227–264.

Koch, C., Poggio, T., Torre, V. 1983. Nonlinear interaction in a dendritic tree: localization, timing and role in information processing. *Proc. Natl. Acad. Sci. USA* 80: 2799–2802.

Koch, C., Torre, V., Poggio, T. 1986b. Computations in the vertebrate retina: motion discrimination, gain enhancement and differentiation. *Trends Neurosci.* 9:204–211.

Koenderink, J. J., van Doorn, A. J. 1976. Local structure of movement parallax of the plane. *J. Opt. Soc. Am.* 66:717–723.

Koenderink, J. J., van Doorn, A. J. 1986. Depth and shape from differential perspective in the presence of bending deformations. *J. Opt. Soc. Am. A* 3:242–249.

Kolers, P. A. 1972. *Aspects of Motion Perception.* New York: Pergamon.

Lappin, J. S., Bell, H. H. 1976. The detection of coherence in moving random dot patterns. *Vision Res.* 16:161–168.

Lawton, D. T. 1983. Processing translational motion sequences. *Comput. Vision Graph. Image Proc.* 22:116–144.

Lee, D. N. 1980. The optic flow field: The foundation of vision. *Philos. Trans. R. Soc. London B* 290:169–179.

Lee, D. N., Aronson, E. 1974. Visual proprioceptive control of standing in human infants. *Percept. Psychophys.* 15:529–532.

Levinson, E., Sekuler, R. 1976. Adaptation alters perceived direction of motion. *Vision Res.* 16:779–781.

Limb, J. O., Murphy, J. A. 1975. Estimating the velocity of moving images in television signals. *Comput. Graph. Image Proc.* 4:311–327.

Longuet-Higgins, H. C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature* 293:133–135.

Longuet-Higgins, H. C. 1983. The role of the vertical dimension in stereoscopic vision. *Perception* 11:377–386.

Longuet-Higgins, H. C. 1984. The visual ambiguity of a moving plane. *Proc. R. Soc. Lond. B.* 223:165–175.

Longuet–Higgins, H. C., Prazdny, K. 1981. The interpretation of moving retinal images. *Proc. R. Soc. London Ser. B* 208:385–397.

Marchiafava, P. L. 1979. The responses of retinal ganglion cells to stationary and moving visual stimuli. *Vis. Res.* 19:1203–1211.

Marr, D. 1969. A theory of cerebellar cortex. *J. Physiol.* 202:437–470.

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

Marr, D., Hildreth, E. C. 1980. Theory of edge detection. *Proc. R. Soc. London Ser. B* 207:187–217.

Marr, D., Poggio, T. 1977. From understanding computation to understanding neural circuitry. *Neurosci. Res. Prog. Bull.* 15:470–488.

Marr, D., Ullman, S. 1981. Directional selectivity and its use in early visual processing. *Proc. R. Soc. London Ser. B* 211:151–180.

Masland, R.H., Mills, W., Cassidy, C. 1984. The functions of acetylcholine in the rabbit retina. *Proc. Roy. Soc. Lond. B*, 223:121–139.

Massey, S.C., Neal, M. J. 1979. The light evoked release of ACh from the rabbit retina *in vivo* and its inhibition by GABA. *J. Neurochem.* 32:1327–1329.

Mather, G., Cavanagh, P., Anstis, S. M. 1985. A moving display which opposes short–range and long–range signals. *Perception* 14:163–166.

Maturana, H., Frenk, S. 1963. Directional movement and horizontal edge detectors in the pigeon retina. *Science*, 142:977–979.

Maturana, H. R., Lettvin, J. Y., McCulloch, W. S., Pitts, W. H. 1960. Anatomy and physiology of vision in the frog (*Rana pipiens*). *J. Gen. Physiol.* 43:(Suppl. 2):129–171.

Maunsell, J. H. R., Van Essen, D. C. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed and orientation. *J. Neurophysiol.* 49:1127–1147.

Michael, C. R. 1966. Receptive fields of directionally selective units in the optic nerve of the ground squirrel. *Science*, 152:1092–1095.

McCulloch, W. S., Pitts, W. 1943. A logical calculus of ideas immanent in neural nets. *Bull. Math. Biophys.* 5:115–137.

McKee, S. 1981. A local mechanism for differential velocity detection. *Vision Res.* 21:491–500.

Miezin, F., McGuinness, E., Allman, J. 1982. Antagonistic direction specific mechanisms in area MT in the owl monkey. *Neurosci. Abstr.* 8:681.

Mikami, A., Newsome, W. T., Wurtz, R. H. 1986. Motion selectivity in macaque visual cortex: II. Spatio–temporal range of directional interactions in MT and V1. *J. Neurophys.* 55:1308–1327.

Miller, R. F., Bloomfield, S. A. 1983 Electroanatomy of an unique amacrine cell in the rabbit retina. *Proc. Natl. Acad. Sci. USA*, 80: 3069–3073.

Minsky, M., Papert, S. 1969. *Perceptrons: an Introduction to Computational Geometry.* Cambridge: MIT Press.

Mitiche, A. 1984. Computation of optical flow and rigid motion. *Proc. Workshop on Computer Vision: Representation and Control*, Annapolis, MD, pp. 63–71.

Mitiche, A. 1986. On kineopsis and computation of structure and motion. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI–8:109–112.

Mitiche, A., Seida, S., Aggarwal, J. K. 1985. Determining position and displacement in space from images. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, June, pp. 504–509.

Morgan, M. J. 1985. Computational vision. In *Textbook Series in Psychology*, London: British Psychological Society.

Motter, B. C., Mountcastle, V. B. 1981. The functional properties of the light–sensitive neurons in the posterior parietal cortex studied in waking monkeys: foveal spacing and opponent vector organization. *J. Neurosci.* 1:3–26.

Moulden, B., Begg, H. 1984. *Perception*, 13:A18.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., Newsome, W. T. 1985. The analysis of moving visual patterns. In *Pattern Recognition Mechanisms*, ed. C. Chagas, R. Gattas, C. G. Gross. Rome: Vatican Press.

Murray, D. W., Buxton, B. F. 1984. Reconstructing the optic flow field from edge motion: an examination of two different approaches. *Proc. First Conf. AI Applications*, Denver.

Mutch, K., Smith, I. M., Yonas, A. 1983. The effect of two–dimensional and three–dimensional distance on apparent motion. *Perception* 12:305–312.

Mutch, K. M., Thompson, W. B., 1985. Analysis of accretion and deletion at boundaries in dynamic scenes. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI–7:133–138.

Nagel, H.–H. 1984. Recent advances in image sequence analysis. *Proc. Premier Colloque Image — Traitement, Synthese, Technologie et Applications*, Biarritz, France, May, pp. 545–558.

Nagel, H.–H., Enkelmann, W. 1984. Towards the estimation of displacement vector fields by "oriented smoothness" constraints. *Proc. 7th Int. Conf. on Pattern Recognition*, Montreal, Canada, July, pp. 6–8.

Nagel, H.–H., Enkelmann, W. 1986. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Patt. Anal. Machine Intell.* PAMI–8:565–593.

Nakayama, K. 1981. Differential motion hyperacuity under conditions of common image motion. *Vision Res.* 21:1475–1482.

Nakayama, K. 1985. Biological motion processing: a review. *Vision Res.* 25:625–660.

Nakayama, K., Loomis, J. M. 1974. Optical velocity patterns, velocity–sensitive neurons, and space perception: a hypothesis. *Perception* 3, 63–80.

Nakayama, K., Silverman, G. H. 1983. Perception of moving sinusoidal lines. *J. Opt. Soc. Am. A* 72.

Nakayama, K., Silverman, G. H. 1984a. Propagation of velocity information along moving contours. *J. Opt. Soc. Am. A* 1:1266.

Nakayama, K., Silverman, G. H. 1984b. Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Res.* 24:293–299.

Navon, D. 1976. Irrelevance of figural identity for resolving ambiguities in apparent motion. *J. Exp. Psychol., Human Percep. Perform.* 2:130–138.

Negahdaripour, S., Horn, B. K. P. 1985. Direct passive navigation. *MIT Artif. Intell. Memo* 821.

Newsome, W. T., Mikami, A., Wurtz, R. H. 1982. Direction selective responses to sequentially flashed stimuli in extrastriate area MT in the awake macaque monkey. *Neurosci. Abstr.* 8:812.

Newsome, W. T., Mikama, A., Wurtz, R. H. 1986. Motion selectivity in macaque visual cortex: III. Psychophysics and physiology of apparent motion. *J. Neurophys.* 55:1340–1358.

Newsome, W. T., Wurtz, R. H., Dursteler, M. R., Mikami, A. 1985. Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *J. Neurosci.* 5:825–840.

O'Donnell, P., Koch, C., Poggio, T. 1985. Demonstrating the nonlinear interaction between excitation and inhibition in dendritic trees using computer-generated color graphics: a film. *Neurosci. Abstr.* 11:142.1.

Orban, G. A., Gulyas, B., Vogels, R. 1986. Influence of a moving textured background on direction selectivity of cat striate neurons. *J. Neurophsyiology, in press.*

Orban, G. A., Kennedy, H., Maes, H. 1981. Responses to movement of neurons in areas 17 and 18 of the cat: direction selectivity. *J. Neurophysiol.*, 45:1059–1073.

Oyster, C. W., Takahashi, E., Collewijn, H. 1972. Direction-selective retinal ganglion cells and control of optokinetic nystagmus in the rabbit. *Vision Res.*, 13:183–193.

Pantle, A. J., Picciano, L. 1976. A multistable display: evidence for two separate motion systems in human vision. *Science* 193:500–502.

Petersik, J. T. 1980. The effect of spatial and temporal factors on the perception of stroboscopic rotation stimulations. *Perception* 9:271–283.

Petersik, J. T., Pantle, A. 1979. Factors controlling the competing sensations produced by a bistable stroboscopic motion display. *Vision Res.* 19:143–154.

Poggio, T. 1982. Visual algorithms. In *Physical and Biological Processing of Images*, eds. O. J. Braddick, A. C. Sleigh, Berlin: Springer–Verlag, pp. 128–153.

Poggio, T. 1984. Vision by man and machine. *Sci. Am.* 250:106–115.

Poggio, G. F., Talbot, W. H. 1981. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J. Physiol.* 315:469–492.

Poggio, T., Koch, C. 1985. Ill–Posed Problems in Early Vision: from Computational Theory to Analog Networks. *Proc. R. Soc. London B* 226:303–323.

Poggio, T., Reichardt, W. E. 1973. Considerations on models of movement detection. *Kybernetik* 13:223–227.

Poggio, T., Reichardt, W. 1976. Visual control of orientation behaviour in the fly: Part II: Towards the underlying neural interactions. *Quart. Rev. Biophysics*, 9:377–438.

Poggio, T., Torre, V., Koch, C. 1985. Computational Vision and Regularization Theory. *Nature*, 317:314–319.

Potter, J. L. 1977. Scene segmentation using motion information. *Comput. Graph. Image Proc.* 6:558–581.

Prazdny, K. 1980. Egomotion and relative depth map from optical flow. *Biol. Cybern.* 36:87–102.

Prazdny, K. 1983. On the information in optical flows. *Comput. Vision Graph. Image Proc.* 22:239–259.

Ramachandran, V. S. 1985. Apparent motion of subjective surfaces. *Perception* 14:127–134.

Ramachandran, V. S., Anstis, S. M. 1983. Extrapolation of motion path in human visual perception. *Vision Res.* 23:83–85.

Ramachandran, V. S., Anstis, S. M. 1985. Perceptual organization in multistable apparent motion. *Perception* 14:135–143.

Ramachandran, V. S., Gregory, R. L. 1978. *Nature* 275:55–56.

Ramachandran, V. S., Rao, V. M., Vidyasagar, T. R. 1973. Apparent motion with subjective contours. *Vision Res.* 13:1399–1401.

Regan, D., Beverley, K. I. 1979. Visually guided locomotion: Psychophysical evidence for neural mechanisms sensitive to flow patterns. *Science* 205:311–313.

Regan, D., Beverley, K. I. 1983. Visual fields for frontal plane motion and changing size. *Vision Res.* 23:673–676.

Regan, D., Spekreijse, H. 1970. Electrophysiological correlate of binocular depth perception in man. *Nature* 225:92–94.

Reichardt, W. 1969. Movement perception in insects. In: *Processing of optical data by organisms and machines*, W. Reichardt, ed., London, New York: Academic press, pp. 465–493.

Reichardt, W., Guo, A.-K. 1986. Elementary pattern discrimination (behavioral experiments with the fly *Musca domestica*). *Biol. Cybern.* 53:285–306.

Reichardt, W., Poggio, T. 1979. Figure–ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.* 35:81–100.

Reichardt, W., Poggio, T., Hausen, K. 1983. Figure–ground discrimination by relative movement in the visual system of the fly. Part II: Towards the neural circuitry. *Biol. Cybern.* 46:1–30.

Richards, W., Lieberman, H. R. 1982. Velocity blindness during shearing motion. *Vision Res.* 22:97–100.

Riley, M. D. 1981. The representation of image texture. *MIT Artif. Intell. Tech. Rep.* AI–TR–649.

Roach, J. W., Aggarwal, J. K. 1980. Determining the movement of objects from a sequence of images. *IEEE Proc. Pattern Anal. Machine Intell.* PAMI–2:554–562.

Rodieck, R. W. 1979. Visual pathways. *Ann. Rev. Neurosci.*, 2:193–226.

Rogers, B. J., Graham, M. 1979. Motion parallax as an independent cue for depth perception. *Perception* 8:125–134.

Saito, H.-A., Yukie, M., Tanaka, K., Hikosaka, K., Fukuda, Y., Iwai, E. 1986. Interation of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci.*, 6:145–157.

Sakata, H., Shibutani, H., Kawano, K., Harrington, T. L. 1985. Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vis. Res., in press.*

Schiller, P. H. 1982. Central connections of the retinal ON and OFF pathways. *Nature* 297:580–583.

Schiller, P. H., Finlay, B. L., Volman, S. F. 1976. Quantitative studies of single–cell properties in monkey striate cortex. I Spatiotemporal organization of receptive fields. *J. Neurophysiol.*, 49:1288–1319.

Schunck, B. G. 1984. The motion constraint equation for optical flow. *Intl. Conf. Patt. Recog.*, Montreal, Canada.

Schunck, B. G., Horn, B. K. P. 1981. Constraints on optical flow computation. *Proc. IEEE Conf. Patt. Recog. Image Proc.*, August, pp. 205–210.

Scott, G. L. 1986. Local and global interpretation of moving images. PhD. Thesis, Univ. Sussex, Cognitive Studies Program.

Segev, I., Fleshman, J. W., Miller, J. P., Bunow, B. 1985. Modeling the electrical behavior of anatomically complex neurons using a network analysis program: passive membrane. *Biol. Cybern.*, 53:27–40.

Segev, I., Parnas, I. 1983. Synaptic integration mechanisms. *Biophys. J.*, 41:41–50.

Shepard, R. N., Judd, S. A. 1976. Perceptual illusion of rotation of three–dimensional objects. *Science* 191:952–954.

Sherk, H., Horton, J. C. 1984. Receptive–field properties in the cat's area 17 in the absence of on–center geniculate input. *J. Neurosci.*, 4:374–380.

Siegel, R. M., Andersen, R. A. 1986. Motion perceptual deficits following ibotenic acid lesions of the middle temporal area (MT) in the behaving rhesus monkey. *Neurosc. Abstr.* 12:324.8.

Sillito, A. M. 1977. Inhibitory processes underlying the directional specificity of simple, complex and hypercomplex cells in the cat's visual cortex. *J. Physiol.*, 271:699–720.

Sillito, A. M., Kemp, J. A., Milson, J. A., Berardi, N. 1980. A re–evaluation of the mechanism underlying simple cell orientation selectivity. *Brain Res.*, 194:517–520.

Simpson, J. I. 1984. The accessory optic system. *Ann. Rev. Neurosci.*, 7:13–41.

Sloper, J. J., Powell, T. P. S. 1979. An experimental electron microscopic study of afferent connections to the primate motor and somatic sensory cortices. *Philos. Trans. Roy. Soc. Lond. B*, 285:199–226.

Sperling, G., Pavel, M., Cohen, Y., Landy, M. S., Schwartz, B. J. 1983. Image processing in perception and cognition. In *Physical and Biological Processing of Images*, eds. O. J. Braddick, A. C. Sleigh, Berlin: Springer–Verlag.

Sterling, P., Wickelgren, B. G. 1969. Visual receptive fields in the superior colliculus of the cat. *J. Neurophysiol.* 32:1–15.

Stone, J., Fabian, M. Specialized receptive fields of the cat's retina. *Science*, 152:1277–1279.

Strauss, G., von Seelen, W. 1986. Contribution of suprasylvian cortex to pattern recognition in the cat. *In preparation.*

Subbarao, M., Waxman, A. M. 1985. On the uniqueness of image flow solutions for planar surfaces in motion. *Proc. Workshop on Computer Vision: Representation and Control*, Bellaire, MI, October, pp. 129–140.

Sugie, N., Inagaki, H. 1984. A computational aspect of kinetic depth effect. *Biol. Cybern.* 50:431–436.

Tarr, M. J., Pinker, S. P. 1985. Nearest neighbors in apparent motion: Two or three dimensions? *Proc. Annual Meeting Psychonomics Soc.*, Boston, November, p. 19.

Tauchi, M., Masland, R. H. 1984. The shape and arrangement of the cholinergic neurons in the rabbit retina. *Proc. R. Soc. Lond. B*, 223:101–119.

Ternus, J. 1926. Experimentelle Untersuchung über phänomenale Identität. *Psychologische Forschung* 7:81–136. Translated in *A Source Book of Gestalt Psychology*, ed. W. D. Ellis, New York: Humanities Press, 1967.

Thompson, W. B. 1980. Combining motion and contrast for segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI–2:543–549.

Thompson, W. B., Barnard, S. T. 1981. Lower–level estimation and interpretation of visual motion. *IEEE Computer*, August, pp. 20–28.

Thompson, W. B., Mutch, K. M., Berzins, V. 1982. Edge detection in optical flow fields. *Proc. Second Nat. Conf. Artif. Intell.*, August, 26–29.

Thompson, W. B., Mutch, K. M., Berzins, V. 1985. Dynamic occlusion analysis in optical flow fields. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI–7:374–383.

Todd, J. T. 1982. Visual information about rigid and nonrigid motion: A geometric analysis. *J. Exp. Psychol.* 8:238–252.

Todd, J. T. 1984. The perception of three–dimensional structure from rigid and nonrigid motion. *Percept. Psychophys.* 36:97–103.

Torre, V., Poggio, T. 1978 A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. London Ser. B* 202:409–416.

Traub, R. D., Knowles, W. D., Miles, R., Wong, R. K. S. 1984. Synchronized afterdischarges in the hippocampus: simulation studies of the cellular mechanism. *Neurosci.*, 12:1191–1200.

Tsai, R. Y., Huang, T. S. 1981. Uniqueness and estimation of three–dimensional motion parameters of rigid objects with curved surfaces. *Univ. Illinois Urbana–Champaign, Coordinated Science Laboratory Report* R–921.

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

Ullman, S. 1981a. Analysis of visual motion by biological and computer systems. *IEEE Computer*, August, pp. 57–69.

Ullman, S. 1981b. The effect of similarity between line segments on the correspondence strength in apparent motion. *Perception* 9:617–626.

Ullman, S. 1983. Computational studies in the interpretation of structure and motion: summary and extension. In *Human and Machine Vision*, ed. J. Beck, B. Hope, A. Rosenfeld. New York: Academic, 459–480.

Ullman, S. 1984. Maximizing rigidity: the incremental recovery of 3-D structure from rigid and rubbery motion. *Perception* 13:255–274.

Ullman, S. 1985. The optical flow of planar surfaces. *MIT Artif. Intell. Memo* 870.

Ullman, S. 1986. Artificial Intelligence and the brain: Computational studies of the visual system. *Ann. Rev. Neurosci.*, 9:1–26.

Van Essen, D. C., Maunsell, J. H. R. 1983. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci.*, 6:370–375.

Van Santen, J. P. H., Sperling, G. 1984. A temporal covariance model of motion perception. *J. Opt. Soc. Am. A*, 1:451–473.

Van Santen, J. P. H., Sperling, G. 1985. Elaborated Reichardt detectors. *J. Opt. Soc. Am. A*, 2:300–320.

Varju, D., Reichardt, W. 1967. Übertragungseigenschaften im Auswertesystem für das Bewegungssehen. *Z. Naturforsch.*, 22b:1343–1351.

Vaughn, J. E., Famiglietti, E. V., Barber, R. P., Saito, K., Roberts, E., Ribak, C. E. 1981. GABAergic amacrine cells in rat retina: immunocytochemical identification and synaptic connectivity. *J. comp. Neurol.*, 197:113–127.

Wallach, H. 1976. On perceived identity: 1. The direction of motion of straight lines. In *On Perception*, ed. H. Wallach. New York: Quadrangle.

Wallach, H., O'Connell, D. N. 1953. The kinetic depth effect. *J. Exp. Psych.* 45:205–217.

Wallach, H., Weisz, A., Adams, P. A. 1956. Circles and derived figures in rotation. *Am. J. Psych.* 69:48–59.

Warren, S., Hamalainen, H. A., Gardner, E. P. 1986. Coding of the spatial period of gratings rolled across the receptive fields of somatosensory cortical neurons in awake monkeys. *J. Neurophysiol., in press.*

Watanabe, S.-I., Murakami, M. 1984. Synaptic mechanisms of directional selectivity in ganglion cells of frog retina as revealed by intracellular recordings. *Jap. J. Physiol.* 34:497–511.

Watson, A. B., Ahumada, A. J. 1985. Model of human visual–motion sensing. *J. Opt. Soc. Am. A* 2:322–341.

Waxman, A. M. 1986. Image flow theory: A framework for 3–D inference from time–varying imagery. In *Advances in Computer Vision*, ed. C. Brown, New Jersey: Erlbaum, *in press.*

Waxman, A. M., Ullman, S. 1985. Surface structure and three–dimensional motion from image flow kinematics. *J. Robotics Res.* 4:72–94.

Waxman, A. M., Wohn, K. 1985. Contour evolution, neighborhood deformation and global image flow: planar surfaces in motion. *Int. J. Robotics Res.* 4:95–108.

Webb, J. A., Aggarwal, J. K. 1981. Visually interpreting the motions of objects in space. *Computer* 14:40–49.

Werblin, F. S., 1970. Response of retinal cells to moving spots: intracellular recording in *Necturus Maculosus. J. Neurophysiol.*, 33:342–350.

White, B. W., Mueser, G. E. 1960. Accuracy in reconstructing the arrangement of elements generating kinetic depth displays. *J. Exp. Psychol.* 60:1–11.

Wohn, K. 1984. *A contour–based approach to image flow.* PhD. Thesis, Univ. Maryland, Dept. Computer Science, September.

Wohn, K., Waxman, A. M. 1985. Contour evolution, neighborhood deformation and local image flow: curved surfaces in motion. *Univ. Maryland Center for Automation Research Tec. Rep.* 134, July.

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 10/26/95

Report # AIM-919

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☒ Technical Report (TR) ☐ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 61 (66-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number): FOLLOWS TITLE PAGE

Photographs/Tonal Material (by page number):_____

Other (note description/page number):

Description :	Page Number:
IMAGE MAP · (1-60) UNNUMBERED TITLE PAGE, BLANK	1-59
(62-66) SCANCONTROL, DOD, TRGT'S (3)	

Scanning Agent Signoff:

Date Received: 10/26/95 Date Scanned: 11/6/95 Date Returned: 11/9/95

Scanning Agent Signature: Michael N. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AI Memo 919	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER: AD-A190388	
4. TITLE (and Subtitle): The Analysis of Visual Motion: From Computational Theory to Neuronal Mechanisms	5. TYPE OF REPORT & PERIOD COVERED: memorandum
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Ellen C. Hildreth and Christof Koch	8. CONTRACT OR GRANT NUMBER(s): N00014-85-C-0038
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Inteligence Laboratory, 545 Technology Square, Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, VA 22209	12. REPORT DATE: December 1986 13. NUMBER OF PAGES: 59
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, VA 22217	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

motion analysis
3-D vision
object segmentation
velocity field
motion measurement
image analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reviews a number of aspects of visual motion analysis in biological systems, from a computational perspective. We illustrate the kinds of insights that have been gained through computational studies and how these observations can be integrated with experimental studies from psychology and the neurosciences, to understand the particular computations used by biological systems to analyze motion. The particular areas of motion analysis that we discuss include early motion detection and measurement, the optical flow computation, motion correspondence, the detection of motion discontinuities, and the recovery of three-dimensional structure from motion.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 11/06/1995

M.I.T. Libraries
Document Services