

Transparent Motion Perception as Detection of Unbalanced Motion Signals. III. Modeling

Ning Qian,^a Richard A. Andersen, and Edward H. Adelson

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

In the preceding two companion articles we studied the conditions under which transparent motion perception occurs through psychophysical experiments, and investigated the underlying neural mechanisms through physiological recordings. The main finding of our perceptual experiments was that whenever a display has finely balanced motion signals in all local areas, it is perceptually nontransparent, and that transparent displays always contain motion signals in different directions that are either spatially unbalanced, or unbalanced in their disparity or spatial frequency contents. In the physiological experiments, we found two stages in the processing of transparent stimuli. The first stage is located primarily in area V1. At this stage motion measurements are made and V1 cells respond well to both the balanced, nontransparent stimuli and the unbalanced, perceptually transparent stimuli. The second stage is located primarily in area MT. MT cells show strong suppression between opposite directions of motion. The suppression for the unbalanced, transparent stimuli is significantly less than that for the balanced, nontransparent stimuli. Therefore, the activity in the second, MT stage correlates better with the perception of motion transparency than the first, V1 stage, which does not distinguish reliably between transparent and nontransparent motion.

The above experiments suggest a two-stage model of motion perception with a motion measurement stage in V1 and an opponent-direction suppression stage in area MT. In this article we explicitly test this model through analysis and computer simulations, and compare the response of the model to the perceptual and physiological results using the same balanced and unbalanced stimuli we used in the experiments. In the first stage of the computational model, motion energies in different spatial frequency and disparity ranges are extracted from each local region. Similar to V1, this stage does not distinguish between the balanced and unbalanced stimuli. In the subsequent stage motion energies of opposite directions but with same spatial frequency and disparity contents suppress each

other using subtractive or divisive inhibition. This stage responds significantly better to the transparent stimuli than to the nontransparent ones, in agreement with MT activity.

[Key words: motion transparency, motion energy models, stereopsis, motion-stereo integration, spatial frequency, computer modeling]

Motion is a rich source of various types of useful information. For example, it allows us to determine three-dimensional structures of moving objects, and to segment a complex scene into its meaningful parts (see Nakayama, 1985, for a review). The task of motion detection is relatively easy when there is only a single point-like object moving on a blank background. Our visual system, however, has to handle (and is able to handle) much more complicated situations. For instance, when an object with differently oriented boundaries is moving in a certain direction, it generates local motion vectors that are perpendicular to the boundaries (the "aperture problem"). These vectors may not be pointing in the true direction of motion. Even more complex is the situation where there are partial occlusions and translucent surfaces in a scene with moving objects. In these cases the visual system has to represent more than one motion in the same part of space—the problem of transparent motion perception. A laboratory demonstration of motion transparency uses two independent sets of random dots moving in opposite directions in the same part of a video monitor. Two transparent surfaces, one defined by each set of dots, are seen as continuously and independently moving across each other.

In the first of this set of three studies, we performed psychophysical experiments for determining the conditions under which transparent motion perception occurs (see preceding companion article, Qian et al., 1994). We found that displays with locally well-balanced motion signals in opposite directions are perceptually nontransparent. The transparent displays, on the other hand, contain locally unbalanced motion signals in different directions. Furthermore, if the two components of the spatially balanced displays are at different depths, or contain very different spatial frequency contents, the displays appear transparent. These displays contain motion signals that are unbalanced in binocular disparity or spatial frequency. Based on these results, we proposed that local suppression among different directions of motion within each disparity and spatial frequency channel could be the mechanism for distinguishing transparent displays from nontransparent ones. Nontransparent displays presumably maximize the suppression in all frequency and disparity channels and therefore evoke relatively weak responses.

We also investigated motion transparency physiologically

Received Aug. 18, 1993; revised Apr. 18, 1994; accepted Apr. 27, 1994.

We are grateful to Eero Simoncelli and John Wang for their help with OBVIOUS program. We would also like to thank Bard Geesaman and two anonymous reviewers for their comments on earlier versions of the manuscript. The research is supported by Office of Naval Research Contract N00014-89-J1236 and NIH Grant EY07492, both to R.A.A. N.Q. was supported by a McDonnell-Pew postdoctoral fellowship during the early phase of this work.

Correspondence should be addressed to Dr. Richard A. Andersen, Division of Biology, 217-76, California Institute of Technology, Pasadena, CA 91125.

^aPresent address: Center for Neurobiology and Behavior, Columbia University, New York, NY 10032.

Copyright © 1994 Society for Neuroscience 0270-6474/94/147381-12\$05.00/0

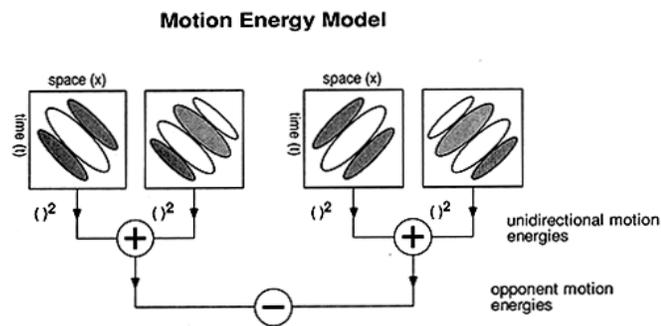


Figure 1. The motion energy model used in most of our simulations for explaining the perceptual difference of transparent and nontransparent displays. Motion sensitivity is generated by the spatiotemporally oriented filters shown schematically in the square boxes. There are two stages in the model. The first stage computes unidirectional motion energies by squaring and then summing the outputs of a quadrature pair of oriented filters. In the second stage, motion energies from opposite directions suppresses each other. Subtractive inhibition is indicated in the figure, but we have also considered divisive inhibition (see text).

through single-unit recordings from V1 and MT (see preceding companion article, Qian and Andersen, 1994). We found that the directionally selective V1 cells responded quite well to both transparent and the nontransparent stimuli. On the average, they could not reliably distinguish the two types of displays. On the other hand, MT cells' responses to displays with two components moving in opposite directions are strongly suppressed in comparison to their preferred direction responses. More importantly, the suppression is significantly higher (or the response weaker) for the nontransparent stimuli than for the perceptually transparent ones.

Our physiological experiments indicate a two-stage model of motion perception with a motion measurement stage in V1 and an opponent-direction suppression stage in area MT. Our psychophysical experiments further suggest that the suppression in the second stage should be spatial frequency and disparity specific. In this article, we present analysis and computer simulations using such a two-stage model in order to demonstrate more quantitatively that a disparity- and spatial frequency-specific suppressive mechanism can indeed account for the difference in the perceptual transparency of our displays.

We use motion energy models (Adelson and Bergen, 1985; Watson and Ahumada, 1985) and their extension to disparity sensitivity (Qian, 1994) for motion measurements in the first stage. Many models for biological motion processing have been proposed. We choose motion energy models because of their biological plausibility. Although certain versions of this class of models have been shown to be equivalent to the Reichardt motion detectors (Reichardt, 1961; Adelson and Bergen, 1985; van Santen and Sperling, 1985), the unidirectional motion energy stage of the model is not equivalent to any stage in the Reichardt detector (Emerson et al., 1992). There is physiological evidence suggesting that directionally selective cells in the primary visual cortex behave as if they compute unidirectional motion energies (Reid et al., 1987; McLean and Palmer, 1989; Snowden et al., 1991; Emerson et al., 1992). For the second, suppression stage of the model, we consider both subtractive (Adelson and Bergen, 1985) and divisive (Snowden et al., 1991; Heeger, 1992) types of inhibition among different directions of motion.

We first apply the model to those displays without disparity cues. The spatial frequency effect can be explained by the energy models since energy detectors naturally contain frequency selectivity. We then consider the effect of binocular disparity. Since standard energy models do not contain disparity tuning, we have recently developed a model for biological stereopsis and combined it with motion energy models into a common framework (Qian, 1994). We show here through computer simulations that the extended model can account for the contribution of disparity to motion transparency.

Preliminary versions of the results presented here have appeared previously (Qian et al., 1991, 1992).

Analysis and Simulations

We used the version of the energy models proposed by Adelson and Bergen (1985) in our analysis and simulations. An essential idea behind the model is that motion of an object through space over time can be described by an orientation in the spatiotemporal space (Fatale and Poggio, 1981; Adelson and Bergen, 1985). The model uses spatiotemporally oriented linear filters to detect motion. The linear mechanism for generating motion sensitivity is supported by recent intracellular studies of directionally selective cells in cat visual cortex (Jagadeesh et al., 1993). The outputs of two linear filters with 90° phase difference are squared and then summed to form the phase-independent unidirectional motion energy detector. Such a detector simply measures the Fourier power within a certain spatiotemporal frequency window specified by its parameters.

We will examine whether our psychophysical and physiological results can be explained by assuming a suppressive stage in the motion pathway at which motion signals in different directions from each small region inhibit each other. Different suppression mechanisms have been proposed in the past. They include subtractive opponency (Adelson and Bergen, 1985) and divisive normalization (Adelson and Bergen, 1986; Snowden et al., 1991; Heeger, 1992). While there are important differences between these two mechanisms (see Divisive inhibition, below), our psychophysical experiments are not designed to differentiate between them. In fact, both mechanisms can explain our results well. A schematic drawing of the motion energy model with subtractive inhibition is shown in Figure 1.

We now give an explicit description of our model for motion transparency. We assume that for each spatial location there is a population of motion energy detectors tuned to different ranges of spatiotemporal frequency (and thus to different directions and speeds of motion) and disparity. To the first approximation, this stage can be identified with V1 cells. At the second (suppressive) stage either the opponent energy (with subtractive inhibition) or the normalized energy (with divisive inhibition) is computed within each spatial frequency and disparity channel from the initial motion energy measurements. We propose that this second stage is performed by subunits in MT cells' receptive fields. We hypothesize that if a visual display generates large energies for both left and right directions at the suppression stage and if these energies in different directions are spatially mixed, the pattern is perceptually transparent. Note that for patterns with a richer spectrum of frequencies, more unidirectional energy detectors will be activated. Other things being equal, these patterns will activate more opponent or normalized energy detectors and will therefore be more likely to appear transparent. This is consistent with our psychophysical observation that random dot patterns look transparent over a

wider range of parameters than randomly spaced parallel line patterns and that counterphase gratings and square wave gratings look nontransparent over a wider range of parameters than equally spaced line patterns (Qian et al., 1994).

In the above description, we assume that motion suppression occurs at the subunit level of MT cells. For completeness we further propose that all subunits of an MT cell have similar tuning properties and that the overall response of an MT cell is equal to the summation of the thresholded output of all its subunits. This step could be viewed as a part of the spatial integration process that may be responsible for the coherent percept of transparent motion (see Discussion), and will be explored in our future publications. For our present purpose of explaining the perceptual differences of our psychophysical stimuli, the activity at the MT subunit level is sufficient.

For simplicity, we first consider the case of one spatial dimension and no disparity, and use subtractive inhibition for suppression. We use Gabor filters for spatiotemporal orientation detection. These filters correspond to V1 simple cell. A quadrature pair of such filters with even and odd phases tuned to leftward (-) and rightward (+) directions of motion are given by

$$g_e^\pm = \frac{1}{2\pi\sigma_x\sigma_t} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}\right) \cos(\omega_x x \pm \omega_t t) \quad (1)$$

$$g_o^\pm = \frac{1}{2\pi\sigma_x\sigma_t} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}\right) \sin(\omega_x x \pm \omega_t t) \quad (2)$$

where σ_x and σ_t determine the widths of the filters in the spatial and temporal domain, respectively, and ω_x and ω_t are the central angular spatial and temporal frequencies. The ratio of the latter two parameters determines the orientation of the filters in the spatiotemporal space. All these variables are assumed to be positive. Also note that the areas under the Gaussian envelopes of the filters are normalized. This is important for comparing results from filters of different scales. The responses of these filters to visual stimuli are given by the convolution operation.

The phase-insensitive leftward and rightward motion energies for a spatiotemporal pattern $f(x,t)$ are defined as

$$E^+(x,t) = [f * g_e^+]^2 + [f * g_o^+]^2, \quad (3)$$

$$E^-(x,t) = [f * g_e^-]^2 + [f * g_o^-]^2, \quad (4)$$

where * denotes convolution. Responses at this stage correspond to directionally selective complex cells. The opponent motion energy is defined as the difference of the two:

$$E(x,t) = E^+(x,t) - E^-(x,t) \quad (5)$$

Under these definitions, positive (negative) opponent energy indicates the rightward (leftward) motion and a value around zero means that no motion is detected. Of course, neurons can only fire positively. In reality, the leftward and rightward opponent motion energies have to be carried by two different populations of MT cells with subunits having threshold nonlinearity. It is obvious that the introduction of opponency will cause substantial cancellation of motion energies from opposite directions. What we are interested in here, however, is whether the residual responses after the cancellation would indeed be quite different for our transparent and nontransparent patterns. We now make some explicit calculations and computer simulations for some of these patterns.

Counterphase gratings

We start with counterphase gratings, which are composed of two identical sine wave gratings moving across each other in opposite directions. Counterphase gratings are perceptually nontransparent. They can be represented mathematically as

$$f_{cp}(x,t) = \sin(\Omega_x x - \Omega_t) + \sin(\Omega_x x + \Omega_t) \quad (6)$$

where Ω_x and $\pm\Omega_t$ are the spatial and temporal frequencies. Assume that both Ω_x and Ω_t are positive, then the two terms on the righthand side of Equation 6 represents two identical sine wave gratings moving to the right and left, respectively. As is shown in the Appendix, the opponent energy for the counterphase grating is exactly

$$E_{cp}(x,t) = 0, \quad (7)$$

for all (x,t) (i.e., for filters located at any position at any time), independent of the parameters for the Gabor filters and the grating. This result explains the lack of transparent motion perception for counterphase gratings. Note that while it is intuitively obvious that opponent energy should be small for the counterphase gratings, the exact null result in Equation 7 is a consequence of using quadrature Gabor filters and subtractive inhibition.

Spatial frequency specificity

While two identical sine wave gratings moving across each other look like flicker, two sine gratings with very different spatial frequencies are perceptually transparent (Qian et al., 1994). This suggests that the suppression between different directions of motion is limited within each spatial frequency channel. The motion detectors in energy models already have frequency selectivity built into it. From the Fourier transformation of the Gabor filters in Equations 1 and 2, it can be shown that the frequency responses of these filters are centered around $(\pm\omega_x, \pm\omega_t)$, with the bandwidth (defined at halfpeak amplitude) along each dimension equal to

$$bw = \log_2 \left(\frac{\omega\sigma + \sqrt{2\ln 2}}{\omega\sigma - \sqrt{2\ln 2}} \right) \quad (8)$$

where ω and σ represent the central angular frequency and Gaussian width of the given dimension and \ln stands for natural logarithm. To model frequency specificity of directional suppression, we therefore apply opponency only to the two unidirectional motion energies computed with filters of identical frequency selectivity but tuned to opposite directions of motion.

We performed computer simulations on a display composed of two different sine wave gratings with spatial frequencies equal to 1.5 cycles/degree and 6 cycles/degree, respectively. We also considered a counterphase grating with spatial frequency equal to 3 cycles/degree for comparison. We used three sets of filters with spatial frequencies centered around 1.5, 3, and 6 cycles/degree. They represent three spatial frequency channels. The Gaussian widths (σ) of the filters were 1/3, 1/6, and 1/12 of a degree, respectively (the actual widths of the filters, defined at the half-amplitude of the Gaussian envelope, are 0.78°, 0.39°, and 0.20°, respectively). Under these parameters, the bandwidths of all filters are 1.14 octaves according to Equation 8. The results of our simulations are shown in Figure 2. The spatiotemporal representations of the display composed of two different sine wave gratings and that of the counterphase grating are shown in Figure 2, *a* and *b*. Figures 2*c-h* represents the opponent energies from

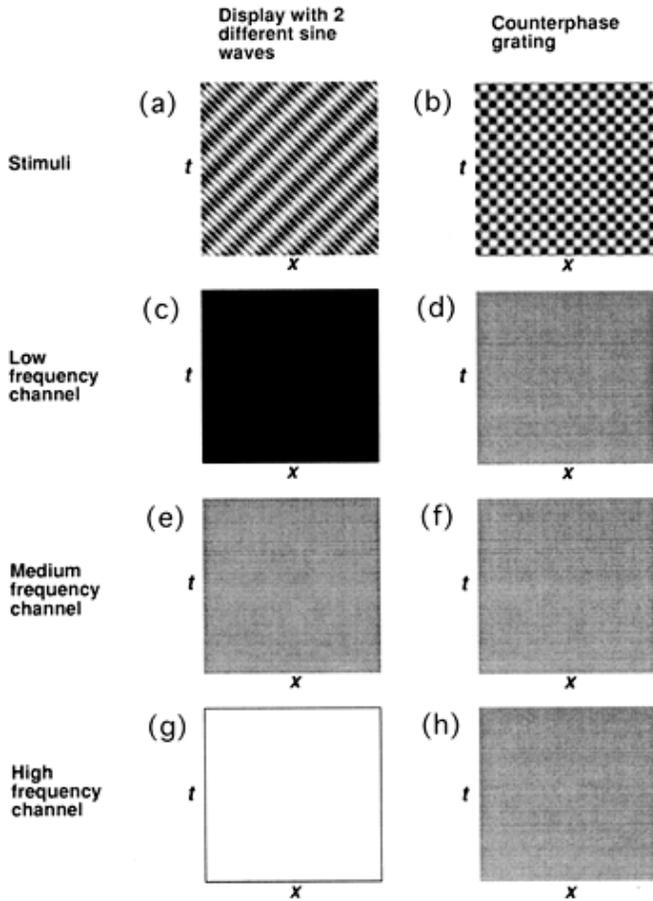


Figure 2. Simulation for a counterphase grating and two sine wave gratings with different spatial frequencies. All figures are shown in spatiotemporal space. *a* and *b* are the spatiotemporal representations of the two types of patterns. *c-h* show the opponent energies in three different frequency channels for the two patterns, all shown with the same gray scale. Gray indicates little motion energy and white and black code for rightward and leftward opponent energies, respectively. See text for the details of the parameters used.

the three frequency channels for the two types of displays. In this figure gray indicates little motion energy, and white and black code for rightward and leftward opponent energies, respectively. It is clear from Figure 2*c-h* that while the counterphase grating gives no opponent motion energy in any of the three channels, the display composed of two different sine wave gratings generates rightward opponent motion energy in the high-frequency channel and leftward energy in the low-frequency channel. These results correlate well with the observation that the former is not perceptually transparent while the latter is. If the inhibition were between the low- and high-frequency channels, the motion energies in opposite directions would strongly cancel each other, and the pattern would appear non-transparent, just like counterphase gratings.

Line patterns

We next turn to the line patterns. Our psychophysical experiments indicate that over a wide range of parameters, displays composed of two sets of randomly spaced parallel lines moving across each other in opposite directions are perceptually transparent. If the spacings between every two adjacent lines are

made equal, however, the resulting equally spaced line patterns are nontransparent.

Two sets of N parallel lines with initial positions x_n^+ and x_n^- ($n = 1, 2, \dots, N$), respectively, moving in the opposite directions with speed v , can be represented by

$$f_L(x, t) = c \sum_{n=1}^N [\delta(x - x_n^+ - vt) + \delta(x - x_n^- + vt)], \quad (9)$$

where $\delta()$ is the Dirac function and c is a constant with the dimension of the inverse of length. Assuming that v is positive, the two terms in the square bracket on the right-hand side represent lines moving to the right and left, respectively. It can be shown (see Appendix) that under appropriate assumptions, the opponent energy for such a line pattern is given by

$$E_L \approx c' \sum_{n=1}^N \left\{ \exp \left[-\frac{(x - x_n^+ - vt)^2}{v^2 \sigma_t^2 + \sigma_x^2} \right] - \exp \left[-\frac{(x - x_n^- + vt)^2}{v^2 \sigma_t^2 + \sigma_x^2} \right] \right\} + c' \sum_{n,m}^N \sum_{n \neq m}^N \left\{ \exp \left[-\frac{(x - x_n^+ - vt)^2 + (x - x_m^+ - vt)^2}{2(v^2 \sigma_t^2 + \sigma_x^2)} \right] \cdot \cos[\omega_x(x_m^+ - x_n^+)] - \exp \left[-\frac{(x - x_n^- + vt)^2 + (x - x_m^- + vt)^2}{2(v^2 \sigma_t^2 + \sigma_x^2)} \right] \cdot \cos[\omega_x(x_m^- - x_n^-)] \right\}, \quad (10)$$

where the constant c' is equal to $c^2/2\pi(v^2\sigma_t^2 + \sigma_x^2)$.

The second summation is unlikely to achieve large magnitude because $x - x_n^+ - vt$ and $x - x_m^+ - vt$ (and similarly, $x - x_n^- + vt$ and $x - x_m^- + vt$) cannot both be zero at a given (x, t) for $n \neq m$. For randomly spaced line patterns the second summation is much smaller in magnitude than the first for the additional reason that the cosine terms are equally likely to be positive or negative. We therefore only need to consider the contribution of the first summation. The first term in the summation will generate a large positive contribution to the opponent energy around those positions in the $x-t$ plane such that $x - x_n^+ - vt$ is close to 0, that is, around the trajectory of each right-going line in the spatiotemporal space. Similarly, the second term in the first summation will have a large negative contribution to the opponent energy when $x - x_n^- + vt$ is very small in magnitude. These two terms will not sufficiently cancel each other due to the random location of the lines. There will be both large positive and large negative opponent energies across the pattern at any instance of time, indicating both the rightward and leftward motion (the presence of transparent motion).

For equally spaced line patterns we have

$$x_m^+ - x_n^+ = x_m^- - x_n^- = (m - n)\Delta x, \quad (11)$$

where Δx is the spacing between two adjacent lines, so the two cosine terms in the second summation of Equation 10 are equal. At periodic time intervals, each line in one set is spatially very close to one, and only one, line in the other set and the two corresponding terms in both summations of Equation 10 will strongly cancel each other. This periodic loss of motion signal helps to explain the oscillatory perception for these patterns.

The above argument is based on the assumption that filters of reasonable sizes (σ_t and σ_x) are used. The difference between the opponent energy of an equally spaced line pattern and that of the

corresponding randomly spaced line pattern certainly depends on the choice of σ_x and σ_t . We consider here two extreme cases. First, let $\sigma_x \rightarrow \infty$ and $\sigma_t \rightarrow \infty$, that is, use extremely wide filters. It is clear from Equation 10 that the opponent energy $E_L \rightarrow 0$ no matter whether the lines are equally spaced or randomly spaced, indicating that both types of patterns should appear nontransparent. This agrees with our informal observation that a pattern always become less transparent or nontransparent when it is viewed with more eccentric part of retina, where the filter sizes are presumably larger. We next consider the case when the filters are extremely narrow, that is, $\sigma_x \rightarrow 0$ and $\sigma_t \rightarrow 0$. Under this condition, Equation 10 reduces to

$$E_L(x, t) = c \sum_{n=1}^N [\delta(x - x_n^+ - vt) - \delta(x - x_n^- + vt)]. \quad (12)$$

We basically recover the original line patterns with one set of lines contributing positive opponent energy, and the other negative energy. Thus, there are motion signals in both directions even for the equally spaced line patterns, except at the moment when the two opposite-going components are about to superimpose. At that time, $x_n^+ - vt = x_n^- + vt$ for all n , and the opponent energy in Equation 12 is 0. This corresponds closely to what we observed for the equally spaced line patterns when the number of lines was small and therefore the spacing between lines was large (increasing the spacing between lines is mathematically equivalent to reducing σ_x and σ_t). Physiologically, the motion sensitive filters are not arbitrarily large or small. As a result, we see randomly spaced line patterns as much more transparent than the corresponding equally spaced line patterns over a range of parameters.

We did computer simulations with randomly spaced and equally spaced parallel line patterns. In Figure 3, *a* and *b* are the spatiotemporal representations of two such line patterns. Each figure represents 5° in the spatial dimension and 2.5 sec in the temporal dimension, respectively. The speed of all the lines is $2^\circ/\text{sec}$. Since each line is wrapped around when it moves out of the spatial window, there are 15 lines moving in each of the two opposite directions at any instance of time for both line patterns. With these parameters the randomly spaced line pattern is perceptually transparent while the equally spaced line pattern is not, according to our psychophysical observations. We used the same three sets of Gabor filters as in Figure 2. They represent low-, medium-, and high-frequency channels, or equivalently, wide, medium, and narrow spatial scales. The central spatial and temporal frequencies of the medium channel are equal to the fundamental spatial and temporal frequencies of the equally spaced line patterns. The opponent energies from the three channels for the two line patterns are shown in Figure 3*c-h*. Again, in this figure gray indicates little motion energy and white and black code for rightward and leftward opponent energies, respectively. It is clear that the opponent energy for the randomly spaced line pattern contains both rightward and leftward motion signals while that for the equally spaced line patterns is much weaker. This corresponds well with the presence and absence of perceptual motion transparency in the two displays. Also, the way opponent motion energies change across the three scales of the filters follows what we predicted based on Equation 10: as the filters get wider, the opponent energies become weaker. Likewise, as the filters get narrower, the difference between the two types of patterns becomes smaller, and the equally spaced line patterns produce some opponent energy.

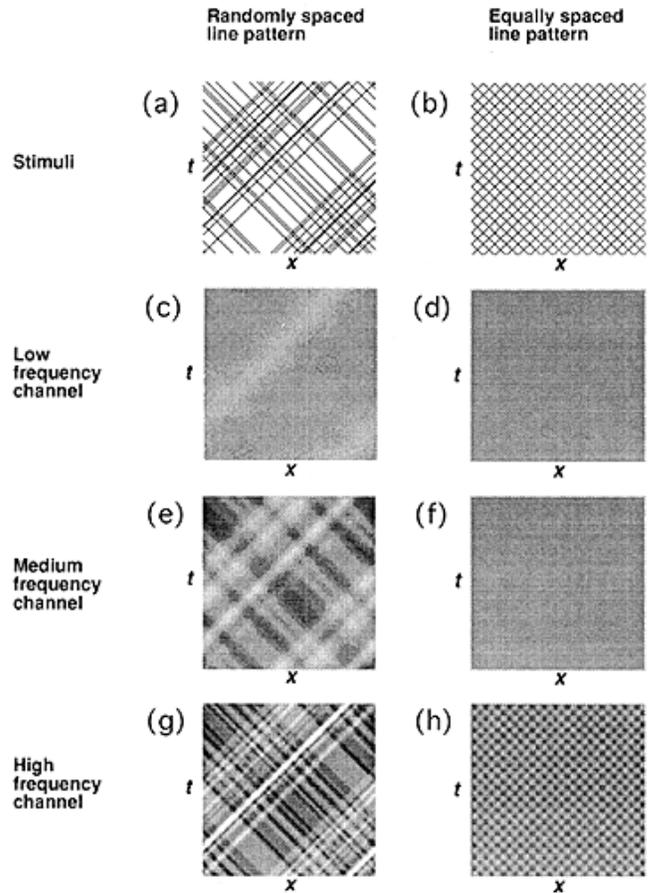


Figure 3. Simulation for an equally spaced and the corresponding randomly spaced parallel line patterns. All figures are shown in spatiotemporal space. *a* and *b* are the spatiotemporal representations of the two patterns. The total number of black pixels is exactly the same in these two patterns. *c-h* show the opponent energies in three different frequency channels for the two patterns, all shown with the same gray scale. Gray indicates little motion energy and white and black code for rightward and leftward opponent energies, respectively. See text for the details of the parameters used.

This corresponds well with the perception that as the number of lines in an equally spaced line patterns decreases, the patterns become a little more transparent.

Dot patterns

We now consider the paired and the unpaired dot patterns, which are perceptual nontransparent and transparent, respectively. Since these dot patterns have contrast variations along the y -dimension as well as along the x - and t -dimensions, we need three-dimensional Gabor filters for computing their opponent energies. As is shown in the Appendix, the results of analysis are rather similar to those for the line patterns described in the previous section. The introduction of the y -dimension simply adds an amplitude term and a phase term to the response for each dot. There is an additional amplitude term due to the limited lifetime of the dots. The presence and absence of motion transparency in paired and unpaired dot patterns could be explained in a similar way as for the line patterns. For the paired dot patterns with two dots in each pair having different signs of contrast (one black, the other white), the results are similar

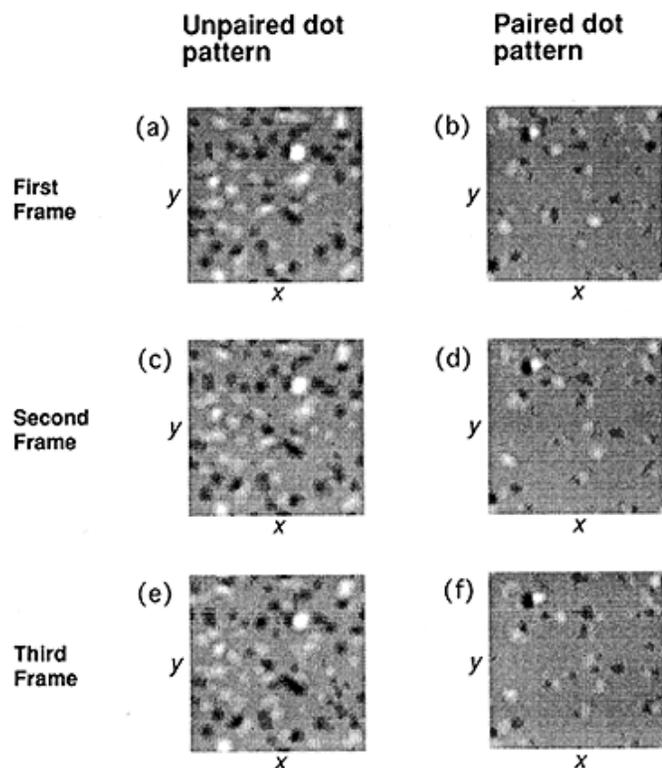


Figure 4. Simulation for a paired and the corresponding unpaired dot patterns. The two columns show the results for the unpaired and the paired dot patterns, respectively. The three rows represent the spatial distribution of the opponent energies for the two types of patterns at three successive time frames. All figures are displayed with the same gray scale. Gray indicates little motion energy and white and black code for rightward and leftward opponent energies, respectively. See text for the details of the parameters used.

because the unidirectional energies do not depend on the signs of contrast due to the squaring nonlinearity. For paired dot patterns with large vertical offsets, the two dots in each pair will generate quite different motion energies in opposite directions due to the y -dependent amplitude term. Their contributions will not cancel well locally and the patterns will thus appear transparent.

We have carried out computer simulations for the paired and the unpaired dot patterns. An example is shown in Figure 4. The simulation was done with a set of filters with central frequency equal to 2.23 cycles/degree, and Gaussian width 0.1° . Note that in Figure 4 the two axes of each figure represent two spatial dimensions, instead of one spatial and one temporal dimension as in Figures 2 and 3. The temporal dimension in Figure 4 is represented by showing three successive time frames in three rows. It is clear from the figure that the unpaired dot pattern contains much stronger leftward and rightward opponent motion energies than the paired dot patterns, in agreement with our perception.

It is important to note that the Fourier power spectra of the paired and the unpaired dot patterns are rather similar. This indicates that the perceptual difference of the two types of patterns is unlikely to be explained by the motion energy measurements alone without the introduction of the suppression stage. We have carried out the Fourier transformation on the paired and the unpaired dot patterns. For simplicity, we considered dots moving in the x -dimension and ignored the y -di-

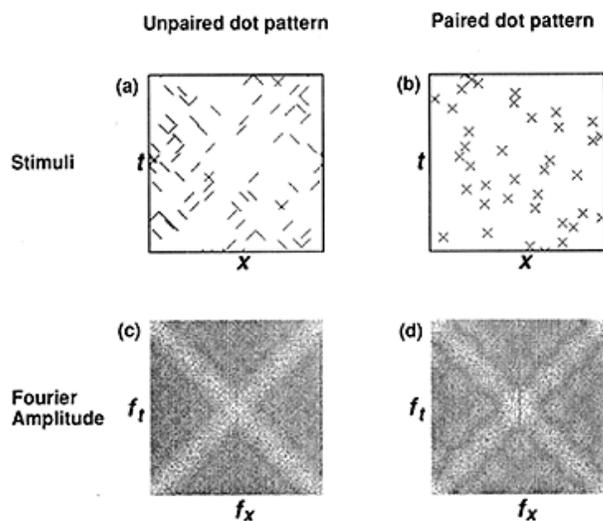


Figure 5. Fourier transforms of a paired and an unpaired dot patterns. *a* and *b* are the spatiotemporal representations of an unpaired dot pattern and its corresponding paired dot pattern. The y -dimension is not shown. *c* and *d* are the amplitudes of the Fourier transforms of the two patterns displayed with the same gray scales. The zero spatiotemporal frequency points are located at the centers of the diagrams.

mension of the patterns. In Figure 5, *a* and *b*, show the spatiotemporal representations of a paired dot pattern and its corresponding unpaired dot pattern. The amplitudes of the Fourier transforms of the two patterns are shown in Figure 5, *c* and *d*. The points with zero spatial and temporal frequencies are located at the centers of both diagrams. It is clear from these figures that both patterns have their main Fourier power concentrated along the two diagonal lines going through the origin. These two lines are generated by the dots moving in the two opposite directions of motion (Watson and Ahumada, 1985). If one tries to detect motion by doing something equivalent to fitting lines (through origin) to these two spectra (Heeger, 1988; Shizawa and Mase, 1990), then the paired and the unpaired dot patterns will both be considered transparent by such a procedure. Since perceptually the unpaired dot patterns are much more transparent than the paired ones, we conclude that the suppression stage is essential for determining the perceptual transparency of a display.

Disparity specificity

We showed that the paired dot patterns can be made perceptually transparent if a certain amount of binocular disparity is introduced between the dots in each pair (Qian et al., 1994). To model this interaction between motion and stereo vision, we need to incorporate disparity sensitivity into the motion energy model. We have recently developed a model of stereo vision based on known properties of binocular cells in the visual cortex and have shown that this model can be naturally combined with motion energy models (Qian, 1994). Here we show through computer simulations how the combined model can be used to explain the effect of disparity on perceptual motion transparency. Our model assumes that the left and right receptive fields of a binocular cell are given by

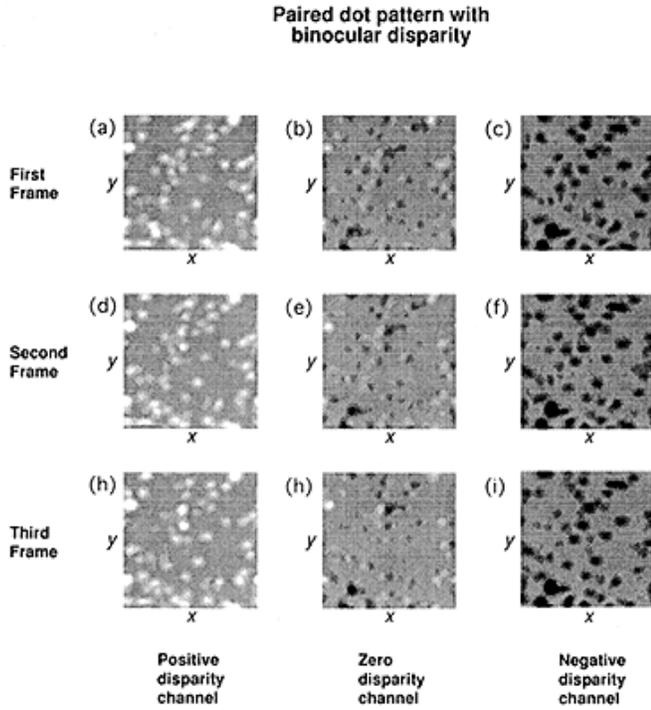


Figure 6. Simulation for a paired dot pattern with one set of dots moving in one direction with positive disparity, and the other set of dots in the opposite direction with negative disparity. The three rows show the spatial distribution of the opponent energies for the pattern at three successive time frames. The three columns correspond to the opponent energies in three disparity channels. All figures are displayed with the same gray scale. Gray indicates little motion energy and white and black code for rightward and leftward opponent energies, respectively. See text for the details of the parameters used.

$$f_l(x, y, t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right) \cdot \cos(\omega_x x + \omega_y y + \omega_t t + \phi_l), \quad (13)$$

$$f_r(x, y, t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right) \cdot \cos(\omega_x x + \omega_y y + \omega_t t + \phi_r), \quad (14)$$

where σ and ω are the widths of the Gaussians and angular frequencies along the spatial and temporal dimensions, and ϕ_l and ϕ_r are the phase parameters. For a stimulus with a constant disparity D and moving at speed v_x and v_y along the horizontal and vertical directions, it can be shown that the energy computed with a quadrature pair of such filters is approximately

$$E \approx 4\rho^2 \delta^2(\omega_t + \omega_x v_x + \omega_y v_y) \cos^2\left(\frac{\phi_l - \phi_r}{2} + \frac{\omega_x D}{2}\right), \quad (15)$$

where $\delta()$ is the delta function, and ρ is the amplitude of the Fourier transformation of the stimulus (Qian, 1994). Equation 15 indicates that the cell is indeed sensitive to both motion and stereo disparity, similar to some real cortical cells. While ω_x , ω_y , and ω_t determine the motion selectivity, $(\phi_l - \phi_r)$ determines the disparity sensitivity. The width of disparity tuning (defined at the half peak amplitude) is equal to

$$\Delta D = \frac{\pi}{\omega_x}. \quad (16)$$

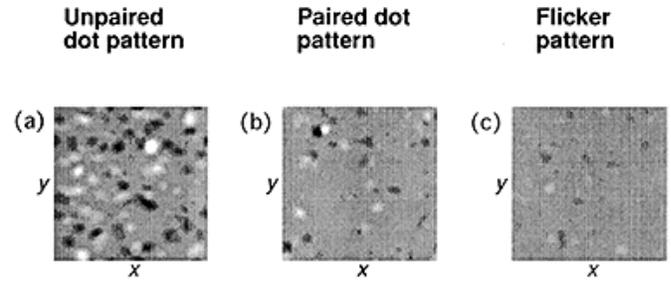


Figure 7. Opponent energies for the same paired and the unpaired dot patterns in Figure 4, and the corresponding flicker pattern. Only one time frame for each spatial energy distribution is shown. All figures are displayed with the same gray scale. Gray indicates little motion energy and white and black code for rightward and leftward opponent energies, respectively. See text for the details of the parameters used.

We can thus explain the effect of both frequency and disparity cues in our psychophysical experiments by restricting opponency to unidirectional energies computed with filters with identical ω_x , ω_y , and ω_t , and $(\phi_l - \phi_r)$, but tuned to opposite directions of motion.

We carried out a simulation on a paired dot pattern with one set of dots moving in one direction with disparity 0.11° , and the other set of dots in the opposite direction with disparity -0.11° . The results are shown in Figure 6. Again, the three columns represent three successive time frames. The three columns represent the opponent energies computed with three sets of filters with their $(\phi_l - \phi_r)$ equal to $-\pi/2$, 0 , and $\pi/2$, respectively. The central spatial frequency in horizontal dimension, $(\omega_x/2\pi)$, was equal to 2.23 cycles/degree (or the angular frequency ω_x equal to 14.0 radians/degree). With this choice of ω_x , the three sets of filters have their peak disparity tuning around 0.11° , 0° , and -0.11° , respectively, and the widths of tuning are 0.22° . We see from Figure 6 that the leftward and rightward opponent motion energies are now segregated into different disparity channels. Unlike the paired dot pattern without disparity shown in Figure 4, *b*, *d*, and *f*; these energies do not cancel each other out. This accounts for the perceptual transparency of the pattern.

Flicker responses

Since nontransparent patterns such as counterphase gratings and paired dot patterns look rather like flicker, we also generated a flicker pattern for comparison. The pattern was derived from the unpaired dot pattern in Figure 4 by setting the dot speed to zero. Other parameters including dot lifetime are the same. The computed opponent motion energy of the flicker pattern, together with the opponent energies for the paired and the unpaired dot patterns from Figure 4, are shown in Figure 7*a-c*. The opponent motion energy of the paired dot pattern is only slightly stronger than that of the flicker pattern, and both are significantly weaker than the opponent energy of the unpaired dot pattern. These simulations conform to our finding that MT cells, responses to the paired dot patterns and to the flicker noise are not significantly different from each other, and that both responses are significantly weaker than the response to the unpaired dot patterns (Qian et al., 1994).

Divisive inhibition

We have shown above that simple subtractive inhibition following motion energy computation can account for our psy-

chophysical and physiological results on transparent and nontransparent patterns. Single-unit recordings from behaving monkeys, on the other hand, indicate that the directional inhibition in area MT has a divisive component (Snowden et al., 1991). Heeger (1992) also demonstrated that divisive normalization can account for a large amount of psychophysical and physiological results, some of which are inconsistent with pure subtractive inhibition. For example, the opponent energy computed with subtractive inhibition is always proportional to the square of stimulus contrast while real visual cells exhibit contrast saturation. It is therefore important to demonstrate that our psychophysical and physiological experiments can also be explained by divisive suppression.

We have repeated all our simulations described above using divisive suppression and similar results (not shown) have been obtained. We considered two different normalization schemes. The first one normalizes the opponent energy by the static energy (Adelson and Bergen, 1986):

$$NE(x,t) = \frac{E^+(x,t) - E^-(x,t)}{E^0(x,t) + \varepsilon}, \quad (17)$$

where $E^+(x,t)$ and $E^-(x,t)$ are the left and the right motion energies, respectively, $E^0(x,t)$ is the static energy, and ε is a small number for stabilizing the division. This approach is closely related to the least-squares method (Lucas and Kanade, 1981) and gradient method (Horn and Schunck, 1981) for velocity estimation (Adelson and Bergen, 1986; Simoncelli, 1993). It is not surprising that Equation 17 can also explain the perceptual difference between the transparent and the nontransparent patterns since it is proportional to the opponent energy.

The second approach normalizes the output of each unidirectional energy detector by the sum of the outputs of all detectors tuned to the same spatial frequency range. For displays containing only opposite directions of motion, the normalized left and right motion energies within a given frequency band are

$$NE^\pm(x,t) = \frac{E^\pm(x,t)}{E^+(x,t) + E^-(x,t) + c}, \quad (18)$$

where c represents contributions from filters tuned to directions other than left and right. This approach is modified from Heeger (1992), who used the total output of detectors tuned to all frequency ranges as the normalization factor. We restricted normalization to be within each frequency channel because our psychophysical experiments suggest that the suppression is spatial frequency specific (Qian et al., 1994). For the convenience of the description we let c be equal to zero. For the nontransparent patterns with locally well-balanced motion signals, $E^\pm(x,t)$ is approximately equal to $E(x,t)$ in Equation 18 so that the normalized motion energies are very close to 0.5. For the unbalanced patterns that are perceptually transparent, on the other hand, there are many locations at which $E^\pm(x,t)$ is much larger than $E(x,t)$ or vice versa. The normalized left or right motion energies at these locations will be close to 1.0. This twofold difference in normalized energies between the balanced and the unbalanced patterns could account for the difference in their perceptual transparency. Notice that this difference cannot be reduced or reversed by changing the contrast of the patterns because the normalized energy is independent of contrast. Also note that the normalized energies of nontransparent patterns are similar to those of flicker patterns. The latter are also close to

0.5 since flicker patterns contain equal amounts of left and right motion energy.

Discussion

We have shown in this article that a motion energy computation followed by disparity- and spatial frequency-specific suppression among different directions of motion can indeed explain the perceptual difference of the transparent and nontransparent displays used in our psychophysical experiments. Specifically, we found that the nontransparent displays generate relatively weak opponent or normalized energies at the suppression stage. In fact, these energies are not higher than those generated by flicker patterns. On the other hand, the perceptually transparent displays generate much stronger opponent or normalized motion energies along more than one direction of motion. These energies in different directions are located either in different but mixed small areas (as in the randomly spaced line pattern and the unpaired dot pattern), or in different disparity or spatial frequency channels over the same spatial regions (as in the display made of two different sine wave gratings, and in the paired dot pattern with binocular disparity). We hypothesize that a later stage could integrate these energies in different directions separately to form two over-lapping transparent surfaces. Note that a pattern moving in a single direction will not appear transparent because it will only generate strong opponent or normalized motion energies in one direction. A display containing two widely separated objects moving in opposite directions will not appear transparent either. Although such a display will generate strong opponent or normalized motion energies in two different directions, these energies are not spatially mixed and therefore cannot be integrated into two overlapping surfaces.

Previous physiological experiments indicate that MT cells show strong suppression among different directions of motion (Snowden et al., 1991). MT could therefore be the physiological equivalent of the suppression stage in our simulation, where transparent and nontransparent displays can be distinguished. In fact, our physiological recordings in a preceding companion article demonstrate that average MT activity to the transparent unpaired dot patterns is significantly higher than that to the nontransparent paired dot patterns (Qian and Andersen, 1994). In addition, our psychophysical experiments with the paired and the unpaired dot patterns indicate that under the foveation condition, an alignment of opposing motion signals on the scale of 0.4° can generate large difference in perceptual transparency. We therefore propose that the suppression stage for differentiating transparent and nontransparent occurs at the subunit level of MT receptive fields. V1 cells, on the other hand, show relatively weak suppression among different directions of motion (Snowden et al., 1991; Qian and Andersen, 1994). Many of them behave rather like unidirectional motion energy detectors. They respond quite well to both transparent and nontransparent patterns, and the average V1 activity could not reliably tell the two types of patterns apart (Qian and Andersen, 1994). V1 and MT therefore approximately correspond to the energy computation and suppression stages in our simulations.

We used Gabor filters along both spatial and temporal dimensions for motion detection in our simulations. While the spatial receptive field structures of simple cells are known to be described by Gabor functions well (Jones and Palmer, 1987), it is not physiologically plausible to use Gabor functions for the temporal responses. Temporal Gabor filters are nonzero on the negative time axis. They are thus noncausal. This is, however,

not a major problem because these filters decay to zero exponentially due to the Gaussian envelopes. We could practically make the filters causal by shifting them toward the positive time direction by 3σ . Such a shift will only generate a phase term and will not affect energy measures. A more serious problem with using temporal Gabor filters is that the temporal response of real simple cells is skewed, with its envelope having a longer decay time than rise time. Also, zero-crossing intervals in the temporal dimension are not equally spaced (DeAngelis, et al., 1993). In order to make our model more biologically relevant, we have repeated our simulations using more realistic temporal response functions given in Equation 1 of Adelson and Bergen (1985), and obtained the similar results. We believe that our results do not depend on the details of the receptive field shapes of the energy detectors as long as they approximately measure the local Fourier power within a certain spatiotemporal frequency window specified by their parameters.

The suppression among different directions of motion used in our simulation makes it impossible for a stimulus to generate strong responses along more than one direction of motion in each small spatial area at the opponent stage when there are no other cues in the stimulus, such as disparity or spatial frequency (the weak opponent energies for the paired dot pattern in Fig. 4h, *d-f* give an example). The size of the small area is determined by the size of the front-end filters. In this regard, the suppression stage is rather like the pooling or regularization step commonly used in machine vision systems (Horn and Schunck, 1981; Hildreth, 1984; Heeger, 1987; Wang et al., 1989; Grzywacz and Yuille, 1990). Such a step is required to solve the aperture problem and to average out noise while at the same time it prevents those models from having more than one velocity estimation over each area covered by the pooling operator. In this connection, it is interesting to note that some versions of the pooling procedures for combining local gradient constraints (Horn and Schunck, 1981; Lucas and Kanade, 1981) are equivalent to a mixture of subtractive and divisive types of suppression (Simoncelli, 1993). The agreement between our simulations and the psychophysical observations implies that machine vision systems can be made more consistent with transparent motion perception if the pooling operation in these systems is restricted to small areas and to each frequency and disparity channel. We suggest that the difficulty most machine vision systems have with motion transparency can be partly attributed to the fact that these systems typically apply pooling operations over a relatively large region and that they usually do not explore other cues such as disparity and spatial frequency to restrict the scope of pooling.

The displays we used in our psychophysical and physiological experiments and computer simulations are highly artificial and are unlikely to be found in the natural environment. What, then, is the advantage of having a suppression stage, such as MT, in the motion pathway if it is not just for making counterphase gratings or paired dot patterns appear nontransparent? In fact, if a subpopulation of V1 cells act like unidirectional motion energy detectors, why doesn't the brain use a family of these cells tuned to different directions of motion to represent the perception of multiple motions? Why instead should perception be derived from the suppression stage in MT, which reduces the system's acuity to transparent motion (Snowden, 1989) and at the same time makes the well-balanced patterns appear nontransparent? The identification of the suppression stage with the pooling operation in machine vision systems discussed above provides

an answer. Since the function of the pooling operation is to solve the aperture problem and to average out noise, we suggest that directional suppression has similar functions. Indeed, unidirectional energy detectors like V1 cells suffer the aperture problem; that is, they seem to respond only to the component of the motion that is perpendicular to the local spatial orientations of the stimulus contrast. Also, V1 cells are very responsive to dynamic noise patterns made of flickering dots (Qian and Andersen, 1994). The suppression stage in the motion pathway could help to solve these problems, just as the pooling operations do in machine vision systems. In fact, we found that the noise response of MT cells is much lower than that of V1 cells (Qian and Andersen, 1994). There is also evidence that the human visual system may solve the aperture problem by averaging local motion measurements (Ferrera and Wilson, 1990, 1991; Yo and Wilson, 1992; Rubin and Hochstein, 1993). Suppression among different directions of motion could be viewed as a kind of averaging operation and thus could be used to solve the aperture problem. A negative effect of the suppression is the reduced acuity for transparent motion. This problem is minimized, however, by applying suppression locally and by restricting it within each disparity and spatial frequency channel, since multiple motions in the real world are usually not precisely balanced in each local area and different objects tend to have different disparity and spatial frequency distributions. While the inhibition among the cells within each disparity and spatial frequency channel could help to combine V1 outputs into a single motion signal at each location in order to solve the aperture problem and to reduce noise, different disparity and spatial frequency channels could represent multiple motions at the same spatial location.

It also seems reasonable to assume that at each spatial location of the visual field, cells at the suppression stage that are tuned to the same direction of motion, but different spatiotemporal frequency bands, should facilitate each other. These cells all carry consistent motion signals from different frequency ranges at the same spatial location and these signals are likely generated by the same moving object.

The model we have proposed for motion transparency is incomplete in several ways. One problem is that while we perceive each transparent surface as a coherent whole the output of the model at the suppression stage contains many isolated patches (see Figs. 4, 6). This problem can be solved with the introduction of spatial integration. We can assume that cells tuned to similar directions of motion at nearby spatial locations have excitatory connections between them. When enough of the spatially mixed cells tuned to a given direction are active, the activity will spread across the whole layer of cells tuned to the same direction of motion. Transparency will then correspond to a multi-peaked distribution of the direction of motion at all spatial locations. Spatial summation could also occur within individual MT cells through facilitation between subunits in a cell's receptive field. Another issue we have not addressed is whether explicit speed estimation has an important role in motion transparency. In fact, the outputs of our model are motion energies in different frequency and disparity channels instead of explicit velocity estimations. In grouping together elements to form surfaces, does the visual system consider the magnitudes of the velocity vectors of these elements, or just their directions? To explore this question we used transparent random dot patterns and assigned the speeds of the dots composing each surface according to a uniform probability distribution. We found that when the

width of the distribution was as large as half of the average speed of the dots, motion transparency could still be observed without much difficulty. This suggests that transparent motion perception does not require precise knowledge of the relative speed of the individual dots. This observation is consistent with a recent result by Watamaniuk and Duchon (1992), who found that human visual system tends to average speed information.

We can now sketch a more complete model for biological motion detection based on the above discussions. At each spatial location there is a population of unidirectional motion energy detection cells tuned to different disparity and different three-dimensional spatiotemporal frequency bands. Each cell receives input from the moving pattern in its receptive field according to the amount of Fourier power of the pattern falling on its disparity and frequency window. At this stage families of V1 cells tuned to different directions of motion may be activated, but this could be caused by the aperture problem of a single moving object, by noise in the environment, or by true transparent motion of different objects. Further processing has to be carried out by connections among the cells. At each spatial location the cells with similar disparity and frequency tuning, but different direction preferences, are mutually inhibitory while those with different frequency tuning but similar directional preference are mutually excitatory. This is a generalization of the suppression between opposite directions of motion described in this article and it combines raw measurements within each disparity and frequency channel at each spatial location into a single motion representation (or more accurately a unimodal distribution). The process solves the noise and the aperture problem and still allows multiple motions to be represented in different small areas or among different disparity and frequency channels. Finally, the cells at nearby spatial locations with similar directional preferences have excitatory connections between them to facilitate spatial integration. This process combines consistent local measurements into coherent surfaces. We are currently implementing this more complete model of biological motion processing.

Appendix

We derive the opponent energy expressions for the counterphase gratings, line patterns, and dot patterns in this section. For the convenience of subsequent calculations and presentation, we introduce the complex Gabor filters,

$$g_c^\pm(x, t) = \frac{1}{2\pi\sigma_x\sigma_t} \exp\left[-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2} + i(\omega_x x \pm \omega_t t)\right], \quad (19)$$

whose real and imaginary parts are the even and odd Gabor functions in Equations 1 and 2. Since the convolution of any real function $f(x, t)$ with the even and odd Gabor filters is equal to the real and imaginary parts of the convolution of that function with the complex filters, we will only give responses of the complex filters for brevity. Using the complex Gabor function, the unidirectional and opponent motion energies can be written as

$$E^+(x, t) = |f * g_c^+|^2, \quad (20)$$

$$E^-(x, t) = |f * g_c^-|^2, \quad (21)$$

and,

$$E(x, t) = |f * g_c^+|^2 - |f * g_c^-|^2, \quad (22)$$

respectively.

Counterphase gratings

A single sine wave grating with spatial frequency Ω_x and temporal frequencies Ω_t , and drifting to the right is represented by

$$f_s^+(x, t) = \sin(\Omega_x x - \Omega_t t), \quad (23)$$

where both Ω_x and Ω_t are assumed to be positive. It can be shown that the responses of the complex Gabor filters in Equation 19 to the sine wave grating are given by

$$f_s^+ * g_c^\pm = \frac{1}{2i} \left\{ u(-\omega_x, \pm\omega_t) \exp[i(\Omega_x x - \Omega_t t)] - u(+\omega_x, \pm\omega_t) \exp[i(-\Omega_x x + \Omega_t t)] \right\}, \quad (24)$$

where $u(\omega_x, \omega_t)$ are defined as

$$u(\omega_x, \omega_t) = \exp\left[-\frac{\sigma_x^2}{2}(\Omega_x + \omega_x)^2 - \frac{\sigma_t^2}{2}(\Omega_t + \omega_t)^2\right]. \quad (25)$$

Note that $f_s^+ * g_c^+$ can be obtained from $f_s^+ * g_c^-$ by replacing ω_t with $-\omega_t$. Since $\Omega_x, \Omega_t, \omega_x,$ and ω_t are all assumed to be positive, $u(-\omega_x, -\omega_t)$ is usually much larger than $u(\omega_x, -\omega_t)$, $u(-\omega_x, \omega_t)$, and $u(\omega_x, \omega_t)$. It achieves its maximum value when the central spatial and temporal frequencies of the filters match those of the counterphase grating. It is clear from Equation 24 that the rightward motion detector (g_c^+) responds to the right-going sine wave much better than the leftward motion detector (g_c^-) does.

The sine wave grating drifting to the left with the same spatial and temporal frequencies as in Equation 23 can be obtained by replacing Ω_t with $-\Omega_t$ in Equation 23. The responses of the complex Gabor filters to this grating can be obtained by the same replacement in Equation 24. Using these expressions, the opponent energy for the counterphase grating (composed of the above two sine wave gratings moving in the opposite directions) can then be shown to equal to 0 independent of the parameters for the filters and the grating.

Line patterns

A single line with initial position x_0 and moving to the right with speed v can be represented by

$$f_l^+(x, t) = c\delta(x - x_0 - vt), \quad (26)$$

where δ is the Dirac function and c is a constant with the dimension of the inverse of length. The responses of the complex Gabor filters to the line can be shown to be

$$f_l^+ * g_c^\pm = \frac{c}{\sqrt{2\pi(v^2\sigma_t^2 + \sigma_x^2)}} \cdot \exp\left[-\frac{(x - x_0 - vt)^2 + \sigma_x^2\sigma_t^2(\omega_t \pm \omega_x v)^2}{2(v^2\sigma_t^2 + \sigma_x^2)}\right] \cdot \exp\left[i(x - x_0 - vt)\frac{\omega_x\sigma_x^2 \pm \omega_t\sigma_t^2 v}{v^2\sigma_t^2 + \sigma_x^2}\right]. \quad (27)$$

To simplify subsequent calculations, we assume that

$$\omega_t = \omega_x v. \quad (28)$$

That is, the spatial and temporal frequencies of the filters are such that they are tuned to the speed of the line. Also, $f_l^+ * g_c^+$ is usually much smaller in magnitude than $f_l^+ * g_c^-$ because the preferred direction of the filter is opposite to that of the line motion. Then, Equation 27 can be simplified to

$$f_l^+ * g_c^+ = \frac{c}{\sqrt{2\pi(v^2\sigma_t^2 + \sigma_x^2)}} \cdot \exp\left[-\frac{(x-x_0-vt)^2}{2(v^2\sigma_t^2 + \sigma_x^2)}\right] \exp[i\omega_x(x-x_0-vt)], \quad (29)$$

$$f_l^+ * g_c^- \approx 0. \quad (30)$$

We can now consider two sets of N lines represented by Equation 9. Using Equations 29 and 30, we obtain the opponent energy for the line pattern as shown in Equation 10.

Dot patterns

A dot with initial position (x_0, y_0) and moving in the $+x$ direction with speed v from time t_1 to t_2 ($t_2 > t_1$) can be represented by

$$f_d^+(x, t) = c\delta(x-x_0-vt)\delta(y-y_0)H(t-t_2)H(t-t_1), \quad (31)$$

where H is the step function and c is a constant with the dimension of the inverse of length squared. With the three-dimensional extension of the Gabor filters in Equation 19, it can be shown that

$$f_d^+ * g_c^+ = \exp\left[-\frac{(x-x_0-vt)^2 + \sigma_x^2\sigma_t^2(\omega_x \pm \omega_x v)^2}{2(v^2\sigma_t^2 + \sigma_x^2)} - \frac{(y-y_0)^2}{2\sigma_y^2}\right] \cdot \exp\left[-i(x-x_0-vt)\frac{\omega_x\sigma_x^2 \pm \omega_t\sigma_t^2 v}{v^2\sigma_t^2 + \sigma_x^2} + i\omega_y(y-y_0)\right] \cdot \frac{c}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_t} \int_{t-t_1+b/2a}^{t-t_2+b/2a} \exp(-ax^2)dx, \quad (32)$$

where

$$a = \frac{1}{2}\left(\frac{v^2}{\sigma_x^2} + \frac{1}{\sigma_t^2}\right), \quad (33)$$

$$b = \frac{(x-x_0-vt)v}{\sigma_x^2} + i(\omega_t - \omega_x v). \quad (34)$$

Again, if we assume Equation 28 and neglect $f_d^+ * g_c^-$, Equation 32 can be simplified to

$$f_d^+ * g_c^+ = \frac{c}{2\pi\sqrt{v^2\sigma_t^2 + \sigma_x^2}} \exp\left[-\frac{(x-x_0-vt)^2}{2(v^2\sigma_t^2 + \sigma_x^2)} - \frac{(y-y_0)^2}{2\sigma_y^2}\right] \cdot \exp[i\omega_x(x-x_0-vt) + i\omega_y(y-y_0)] \cdot \{erf[\sqrt{a}(t-t_2+b/2a)] - erf[\sqrt{a}(t-t_1+b/2a)]\}, \quad (35)$$

$$f_d^+ * g_c^- \approx 0, \quad (36)$$

where

$$b' = \frac{(x-x_0-vt)v}{\sigma_x^2}. \quad (37)$$

Comparing Equation 35 with Equation 29, we see that the two expressions are rather similar. The introduction of the y -dimension simply adds an amplitude term and a phase term to the response. The term containing two error functions is due to the limited lifetime of the dot. We could derive the opponent energy expression for the paired and unpaired dot patterns using Equation 35 but we omit it here because it would be very similar to Equation 10 for the line patterns.

References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am [A]* 2:284-299.
- Adelson EH, Bergen JR (1986) The extraction of spatiotemporal energy in human and machine vision. In: *Proceedings of the IEEE Workshop on Visual Motion*, pp 151-156.
- DeAngelis GC, Ohzawa I, Freeman RD (1993) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J Neurophysiol* 69:1091-1117.
- Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Res* 32:203-218.
- Fahle M, Poggio T (1981) Visual hyperacuity: spatiotemporal interpolation in human vision. *Proc R Soc Lond [Biol]* 213:451-477.
- Ferrera VP, Wilson HR (1990) Perceived direction of moving two-dimensional patterns. *Vision Res* 30:273-287.
- Ferrera VP, Wilson HR (1991) Perceived speed of moving two-dimensional patterns. *Vision Res* 31:877-894.
- Grzywacz NM, Yuille AL (1990) A model for the estimate of local image velocity by cells in the visual cortex. *Proc R Soc Lond [A]* 239: 129-161.
- Heeger DJ (1987) Model for the extraction of image flow. *J Opt Soc Am [A]* 4:1455-1471.
- Heeger DJ (1988) Optical flow using spatiotemporal filters. *Int J Comp Vis* 2:279-302.
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* in press.
- Hildreth EC (1984) Computations underlying the measurement of visual motion. *Artif Intell* 23:309-355.
- Horn BKP, Schunck BG (1981) Determining optical flow. *Artif Intell* 17:185-203.
- Jagadeesh B, Wheat HS, Ferster D (1993) Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science* 262:1901-1904.
- Jones JP, Palmer LA (1987) The two-dimensional spatial structure of simple receptive fields in the cat striate cortex. *J Neurophysiol* 58: 1187-1211.
- Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pp 674-679.
- McLean J, Palmer LA (1989) Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Res* 29:675-679.
- Nakayama K (1985) Biological image motion processing: a review. *Vision Res* 25:625-660.
- Qian N (1994) Computing stereo disparity and motion with known binocular cell properties. *Neural Comput*, in press.
- Qian N, Andersen RA (1994) Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *J Neurosci* 14: 7367-7380.
- Qian N, Andersen RA, Adelson EH (1991) V1 responses to two-surface transparent and non-transparent motion. *Soc Neurosci Abstr* 17:177.
- Qian N, Andersen RA, Adelson EH (1992) MT responses to transparent and non-transparent motion. *Soc Neurosci Abstr* 18:1101.
- Qian N, Andersen RA, Adelson EH (1994) Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *J Neurosci* 14:7357-7366.
- Reichardt W (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: *Sensory communication* (Rosenblith WA, ed). New York: Wiley.
- Reid RC, Soodak RE, Shapley RM (1987) Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proc Natl Acad Sci USA* 84:8740-8744.
- Rubin N, Hochstein S (1993) Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Res* 33:1385-1396.
- Shizawa M, Mase K (1990) Simultaneous multiple optical flow estimation. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Atlantic City.
- Simoncelli EP (1993) Distributed analysis and representation of visual motion. PhD thesis Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology.

- Snowden RJ (1989) Motions in orthogonal directions are mutually suppressive. *J Opt Soc Am [A]* 7:1096-1101.
- Snowden RJ, Treue S, Erickson RE, Andersen RA (1991) The response of area MT and V1 neurons to transparent motion. *J Neurosci* 11: 2768-2785.
- van Santen JPH, Sperling G (1985) Elaborated Reichardt detectors. *J Opt Soc Am [A]* 2:300-321.
- Wang HT, Mathur M, Koch C (1989) Computing optical flow in the primate visual system. *Neural Comput* 1:92-103.
- Watamaniuk SNJ, Duchon A (1992) The human visual system averages speed information. *Vision Res* 32:931-941.
- Watson AB, Ahumada AJ (1985) Model of human visual-motion sensing. *J Opt Soc Am [A]* 2:322-342.
- Yo C, Wilson HR (1992) Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Res* 32:135-147.