Biol. Cybern. 66, 327–333 (1992)

Biological Cybernetics

Structure-from-motion based on information at surface boundaries

William B. Thompson [1], Daniel Kersten [2], and William R. Knecht [2]

[1] Department of Computer Science, University of Utah, Salt Lake City, UT 84112, USA
[2] Psychology Department, University of Minnesota, Minneapolis, MN 55455, USA

Received February 2, 1991/Accepted in revised form August 21, 1991

Abstract. Existing computational models of structure-from-motion – the appearance of three-dimensional motion generated by moving two-dimensional patterns – are all based on variations of optical flow or feature point correspondence within the interior of single objects. Three separate phenomena provide strong evidence that in human vision, structure-from-motion is significantly affected by surface boundary cues. In the first, a rotating cylinder is seen, though no variation in optical flow exists across the apparent cylinder. In the second, the shape of the bounding contour of a moving pattern dominates the actual differential motion within the pattern. In the third, the appearance of independently moving objects changes significantly when the boundary between them becomes indistinct. We describe a simple computational model sufficient to account for these effects. The model is based on qualitative constraints relating possible object motions to patterns of flow, together with an understanding of the patterns of flow that can be discriminated in practice.

1 Introduction

The inference of three-dimensional structure and motion from changes in a two-dimensional image has for many years been an important topic of research in both the perceptual psychology and computational vision communities. Wallach and O'Connell (1953) demonstrated the "kinetic depth effect" in which people were able to recognize the three-dimensional shape of unfamiliar moving objects based on the changing projection of the objects onto a screen. More recently, Ullman (1979) did the first extensive computational analysis of this problem, showing that correspondences between a small number of image features over a few frames can be used to recover the relative three-dimensional position of the corresponding object points and the object motion occurring between frames. Many hundreds of papers on aspects of this *structure-from-motion* (SFM) problem followed.

Except for special cases involving jointed motion, all computational models of structure-from-motion have been based on an analysis of spatial and/or temporal variations in either optical flow or feature point correspondences within the interior of a single object. No methods yet described explicitly use information available at surface boundaries. If boundaries are mentioned at all, it is to claim that such boundaries can somehow be used to segment the scene into components involving a single rigid object before the SFM algorithms are applied.

In this paper, we demonstrate that rotational motion and curvature in depth can be perceived in situations where there are no spatial or temporal changes in optical flow across the surface of the object. We then provide a simple compuational model which explains how information obtained from patterns of optical flow at surface boundaries can indicate that such object motion is occurring. Our analysis starts with observations about the patterns of flow at boundaries and over surfaces that can be discriminated in practice. We then show that these discriminable patterns lead to qualitative constraints on the kinds of object motion that can be occurring. When appropriately combined, these constraints predict the motion and surface shape that is actually seen.

2 Boundary effects can cause the perception of rotation

The *only* visual cues for rotation in depth described in the computational vision literature involve spatial variations in optical flow or discrete point correspondences.[1] These computational models are thus unable to explain illusions that have been reported in which rotation in depth can be seen when there is no spatial variation in flow over the surface which appears to be part of a rotating object. Kaplan (1969), in a study of

[1] "Rotation in depth" involves object rotation around an axis perpendicular to the line of sight

the depth effect induced by the dynamic accretion or deletion of surface texture at a boundary, reported that when accretion/deletion was occurring on *both* sides of a boundary most subjects saw what appeared to be surfaces moving over a roller. All of Kaplan's displays were generated using uniform translational motion and hence contained none of the traditional image cues for rotation. Royden et al. (1988) briefly described a similar effect, suggesting that it involved an interaction between optical flow perpendicular to some region boundaries and parallel to others. Ramachandran et al. (1988) showed that two transparent planes of dots superimposed on each other and moving in opposite directions at constant velocities appeared as a rotating cylinder when viewed through the appropriate aperture.

Ramachandran et al. (1988) reported on a series of experiments in which moving dots consistent with actual rotation in depth interact with boundary effects. For example, they reported that when a pattern of dots representing a rotating, transparent, 3-dimensional cylinder was viewed through a triangular aperture, it appeared to look like a 3-D cone. They concluded that "...the *segmentation boundaries* that delineate the object in motion (i.e., the edges of the triangular window) seem to have a strong influence on the magnitude of perceived depth."

The same paper went on to describe another effect that may be even more significant. The display consisted of dot patterns consistent with two, vertically oriented transparent cylinders undergoing a short clockwise-counterclockwise, in-phase rocking motion. When sufficiently distant horizontally, the two dot clusters were seen as separate, rocking cylinders. When the dot clusters were moved towards each other until they almost touched, they appeared to fuse into a single larger cylinder rocking on an axis centered between the two constituent clusters. Ramachandran et al. observed that this effect is striking for two reasons. First of all, even though a single cylinder was seen there was no single rigid interpretation possible for the pattern as presented. Secondly, the range of depth perceived when the dot clusters were close was significantly greater than what would be predicted by an analysis of the clusters in isolation.

Neither Kaplan, Royden et al., nor Ramachandran et al. provided a computational model for the processes that might be involved in the effects described above.

Figure 1 illustrates an illusion that we investigate quantitatively in the next section. It is similar to that described in Royden et al. (1988). This stimulus captures the essential features that a computational model must address. A random dot kinematogram is divided up into two regions – a vertically oriented central slit and a larger surrounding region. Random dots are rendered onto both regions. Dots in the central region move horizontally in one direction. Dots in the outer region move horizontally in the opposite direction. Dots moving towards a boundary between regions disappear at the boundary. Dots moving away from a boundary first appear at the boundary. The boundary is otherwise not marked. (The dashed line in Fig. 1 is for illustration only.)

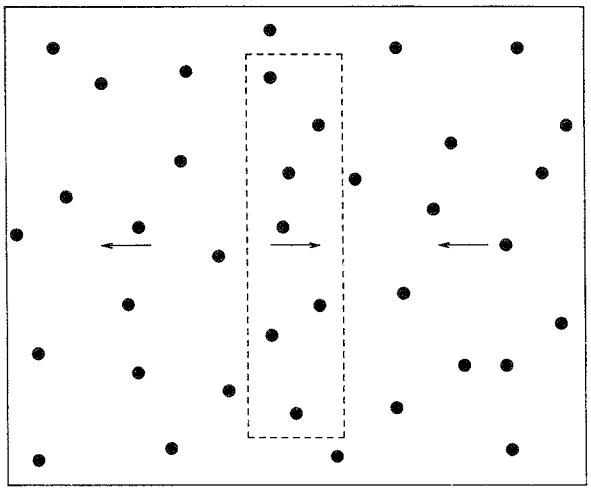

Fig. 1. Certain translational flow patterns appear as a spinning cylinder

Under appropriate conditions of slit size and dot velocities, the central region in Fig. 1 is usually seen as a cylinder rotating in front of a translating background.

2.1 Psychophysics

In order to explore conditions for which a spinning cylinder is likely to be seen in moving dot patterns such as Fig. 1, we measured subject responses for a range of inside and outside speeds, and widths.

2.1.1 Methods. Random moving dot patterns were generated with a Sun SPARCstation computer and displayed on a monochrome raster monitor. The outside square was 10.5×10.5 deg. The height of the inside rectangle was 9.48 deg, and the width varied from 0.11 to 6.05 deg in half octave steps. The dot density was 4%, and the dot size was 2×2 pixels, and subtended 0.06×0.06 deg. The speeds of both inside and outside were: 0, ± 1.44, ± 2.89, ± 4.3, ± 5.77, and ± 7.22 deg/s. All pairs of speed combinations, except for identical inside and outside velocities, were tested. The frame rate was 47.4 Hz, with less than 0.5% standard variation from trial to trial. The screen was viewed monocularly from a distance of 57 cm.

On a given trial, observers fixated a small stationary mark at the center of the screen. The moving dot pattern was shown for 4.2 s (200 frames) after which the subject had to decide whether the central rectangle appeared to be a cylinder, an open slot, a flat plane, or none of these. There was an inter-stimulus interval of 3 s, after which the next trial began. There were four observers. Two of the subjects were authors, and two were naive as to the purposes of the experiment. All had normal or corrected normal vision.

2.1.2 Results. As one would expect, when the central patch was stationary, observers reported a flat plane in front of a moving background. And when the background was stationary, observers generally saw moving central patches appear as slots. Generally, as the outside

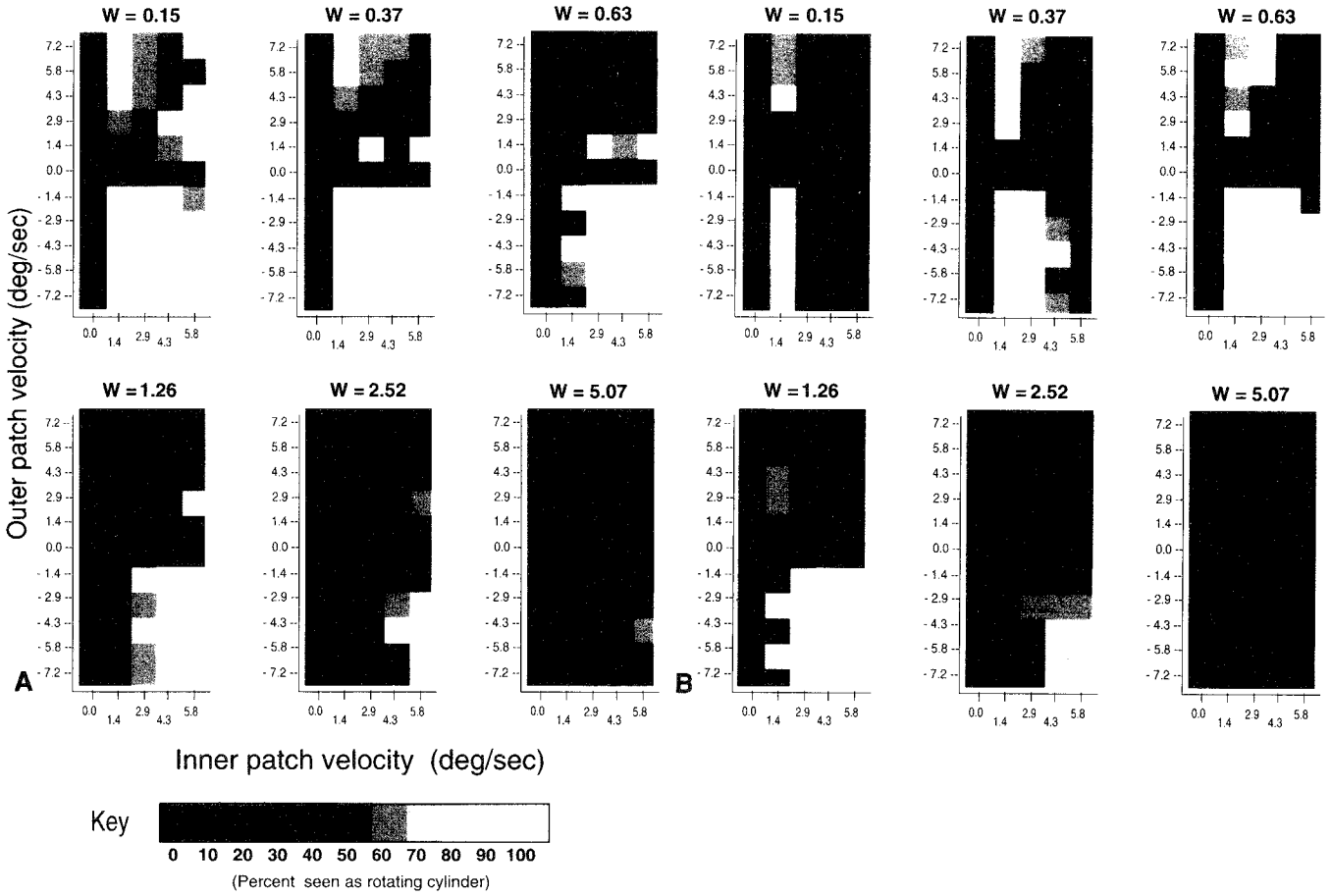

Fig. 2. Proportion of times central patch perceived as a rotating cylinder as a function of outer and inner patch velocities for six patch widths. Panel **A** shows data for naive observers. Panel **B** shows data for experienced observers

speed slowed relative to the inside, observers were likely to report seeing an open slot. Flat planes were increasingly likely to be seen when the inside speed was slow.

When both inner and outer regions were moving, observers reported seeing cylinders, flat planes or slots depending on the condition. There are widths (from 0.21 to 1.07 deg) in which cylinder judgments were reported for all speed combinations, except zero inside velocities and where the inner and outer patch speeds were the same. This includes some cases in which inner and outer velocities are in the same direction, but with different speeds. Because the case of the spinning cylinder is the major focus of this paper, we will examine the proportion of cylinder judgments in detail.

The central observations pertaining to cylinder judgments can be seen in Fig. 2A and B showing the data for naive and experienced subjects, respectively. Each figure consists of 6 panels corresponding to 6 central patch widths. Because there were negligible differences in the data between neighboring widths, the data for pairs of the original 12 widths were averaged and labeled with the geometric mean of the two widths. In each panel, the proportion of cylinder judgments averaged over two observers is shown proportional to lightness. The lightness ranges from black to white and indicates from 0 (black) to 100 (white) per cent cylinder judgments. The outer speed is shown on the ordinate, and ranges from -8 to $+8$ deg/s. The inner speed is shown on the abscissa, and ranges from 0 to 8 deg/s. Apart from the zero speed points, each square datum point represents a total of 8 judgments. (There were four judgments for each speed combination for all but the zero velocity conditions for each observer.) The lower and upper halves of a panel represent data for opposite and same velocity directions, respectively. The zero speed conditions show up as black, indicating 0% cylinder judgments. The black diagonal apparent in the upper half of the first 4 panels shows where inner and outer velocities are identical.

There are two main features of the data to emphasize. First, when the inner and outer velocities are opposing (lower half of panels in Fig. 2), spinning cylinders can be seen over the entire range of central patch widths. However, the number of cylinder judgments, averaged over all speed combinations, drops off to near zero as widths increase towards the maximum of 6 deg. This is shown in Fig. 3. The highest percentage of cylinder judgments was obtained near 0.44 deg. The error bars show the maximum and minimum range of proportions across the four observers.

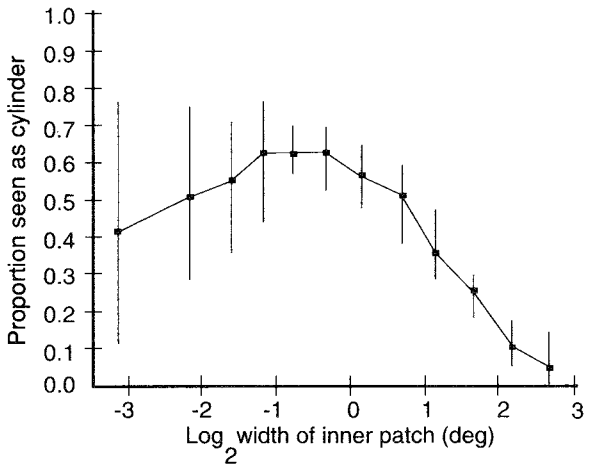

Fig. 3. Proportion of rotating cylinder judgments as a function of patch width

Second, it is also apparent from Fig. 2 that as the width of the central patch increases, higher inside speeds are required to see spinning cylinders. In Fig. 4A and B, we have plotted the proportion of cylinder judgments for naive and experienced observers, respectively. The data were averaged over the opposing outer speeds, as a function of central patch speed, for the 12 widths. The proportion of cylinder judgments gradually increases with velocity for widths above 1 deg (filled symbols). Although there are some differences between naive and experienced observers for the narrowest widths, it is clear that the inner speed corresponding to the peak shifts right as width increases.

Under the relatively short presentation times used in this experiment, a rotating cylinder was seldom seen if the outer patch was stationary. If such displays were viewed for longer intervals, however, some observers reported a bistable effect in which the central patch is sometimes seen as a flat surface viewed through a slit in the surrounding patch and sometimes as a rotating cylinder in front of the surrounding patch.

What are the essential features of the data that a computational model must address? We have seen that under appropriate conditions of slit size and dot velocities, the central region in Fig. 1 is usually seen as a cylinder rotating in front of a translating background. Further, if the width of the central region is increased, the effect is gradually reduced, to be replaced with an uncomfortable sensation that surfaces are jumping around on the screen. Since there is no differential motion within either of the regions, the sense of rotation must be due at least in part to effects of the discontinuous motion at the region boundaries. Since the effect occurs for motions that are equal in speed but opposite in direction, the cause cannot be simply occlusion cues at the boundaries. In such cases, motion is symmetric on either side of the boundary – either both surfaces are moving away from the boundary or both surfaces are moving towards the boundary. The effect even persists, at least to an extent, if a conflicting occlusion cue is introduced. When the outer region is held stationary, there is a strong occlusion cue indicating that the central region corresponds to a surface moving behind a slit in another surface corresponding to the outer region. In fact, this is often seen. The display is bistable, however, and a spinning cylinder is also sometimes seen.

In the next section, we provide a computational account for these data and those of Kaplan (1969); Royden et al. (1988); and Ramachandran et al. (1988).

3 Computational theory

If optical flow is known exactly, then a simple qualitative test can be used to determine whether or not a smooth surface is rotating in depth. First of all, we observe that spatial discontinuities in optical flow can only occur due to discontinuities in depth between two surfaces moving rigidly with respect to the observation point and/or one surface moving relative to another. Discontinuities in flow signal more than just the presence of surface boundaries, however. If the image plane motion of the discontinuity itself (the *boundary flow*) is considered along with the optical flow observed on either side, it is possible to determine information about occlusion. In particular, the *boundary flow constraint* states that the flow associated with the occluding surface immediately adjacent to the boundary will be equal

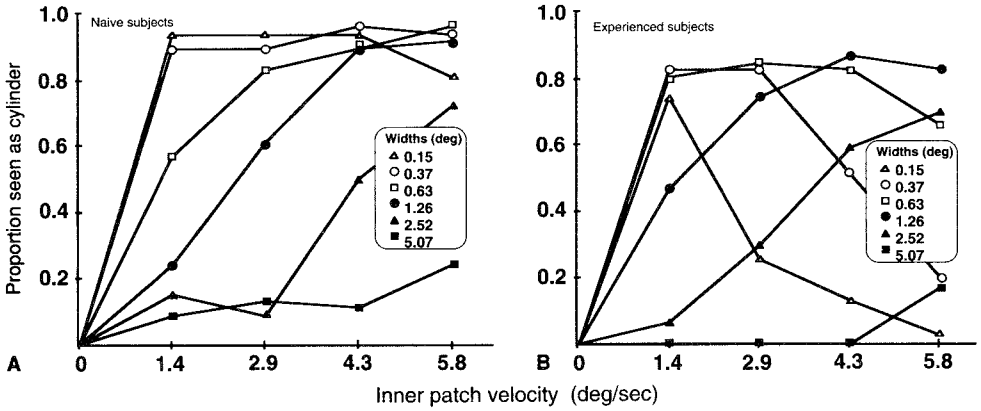

Fig. 4. Proportion of rotating cylinder judgments as a function of inner patch velocity for six patch widths. Panel A shows averaged data for two naive observers. Panel B shows averaged data for two experienced observers

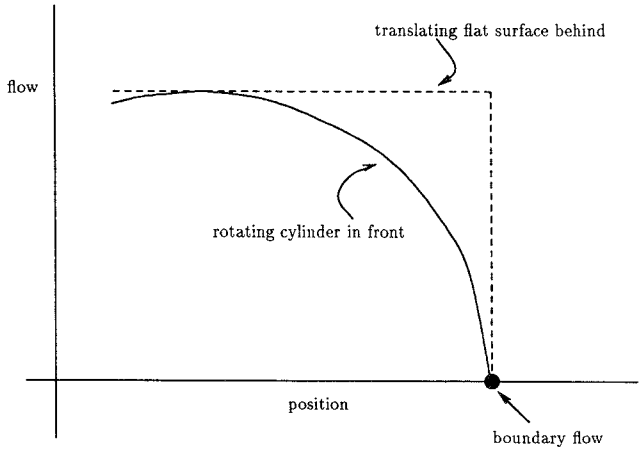

Fig. 5. Optical flow for rotating cylinders and translating flat surface

to the flow of the boundary (Thompson et al. 1985).[2] The boundary flow constraint holds whether or not the surface boundary is due to a sharp edge or to a smooth surface rotating in depth and thus progressively occluding or disoccluding itself. For sharp edges, the validity of the constraint is obvious since the motion of the boundary and the motion of the surface adjacent to the boundary are one and the same. For curved surfaces rotating in depth, the situation is more complicated. Somewhere on the curved surface will be a *generating contour* that projects to the edge of the surface region in the image. Three-dimensional surface motion at the generating contour will in general not be the same as the 3-D motion of the contour. Nevertheless, when projected into the image the optical flows are the same.

Now, consider the pattern of optical flow in Fig. 1 at and immediately to the left of the right side of the slit. Figure 5 shows two possibilities. The solid line shows the flow that would arise from a cylinder rotating in depth. Since the front of the cylinder is a (self) occluding surface, it satisfies the boundary flow constraint. In the limit as the position of the boundary is approached, the surface flow is equal to the boundary flow.[3] The dashed line shows the flow that would arise from a flat, translating surface moving behind an occluding contour. In this case, the surface flow is not equal to the boundary flow. This sort of pattern can only arise in situations where the flat surface is translating behind some other occluding surface. (For our purposes, rotation of objects with sharp occluding contours is effectively equivalent to translation, though the flow would not be constant over the slit.)

This discussion suggests that a simple test involving boundary flow and the patterns of surface flow immediately adjacent should be sufficient to distinguish whether situations involving differential optical flow correspond to translating or rotating motion. In fact, this is not the case. For curved surfaces rotating in depth, the boundary flow constraint predicts that surface flow will approach boundary flow in the limit as the surface point at which flow is evaluated approaches the generating contour for the region boundary. For surfaces with a small radius of curvature, the rapid change in flow near the generating contour may make it impossible in practice to determine that surface flow is actually approaching boundary flow.

To see this, consider the orthographic projection of a unit radius cylinder with rotation axis coincident with the y axis and rotational velocity such that the magnitude of maximum optical flow is $f_{\max}$. Over the projection of the cylinder, the optical flow at image location x is given by:

$$flow = f_{\max}\sqrt{1-x^2}$$

The derivative of the flow with respect to x is:

$$\frac{\mathrm{d}\,flow}{\mathrm{d}x} = -\frac{f_{\max}x}{\sqrt{1-x^2}}$$

For $x = 0.95$ (a position 5% of the radius from the boundary), this leads to a derivative of flow with a magnitude in excess of $3f_{\max}$. All techniques for estimating optical flow at a point explicitly or implicitly assume that flow is either constant or slowly varying within some local neighborhood around the point. These methods will have substantial difficulty dealing with flow fields which change this rapidly. In particular, it is likely that surfaces rotating in depth and with a small radius of curvature relative to the neighborhood over which flow is determined may generate apparent flow patterns distinctly different from the flow of the boundary.[4] In such situations, rotation in depth of a curved surface in front of some other surface may not be easily distinguishable from two flat surfaces translating with respect to one another.

To be useful, a computational model of motion perception must be based on image properties that can actually be discriminated in practice. As we have just argued, a pattern of optical flow near a flow discontinuity which *appears* to have values significantly different from the flow of the boundary itself can be due either to the self occlusion of a rotating smooth object or to the translation of one surface behind another. Fortunately, the situation is not completely ambiguous. By considering the flow patterns on both sides of a boundary, constraints on possible interpretations can be discovered. Figure 6 gives the possible relationships between recognizable flow patterns at a boundary and

[2] It is easiest to consider only the component of flow perpendicular to the orientation of the boundary, since flow parallel to the boundary provides no additional constraints on spatial organization unless additional information is known

[3] To determine unambiguously that the pattern shown corresponds to rotation in depth of a smooth surface, it is useful to consider both boundary flow and the accretion/deletion of surface texture on the cylinder. In particular, rotation in depth is signaled by a surface which satisfies the boundary flow constraint yet which is progressively appearing or disappearing at the boundary (Thompson et al. 1985)

[4] Quantitative predictions of the flow magnitude and radius of curvature for which this problem becomes significant requires an analysis of optical flow estimation that is beyond the scope of this article

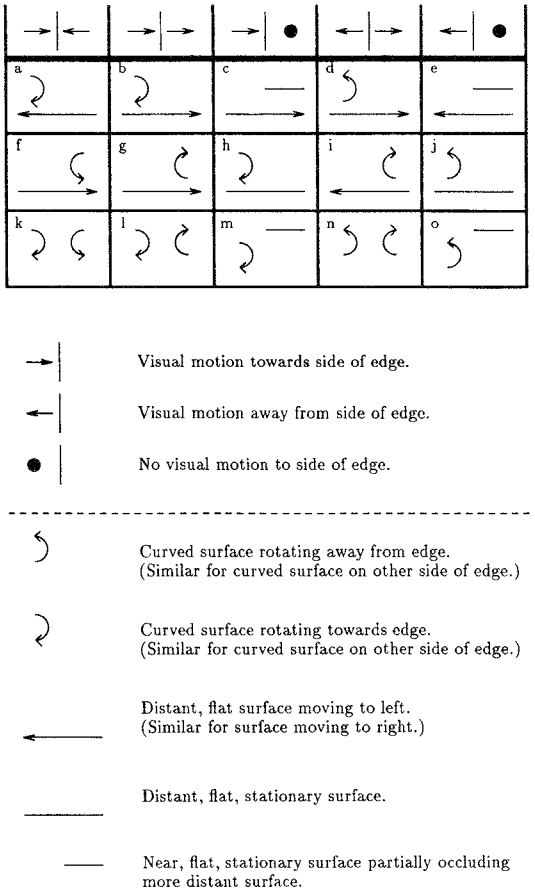

Fig. 6. Optical flow constraints on translation/rotation at a boundary

the surface motions generating the flow. The top row of the figure shows all possible combinations of flow moving towards, away from, or with a boundary. (For simplicity, the coordinate system has been chosen so that the boundary appears stationary and symmetric cases have been removed.) The columns give the possible surface motions corresponding to the flow pattern, under the assumption that rotation in depth of curved surfaces will generate a flow pattern that appears to differ from the boundary flow. Figure 6 helps to limit the combinatorics associated with determining surface motion, since certain combinations are not allowed. For example, the pattern shown in the first column cannot be due to two flat surfaces translating with respect to one another.

The effects seen in Fig. 1 can be explained by a model based on combinining two distinct visual cues:

- Patterns of flow at boundaries provide constraints on possible motions, surface orientations and curvatures, and relative depth.

- Uniform flow across surfaces provides a constraint limiting surface curvature.

(Since the model presented here is qualitative, we will not discuss the effects of absolute flow magnitude in the perception of shape and motion.)

In the case of Fig. 1, the constraints from Fig. 6 allow only possibilities **d**, **i**, and **n** at the left edge of the central region and possibilities **a**, **f**, and **k** at the right edge of the central region. The constant optical flow over the large, surrounding region gives strong evidence that this region is flat, eliminating possibilities **d** and **n** from the left and **f** and **k** from the right edge. Estimates of differential flow must be made over a neighborhood. As a result, it is less easy to determine how, if at all, flow is changing within the central region. This leads to a weaker (or non-existent) constraint indicating surface flatness for this region, allowing interpretation **i** on the left edge and **a** on the right edge to remain viable. The result is a perception in which the central region is seen as rotating in front of flat, translating regions which are visible to either side.

As the width of the central region increases, it becomes easier to determine that there is no differential flow within. The result is a constraint that the central region is likely to be flat. This is inconsistent with *any* of the possibilities in Fig. 6 and explains why the pattern appears to look so strange.

In the case where the outside region is stationary, the flow patterns at the left edge of the central region allow for possibilities **e**, **j**, and **o** in Fig. 6 (flipped left-for-right to account for the stationary surface being on the left). At the right edge, possibilities **c**, **h**, and **m** are allowed. The lack of differential flow in the outer region does not further constrain the interpretation. Weak evidence for the flatness of the surface corresponding to the central region favors possibility **e** on the left and **c** on the right, corresponding to a translating, flat surface seen through a slit in a stationary surface that is in front. Interpretations **j** on the left and **h** on the right are still possible, corresponding to the occasionally seen situation of a cylinder rotating in front of a flat, stationary surface. The interpretation with **o** on the left and **m** on the right, corresponding to a rotating cylinder seen through a slit is seldom if ever seen. Perhaps that is because this situation involves a singular and therefore unlikely viewpoint. If so, possibilities **m** and **o** should be removed from Fig. 6.

4 Discussion

As described above, current computational models of structure-from-motion are unable to account for the sense of three-dimensional objects rotating in depth produced by three related classes of displays:

- Translational motion viewed through a rectangular aperture.

For sufficiently narrow apertures, a rotating cylinder is seen. The effect occurs for both dark surroundings (Ramachandran et al. 1988) and for a surround moving with constant velocity in the opposite direction (Royden et al. 1988, and our own work).

- Cylindrical motion, view through a triangular aperture.

A spinning cone is seen, even though the motion of the dots is not consistent with such an object except at the widest portion of the aperture.

- Adjacent, rotating cylinders.

Unlike the previous two cases, this pattern does have an interpretation in terms of rigidly moving objects. This is *not* what is actually seen, however. Instead, the perception is of a single larger rotating cylinder.

The section on Computational Theory shows how our model accounts for the first of these effects. The case of the triangular aperture suggests that the vision system weights the qualitative constraints present at boundaries more than the quantitative information provided by the differential optical flow within the aperture. The case of the two adjacent cylinders is perhaps the most interesting. Ramachandran et al. (1988) suggest that the width of perceived objects affects the magnitude of perceived depth. We can now provide a more complete explanation. Local spatial gradients of optical flow provide a strong cue that the surfaces of the individual cylinders are curved, a weak cue as to the magnitude of the curvature, but little or no information about the sign of curvature or the actual surface orientation. Significant, detectable discontinuities in optical flow are seen to the outsides of the cylinder pair, but not along the line where the two cylinders touch. (The direction and magnitude of flow near this line is the same on both sides.) As a result, the surface/boundary constraints given in Fig. 6 apply only at the outsides of the merged patterns of dots. These boundary constraints, together with the spatial gradient of flow, suggests a single curved surface rotating in depth. The vision system, absent accurate information about the radius of curvature, approximates a single smooth surface between the two apparent boundaries.

A number of important open questions remain. All the displays used by Ramachandran et al. are transparent. Our own experience also suggests that overlaying dot patterns with motions of equal magnitude but opposite direction strengthens the sense of rotation. Transparency thus needs to be accounted for in a more complete computational model.[5] The sense of rotation in depth for displays such as Fig. 1 is made more compelling when the motion of the central region is made to truly correspond to a rotating cylinder. A more complete account of this effect will require a theory that embodies both the qualitative constraints we have described and more traditional quantitative structure-from-motion analysis. The qualitative constraints are important in their own right, however, since they are simple, easily and reliably computed, informative, and thus likely to be part of many possible quantitative models.

Acknowledgements. This work was supported by NSF Grants IRI-8722576 and BNS-8708532 and AFOSR contract AFOSR-90-0274. John Allman first drew our attention to the effect produced by Fig. 1. Bosco Tjan discovered the relationship between velocity and the widths at which rotation is seen. Ellen Hildreth, Suthep Madarasmi, and Albert Yonas provided valuable comments on our work.

References

Kaplan GA (1969) Kinetic disruption of optical texture: The perception of depth at an edge. Percept Psychophys 6:193–198

Prazdny K (1985) Detection of binocular disparities. Biol Cybern 52:93–99

Ramachandran VS, Cobb S, Rogers-Ramachandran D (1988) Perception of 3-d structure from motion: the role of velocity gradients and segmentation boundaries. Percept Psychophys 44:390–393

Royden CS, Baker JF, Allman J (1988) Perceptions of depth elicited by occluded and shearing motions of random dots. Perception 17:289–296

Thompson WB, Mutch KM, Berzins VA (1985) Dynamic occlusion analysis in optical flow fields. IEEE Trans Pattern Anal Machine Intell PAMI-7:374–383

Ullman S (1975) The interpretation of visual motion. MIT Press, Cambridge Mass

Wallach H, O'Connell DN (1983) The kinetic depth effect. J Exp Psychol 45:205–217

Dr. William B. Thompson
Department of Computer Science
University of Utah
3190 Merrill Engineering Building
Salt Lake City, UT 84112
USA

[5] In fact, almost all of the computational models for surface recovery yet developed fail to account for transparency (see Prazdny 1985)