REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

12

2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
January 1992	memorandum

4. TITLE AND SUBTITLE

Apparent Opacity Affects Perception of Structure from Motion

5. FUNDING NUMBERS

DACA76-85-C-0010
N00014-85-K-0124

6. AUTHOR(S)

Daniel Kersten and Heirich Bülthoff

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Artificial Intelligence Laboratory
545 Technology Square
Cambridge, Massachusetts 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

AIM 1285
C.B.I.P. 34

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Information Systems
Arlington, Virginia 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution of this document is unlimited

12b DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

It is well known that the human visual system can reconstruct depth from simple random-dot displays given motion information. This fact has lent support to the notion that structure from stereo and motion systems rely on low-level primitives or tokens, such as edges, derived from image intensities. In contrast, the judgment of surface attributes such as transparency or opacity is often considered to be a higher-level visual process that would make use of low-level stereo or motion information, and perhaps attention or later recognition to tease apart the transparent from the opaque parts. This is exemplified by the lack of computational studies dealing with transparency, compared with the at least limited success of a number

(continued on back)

14. SUBJECT TERMS (key words)	15. NUMBER OF PAGES
	14
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

of algorithms to solve structure from motion or stereo. In this study, we describe a new illusion and some results that question the above view by showing that depth from transparency and opacity can override the rigidity bias in perceiving depth from motion. This provides support for the idea that the brain's computation of the surface material attribute of transparency may have to be done either before, or in parallel with the computation of structure from motion.

12

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 1285 January 1991
C.B.I.P Memo No. 34

Apparent Opacity affects Perception of Structure from Motion

Daniel Kersten Heinrich H. Bülthoff

Abstract

It is well known that the human visual system can reconstruct depth from simple random-dot displays given motion information. This fact has lent support to the notion that structure from stereo and motion systems rely on low-level primitives or tokens, such as edges, derived from image intensities. In contrast, the judgment of surface attributes such as transparency or opacity is often considered to be a higher-level visual process that would make use of low-level stereo or motion information, and perhaps attention or later recognition to tease apart the transparent from the opaque parts. This is exemplified by the lack of computational studies dealing with transparency, compared with the at least limited success of a number of algorithms to solve structure from motion or stereo. In this study, we describe a new illusion and some results that question the above view by showing that depth from transparency and opacity can override the rigidity bias *in perceiving depth* from motion. This provides support for the idea that the brain's computation of the surface material attribute of transparency may have to be done either before, or in parallel with the computation of structure from motion.

This report describes research done at the Massachusetts Institute of Technology within the Artificial Intelligence Laboratory and the Center for Biological Information Processing in the Department of Brain and Cognitive Sciences and Whitaker College. The Center's research is sponsored by grant N00014-91-J-1270 from the Office of Naval Research (ONR), Cognitive and Neural Sciences Division; and by National Science Foundation grant IRI-8719394. The Artificial Intelligence Laboratory's research is sponsored by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76-85-C-0010 and in part by ONR contract N00014-85-K-0124. Daniel Kersten was supported by NSF grant BNS-8708532 to. His present address is: Department of Psychology, University of Minnesota Minneapolis, Minnesota 55455. Heinrich H. Bülthoff is now at the Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912.

17

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 modules such as surface-color-from-radiance, shape-from-shading, or structure-from-motion (Land, 1959; Horn, 1975; Ullman, 1979). A major result of these studies is that scene reconstruction from image data is often under-constrained—there are many solutions that satisfy the data. Prior constraints then have to be sought to find a unique interpretation of the environment from the image intensities. One promising avenue of research to reduce the strength of prior assumptions required is *integration*—the combination of visual information from multiple sources, such as stereo and motion. Poggio (1985) proposed a theory based on a Bayesian approach that attempts to estimate the posterior probability of, say, depth, given all the data from different sensors and algorithms and *a priori* knowledge, embedded in an appropriate *prior* distribution. The theory assumes a specific model for the underlying probabilities, the MRF model, and uses a number of techniques –deterministic and stochastic – to estimate the appropriate quantities associated with the posterior probability, given the data, such as its maximizer or its mean(Little et al., 1988). This theory formed the basis of the MIT Vision Machine project (see eg., (Poggio et al., 1990)).

A second approach is *cooperative coupling* of the estimates of various scene attributes to achieve the consistency required by the laws of image formation.[1] Consistent with the methodology of computer vision, current physiological, anatomical and psychophysical research indicates modular and concurrent processing, such as for motion, as distinct from form and color (Zeki and Shipp, 1978; Livingstone and Hubel, 1987; Cavanagh, 1987). The number of distinct visual cortical areas is thought to be over twenty, each with a potentially different function, and with both feedforward and feedback connection between many of them (Essen, 1985). At this point, however, there are only vague ideas of the relationship between the processing streams in the brain, the modules of computational analysis, and perception as they pertain to integration and cooperative coupling of visual information.

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 visual ambiguity one expects from weak prior constraints is the exception, rather than the rule. In the 19th century, Ernst Mach demonstrated

[1]Cooperative coupling refers to the interaction between two perceptual representations of scene attributes (such as surface depth and reflectance) in order to satisfy a mutual consistency constraint usually imposed by how the image could be formed physically. The Mach card is an example of the cooperative coupling of perceived reflectance and relative depth. See D. J. Kersten, in "Computational Models of Visual Processing" M. Landy, A. Movshon, Eds. (M.I.T. Press, Cambridge, Massachusetts, 1991), and H. Bülthoff and A. Yuille, SPIE Visual Communication and Image Processing (1990) for a discussion of coupling of visual information.

that perceptual representations of the environment do interact in human perception and interact in such a way as to produce a consistent perception of the state of the scene that is unambiguous at a given moment, but bistable over time (Mach, 1959). In his well-known Mach-card illusion, the perceived surface color or lightness of a simple folded card, placed on a table, depends on light source direction, and the bistably perceived geometry of the card. We describe a new illusion, that like the Mach Card has a bistable 3D interpretation; but the bistability is induced through motion parallax, and rather than interacting with the lightness of a surface, the perceived depth affects the phenomenal transparency.[2] Using this stimulus, we have studied how the human perception of depth from motion interacts with the perceived surface attribute of opacity.

It is well-known that motion provides information about relative depth relationships between surfaces in the world. Interactions between depth from motion and and other depth sources, such as proximity luminance, have been studied before (Dosher et al., 1986). It has recently been discovered that degree of transparency determines whether two superimposed and independently moving square wave patterns are seen as moving in a single direction or in two independent directions (Ramachandran, 1989; Stoner et al., 1990). Less well appreciated is that fact that transparency cues also provide depth information. Particular intensity relationships not only determine whether transparency is seen (Metelli, 1974; Beck et al., 1984), but also bias which of two overlapping surfaces is seen in front. We call this depth from transparency. Perception of transparency can lead to neon-color spreading, and loss of stereoscopic capture (Nakayama et al., 1989). It has also been shown that perception of incorrect depth from transparency can lead to a delay in seeing the correct depth relationships between surfaces based on stereo or motion information (Kersten et al., 1989). In this paper we specifically address the question: "When motion and transparency contradict, which takes precedence—motion or transparency information?"

2 Method

In an attempt to answer the above question, we simulated an object consisting of two square planar parallel surfaces that could rigidly rock back and forth about a vertical axis perpendicular to the line of site (Fig. 1). Animated sequences of images corresponding to a perspective view of two planar and possibly transparent faces (each a simulated 5 x 5 cm square) were generated with a Macintosh II computer and displayed on a CRT monitor with a 256 gray-level capacity. The bias of the apparent depth of the two faces was controlled by motion and the intensity relations in the display that invoke various types of transparency. To provide motion information about the relative depths of the two faces, they were rocked back and forth rigidly about the vertical axis passing between the two surfaces and passing through a point equidistant to both. Like the Necker cube

[2]Phenomenal transparency of a surface means we can see through it to another background surface. A perceptual consequence of phenomenal transparency is interpreting the transparent surface as being in front of the background.

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 bias of the apparent depth of the two faces was controlled by motion and transparency pattern. To provide motion information about the relative depths of the two faces, the planes oscillated sinusoidally back and forth by 40 deg about the vertical axis at 0.48 Hz. The distance between the point equi-distant between the two faces and the observer's eye-point was 57 cm. There were 21 frames per period. The planes could be seen as square when in a head-on view, but typically appeared trapezoidal due to perspective. The top (or bottom) face, could either appear in front or behind the other. The depth relation seen depends on perceived transparency and 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 together with a bias toward rigidity also affects the depth one sees (Wallach and O'Connell, 1953; Ullman, 1979). Depth also depends on the a priori bias of the observer to see a a rigid body in perspective with the front face larger than the rear face, or alternatively, with the front face smaller than the rear face, but we do not study this here.

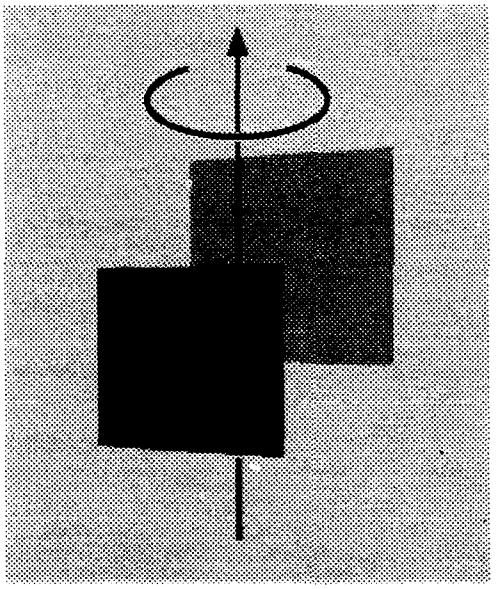

Figure 1: Animated sequences of images corresponding to a perspective view of two rigidly coupled planar and possibly transparent faces were displayed on a 8-bit CRT monitor. The object was rocked back and forth rigidly about the vertical axis passing between the two surfaces and through a point equidistant to both. Like the Necker cube, a particular image frame can give rise to an ambiguous depth percept: the top face can appear in front of or behind the bottom face.

3 Basic Perceptual Phenomena

In the following sections we will describe the basic perceptual phenomena, and then detail the results of some quantitative measurements. In all three of the demonstrations discussed 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. The basic phenomena are unaffected by placing the top face in front.

3.1 Opaque Surfaces

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. Next the intensities were adjusted so that the top patch appeared to 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 slipping and sliding over one another. This percept persists for many minutes. After awhile, some observers report that they can see the outside edges of the two surfaces move as if rigidly coupled if they consciously discount the "T" junctions indicating occlusion. Observers—seven out of seven informally queried as to whether they saw them or not—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.

3.2 Relaxed Occlusion

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 Fig. 2f and 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. From a formal point of view, one might expect this when the image results from multiplying two source images. Multiplication is commutative, so there is no way to decide which surface is in front. It is curious to note that the plausible alternative of both surfaces being transparent is never reported. One can also adjust the intensities of the top and bottom squares to be equal in which case the only biases to favoring front are to prefer the bottom over the top, and the larger over the smaller (Fig. 2g). In either case, when the two planes were rocked back and forth, we saw a striking bistability. If the bottom face was seen in front in an initial static view, we saw both planes rigidly rocking back and forth with the bottom face appearing transparent, and the top face opaque. After watching this for anywhere between 2 to 30 seconds, suddenly the top face

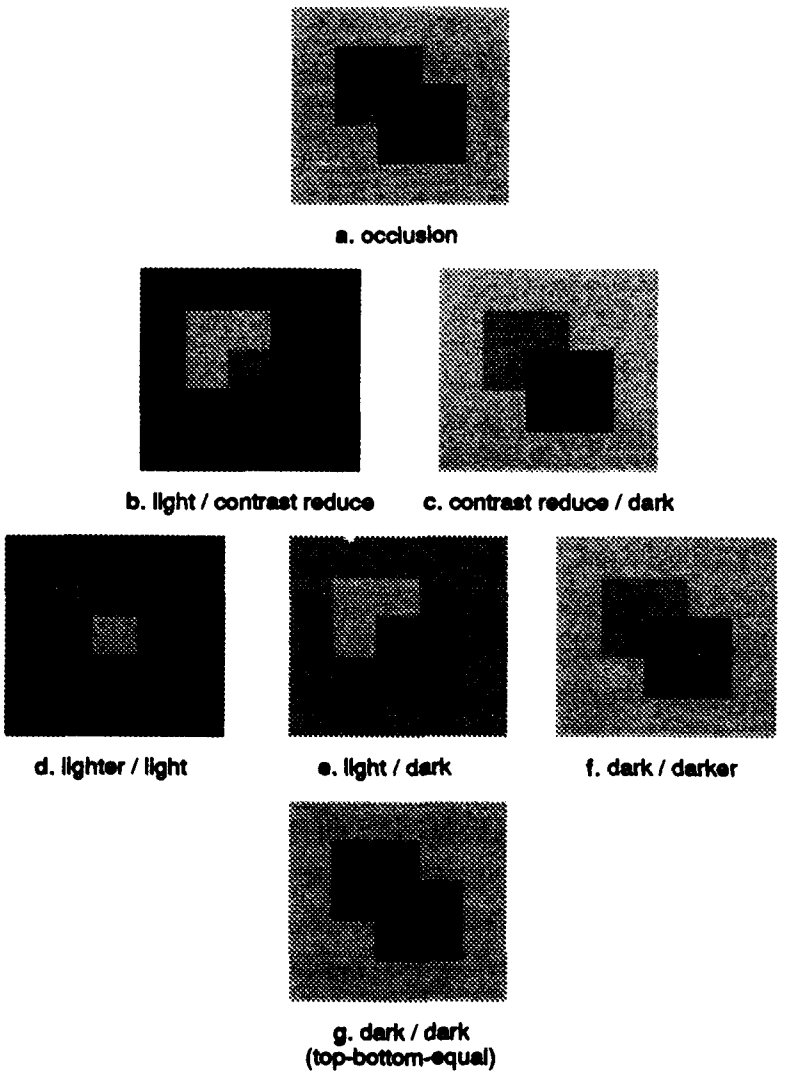

Figure 2: Five transparency types were used to induce different strengths of depth-from-transparency cues in which the top-bottom squares could have the following effect on the intensities that they covered: dark/darker, contrast reduce/dark, light/dark, light/contrast reduce, lighter/light. These five types were built from permutations of four intensities : 16, 26, 38 and 51 cd/m^2 for a high contrast condition. We also tested responses to 5 low contrast versions of these five types, an occlusion case, and a balanced dark/dark condition in which the top and bottom were both equal in intensity was included (see Table 1).

would appear in front and then the 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. The fact that these multistable percepts are still seen when the transparency cues to depth were exactly balanced (Fig. 2g) shows that a default assignment of relative depth (as with a stationary Necker cube) and transparency is made which interacts with depth

from motion.

3.3 Diaphanous 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, as shown below, bias its relative depth to be in front of the other square. This can be physically realized by a perforated screen whose holes are below the spatial resolution limit and which transmits a fraction of the light coming from behind, and reflects a fraction coming from the front (Richards and Witkin, 1979; Kersten, 1991). 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,c). In the demonstration, the top square was made to appear contrast reducing. The bottom square was made to appear as a dark milky film behind the high 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.

4 Interaction between Transparency and Structure from Motion

In order 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* conditions) or agree (*consistent* condition) with the subsequent 3D rigid motion. The experimental set-up was as before.

By specifying the gray-levels of the four image regions, it was possible to control apparent transparency, and thus bias whether the top face or the bottom face appeared in front. We chose 12 different transparency types summarized in Table 1. The notion of the transparency type indicates how the top and bottom patches affect the brightness of the background. 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. If both faces lighten the background, one of them still appears lighter and is indicated in the label. The same rule is used when both faces darken the background. For example, a "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. There

Transparency type	Luminance $[cd/m^2]$				
	Top	Center	Bottom	Background	Contrast (%)
top-bottom-equal dark/dark	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

Table 1: Intensity values of the center, bottom, top and background regions of the two planes are shown in cd/m^2. HC and LC refer to high and low contrast conditions, respectively.

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. In order to further increase the range of transparency types, we also added five stimuli in which the local edge contrast (Michelson contrast) of the lower right hand corner of the central patch was smaller.

To understand our selection better, consider the top horizontal edge of the bottom patch of one of the transparencies in Figure 2. It crosses a vertical boundary of the top patch. If the bottom patch is not seen as a hole, the horizontal edge is attached to, or "intrinsic" to the bottom patch. This bottom film can either preserve or reverse the contrast polarity of the two regions separated by the vertical edge. A high contrast reversing surface does not in general appear transparent. Suppose the horizontal edge is contrast preserving. Then it can either lighten or darken the underlying regions, or it can reduce or enhance the contrast at the vertical edge. When the horizontal edge of a neutral density filter with transmittance less than 100% crosses the vertical boundary, it darkens the intensity on both sides of this edge (see "dark/darker" condition). A purely positive additive transparency lightens both regions that it covers. Of particular interest here is an edge that reduces contrast in the sense that it lightens the darker of the two regions it covers, and darkens the lighter without reversing the contrast polarity ("contrast reducing" condition). If the horizontal edge reduces contrast, there must be a vertical edge that darkens both regions while reversing contrast. Further, the horizontal edge, if considered attached to the top region, is contrast enhancing in the sense of

darkening the darker of two regions it covers, and lightening the lighter without changing contrast polarity. Surfaces attached to contrast enhancing edges are not likely to be seen as transparent surface discontinuities. This provides a cue to edge attachment, and thus occlusion.

4.1 Perceptual biases

We wanted to find out how the degree of bias to see a particular surface as transparent would affect the time to see rigid motion when the motion either agreed or disagreed with the depth from transparency cues.

In order to increase the number of stimuli, we included the five additional transparencies, similar to those in Figure 2 in which the local edge contrast of the lower right hand corner of the central patch was smaller. The high and low contrast groups had contrasts whose absolute values were above 19% and below 8.6%, respectively. On half of the trials, the top face 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). Further, because the perspective view made the image of the front patch larger than the back, the observers were shown the stimuli with the top and bottom intensities "normal" or "exchanged" for each of the front-top and front-bottom conditions. Subjects first viewed a static head-on view of the two faces from a distance of 57 cm. Because we could not guarantee, for example, that a given transparency condition would generate a consistent depth ordering, the observer was asked to indicate whether the top or bottom surface appeared in front by pushing a button. This button press also initiated the animation of the object. The subject was to push another button once rigid motion was seen. The time to see rigid motion was measured. There were 5 subjects, 1 of which was naive to psychophysical experiments. Each subject saw each stimulus eight times. The presentation order was randomized.

A five way ANOVA on reaction times (subjects vs. normal/exchanged vs. front-top/front-bottom vs. contrast vs. transparency type) showed a significant three-way interaction between transparency type, normal/exchanged, and front-top/front-bottom factors ($p < 0.0001$) indicating that there was a preferred face to be seen in front in a static view that interacted with the subsequent motion. There was also a significant difference in the range of observers reaction times, between 0.5 and 3 seconds for one observer, and between 1 and 30 seconds for the second. There was no significant main effect of high vs. low contrast on the interaction.

Figure 3 presents the main observation of reaction time for two observers in a simpler way by averaging the reaction times over conditions in which the depth from transparency is either consistent or inconsistent with depth from motion. Motion and transparency information could be consistent (or inconsistent) in two ways. For example, the transparency information could either indicate that the bottom square was in front when rigid motion concurred, or that it was behind when rigid motion concurred. Figure 3 shows that the reaction times were substantially longer when the transparency cues gave depth relations inconsistent with the subsequent rigid motion for all transparency conditions for

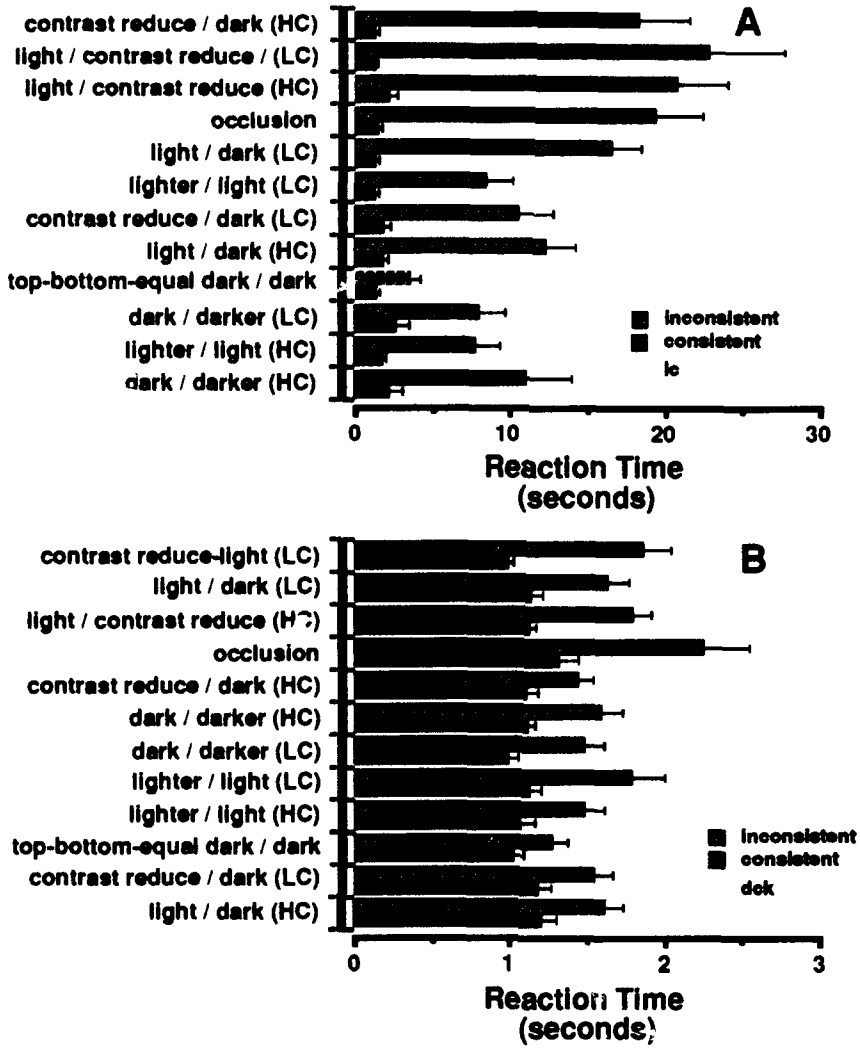

Figure 3: The times to see rigid motion of the front and back faces was measured for 12 different opacity conditions are shown here for two observers. In all cases the time to see rigid motion when the initial static opacity or transparency cues indicated a relative depth that was inconsistent with the subsequent rocking motion was longer than when the cues were consistent. The transparency types are arranged from bottom to top in order of increasing likelihood that a particular plane consistently appears in front (or behind) the other face (see Table 2).

two observers. We have tested 5 other observers on 15 other variations of transparency relations and this pattern of results has held for all—the consistent reaction times are shorter than the inconsistent times, although as in Figure 3 there are substantial individual differences in the values of the average times.

There was also an effect of the type of transparency on the preferred depth relation seen. In Figure 4 the same data are replotted in a different way in order to visualize the gradual increase in the reaction time with the strength of inconsistency given by a

Subject⇒	dck		lc		zl		sk		pm	
Tranparency type ⇓	Plane favored to be seen in front	%	Plane favored to be seen in front	%	Plane favored to be seen in front	%	Piane favored to be seen in front	%	Plane favored to be seen in front	%
top-bottom-equal	bigger	69	bigger	75	bigger	56	smaller	56	bigger	63
occlusion	occluder	100	occluder	97	occluder	100	occluder	100	occluder	91
dark / darker (HC)	darker	88	neither	50	dark	62	darker	53	neither	50
cont. rd. / dark (HC)	cont. rd.	97	cont. rd.	100	cont. rd.	97	cont. rd.	100	cont. rd.	97
light / dark (HC)	dark	63	light	84	light	91	light	81	light	56
light / cont. rd. (HC)	cont. rd.	100	cont. rd.	97	cont. rd.	53	cont. rd.	59	cont. rd.	59
lighter / light (HC)	lighter	75	lighter	69	lighter	66	lighter	63	lighter	53
dark / darker (LC)	darker	84	darker	72	dark	69	darker	59	darker	69
cont. rd. / dark (LC)	cont. rd.	69	cont. rd.	84	cont. rd.	66	cont. rd.	62	neither	50
light / dark (LC)	dark	100	dark	94	dark	72	dark	72	dark	78
light / cont. rd. (LC)	cont. rd.	100	cont. rd.	97	cont. rd.	100	cont. rd.	100	cont. rd.	100
lighter / light (LC)	light	78	light	84	lighter	62	lighter	75	lighter	78

Table 2: The face-in-front bias for different transparency types is shown for five subjects. The bias is measured as the percentage of time a particular face appears in front in a static view.

face-in-front bias. This bias is the proportion of times a particular face was perceived in front in the initial static view. Apart from occlusion and contrast-reducing transparency, there was no general rule to predict the face-in-front bias across observers. However, in all four of the contrast reducing conditions, the contrast reducing face appeared in front of any other type of face in the initial static view at least 50% of the time, or more (Table 2). In two of the conditions ("light/contrast reducing" and "contrast reducing/ dark") the probability of seeing the contrast reducing face in front was 97% or more for all five observers[3]

5 Discussion

Evidence has been presented elsewhere that surface occlusion information may be represented early in the visual system. In particular, occlusion can override stereo (Ra-

[3]The probability was estimated by averaging over 16 presentations each for all five observers.

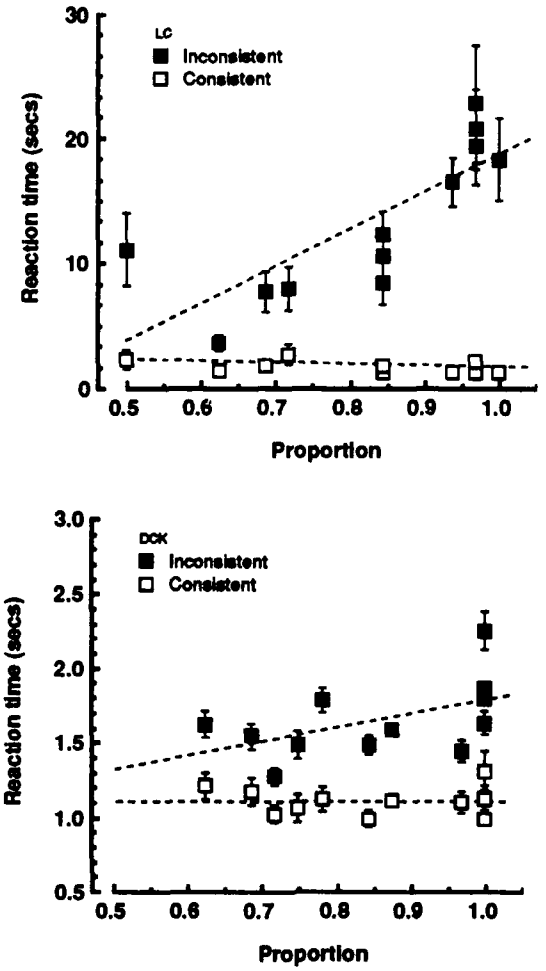

Figure 4: Mean time (±SEM) to see rigid motion plotted against the face-in-front bias for two observers. The face-in-front bias is the proportion of times a particular face appeared in front in the initial static view. Results from 12 transparency conditions are plotted. Each point is the mean of 16 measurements, averaged over conditions in which the top and bottom intensities were exchanged.

machandran and Cavanagh, 1985), raise recognition performance for faces (Nakayama et al., 1988), and affect motion perception (Shimojo et al., 1989). Our results are consistent with the idea that the determination of what regions the boundary of a surface belongs (i.e. intrinsic or extrinsic edges) is done early. We add to this that the attachment of an edge to a region is influenced by transparency, and is also done early enough to affect the perceived relative motion between two surfaces.

Computational vision research has underlined the importance of questions of representation, modularity and algorithm (Marr, 1982). In addition, we need to know what to

compute when. The striking bistability of the perceived motion together with the quantitative increase in reaction time when motion and transparency cues are in opposition strongly suggest that surface transparency and relative depth are explicitly represented in the brain, and that they are computed cooperatively, rather than in strict sequence. These results point to central problems of depth integration and representation, and cooperative computation of multiple scene attributes. In previous studies (Bülthoff and Mallot, 1987; Bülthoff and Mallot, 1988), depth from shading and stereo was shown to accumulate, gradually increasing the perceived curvature of a smooth convex surface when the cues were consistent. As here, however, 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. What kind of mechanism can explain this? 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 modes of the probability distribution. This formulation, however, does not answer the mystery of how the switch is made from one mode to the next. A number of the properties of simulated neural-like networks parallel properties of perceptual multistability (Kawamoto and Anderson, 1985), but whether this is how the computation is realized in the brain remains a challenging problem for the future.

Acknowledgements

The authors would like to thank Momi Furuya, Ken Kurtz and Bennett Schwartz for data collected in pilot studies, and David Knill for many useful comments and suggestions, and Edward Adelson who first showed us the striking effect certain types of contrast reduction has on depth ordering of transparencies.

References

Barrow, H. G. and Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In Hanson, A. R. and Riseman, E. M., editors, *Computer Vision Systems*, pages 3–26. Academic Press, New York, NY.

Beck, J., Prazdny, S., and Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception and Psychophysics*, 35:407–422.

Bülthoff, H. H. and Mallot, H. A. (1987). Interaction of different modules in depth perception. In *Proceedings of the 1st International Conference on Computer Vision*, pages 295–305.

Bülthoff, H. H. and Mallot, H. A. (1988). Interaction of depth modules: stereo and shading. *Journal of the Optical Society of America*, 5:1749–1758.

Cavanagh, P. (1987). Reconstructing the third dimension: interactions between color, texture, motion, binocular disparity and shape. *Computer Vision, Graphics, and Image Processing*, 37:171–195.

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

Essen, D. C. V. (1985). Lightness and retinex theory. *Cerebral Cortex*, 3:259–329.

Horn, B. K. P. (1975). Obtaining shape from shading information. In Winston, P. H., editor, *The Psychology of Computer Vision*, pages 115–155. McGraw-Hill, New York, NY.

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

Kersten, D. (1991). Mechanisms of attention. In Landy, M. and Movshon, A., editors, *Computational Models of Visual Processing*. M.I.T. Press, Cambridge, MA. in press.

Kersten, D., Bülthoff, H. H., and Furuya, M. (1989). Apparent opacity affects perception of structure from motion and stereo. *Supplement to Investigative Ophthalmology and Visual Science*, 30:264.

Land, E. H. (1959). Color vision and the natural image. *Proceedings of the National Academy of Science*, 45:115–129.

Little, J. J., Poggio, T., and Gamble Jr., E. B. (1988). Seeing in Parallel: The Vision Machine. *International Journal of Supercomputing Applications*, 2:13–28.

Livingstone, M. S. and Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, colour, movement and depth. *The Journal of Neuroiscience*, 7:3416–3468.

Mach, E. (1886,1959). *The Analysis of Sensations.* Dover, New York.

Metelli, F. (1974). The perception of transparency. *Scientific American*, 230:91–98.

Nakayama, K., Shimojo, S., and Ramachandran, V. S. (1989). Depth, subjective contours and transparency: relation to neon color spreading. *Supplement to Investigative Ophthalmology and Visual Science*, 30:255.

Nakayama, K., Shimojo, S., and Silverman, G. H. (1988). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18:739–746.

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

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

Ramachandran, V. S. (1989). Constraints imposed by occlusion and image segmentation. Conference on *Vision and Three-dimensional Representation*, Minneapolis, Minnesota.

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

Richards, W. and Witkin, A. P. (1979). How to play twenty questions with nature and win. AFOSR Report Contract Number AFOSR-79-0020, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

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

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

Ullman, S. (1979). *The interpretation of visual motion.* MIT Press, Cambridge, MA.

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