Model of Visual Motion Sensing

David J. Heeger
Department of Psychology
Stanford University
Stanford, California 94305
(415) 723-4048

Eero P. Simoncelli
The Media Laboratory
and
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

To appear in Spatial Vision in Humans and Robots
L. Harris and M. Jenkin, editors
Cambridge University Press

March, 1992

Published in: Spatial Vision in Humans and Robots,
eds. L. Harris and M. Jenkin,
chapter 19, pages 367-392,
Cambridge University Press, 1993.

Abstract

A number of researchers have proposed models of early motion sensing based on direction-selective, spatiotemporal linear operators. Others have formalized the problem of measuring optical flow in terms of the spatial and temporal derivatives of stimulus intensity. Recently, the spatiotemporal filter models and the gradient-based methods have been placed into a common framework. In this chapter, we review that framework and we extend it to develop a new model for the computation and representation of velocity information in the visual system. We use the model to simulate psychophysical data on perceived velocity of sine-grating plaid patterns, and to simulate physiological data on responses of simple cells in primary (striate) visual cortex.

1 Introduction

More than forty years ago, Gibson (1950, 1957) noted that visual motion perception is essential for an observer's ability to explore and interact with his/her environment. As an observer moves and explores the environment, the visual stimulation in his/her eye is constantly changing. Somehow he/she is able to perceive the spatial layout of the scene, and to discern his/her movement through space. Imagine, for example, that you are watching a scene from a movie that was shot with the camera in motion. The visual stimulation in your eye is an array of light that changes over time, yet you experience a sense of moving through a three dimensional space.

Since Gibson's initial work, perception of motion has been studied extensively by researchers in the fields of visual psychophysics, visual neurophysiology, and computational vision. It is now well-known that the visual system has mechanisms that are specifically suited for analyzing motion (see Nakayama 1985, for review), and that human observers are capable of recovering accurate information about the world (e.g., three-dimensional trajectory, relative distance, shape) from visual motion (e.g., Wallach and O'Connell 1953; Johansson 1975; Warren and Hannon 1988, 1990).

The first stage of motion perception is generally believed to be the measurement of optical flow. Optical flow is a field of two-dimensional velocity vectors, indicating the speed and direction of motion for each small region of the visual field.

A number of machine vision algorithms have been developed for measuring optical flow fields from sequences of (e.g., video) images. At the same time, psychophysicists and neurophysiologists have performed experiments to study the manner by which people and animals sense velocity. Little effort, however, has gone into integrating the results from the three disciplines of computational vision, visual psychophysics, and visual neurophysiology.

In this chapter, we describe a model for the computation and representation of velocity information in the primate visual system that accounts for a variety of psychophysical and physiological observations. We use the model to simulate psychophysical data on perceived velocity of sine-grating plaid patterns, and to simulate physiological data on responses of simple cells in primary (striate) visual cortex.

2 The Model

In this section, we review two algorithms for measuring flow fields, the gradient-based methods and the spatiotomporal filtering methods. Following Adelson and Bergen (1986), and Simoncelli and Adelson (1991a, 1991b), we show that these two methods can be expressed in a common mathematical framework. Finally, we introduce some extensions to this framework to develop our new model of biological motion sensing.

2.1 Gradient-Based Methods

Researchers (Horn and Schunk 1981; Lucas and Kanade 1981; Nagel 1987; and others) have proposed algorithms that compute flow from the spatial and temporal derivatives of intensity. Following the standard gradient formulation, we assume that the stimulus is shifted (locally translated) over time, and that the shifted intensity values are conserved. This intensity conservation assumption is expressed as follows:

$$f(x, y, t) = f(x + v_1, y + v_2, t + 1), \tag{1}$$

where $f(x, y, t)$ is stimulus intensity as a function of space and time, and $\mathbf{v} = (v_1, v_2)$ is velocity. Note that this intensity conservation assumption is only approximately true in practice. For example, it ignores possible changes in intensity due to lighting changes.

We further assume that the time-varying stimulus intensity is well approximated by a first-order Taylor series expansion:

$$f(x + v_1, y + v_2, t + 1) \approx f(x, y, t) + v_1 f_x(x, y, t) + v_2 f_y(x, y, t) + f_t(x, y, t),$$

where f_x, f_y, and f_t are the spatial and temporal derivatives of stimulus intensity. Substituting this approximation into equation (1) gives:

$$v_1 f_x(x, y, t) + v_2 f_y(x, y, t) + f_t(x, y, t) = 0. \tag{2}$$

This equation relates the velocity, at one point in the visual field, to the spatial and temporal derivatives of stimulus intensity. We refer to equation (2) as the *gradient constraint*.

Combining Constraints. It is impossible to recover velocity, given the gradient constraint at only a single position, since equation (2) offers only one linear constraint to solve for the two unknown components of velocity. Gradient-based methods solve for velocity by combining information over a spatial region. The different gradient-based methods use different combination rules. A particularly simple rule for combining constraints from two nearby spatial positions is:

$$\begin{bmatrix} f_x(x_1, y_1, t) & f_y(x_1, y_1, t) \\ f_x(x_2, y_2, t) & f_y(x_2, y_2, t) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} f_t(x_1, y_1, t) \\ f_t(x_2, y_2, t) \end{bmatrix} = \mathbf{0}, \tag{3}$$

where the two coordinate pairs (x_i, y_i) correspond to the two spatial positions. Each row of equation (3) is the gradient constraint for one spatial position. Solving this equation simultaneously for both positions gives the velocity that is consistent with both constraints.

Lucas and Kanade (1981) suggested combining constraints from more than just two spatial positions, by squaring and summing:

$$R(v_1, v_2) = \sum_{x,y} [v_1 f_x(x, y, t) + v_2 f_y(x, y, t) + f_t(x, y, t)]^2. \tag{4}$$

Each squared term in the summation is a constraint on the flow from a different (nearby) position. The summation is taken over a local spatial region, e.g., in a Gaussian weighted window. Since there are now more constraints than unknowns, there may not be a solution that satisfies all of the constraints simultaneously. In other words, $R(v_1, v_2)$ will typically be non-zero for all (v_1, v_2). The choice of (v_1, v_2) that minimizes $R(v_1, v_2)$ is the least squares estimate of velocity.

Least Squares Estimate. One way to find the minimum of $R(v_1, v_2)$ is to evaluate the function at a number of points (say, on a fixed square grid) and to pick the smallest result. Figure 1 shows some examples. Figures 1(a) and (b) depict sine-grating plaid stimuli. The component gratings in the two stimuli have different orientations and spatial frequencies, but the speeds of the component gratings were chosen so that both plaids moved rightward with the same velocity. Figures 1(c) and (d) show $R(v_1, v_2)$ for (a) and (b), respectively. Each point in (c) and (d) corresponds to a different velocity, with the center of each image corresponding to zero velocity. Brightness at each point is inversely proportional to $R(v_1, v_2)$, and the locations of the peaks correspond to the velocity estimates. The peaks correspond to the correct velocity in both cases, despite of the difference in the spatial structures of the two stimuli[1].

** Figure 1 About Here **

Since equation (4) is a quadratic expression, there is a simple analytical expression for the velocity estimate. The solution is derived by taking derivatives of equation (4) with respect to v_1 and v_2, and setting them equal to zero:

$$\begin{aligned} \frac{\partial R(v_1, v_2)}{\partial v_1} = \sum_{xy} [v_1 (f_x)^2 + v_2 (f_x f_y) + (f_x f_t)] &= 0 \\ \frac{\partial R(v_1, v_2)}{\partial v_2} = \sum_{xy} [v_2 (f_y)^2 + v_1 (f_x f_y) + (f_y f_t)] &= 0 \end{aligned}$$

These equations may be rewritten as a single equation in matrix notation:

$$\mathbf{M} \cdot \mathbf{v} + \mathbf{b} = \mathbf{0},$$

where

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

and where

$$\begin{aligned} m_{11} &= \sum (f_x)^2 \\ m_{22} &= \sum (f_y)^2 \\ m_{12} &= \sum (f_x f_y) \\ b_1 &= \sum (f_x f_t) \\ b_2 &= \sum (f_y f_t). \end{aligned}$$

The least-squares solution is then given by

$$\hat{\mathbf{v}} = -\mathbf{M}^{-1}\mathbf{b}, \tag{5}$$

presuming that $\mathbf{M}$ is invertible.

Aperture Problem. When the matrix $\mathbf{M}$ in equation (5) is singular (or ill-conditioned), there are not enough constraints to solve for both unknowns. This situation corresponds to what has been called the aperture problem. For some patterns (e.g., a very gradual curve) there is not enough information in a local region (small aperture) to disambiguate the true direction of motion. For other patterns (e.g., an extended grating or edge) the information is insufficient regardless of the aperture size.

The latter case is illustrated in figure 2(a). The diagonal line indicates the locus of velocities compatible with the motion of the grating. At best, we may extract only one of the two velocity components. Figure 2(b) shows how the motion is disambiguated when there is more spatial structure. The plaid pattern illustrated in figure 2(b) is composed of two moving gratings. The lines give the possible motion of each grating alone. Their intersection is the only shared motion.

** Figure 2 About Here **

Combining the gradient constraints according to the summation in equation (3) is related to the intersection of constraints rule depicted in figure 2. The gradient constraint, equation (2), is linear in both v_1 and v_2. Given measurements of the derivatives, (f_x, f_y, f_t), there is a line of possible solutions for (v_1, v_2), analogous to the constraint line illustrated in figure 2(a). For each different position, there will generally be a different constraint line. Equation (3) gives the intersection of these constraint lines, analogous to figure 2(b).

Prior Bias. To deal with the aperture problem, we could consider combining constraints over a larger spatial area (e.g., Horn and Schunk 1981). Instead, we add a slight prior preference for slower speeds. The resulting velocity estimate is approximately equal to the *normal flow*, the component of motion parallel to the spatial intensity gradient.

The prior preference is implemented by adding a small offset to each of the diagonal entries of $\mathbf{M}$. Elsewhere (Simoncelli, Adelson, and Heeger, 1991), we formally prove that adding this offset gives a Bayesian estimate for velocity. The Bayesian estimator incorporates a prior likelihood for each possible velocity. The offset is the (inverse) variance of this prior probability distribution. Adding the offset yields a slight bias toward lower speeds. The bias is greater for low contrast stimuli, i.e., when the entries of $\mathbf{M}$ are small.

** Figure 3 About Here **

Figure 3 illustrates the effect of the prior. Figure 3(a) shows $-R(v_1, v_2)$, from equation (4), for a drifting sine grating stimulus. Since the velocity of the grating is ambiguous (due to the aperture problem), there is no peak in the distribution. Rather the distribution is shaped like a ridge. Any velocity along this ridge is an equally good interpretation of the stimulus' motion. Figure 3(b) shows that including the prior gives a distribution with a broad peak. The location of the peak corresponds approximately to the normal flow, but at a very slightly slower speed.

2.2 Space-Time Filtering Methods

In this section, we reformulate the gradient-based flow algorithm, this time in terms of biological mechanisms. We first review the spatiotemporal linear model of biological motion sensing. Then we relate that model to the gradient method.

Space-Time Orientation A number of authors have proposed models of biological motion sensing based on direction selective, spatiotemporal linear operators (Fahle and Poggio 1981; Watson and Ahumada 1983, 1985; Adelson and Bergen 1985; van Santen and Sperling 1985; Heeger 1987, 1988; Grzywacz and Yuille 1990). . These authors have explained that visual motion is like orientation in space-time, and that spatiotemporally-oriented, linear operators can be used to detect and measure it.

** Figure 4 About Here **

Figure 4 shows a simple example. Figure 4(a) depicts a vertical bar moving to the right over time. Imagine that we film a movie of this stimulus and stack the consecutive frames one after the next. We end up with a three-dimensional volume (space-time cube) of intensity data like that shown in figure 4(b). Figure 4(c) shows an x-t slice through this space-time cube. The slope of the edges in the x-t slice equals the horizontal component of the bar's velocity (change in position over time). Different speeds correspond to different slopes.

Spatiotemporal Linear Operators. The response of a linear operator is expressed as a weighted sum, over local space and recently past time, of the stimulus intensities. Specifically, the response, $L(t)$, is the inner product in space and the convolution in time of a stimulus, $f(x, y, t)$, with the spatiotemporal weighting function of the operator, $g(x, y, t)$:

$$L(t) = \int \int \int_{-\infty}^{\infty} g(x, y, \tau) f(x, y, \tau - t)\, dx\, dy\, d\tau. \tag{6}$$

The triple integral in the above equation is simply a weighted sum of the stimulus intensities over space and time.

The linear operators that we consider in this chapter have weighting functions with positive and negative subregions. The positive and negative weights are balanced, so the operators give no output for a constant intensity stimulus. Rather, their responses are proportional to stimulus contrast, for stimuli that vary in intensity over space and/or time.

The spatiotemporal weighting function of a linear operator determines its selectivity (e.g., for orientation or direction of motion). A linear operator is direction selective if its subregions are tilted along an oblique axes in space-time. For example, figure 4(c) illustrates the weighting function of a direction selective operator, that responds preferentially to rightward motion.

A spatial array of identical linear operators (sampling the entire visual field) can be thought of as a linear filter that performs a convolution (over both space and time) with the stimulus,

$$g(x,y,t) * f(x,y,t) = \int\int\int_{-\infty}^{\infty} g(\xi,\eta,\tau) f(\xi - x, \eta - y, \tau - t)\, d\xi\, d\eta\, d\tau,$$

where $*$ means convolution.

Space-time Filters and the Gradient Method. Following Adelson and Bergen (1986), and Simoncelli and Adelson (1991a, 1991b), we now show that the gradient-based solution can be expressed in terms of the outputs of a set of space-time oriented linear operators. To this end, note that the derivative operators may be written as convolutions. Furthermore, we can prefilter the stimuli to extract some spatiotemporal subband, and perform the analysis on that subband. Consider, for example, prefiltering with a space-time Gaussian function. Abusing the notation somewhat, we define:

$$f_x(x,y,t) \equiv \tfrac{\partial}{\partial x}[g(x,y,t) * f(x,y,t)] = g_x(x,y,t) * f(x,y,t),$$

where $*$ is convolution and g_x is the x-derivative of a Gaussian. In words, we compute f_x by convolving with g_x, a spatiotemporal linear filter. We compute f_y and f_t similarly.

Note also that derivatives in oblique space-time orientations can be expressed as linear sums of f_x, f_y, and f_t. For example, the derivative of a Gaussian in a diagonal spatial orientation is given by:

$$g_p = (g_x + g_y),$$

where g_p is a diagonally oriented derivative operator. Finally, note that products of derivatives in the x-, y-, and t- directions can be written as combinations of the obliquely oriented derivatives. For example,

$$\begin{aligned} 4 f_x f_y &= (f_x + f_y)^2 - (f_x - f_y)^2 \\ &= [(g_x + g_y) * f]^2 - [(g_x - g_y) * f]^2. \end{aligned}$$

Now we rewrite the entries of $\mathbf{M}$ and $\mathbf{b}$ in terms of a set of squared linear filter outputs:

$$\begin{aligned} m_{11} &= \sum (f_x)^2 \\ m_{22} &= \sum (f_y)^2 \\ m_{12} &= \tfrac{1}{4} \sum [(f_x + f_y)^2 - (f_x - f_y)^2] \\ b_1 &= \tfrac{1}{4} \sum [(f_x + f_t)^2 - (f_x - f_t)^2] \\ b_2 &= \tfrac{1}{4} \sum [(f_y + f_t)^2 - (f_y - f_t)^2]. \end{aligned} \tag{7}$$

In primary visual cortex, there are no cells with receptive fields that behave like products of derivatives (e.g., $f_x f_y$). Thus, rewriting the solution as in equation (7) brings us closer to a model of the physiology. Each linear filter in equation (7) is orientation tuned, with oriented spatial subregions. Four of the operators are direction selective with weighting functions that are tilted obliquely in space-time, e.g., $(g_x + g_t)$ and $(g_x - g_t)$ are selective for leftward and rightward motion.

The linear operators in equation (7) are, therefore, similar to the receptive fields of cortical cells. There are, however, some important differences. As we shall see (Section 3.1), higher order derivative operators are a better model of cortical receptive fields.

2.3 Using Higher Order Derivatives

In this section, we extend the spatiotemporal filter method to use higher order derivative operators.

Consider using g_{xx}, g_{xy} and g_{yy} as prefilters and writing three gradient constraint equations, in terms of derivatives of each these prefilters:

$$\begin{aligned} v_1 f_{xxx} + v_2 f_{xxy} + f_{xxt} &= 0 \\ v_1 f_{xxy} + v_2 f_{xyy} + f_{xyt} &= 0 \\ v_1 f_{xyy} + v_2 f_{yyy} + f_{yyt} &= 0, \end{aligned} \tag{8}$$

where f_{xxx} should be interpreted as $f * g_{xxx}$, and likewise for the other derivatives. Equation (8), written in terms of third derivatives, gives three constraints on velocity.

The gradient constraint, equation (2), is based on the intensity conservation assumption; i.e., it assumes that the stimulus intensity shifts (locally translates) from location to location over time. The third derivative constraints, equation (8), are based on conservation of the second spatial derivatives of intensity, i.e., that (f_{xx}, f_{xy}, f_{yy}) shifts over time.

An advantage of using higher order derivatives is that, in principle, there are enough constraints at a single spatial position. Even so, there are stimuli for which there will

not be enough constraints locally. There is still a need, therefore, to combine constraints over a local spatial region.

Combining constraints over a local spatial region gives:

$$\begin{aligned} R(v_1, v_2) &= \sum_{x,y} [v_1 f_{xxx} + v_2 f_{xxy} + f_{xxt}]^2 \\ &+ \sum_{x,y} [v_1 f_{xxy} + v_2 f_{xyy} + f_{xyt}]^2 \\ &+ \sum_{x,y} [v_1 f_{xyy} + v_2 f_{yyy} + f_{yyt}]^2 \end{aligned}$$

The least-squares estimate of velocity, minimizing this expression is

$$\hat{\mathbf{v}} = -\mathbf{M}^{-1}\mathbf{b},$$

where $\mathbf{M}$ and $\mathbf{b}$ are now defined as:

$$\begin{aligned} m_{11} &= \sum[(f_{xxx})^2 + (f_{xxy})^2 + (f_{xyy})^2] \\ m_{22} &= \sum[(f_{xxy})^2 + (f_{xyy})^2 + (f_{yyy})^2] \\ m_{12} &= \sum[(f_{xxx})(f_{xxy}) + (f_{xxy})(f_{xyy}) + (f_{xyy})(f_{yyy})] \\ b_1 &= \sum[(f_{xxx})(f_{xxt}) + (f_{xxy})(f_{xyt}) + (f_{xyy})(f_{yyt})] \\ b_2 &= \sum[(f_{xxy})(f_{xxt}) + (f_{xyy})(f_{xyt}) + (f_{yyy})(f_{yyt})]. \end{aligned}$$

Note the similarity with equation (5). The solutions using first and third derivates are essentially the same. The main differences are: (1) that the third derivative solution uses a greater number of linear operators, and (2) that the third derivative operators are more narrowly tuned (with more subregions) for spatiotemporal orientation.

As with first derivatives, each element of $\mathbf{M}$ and $\mathbf{b}$ may be rewritten as a sum of squared outputs of spatiotemporally-oriented operators. As above, we rewrite the products, e.g.,

$$(f_{xxx})(f_{xxt}) = \tfrac{1}{4}[(f_{xxx} + f_{xxt})^2 - (f_{xxx} - f_{xxt})^2].$$

For the third derivative operators, we also rewrite the spatial cross-derivatives (e.g., f_{xxy} and f_{xyy}) in terms of spatially oriented operators. To this end, we define g_p and g_q to be to be derivative operators in diagonal orientations,

$$\begin{aligned} g_p &= g_x + g_y \\ g_q &= g_x - g_y. \end{aligned}$$

The third derivatives in the diagonal orientations are:

$$\begin{aligned} g_{ppp} &= g_{xxx} + 3g_{xxy} + 3g_{xyy} + g_{yyy} \\ g_{qqq} &= g_{xxx} - 3g_{xxy} + 3g_{xyy} - g_{yyy}. \end{aligned}$$

Spatial cross-derivative operators may then be expressed in terms of the oriented operators:

$$\begin{aligned} g_{xxy} &= \tfrac{1}{6}\left[g_{ppp} - g_{qqq} + 2g_{yyy}\right] \\ g_{xyy} &= \tfrac{1}{6}\left[g_{ppp} + g_{qqq} + 2g_{xxx}\right]. \end{aligned}$$

Using a set of identities like these, we can express the velocity estimate in terms of the squared outputs of a set of spatiotemporally-oriented operators. Figure 6 shows the spatiotemporal weighting functions of a representative set of those operators.

On the other hand, we have no *a priori* theoretical basis for choosing Gaussian third derivatives. Other operators could be used just as well (e.g., third or fourth derivatives of some smooth, unimodal, non-Gaussian function). One set of operators or another may provide a stronger constraint on velocity in different situations, depending on the local image structure. For machine vision applications, we advocate using several prefilters, with different preferences for spatial frequency (scale), different orientation tuning widths, and different (e.g., even and odd) phases.

2.4 Normalization and Rectification

The model that we advocate in this chapter is an extension of the spatiotemporal filter method described above. In this section, we briefly describe two additional steps in the computation of the model — normalization and rectification. Both extensions are needed for a realistic model of physiological data.

Rectification. The linear model of simple cell physiology is attractive because the response of a linear operator can be completely characterized with a relatively small number of measurements. Unfortunately, the linear model falls short of a complete account of simple cell responses. One major fault with the linear model is that cell firing rates are by definition positive, whereas linear operators can have positive or negative outputs.

A linear cell with a high maintained firing rate could encode the positive and negative values by responding either more or less than the maintained rate. Cells in primary visual cortex, however, have very little maintained discharge so they can not truly act as linear operators.

Rather, the positive and negative outputs can be encoded by two halfwave-rectified operators. One mechanism encodes the positive outputs of the underlying linear operator, and the other one encodes the negative outputs. These two mechanisms are complements of one another, that is, the positive weights of one weighting function are replaced by negative weights in the other. Due to the rectification, only one of the two has a non-zero response at any given time.

In this chapter, we consider half-squaring as an alternative form for the rectification. The output of a half-squared linear operator is given by:

$$A(t) = \lfloor L(t) \rfloor^2 , \tag{9}$$

where $\lfloor x \rfloor = \max(x, 0)$ is halfwave-rectification, and $L(t)$ is the linear response defined in equation (6).

Normalization. A second major fault with the linear model of simple cells is the fact that cell responses saturate at high contrasts. The responses of ideal linear operators, on the other hand, increase proportionally to stimulus contrast over the entire range of contrasts. To explain response saturation, several researchers (Robson, 1988; Bonds, 1989; Heeger, 1992a) have suggested that cells in primary visual cortex mutually inhibit one another, effectively normalizing their responses with respect to stimulus contrast.

Normalization of striate cell responses is also motivated from a theoretical point of view. It is commonly believed that information about a visual stimulus, other than its contrast, is represented as the relative responses of collections of cells. For example, the orientation of a grating might be represented as the ratio of the responses of two cells, each with a different orientation tuning. Indeed physiologists have found that the ratio of a cell's responses to two stimuli is largely independent of stimulus contrast (see Section 3.1). But cortical cells, unlike linear operators, have a limited dynamic range: their responses saturate for high contrasts. Normalization makes it possible for response ratios to be independent of stimulus contrast, even in the face of response saturation.

Consider a collection of half-squared linear operators with various receptive field centers (covering the visual field) and with various spatiotemporal frequency tunings. Let $A_i(t)$ be the squared output of mechanism i. Normalization is achieved by dividing each output by the sum of all of the outputs:

$$\overline{A_i}(t) = \frac{A_i(t)}{\sigma^2 + \sum_i A_i(t)}, \tag{10}$$

where σ^2 is called the semi-saturation constant. As long as σ is nonzero, the normalized output will always be a value between 0 and 1, saturating for high contrasts.

The underlying linear operators can be chosen so that they tile the frequency domain, i.e., the sum of their squared frequency responses is the unit constant function (everywhere equal to one). In that case, summing the squared outputs over all spatial positions and all frequencies gives the total Fourier energy of the stimulus. The normalization can also be computed "locally" by summing over a limited region of space and a limited range of frequencies.

There is a problem with normalization, as it has been presented thus far. Equations (9) and (10) express the normalization in a feed-forward manner. First, the half-squared outputs are computed, using equation (9). Then the half-squared outputs are combined

to give the normalized outputs, using equation (10). However, the unnormalized outputs can not be represented by mechanisms with limited dynamic range (e.g., neurons). The solution is to use a feedback network to do the normalization so that the unnormalized outputs need not be explicitly represented as cell output firing rates (see Heeger 1992a, for details).

3 Results

In the previous section, we describe a model for computing velocity from visual stimuli. In this model, velocity estimates are computed from the outputs of a set of normalized, half-squared, linear operators. The normalized outputs are summed to get the entries of $\mathbf{M}$ and $\mathbf{b}$. In addition, a small offset (the prior) is added to the diagonal entries of $\mathbf{M}$. Finally, the velocity estimate is given $-\mathbf{M}^{-1}\mathbf{b}$.

This section reports on simulations of both physiological and psychophysical experiments. We show that our model explains a variety of experimental results.

3.1 Simple Cell Physiology

For over thirty years, physiologists have been measuring response properties of simple cells in primary (striate) visual cortex. A longstanding view of simple cells is that their responses can be characterized as a weighted sum (over local space) of the intensity values in a visual stimulus (Hubel and Wiesel, 1962; Campbell et al., 1968, 1969). A currently popular model of simple cells is that they act like halfwave-rectified, spatiotemporal linear operators. However, some experiments have revealed blatant violations of linearity.

The model that we advocate is based on spatiotemporal linear operators, but with two important modifications. First, the outputs of the linear operators are half-squared (not halfwave-rectified). Second, the responses are normalized. Heeger (1992a, 1992b) has demonstrated that this new model, with half-squaring and normalization, is qualitatively consistent with a significantly larger body of physiological data.

In this section, we review some measurements of simple cell responses. First, we compare physiological data with the Gaussian third derivative operators. We conclude that the third derivative operators are a reasonable model for the linear weighting functions that underlie simple cells responses. Then, we demonstrate that response saturation can be explained by the nonlinearities (half-squaring and normalization) in the model.

Responses to Impulses. Many researchers have used impulses (flashed spots or bars) and white noise stimuli to map simple cell weighting functions (e.g., Hubel and Wiesel, 1962; Heggelund, 1981; Jones and Palmer, 1987; McLean and Palmer, 1989;

Shapley et al. 1991). Here, we compare physiological data with the weighting functions of the third derivative operators.

Hubel and Weisel (1962) discovered that simple cells have clearly defined excitatory and inhibitory spatial subregions. Bright (brighter than the mean intensity) light in an excitatory region or dim (darker than the mean) light in an inhibitory region enhances the cell's response, whereas bright light in an inhibitory region or dim light in excitatory region inhibits its response. These results are readily explained by the model. The underlying linear stage of the model predicts that excitation to a bright light is complemented by inhibition to a dim light. Due to rectification, the inhibition can be measured only by first driving the operator to a nonzero response with an excitatory stimulus.

According to the model, cells are direction selective because of the underlying linear stage. McLean and Palmer (1989) and Shapley et al. (1991) measured full 3D spatiotemporal weighting functions of simple cells using white-noise stimuli. They found some simple cells with weighting functions tilted along an oblique axis in space-time, like that illustrated in figure 4(c). The model predicts that these cells be direction selective, that is, that they prefer motion in one direction over the other. In fact, since a spatiotemporal linear operator is completely characterized by its impulse response, the model allows one to predict a cell's preferred direction and speed of motion from the cell's spatiotemporal weighting function. When McLean and Palmer (1989) measured simple cell responses to moving bars, they could, for most cells, correctly predict the preferred bar motion from the weighting function.

** Figure 5 About Here **

McLean and Palmer (1989) and Shapley et al. (1991) also found some simple cells with space-time separable weighting functions. Space-time separable functions can be expressed as the product of a spatial function multiplied by a temporal function. In the model, the direction selective linear operators are constructed by summing space-time separable operators. For example, $(g_{xxx} + g_{xxt})$ is a third derivative operator that is selective for leftward motion. This operator is constructed by summing the outputs of two linear operators, g_{xxx} and g_{xxt}. Figure 5 shows space-time slices through the weighting functions of each of these three operators. Although g_{xxx} and g_{xxt} are each space-time separable, their sum is tilted in space-time (not space-time separable)[2].

** Figure 6 About Here **

Figure 6 shows examples of other linear operators used in the model. The top row shows spatial slices through the weighting functions, and the bottom row shows space-time slices through the weighting functions. The operators depicted in figure 6 are representative of all of the operators used in the model. Some of these operators are Gaussian third derivatives (like g_{xxx} and g_{xxt}), while others are constructed by summing

third derivatives (like $g_{xxx} + g_{xxt}$). The outputs of these operators (and others like them, there are a total of 46) are half-squared, normalized, and then summed to give the entries of $\mathbf{M}$ and $\mathbf{b}$.

For the most part, these linear operators resemble physiological measurements of simple cell weighting functions, such as those measured by McLean and Palmer (1989), Shapley et al. (1991), and others. First, all of the operators in figure 6 are spatially oriented, with two or more spatial subregions. Second, some of the operators are direction selective (they are tilted in space-time), while others are not direction selective (they are space-time separable). Third, the operators have temporal responses that are either monophasic (like g_{xxx}) or biphasic (like g_{xxt} or $g_{xxx} + g_{xxt}$). And fourth, there is quite a lot of variability in the model's weighting functions[3].

There are, however, some differences between the model operators and simple cell weighting functions. First, some of the operators have irregular spatial structure (e.g., second and third from the left in figure 6). Second, some of the operators have impulse responses that rotate slightly over time. For example, the operator farthest to the right in figure 6 has an impulse response that rotates first clockwise by $\pi/4$ radians, and then counter-clockwise by the same amount. Simple cells with this property have not been reported in the literature.

Responses to Gratings. The response of a spatiotemporal linear operator, to a drifting grating, varies sinusoidally over time with the same temporal frequency as that of the stimulus. A halfwave-rectified linear operator responds over only half of each cycle, remaining silent during the other half-cycle. A half-squared operator also responds over only half of each cycle, but the shape of the response waveform is distorted. Simple cells, like rectified linear operators, also respond over approximately half of each cycle (Movshon et al., 1978; Andrews and Pollen, 1979; Kulikowski and Bishop, 1981b).

Spatiotemporal linear operators, like the linear operators in the model, respond preferentially to gratings with certain orientations, spatial frequencies, and temporal frequencies. In other words, the linear operators are tuned for spatial frequency, temporal frequency, and orientation. The tuning curves of the operators can be computed by taking the Fourier transform of the operator's weighting functions. In this section, we compare simple cell tuning curves with those of first and third derivative operators. In contrast with the third derivative operators, the first derivative operators are not a satisfactory model of simple cell weighting functions for two reasons:

1. There are two few spatial subregions in the first derivative operators. In other words, they are too broadly tuned for orientation and spatial frequency.

2. Researchers have found that a simple cell's spatial frequency tuning (measured with gratings drifting only in one direction) is largely independent of the stimulus temporal frequency. This is not the case for the first derivative operators, but it

is very nearly true for the third derivative operators.

** Figure 7 About Here **

Figure 7(a) shows a series of spatial frequency tuning curves, measured from a simple cell (data replotted from Hamilton et al., 1989). Note that the shape of the spatial frequency curves are largely independent of temporal frequency. Other physiologists (Tolhurst and Movshon, 1975; Holub and Morton-Gobson, 1981; Ikeda and Wright, 1975; Foster et al., 1985) have noted this same result, that spatial frequency and temporal frequency tuning curves (measured with gratings drifting only in one direction) are independent of one another[4].

Figure 7(b) shows an analogous series of spatial frequency tuning curves for one of the model's third derivative operators. Like the simple cell data, these simulated tuning curves are largely independent of temporal frequency. Figure 7(c), on the other hand, shows a series of tuning curves for one of the first derivative operators. The spatial frequency tuning of the first derivative operator is much broader and it shifts systematically as a function of temporal frequency.

** Figure 8 About Here **

Figure 8(a) shows the orientation/direction tuning of a simple cell (data replotted from Movshon et al., 1986). Figure 8(b) shows an analogous tuning curve for one of the model's third derivative operators. Although there are some differences (the model operator responds slightly to motion in the non-preferred direction), the shape of the tuning curve is quite similar. Figure 8(c), on the other hand, shows that for a first derivative operator, the tuning curve is much broader.

Responses to Plaids. Movshon et al. (1986) also measured direction tuning curves for sine-grating plaid patterns. Figure 9(a) shows an example of their results, for a typical cell in primary visual cortex. The plaid stimuli consisted of a pair of orthogonal gratings, each of the cell's preferred spatial and temporal frequency. For each different stimulus condition, the entire plaid pattern was rotated so that it moved in a different direction. Figure 9(b) shows an analogous tuning curve for one of the third derivative operators, and figure 9(c) shows the tuning curve for a first derivative operator. The first derivative operator is so broadly tuned that it does not respond independently to the two component gratings.

** Figure 9 About Here **

Movshon et al. (1986) classified cells into two types (component-flow and pattern-flow) by observing their responses to sine-grating plaid stimuli. Component-flow cells

respond independently to each of the component gratings. Pattern-flow cells do not respond independently to the components. According to this classification, the third derivative operator would be classified as a component-flow cells, and the first derivative operator would be classified as a pattern-flow cell. Movshon et al., however, found that all cells in primary visual cortex are of the component-flow type. Pattern-flow cells were found in a different area of primate visual cortex, area MT.

Figure 9 raises some doubt about the interpretation of these experimental results. Movshon et al.'s (1986) argued that pattern-flow cells respond to the direction of motion of the plaid as a whole, i.e., to the intersection of constraints direction. The result in figure 9(c) suggests that this might not be the case. The first derivative operator is not solving for the intersection of constraints. Rather since it is very broadly tuned for orientation/direction, the first derivative operator responds to the average of the two component directions. Moreover, Movshon et al. (1986) did not find a sharp dichotomy between component- and pattern-flow cells. It might be that the continuum of component/pattern types reflects a continuum of orientation tuning widths.

Response Saturation. The contrast-response function is a plot of response as a function of contrast, typically measured using sine-grating stimuli. Here, we demonstrate that contrast-response of simple cells can be explained by the nonlinearities (divisive normalization and half-squaring) in our model.

** Figure 10 About Here **

Figure 10(a) plots typical experimental contrast-response data, and figure 10(b) shows results of model simulations. The simulated responses saturate with increased contrast because of normalization.

The hyperbolic ratio function,

$$R = R_{\max}\frac{c^n}{\sigma^n + c^n} + M, \tag{11}$$

has been used to fit contrast-response data, for cells in both cat and primate (Albrecht and Hamilton, 1982; Chao-yi and Creutzfeldt, 1984; Sclar et al., 1990): R, in equation 11 is the evoked response, c is the contrast of the test grating, M is maintained discharge, n is a constant exponent, σ^n is the semi-saturation constant, and $R_{\max}$ is the maximum attainable response. From the fits, experimenters have found that the exponent, n, is 2 on average (Albrecht and Hamilton, 1982; Sclar et al., 1990).

The contrast-response of a model cell is given exactly by the hyperbolic ratio with parameters $n = 2$ and $M = 0$. This is easily demonstrated by recalling that the summation, $\sum A_i(t)$, in the denominator of equation (10) is proportional to c^2. The exponent is 2 in the model because of half-squaring.

In addition, it can be shown that the contrast-response curve of a model cell shifts mostly downward (on log-log axes) if the orientation or frequency of the test grating is non-optimal (see Heeger, 1992a for details). This downward shift is again due to divisive normalization in the model. Downward shifts of contrast-response have been measured physiologically in several labs. Albrecht and Hamilton (1982), for example, measured contrast-response curves for stimuli of non-optimal spatial frequency. Their data is replotted in figure 10(a), and figure 10(b) shows the contrast-response curves of a model cell. For both model cells and real cells, the curves shift mostly downward.

This downward shift of contrast-response has important consequences. Consider the response of a linear operator when presented with two different stimuli. If both stimuli are multiplied by the same factor then the ratio of the responses to the two stimuli remains unchanged. The downward shift in figure 10 demonstrates that this is also true for normalized operators and for real cells. In spite of saturation, the response ratio to two different stimuli is largely independent of stimulus contrast. In this way, information about a visual stimulus, other than its contrast, is represented as the relative responses of a collection of cells.

3.2 Perceived Velocity of Plaids

The perceived velocity of a moving pattern depends on its spatial structure. Adelson and Movshon (1982) conducted psychophysical experiments to study this dependence using sine-grating plaid patterns. Since then, a number of other psychophysicists have measure human velocity judgements using plaids. Stone et al. (1990), in particular, measured the effect of contrast on perceived direction. By varying the relative contrasts of the two component gratings, they found that the plaid motion direction is biased away from the intersection of constraints rule (illustrated in figure 2b), toward the higher contrast grating. In this section, we show that our model is consistent with their data.

The nominal stimulus in this experiment was a sine-grating plaid made up of two component gratings with equal contrasts and temporal frequencies. This plaid stimulus appeared to move directly upward (in accordance with the intersection of constraints rule). Stone et al. varied both the relative contrast and the relative temporal frequency of the two gratings. These stimuli (with different contrasts or temporal frequencies) appeared to move either slightly right of vertical or slightly left of vertical. The subject's task was to indicate, for each stimulus presentation, whether the plaid appeared to move rightward or leftward.

The total contrast of the plaid was also varied (total contrast was defined by Stone et al. to be the sum of the contrasts of the two components). For each total contrast, Stone et al. varied the contrast ratio of the two components. For each contrast ratio, they adjusted the relative temporal frequency (in a staircase procedure) until the pattern appeared to move directly upward. In other words, they varied the relative temporal frequency to compensate for the bias introduced by the relative contrast difference.

Figure 11(a) shows data from Stone et al. 1990, averaged over four subjects. Each curve is the inferred bias, for a fixed total contrast, as a function of contrast ratio.

** Figure 11 About Here **

There are two parameters in the model, the semi-saturation constant for the normalization, and the prior. Both of the parameters in the model contribute to deviations from the intersection of constraints rule. If the normalized responses are small relative to the prior, then there is a large bias. If the normalized responses are large, then there is a small bias. For appropriate values of the two parameters the model behaves like human observers, as shown in figure 11(b).

On the other hand, there are differences between the simulation results and the actual data. At the highest total contrast (40%) and for small contrast ratios, the human observers often saw the the plaid motion direction biased toward the *lower* contrast grating. This is evident in figure 11(a) where the 40% curve dips below zero bias. For all of the conditions that we have simulated, the model predicts a bias toward the higher contrast grating.

4 Summary

This chapter presents a model for the computation and representation of velocity information in the primate visual system that accounts for a variety psychophysical and physiological observations. The first stage of the model uses spatiotemporal linear operators to compute a linear sum of the stimulus intensities over a local region of space and recently past time. The outputs of the linear operators are half-squared and then normalized. A slight prior preference for slower speeds is introduced by adding a small offset to two of the normalized outputs. The normalized outputs are then combined, according to a simple formula, to give final velocity estimates.

Our model is consistent with recent psychophysical experiments by Stone et al. (1990) on the perception of sine-grating plaid velocities. When the component grating contrasts are unequal, the velocity estimated by the model is biased toward the higher contrast grating. The bias occurs in the model because the model includes a slight prior (preference) for slower speeds. For appropriate values of the model's two parameters, the model behaves like human observers (figure 11).

Ferrera and Wilson (1990, 1991) have also measured perceived speed and direction of plaids. We are currently working toward explaining their psychophysical results with the same model (Simoncelli and Heeger, 1992).

Our model is also consistent with physiological data on responses of simple cells in primary (striate) visual cortex. In this chapter, simple cells are modeled as normalized,

half-squared linear operators. We consider two sets of linear operators, first and third spatiotemporal derivatives of a Gaussian. Although somewhat more cumbersome because of the larger number of filters, the third derivative operators are a better model of simple cells than the first derivative operators. The third derivative operators are consistent with a variety of physiological results:

- According to the model, a simple cell's selectivity is due to an underlying spatiotemporal, linear stage. There are a variety of physiological results that are consistent with the linear hypothesis (see Heeger, 1992b for review). McLean and Palmer (1989), in particular, were able to predict a cell's preferred speed and direction of motion from measurements of its underlying spatiotemporal weighting function.

- The third derivative operators in the model resemble simple cell weighting functions (figures 5 and 6). First, all of the operators are spatially oriented, with two or more spatial subregions. Second, some of the operators are direction selective (they are tilted in space-time), while others are not direction selective (they are space-time separable). Third, the operators have temporal responses that are either monophasic or biphasic. And fourth, there is quite a lot of variability in the model's weighting functions.

- Researchers (e.g., Hamilton et al., 1990) have found that spatial frequency and temporal frequency tuning curves are largely independent of one another. This is approximately true of the third derivative operators as well (figure 7).

- The third derivative operators have orientation tuning curves that resemble those of real cells (figure 8).

- The third derivative operators are sufficiently narrowly tuned for orientation, so that they act like "component-flow" cells (figure 9), responding independently to each component of a sine-grating plaid stimulus.

- Responses of both model cells and real cells saturate at high contrasts, according to the hyperbolic ratio function (figure 10).

- The contrast-response curve, for either a model cell or a real cell, shifts mostly downward for non-optimal stimuli (figure 10). In other words, the ratio of responses produced by two different stimuli is largely invariant with respect to stimulus contrast. In this way, information about a visual stimulus, other than its contrast, is represented as the relative responses of a collection of cells.

Our model has also been used to compute optical flow fields from image sequences (Simoncelli, Adelson, and Heeger 1991). It is important to keep in mind, however, that the gradient constraint, equation (2), is only approximately valid. The constraint is based on the intensity conservation assumption, that changes in intensity are due

only to local translation. This ignores possible changes in lighting and reflectance. Moreover, the assumption of local translation is not valid near motion boundaries nor for transparent motions. The gradient constraint is also based on a planar approximation to the (prefiltered) intensity values. The velocity estimated by the model is in error when these assumptions are not satisfied.

In our future research, we plan to extend the model to make it more robust with respect to these assumptions. We also plan to use the model to explain further experimental results. From our point of view, fitting psychophysical or physiological data is not, by itself, a satisfactory goal of computational modeling. The model must also give reliable velocity estimates. Although primates do not always perceive velocity veridically (e.g., figure 11), we do quite well for most stimuli.

Footnotes

1. Some algorithms do not always compute the correct velocity for sine grating plaid patterns. In particular, models proposed by Watson and Ahumada (1985), by Heeger (1987), and by Grzywacz and Yuille (1990) give the wrong solution unless the spatial frequencies of the gratings equal the preferred spatial frequency of the filters. Grzywacz and Yuille (1990) claim that their method does not depend on the spatial frequency content of the stimulus, but in fact that claim is not true for sine grating stimuli.

2. Watson and Ahumada (1983, 1985) and Adelson and Bergen (1985) proposed the quadrature model of direction selectivity, in which direction selective linear operators are constructed by summing the outputs of two space-time separable subunits. These subunits are related to one another by a quadrature phase shift both in space and in time. In our model, the direction selective operators are also constructed by summing the outputs of two space-time separable subunits, but the subunits are not quadrature pairs.

3. An extension of the model would predict even greater variability in the weighting functions. The model operators need not be Gaussian derivatives. Other operators could also be used (e.g., third or fourth derivatives of some smooth, unimodal, non-Gaussian function). Moreover, different prefilters could be used at different spatial positions. At a given position, the operators must all be derivatives of a common prefilter, but the prefilters at different spatial positions need not be the same.

4. Figure 7 demonstrates that spatial and temporal frequency tuning curves (measured with gratings drifting only in one direction) are largely independent of one another. Some researchers have summarized this result by saying that the spatiotemporal frequency tuning is "space-time separable". Note, however, that this is different from requiring space-time separability of an operator's weighting function. The spatiotemporal frequency tuning (for gratings drifting only in one direction) can be separable even if the weighting function is inseparable. The frequency domain measurements (in figure 7) are separable only when considering one direction of motion. The full spatiotemporal frequency tuning (for gratings drifting in all directions) is space-time separable if and only if the weighting function is space-time separable (i.e., nondirection selective).

Acknowledgements

This research was supported by NASA RTOP 506-71-51, by NASA-Ames grant NCC2-307, and by the MIT Media Lab. This paper benefited greatly from discussions with Misha Pavel and Lee Stone.

Figure Captions

Figure 1: Distributed representations of velocity for rightward moving plaid stimuli. (a) and (b) Plaid stimuli made from pairs of gratings. Both plaids moved rightward with the same velocity. (c) and (d) Distributed representations corresponding to stimuli in (a) and (b), respectively. Each point corresponds to a different velocity (center corresponds to zero velocity). Brightness at each point is inversely proportional to $R(v_1, v_2)$ in equation (4). Locations of the peaks correspond to the correctly perceived velocities.

Figure 2: (a) Single moving grating. The diagonal line indicates the locus of velocities compatible with the motion of the grating. (b) Plaid composed of two moving gratings. The lines give the possible motion of each grating alone. Their intersection is the only shared motion.

Figure 3: Distributed representations of velocity for a vertical grating stimulus moving to the right. (a) Since the velocity is ambiguous, there is no peak in the distribution. (b) Responses are biased slightly by adding a small offset to the diagonal elements of **M**. This corresponds to a broad prior probability distribution centered at zero. Including the prior gives a broad peak in the distribution.

Figure 4: Orientation in space-time (based on an illustration by Adelson and Bergen, 1985). (a) A vertical bar translating to the right. (b) The space-time cube of stimulus intensities corresponding to motion of the vertical bar. (c) An x-t slice through the space-time cube. Orientation in the x-t slice is the horizontal component of velocity. Motion is like orientation in space-time, and spatiotemporally oriented filters can be used to detect and measure it.

Figure 5: Space-time slices through weighting functions of third derivative operators. (a) g_{xxx}, the third spatial derivative of a Gaussian, is monophasic and space-time separable. (b) g_{xxt} is biphasic and space-time separable. (c) $(g_{xxx} + g_{xxt})$ is tilted in space-time (not space-time separable), and selective for leftward motion.

Figure 6: Spatial slices (top row) and space-time slices (bottom row) through weighting functions of linear operators representative of the 46 operators used in the model. Some of these operators are Gaussian third derivatives, while others are constructed by summing third derivatives. The operators resemble physiological measurements of simple cell weighting functions.

Figure 7: (a) Spatial frequency tuning of a simple cell, measured with sine-grating stimuli drifting in the cell's preferred orientation (data replotted from Hamilton et al., 1989). Each curve is for a different stimulus temporal frequency, and each was shifted vertically for ease of viewing. Spatial frequency tuning is largely independent of temporal frequency. (b) Spatial frequency tuning for third derivative operator $(g_{xxx} + g_{xxt})$ is likewise independent of temporal frequency. (c) Spatial

frequency tuning for first derivative operator ($g_x + g_t$) is much broader and varies systematically with temporal frequency.

Figure 8: (a) Orientation/direction tuning of a simple cell, measured with sine-grating stimuli of preferred spatial and temporal frequency (data replotted from Movshon et al., 1986). Direction of motion is represented by the angular coordinate and relative response is plotted radially. (b) Orientation/direction tuning for third derivative operator ($g_{xxx} + g_{xxt}$) is similar. (c) Orientation/direction tuning for first derivative operator ($g_x + g_t$) is much broader.

Figure 9: (a) Direction tuning of a simple cell, measured with sine-grating plaid stimuli (data replotted from Movshon et al., 1986). Direction of motion of the plaid pattern is represented by the angular coordinate and relative response is plotted radially. (b) Plaid direction tuning for third derivative operator ($g_{xxx} + g_{xxt}$) is similar. (c) Plaid direction tuning for first derivative operator ($g_x + g_t$) is so broad that it does not respond independently to the two component gratings.

Figure 10: Response versus contrast as the spatial frequency, ω, of the stimulus is varied. (a) Data replotted from Albrecht and Hamilton (1982). (b) Model simulation. For both model cells and real cells, the contrast-response curve shifts mostly downward in the log-log plot if the spatial frequency of the test grating is non-optimal.

Figure 11: Bias of human velocity judgements for sine-grating plaids, as a function of contrast ratio of the two component gratings. (a) Data averaged from four subjects, replotted from Stone et al. (1990). Each curve is for a different total contrast. Relative temporal frequency was varied to compensate for the bias introduced by the relative contrast difference. Inferred bias plotted on the vertical axes is directly related to relative temporal frequency. Inferred bias is the direction that would be seen for that relative temporal frequency (according to the intersection of constraints rule) if both gratings had the same contrast. (b) Results from model simulations. The two parameters of the model were chosen to give the best (least-squares) fit to the data. For these parameter values the model behaves like human observers.

References

[1] E H Adelson and J R Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2:284–299, 1985.

[2] E H Adelson and J R Bergen. The extraction of spatio-temporal energy in human and machine vision. In *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, pages 151–156, Charleston, S Carolina, 1986.

[3] E H Adelson and J A Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300(5892):523–525, 1982.

[4] D G Albrecht and D B Hamilton. Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, 48:217–237, 1982.

[5] B W Andrews and D A Pollen. Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology (London)*, 287:163–176, 1979.

[6] A B Bonds. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2:41–55, 1989.

[7] F W Campbell, G F Cooper, and C Enroth-Cugell. The angular selectivity of visual cortical cells to moving gratings. *Journal of Physiology (London)*, 198:237–250, 1968.

[8] F W Campbell, G F Cooper, and C Enroth-Cugell. The spatial selectivity of visual cells of the cat. *Journal of Physiology (London)*, 203:223–235, 1969.

[9] Li Chao-yi and O Creutzfeldt. The representation of contrast and other stimulus parameters by single neurons in area 17 of the cat. *Pflugers Archives*, 401:304–314, 1984.

[10] M Fahle and T Poggio. Visual hyperacuity: spatiotemporal interpolation in human vision. *Proceedings of the Royal Society of London, B*, 213:451–477, 1981.

[11] V P Ferrara and H R Wilson. Perceived direction of moving two-dimensional patterns. *Vision Research*, 30:273–287, 1990.

[12] V P Ferrara and H R Wilson. Perceived speed of moving two-dimensional patterns. *Vision Research*, 31:877–893, 1991.

[13] K H Foster, J P Gaska, M Nagler, and D A Pollen. Spatial and temporal frequency selectivity of neurons in visual cortical areas V1 and V2 of the macaque monkey. *Journal of Physiology (London)*, 365:331–363, 1985.

[14] J J Gibson. *The Perception of the Visual World*. Houghton Mifflin, Boston, 1950.

[15] J J Gibson and E J Gibson. Continuous perspective transformations and the perception of rigid motions. *Journal of Experimental Psychology*, 54:129–138, 1957.

[16] N M Grzywacz and A L Yuille. A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London A*, 239:129–161, 1990.

[17] D B Hamilton, D G Albrecht, and W S Geisler. Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Research*, 29:1285–1308, 1989.

[18] D J Heeger. Model for the extraction of image flow. *Journal of the Optical Society of America A*, 4:1455–1471, 1987.

[19] D J Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1:279–302, 1988.

[20] D J Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:in press, 1992a.

[21] D J Heeger. Half-squaring in responses of cat simple cells. *Visual Neuroscience*, in press, 1992b.

[22] P Heggelund. Receptive-field organization of simple cells in cat striate cortex. *Experimental Brain Research*, 42:89–98, 1981.

[23] R A Holub and M Morton-Gibson. Response of visual cortical neurons of the cat to moving sinusoidal gratings: Response-contrast functions and spatiotemporal integration. *Journal of Neurophysiology*, 46:1244–1259, 1981.

[24] B K P Horn and B G Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[25] D Hubel and T Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106–154, 1962.

[26] H Ikeda and M J Wright. Spatial and temporal properties of 'sustained' and 'transient' neurones in area 17 of the cat's visual cortex. *Experimental Brain Research*, 22:363–383, 1975.

[27] G Johansson. Visual motion perception. *Scientific American*, 232:76–88, 1975.

[28] J P Jones and L A Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1187–1211, 1987.

[29] J J Kulikowski and P O Bishop. Linear analysis of the response of simple cells in the cat visual cortex. *Experimental Brain Research*, 44:386–400, 1981a.

[30] B D Lucas and T Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.

[31] J McLean and L A Palmer. Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research*, 29:675–679, 1989.

[32] J A Movshon, E H Adelson, M S Gizzi, and W T Newsome. The analysis of moving visual patterns. In C Chagas, R Gattass, and C Gross, editors, *Experimental Brain Research Supplementum II: Pattern Recognition Mechanisms*, pages 117–151. Springer-Verlag, New York, 1986.

[33] J A Movshon, I D Thompson, and D J Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology (London)*, 283:53–77, 1978.

[34] H H Nagel. On the estimation of optical flow: relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[35] K Nakayama. Biological image motion processing: A review. *Vision Research*, 25:625–660, 1985.

[36] J G Robson. Linear and nonlinear operations in the visual system. *Investigative Opthalmology and Visual Science Supplement*, 29:117, 1988.

[37] G Sclar, J H R Maunsell, and P Lennie. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, 30:1–10, 1990.

[38] R Shapley, R C Reid, and R Soodak. Spatiotemporal receptive fields and direction selectivity. In M Landy and J A Movshon, editors, *Computational Models of Visual Processing*, pages 109–118. MIT Press, Cambridge, MA, 1991.

[39] E P Simoncelli and E H Adelson. Computation of optical flow: Relationship between several standard techniques. Technical Report 165, Vision and Modeling Group, MIT Media Lab, 1991a.

[40] E P Simoncelli and E H Adelson. Relationship between gradient, spatio-temporal energy, and regression models for motion perception. *Investigative Opthalmology and Visual Science Supplement*, 32:893, 1991.

[41] E P Simoncelli, E H Adelson, and D J Heeger. Probability distributions of optical flow. In *Proceedings of Computer Vision and Pattern Recognition*, pages 310–315, Mauii, HI, June 1991.

[42] E P Simoncelli and D J Heeger. A computational model for perception of two-dimensional pattern velocities. *Investigative Opthalmology and Visual Science Supplement*, 33, in press, 1992.

[43] L S Stone, A B Watson, and J B Mulligan. Effect of contrast on the perceived direction of a moving plaid. *Vision Research*, 30:1049–1067, 1990.

[44] D J Tolhurst and J A Movshon. Spatial and temporal contrast sensitivity of striate cortical neurons. *Nature*, 257:674–675, 1975.

[45] J P H van Santen and G Sperling. Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, 2:300–321, 1985.

[46] H Wallach and D N O'Connell. The kinetic depth effect. *Journal of Experimental Psychology*, 45:205–217, 1953.

[47] W H Warren and D J Hannon. Direction of self-motion is perceived from optical flow. *Nature*, 336:162–163, 1988.

[48] W H Warren and D J Hannon. Eye movements and optical flow. *Journal of the Optical Society of America A*, 7:160–169, 1990.

[49] A B Watson and A J Ahumada. A look at motion in the frequency domain. In J K Tsotsos, editor, *Motion: Perception and representation*, pages 1–10. Association for Computing Machinery, New York, 1983.

[50] A B Watson and A J Ahumada. Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2:322–342, 1985.

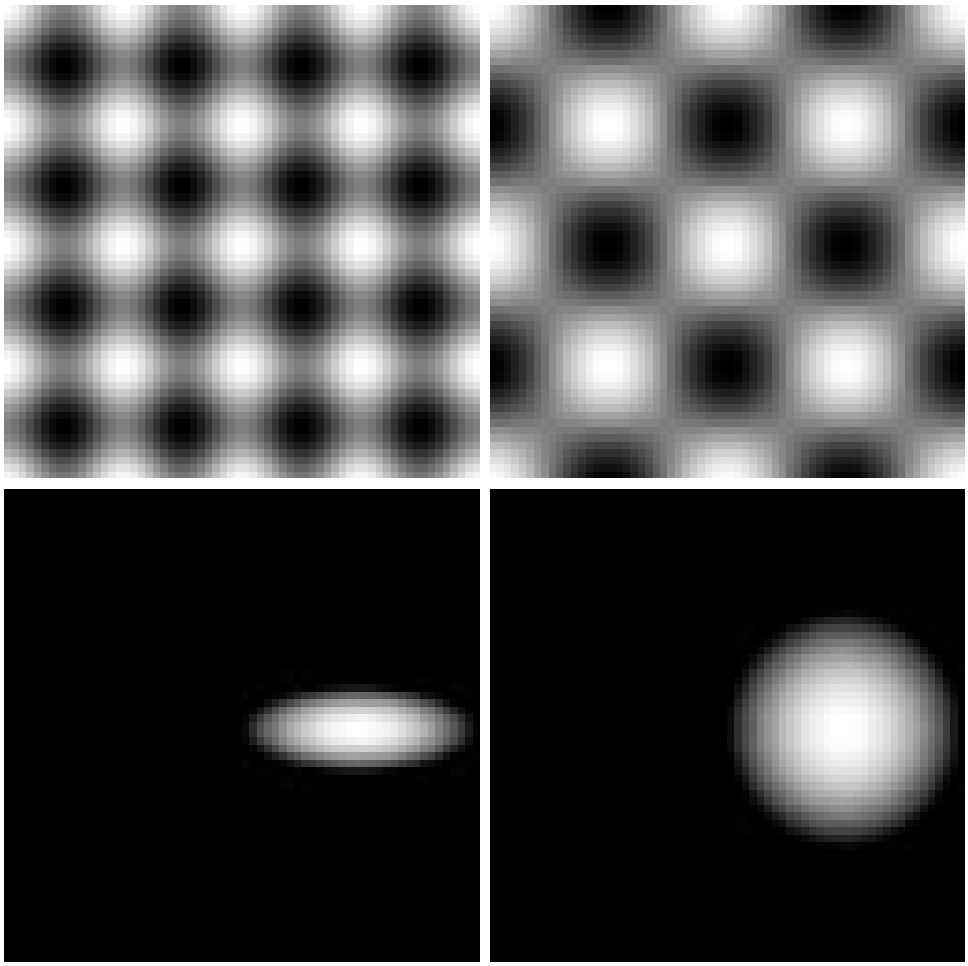

Figure 1: Distributed representations of velocity for rightward moving plaid stimuli. (a) and (b) Plaid stimuli made from pairs of gratings. Both plaids moved rightward with the same velocity. (c) and (d) Distributed representations corresponding to stimuli in (a) and (b), respectively. Each point corresponds to a different velocity (center corresponds to zero velocity). Brightness at each point is inversely proportional to $R(v_1, v_2)$ in equation (4). Locations of the peaks correspond to the correctly perceived velocities.

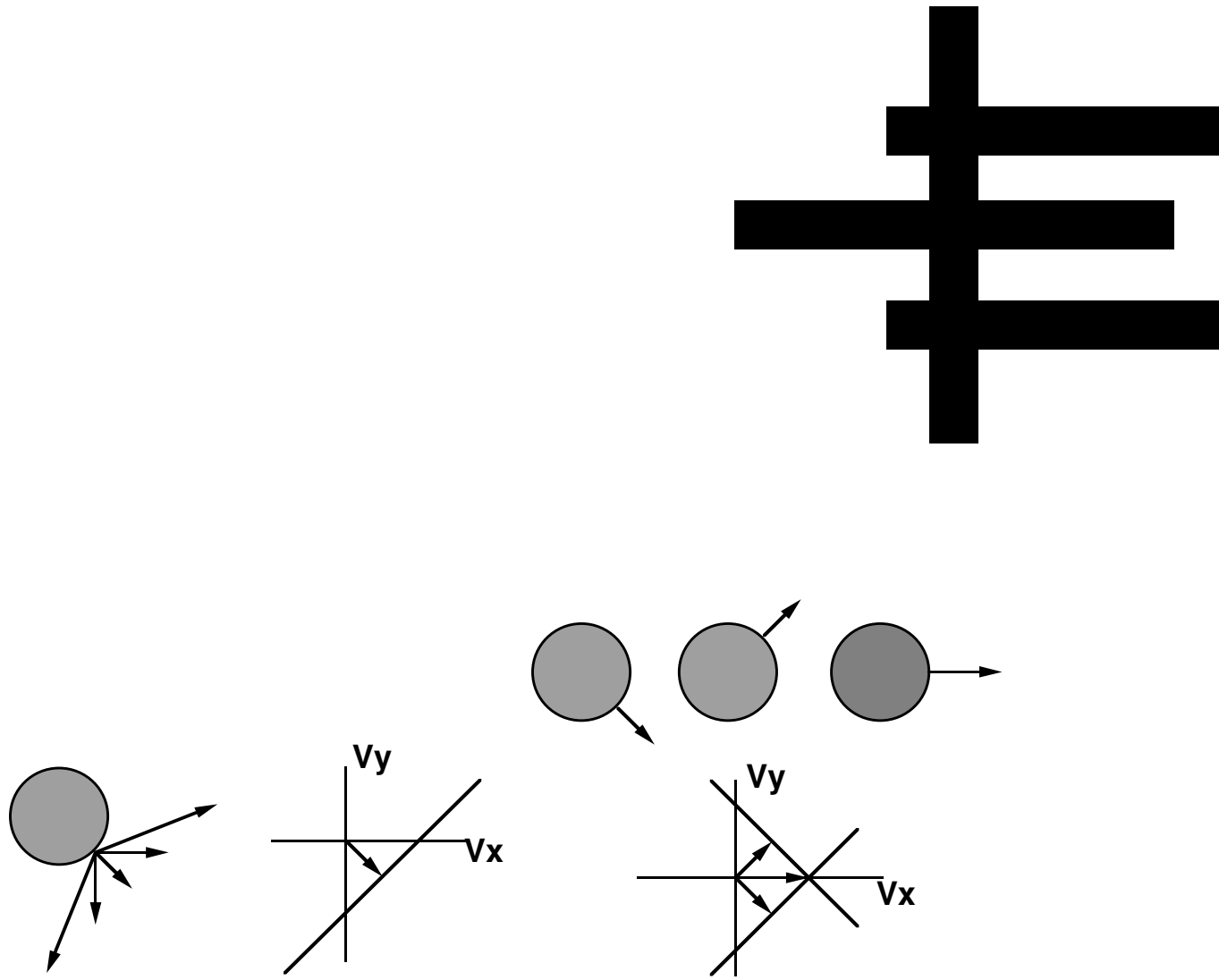

Figure 2: (a) Single moving grating. The diagonal line indicates the locus of velocities compatible with the motion of the grating. (b) Plaid composed of two moving gratings. The lines give the possible motion of each grating alone. Their intersection is the only shared motion.

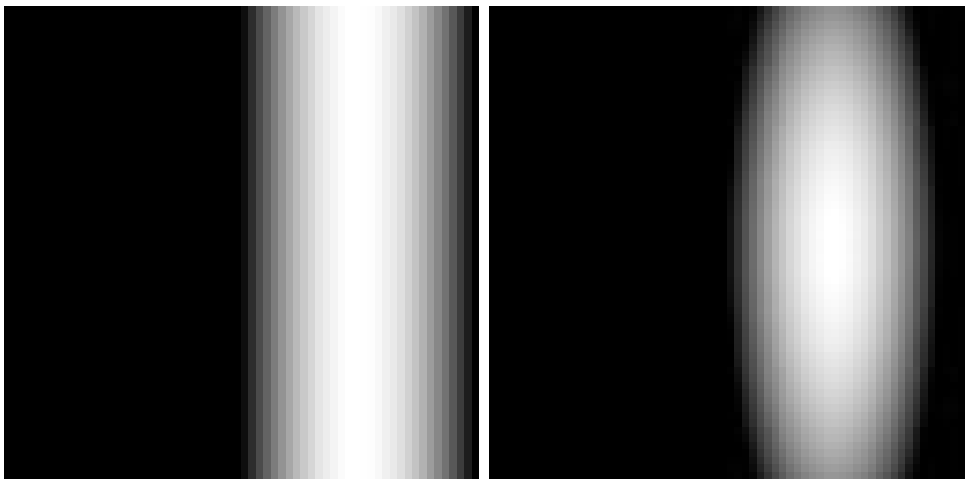

Figure 3: Distributed representations of velocity for a vertical grating stimulus moving to the right. (a) Since the velocity is ambiguous, there is no peak in the distribution. (b) Responses are biased slightly by adding a small offset to the diagonal elements of $\mathbf{M}$. This corresponds to a broad prior probability distribution centered at zero. Including the prior gives a broad peak in the distribution.

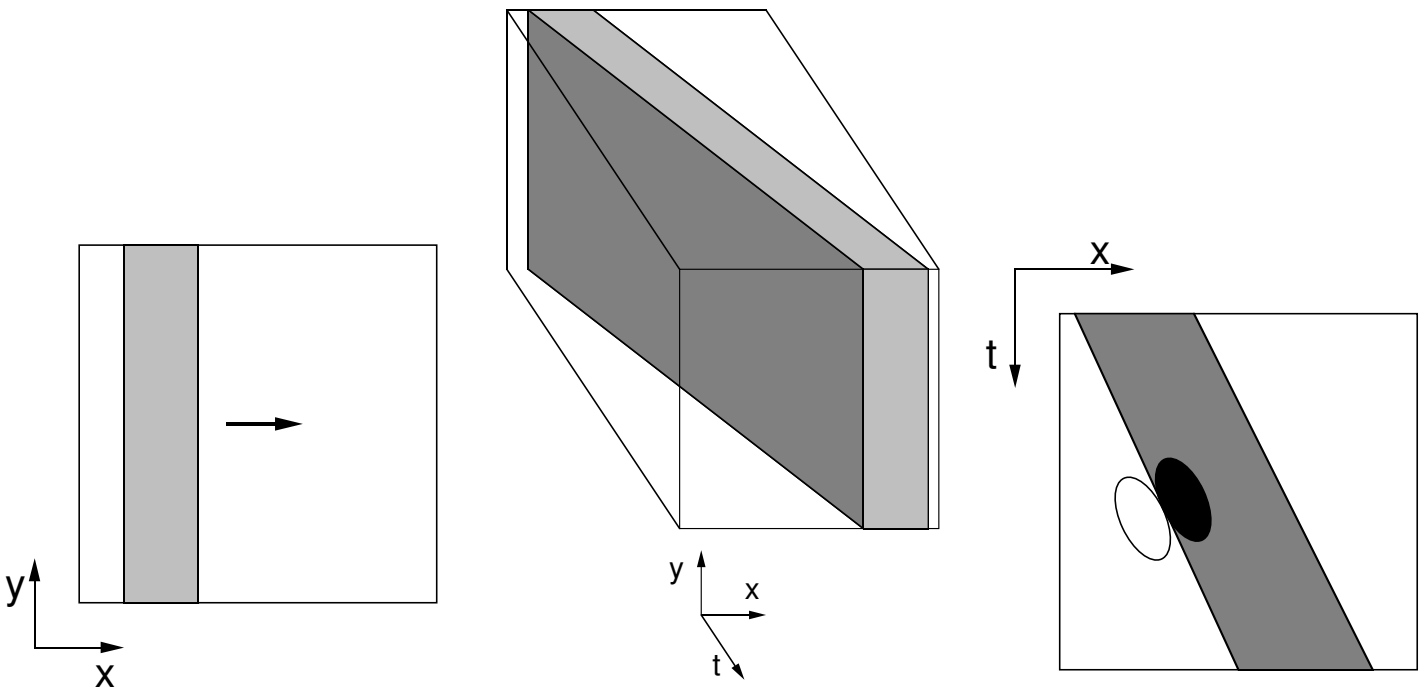

Figure 4: Orientation in space-time (based on an illustration by Adelson and Bergen, 1985). (a) A vertical bar translating to the right. (b) The space-time cube of stimulus intensities corresponding to motion of the vertical bar. (c) An x-t slice through the space-time cube. Orientation in the x-t slice is the horizontal component of velocity. Motion is like orientation in space-time, and spatiotemporally oriented filters can be used to detect and measure it.

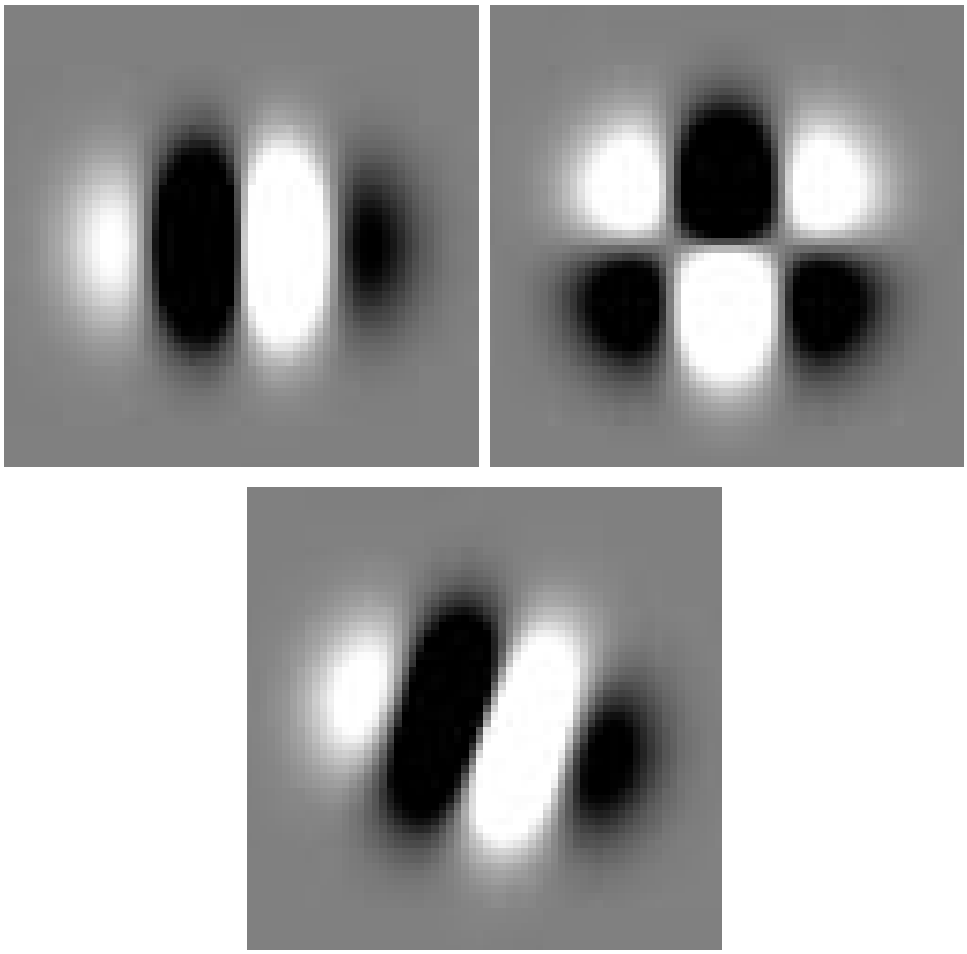

Figure 5: Space-time slices through weighting functions of third derivative operators. (a) g_{xxx}, the third spatial derivative of a Gaussian, is monophasic and space-time separable. (b) g_{xxt} is biphasic and space-time separable. (c) $(g_{xxx} + g_{xxt})$ is tilted in space-time (not space-time separable), and selective for leftward motion.

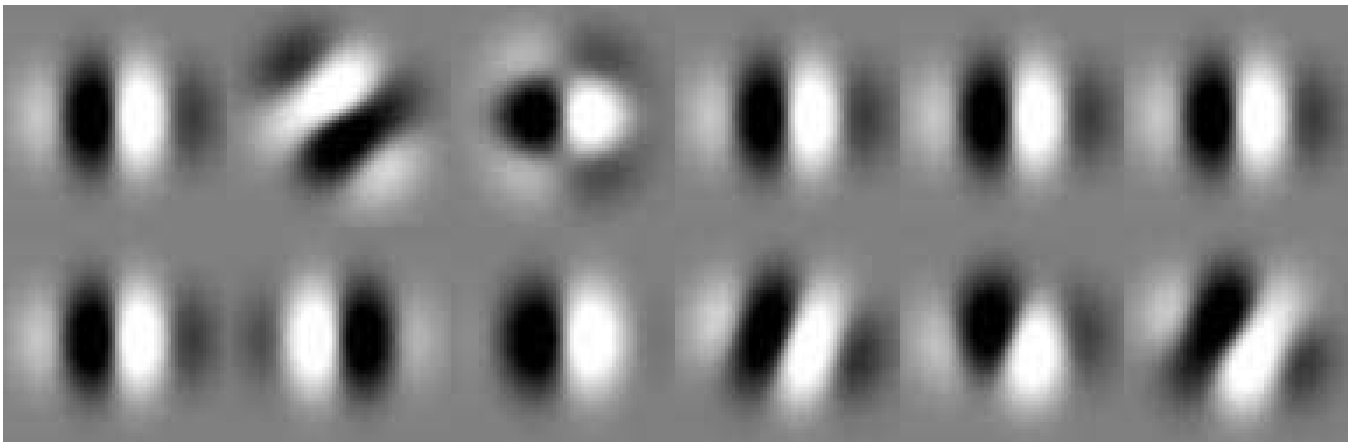

Figure 6: Spatial slices (top row) and space-time slices (bottom row) through weighting functions of linear operators representative of the 46 operators used in the model. Some of these operators are Gaussian third derivatives, while others are constructed by summing third derivatives. The operators resemble physiological measurements of simple cell weighting functions.

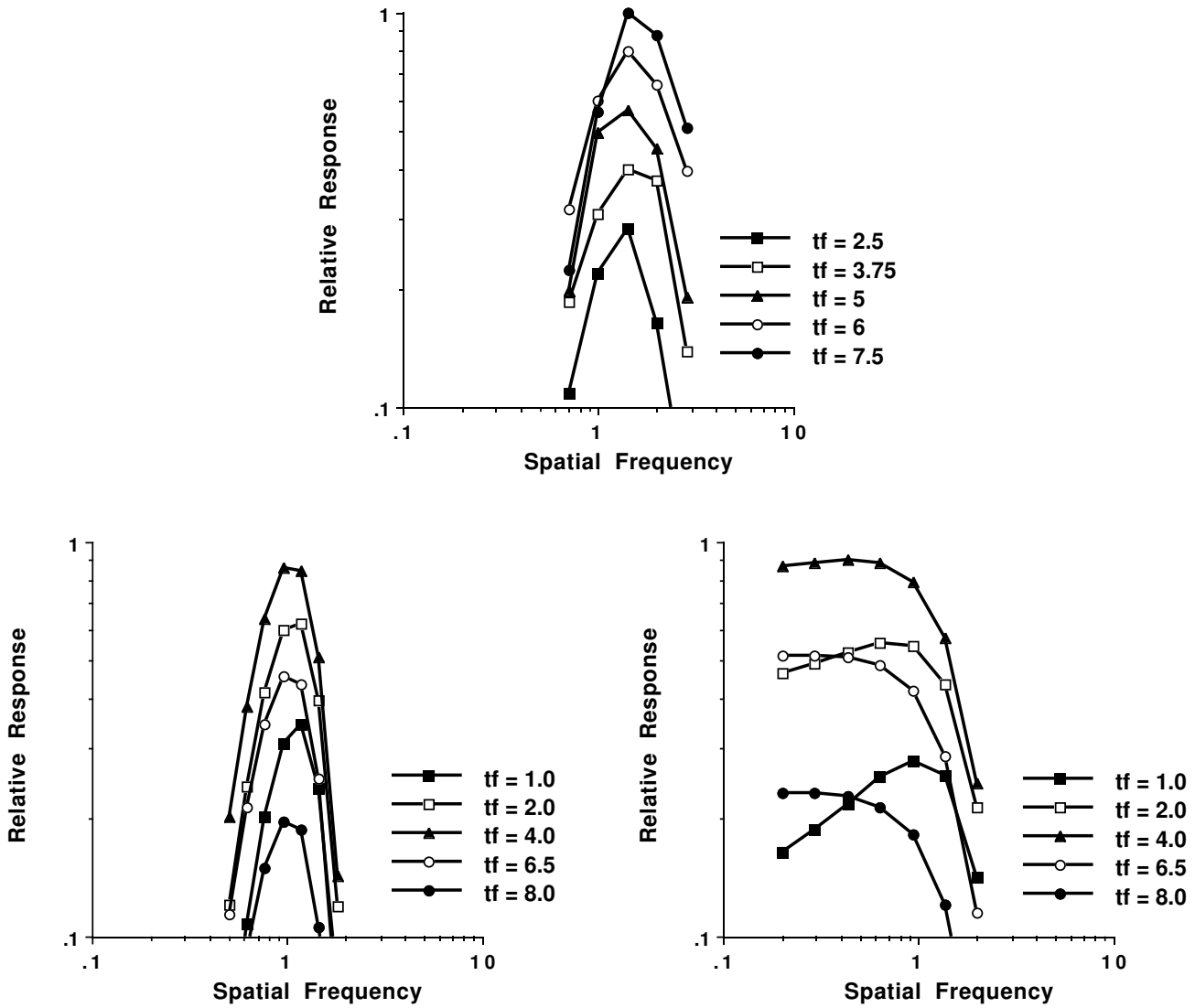

Figure 7: (a) Spatial frequency tuning of a simple cell, measured with sine-grating stimuli drifting in the cell's preferred orientation (data replotted from Hamilton et al., 1989). Each curve is for a different stimulus temporal frequency, and each was shifted vertically for ease of viewing. Spatial frequency tuning is largely independent of temporal frequency. (b) Spatial frequency tuning for third derivative operator ($g_{xxx} + g_{xxt}$) is likewise independent of temporal frequency. (c) Spatial frequency tuning for first derivative operator ($g_x + g_t$) is much broader and varies systematically with temporal frequency.

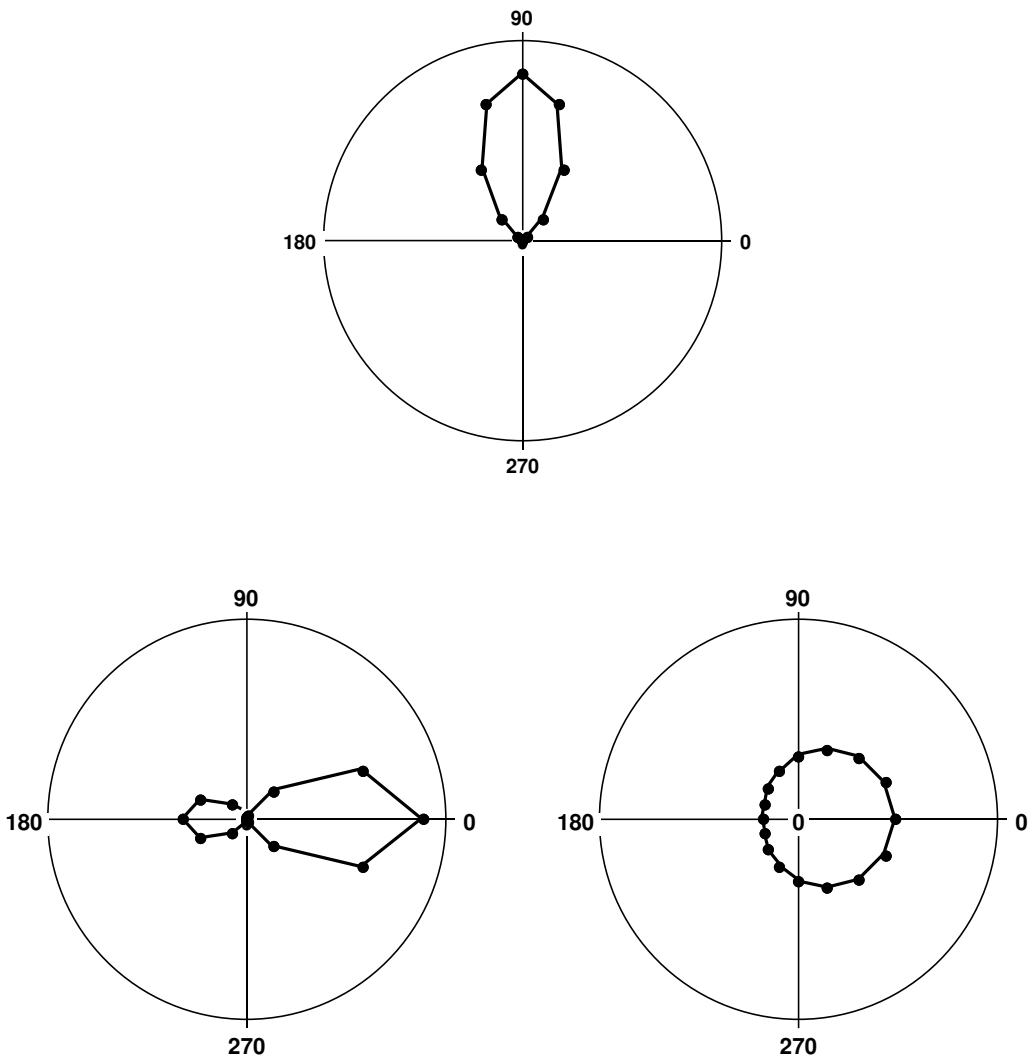

Figure 8: (a) Orientation/direction tuning of a simple cell, measured with sine-grating stimuli of preferred spatial and temporal frequency (data replotted from Movshon et al., 1986). Direction of motion is represented by the angular coordinate and relative response is plotted radially. (b) Orientation/direction tuning for third derivative operator ($g_{xxx} + g_{xxt}$) is similar. (c) Orientation/direction tuning for first derivative operator ($g_x + g_t$) is much broader.

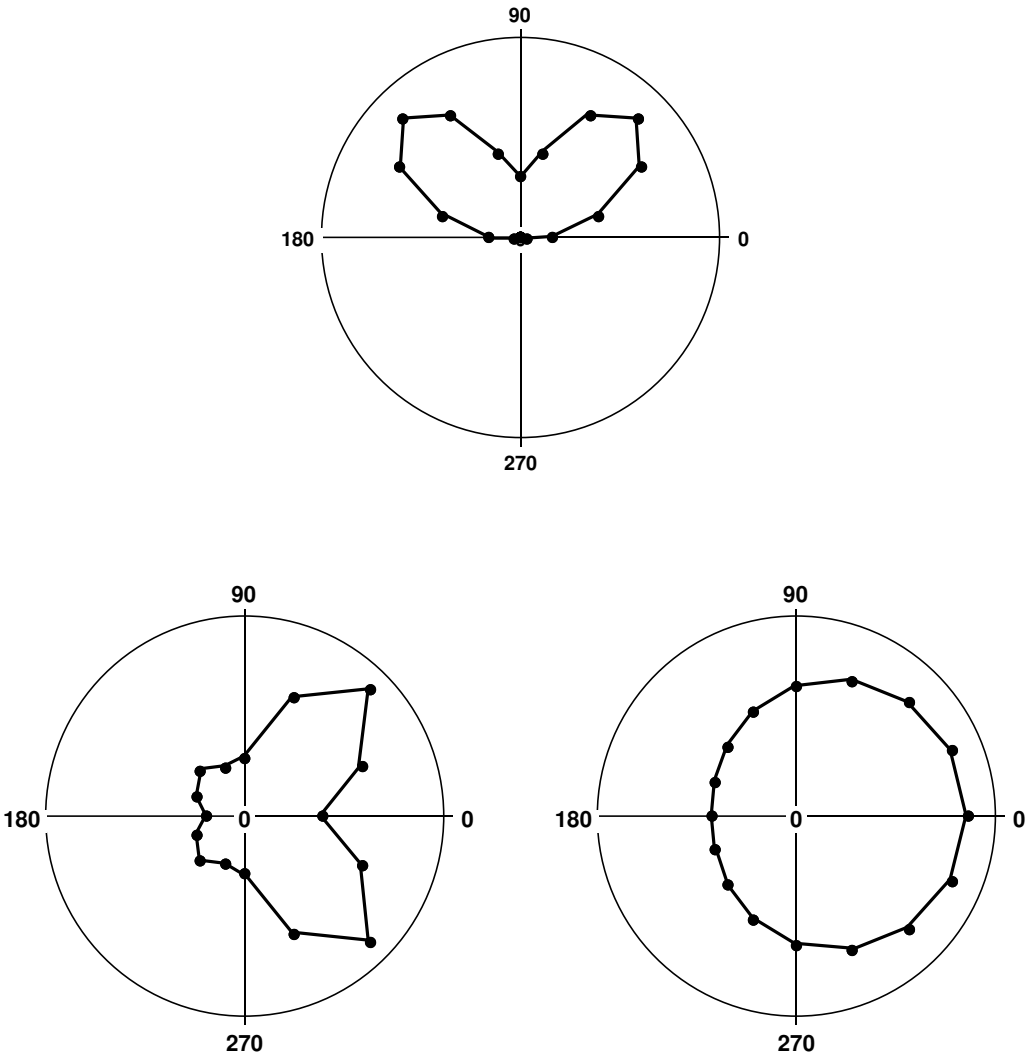

Figure 9: (a) Direction tuning of a simple cell, measured with sine-grating plaid stimuli (data replotted from Movshon et al., 1986). Direction of motion of the plaid pattern is represented by the angular coordinate and relative response is plotted radially. (b) Plaid direction tuning for third derivative operator ($g_{xxx} + g_{xxt}$) is similar. (c) Plaid direction tuning for first derivative operator ($g_x + g_t$) is so broad that it does not respond independently to the two component gratings.

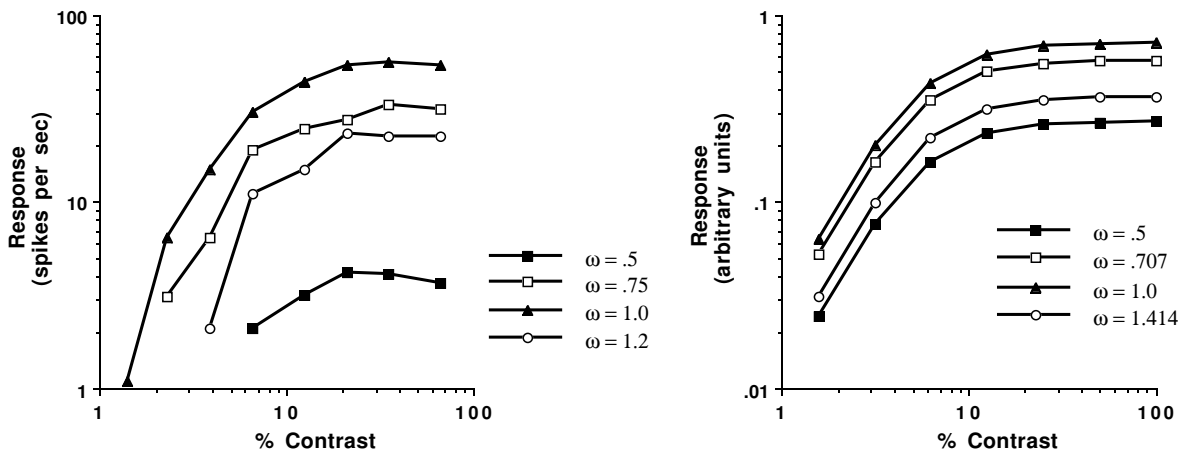

Figure 10: Response versus contrast as the spatial frequency, ω, of the stimulus is varied. (a) Data replotted from Albrecht and Hamilton (1982). (b) Model simulation. For both model cells and real cells, the contrast-response curve shifts mostly downward in the log-log plot if the spatial frequency of the test grating is non-optimal.

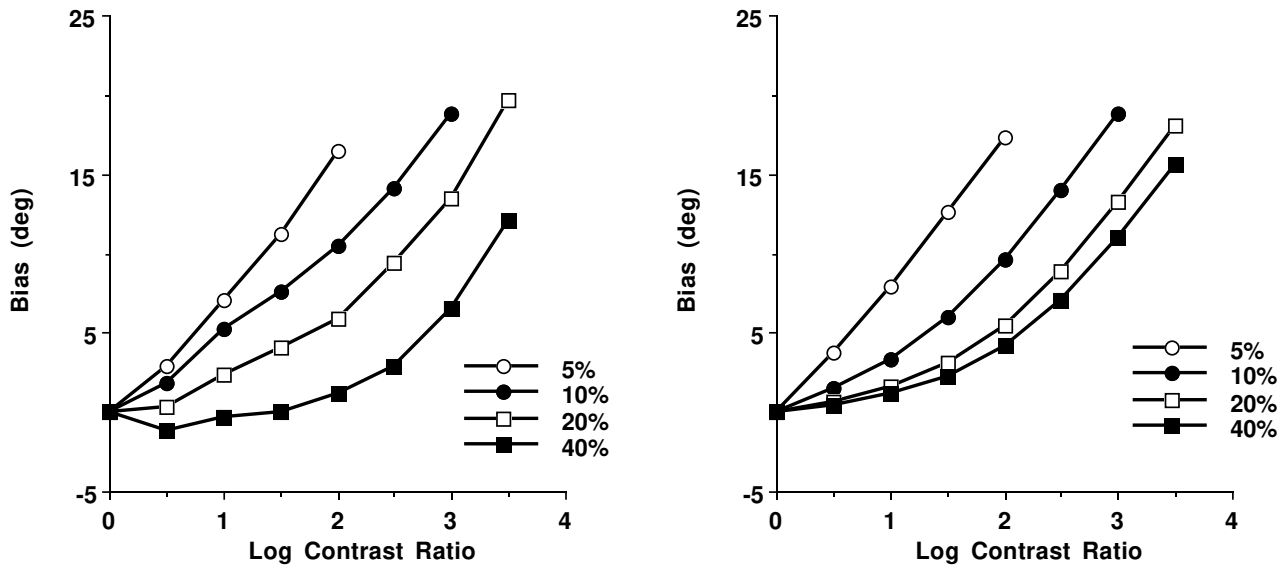

Figure 11: Bias of human velocity judgements for sine-grating plaids, as a function of contrast ratio of the two component gratings. (a) Data averaged from four subjects, replotted from Stone et al. (1990). Each curve is for a different total contrast. Relative temporal frequency was varied to compensate for the bias introduced by the relative contrast difference. Inferred bias plotted on the vertical axes is directly related to relative temporal frequency. Inferred bias is the direction that would be seen for that relative temporal frequency (according to the intersection of constraints rule) if both gratings had the same contrast. (b) Results from model simulations. The two parameters of the model were chosen to give the best (least-squares) fit to the data. For these parameter values the model behaves like human observers.