Transparent Motion Perception as Detection of Unbalanced Motion Signals. I. Psychophysics

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

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

Our visual system can solve the difficult problem of representing multiple motions in the same part of the visual space, the motion transparency problem. We investigated the conditions under which transparent motion perception occurs through psychophysical observations, using a series of visual displays composed of two simple patterns moving in opposite directions. We found that whenever a display has finely balanced opposing motion signals in all local regions, it is perceptually nontransparent. The displays that appeared transparent always contain locally unbalanced motion signals, with some local regions having net motion signals in one direction and some other regions in the opposite direction. These interdigitating net motion signals in both directions appear to be integrated separately to form two overlapping transparent surfaces. Displays that were spatially balanced could be made perceptually transparent if the two components moving in opposite directions were at different stereo depth planes or had different spatial frequency contents. Our results can be explained by proposing a disparity and spatial frequency-specific suppression stage in the motion pathway, at which motion signals of different directions, but of the same disparity and spatial frequency contents, locally inhibit each other. Such a mechanism would suppress noise input to the motion system, which generally activates several direction channels simultaneously, and would still not eliminate activity evoked by transparent surfaces that are at different depths or have different textures.

[Key words: motion transparency, visual psychophysics, stereopsis, motion-stereo interaction, random dot patterns, counterphase gratings]

Motion transparency is defined as the perception of more than one velocity field in the same part of the visual space. It is an important problem to study because transparent motion occurs frequently in the natural environment, due to either partial occlusions of moving objects or motion of overlapping semitransparent surfaces. For example, when one is looking at an animal moving behind bushes on a windy day, his visual system has to represent the velocity fields of both the animal and the bushes at the same time in the same part of space. Other examples are shadows moving across textured backgrounds, or stationary specular reflections from turning objects. A demonstration of transparent motion perception is a display with two independent sets of random dots moving in opposite directions in the same location in the visual field. Two transparent surfaces, one defined by each set of dots, are seen as continuously and independently moving across each other.

Not all patterns with two components moving in opposite directions give the percept of motion transparency. A well-known example is the so-called counterphase grating composed of two identical sine wave gratings moving across each other in opposite directions and with equal speed (Levinson and Sekuler, 1975). Instead of perceiving two coherent transparent sine wave gratings, one usually sees flicker or oscillation. It is therefore of interest to determine the conditions under which a composite pattern appears transparent. One possible explanation for the lack of transparency in counterphase gratings relies on the fact that a counterphase grating is mathematically equivalent to a single standing sine wave with temporally modulated contrast. However, we present in this article some other nontransparent displays that cannot be explained this way.

Motion transparency is also a challenging problem from the computational point of view, because most machine vision algorithms have difficulties with it. There are typically two steps of computation in these models: an initial measurement step that detects in each location some measure of the motion signal, and a subsequent pooling or regularization step that combines all the measurements in an area to obtain a single motion estimation for that area (Horn and Schunck, 1981; Lucas and Kanade, 1981; Hildreth, 1984; Heeger, 1987; Poggio et al., 1988; Wang et al., 1989; Grzywacz and Yuille, 1990). The measurements in the first step normally have the aperture problem (Hildreth, 1984): they are only sensitive to motion perpendicular to local spatial orientations in the stimuli. The second pooling step is required in order to solve the aperture problem. The second stage also helps to reduce noise and to interpolate motion across areas with sparse initial measurements. The pooling is usually based on some smoothness assumptions about the velocity field, or on the least-squares method, which finds the solution most consistent with the initial measurements from several nearby locations. These models work well for one motion but typically fail to recover multiple motions under transparent conditions [but see Shizawa and Mase (1990) for an exception, to be considered in Discussion]. By using an array of motion detectors with various preferred directions and speeds, the first step of these models can measure the different directions of motion of the transparent stimuli (Zucker et al., 1990). However, the second step poses a

major problem: there does not seem to be an easy way of combining these initial measurements from an area to obtain more than one motion vector. One might think that this could be achieved by first grouping the initial measurements into two sets and then performing the pooling operation within each set. But without extra cues other than motion, there is no basis for the grouping. A natural question is how the human visual system is able to solve the motion transparency problem, which has proved so difficult in machine vision.

In order to investigate the physiological basis of multiple motion representation, transparent random dot patterns have previously been used in our laboratory as stimuli in recording experiments from V1 and MT cells in behaving monkeys (Snowden et al., 1991). It was found that the responses of most MT cells to a transparent pattern, with one set of dots moving in their preferred directions and the other in their antipreferred directions, were significantly reduced compared to a single set of dots moving in their preferred directions alone. This result indicates strong suppression in MT between the preferred and the antipreferred directions of motion. This type of suppression was found to be much weaker in area V1. While MT is usually considered to be a major site for motion analysis, the role of the directional suppression of MT cells in transparent motion perception is not clear. Indeed, one would think that the subpopulation of directionally selective V1 cells whose responses to transparent displays are not much suppressed could better solve the problem of motion transparency.

To address the questions mentioned above, we have investigated the problem of motion transparency using a combination of psychophysical, computational, and physiological approaches. Our results are presented in the following series of three articles. The psychophysical findings, delineating the conditions for transparent motion perception, are reported in this article. In the second article, we present our physiological recordings from V1 and MT using both transparent and nontransparent displays (Qian and Andersen, 1994). The results of our computational modeling for explaining the perceptual transparency of various displays are presented in the third article (Qian et al., 1994).

In this article, we describe a series of new patterns that lie between the transparent random dot patterns and the nontransparent counterphase gratings. The main finding is that whenever a pattern has finely balanced opposing motion signals in each local area, it is not perceptually transparent. This leads us to propose that motion transparency could be explained by the presence of locally unbalanced motion signals in different directions. We test the hypothesis by generating what we call paired and unpaired dot patterns. We then introduce displays containing two moving patterns with different disparities or spatial frequency ranges and show that these extra cues facilitate transparent motion perception. These cues are also present in the real world because different objects tend to exist at different depths and/or have different spatial frequency contents.

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

Materials and Methods

The visual displays were generated on an AST 386 PC with a Number Nine SGT graphics board and displayed on a Monitronix video monitor. The graphics board has a resolution of 640 by 480 pixels. The vertical refresh rate is 60 Hz noninterlaced. Each pixel extends a visual angle of 0.028° when viewed at a distance of 114 cm or 0.056° at 57 cm. The monitor was calibrated with an EG&G Gammer Scientific photometer (model 450-1) with capacitors added across the output resistors in order to integrate over the monitor's refresh period to obtain time-averaged readings. In all experiments, a fixation point of 0.1° diameter was always shown at the center of the monitor for the viewer to fixate. For all displays, the directions of motion of the two moving components were always leftward and rightward. The speeds of the two components were equal unless indicated otherwise. If part of a pattern moved off the display window, it was wrapped around and appeared on the other side of the window. The orientation for all gratings and line patterns was vertical. All experiments were conducted in a dimly lit room. In most of our experiments (exceptions will be described in the next paragraph), the patterns were viewed at a distance of 114 cm and each pattern covered a 5° by 5° square display window. In qualitative experiments exploring parameter ranges, each pattern was presented at the center of the monitor for 5 sec. In two-alternative forced choice (2AFC) experiments, two 5° by 5° movies were shown side by side with a horizontal separation of 0.6° between the closest edges. Each display lasted for 1 sec, and the left and right positions of the two patterns were randomized. The fixation point was halfway between the two patterns. The subjects were asked to indicate which one of the patterns looked more like two coherent motion patterns moving in opposite directions.

The experimental settings for testing disparity and spatial frequency specificity were somewhat different from those described above. In the 2AFC experiments, two patterns in each trial were presented sequentially, one after another in random order, with a temporal separation of 0.5 sec. Stereo dot patterns were generated using red and green dots and viewed with red and green filters for the left and the right eyes, respectively. These patterns were viewed at a distance of 57 cm and they covered a square of 5° by 5° at this distance. When testing spatial frequency specificity, gray scale rectangular patterns 9° wide and 2.5° high were generated and viewed at a distance of 114 cm. All the other parameters were the same as stated in the above paragraph.

Results

We will refer to each grating or each set of dots or lines moving in a given direction as a component, rather than a surface, since under some conditions it will not produce the perception of a surface. All our stimuli consisted of two components moving in opposite directions, with each component covering the whole display window. We define motion transparency in this study as the perception of two transparent surfaces moving coherently across each other over the window. We would like to emphasize that all the observations reported in this article were made while subjects fixated a stationary point. The percept of a display under conditions allowing eye movements can often be very different, since subjects tend to track one of the moving components.

We mentioned in the introduction that some displays composed of two components moving in opposite directions give the perception of motion transparency while others appear nontransparent, as exemplified by the random dot patterns and the counterphase gratings, respectively. We first report on more systematic examination of these two patterns. We found that over a wide range of parameters counterphase gratings always appear nontransparent and random dot patterns always appear transparent. We then describe novel patterns created to elucidate the critical factors accounting for transparent motion perception.

Counterphase gratings

None of the counterphase gratings we examined showed motion transparency. For the spatial frequencies (from 0.3 to 8 cycles/degree) and temporal frequencies (from 0.25 to 15 Hz) tested, no transparency could be perceived in the counterphase gratings. One typically saw either periodic oscillation of the whole patterns, or flicker when the temporal frequencies were high. Most

of the gratings had a contrast of 0.5. We also tried other contrast values and obtained the same results.

Random dot patterns

We observed motion transparency for the displays with two sets of random dots moving across each other in opposite directions over a wide range of parameters. The two sets of dots were generated independently according to a uniform probability distribution over the display window. We first fixed the speed of both sets of dots at 2°/sec. For as few as 10 dots in each set, two well-defined transparent surfaces, one for each set of dots, could be clearly seen. When the number of dots in each set was below 5, one tended to see individual dots moving instead of coherent surfaces. We next fixed the number of dots in each set at 50 and varied the speed of the dots. For speeds lower than 10°/sec transparency was seen. For higher speeds the percept was not well defined.

Can differences in the Fourier spectrum account for differences in perceived transparency?

What makes the random dot patterns perceptually transparent and the counterphase gratings nontransparent? We examined the differences between these two types of patterns. An ideal counterphase grating contains only one spatial frequency and two temporal frequencies of equal magnitude and opposite signs, while a random dot pattern with small dot size has a rather broad Fourier spectrum. With this consideration, we generated displays with two square wave gratings to see if an increase of the spatial and temporal frequency contents could make the displays look transparent. Another difference between the random dot patterns and the counterphase gratings is that the former have luminance variations along both spatial dimensions while the latter are one dimensional. This can also be considered as the difference in the richness of the two patterns, spatial frequency contents as their spatial Fourier spectra are one- and two-dimensional, respectively. In order to investigate whether dimensionality is an essential factor for motion transparency, we generated displays with two sets of randomly spaced parallel line patterns, which were essentially the one-dimensional versions of the random dot patterns. We then created displays with two sets of equally spaced parallel line patterns, which were between the square wave gratings and the randomly spaced parallel line patterns in terms of their spatiotemporal frequency contents. All these patterns are shown schematically in Figure 1, and they form a gradual transition from one end of the diagram to the other.
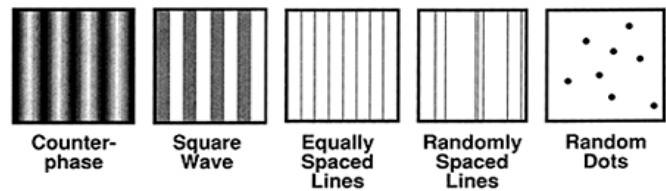
Square wave gratings

The results for displays with two square wave gratings were very similar to those for the counterphase gratings: no motion transparency was perceived when the (fundamental) spatial and temporal frequencies were varied over the same range as for the counterphase gratings discussed above. As with the counterphase gratings, one saw only oscillation and/or flicker.

Randomly spaced parallel line patterns

We saw motion transparency in displays with two sets of randomly spaced parallel line patterns. This was observed over a wide range of parameters, suggesting that the dimensionality of a pattern is not the determining factor for transparency perception. The two sets of vertical lines were independently generated, each with a uniform probability distribution for the horizontal posi-

Patterns Containing 2 Component. Moving in Opposite Directions

Figure 1. Schematic drawings of the family of patterns used in this study to determine the conditions under which transparent motion perception occurs. The patterns are ordered according to the richness of their Fourier spectra. The actual displays contain two components moving across each other in opposite directions.

tion of the lines. We first set the line speed at 2°/sec. When the number of lines in each set was below 25, we saw motion transparency. With the number of lines per set adjusted above 30, the display seemed too crowded and there appeared to be a lot of random motion instead of two coherent transparent surfaces. We next fixed the number of lines in each set at 15 and varied the speed of the motion. For speeds below 4°/sec, we saw transparency; for speeds higher than 6°/sec, we did not. These experiments demonstrate that one-dimensional patterns can be seen as transparent but within a smaller parameter range compared with the results from the two-dimensional random dot patterns discussed above.

Equally spaced parallel line patterns

The patterns used in this set of experiments were the same as those described in the previous paragraph except the spacings between adjacent lines in each component pattern were made equal. We therefore had two identical sets of equally spaced parallel line patterns moving across each other. Interestingly, over a wide range of parameters in which the randomly spaced line patterns appeared transparent, the corresponding equally spaced line patterns showed no motion transparency. In fact, the percept was rather similar to that for the counterphase gratings. When the number of lines per set was 15, no transparency was observed for all line speeds tested, ranging from 0.2° to 8°/sec. At a speed of 2°/sec we saw no clear transparent motion when the number of lines in each set was more than 5. With the number of lines below 5, we could see individual lines moving in both directions when the two sets of interdigitating lines were well separated. As the two sets of lines were about to superimpose, however, the percept of transparent motion was lost. It should be emphasized that the sharp contrast in perceptual transparency between the randomly spaced and the equally spaced line patterns, under identical conditions, is more important than the absolute transparency of individual patterns. As the number of lines was decreased, the equally spaced line patterns did appear slightly transparent.

The lack of motion transparency in the equally spaced line patterns is not due to temporal aliasing. Consider, for example, the pattern with 10 lines moving in each direction at the speed of 2°/sec. Since the pattern is 5° wide, the fundamental spatial period of the pattern (the spacing between two adjacent lines) is 0.5°. For the 60 Hz monitor we used, the spatial displacement

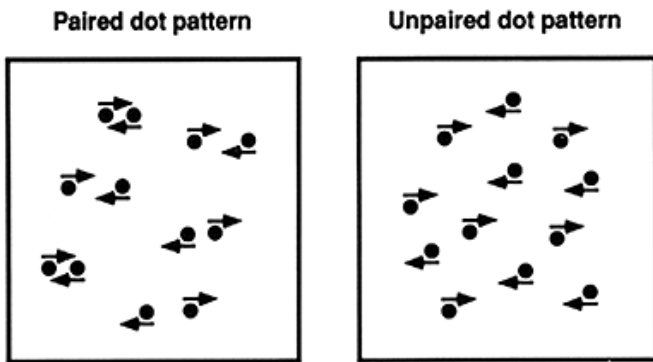**Paired dot pattern**   **Unpaired dot pattern**

Figure 2. Schematic drawings of the paired and the unpaired dot patterns. Note that dot pairs are asynchronized with respect to each other. A typical display contains 50-100 pairs of dots. Only six pairs of dots are shown here.

between two successive frames at the speed of 2°/sec is only 0.03°, much smaller than half of the fundamental spatial period.

Suppression stage hypothesis

To summarize, displays with two sine wave gratings (i.e., counterphase gratings), two square wave gratings, and two sets of equally spaced parallel lines did not give the percept of transparent motion while those with two sets of random dots and randomly spaced parallel lines did over a wide range of parameters. These observations suggest that neither spatiotemporal frequency content nor dimensionality of a pattern is the determining factor of its perceptual transparency.

A simple and biologically plausible explanation of our observations is to assume a suppression stage in the motion pathway at which motion signals in different directions in each small region locally inhibit each other. For the nontransparent patterns (counterphase and square wave gratings, and equally spaced parallel line patterns) with high spatial and temporal frequencies, the motion signals in the two opposite directions were very well balanced at all times, and they would strongly cancel each other at the suppression stage. For those with low spatial and temporal frequencies the motion signals in opposite directions were well balanced at regular time intervals when the two opposite-going component patterns were spatially in phase and this periodic loss of motion signal across the whole pattern may be the basis for the oscillatory perception. For the random dot and the randomly spaced parallel line patterns, on the other hand, there is considerable locally unbalanced unidirectional motion signals across the pattern due to the fluctuations of the local dot or line densities. For example, in a small area, there may be three dots moving in one direction while only one moving in the opposite direction. The cancellation of motion signals in each local region is therefore much weaker and the suppressive stage consequently responds more rigorously. A more quantitative explanation with computer simulations will be given in a following companion article (Qian et al., 1994).

To test further the idea that the perception of motion transparency corresponds to the detection of locally unbalanced motion signals, we designed the paired and the unpaired dot patterns to be described next.
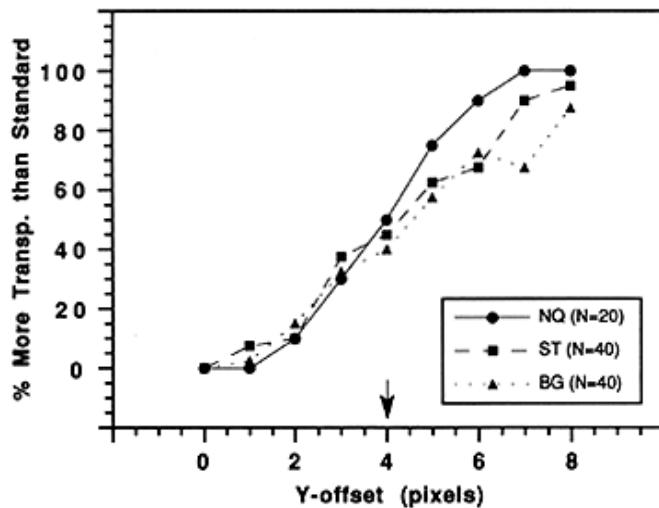
Paired and unpaired dot patterns

We generated paired random dot patterns that consisted of many randomly located pairs of dots (see Fig. 2). The two dots in each

pair moved across each other over a certain distance and then disappeared and reappeared in a new and randomly chosen location. The disappearances and reappearances of dot pairs were asynchronized with respect to each other. We reasoned that if motion transparency can indeed be explained by the detection of the unbalanced motion energies in opposite directions, we should not see transparent motion in a paired dot pattern when the distance over which dots in each pair travel across each other is small. This is indeed what we observed. With 50 such pairs of dots moving at a speed of 2°/sec, we found that no transparency was evident if the dots traveled less than 0.40 (or 14 pixels) on the monitor. This corresponded to a dot lifetime of 240 msec. In fact, the displays look more like flicker. The lack of motion transparency was not because that the dots did not travel far enough for any motion to be detected. If one concentrated on a small area, the opposed motion of both dots in a pair could be easily seen. What was destroyed by the process of pairing was the global coherent motion percept. More importantly, if instead of crossing over each other, the two dots in each pair had a vertical offset of 0.2° or more, or if the dots were simply unpaired by positioning them independently, we saw clear motion transparency. With the distance dots traveled fixed at 0.2°, we saw no transparency when the speed of dots was above 1°/sec. For speeds below 0.5°/sec, some transparency could be seen for a short period of time after the onset of the display, but then disappeared. We also generated paired dot patterns with dark dots moving on bright background as a control for phosphor persistence on the monitor, and similar results were obtained. These experiments show that even two-dimensional patterns can be made nontransparent if the motion signals in the opposite directions are carefully balanced.

To quantify the degree of motion transparency in paired dot patterns, we measured perceptual performance in two tasks using a two-alternative forced choice (2AFC) paradigm. One way to manipulate the amount of perceptual transparency is to vary the vertical offset of the two dots in each pair. We generated nine paired dot patterns with the vertical offset between two dots in each pair equal to 0-8 pixels, respectively (1 pixel corresponds to 0.028° of visual angle at the viewing distance of 114 cm used in this experiment). There were 100 pairs of dots and each dot moved at a speed of 2°/sec over a distance of 0.2°. Each pattern was then displayed beside a standard pattern and subjects were asked to indicate which of the two patterns was more transparent. We used the pattern with a vertical offset of 4 pixels as the standard since it appeared to be halfway between clearly transparent and nontransparent. Figure 3 plots the probability that a particular pattern appeared more transparent than the standard one as a function of the vertical offset of that pattern for three subjects. It is clear from this curve that as the offset between two dots in each pair became larger, and therefore the motion signals became more unbalanced, the pattern looked more transparent.

In the second 2AFC experiment, we generated another series of paired dot patterns, each with a certain percentage of dots unpaired and the rest paired with 0 vertical offset. The unpaired dots were simply positioned independently and randomly over the display window. Again, there were 100 dots moving in each direction and each dot moved at a speed of 2°/sec over a distance of 0.2°. We used the pattern with 50% paired and 50% unpaired dots as the standard for all patterns to compare with. Figure 4 plots the percentage of the time a particular pattern was deemed more transparent than the standard one as a function of the per-
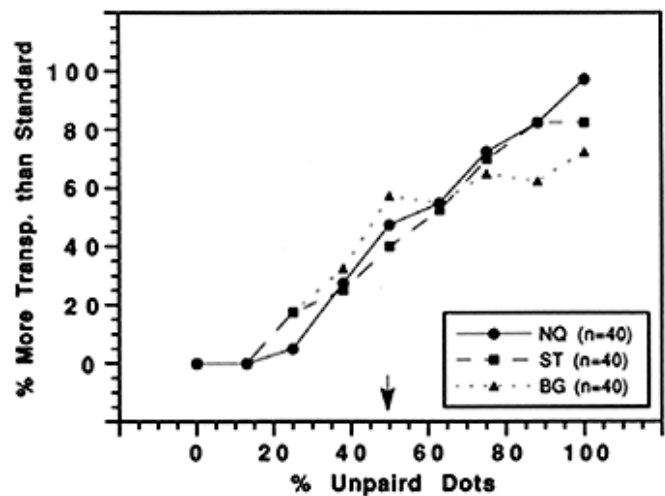
Figure 3. Two-alternative forced choice experiment with different vertical offsets. The three curves are results from three different subjects. We generated nine paired dot patterns with vertical offset between two dots in each pair equal to 0-8 pixels (1 pixel corresponds to 0.028° of visual angle at the viewing distance of 114 cm used in this experiment). The pattern with offset equal to 4 pixels was chosen as the standard for all the other patterns to compare with. The location of the standard pattern on the x-axis is indicated by an arrow. The percentage of times a pattern appeared more transparent than the standard is plotted against the amount of offset in that pattern.

Figure 4. Two-alternative forced choice experiment with different percentage of paired and unpaired dots. The three curves are results from three different subjects. We generated nine paired dot patterns with percentage of unpaired dots equal to 0-100%. The pattern with 50% unpaired dots was chosen as the standard for all the other patterns to compare with. The location of the standard pattern on the x-axis is indicated by an arrow. The percentage of times a pattern appeared more transparent than the standard is plotted against the percentage of unpaired dots in that pattern.

centage of unpaired dots in that pattern. Not surprisingly, the higher the percentage of paired dots in a pattern, the less transparent the pattern looked.

There is an alternative explanation for the lack of motion transparency in the paired dot patterns. It could be argued that at the point of crossing over, the identity of the two dots in each pair is lost. As a result, the two dots in each pair may appear to bounce back rather than move across one another. When this happens the effective displacement of dots is halved and thus the motion signal is weaker, which could account for the lack of transparency. There are several reasons to reject this possibility. First, when a vertical offset of one or two pixels is introduced between the two dots in each pair, and therefore the above miscorrespondence problem is greatly reduced, the patterns are still largely nontransparent. Second, for a given paired dot pattern, if we halve the distance the dots travel and unpair the dots, the resulting pattern looks much more transparent. Finally, we generated paired dot patterns in which two dots in each pair had either different colors (one red, the other green) or different signs of contrast (one bright, the other dark, both moving on a gray background). All dots moving in the same direction were of the same color or same sign of contrast. Again, the mis-correspondence problem was greatly reduced in these patterns. After proper adjustment so that the red and green dots appeared equally bright or that the bright and the dark dots appeared to have equal absolute contrast, the patterns were again nontransparent.

The fact that paired dot patterns with red and green colors can also be made nontransparent by properly adjusting the relative luminances of the two colors deserves some further comments. The result should not be taken as indicating that motion analysis is color blind. It merely suggests that local motion suppression between opposite directions of motion is not color specific; that is the suppression is not restricted within individual color
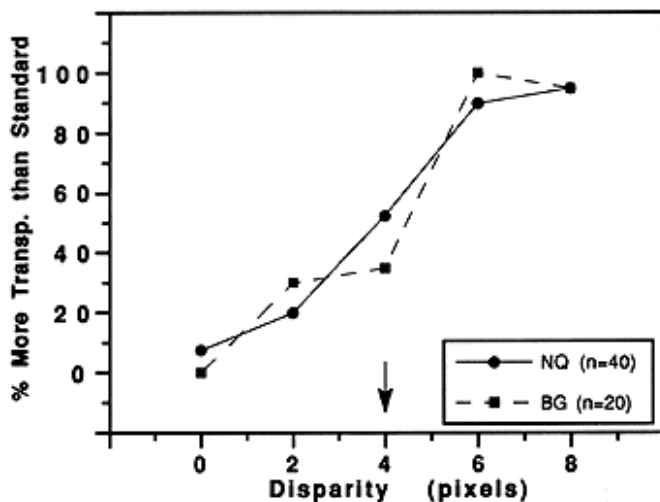
channels. It has been demonstrated convincingly that color information does contribute to motion analysis (Krauskopf and Farell, 1990; Dobkins and Albright, 1993).

Disparity specificity

So far we have considered stimuli with two components of motion at the same disparity, and explained their perceptual transparency by proposing a suppression stage in the motion pathway, in which motion signals of different directions locally inhibit each other. We wondered if the suppression is disparity specific; that is, if the two components of a stimulus are in different depth planes, does the stimulus become perceptually more transparent? We generated stereoscopic paired dot patterns in which two dots in each pair had different binocular disparities and therefore appeared to move in different depth planes. All dots moving in the same direction had the same disparity. When viewed binocularly, such patterns appeared much more transparent than when viewed monocularly. Also, the larger the disparity, the more transparent a pattern appeared to be. This suggests that suppression at the motion opponency stage is strongest among motion detectors tuned to similar disparities.

To quantify this effect, we generated five paired dot patterns with stereo disparities between the two dots in each pair equal to 0, 2, 4, 6, and 8 pixels, respectively (1 pixel corresponds to 0.056° of visual angle at the viewing distance of 57 cm used in this experiment). The fixation point had zero disparity and it lay midway in depth between the two disparity planes defined by the two sets of dots moving in opposite directions. There were 50 pairs of dots and each dot moved at a speed of 3.4°/ sec over a distance of 0.4°. We measured the psychometric curves using a 2AFC paradigm. The pattern with 4 pixel disparity was chosen as the standard and each pattern was compared with it. In each trial, two patterns for comparison were shown one after another in a random order with a 0.5 sec gap between them and the subjects
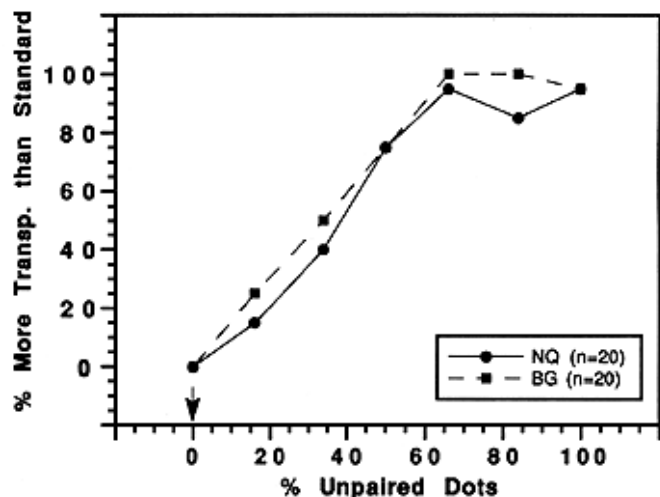
Figure 5. Two-alternative forced choice experiment with different disparity. The two curves are results from two different subjects. We generated five paired dot patterns with stereo disparities between the two dots in each pair equal to 0, 2, 4, 6, and 8 pixels, respectively (1 pixel corresponds to 0.056° of visual angle at the viewing distance of 57 cm used in this experiment). The pattern with 4 pixel disparity was chosen as the standard and each pattern was compared with it. The location of the standard pattern on the x-axis is indicated by an *arrow*. The percentage of times a pattern appeared more transparent than the standard is plotted against the disparity of that pattern.

Figure 6. Two-alternative forced choice experiment with monocular and binocular patterns. Seven monocular patterns with different percentages of paired and unpaired dots were generated by modifying (see Results for details) the corresponding binocular patterns with disparity of 2 pixels (for subject BG) or 4 pixels (for subject NQ). One pixel corresponds to 0.056° of visual angle at the viewing distance of 57 cm used in this experiment. The binocular pattern with 0% unpaired dots was used as the standard for all the monocular patterns to compare with. An arrow was marked at 0 along the x-axis as a reminder of this fact. The percentage of times a monocular pattern appeared more transparent than the binocular standard pattern is plotted against the percentage of unpaired dots in the monocular pattern.

were asked to indicate which pattern appeared more transparent. It can be seen from the results in Figure 5 that increasing the amount of disparity makes a paired dot pattern look more transparent.

In the above experiment, the disparity between two dots in each pair was generated by horizontally displacing the red and green images (for the left and right eyes, respectively) of the two dots in opposite directions. As a consequence, the pairing of dots became less precise for patterns with larger disparities, due to the larger displacements. The decreased degree of pairing with increasing disparity could be argued to account for the increase of perceptual transparency in Figure 5. To be certain that binocular stereo mechanisms really contribute to the emergence of transparency perception, we need to compare the perceptual transparency of a stereo paired dot pattern viewed monocularly with the same pattern viewed binocularly. As it would be inconvenient to ask the subjects to sometimes close one eye during the experiments, we generated the monocular patterns from the corresponding binocular ones by making all the red dots in these patterns invisible and all green dots yellow so that they could be seen through the filters on both eyes.

The following experiment was designed, based on the above consideration, (1) to quantify further the contribution of binocular disparity to transparency perception, and (2) to control for the decreased pairing of dots that results from the introduction of disparity. This was done by comparing the effect of disparity (on transparency) in binocular paired dot patterns with the effect of unpairing dots in the corresponding monocular paired dot patterns. We generated seven monocular patterns with different percentages of paired and unpaired dots similar to what we did in Figure 4. They were derived from the seven corresponding binocular patterns, all having a fixed disparity between the two dots in each pair. We then asked our subjects to compare these monocular patterns with a standard binocular pattern, which was

the same binocular pattern used to derive the monocular pattern with 0% of unpaired dots. In each trial, two patterns for comparison were shown one after another in a random order with a 0.5 sec gap between them, and the subjects were asked to indicate which pattern appeared more transparent.

The results are shown in Figure 6. Due to the different stereo acuity of the two subjects, we used a fixed disparity of 2 pixels for subject BG and 4 pixels for subject NQ (again, 1 pixel subtends 0.056° in these experiments). The leftmost point on the figure for each subject was the result of comparing the monocular and binocular versions of the same pattern. If the stereo cues did not contribute to the perceptual transparency, the point would be around 50% along the y-axis. Instead, the subjects always reported that the binocular versions of the patterns were more transparent than the monocular ones in all trials. By locating the point corresponding to the chance level (50%) in Figure 6, we conclude that the contribution of stereo cues generated by 2 pixel disparity for subject BG and 4 pixel disparity for subject NQ was equivalent to unpairing about 40% of the dots. For subject NQ, 2 pixel disparity was not enough to generate any significant effect while, for subject BG, 4 pixel disparity was so effective that the binocular version of the pattern with all dots paired was more transparent than all the monocular patterns, including the one with all dots unpaired.

Spatial frequency specificity

It is well known that for any direction of motion there are cells or channels tuned to different spatial and/or temporal frequency ranges (Campbell and Robson, 1968; Graham and Nachmias, 1971; Shapley and Lennie, 1985). It is therefore interesting to know if suppression occurs between any two channels with opposite directional preference, or if there is more specificity.

The well-known observation that two oppositely directed sine wave gratings with very different spatial frequencies appear transparent suggests that the suppression is spatial frequency specific. We have generated such displays and found that when the spatial frequencies of the two gratings differ by more than 2 octaves we see transparent motion.

Consider two sine wave patterns with identical amplitude, different spatial frequencies, and with temporal frequencies of equal magnitude and opposite signs (therefore moving in opposite directions). Mathematically, the summation of the two gratings is given by

$$\sin(k_1 x + \omega t) + \sin(k_2 x - \omega t)$$
$$= 2 \sin\left(\frac{k_1 + k_2}{2} x\right) \cos\left(\frac{k_1 - k_2}{2} x - \omega t\right), \quad (1)$$

where k_1 and k_2, and ω and $-\omega$ are the angular spatial and temporal frequencies of the two sine wave gratings, respectively. For $k_1 \neq k_2$, the right-hand side of Equation 1 represents a lower spatial frequency $((k_1 - k_2)/2)$ envelope moving on a stationary higher spatial frequency $((k_1 + k_2)/2)$ background. Perceptually, however, one sees two transparent gratings with spatial frequencies k_1 and k_2 moving across each other if the difference between k_1 and k_2 is more than 2 octaves. This paradox can be explained by assuming that the display is first decomposed into the two terms on the left-hand side of Equation 1 by two separate spatial frequency channels early in the motion pathway and the two channels do not strongly interact with each other at the suppression stage.

We also generated displays composed of two band-limited patterns moving across each other in opposite directions. Each band-limited pattern was a summation of several equal-speed vertical sine wave gratings with their spatial frequencies chosen from a certain range and with arbitrary relative amplitudes and phases among the gratings. We found that transparent motion was clearly observed when the spatial frequency bands of the two patterns were nonoverlapping and when their central frequencies were more than 2 octaves apart, suggesting little suppression between different spatial frequency channels.

To study this perceptual phenomenon more quantitatively, we generated five patterns composed of two sine wave gratings moving across each other in opposite directions. Each pattern had one sine wave grating fixed at the spatial frequency of 0.36 cycle/degree and the other grating had a spatial frequency that was 0.5, 1, 1.5, 2, or 2.5 octaves higher than the first one. All gratings had temporal frequency magnitudes of 2.5 cycles/sec and a contrast of 0.5. The pattern with two gratings 1.5 octaves apart was chosen as the standard for all the patterns to compare with in a 2AFC experiment. The results shown in Figure 7 indicates that the larger the spatial frequency difference between the two gratings in a pattern, the more transparent it appeared.

In the above experiment, we fixed one sine wave component in every pattern at a relatively low spatial frequency and increased the spatial frequency of the other to make the pattern look progressively more transparent. We also generated a complementary set of patterns, each of which contained a fixed sine wave grating with a relatively high spatial frequency and the other with gradually decreasing spatial frequencies. The results (not shown) were the same as above: patterns containing two sine wave gratings with larger spatial frequency difference moving in opposite directions appeared more transparent.

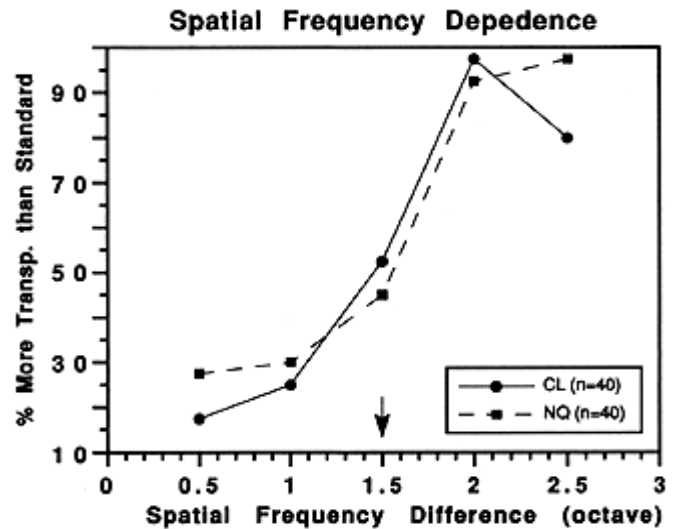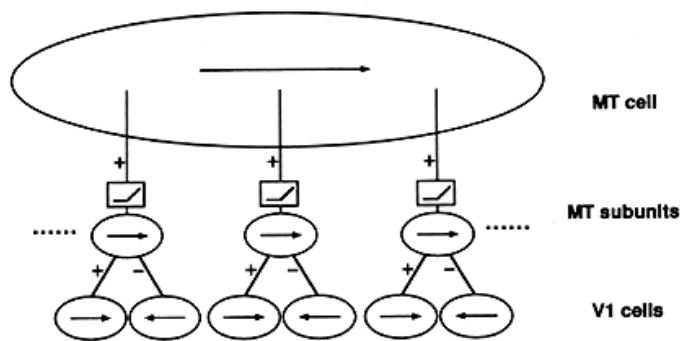We interpret the above results as suggesting a lack of direc-

Figure 7. Two-alternative forced choice experiment with various spatial frequency differences. Five patterns composed of two sine wave gratings moving across each other in opposite directions were generated. Each pattern had one sine wave grating fixed at the spatial frequency of 0.36 cycle/degree and the other grating had spatial frequency that was 0.5, 1, 1.5, 2, and 2.5 octaves higher than the first one, respectively. The pattern with two gratings 1.5 octaves apart was chosen as the standard for all the patterns to compare with. The location of the standard pattern on the x-axis is indicated by an arrow. The percentage of times a pattern appeared more transparent than the standard is plotted against the spatial frequency difference between the two sine wave gratings of that pattern.

tional inhibition between different spatial frequency channels. An alternative explanation is that patterns containing two sine wave gratings with large spatial frequency differences appear more transparent simply because of the presence of these very different spatial frequencies, having nothing to do with the lack of inhibition between them. To rule out this possibility, we generated three patterns all containing the same two spatial frequencies two octaves apart. We will refer to these frequencies as $1f$ and $4f$ for the convenience of the following description. Pattern 1 consisted of a $1f$ sine wave grating moving to the left and a $4f$ grating moving to the right. It was identical to one of the patterns used above in Figure 7. Pattern 2 contained a $1f$ and a $4f$ grating moving to the left, and a $1f$ and a $4f$ grating moving to the right. It is equivalent to an addition of two different counterphase gratings. Since this pattern was a sum of four sine wave gratings (two in each counterphase grating), its mean luminance was two times that of the first pattern. Pattern 3 was identical to pattern 2 except that the mean luminance was halved. We compared these patterns using a 2AFC paradigm and pattern 1 was considered more transparent than patterns 2 and 3 in all 20 trails for both subjects CL and NQ. Although patterns 2 and 3 contained two very different spatial frequencies compared to pattern 1, the motion signals at each of these spatial frequencies were canceled out within each of the counterphase gratings contained in these patterns.

Discussion

The work presented here was inspired by the issues we raised in the introduction of this paper. Briefly, these issues are as follows. (1) Not all displays with two overlapping components moving in opposite directions give a percept of motion transpar-

Figure 8. A schematic drawing of the proposed subunit structure for MT cells. The receptive field of an MT cell is assumed to be composed of many subunits (only three are shown). Motion suppression occurs at the level of MT subunits, each of which receives mutually inhibitory inputs from V1 cells tuned to different directions of motion (only two opposite directions of motion are shown). The overall response of the MT cell is assumed to equal to the sum of the thresholded responses of all of its subunits. *Arrows* in the figure represent preferred directions. The *curves* inside the *rectangular boxes* represent threshold nonlinearity. The *plus* and *minus signs* represent excitatory and inhibitory connections, respectively.

ency. What, then, are the conditions under which transparent motion perception occurs? (2) Pooling procedures commonly used in computer models of motion detection allow only one motion estimation over the area of pooling. Is there a stage in the human visual system that has a similar limitation? (3) Single-unit recordings indicate that MT cells show strong suppression among different directions of motion (Snowden et al., 1991). What is the role of this suppression in transparent motion perception? Since V1 cells respond better to transparent patterns, does this mean that V1 activity forms the basis of transparent motion perception? Although this article is focused on the first question, namely, the conditions for transparent motion perception, the results also shed light on the other two questions.

As we have already mentioned, our results suggest that a suppressive mechanism such as the one found in MT may actually determine the perceptual transparency of a display. Specifically, we found that displays with locally well-balanced motion signals in opposite directions are perceptually nontransparent. These displays presumably maximize the degree of local directional suppression and therefore evoke minimal responses at the suppression stage. The transparent displays we studied, on the otherhand, always contain locally unbalanced motion signals. For example, the unpaired dot patterns are locally unbalanced because of the random nature of dot distributions. The paired dot patterns with a fixed amount of disparity between the two dots in all pairs are also unbalanced because the dots moving in opposite directions excite different disparity channels at the same location. Likewise, stimuli composed of two oppositely directed sine wave gratings with very different spatial frequencies are unbalanced because the two gratings activate different spatial frequency channels. We propose that these displays are perceptually transparent because the unbalanced motion signals generate significant residual responses at the suppression stage. We further propose that the reason those displays with unbalanced disparity or spatial frequency are transparent is that suppression is greater between similar disparity and spatial frequency channels. These ideas will be discussed further in our following articles concerning physiology and modeling (Qian

and Andersen, 1994; Qian et al., 1994). We will also show in the physiology article that although V1 cells respond well to transparent patterns, they also give good responses to nontransparent ones. In fact, the average Vl activity could not reliably distinguish the paired dot patterns from nonpaired ones while the average MT activity could.

Our results with the paired and the unpaired dot patterns indicate that small differences in alignment of the opposing motion on the order of a fraction of a degree can cause major differences in perceived transparency. The two types of patterns can therefore be best distinguished by units of similarly small sizes. The receptive field sizes of MT cells, however, are much too large. If MT is indeed the suppression stage that could differentiate the paired and the unpaired dot patterns, one has to assume that the receptive field of an MT cell is composed of small subunits of the size of a fraction of a degree and that motion suppression occurs at the subunit level. If it is further assumed that the response of each subunit is thresholded and that the total response of an MT cell is determined by the sum of the thresholded outputs of all its subunits, the cell will then respond to the two types of patterns differently. We suggest that the hypothesized MT subunits correspond to V1 inputs to MT and therefore the sizes of subunits should be comparable to V1 receptive field sizes. The idea is illustrated schematically in Figure 8. The subunit assumption is consistent with a recent finding by Shadlen et al. (1993) that directional tuning of MT cells is determined by local motion instead of global apparent motion.

We found that the paired dot patterns without disparity are perceptually nontransparent if the dots travel less than 0.4°, under the foveation condition, and argue that it is a consequence of local suppression among different directions of motion. The result implies that in the absence of any extra cues such as disparity or spatial frequency, the suppressive mechanism only allows one motion to be represented in a small area of about 0.4°. This reminds us of the pooling or regularization procedures used in most computer models, which inevitably combine initial measurements from an area into a single motion estimation. It therefore seems reasonable to identify the proposed suppression stage in the motion pathway with the pooling step in computer models. This identification suggests that the pooling procedure in computer models should be restricted to areas in the range of 0.4°. More importantly, it also suggests that functions of suppression in MT could be similar to the functions of pooling in computer models. These functions include solving the aperture problem and reducing noise. We will discuss these issues in more detail in our following physiology and modeling articles. Finally, the results suggest that transparency can be extracted more efficiently if additional cues are used, such as disparity and spatial frequency, to disambiguate noise stimuli from transparent stimuli. These cues take advantage of the fact that transparent surfaces in the real world generally are at different depths and have different textures.

The lack of transparency in the paired dot patterns (without disparity) also indicates that the motion vectors in different directions cannot be too close to each other in order to generate the percept of motion transparency: when they are too close, they simply cancel each other out due to local suppression or pooling. On the other hand, it is also obvious that these different motion vectors should not be too far apart either, for otherwise one would simply see different motions in spatially separated regions without being able to integrate them into transparent

surfaces over the same region (van Doorn and Koenderink, 1982). Somewhere between these two extreme situations lie the conditions for transparent motion perception. The motion vectors in different directions should be locally separated enough (and therefore unbalanced) to survive the local suppression among different directions of motion, while at the same time globally mixed enough to allow a presumed later stage to integrate the net signals spatially after suppression into coherent transparent surfaces. Clearly, the integration should be carried out for each direction of motion separately in order to obtain multiple motion fields.

The perceptual difference between the paired and the unpaired dot patterns can also be explained from some other perspectives. For example, one might argue that the difference in the statistical regularities of these patterns may be responsible for the difference in their perceptual transparency. Or one could imagine that there might be a grouping process that tends to group together motion elements that are spatially proximal to each other, thus differentiating the paired dot patterns from the unpaired ones. These alternative explanations are not in contradiction with the hypothesis of local motion suppression we propose in this article. Our hypothesis is at the level of neural mechanism, while the above two alternatives are higher-level explanations. In fact, local motion suppression followed by an integration stage (see below) can be viewed as a neural implementation of these higher-level explanations.

The concept of suppression among different directions of motion is not new (see, e.g., Stromeyer et al., 1984), although to our knowledge, its role in motion transparency has not been discussed before. This concept can be traced back to the well-known illusion of motion aftereffect: when a part of visual field is adapted to motion in a given direction, a subsequently presented pattern without net motion appears to move in the opposite direction. The standard explanation for such an illusion is to assume an opponent stage in the motion pathway. The adaptation to motion in a certain direction fatigues input for that direction at the opponent stage and therefore a net response in the opposite direction to a pattern without net motion signal results. It seems reasonable to assume that the suppression stage we propose in this article for determining motion transparency contributes to the motion aftereffect. This assumption, together with our finding in this article that the suppression is disparity and spatial frequency specific, predicts that motion aftereffects should be disparity and spatial frequency specific as well. In fact these experiments have been performed and the results are consistent with the prediction (Regan and Beverley, 1973; Anstis and Hassis, 1974; Cameron et al., 1992).

The problem of motion transparency has been studied previously with plaid patterns composed of two gratings with different orientations and directions of motion (Adelson and Movshon, 1982; Stoner et al., 1990). In these studies the patterns were perceived as either transparent or coherent depending on the parameters chosen. In the present work, we used patterns with either no spatial orientation or a single spatial orientation and the motions in the patterns were always in opposite directions. The problem of motion coherence present in displays with multiple spatial orientations was thus avoided. This approach has the advantage of allowing us to determine the conditions for transparent motion perception without the confounding effect of motion coherence. This simplification helped us to pinpoint local suppression as a potential mechanism for determining the perceptual transparency of a display.

We have considered how spatial frequency and binocular disparity contribute to transparent motion perception and find that differences in these cues between patterns moving in opposite directions increased the perceived motion transparency. We interpret our results by suggesting that the suppression between opposite directions of motion is spatial frequency and disparity specific. It is well known that motion transparency can also be affected by other visual cues present in the world, such as form, color, and static transparency (Krauskopf and Farell, 1990; Stoner et al., 1990; Kooi et al., 1992). For example, Stoner et al. (1990) varied the intersection luminance of square wave plaid patterns to change the degree of static transparency of the overlapping regions and found that transparent motion perception dramatically depends on this manipulation. Our model needs to be extended in order to explain the effect of these other cues. One possible way to explain the observation of Stoner et al. (1990) is to introduce a mechanism that is sensitive to static transparency and to have this mechanism gate the suppressive interaction among local motion measurements.

We also made some observations on the interaction between two directions of motion other than 180° apart. We generated paired dot patterns with the two dots in each pair moving across each other at four different angles (45°, 90°, 135°, and 180°) over short distances. Note that the luminance intersections between different orientations found in the plaid patterns are absent in these dot patterns. Our preliminary results indicate that when the difference of the two directions of motion is 90°, 135°, or 180°, neither motion transparency nor coherence is perceived. If the difference of the direction of motion is 45°, however, we saw a single coherent motion that was the average of the two motion vectors. At all four angles we saw transparency when the dots are not paired. These observations indicate that the inhibitory effect is not restricted to the two opposite directions of motion but extends at least to any two directions that are 90° or more apart. This result agrees with the physiological observations by Snowden et al. (1991), who found that for MT cells, there is significant suppression between two directions of motion 90° apart. When the directional difference is 45° between the two sets of moving dots the inhibition is presumably weak and motion coherence results.

We argue that the proposed suppression stage of the visual system or the pooling operation used in most computer models could not represent more than one motion in a small area, in the absence of other cues such as disparity and spatial frequency. This is because the suppression or the pooling operation combines all initial motion measurements from a local area into a single motion estimation, in order to solve the aperture problem and to reduce noise. Without additional cues other than motion, there is no basis for dividing the initial motion measurements from a location into more than one set and restricting pooling within each set to obtain multiple motion estimations. Of course, as we have already pointed out, multiple motion representation at a given location can still be achieved at a later stage through spatial integration if there are net motion signals along different directions in different but mixed spatial locations at the suppression stage. It is important to note that the above argument applies only to those methods that first measure motion energies along different directions separately and then apply suppression or pooling on these measurements in the subsequent stage without further reference to the original stimulus. While such a two-step model is biologically plausible (Snowden et al., 1991; Qian and Andersen, 1994), it does not

make the best use of a stimulus: one loses information by computing motion energies from a stimulus and then discarding the original stimulus in the subsequent steps. The original stimulus cannot be reconstructed from the energy measures. When there are no other cues available, the suppression or pooling stage has no choice but to group together all energy measurements from a local area into a single estimate.

On the other hand, if one is not constrained by the biologically plausible procedures mentioned above, it is then possible to estimate more than one motion at a single location, in the absence of disparity or spatial frequency cues, as demonstrated recently by Shizawa and Mase (1990). They considered stimuli that can be expressed as summation of two or more brightness distributions, each generated by a moving object. Each distribution is assumed to satisfy the constraint that the total brightness is conserved over time (Horn and Schunck, 1981). They then derived a constraint equation, parameterized by multiple motion velocities, that is satisfied by the stimuli. Multiple motion fields could therefore be recovered from a stimulus by minimizing the violation of the constraint. A heuristic was used in their model to determine the number of motion fields in a given stimulus. The reason that their algorithm could obtain multiple motion at a single location is that they did not first measure the motion signals in different directions separarely. Instead, their method recovers all final motion parameters at the same time directly from the stimuli. Such an approach is nonphysiological since it does not make use of the directionally selective cells in the primary visual cortex, which appear to measure motion energies in different directions instead of coding the final motion velocities (Emerson et al., 1992). Furthermore, we suspect that their model could not explain the difference in perceptual transparency between the paired and the unpaired dot patterns. Both types of patterns can be expressed as summation of two opposite motions. Their method would recover the two motions from these patterns and thus consider both patterns as transparent.

Our psychophysical studies presented in this article suggest that there are two ways of representing transparent motion in the human visual system. When there is no additional cues other than motion, only one motion can be estimated in each small visual area due to the local suppressive mechanisms. In this case, different directions of motion are represented on different spatial locations. These motion signals in different directions are spatially mixed and can therefore be integrated into multiple coherent surfaces. When there are extra cues other than motion available, one can group the initial measurements from each small area according to these cues, and restrict suppression among different directions to be within each group to obtain multiple motion estimations. Movements of objects in the real world are usually not as well balanced as in the paired dot pattern. Furthermore, different objects tend to exist at different depths and have different surface textures. They therefore generate different disparities and usually contain different spatial frequency contents. Both mechanisms can therefore be used for representing multiple motions in the natural world.

References

Adelson EH, Movshon JA (1982) Phenomenal coherence of moving visual patterns. Nature 300:523-525.

Anstis SM, Hassis JP (1974) Movement aftereffects contingent on binocular disparity. Perception 3: 153-168.

Cameron EL, Baker CL, Poulton JC (1992) Spatial frequency selective mechanisms underlying the motion aftereffect. Vision Res 32:561568.

Campbell FW, Robson J (1968) Application of Fourier analysis to the visibility of gratings. J Physiol (Lond) 197:551-566.

Dobkins KR, Albright TD (1993) What happens if it changes color when it moves? Psychophysical experiments on the nature of chromatic input to motion detectors. Vision Res 33:1019-1036.

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.

Graham N, Nachmias J (1971) Detection of gratings patterns containing two spatial frequencies: a comparison of singlechannel and multiple channel models. Vision Res 11:251-259.

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.

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.

Kooi FL, De Valois KK, Switkes E, Grosof DH (1992) Higher-order factors influencing the perception of sliding and coherence of a plaid. Perception 21:583-598.

Krauskopf J, Farell B (1990) Influence of colour on the perception of coherent motion. Nature 348:328-331.

Levinson E, Sekuler R (1975) The independence of channels in human vision selective for direction of movement. J Physiol (Lond) 250: 347-366.

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.

Poggio T, Yang NV, Torre V (1988) Optical flow: computational properties and networks, biological and analog. In: The computing neuron (Durban R, Miall C, Mitcheson G, eds), pp 355-370. Wokingham, UK: Addison-Wesley.

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) Vl responses to two-surface transparent and non-transparent motion. Soc Neurosci Abstr 17:177.

Qian N, Andersen RA, Adelson EH (1994) Transparent motion perception as detection of unbalanced motion signals. III. Modeling. J Neurosci 14:7381-7392.

Regan D, Beverley KI (1973) Disparity detectors in human depth perception: evidence for directional selectivity. Nature 181:877879.

Shadlen MN, Newsome WT, Zohary E, Britten KH (1993) Integration of local motion signals in area MT. Soc Neurosci Abstr 19:1282.

Shapley R, Lennie P (1985) Spatial frequency analysis in the visual system. Annu Rev Neurosci 8:547-583.

Shizawa M, Mase K (1990) Simultaneous multiple optical flow estimation. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Atlantic City.

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.

Stoner GR, Albright TD, Ramachandran VS (1990) Transparency and coherence in human motion perception. Nature 344:153-155.

Stromeyer CF III, Kronauer RE, Madsen JC, Klein SA (1984) Opponent-movement mechanisms in human vision J Opt Soc Am [A] 1 :876-884.

van Doorn AJ, Koenderink JJ (1982) Spatial properties of the visual detectability of moving spatial white noise. Exp Brain Res 45:189195.

Wang HT, Mathur M, Koch C (1989) Computing optical flow in the primate visual system. Neural Comput 1:92-103.

Zucker SW, Iverson L, Hummel RA (1990) Coherent compound motion: corners and nonrigid configurations. Neural Comput 2:44-57.