Interaction between Transparency and Structure from Motion

Daniel Kersten
Department of Psychology, University of Minnesota,
Minneapolis, MN 55455 USA

Heinrich H. Bülthoff
Bennett L. Schwartz
Kenneth J. Kurtz
Department of Cognitive and Linguistic Sciences, Brown University,
Providence, RI 02912 USA

It is well known that the human visual system can reconstruct depth from simple random-dot displays given binocular disparity or motion information. This fact has lent support to the notion that stereo and structure from motion systems rely on low-level primitives derived from image intensities. In contrast, the judgment of surface transparency is often considered to be a higher-level visual process that, in addition to pictorial cues, utilizes stereo and motion information to separate the transparent from the opaque parts. We describe a new illusion and present psychophysical results that question this sequential view by showing that depth from transparency and opacity can override the bias to see rigid motion. The brain's computation of transparency may involve a two-way interaction with the computation of structure from motion.

1 Introduction

One of the major challenges of vision research is to understand how the brain constructs a model of the visual environment from the pattern of changing retinal light intensities. With relatively few exceptions (Poggio *et al.* 1988; Barrow and Tenenbaum 1978), computational research has sought to first divide the problem into noninteracting modules such as surface color from radiance, shape from shading, or structure from motion (Land 1983; Horn and Brooks 1989; Ullman 1979; Martin and Aggarwal 1988). Consistent with the methodology of computer vision, current physiological and psychophysical research indicates modular and concurrent processing for some sources, such as motion, or form and color (Livingstone and Hubel 1987; Cavanagh 1987; Zeki 1978; Van Essen 1985).

In contrast to the modularity of vision research, it is phenomenally apparent that visual information is eventually integrated to provide a strikingly singular description of the visual environment. The simple unity of visual experience belies some difficult problems of neural computation. One problem is *cue integration*—the combination (possibly linear) of visual information from multiple sources, such as stereo and motion, to compute a single attribute such as depth. A second and theoretically more difficult problem is the *cooperative coupling* (typically nonlinear) of the perceptual representations of two or more scene attributes (such as surface depth and material property) to achieve the consistency required by the laws of image formation (Kersten 1991; Bülthoff 1991; Bülthoff and Yuille 1991). Because the outputs from two modules are not independent, algorithms require feedback between modules and thus are open to the problems of convergence and instability (Clark and Yuille 1990).[1]

We describe a new illusion, that has a bistable three-dimensional (3D) interpretation. The bistability is interpreted as the result of cooperative coupling between depth from motion and phenomenal surface transparency. Phenomenal transparency of a surface means we can see it and through it to another background surface. This illusion poses a problem for computational models of structure from motion because it seems to require cooperative coupling or strong fusion between representations of relative surface depth and surface material.

Motion provides information about relative depth relationships between surfaces in the world. Information for depth is available from motion parallax and motion disparity (Wallach and O'Connell 1953). Theoretical work has shown how structure from motion can be reconstructed with *a priori* structural biases, such as assuming that the object viewed is rigid (Martin and Aggarwal 1988).[2] Interactions between depth from motion and other depth sources, such as stereo and proximity luminance, have been studied before (Dosher *et al.* 1986; Nawrot and Blake 1989). With respect to transparency, it has recently been discovered that degree of transparency determines whether two superimposed and independently moving square wave grating patterns at right angles to each other are seen as moving in a single direction or in two independent directions (Ramachandran 1989; Stoner *et al.* 1990). In these experiments, when the luminance of the intersection of the two gratings was consistent with that derived from a physically transparent grating, the motion of the two gratings was seen to be independent—they appeared to be

---

[1]Our distinction between integration and cooperative coupling closely corresponds to the distinction between weak data fusion and strong fusion with recurrency made by Clark and Yuille (1990).

[2]Motion parallax typically refers to the differences in image speed, due to viewpoint changes, between points at different distances under perspective projection. Because the same retinal optic flow pattern can be induced by moving the object, the term motion parallax is sometimes used in this case too. Structure from motion typically refers to the reconstruction of object geometry due to its motion under orthographic or perspective projection.

sliding over each other. However, if the intersection luminance was not consistent with a physically plausible transparency or with occlusion, the gratings appeared to cohere as a single pattern moving in a unique direction. The conclusion was that motion detecting mechanisms may have tacit knowledge of the physics of transparency. In these studies, the relationship between depth and transparency is only implicit. Although the transparency of a surface implies that it is closer than the surface it covers, one would like to know whether transparency can provide specific depth information that could affect three-dimensional structure from motion. It turns out that particular intensity relationships not only determine whether transparency is seen (Metelli 1974; Richards and Witkin 1979; Beck *et al.* 1984), but as is shown below, also bias which of two overlapping surfaces is seen in front. We call this depth from transparency. How do depth from motion and transparency interact? In particular, when depth from motion and depth from transparency contradict, which takes precedence—motion or transparency information?

2 Perceptual Observations

In an attempt to answer the above questions we simulated an object consisting of two square planar parallel surfaces that could rigidly rock back and forth at $\pm 40°$ about a common vertical axis midway between them (see Fig. 1 for more details). The planes could be seen as square when in a head-on view, but typically appeared trapezoidal due to perspective. Either the top or bottom face could be made to appear in front of the other depending on apparent transparency and depth from motion. The particular intensity relationships of the four regions bias the apparent transparency of a face, and thus determine the relative depth of the front and back planes. The motion parallax, together with a bias toward rigidity (Wallach *et al.* 1953) also affects the depth one sees.

In the following we will describe the basic perceptual phenomena, and then detail the results of some quantitative psychophysical measurements. In all three of the demonstrations described below, the rigid motion is described as being consistent with the bottom face being in front of the top face and only the intensities of the various regions are changed. As described in Section 3, the basic observations are unaffected by whether the top or bottom face is in front.

First we looked at the case in which both surfaces have zero transparency—that is, they are both opaque with the bottom square in front, and partially occluding the top (Fig. 2a). When the object was rocked back and forth, not surprisingly, observers saw rigid motion that was consistent with both the motion and occlusion cues. The surfaces do and appear to share a common rotation axis that is behind the bottom square, and in front of the top square. Next the intensity of the center patch of overlap was changed to match the top face. Thus the top patch appeared to
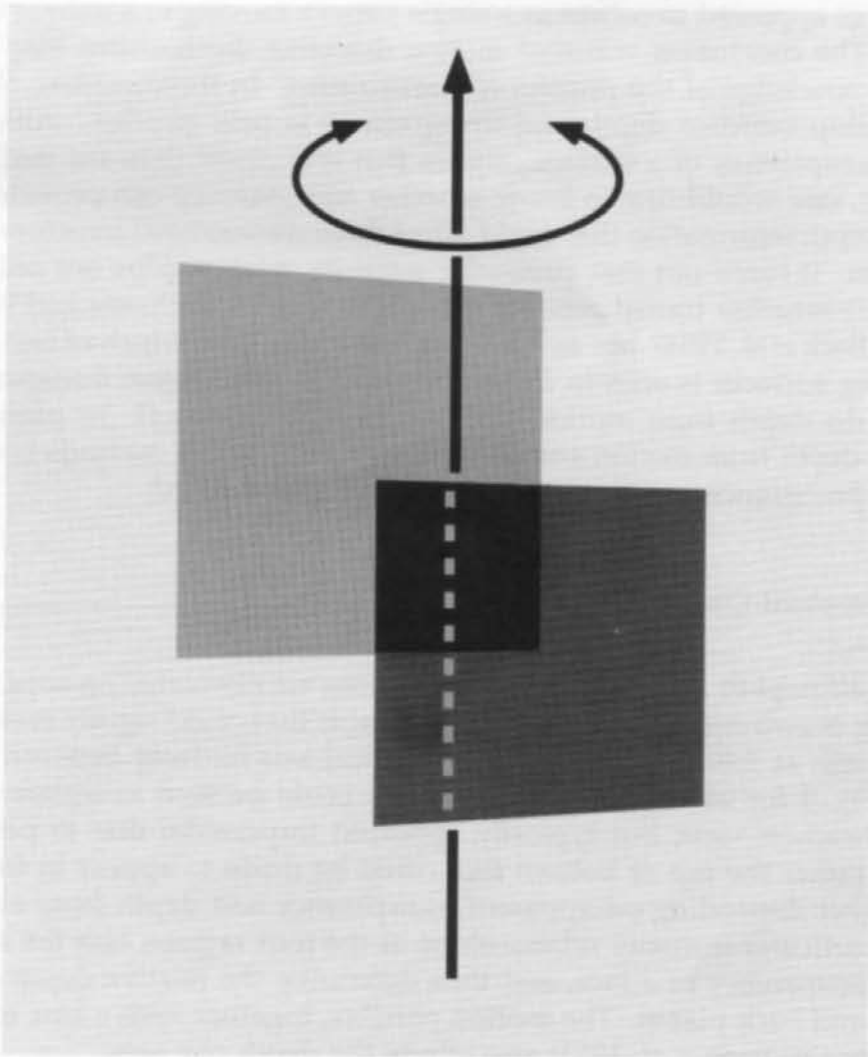
Figure 1: Animated sequences of images corresponding to a perspective view of two rigidly coupled planar faces (each a simulated 5×5 cm square) were generated with a Macintosh II computer and displayed on a CRT monitor with a 256 gray-level capacity. The object was rocked back and forth rigidly about the vertical axis passing between the two surfaces and through a point equidistant to both. Like the Necker cube, which is an orthographic projection of a wire cube, a particular image frame can give rise to an ambiguous depth percept: the top face can appear in front or behind the bottom face. The planes oscillated sinusoidally over an amplitude of $\pm 40°$ at 0.48 Hz. The front and back faces were separated by a simulated depth of 6 cm. The centers of the two faces were separated by 2.5 cm in the horizontal and vertical directions. The distance between the point equidistant between the two faces and the observer's eye-point was 57 cm. There were 21 frames per period.
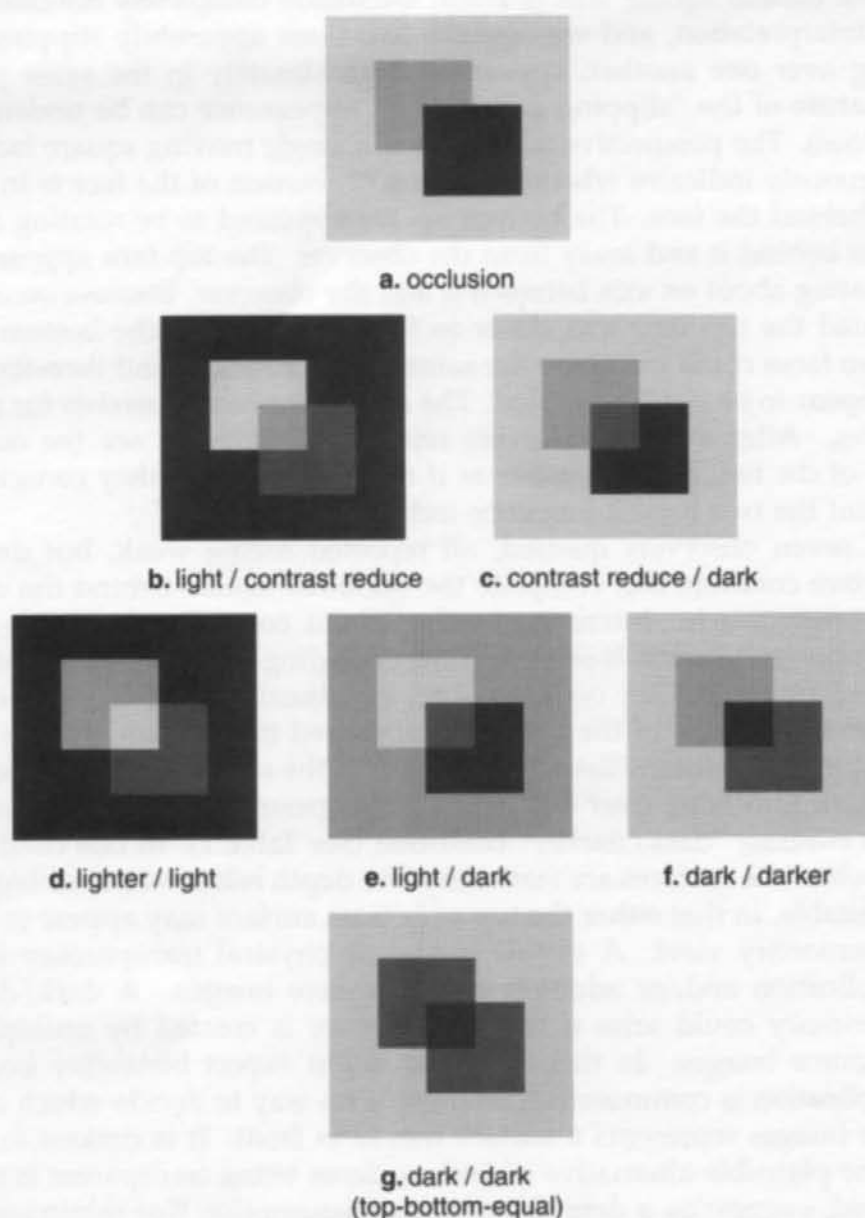
Figure 2: Different transparency types. The first and second words on the label for a transparency type indicate how the top and bottom faces affect the brightness of the patches they cover, respectively. For example, the dark/darker transparency means that both the top and bottom faces darkened what they cover, and that the bottom one was darker than the top. Note that in the actual animation, the nearer square appeared slightly larger because of the perspective projection. Note how the contrast reducing faces (bottom square in b and top square in c) tend to appear nearer the observer.

occlude the bottom, in contradiction to the rigid motion, which indicated that the bottom square was in front. Occlusion completely inhibited the rigid interpretation, and we saw the two faces apparently slipping and sliding over one another, appearing approximately in the same plane. The nature of the "slipping and sliding" appearance can be understood as follows. The perspective projection of a single moving square face unambiguously indicates whether the axis of rotation of the face is in front of or behind the face. The bottom square appeared to be rotating about an axis behind it and away from the observer. The top face appeared to be rotating about an axis between it and the observer. Because occlusion indicated the top face was closer to the observer than the bottom face, the two faces could not share the same axis of rotation, and therefore did not appear to be rigidly coupled. The nonrigid percept persists for many minutes. After awhile, observers report that they can see the outside edges of the two surfaces move as if rigidly coupled if they consciously discount the two local T-junctions indicating occlusion.[3]

Of seven observers queried, all reported seeing weak, but definite subjective contours that complete the occluded square behind the center overlapping patch. Interestingly, these faint contours are visible even when nonrigid motion is seen, as if the occluding patch were transparent.

Next we relaxed the occlusion cue, by adjusting the intensities of the patches so that one of the two faces appeared transparent. In one case, we adjusted the intensities so that either of the surfaces could appear to be a dark film lying over a light gray background, referred to below as a high contrast "dark/darker" condition (see Table 1). In this condition, even when the surfaces are stationary, the depth relations are ambiguous and bistable, in that either the top or bottom surface may appear in front in a stationary view. A simple model of physical transparency is the multiplication and/or addition of two source images. A dark/darker transparency could arise if the transparency is created by multiplying two source images. In this case, one might expect bistability because multiplication is commutative, so there is no way to decide which of the source images represents a surface that is in front. It is curious to note that the plausible alternative of both surfaces being transparent is never reported, suggesting a default perceptual assumption that minimizes the number of transparent surfaces. One can also adjust the intensities of the top and bottom squares to be equal. In this case the only biases to favor seeing a plane in front are to prefer the bottom over the top, and the larger over the smaller (Fig. 2g). In either the dark/darker or the top-bottom-equal condition, when the two faces were rocked back and forth, we saw a striking bistability. With the bottom face in front, we saw both planes rigidly rocking back and forth with the bottom face appearing transparent, and the top face opaque. After watching this for 2 to 30 sec, suddenly the top face would appear in front and then the

---

[3]A T-junction occurs where the edge of the occluder covers the edge of the occluded surface. An X-junction is the image point where transparent and opaque contours cross.

Table 1: Intensity Values (cd/m^2) for the transparency types.[a]

Transparency type	Top patch	Center patch	Bottom patch	Background Background	Contrast (%)
Dark/dark with top-bottom-equal	26	16	26	51	−24
Occlusion	38	16	16	51	0
Dark/darker (HC)	38	16	26	51	−24
Contrast reduce/dark (HC)	38	26	16	51	24
Light/dark (HC)	51	26	16	38	24
Light/contrast reduce (HC)	51	38	26	16	19
Lighter/light (HC)	38	51	26	16	32
Dark/darker (LC)	38	16	19	51	−8.6
Contrast reduce/dark (LC)	38	26	23	51	6.1
Light/dark (LC)	51	26	23	38	6.1
Light/contrast reduce (LC)	51	38	34	16	5.6
Lighter/light (LC)	38	51	46	16	5.2

[a]In addition to occlusion and a dark/dark top-bottom-equal transparency, the choice of transparency types was motivated by a consideration of the possible transparencies one can generate permuting four intensities. There are twenty-four possible permutations, but these can be reduced to just six by excluding top/bottom symmetry and the physically implausible contrast reversing and contrast enhancing pairs. Of these six, two involve faces that both darkened the underlying surfaces, so one was eliminated, leaving five. To further increase the range of transparency types, we also added five stimuli in which the Michelson contrast of the lower right-hand corner of the central patch was smaller. The high (HC) and low contrast (LC) groups had contrasts whose absolute values were above 19% and below 8.6%, respectively.

perceived motion was one of two faces slipping and sliding over each other. Simultaneous with this reversal of depth, there was an exchange of surface property—the top face now appeared transparent and the bottom opaque. In the top-bottom-equal condition, there is no depth from transparency bias to see one or the other face in front. Nevertheless, at a given moment, the visual system makes a commitment to one of two plausible relative depth and transparency assignments. The specific default assignment of depth, which lasts for a while and then changes, is similar to what happens when viewing a stationary Necker cube. Our demonstration, in addition, clearly demonstrates that relative depth interacts with apparent surface transparency.

In a third demonstration, we sought a condition intermediate between the symmetric transparency of a dark/dark combination and complete occlusion by constructing a transparent overlay that appears diaphanous. A diaphanous transparent square has both additive and multiplicative components that bias its relative depth to be in front of the other square. This can be physically realized by a finely perforated screen whose holes are below the spatial resolution limit and that transmits a fraction of the light coming from behind, and reflects a fraction coming from the front (Kersten 1991; Richards and Witkin 1979). Consistent with the interpretation of a perforated screen, a film that reduces the contrast of the edges

it overlays by lightening the darker region, and darkening the lighter, without changing contrast polarity tends to be seen in front (Fig. 2b and c). In the demonstration, the top square was made to appear contrast reducing. The bottom square was made to appear as a dark patch behind the contrast-reducing top square (the high contrast "contrast reduce/dark" condition in Table 1, Fig. 2c). When the two faces were rocked back and forth, we saw the wrong motion. Just as in the case of occlusion, the surfaces appeared to slip nonrigidly over one another with the top face appearing in front. After several seconds of observation, suddenly rigid motion is seen at which time the top contrast-reducing square is seen behind a dark bottom film. Again there was a simultaneous and unambiguous reversal of apparent transparency—the contrast-reducing top square suddenly appeared opaque and behind a dark film at the bottom.

Our paradigm is potentially useful for studying other interactions between structure from motion and transparency. For example, informal observations have shown that connecting the corners of the two faces with lines (e.g., a wire-frame outline of the cube) can override the nonrigid interpretation induced by weak perspective information. Further, even without connecting wires, an orthographic projection eliminates the nonrigid interpretation. Under orthographic projection, the appearance of the two faces flips between two rigid interpretations, like a Necker cube. In this case, the transparency biases which of the two rigid interpretations are seen.

3 Psychophysics: Response Time vs. Face-in-Front Bias _____

To quantify the interaction between transparency cues on depth and structure from motion, we made measurements of the reaction time to see rigid motion conditional on the perceived depth relations seen in an initial static view. The time to see rigid motion was measured in two basic conditions in which the initial depth perception, based on transparency, could either conflict (*inconsistent* condition) or agree (*consistent* condition) with the subsequent 3D motion. The experimental set-up was as before.

By specifying the gray levels of the four image regions, it was possible to vary the transparency, and thus produce a range of biases for whether a face of a particular transparency type appeared in front. This range of biases can be seen in the abscissa of Figure 3 where the proportion of times a particular face (e.g., the darker of a dark/darker combination) is seen in front (the face-in-front bias) ranges from 50 to 100% of the time depending on the transparency type for a given observer. We chose 12 different transparency types summarized in Table 1. The notation for the transparency type indicates how the top and bottom patches affect the brightness of the background. On half of the trials, the top face
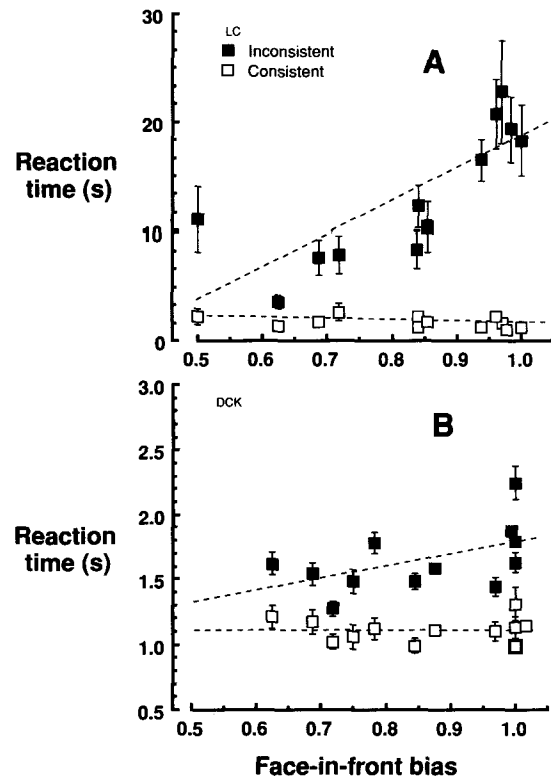
Figure 3: Mean time (±SEM) to see rigid motion plotted against the face-in-front bias for two observers (A and B). The face-in-front bias is the proportion of times a particular face appeared in front in the initial static view (e.g., the contrast-reducing face tended to appear in front of whatever face it overlapped). Results from the 12 different transparency conditions are shown. Each point is the mean of 16 measurements, averaged over conditions in which the top and bottom intensities were exchanged.

was in front of the bottom face (front-top), as defined by the subsequent motion, and on the other half of the trials, it was behind the bottom face (front-bottom). Because the perspective view made the image of the front patch larger than the back, we balanced for a possible bias by showing the observers the stimuli with the top and bottom intensities "normal" or "exchanged" for each of the front-top and front-bottom conditions. Because we could not guarantee that a given transparency condition would

1989; Kersten 1991; Trueswell and Hayhoe 1991; see also Nakayama *et al.* 1989). What are the implications of our psychophysical observations for neural computation?

The perceptual observations presented above suggest criteria that a model for the perception of surface transparency must eventually have to satisfy. Specifically, the model should handle (1) resolution of multiple locally ambiguous image data, (2) integration of depth from transparency and depth from motion, (3) multistable perception, and (4) cooperative computation of surface material and relative depth. Although there are models that incorporate some of these constraints, we know of no model that satisfies all of these criteria. Ultimately, a good neural model should also make some predictions concerning the neurophysiology.

4.1 Resolution of Multiple Locally Ambiguous Image Data. It seems unlikely that any neural model that works by propagating numerous local constraints could account for our psychophysical data, and the perception of transparency in general. The basic fact to contend with is that the local ambiguity at the two distantly separated X-junctions always seems to be resolved by the human visual system. For example, at a given instant, the relative depth of an edge in each X-junction of the top-bottom-equal stimulus (Fig. 2g), has a 50% chance of being in front of (or behind) the other edge it crosses. Any model that propagates local constraints along a contour, such as a Markov Random Field model for transparency (Kersten 1991), has a 50% chance that the two X-junctions are inconsistent. That is the vertical edge of one square is transparent, and the horizontal edge of the same square is opaque. This corresponds to a perception that is never seen. Iterative methods using various convergence tricks can avoid this problem and solve simple transparencies (Kersten 1991). However, any method that requires a large number of iterations to converge is an improbable algorithm for a perception that is fast enough to interact with structure from motion. A possible solution is to compute a relatively small number of intermediate surface representations, and then choose the most probable configurations.

4.2 Integration of Depth Cues. Part of the problem of the interaction between depth from motion and transparency is a cue integration problem (Terzopoulos 1986; Nawrot and Blake 1991; Bülthoff and Mallot 1988; Maloney and Landy 1989). In Bülthoff and Mallot's studies of depth integration, depth from shading and stereo was shown to accumulate, gradually increasing the perceived curvature of a smooth convex surface when the cues were consistent; but as in the present study, inconsistent cues were not resolved by averaging. One could imagine an accumulation of depth from transparency—a gradual increase in the contrast reduction of a planar surface mixing with the depth from motion to produce an intermediate relative depth. But this does not happen. The perceived depth

is fixed until suddenly it flips. This is analogous to the flip in the perception of the rotation of a clear transparent sphere that is covered with dots, and viewed under orthographic projection. The rotation ambiguity is resolved by stereo (Nawrot and Blake 1989). Nawrot and Blake have designed a network model of integration in which the sources of depth information (binocular and motion disparity) are at the same or corresponding image locations (Nawrot and Blake 1991). However, in our transparent stimuli, the transparency cues are at the X-junctions, and the depth from motion (and perspective) involves the movement of the X-junctions as well as the distant corners. There is no local motion or static intensity information at an X-junction to disambiguate the relative depth. Again, this argues for an intermediate explicit surface representation in a model of cooperativity involving material and depth attributes.

4.3 Multistability. One way of viewing multistability is in terms of the brain constructing an *a posteriori* probability of the world's state of affairs conditional on the image data (Kersten 1991). Multistability is reflected in multiple discrete modes of the probability distribution. These modes can also be identified with local minima in an energy function. But computing even a single mode can be problematic because of the local minima and convergence problems faced with nonlinear coupled networks. It is not too difficult to construct a posterior distribution or energy function that has the right modes, but it is difficult to construct one that has only the few correct modes. Further this Bayesian formulation does not answer the mystery of how the switch is made from one mode to the next. A number of the properties of perceptual multistability have been paralleled in simulated neural networks and algorithms (Kawamoto and Anderson 1985; Ditzinger and Haken 1989; Nawrot and Blake 1991). These models employ oscillating nonlinear dynamical systems, adaptation in unit activity level, or synaptic unlearning (e.g., via anti-Hebbian modification). However, exactly what happens in the brain during the flip between opacity and transparency is a puzzle for the future.

4.4 Cooperative Computation of Surface Material and Relative Depth. Our main point in this paper is that the striking bistability of the perceived motion together with the perceptual interchange of transparency and depth strongly suggest that surface transparency and relative depth are explicitly represented in the brain, and that they are computed cooperatively. The need for computing multiple interacting representations in vision has been pointed out by Barrow and Tenenbaum (1978), studied in the "Vision Machine" project at M.I.T. (Poggio *et al.* 1990) and supported psychophysically by studies in the perception of lightness (Gilchrist 1977). There have been only a few attempts at relating cooperative computation of scene attributes to psychophysics (Kersten and Plummer 1988; Adelson *et al.* 1989; Kersten 1991).

Both the Bayesian approach, and our results are consistent with the general notion of the visual system picking rational and plausible interpretations of scene properties causing the image. A good example for a rational interpretation of scene attributes is given by Blake and Bülthoff (1990). They showed that a binocularly viewed curved surface is only perceived as glossy if the specular highlight is close to the correct distance (according to ray optics and differential geometry) from the surface. Their work is also an excellent example for the disambiguation power of cue integration—the convex/concave ambiguity of shape-from-shading can be disambiguated by information about the position of specular highlights (Blake and Bülthoff 1990, 1991). Jepson and Richards (1991) propose a logical framework for cooperative computation between modules. The idea is to choose that interpretation of the world which jointly satisfies the premises about scene structure (e.g., rigidity) from the various vision modules, while avoiding unnecessary faulting of these premises. This is related to the work of Clark and Yuille (1990), who give a thorough discussion of the problems of data fusion between modules within a probabilistic framework.

4.5 Neurophysiology. Neurons in V2 of monkey visual cortex have been shown to respond to subjective contours, suggesting that these cells may have an important role in detecting surface occlusion (von der Heydt et al. 1984; von der Heydt and Peterhans 1989; Nakayama and Shimojo 1990). These V2 neurons are hypothesized to detect conjunctions in the outputs of spatially separate end-stopped neurons with the same preferred orientation (Peterhans et al. 1986). To signal objective contours, the model neuron also receives additive input from a complex cell whose orientation is perpendicular to the preferred orientation of the end-stopped cells. In light of the above arguments for explicit representation of transparent surfaces, it might be reasonable to look for neural responses to edges that depend on the monkey's perception of what surface region belongs to the edge. For example, some neurons (perhaps as early as V2) may be responsive to an edge constituting a transparency X-junction only when that edge belongs to a surface that appears in front of a second surface. There might be a bias to respond to the contrast reducing edge of an X-junction because it is likely to be in front of the background surface. These responses, if found, would presumably be influenced by the global structure of the surface (e.g., X-junctions outside of the classic receptive field). Responses to a local X-junction produced by a pair of moving transparent surfaces (as in Fig. 1) may also be expected to depend on changes in perceived depth induced by structure from motion. This is certainly possible, given the existence of feedback connections from MT to V2. MT is an area implicated in global motion processing (Allman et al. 1985; Movshon et al. 1985). One could also look for neural responses contingent on surface depth from motion. It has been shown by Snow-

den *et al.* that MT may play a role in the perception of transparent surface motion in random dot displays (Snowden *et al.* 1991).

Eventually the brain's computation of visual surfaces must incorporate information regarding the relations between surfaces. One way to do this would be to assume that the output of the above hypothetical V2 subjective contour unit combine with units signaling relative depth information. Alternatively, one could abandon the strictly bottom-up subjective contour model and incorporate feedback involving global surface information, such as that obtained from a motion module. In Section 2, we described the presence of faint subjective contours that completed the occluded square during motion. This observation is also suggestive of a possible role of feedback from a motion area, such as MT, to V2 in the process of surface computation.

Acknowledgments

References

Adelson, E. H., Pentland, A., and Juo, J. 1989. The extraction of shading and reflectance. *Invest. Ophthalmol. Visual Sci. Suppl.* **30**(3), 262.

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

Barrow, H. G., and Tenenbaum, J. M. 1978. Recovering Intrinsic Scene Characteristics from Images. In *Computer Vision Systems*, A. R. Hanson, and E. M. Riseman, (eds.), pp. 3–26. Academic Press, New York.

Beck, J., Prazdny, K., and Ivry, R. 1984. The perception of transparency with achromatic colors. *Percept. Psychophys.* **35**(4), 407–422.

Blake, A., and Bülthoff, H. H. 1990. Does the brain know the physics of specular reflection? *Nature (London)* **343**, 165–169.

Blake, A., and Bülthoff, H. H. 1991. Shape from specularities: Computation and Psychophysics. *Phil. Transact. R. Soc. (London) Ser. B* **331**, 237–252.

Bülthoff, H. H. 1991. Shape from X: Stereo, shading, texture, specularity. In *Computational Models of Visual Processing*. MIT Press, Cambridge, MA.

Bülthoff, H. H., and Mallot, H. A. 1988. Integration of depth modules: Stereo and shading. *J. Opt. Soc. Am. A* 5(10), 1749–1758.

Bülthoff, H. H., and Yuille, A. 1991. Bayesian models for seeing surfaces and depth. *Comments Theor. Biol.* 2(4), 283–314.

Cavanagh, P. 1987. Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity and shape. *Comput. Vision Graphics Image Process.* 37, 171–195.

Clark, J. J., and Yuille, A. L. 1990. *Data Fusion for Sensory Information Processing*. Kluwer Academic Publishers, Boston.

Ditzinger, T., and Haken, H. 1989. Oscillations in the perception of ambiguous patterns. *Biol. Cybernet.* 61, 279–287.

Dosher, B. A., Sperling, G., and Wurst, S. A. 1986. Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Res.* 26(6), 973–990.

Gilchrist, A. L. 1977. Perceived Lightness Depends on Perceived Spatial Arrangement. *Science* 195, 185–187.

Horn, B. K. P., and Brooks, M. J. 1989. *Shape from Shading*. MIT Press, Cambridge, MA.

Jepson, A., and Richards, W. 1991. Lattice framework for integrating vision modules. To appear in June issue of *IEEE Syst. Man Cybern.*

Kawamoto, A. H., and Anderson, J. A. 1985. A neural network model of multistable perception. *Acta Psychol.* 59, 35–65.

Kersten, D. 1991. Transparency and the cooperative computation of scene attributes. In *Computational Models of Visual Processing*, M. Landy, and A. Movshon, eds., pp. 209–228. MIT Press, Cambridge, MA.

Kersten, D., and Bülthoff, H. H. 1991. Transparency affects structure from motion. Tech. Rep. 34, Center for Biological Information Processing, Massachusetts Institute of Technology.

Kersten, D., and Plummer, D. J. 1988. Reflectance estimation in the presence of sharp shadows or transparency. *International Neural Network Society*. Boston.

Kersten, D., Bülthoff, H. H., and Furuya, M. 1989. Apparent opacity affects perception of structure from motion and stereo. *Investi. Ophthalmol. Visual Sci. Suppl.* 30(3), 264.

Land, E. H. 1983. Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image. *Proc. Natl. Acad. Sci. U.S.A.* 80, 5163–5169.

Livingstone, M. S., and Hubel, D. H. 1987. Psychophysical evidence for separate channels for the perception of form, color, movement and depth. *J. Neurosci.* 7(11), 3416–3468.

Maloney, L. T., and Landy, M. S. 1989. A statistical framework for robust fusion of depth information. *SPIE Visual Commun. Image Process.* 1199, 1154–1163.

Martin, W. N., and Aggarwal, J. K. 1988. *Motion Understanding: Robot and Human Vision*. Kluwer Academic Publishers, Boston.

Metelli, F. 1974. The perception of transparency. *Sci. Am.* 230(4), 91–98.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., and Newsome, W. T. 1986. The anal-

ysis of moving visual patterns. In *Pattern Recognition Mechanisms*, C. Chagas, R. Gattas, and C. G. Gross, eds., pp. 117–151. Vatican Press, Rome, Italy.

Nakayama, K., and Shimojo, S. 1990. Toward a neural understanding of visual surface representation. *Brain: Proc. Cold Spring Harbor Symp. Quant. Biol.* **55**, 911–924.

Nakayama, K., Shimojo, S., and Ramachandran, V. S. 1989. Depth, subjective contours and transparency: Relation to neon color spreading. *Invest. Ophthalmol. Visual Sci. Suppl.* **30**(3), 255.

Nawrot, M., and Blake, R. 1989. Neural integration of information specifying structure from stereopsis and motion. *Science* **244**, 716–718.

Nawrot, M., and Blake, R. 1991. A neural network model of kinetic depth. *Visual Neurosci.* **6**, 219–227.

Peterhans, E., von der Heydt, R., and Baumgartner, G. 1986. Neuronal responses to illusory contours stimuli reveal stages of visual cortical processing. In *Visual Neuroscience*, J. Pettigrew, K. J. Sanderson, and W. R. Levick, eds., pp. 343–351. Cambridge University Press, Cambridge, England.

Poggio, T., Gamble, E. B., and Little, J. J. 1988. Parallel integration of vision modules. *Science* **242**, 436–440.

Poggio, T., Little, J., Gamble, E., Gillett, W., Geiger, D., Weinshall, D., Villalba, M., Larson, N., Cass, T., Bülthoff, H., Drumheller, M., Oppenheimer, P., Yang, W., and Hurlbert, A. 1990. The M. I. T. Vision Machine. In *Artificial Intelligence at MIT: Expanding Frontiers—Volume 2*, P. H. Winston, and S. A. Shellard, eds., pp. 492–529. MIT Press, Cambridge, MA.

Ramachandran, V. 1989. Constraints imposed by occlusion and image segmentation. *Vision and Three-dimensional Representation*. University of Minnesota, Minneapolis, MN.

Ramachandran, V. S., and Cavanagh, P. 1985. Subjective contours capture stereopsis. *Nature (London)* **317**, 527–530.

Richards, W., and Witkin, A. P. 1979. *Efficient Computations and Representations of Visible Surfaces*. Tech. Rep. AFOSR-79-0020. Massachusetts Institute of Technology.

Shimojo, S., Silverman, G. H., and Nakayama, K. 1989. Occlusion and the solution to the aperture problem for motion. *Vision Res.* **29**(5), 619–626.

Snowden, R. J., Treue, S., Erickson, R. G., and Andersen, R. A. 1991. The response of area MT and V1 neurons to transparent motion. *J. Neurosci.* **11**(9), 2768–2785.

Stoner, G. R., Albright, T. D., and Ramachandran, V. S. 1990. Transparency and coherence in human motion perception. *Nature (London)* **344**, 155–157.

Terzopoulos, D. 1986. Integrating visual information from multiple sources. In *From Pixels to Predicates*, A. Pentland, ed., pp. 111–142. Ablex Publishing Corporation, Norwood, NH.

Trueswell, J. C., and Hayhoe, M. 1991. Transparency interacts with binocular disparity to determine perceived depth. *Invest. Ophthalmol. Visual Sci. Suppl.* **32**(4), 694.

Ullman, S. 1979. The interpretation of structure from motion. *Proc. R. Soc. London Ser. B* **203**, 405–426.

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

von der Heydt, R., and Peterhans, E. 1989. Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J. Neurosci.* **9**, 1731–1748.

von der Heydt, R., Peterhans, E., and Baumgartner, G. 1984. Illusory contours and cortical neuron responses. *Science* **224**, 1260–1262.

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

Zeki, S. 1978. Functional specialization in the visual cortex of the Rhesus monkey. *Nature (London)* **274**, 423–428.