Vision Res. Vol. 33, No. 3, pp. 313–328, 1993
Printed in Great Britain. All rights reserved

0042-6989/93 $6.00 + 0.00

Surface Segmentation Mechanisms and Motion Perception

JOHN C. TRUESWELL,* MARY M. HAYHOE*

Received 10 January 1992; in revised form 18 June 1992

Two studies are presented which explore how depth information is determined from conditions of transparency and whether this information combines with other depth information to determine the segmenting of motion information on the basis of likely surface boundaries. The first study explored how binocular disparity combines with monocular depth cues associated with transparency to determine whether subjects see one or multiple surfaces in depth in static displays. When transparency provided a depth cue that was consistent with binocular disparity, depth discrimination thresholds were at normal levels. However, if transparency was inconsistent with the binocular disparity, depth discrimination thresholds were elevated, indicating that subjects had difficulty seeing distinct surfaces lying in separate depth planes. Moreover, threshold elevations were found to correspond to the reductions in contrast between the intersecting contours of the stimulus, suggesting that the strength of perceived depth from transparency is the result of attenuated responses from competing contrast sensitive T-junction mechanisms responsible for the detection of opaque occlusion. A second experiment explored whether the grouping of local motion signals relied on surface interpretations that result from the interaction of transparency and disparity. Surface interpretations were manipulated in moving plaids by combining transparent layering and binocular disparity to show that the motion arising from contours is grouped together (pattern motion) when these cues support the existence of a single surface, and is segregated (component motion) when they support separate surfaces. When these cues were consistent, only small disparity differences were required for the gratings to appear as separately moving surfaces. However, when they were inconsistent, greater disparities were required (about a factor of 2 greater). Taken together, these studies demonstrate that the grouping of local motion information is not resolved within the motion system alone. Information seemingly unrelated to motion processing, namely surface segmentation cues, is used to determine whether or not motion information arising from various contours is pooled together to determine a single motion.

Motion Occlusion Transparency Surface interpretation

One of the major tasks of early visual processing is to segment the scene into distinct surfaces lying in three dimensions. For the most part, computational approaches have attempted to understand this problem by first dividing it into smaller, independent problems, such as structure-from-motion and shape-from-shading (e.g. Ullman, 1979; Horn, 1975). Similarly, physiological and psychophysical approaches have supported the early segregation of visual information into non-interactive parallel pathways responsible for features such as motion, color and form (e.g. Livingstone & Hubel, 1987a, b; Van Essen & Maunsell, 1983; Van Essen, 1985; Maunsell & Newsome, 1987). However, as Poggio and others have pointed out, such segregation of information can result in pervasive ambiguity within each subsystem (Poggio, Torre & Koch, 1985; Ullman, 1979). It is often impossible for a subsystem to come up with a single solution to a problem when its input is limited to a restricted stimulus domain. One approach to this problem has been to introduce prior constraints within a domain, such as assuming smoothness constraints on image information. However, an alternative approach for constraining a solution in one domain is to recruit information from another. This takes advantage of the fact that in natural scenes, multiple sources of information are available to determine a perceptual solution which does not rely on possibly incorrect assumptions about the world.

The study of ambiguity resolution in motion processing highlights the distinctions between these approaches. When viewing a moving grating through an aperture, the motion arising from the grating is consistent with many different possible directions and speeds [Fig. 1(A)].

---

*Center for Visual Science, Meliora Hall, University of Rochester, Rochester, NY 14627, U.S.A.

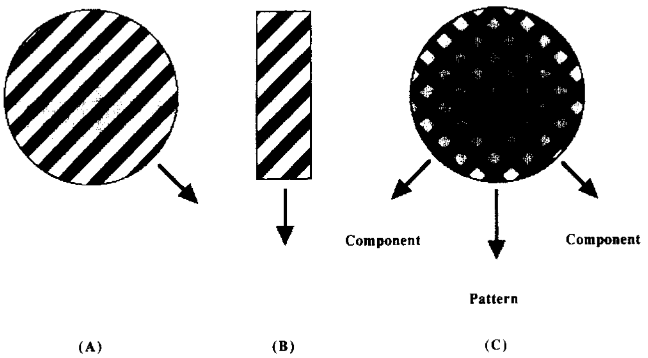

FIGURE 1. The aperture problem. (A) The perceived motion of a grating moving within a circular aperture is typically in a direction orthogonal to the orientation of the grating. (B) The perceived motion of a grating moving within a rectangular aperture is typically in a direction parallel to the longer axis of the aperture (barber-pole illusion). (C) The movement of two overlapping gratings (a plaid) is typically seen as a single unified motion in the direction of the middle arrow (pattern motion). Under certain conditions, the gratings are seen as moving separately, sliding over each other, in the directions of the two outer most arrows (component motion).

Motion is initially encoded in the visual system by an array of motion selective neurons with localized receptive fields, analogous to apertures. Thus, the motion signal arising from each of these elements is also underdetermined (Movshon, Adelson, Gizzi & Newsome, 1985). Although motion in an aperture is ambiguous, a grating moving within a circular aperture appears to move in a direction orthogonal to its orientation, and a grating viewed through a rectangular aperture appears to move in a direction parallel to the long axis of the aperture [the "barber-pole illusion"; Wallach, 1935; see Fig. 1(B)]. Hildreth (1984) and Nakayama and Silverman (1988) provide evidence suggesting that these different solutions arise from constraints imposed by motion detected at the ends of these contours (the contour terminators along the edge of the aperture). The terminators along a circular aperture move on average in a direction orthogonal to the orientation of the grating; whereas the majority of terminators along a rectangular aperture move in a direction parallel to the long axis of the aperture. Indeed, computational models of motion (e.g. Bulthoff, Little & Poggio, 1989; Wang, Marthur & Koch, 1989) have been able to account for aperture effects via smoothing operations in which motion signals arising from terminators propagate towards the interior of the grating.

It seems likely, however, that the resolution process is more complex, and cannot be achieved within the motion system alone (Shimojo, Silverman & Nakayama, 1989). Information unrelated to motion, namely, surface segmentation cues associated with opaque surface occlusion, appear to influence how the visual system treats motion at the ends of contours. Shimojo *et al.* (1989) note that there are two types of contour termination that can occur in a visual image: intrinsic termination, i.e. termination that occurs because the contour actually ends in the world; and extrinsic termination, i.e. termination that occurs because the contour continues behind another surface. They demonstrate that the visual system makes this distinction when determining the direction of motion in the barber-pole illusion. By providing a stereoscopic cue to depth, they show that the effect of the barber pole illusion can be eliminated when the grating is perceived as being *behind* the rectangular aperture. It is argued that, under these conditions, the motion signals from the long edges of the aperture are treated as arising from an extrinsic source and are not used to constrain the perceived motion. When the stripes appear in the same depth plane or a closer depth plane, the motion at the ends of the stripes is viewed as intrinsic to the stripes and is then used to constrain the motion direction, restoring the barber pole illusion. Thus, it appears that the visual system uses depth cues in the stimulus, and real-world constraints concerning depth order arising from occlusion to determine how to treat motion information at a contour termination. Although Shimojo *et al.* (1989) focus on binocular disparity as depth cue, they also provide suggestive evidence that monocular cues to depth (e.g. the existence of T-contour junctions consistent with opaque occlusion) permit distinctions between contour termination. As the authors point out, this suggests that mechanisms responsible for the detection and use of surface segmentation cues related to opaque occlusion may precede those for disambiguating motion signals; raising the possibility that such processes occur relatively early in the visual pathway.

A related paradigm used to study the grouping of local motion information has been to explore motion coherency of two superimposed moving gratings, or plaids. Figure 1(C) presents an example of a plaid stimulus. When viewed separately, each grating appears to move in a direction orthogonal to its orientation. However, when the two gratings are superimposed, viewers typically report motion in a single direction, indicated by the arrow in Fig. 1(C) (i.e. pattern motion) (Adelson & Movshon, 1982; Movshon *et al.*, 1985). It is believed that in a first stage of motion processing (thought to be localized in V1 of the primate visual cortex) the visual system responds to the motion of each component grating, and then in a second stage [thought to be localized in the middle temporal area (MT)] the visual system groups these ambiguous motion signals and determines a single unified direction of motion that satisfies the constraints of these component motions (Adelson & Movshon, 1982; Movshon *et al.*, 1985). Under certain conditions, however, these motion signals are not grouped together, resulting in the percept of two separately sliding gratings (i.e. component motion). For instance, it is well established that component motion occurs when the two gratings have sufficiently different properties, such as differences in spatial frequency, orientation, speed, color, binocular disparity etc. (e.g. Adelson & Movshon, 1982, 1984; Movshon *et al.*, 1985; Krauskopf & Farrell, 1990; Kooi, De Valois & Switkes, 1991; Kin & Wilson, 1992). These effects are typically interpreted as the existence of distinct channels for the motion pathway along these various dimensions. Presumably, these channels assist in determining whether overlapping motion signals arise from the same or different sources, or surfaces (Stoner & Albright, 1993a; Vallortigara & Bressan, 1991).

Like Shimojo *et al.* (1989), Stoner, Albright and Ramachandran, (1990) argue that the grouping of motion signals also arises from mechanisms sensitive to the physical properties related to occlusion. They find that when the luminance relations between two gratings are consistent with that derived from a physically *transparent* grating, motion signals are not grouped together, resulting in component motion. When the luminance relations are *not* consistent with transparency, indicating a single surface, motion signals are grouped together, resulting in pattern motion. Stoner *et al.* (1990) argue that transparency provides a cue to the visual system that the component gratings arise from distinct surfaces, causing the motion signals from each grating to be kept separate. Like Shimojo *et al.*'s (1989) results with opaque occlusion, these results suggest that depth information from transparency is also encoded relatively early in the visual system.

The present paper had two primary goals. The first goal was to explore systematically how various luminance relations in a stimulus are used to determine perceived depth in transparent conditions. It is reasonably clear how depth cues arising from opaque occlusion may be encoded relatively early in the visual system (see below, and Nakayama & Shimojo, 1991). It is less clear, however, how the visual system could rapidly encode depth relations from situations of transparency. In Expt 1, we examine this issue by exploring the interaction between depth cues arising from transparency and another cue to depth, binocular disparity. Related work on depth cue integration has demonstrated that the overall percept of distinct surfaces lying in separate depth planes relies on the combined effects of various perceptual cues to depth and surface boundaries (e.g. Bulthoff & Mallot, 1988; Landy, Maloney & Young, 1991; Maloney & Landy, 1989; Braunstein, Andersen, Rouse & Tittle, 1986; Dosher, Sperling & Wurst, 1986). We demonstrate here that transparency also interacts with disparity information, and that the strength of this interaction corresponds to the contrast relations present at surface boundary intersections. These results suggested that a mechanism sensitive to the constraints of opaque occlusion, similar to that proposed by Nakayama, Shimojo and colleagues, may also provide depth and surface segmentation information in transparent conditions as well. The second goal of this paper was to provide more compelling evidence that the motion effects found in Stoner *et al.* (1990) are indeed the result of depth relations arising from depth cues present in conditions of transparency. In Expt 2 we demonstrate that the perceived depth arising from the *combined* effects of disparity and transparency influences the grouping of motion information in moving plaids. Subjects see component motion when transparency and disparity combine to support the separate gratings in separate depth planes, but see pattern motion when these cues support a single pattern in a one depth plane. These results provide compelling evidence supporting a model of motion grouping which relies on information related to surface segmentation.

EXPERIMENT 1: TRANSPARENCY AS A CUE TO DEPTH

Under certain conditions, perceptual cues related to surface segmentation and depth converge on a solution in which a single image region has a double figural status, i.e. intensities from the same region arise from two separate causes. This perception of transparency can be divided into distinct classes based on the physical conditions under which it may arise. (See Kersten, 1991 for a more complete review.) These classes include: *specular transparency*, arising from surface reflections of "glossy" or "polished" surfaces, such as smooth rock; *diaphanous or sheer transparency*, arising from light passing though a perforated occluder whose holes are below spatial resolution, such as dense foilage; *depth-induced or motion-induced transparency*, which can be best seen when two random-dot stereograms move over each other in different directions.

One of the most frequently studied classes of transparency, and the one of interest here is what might be called *film or filter transparency*. Film transparency typically arises from the presence of a surface that permits some fraction of light to pass through with little or no diffraction, resulting in a darkening of the image beneath it. Figure 2(B, C) provide illustrations consistent with these physical conditions. Present in the illustrations are certain surface segmentation cues which bias a percept of two rectangluar figures on a white background. In both of these figures, the luminance of the overlapping region (R_2) causes one of the rectangles to appear transparent and in front of the other rectangle. In Fig. 2(B) this depth order is ambiguous, causing a bistable percept similar to a Necker cube. In Fig. 2(C), there is an unambiguous depth relation in which only the light gray rectangle appears transparent and in front of the black rectangle.

Perceptual transparency, like that in Fig. 2(B, C), is thought to rely on certain figural and intensity conditions (Gerbino, Stultiens, Troost & de Weert, 1990; Metelli, 1974, 1975; Beck, 1986; Beck & Ivry, 1988). In particular, at least four regions [R_0–R_3 in Fig. 2(A)] are needed to specify transparency. The boundaries of these regions should permit the good continuation of contours across other contours, resulting in some regions having a double figural status (i.e. R_2 belongs both to a rectangle containing R_1 and a rectangle containing R_3). The intensity relations between these four regions should be consistent with the product of both the reflectance and transmittance properties of the physical surfaces corresponding to these regions, as has been specified in models of achromatic transparency, such as Metelli (1974), Beck (1986) and Beck and Ivry (1988). To understand these constraints better, imagine that Fig. 2(B) is a photograph of two rectangular surfaces superimposed on a background (region R_0) whose luminance is 100 cd/m^2. If the rectangles are neutral density

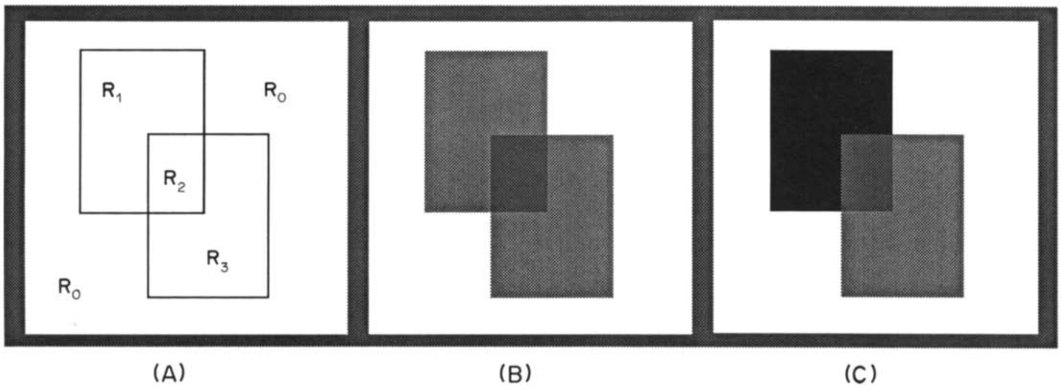

FIGURE 2. Examples of transparency. (A) Two overlapping rectangles with regions marked R_0–R_3. (B) Example of transparency with ambiguous layering. The luminance relations are such that either rectangle can appear in front. This percept is often bistable, much like a Necker-cube. (C) Example of transparency which provides a depth relation. With the luminance of R_1 lowered, the right rectangle appears transparent and in front of the left rectangle.

filters of 50% transmittance, then their luminance would be 50 cd/m^2 (regions R_1 and R_3). However, the luminance of the intersecting region would be 50% of 50 cd/m^2, or 25 cd/m^2 (region R_2). Any intersection luminance above 25 cd/m^2 but below 50 cd/m^2 would still be compatible with the physics of transparency. However, one of the rectangles would be seen as having some surface reflectance of its own.

The first experiment explores how different luminance relations determine whether rectangles like those in Fig. 2 appear to be in separate depth planes. In the study, binocular disparity is provided as a second perceptual cue to depth. If luminance relations consistent with transparency provide a cue to depth, they should interact with disparity to determine the perceived depth between regions. In the study, subjects stereoscopically viewed images similar to those presented in Fig. 2. The disparity between the two rectangles was then manipulated so that one rectangle appeared in front of the other rectangle. Of particular interest were the conditions under which the cues of disparity and transparency were in conflict. Kersten, Bulthoff and Furuya (1989) provided suggestive evidence in a reaction-time study using similar stimuli that it is difficult to perceive depth when transparency and disparity are in conflict.* Based on this, and other studies examining the effect of inconsistent depth cues (Landy *et al.*, 1991; Dosher *et al.*, 1986; Brown & Weistein, 1988; Braunstein *et al.*, 1986), depth discrimination thresholds should be elevated when transparency and disparity support opposite depth relations. Moreover, threshold elevations should provide us with an indirect measure of the strength of the transparency cue for a given luminance relation. A simple model will be presented which uses contrast relations to predict the varying strength of transparency cues to depth. This model demonstrates one way in which the visual system might determine an appropriate depth order from transparency.

Method

Subjects and apparatus. Four subjects participated in the experiment. All had normal stereopsis. The stimuli were generated using a Macintosh 8-bit Video Card (640 × 480 pixels, set to 8-bit pixel depth, 60 Hz non-interlaced) operated in a Macintosh II computer. Stimuli were displayed on a 13-in. AppleColor High Resolution RGB monitor. The viewing distance was 225 cm. A modified Wheatestone stereoscope was used to separate retinal images between the eyes. Each subject was run separately. Before beginning a block, the subject's head was stabilized using a chin rest properly aligned with the stereoscope.

Stimuli and procedure. In the experiment, subjects viewed stereoscopic stimuli consisting of two overlapping rectangles, similar to the stereograms in Fig. 3. Each rectangle subtended 0.95 × 1.43° and appeared within a 2.38 × 2.61° frame. Separate blocks of trials were run for ambiguous and unambiguous transparent layering conditions. From trial to trial the disparity of the right rectangle (the unified regions R_2 and R_3) was set to one of 21 possible disparities, ranging from 5.5 min arc crossed disparities to 5.5 min arc uncrossed disparities. The left rectangle (the unified regions R_1 and R_2) was set to zero disparity, with respect to the rectangular aperture.

To provide disparities that were always consistent with two overlapping rectangles, disparity at the left edge of the intersecting region was always the same as the lower–right rectangle's disparity, whereas disparity at the right edge of the intersecting region was always the same as the upper-left rectangle's disparity (0 min arc). Thus, *uncrossed* disparities were consistent with the left rectangle being closer than the right rectangle, and *crossed* disparities were consistent with the right rectangle being closer than the left rectangle.

For the unambiguous stimuli, the luminance of each region is shown in the first half of Table 1. From trial to trial, the luminance of the intersecting region was set to one of twelve possible luminances, as indicated in the table. The luminance settings for the *ambiguous* transparent conditions is shown in the second half of the

---

*Charles DeWeert (personal communication) has made a similar observation concerning the conflict between transparency and disparity.

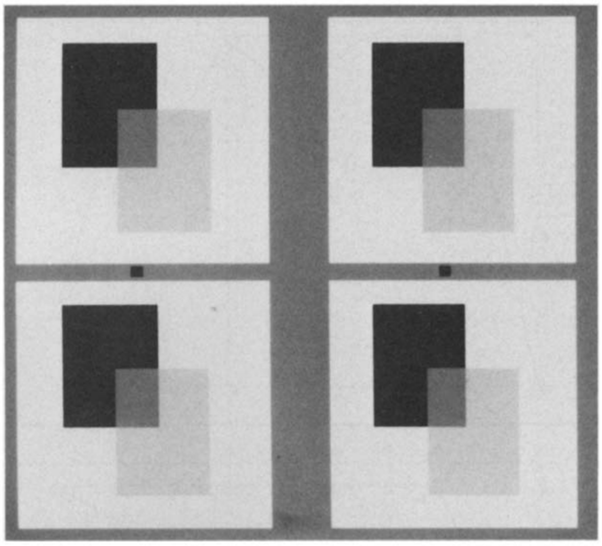

FIGURE 3. Examples of transparency and binocular disparity depth cues. These stimuli can be viewed stereoscopically either by using stereoscopic viewing glasses or by "free-fusing" the images by slightly uncrossing or crossing the eyes. The bottom stereo pair for crossed-fusers and in the top stereo pair for uncrossed-fusers have consistent binocular disparity and transparent depth cues. Monocularly, the light gray rectangle appears transparent and in front of the dark rectangle. Stereoscopically, the binocular disparity cue also supports this depth order. The depth between the rectangles is usually easy to perceive. The top stereo pair for crossed-fusers and the bottom stereo pair for uncrossed-fusers provide an example of inconsistent transparency and binocular disparity cues to depth. Stereoscopically, the binocular disparity cue supports a depth order opposite of the transparent depth cue: the transparent gray rectangle should be behind the dark rectangle. Typically the depth between the rectangles is difficult to perceive or is diminished in magnitude.

table. The luminance of the intersecting region was not manipulated in this condition.

To measure depth discrimination thresholds, subjects pressed a left button if the left rectangle appeared closer or a right button if the right rectangle appeared closer. Subjects viewed the stimuli until a response was made. They were requested to respond as rapidly and as accurately as possible. Subjects were told that under certain conditions, the brightness relations of the surfaces may make it appear as if one surface is in front of the other, even though the disparity cues are in the opposite direction. Subjects were urged to base their judgments concerning perceived depth on the binocular disparity cues alone.

TABLE 1. Luminance of regions in Expt 1

Layering condition	Region	Luminance (cd/m^2)
Unambiguous layering		
	R_0	82
	R_1	1.3
	R_2	0.5–82
	R_3	53
Ambiguous layering		
	R_0	82
	R_1	53
	R_2	35
	R_3	53

RESULTS

Ambiguous transparent layering. For ambiguous transparent layering blocks, subjects accurately judged which rectangle was closer, based on disparity, down to the smallest disparity measured, 0.5 min arc. This result should be expected, given that this transparency cue is ambiguous with respect to depth order. In this case, disparity provides a single cue to depth and is used to disambiguate the two possible layering relationships between the surfaces. This result demonstrates that each subject had normal stereoacuity under these viewing conditions. Furthermore, any possible biases associated with the perceived layering of these Necker-cube-like displays had no effect on the subjects' judgments.

Unambiguous transparent layering. Figure 4 shows depth discrimination thresholds for crossed disparities (open symbols) and uncrossed disparities (solid symbols) as a function of the luminance of the intersecting region (R_2). Each data point represents the smallest binocular disparity at which subjects accurately made a judgment consistent with the disparity cue. Thus, for uncrossed disparities, in which the *left* rectangle was stereoscopically "closer" than the right rectangle, each data point represents the smallest disparity at which subjects judged that the *left* rectangle was in front of the right at least 70% of the time. For crossed disparities, in which the *right* rectangle was stereoscopically "closer" than the left rectangle, each data point represents the smallest disparity at which subjects judged that the *right* rectangle was in front of the left at least 70% of the time.

A marker labeled D on the luminance axis indicates the point at which the luminance of the intersecting region was the same as the dark region, region R_1. At this point the left edge of the intersecting region was not visible, and, monocularly, the display looked like an opaque black rectangle occluding a light gray rectangle (Dark occluding Light). A marker labeled L on this axis indicates the point at which the luminance of the intersecting region was the same as the light gray region, region R_3. At this point the right edge of the intersecting region was not visible, and, monocularly, the display looked like an opaque light gray rectangle occluding a black rectangle (Light occluding Dark). According to models of perceived transparency, the luminances between points D and L should cause the perception of a transparent light gray rectangle occluding a black rectangle. This is indeed the monocular percept when the luminance is noticeably different from points D and L.

Transparent conditions. Depth discrimination thresholds occurring between luminances D and L will be discussed first. Under these conditions, the light gray rectangle appeared to be transparent and in front of the dark rectangle. When binocular disparity supported this same relation (crossed disparities, open triangles), thresholds were at normal levels for all subjects (<0.5 min arc). However, when the disparity information was inconsistent with the transparency information (uncrossed disparities, solid triangles, which support the dark rectangle being *in front of* the light

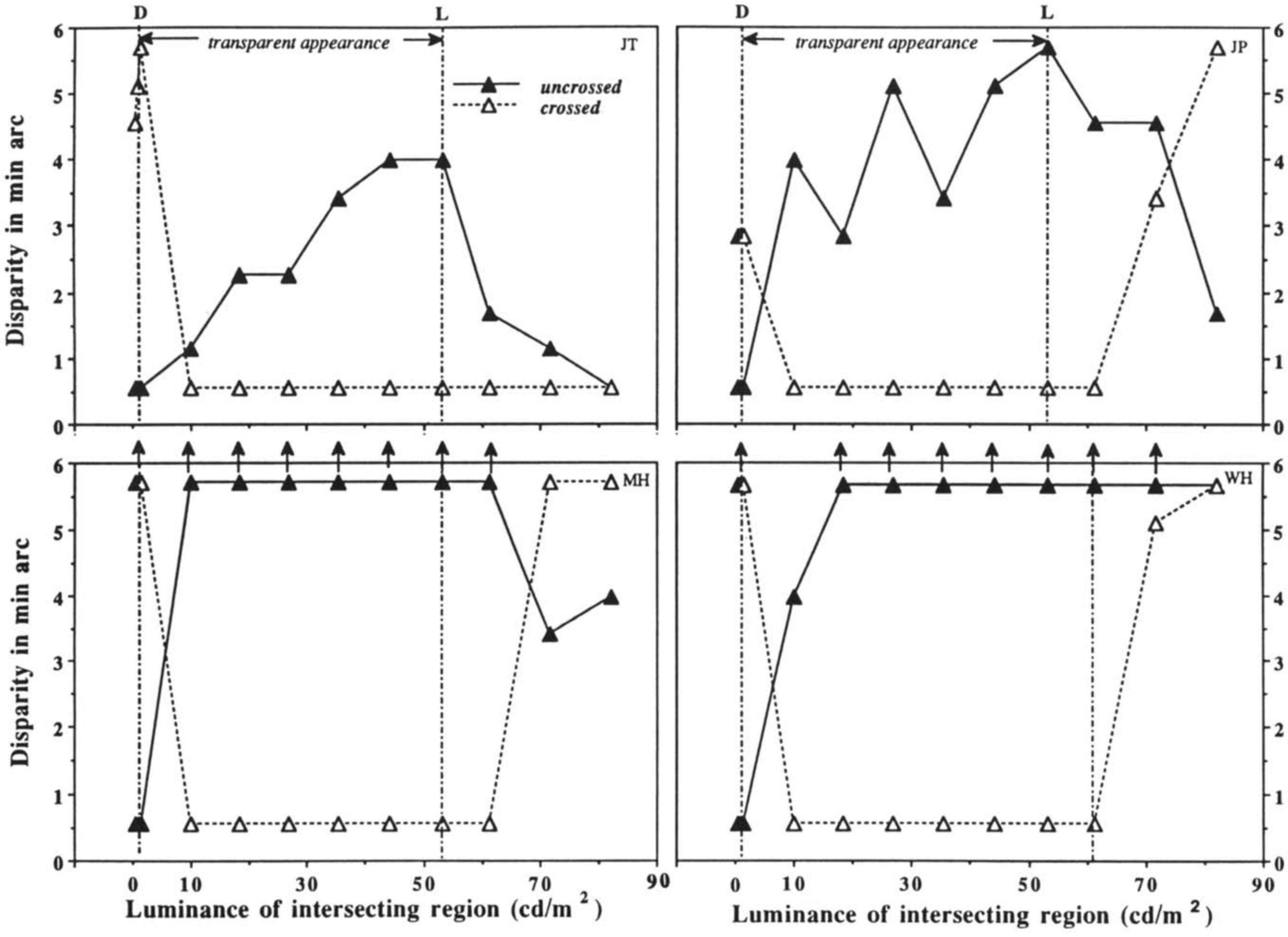

FIGURE 4. Results of Expt 1. Depth discrimination thresholds based on binocular disparity for subjects JT, JP, MH and WH.

rectangle), thresholds were significantly elevated in all subjects.

Opaque conditions. Points D and L mark the opaque occlusion situations. Point D indicates the luminance at which the dark rectangle appears to occlude the light rectangle. When binocular disparity supported this same relation (uncrossed disparities), depth discrimination thresholds were at normal levels for all subjects (<0.5 min arc). However, when the disparity information was inconsistent with the transparency information (crossed disparities), thresholds were significantly elevated in all subjects. Point L marks the luminance at which, the light rectangle appears to occlude opaquely the dark rectangle. Here the situation is reversed, when binocular disparity supported this same relation (crossed disparities), depth discrimination thresholds were at normal levels for all subjects (<0.5 min arc). When the disparity information was inconsistent with the transparency information (uncrossed disparities), thresholds were significantly elevated in all subjects.

Depth discriminations at luminances greater than point L varied somewhat between subjects. These conditions do not provide a clear surface occlusion cue. Thus, the display tends to appear as some kind of rectangular pattern. For three of the four subjects (JT, JP and MH), uncrossed disparity thresholds show a relatively weak decrease as luminance increases. Conversely, crossed disparity thresholds show an increase at the brightest settings for a different group of three subjects (subjects JP, MH and WH).

Discussion

These results clearly indicate that occlusion cues, transparent and opaque, interact with binocular disparity to determine the perceived depth between surfaces. When these cues were in conflict, much greater disparities were required before subjects perceived the depth between surfaces. Furthermore, when these cues supported the same depth relation, thresholds were near normal levels, <0.5 min arc. These results are consistent with other studies concerning the integration of depth cues (Landy *et al.*, 1991; Dosher *et al.*, 1986; Brown & Weistein, 1988; Braunstein *et al.*, 1986; Kersten *et al.*, 1989), which indicate that the perception of depth is reduced and becomes less accurate when cues support opposite depth relations.

Depth cues from transparency. These results hint at a possible mechanism responsible for determining depth from the monocular cues of transparent surfaces. For two subjects, JT and JP, threshold elevations were small enough to determine peaks to their curves. One way of thinking about these elevations is as a measure of the "strength" of the monocular cues to depth order. In both subjects, thresholds peak at the opaque occlusion situations (point D for crossed disparities, and point L for

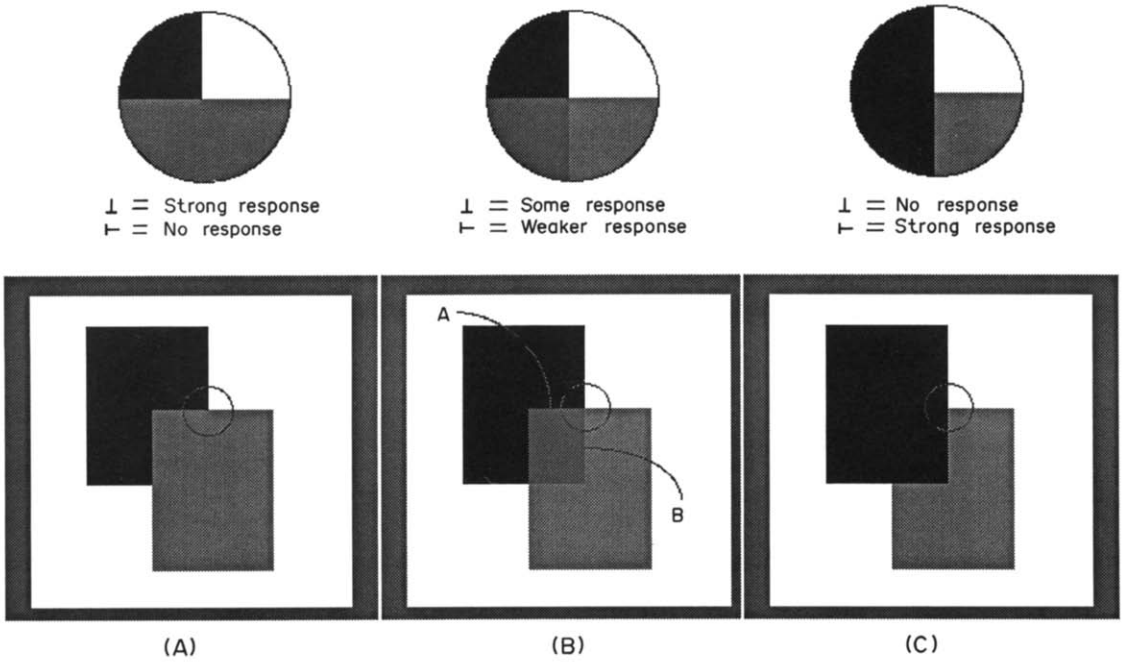

FIGURE 5. Examples of hypothetical T-intersection responses for: (A) a light gray opaque rectangle occluding the dark rectangle, (B) a light gray transparent rectangle occluding the dark, and (C) a dark opaque rectangle occluding the light gray.

uncrossed), suggesting that the strength of perceived depth in transparent conditions may depend on mechanisms sensitive to the monocular depth cues associated with *opaque* occlusion.

As pointed out earlier, the existence of T-contour junctions in a stimulus may provide a monocular cue to depth order (e.g. Helmholtz, 1910; Nakayama & Shimojo, 1991; Shimojo *et al.*, 1989). Typically, a contour that stops at an adjacent contour appears to be in a depth plane behind the other contour. In the presence of other perceptual cues that determine a figure-ground bias, these T junctions may provide evidence of occlusion of one opaque surface by another, with a particular depth order. A strong "response" from such a mechanism would provide a local cue to depth order that may be integrated with other cues, such as binocular disparity.

Consider how this hypothetical T-junction mechanism might respond to contour intersections in our study. Figure 5(A) shows an opaque occlusion situation with a close-up of one of the intersections. Such a mechanism, properly aligned with this intersection, would respond well because the contour clearly stops at another contour—indicating that the horizontal line is in front of the vertical. Now consider the case in which the intersecting region becomes darker [Fig. 5(B), the transparent occlusion situation]. Under these conditions, this same mechanism would be responding in an attenuated fashion, because now the "terminated" vertical contour continues on with a reduced contrast. In addition, an opposing T-intersection mechanism for the perpendicular orientation at this intersection may also start responding (although not very strongly). This mechanism would support the opposite depth relation: namely that the vertical line is front of the horizontal. Since both lines cannot be in front of each other, one might expect the two mechanisms to be involved in some kind of inhibitory competition process. In this case, the stronger of the two T-junction mechanisms should win, which, we will argue in a moment, is the "upside-down-T"-junction mechanism, supporting the horizontal in front of the vertical. As the intersecting region becomes even darker [Fig. 5(C), in which the dark rectangle is an opaque occluder], the "sideways-T"-junction mechanism which was only responding weakly in the transparent case should now respond quite well, whereas the mechanism in the other orientation should not respond at all. Thus, the local cue should now support the vertical being closer than the horizontal.

How well does this proposal fit our data? If a T-intersection mechanism of this type is at work, one would expect that a response from such a mechanism would depend on how much a contour's contrast is diminished as it crosses another, possibly occluding, contour. An estimate can be obtained by comparing the contrasts of the top and right edges of the intersecting region as its luminance changes (e.g. the contours A and B in Fig. 5). As the luminance of the intersecting region changes, the contrast of these two edges change. Because no other regions change in luminance, these two contrasts can be taken as an estimate of how well occlusion sensitive mechanisms might respond at these T intersections. Low contrasts predict a vigorous response from any T-intersection mechanism that has that same orientation. On the other hand, high contrast predicts a poor response. For example, low contrast of the right edge [i.e. Fig. 5(A)] predicts a good response from the "upside-down-T"-mechanism, whereas high contrast [i.e. Fig. 5(C)] predicts a poor response. Therefore, low contrast for the right edge of the intersecting region supports occlusion of the left rectangle, whereas high contrast does not. Conversely, low contrast for the top edge of

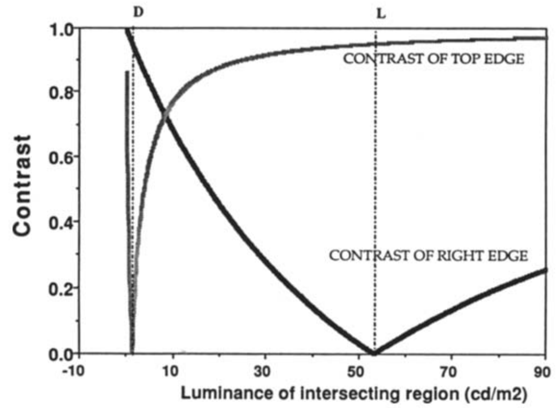

FIGURE 6. Contrast of the right and top edges of the intersecting region as a function of the luminance of the intersecting region.

the intersecting region supports occlusion of the right rectangle, whereas high contrast does not. Figure 6 plots the contrast of these two edges as a function of the luminance of the intersecting region. As expected, the top contour has zero contrast at point D, and the right contour has zero contrast at point L.

It turns out that the difference in contrast between these two contours provides a pattern very similar to our crossed and uncrossed threshold data. Indeed, subtraction of these contrasts is a simple way to model an inhibitory competition process that may occur between possible T-intersection mechanisms. Figure 7(A) shows the top edge subtracted from the right and Fig. 7(B) shows the right edge subtracted from the top. On this scale, values near 1 in Fig. 7(A) support the possibility that the right rectangle is occluding the left, whereas values near 1 in Fig. 7(B) support the opposite relation. Put simply, these data plot the *strength* of the monocular occlusion cue, given these competing T-junction mechanisms. Because these data are being compared with threshold data, differences <0 are plotted as 0 (i.e. thresholds cannot fall below a zero measurement).

Figure 7(C, D) replots the results of subject JT for uncrossed and crossed disparities respectively. When the model predicts that the occlusion cue is strongest for the right rectangle occluding the left, depth discrimination thresholds were elevated when the binocular disparity supported the opposite relation [cf. Fig. 7(A) to JT's uncrossed disparities in Fig. 7(C)]. And, when the model predicts that the occlusion cue is strongest for the left rectangle occluding the right, depth discrimination thresholds were again elevated when the binocular disparity supported the opposite relation [cf. Fig. 7(B) to JT's crossed disparities in Fig. 7(D)].

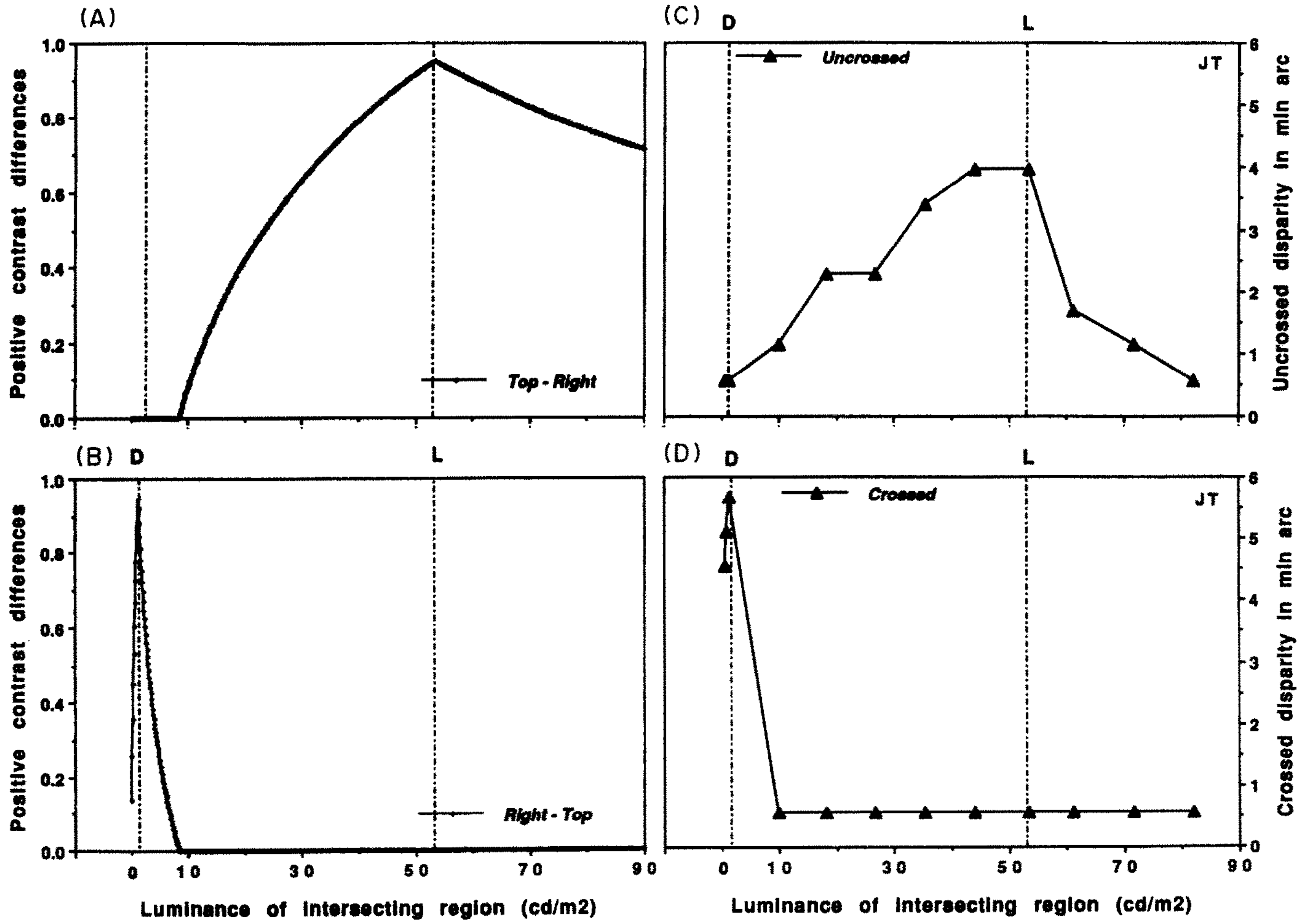

FIGURE 7. Comparison of model and data, as a function of the luminance of the intersecting region. (A) The difference in contrast between the top and right edges of the intersecting region. (B) The difference in contrast between the right and top edges of the intersecting region. (C) Uncrossed disparity thresholds for subject JT. (D) Crossed disparity thresholds for subject JT.

This way of thinking about a monocular depth cue arising from transparency junctions is in many ways complementary to a proposal made by Adelson and Anandan (1990), which suggests a contrast-based classification of "X junctions" [junctions like that in Fig. 5(B)]. The same figural and intensity conditions outlined by Metelli, Beck *et al.* (see Introduction to this experiment) may also be characterized as arising from X-contour junctions in the stimulus (Kersten, 1991; Adelson & Anandan, 1990). Thus, the existence of X-contour junctions with intensity relations that are consistent with the "physical laws of transparency" can be considered a local perceptual cue supporting perceptual transparency. Recently, Adelson and Anandan (1990) have noted that the changes in contrast sign of contours forming an X junction provide a simple heuristic to determine both whether the junction supports perceptual transparency and whether the junction supports a particular depth relation. Returning to the types of transparency shown earlier in Fig. 2, the X junctions in Fig. 2(B) contain contours which *do not* reverse in contrast as they cross each other, supporting the existence of transparency, but no particular depth order. The X junctions in Fig. 2(C) contain one contour that *does not* reverse in contrast and one contour that *does* reverse in contrast, supporting the existence of transparency, but also supporting a particular depth order. X junctions in which both contours reverse contrast sign do not support transparency or a particular depth order. Adelson and Anandan (1990) suggests that the visual system may be sensitive to X junctions as a local surface transparency cue and may also employ these simple heuristics to classify X junctions to determine a particular depth relation. The competitive T junction proposal may be thought of as a possible implementation of the X junction classification described in Adelson and Anandan (1990), since it can appropriately derive depth order in the transparent cases. Adelson and Anandan's (1990) classifications, however, treat X intersections as a separate case from T intersections, suggesting that there are distinct mechanisms for these two cases. The present proposal applies a single mechanism to both T and X junctions.

The only part of our data that the competitive T junction proposal does not capture well is that found for crossed disparities at high intersection luminance levels. Three of the four subjects show elevations in thresholds at this point. At these higher luminances most, if not all models of depth from transparency (except the one proposed here) make no predictions about depth. It is possible that other factors are coming into play at these higher intersection luminances which may influence the figure-ground bias of the intersecting region.

Extrinsic and intrinsic termination in transparency. As Shimojo *et al.* (1989) point out, a T-junction mechanism could be part of a slightly more complex mechanism which would distinguish between extrinsic and intrinsic contour terminators. As mentioned in the Introduction, such a distinction would greatly aid processes related to object recognition and motion processing. We suggest here that the same attenuated and competitively-linked processes outlined above should hold for a terminator classification mechanism when faced with transparent occlusion like that in Fig. 5(B).

Shimojo *et al.* (1989) argue that the distinction between extrinsic and intrinsic terminators depends upon the perceived depth between the two contours in a T junction. If the vertical contour in the T is seen as further away than the horizontal contour, then this is an extrinsic source of termination, and the contour may not actually end in the world (i.e. it may continue behind another surface). If the mere existence of T junction behaves as a monocular depth cue supporting this relation (as Shimojo *et al.* argue), then the "default" is extrinsic termination when faced with a T junction. Other cues to depth (i.e. disparity) also contribute to determine the depth between the contours of the T junction. So, disparity relations will also provide a way to classify termination (Shimojo *et al.*, 1989).

Nakayama and Shimojo (1991), and Shimojo *et al.* (1989) argue that such a classification based on depth information could easily be implemented in the visual system. Cells originally labeled "hypercomplex" by Hubel and Wiesel (1965) and more recently designated as "end stopped," might indicate that a line ended in an image. Such cells respond less vigorously for long lines and fire more vigorously if the line stops in the image plane. However, such cells alone could not signal whether a real line might continue behind an occluder or whether it actually stops. Combining the output of these cells with cells sensitive to depth, however, could resolve the issue for the visual system. Suppose a disparity tuned end-stopped cell responds well to the vertical line of a "right-side-up-T" intersection. If its output was fed to higher order cells which also receive input from cells that encode depth for the horizontal line, such cells could signal whether the a contour termination is intrinsic or extrinsic. If the horizontal contour was coded as closer, then the termination of the vertical contour would be coded as an extrinsic termination.

We suspect that at transparent intersections, hypercomplex cells should respond in an attenuated fashion because their contours continue on with reduced contrast. A mechanism sensitive to T intersections, which receives input from these cells, would presumably also respond in an attenuated fashion. Thus, disregarding any information about depth relations for the moment, an "upside-down T" *extrinsic terminator detector* in Fig. 5(B) should respond more vigorously than a "sideways T" extrinsic terminator detector, since the vertical line has a greater drop off in contrast. It should, therefore, take very little stereoscopic information indicating that the vertical contour is behind the horizontal contour to cause a classification of extrinsic termination for the vertical contour. Conversely, since the "sideways-T" extrinsic terminator detector is not responding very well (and may actually be inhibited), it should require a stronger stereoscopic depth cue supporting occlusion (i.e. the horizontal is behind the vertical) to cause a classification of extrinsic termination for the horizontal contour.* This is consistent with both the asymmetries in disparity thresholds found in our data, and the subjects' resports that it is difficult to arrive a "consistent surface interpretation" when disparity and transparency conflict.

Summary. It is clear from our results that conditions of transparent occlusion provide reliable monocular cues to depth which interact with binocular disparity to determine overall perceived depth. As can be seen in the stereoscopic example of the stimuli (Fig. 3), when binocular disparity and transparency support opposite depth relations, relatively large disparity differences are required to perceive two distinct surfaces lying in separate depth planes. However, when the cues are *not* in conflict, only small disparities are required to perceive two distinct surfaces.

The pattern of disparity thresholds suggest that perceived depth in transparent conditions may arise from the same mechanism responsible for depth from opaque occlusion. We propose that attenuated responses from a hypothetical mechanism sensitive to T-contour junctions in a stimulus may account for threshold elevations in transparent conditions. Thus, local cues at the contour junctions may provide cues to depth order, and possibly to the classification of contour termination as extrinsic or intrinsic. In this way, perceived depth and surface segmentation can be achieved by a common mechanism for both opaque and transparent occlusion of this type.

Finally, it should be made clear that this account of depth under transparent conditions is not intended as a model of how people "perceive" transparent surfaces [see Metelli (1974, 1975), Beck (1986), Beck and Ivry (1988), Gerbino *et al.* (1988) and Kersten (1991) for examples and summaries of such approaches]. Rather, we propose here that information at contour intersections, whether X junctions or T junctions, act as local cues to depth and surface segmentation, and that one possible method of extracting this information is via a single mechanism sensitive T junctions. Indeed, Kersten (1991) provides some striking demonstrations that junctions of this type are neither necessary nor sufficient for the perception of transparent surfaces. Rather, consistent with our results, it is suggested that transparent surfaces arise from the convergence of multiple perceptual cues which support the existence of an image region with double figural status. One cue which can contribute to this interpretation is the existence of X junctions in a stimulus.

EXPERIMENT 2: THE ROLE OF SURFACES IN THE GROUPING OF MOTION SIGNALS

As pointed out earlier, rapid segregation of the visual scene into separate surfaces provides distinct advantages to the visual system for such tasks as motion perception. In particular, motion information arising from various contours needs to be grouped together to determine the overall pattern of motion for an object. However, a scene is usually comprised of multiple moving contours, often belonging to different surfaces. The use of perceptual cues related to surface boundaries in the world could permit the visual system to group local motion signals rapidly and accurately. Shimojo *et al.* (1989) provided evidence in support of this view. The visual system seems to treat the motion arising from contour terminations (intersections) differently depending upon real-world constraints between occlusion and depth order. This suggests that mechanisms sensitive to these constraints may precede those for disambiguating motion signals, raising the possibility that such processes occur relatively early in the visual system.

Stoner *et al.* (1990) provided evidence that depth information from transparency can also be used to influence motion grouping. In their study, subjects viewed plaid stimuli consisting of two overlapping, independently moving, square-wave gratings in which the luminance of the intersecting regions of the two gratings was manipulated. They found that when the luminance of the intersecting regions of the two gratings was consistent with that derived from a physically transparent grating (with bistable ambiguous depth order) the motion of the two gratings was seen as independent, as if two surfaces were sliding over each other (component motion). However, if the luminance of the intersecting region was not consistent with the physics of transparency, the gratings cohered into a unified surface moving in a single direction (pattern motion). The implications for this study are similar to those of Shimojo *et al.* (1989). The visual system appears to be sensitive to perceptual depth cues present in transparent surface occlusion situations, and these cues then contribute to determining the grouping of motion signals. Again, this raises the possibility that these mechanisms may precede those for motion disambiguation.

The results of our first study suggest that a common mechanism may underlie perceived depth from opaque and transparent occlusion of this type, which would lend support to the Stoner *et al.* (1990) proposal for the early extraction of surface boundaries from transparent conditions. However, the Stoner *et al.* (1990) results may be explained without having to recruit additional processes related to transparency (Kim & Wilson, 1992). Rather, the results may arise from a nonlinear component of the motion processing system, such as that independently motivated and proposed by Wilson, Ferrera and Yo (1993). Wilson *et al.* (1993) propose a model of two-dimensional motion perception which incorporates both a simple (linear) motion energy pathway with a "texture boundary motion" (nonlinear) pathway containing a response squaring and before the extraction of motion energy. They suggest that these pathways correspond to Fourier and non-Fourier motion pathways, possibly reflected in the V1-to-MT and V1-to-V2-to-MT

---

*It is likely that this depth cue combination is more complicated than what we describe. For instance, a mechanism sensitive to the presence of T junctions may also depend on other stereoscopic cues related to occlusion, such as the presence of stereoscopically unpaired image points (Shimojo & Nakayama, 1990).

pathways respectively. Indeed, Wilson and Kim (1992) have presented a set of elegant studies using "non-Fourier plaids" which produce counter-intuitive pattern motion perceptions that can easily be accounted for by this two-pathway model. In relation to the Stoner *et al.* (1990) result, Kim and Wilson (1992) argue that the horizontally aligned intersections in the "non-transparent" conditions of the Stoner *et al.* study may result in a strong response in the direction of pattern motion for a nonlinear motion pathway. Thus, when the intersections were considerably darker, or lighter, than the thin stripes, there may have been additional motion in the vertical direction, resulting in the perception of pattern motion. Although Stoner *et al.* (1990) attempt to address a similar argument, they do not take into account adaptational non-linearities which arise early on the visual system (Hayhoe, Levin & Koshel, 1992; Makous, Macleod & Williams, 1985). Such non-linearities would also produce motion energy in the direction of pattern motion. Indeed, even a "non-Fourier" explanation of the Stoner *et al.* results relies on rudimentary surface information—namely the extraction of texture. It is shown here that such an account is not adequate. Rather, our studies indicate that the processing of motion signals relies on more complex surface segmentation mechanisms which extract information relating to both monocular cues to depth, such as opaque and transparent occlusion, and stereoscopic cues, such as binocular disparity.

If the grouping of motion signals in plaids is the result of surface interpretation mechanisms, it should be possible to see the combined effects of transparency and disparity in the perceived motion of plaids. In particular, when transparency and disparity support the same depth relation, the gratings of a plaid should appear as separate surfaces, causing the perception of component motion. However, when these cues conflict, an interpretation in terms of separate surfaces should no longer be possible, and the gratings should cohere into a single surface, causing the perception of pattern motion. Moving plaid stimuli similar to those shown in Fig. 8 were constructed for the experiment. These plaids contain the same binocular disparity and transparent occlusion cues as those in Fig. 3. So, for crossed fusers, the upper stereogram contains inconsistent transparency and disparity cues, whereas the bottom shows *consistent* cues to depth. Again, it should be much easier to perceive distinct surfaces in separate depth planes in the lower stereogram. This can probably be best seen by fusing and fixating on the two dots between the upper and lower stereograms.

There are two possible outcomes one might expect from this experiment. If the surface interpretation mechanisms are *not* guiding motion grouping in plaids, there should be no difference between the percentage of component motion responses for trials that have consistent and inconsistent cues to depth. Thus, there should only be an overall increase in the amount of reported component motion as disparity increased, as has been reported in the early work of Adelson and Movshon (1984). On the other hand, if surface interpretation mechanisms *are* guiding motion grouping, a difference should arise beween the percentage of component motion responses for consistent and inconsistent cue trials. Thus, relatively smaller binocular disparity differences should be required to see component motion when the depth cues are consistent as compared to when they are inconsistent.

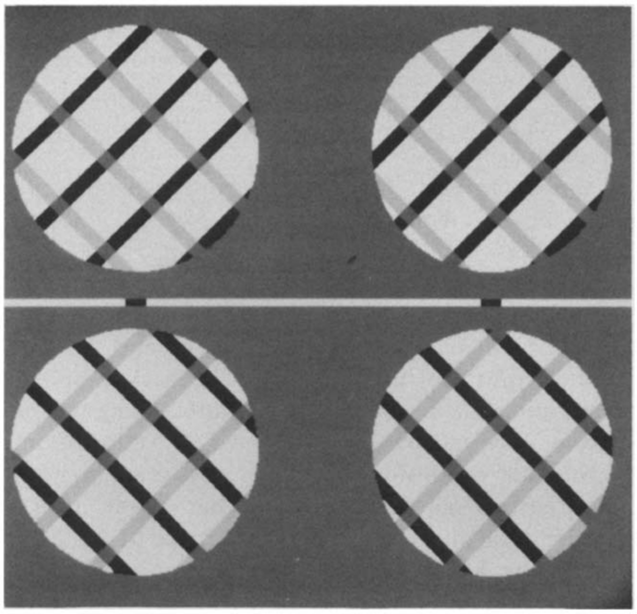

FIGURE 8. Examples of plaid stimuli with transparency and binocular disparity depth cues. The stimuli can be viewed stereoscopically by "free-fusing" the images by slightly crossing the eyes. These plaids contain the same binocular disparity and transparent occlusion cues as those in Fig. 3. In each case, monocularly, the light gray bars of one grating appear transparent and in front of the dark bars of the other grating. In the bottom stereo pair, the binocular disparity relation between the gratings is consistent with the transparent layering cue. Typically, depth is easy to see between the gratings. In the top stereo pair, the binocular disparity relation between the gratings is inconsistent with the transparent layering cue. Typically, is difficult to see the two gratings lying in two distinct depth planes. (Note: this effect is difficult to see when uncross-fusing these images.)

Despite the fact that Stoner *et al.* (1990) find component motion with transparency alone, we expect that, at *zero* disparities, there will be little or no component motion, even with the presence of a transparency cue. The speed and relative orientation of these gratings (described below) were such that they create a strong tendency to see coherent pattern motion. These properties were selected so that effects of disparity could be seen, (i.e. starting at about 0% component motion at zero disparity allows for a larger range of increases in component motion as disparity increases).

Method

Subjects and apparatus. Four subjects with normal stereopsis participated in the experiment. The stimuli were generated using a Macintosh 8-bit Video Card (640 × 480 pixels, set to 2-bit pixel depth, 60 Hz non-interlaced) operated in a Macintosh IIfx computer. Stimuli were displayed on a 13-in. AppleColor High Resolution RGB monitor. The viewing distance was

100 cm, and the aperture diameter was 7 deg. A modified Wheatestone stereoscope was used to separate retinal images between the eyes. Each subjects head was stabilized using a bite bar properly aligned with the stereoscope.

Stimuli. On each trial, the stimulus consisted of a single moving plaid viewed stereoscopically. So, the display was similar to either the upper or lower stereogram in Fig. 5. The display consisted of three regions: (1) the narrow bars of both gratings, which had different luminances, one being 0.1 cd/m^2 and the other being 33 cd/m^2; (2) the wide bars (the background) of both gratings, 79 cd/m^2; and (3) the intersecting regions, 15 cd/m^2. The narrow bar width was 25.5 min arc, and the wide bar width was 83.9 min arc, resulting in a duty cycle [(narrow bar width)/(narrow bar + wide bar)] of 0.23. Thus, on all trials, a transparency cue to depth was present: the light gray bars of one grating appeared transparent and in front of the dark gray bars of the other grating.

On each trial, one component grating was presented with no disparity with respect to the aperture, and the other grating with one of six different uncrossed binocular disparities with respect to the aperture, varying between 0.0 and 26 min arc in increments of 2.6 min arc. The speed of each grating was held constant at 1.8 deg/sec perpendicular to their orientations. The two gratings were perpendicular with respect to each other. Also, the direction of pattern motion was either up or down, and varied randomly from trial to trial. A small fixation dot was presented in the center of the aperture.

In summary, one grating typically had an uncrossed disparity, whereas the other grating had zero disparity (uncrossed disparities usually appear further away than zero disparities). Which of the two gratings appeared transparent and which had an uncrossed disparity varied from trial to trial in random order. It should be pointed out that this design removes the possible confound found in the Stoner *et al.* experiment. In our study, the luminance of the intersecting regions is held constant, only binocular disparity was manipulated. So any "non-Fourier" (or "texture") motion is held constant across conditions.

Only uncrossed disparities were used because crossed disparities cause one grating to be in front of the aperture. Because the aperture occludes the edges of the gratings, it too provides an occlusion cue. We did not want to introduce spurious occlusion cues that were *inconsistent* with disparity, which would have been the case if any grating had a crossed disparity with respect to the aperture. Using various uncrossed disparities caused all occlusion cues from the aperture to be consistent with disparity information across all conditions.

Procedure. Each of the four subjects was run separately on the experiment. At the beginning of each session, the bite bar and stereoscope were alligned. Subjects were instructed to fixate on the dot for the duration of each trial. A trial consisted of a short presentation of the moving plaid stimulus (4 sec), after which the subject had to indicate with a key press whether he or she saw component motion at any time during the presentation. During those times between trials in which the plaid was not being presented, the aperture was displayed, along with the fixation point in front of a white field with a luminance of 79 cd/m^2.

Results and discussion

The results for four subjects are shown in Fig. 9. The stimuli can be split into two groups: those containing consistent cues to depth, and those containing inconsistent cues. Inconsistent stimuli are those in which the lighter "transparent" bars had an uncrossed disparity, whereas consistent stimuli are those in which the darker, partially occluded bars had an uncrossed disparity. Then both cues supported the same depth relation (solid symbols), only small disparity differences were required for the gratings to appear as separately moving surfaces (component motion). Conversely, when the two cues were inconsistent (open symbols), much greater disparities were required. Thus, it appears both the perceived transparent layering order and the binocular disparity between the gratings determines the probability of component motion. This is consistent with a second level of motion processing that determines motion directions by attempting to satisfy multiple constraints relating to surface segmentation and layering cues.

Opaque occlusion. The design of this study contains a possible confound. Consistent stimuli always corresponded to the darker stripes having an uncrossed disparity, whereas inconsistent stimuli always corresponded to the lighter stripes having an uncrossed disparity. It is possible, for instance, that subjects are (for some reason) better able to see stereoscopic depth when the uncrossed disparity is applied darker stripes rather than lighter stripes. To remove this confound, additional blocks of trials were run on the plaid stimuli in which either the light gray stripes appeared as opaque occluders of the dark stripes, or the dark stripes appeared as opaque occluders of the light gray stripes. For trials in which the darker stripes occluded the lighter stripes, the monocular cue to depth was opposite to that present in the transparent conditions. Thus, inconsistent trials now corresponded to the lighter stripes having the uncrossed disparity, rather than the darker stripes (see Fig. 10 for examples of dark bars occluding light bars.) The results were similar to those reported in the transparency conditions: relatively small disparity differences were required for component motion when disparity was consistent with the occlusion cues, whereas much greater disparities were required when occlusion and disparity were in conflict with each other. [For dark gray bars occluding light gray bars, the minimum disparity required for a criterion of 65% component motion was: *consistent* = 6.9 min arc, *inconsistent* = 18.9 min arc; and for light gray bars occluding dark bars it was: *consistent* = 10.3 min arc, *inconsistent* > 26 min arc (the criterion was never met for two subjects and was equal to 20.6 min arc for one subject). These values were

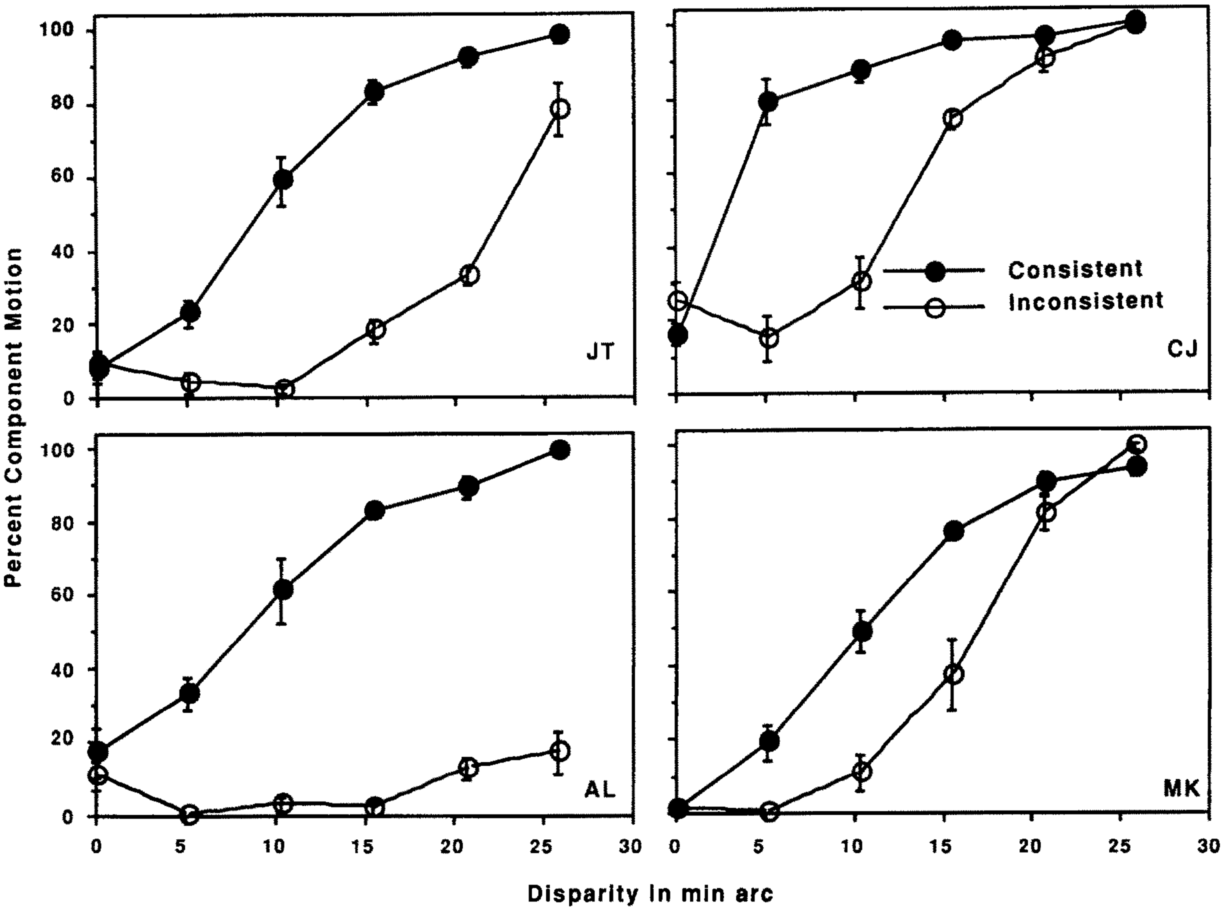

FIGURE 9. Results of Expt 2. The percentage of trials in which subjects JT, CJ, AL and MK reported the percept of component motion.

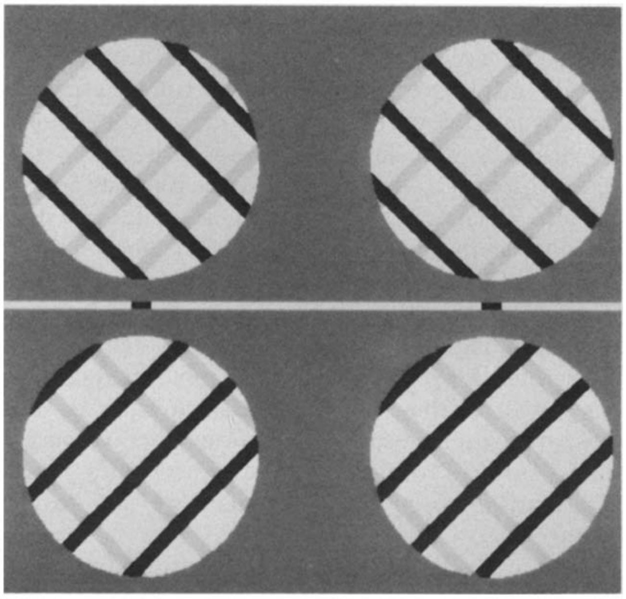

FIGURE 10. Examples of plaid stimuli with *opaque* occlusion and binocular disparity depth cues. The stimuli can be viewed stereoscopically by "free-fusing" the images by slightly crossing the eyes. The dark bars of one grating appear in front of the light bars of the other grating. In the bottom stereo pair, the binocular disparity relation between the gratings is consistent with the occlusion cue. Typically, depth is easy to see between the gratings. In the top stereo pair, the binocular disparity relation between the gratings is inconsistent with the occlusion cue. Typically, is difficult to see the two gratings lying in two distinct depth planes. (Note: this effect is difficult to see when uncross-fusing these images.)

averaged across subjects, JT, CJ and AL.] The data, therefore, do not appear to be sensitive to which grating has uncrossed disparity, but rather to the consistency between surface cues and disparity.

Surface appearance in plaids. To test directly whether these transparency effects were result of subjects seeing single or multiple surfaces, an additional block of trials was run on two of the subjects. In these blocks the plaid were shown *stationary* for 4 sec. The same design and methods were used as above in the transparent occlusion blocks, except not the subject's task was to press one button if the two component gratings appeared as a single surface in a single depth plane and another button if they appeared in separate depth planes. The results for JT and AL are shown in Fig. 11. The figure graphs the percentage of trials under which the subject saw two separate depth planes as a function of binocular disparity, for consistent and inconsistent cue trials. The results are almost identical to the component motion results. Much greater disparities were required to see two depth planes when the cues were inconsistent as compared to when they were inconsistent. In addition, individual differences between JT and AL remained the same in this judgment data: namely, subject AL had more difficulty overriding the transparency cue in the inconsistent trials as compared to subject JT.

Surprisingly, much greater disparities were required to see separate depth planes in this static display

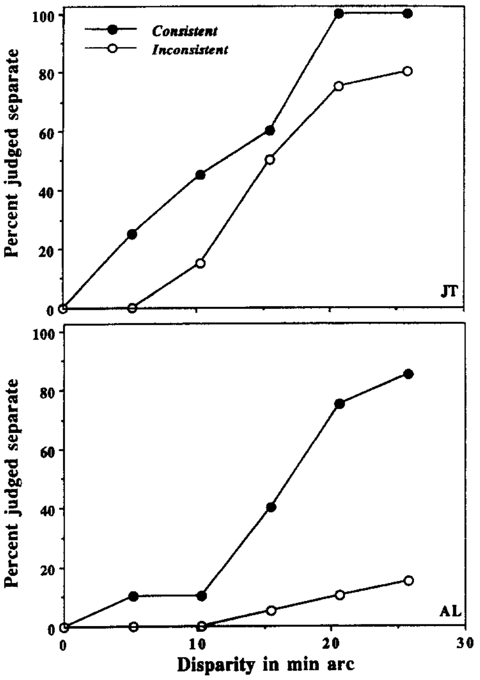

FIGURE 11. Results of depth judgment task on transparent plaids for subjects JT and AL.

experiment as compared to Expt 1. There are several likely reasons. First, in this experiment the stimuli were much more complex patterns. It is likely that having multiple occlusion cues (one for each intersection) would elevate effects with the inconsistent disparities as compared to the first experiment. This does not explain differences in the consistent disparity trials though. Much greater disparities were required to see separate surfaces in this study as compared to the other. One likely reason for this is that manipulating disparity for these gratings introduced vertical disparity differences at the intersecting regions. This occurred because the gratings were at 45 deg with respect to horizontal. To correct this, additional trials were run in which the gratings were horizontal and vertical with respect to the eyes. Thus, horizontal gratings had zero disparity, whereas vertical gratings had some uncrossed disparity. Again, the design was the same. The overall effects for both component motion and depth discrimination were the same, except the differences between inconsistent and consistent cue trials occurred at smaller disparities. Inconsistent cue trials were less likely to cause component motion or appear as separate depth planes, whereas consistent cue trials saw a more rapid increase in component motion and depth discrimination [more than 70% of the trials were judged as separate depth planes at the second smallest disparity measured (2.5 min arc)]. Thus, the vertical disparities caused only an overall decrease in perceived depth, and had little effect on the interaction between depth cues. Further studies will be required to explain remaining differences between these studies.

Other studies on occlusion and plaids. Recently, Vallortigara and Bressan (1992) have presented a series of suggestive demonstrations and experiments which lend support to the view that segmentation cues arising from depth and occlusion influence motion grouping. In one study in particular, they provide evidence suggesting that other cues to depth (e.g. relative contrast) may interact with occlusion cues to determine component motion in plaids. Using plaid stimuli similar to our own, they report that if a high contrast squarewave grating (black stripes) is placed in front of a lower contrast squarewave grating (gray stripes), via an opaque occlusion cue, subjects tend to see component motion more often than if the low contrast grating is placed in front of the high contrast grating. The authors interpret this asymmetry as being the result of cue integration processes related to depth from contrast [e.g. objects that contrast more with a background tend to be seen as closer in depth than objects with lesser contrast (Egusa, 1982)]. In other words, when relative contrast and occlusion support the same depth relation, subjects tend to see component motion. Indeed, our additional results with opaque occlusion cues also show this asymmetry with occlusion (see threshold values in opaque occlusion in the previous section). Thus, consistent with our own studies, Vallortigara and Bressan's (1992) results suggest that the combined effects of depth cues influence motion grouping in plaids. A simpler model of motion grouping based only on the existence of distinct channels in the motion pathway will have difficulty accounting for these interactions.

GENERAL DISCUSSION AND SUMMARY

Our results clearly demonstrate that the grouping of various motion signals depends on surface segmentation cues related to occlusion. The first study addressed how transparent occlusion and binocular disparity combine in *static* displays to determine when subjects perceive single or multiple surfaces. When transparent occlusion and disparity were in conflict with each other, much greater disparities were required to see separate surfaces in depth. When occlusion cues supported the same depth relation, depth discrimination was at normal levels. Experiment 2 demonstrated that the grouping of motion information was determined by the surface interpretations resulting from these cues. Motion signals tended to be grouped together when transparency and disparity cues supported the existence of a single surface. When transparency and disparity supported the existence of distinct surfaces lying in separate depth planes, motion signals arising from those regions were not grouped together.

These results have important implications for various approaches to understanding motion perception. With few exceptions, these approaches have supported a second stage of motion processing that is relatively independent of other visual information (Movshon *et al.*, 1985; Hildreth & Koch, 1987). However, this leads to an indeterminacy concerning how motion signals should be grouped. One solution to such ambiguity is to recruit information from other sources in hopes of constraining the possible solutions. Our results are an example of such a process: information not related to motion processing, namely surface segmentation cues related to perceive depth, is used to disambiguate motion signals.

The results of Kersten, Bulthoff, Schwartz and Kurtz (1993) are consistent with this finding. They have found that transparency also influences the interpretation of structure-from-motion. In their studies, subjects viewed two overlapping rectangles that appeared to rock back and forth in depth. Typically, subjects viewing these kinds of displays perceive rigid motion consisting of two parallel planes rotating about a single axis. However, Kersten *et al.* (1993) show that depth from transparency and opacity can override the bias to see rigid motion. When occlusion cues were present that were inconsistent with relative depth required for rigid motion, subjects reported a strong tendency to see non-rigid motion. Such a result again supports the hypothesis that the grouping of motion information relies heavily on surface segmentation cues such as occlusion. Indeed, it may be useful to treat plaids as a simple case of structure-from-motion, in which the surface structure is two parallel planes. Subjects tend to see the less complex percept of a single surface in a single depth plane, but when surface cues converge on the existence of two separate planes, motion information is segregated and component motion is perceived.

Given that surface occlusion cues influence the grouping of local motion information, at what point in processing does this information converge? It is possible that information about occlusion is derived relatively late in visual processing from complex surface interpretations. This information could be fed back to early motion grouping mechanisms, via neural back-projections, to control which local motion signals are grouped together. An alternative, and perhaps more likely solution, would be for the visual system to derive surface groupings relatively early on in processing. Indeed, as has been mentioned earlier, surface segmentation may begin its implementation at very early stages of cortical visual processing (as early as V1 and V2) via mechanisms sensitive to the physical constraints of occluding and non-occluding edges. Results from the first experiment provided evidence for a single mechanism responsible for depth from both transparent and opaque occlusion. In this model, mechanisms sensitive to opaque occlusion (as described in Shimojo *et al.*, 1989; Nakayama & Shimojo, 1991) may respond in an attenuated, but competitive, process to determine depth order in transparent situations. Furthermore, such a mechanism could be implemented using the output of binocular disparity-tuned cells and end-stopped cells both known to exist area V1 (e.g. Hubel & Wiesel, 1965).

The outputs of such a surface segmentation mechanism could be used to group local motion information. Interestingly, recent models of the physiological pathways of vision indicate two routes projecting to area MT, a site known to process complex motion (DeYoe & Van Essen, 1985; Van Essen, 1985). In addition to a direct route from area V1 to area MT, there are indirect projections via area V2. Indeed, a recent physiological model of two-dimensional motion (Wilson *et al.*, 1993) suggests that this indirect route may be responsible for a second, nonlinear, processing of motion information, to account for results from non-Fourier motion. It is possible that this route may actually be performing a more complex function of early and rapid surface segmentation, and that this information is being fed to areas known to be responsible for motion grouping. Consistent with this, recent physiological evidence indicates that conditions of transparency appear to influence pattern motion cell responses in area MT (Stoner & Albright, 1993a, b).

REFERENCES

Adelson, E. H. & Anandan, P. (1990). Ordinal characteristics of transparency. *Proceedings of the AAAI workshop on qualitative vision* (pp. 77–81). Boston, Mass., July 1990.

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

Adelson, E. H. & Movshon, J. A. (1984). Binocular disparity and the computation of two-dimensional motion. *Optical Society of America, Annual Technical Meeting Program, 56.*

Beck, J. (1986). Perception of transparency in man and machine. In Rosenfeld, A. (Ed.), *Human machine vision II.* London: Academic Press.

Beck, J. & Ivry, R. (1988). On the role of figural organization in perceptual transparency. *Perception and Psychophysics, 44,* 585–594.

Braunstein, M. L., Andersen, G. J., Rouse, M. W. & Tittle, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception and Psychophysics, 40,* 216–224.

Brown, J. M. & Weistein, N. (1988). A spatial frequency effect on perceived depth. *Perception and Psychophysics, 44,* 157–166.

Bulthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America, 5,* 1749–1758.

Bulthoff, H. H., Little, J. & Poggio, T. (1989). A parallel algorithm for real-time computation of optical flow. *Nature, 337,* 549–553.

DeYoe, E. A. & Van Essen, D. C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature, 317,* 58–61.

Dosher, B. A., Sperling, G. & Wurst, S. A. (1986). Trade-offs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990.

Egusa, H. (1982). Effect of brightness on perceived distance as a figure-ground phenomenon. *Perception, 11,* 671–676.

Gerbino, W., Stultiens, C. I. F. H. J., Troost, J. M. & de Weert, C. M. M. (1990). Transparent layer constance. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 3–20.

Hayhoe, M. M., Levin, M. E. & Koshel, R. J. (1992). Subtractive processes in light adaptation. *Vision Research, 32,* 323–333.

Helmholtz, H. (1910). *Treatise on physiological optics.* (Vol. 3). [Southall, J. P. C. (Ed.), translated from the third German edn; reprinted in 1962.] New York: Dover.

Hildreth, E. C. (1984). *The measurement of visual motion.* Cambridge, Mass.: MIT Press.

Hildreth, E. C. & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience, 10,* 477–533.

Horn, B. (1975). Obtaining shape from shading information. In Winston, P. H. (Ed.), *The psychology of computer vision.* New York; McGraw–Hill.

Hubel, D. H. & Wiesel, T. N. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28,* 229.

Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing.* Cambridge, Mass.: MIT Press.

Kersten, D., Bulthoff, H. H. & Furuya, M. (1989). Apparent opacity affects perception of structure from motion and stereo. *Investigative Ophthalmology and Visual Science (Suppl.)*, *30*, 264.

Kersten, D., Bulthoff, H. H., Schwartz, B. & Kurtz, K. J. (1993). Interaction between transparency and structure from motion. *Neural Computation*. In press.

Kim, J. & Wilson, H. R. (1992). Interaction of different spatial frequency scales in motion processing. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *33*, 4.

Kooi, F., De Valois, K. K. & Switkes, E. (1991). Higher order factors influencing the perception of sliding and coherence of a plaid. In *From Pigments to perception: Advances in understanding visual processes*. New York: Plenum Press.

Krauskopf, J. & Farrell, B. (1990). Influence of color on the perception of coherent motion. *Nature*, *348*, 328–331.

Landy, M. S., Maloney, L. T. & Young, M. J. (1991). Psychophysical estimation of the human depth combination rule. *Proceedings of the SPIE: Sensor Fusion III, 3-D Perception and Recognition*, *1383*, 247–254.

Livingstone, M. S. & Hubel, D. H. (1987a). Connections between layer 4b of area 17 and thick cytochrome oxidase stripes of area 18 in the squirrel monkey. *Journal of Neuroscience*, *7*, 3371–3377.

Livingstone, M. S. & Hubel, D. H. (1987b). Psychophysical evidence for separate channels for the perception of form, color, movement and depth. *Journal of Neuroscience*, *7*, 3416–3468.

Makous, W., Macleod, D. I. A. & Williams, D. R. (1985). Nonlinear transformation in human vision. *Journal of the Optical Society of America A*, *32*, 80.

Maloney, L. T. & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. *Proceedings of the SPIE: Visual Communications and Image Processing*, *1199*, 1154–1163.

Maunsell, J. H. R. & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*, 363–401.

Metelli, F. (1974). The perception of transparency. *Scientific American*, *230*, 91–95.

Metelli, F. (1975). On the visual perception of transparency. In Flores d'Arcais, G. B. (Ed.), *Studies in perception: Festschrift for Fabio Metelli*. Milano, Italy: Martello-Giunti.

Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1985). The analysis of moving visual patterns. In Chagas, C., Gattas, R. & Gross, C. (Eds), *Pattern recognition mechanisms*. New York: Springer.

Nakayama, K. & Shimojo, S. (1991). Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, *55*, 911–924.

Nakayama, K. & Silverman, G. H. (1988). The aperture problem II: Spatial integration of velocity information along contours. *Vision Research*, *28*, 747–753.

Poggio, T., Torre, V. & Koch, C. (1985) Visual perception and regularization theory. *Nature*, *317*, 314–319.

Shimojo, S. & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research*, *30*, 69–80.

Shimojo, S., Silverman, G. H. & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, *29*, 619–626.

Stoner, G. R. & Albright, T. D. (1993a). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*. Submitted.

Stoner, G. R. & Albright, T. D. (1993b). Neural correlates of the perceptual coherence of visual motion signals. *Nature*. In press.

Stoner, G. R., Albright, T. D. & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature*, *344*, 153–155.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London B*, *203*, 405–426.

Vallortigara, G. & Bressan, P. (1991). Occlusion and the perception of coherent motion. *Vision Research*, *31*, 1967–1978.

Van Essen, D. C. (1985). Functional organization of primate visual cortex. *Cerebral Cortex*, *3*, 259–329.

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

Wallach, H. (1935). Uber visuell wahrgenommenr Bewegungsrichtung. *Psychologische Forschung*, *20*, 325–380.

Wang, H. T., Marthur, B. & Koch, C. (1989). Computing optical flow in the primate visual system. *Neural Computation*, *1*, 92–103.

Wilson, H. R. & Kim, J. (1992). Non-fourier plaids move in the vector sum direction. *Investigative Ophthalmology and Visual Science (Suppl.)*, *33*, 4.

Wilson, H. R., Ferrera, V. & Yo, C. (1993). A psychophysically motivated model of two-dimensional motion perception. *Visual Neuroscience*. In press.

Acknowledgement—This research was supported by NIH grants EY05729 and EY01319.