CVGIP: IMAGE UNDERSTANDING
Vol. 56, No. 1, July, pp. 69–77, 1992

Qualitative Constraints for Structure-from-Motion*

WILLIAM B. THOMPSON AND JAMES S. PAINTER

Department of Computer Science, University of Utah, Salt Lake City, Utah 84112

Received March 18, 1992; accepted March 18, 1992

Existing computational models of structure-from-motion are all based on a quantitative analysis of variations in optical flow or feature point correspondences within the interiors of single objects. We present an alternative approach effective for objects rotating in depth. The method involves a set of qualitative constraints on shape and motion based on patterns of flow at surface boundaries. These constraints are used in the development of a simple approximation technique for recovering surface shape. The approach is based on large-magnitude effects that are likely to be easily extracted, even from noisy data. In addition, it explains a variety of phenomena described in the literature on human vision that cannot be accounted for by any existing computational model of structure-from-motion.

1. INTRODUCTION

The term *structure-from-motion* (SFM) has come to mean the computational techniques by which visual motion can be used to determine the spatial positions of visible surface points, the orientations and/or curvature of surfaces, and/or the relative motion of objects and camera. Except for special cases involving jointed motion, *all* structure-from-motion techniques depend on spatial and/or temporal variations over surfaces in either optical flow or feature point correspondences. 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.

Motion boundaries do in fact provide relevant information about surface shape. This information is largely qualitative in nature and thus is complementary to numerically oriented structure-from-motion methods based on properties within one surface or a single moving object. In this paper, we deal only with *rotation in depth,* the rotation of an object around an axis perpendicular to the line of sight. SFM methods which presume orthographic projection require that some amount of rotation in depth be present if the problem has a solution. The same is true for perspective projection if the angle of view is narrow and the variation in depth over objects of interest is small with respect to the distance of the object from the camera.[1] Furthermore, it has been demonstrated by several investigators that human vision exploits cues for rotation in depth that are distinctly different from those embodied in any of the structure-from-motion methods yet proposed.

We start by briefly describing results from the literature on human vision that cannot be accounted for by any existing computational model of structure-from-motion. We then propose a set of computational constraints capable of accounting for these effects. Next, we use these constraints to develop an algorithm for estimating surface shape based on boundary information and a very crude estimate of surface curvature. Finally, we demonstrate this algorithm on the original visual patterns and argue why the approach is relevant to more realistic situations.

2. BACKGROUND

The classic structure-from-motion reference is Ullman [1]. Ullman and almost all who followed adopted a presumption that input data (optical flow or point correspondences) came from a single moving object. The methods developed thus do not take into account the visual cues present at boundaries between independently moving surfaces. The *only* visual cues for rotation in depth described in this literature involve spatial or temporal variations in optical flow or discrete point correspondences. We know, however, that under some circumstances people see rotation in depth even when there is no variation in flow over the surface which appears to be part of the rotating object. Kaplan, in a study of 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

* This work was supported in by NSF Grants IRI-8722576 and IRI-9112267.

[1] Note that "rotation in depth" refers to the object, not the camera. In perspective projections, translational camera motion with respect to a distant object or surface produces the equivalent of object rotation in depth for lines of sight different from the direction of camera motion.

1049-9660/92 $5.00

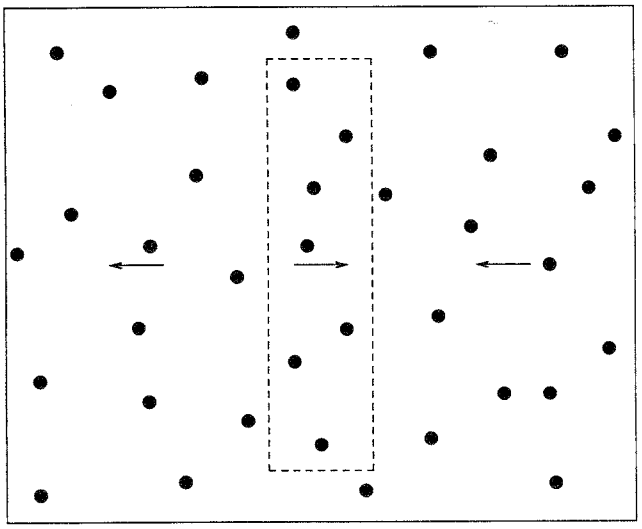


FIG. 1. Certain translational flow patterns appear as spinning cylinders.

roller [2]. 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.* briefly describe a similar effect, suggesting that it involves an interaction between optical flow perpendicular to some region boundaries and parallel to others [3].

Figure 1 shows an example similar to that described in [3]. 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 toward a boundary between regions disappear at the boundary. Dots moving away from a boundary first appear at that boundary. The boundary is otherwise not marked. (The dashed line in Fig. 1 is for illustration only.)

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 [4]. 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 a small 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.

Ramachandran *et al.* described a related series of experiments in which moving dots consistent with actual rotation in depth interact with boundary effects [5]. They argue that the perception results from the combined effects of velocity gradients in optical flow and segmentation boundaries. 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. Second, 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.

3. INFORMATION AT BOUNDARIES

Consider the pattern of optical flow in Fig. 1 at and immediately to the left of the right side of the slit. Figure 2 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 which requires that close to the boundary, the optical flow of the surface is equal to the visual motion of the boundary [6]. 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

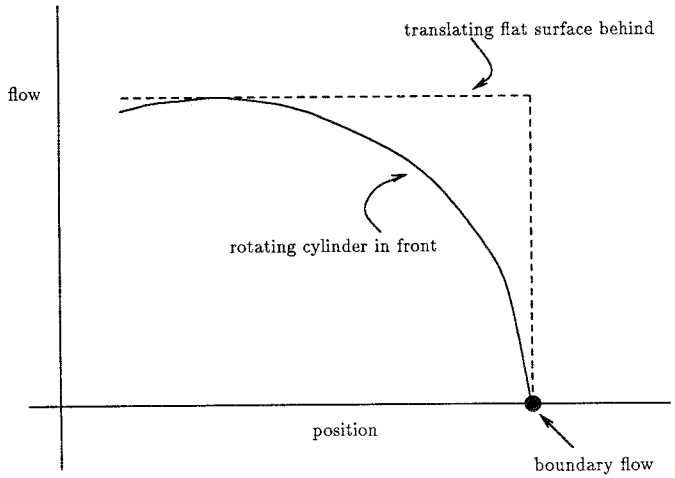


FIG. 2. Optical flow for rotating cylinder and translating flat surface.

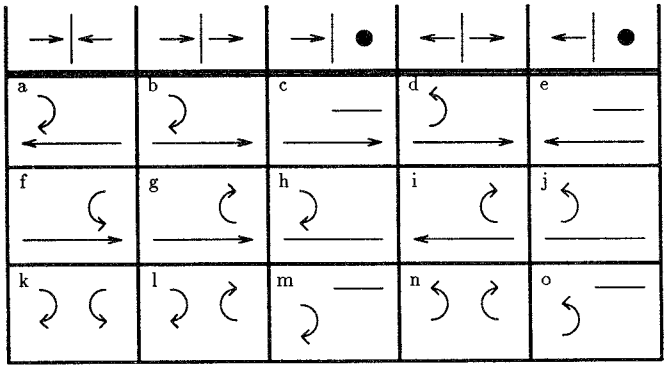
flow. This sort of pattern can only arise in situations where the flat surface is translating behind some other occluding surface.[2]

In principle, it should be possible to distinguish between rotation in depth and dynamic occlusion due to translation by comparing the visual motion of surface boundaries with the optical flow of the surfaces immediately to either side. 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. 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.

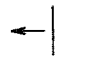
To be useful, a computational model of motion perception must be based on image properties that can actually be discriminated in practice. As we have just argued, a pattern of optical flow near a flow discontinuity which *appears* to have values significantly different from the flow of the boundary itself 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. Table 1 gives the possible relationships between recognizable flow patterns at a boundary and the surface motions generating the flow [4]. The top row of the table 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. Table 1 helps to limit the combinatorics associated with determining surface motion, since certain combinations are not allowed. For

[2] 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.

TABLE 1
Optical Flow Constraints on Translation/Rotation at a Boundary.




Visual motion towards side of edge.


Visual motion away from side of edge.


No visual motion to side of edge.


Curved surface rotating away from edge.
(Similar for curved surface on other side of edge.)


Curved surface rotating towards edge.
(Similar for curved surface on other side of edge.)

Distant, flat surface moving to left.
(Similar for surface moving to right.)

Distant, flat, stationary surface.

Near, flat, stationary surface partially occluding more distant surface.

example, the pattern shown in the first column cannot be due to two flat surfaces translating with respect to one another.

The case of pure shear motion is not included in Table 1. Motion at an edge provides information about occlusion only when there is some component of the motion normal to the edge. Pure shear motion points to a likely depth discontinuity but gives no indication of which surface might be in front and occluding the other.

Even with the constraints in Table 1, several different interpretations are possible for patterns such as those in Fig. 1. In most cases, this ambiguity can be removed by observing that near uniform optical flow over a large region is a good indicator that the surface corresponding to that region is not highly curved. Elsewhere we have presented a more detailed analysis of Fig. 1 in the context of these constraints on surface shape, orientation, and motion [4].

4. RECOVERING SURFACE SHAPE

The analysis described in Section 3 yields a set of labeled edge elements at points in the image where the optical flow is discontinuous. The case analysis show in Table 1 predicts the possible surface orientations on each side of the edge. That is, it predicts the surface normals at each edge element. In addition, the edge analysis predicts relative depth at the edges: which side of the edge is in front and which is behind. In general, the problem is highly underconstrained. There are many possible surfaces which could give rise to the observed flow data. For example, depth differences at occlusion boundaries are indeterminate and must be assigned arbitrarily. Our goal is to find the most qualitatively "plausible" surface, modulo depth differences across discontinuities.

We restrict ourselves first to the cases where only one interpretation is possible for each edge element. The general case will be handled later through repeated application of the single interpretation algorithm. Additional information *may* be provided in the form of directional second derivatives of optical flow. While this is not sufficient to determine local surface shape, it is the case that small magnitude second derivatives are associated with flatter surfaces while larger magnitude derivatives are associated with curved surfaces, though the derivatives alone are not sufficient to estimate the curvature. Since second derivatives of observed optical flow must be computed over a region, the values are reliable only at locations relatively distant from flow discontinuities. As a result, the second derivative information can be assigned an associated confidence level based on proximity to flow edges.

The surface recovery algorithm must reconstruct the depth values at each image pixel from a relatively sparse data set: normals at occlusion boundaries. One approach is based on treating the depth surface as a scalar function of the x and y coordinates of the image plane. A major problem with this approach is that for smooth surfaces, the partial derivatives of depth with respect to x and/or y are singular at occlusion boundaries where the surface normal is parallel to the image plane. Unfortunately, the input data are provided at exactly these points. In light of this problem, we have taken a parametric approach. The surface is viewed as a function of two parameters. Every point on the surface has a well defined surface normal, even at occlusion boundaries. We use the image plane coordinates as the parametric domain, in effect, equivalent to treating z as a scalar function of x and y. Nevertheless, the parameter viewpoint allows us to deal with surface normals rather than x and y partial derivatives which are not defined at the major points of interest: edge points.

Our method is composed of three phases. The first, base level determination, scans the edge element data and assigns an initial base value to each pixel determined from the front/back classification at the edges. This is accomplished via a raster scan of the pixel grid. When an edge is crossed we increment or decrement the depth by a fixed value. The sign is determined by classification of the edge. In the absence of neighboring edges, the depth is held constant. Here we assume that the edge elements form a complete four-connected boundary around the objects.

The second phase, normal interpolation, computes a normal vector at every point in the image plane. Normals are interpolated smoothly between edge elements. The normals are interpolated via a simple relaxation procedure. Each relaxation step updates the normals at non-edge samples by computing the average normals of the four connected neighbors. This process converges rapidly to a stable set of normals that smoothly interpolate the edge normals. Regions where the second directional derivatives of optical flow have small values with large confidences are presumed to be relatively flat. In these regions, a bias is introduced into the relaxation so that orientational variations in surface normals are minimized.

The final phase, surface integration, recovers the surface depth values from the fitted surface normals. Again, a relaxation method is used. Each sample is assigned a new average z value based on the old z values of its neighbors and the surface normal along the edge to that neighbor. A neighbor only contributes to the average if it is on the same surface, that is, there is not an edge between the neighbor and the sample. The surface normal along the edge is estimated by the average of the normals at the end points.

The integration process begins at the edge nodes and propagates away from the edges as relaxation proceeds. This is accomplished by weighting the samples. Initially,

the edge samples have weight 1 and all nonedge samples have weight 0. At each update step, nonedge samples are assigned a new weight which is the average weight of the neighbors which contributed to the sample. As relaxation proceeds, the weights converge to 1.

When multiple interpretations of the edge elements are possible, we exhaustively consider each of the possibilities. The edge classification assignment is assumed to be globally consistent. That is, it is never the case that two adjacent edge elements are assigned inconsistent interpretations. This limits the possible edge interpretations to a few possible cases. In our current system, we fit a surface to each possible configuration of edge interpretations. The surface is evaluated for plausibility and we choose the surface which is most plausible. A surface is deemed plausible when the maximal change in mean curvature is small. Implausible edge interpretations result in surfaces that force large changes in curvature into a small part of the surface.

5. EXPERIMENTAL RESULTS

The method presented in the previous section was applied to classes of motion fields corresponding to general situations described in [3, 5]. The examples were done as simulations to test the approach in a manner that would not be dependent on particular flow estimation and boundary detection algorithms. In all cases, a flat "background" surface was moving horizontally in one direction while the "object" surface was moving horizontally in the opposite direction. Objects were always bounded by fixed, stationary apertures. Input to each test consisted of boundaries labeled as to possible interpretations in Table 1 that were consistent with the particular simulations, together with an indication of regions in the flow field with large second derivatives. No effort was made to model the transparency used in [5].

Figure 3 shows the surface recovered from a simulation modeling a situation in which a narrow cylinder is rotating in front of a translating background and all of the edges have been assigned their correct interpretations. Figure 4 shows the surface recovered from a situation that is identical save that the edges to either side were misclassified. That is, the surface immediately to the left of the left vertical edge was presumed to be oriented parallel to the line of sight and in front of the surface to the right of the left vertical edge, which was presumed to be oriented perpendicular to the line of sight. The right vertical edge was similarly misinterpreted. Note that this case is locally consistent with the constraints in Table 1. Figure 5 is based on input in which the right vertical edge is labeled in a manner consistent with a rotating cylinder, but the left edge is mislabeled as in Fig. 4. Maximum mean curvatures for Fig. 3, 4, and 5 are 0.017, 0.158, and 0.283, respectively, indicating that the reconstruction in Fig. 3 is to be preferred. As should be clear from Fig. 3, the method recovers a reasonable interpretation of the surface in this case.

Figure 6 show the surface reconstructed from data consistent with the displays described in Royden *et al.* [3] and Thompson *et al.* [4]. (In this and all subsequent figures, only the surface associated with the edge classifications minimizing the maximum curvature is shown). In the center of the pattern, uniform translational motion is

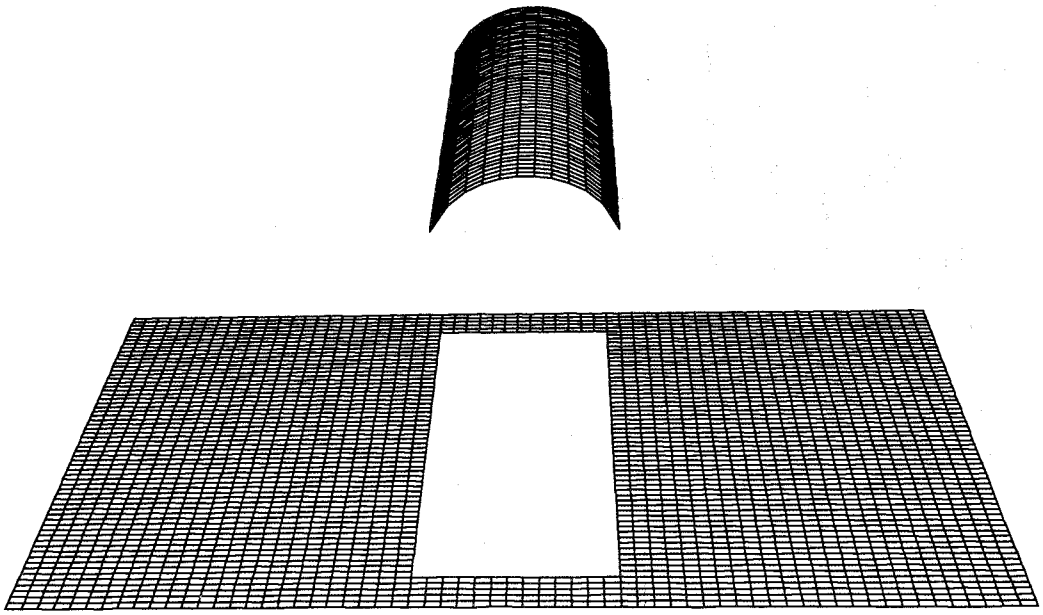


FIG. 3. Reconstruction of rotating cylinder with correctly classified edges.

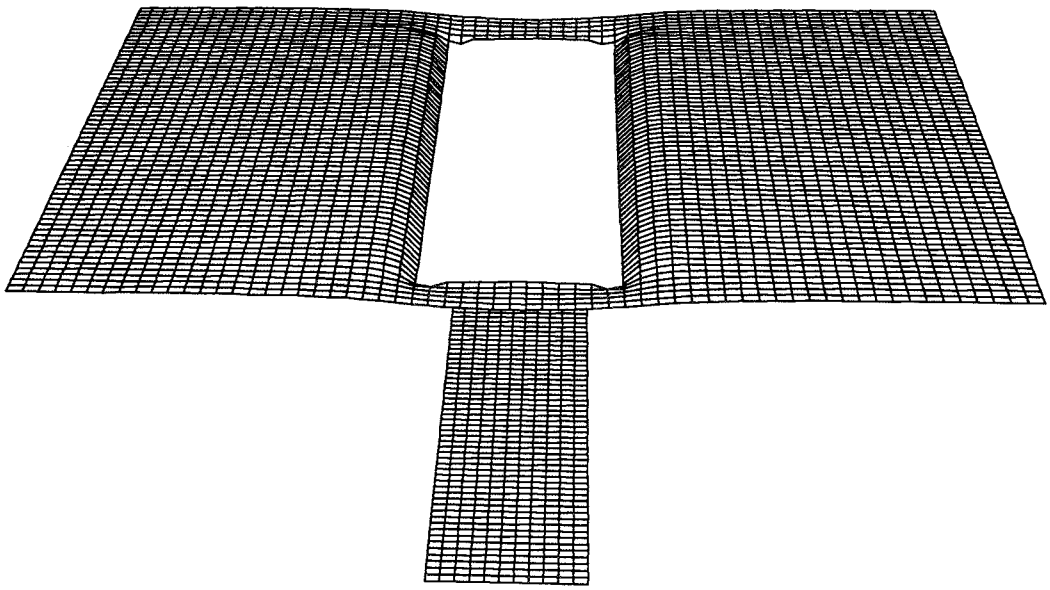


FIG. 4. Reconstruction of rotating cylinder with incorrectly classified edges.

bounded by a stationary aperture. The surrounding area contains uniform translation motion in the opposite direction. As previously indicated, when the aperture is sufficiently narrow this is seen as a rotating cylinder, even though no rigid interpretation is "mathematically" possible. Figure 6 does in fact show a recovered cylinder.

Figure 7 shows the same situation, except that the aperture is much wider. The recovered surface is rather like a conveyor belt, a perception often seen by people when presented with such displays. Figure 8 uses the same aperture width, but now the second derivatives of flow are consistent with a curved object. With this addition, a cylindrical surface is recovered. This demonstrates that the method can make use of quantitative information about local surface shape.

Figure 9 shows the surface recovered from data indi-

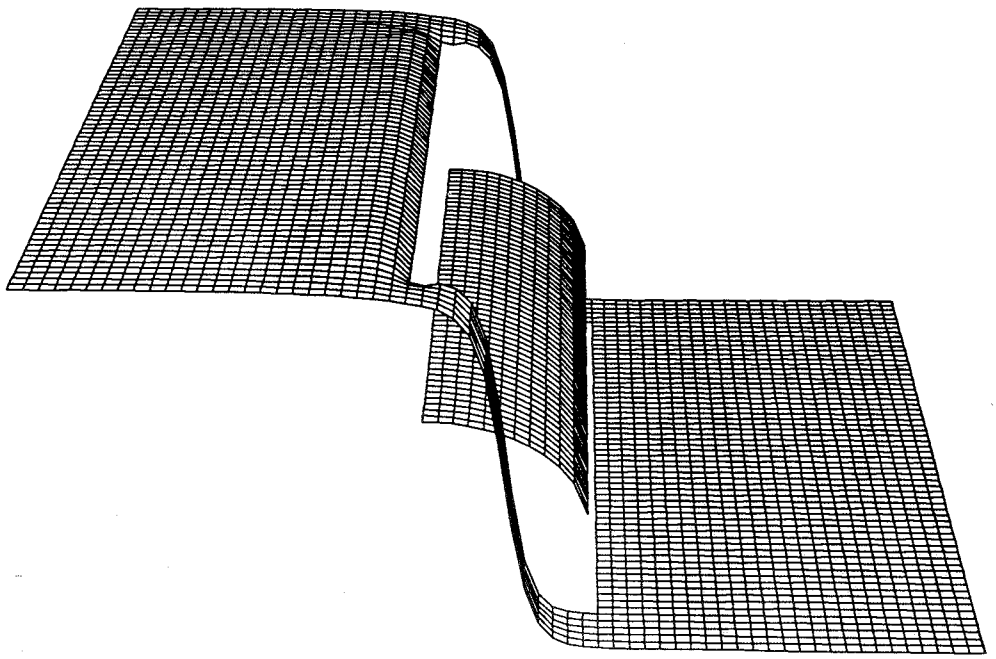


FIG. 5. Reconstruction of rotating cylinder with right edge correctly interpreted but left edge misclassified.

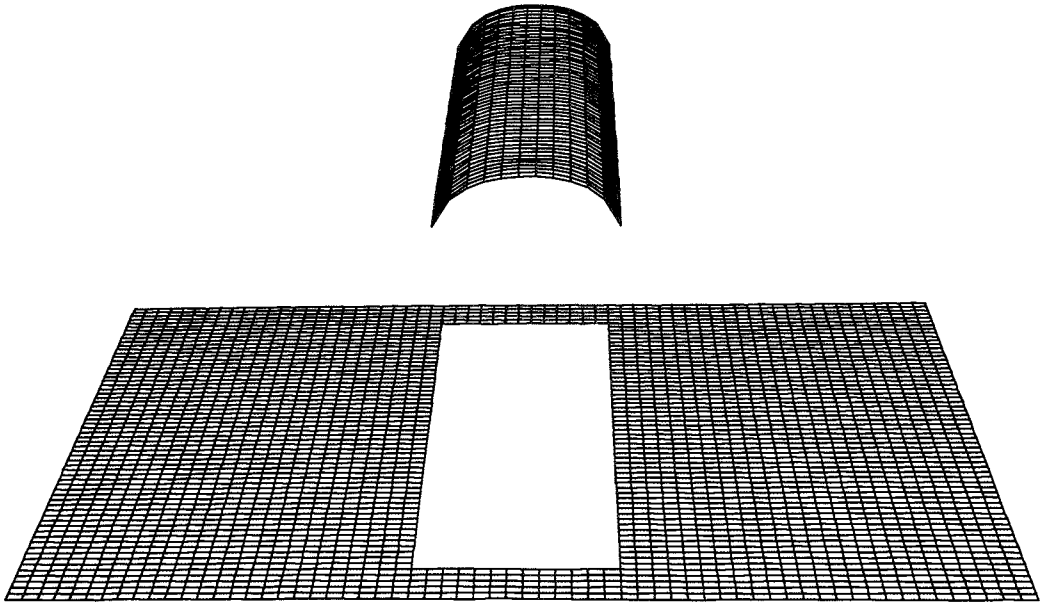


FIG. 6. Reconstruction of translating surface bounded by a narrow aperture.

cating a rotating cylinder surface viewed through a triangular aperture [5]. While the second derivatives indicate no change in curvature vertically, the simple qualitative nature of the recovery algorithm has no difficulty in generating a cone shaped surface. The method uses information about surface curvature in this example, but is not bothered by the fact that the changes in flow are quantitatively inconsistent with the curvature of the cone.

A demonstration involving. Ramachandran *et al.*'s adjacent cylinders would produce essentially the same surface as shown in Fig. 8. In this situation, the edge information disappears where the two cylinders touch, since the flow is continuous at that point. Over most of the region corresponding to the cylinders, second derivatives of flow are high enough to indicate a curved surface. As with perception by humans, the actual amount of curvature in the reconstruction is determined by the nature and location of the edges, not local variations in optical flow.

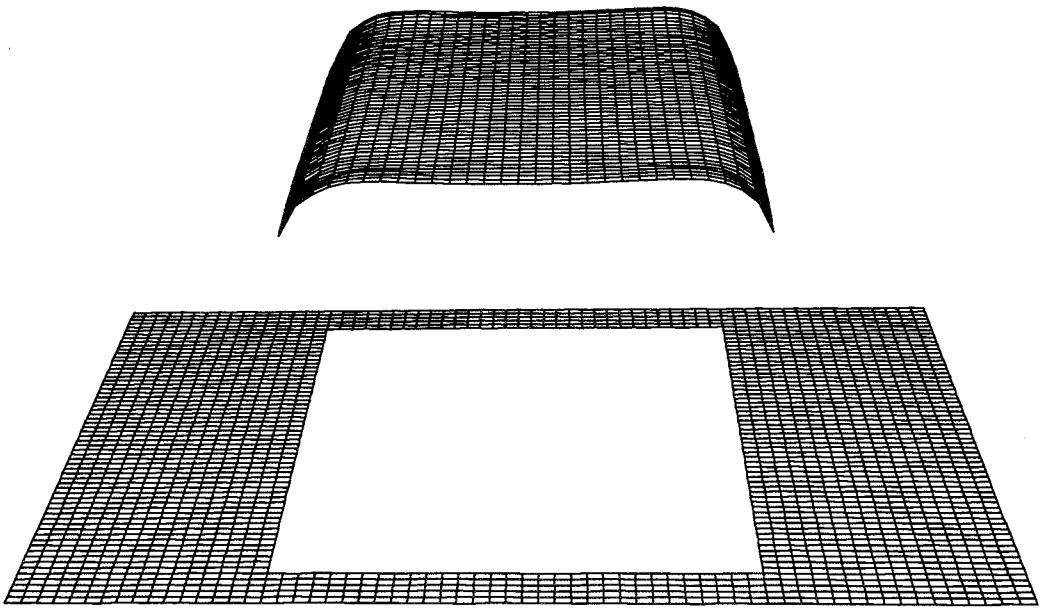


FIG. 7. Reconstruction of translating surface bounded by a wide aperture.

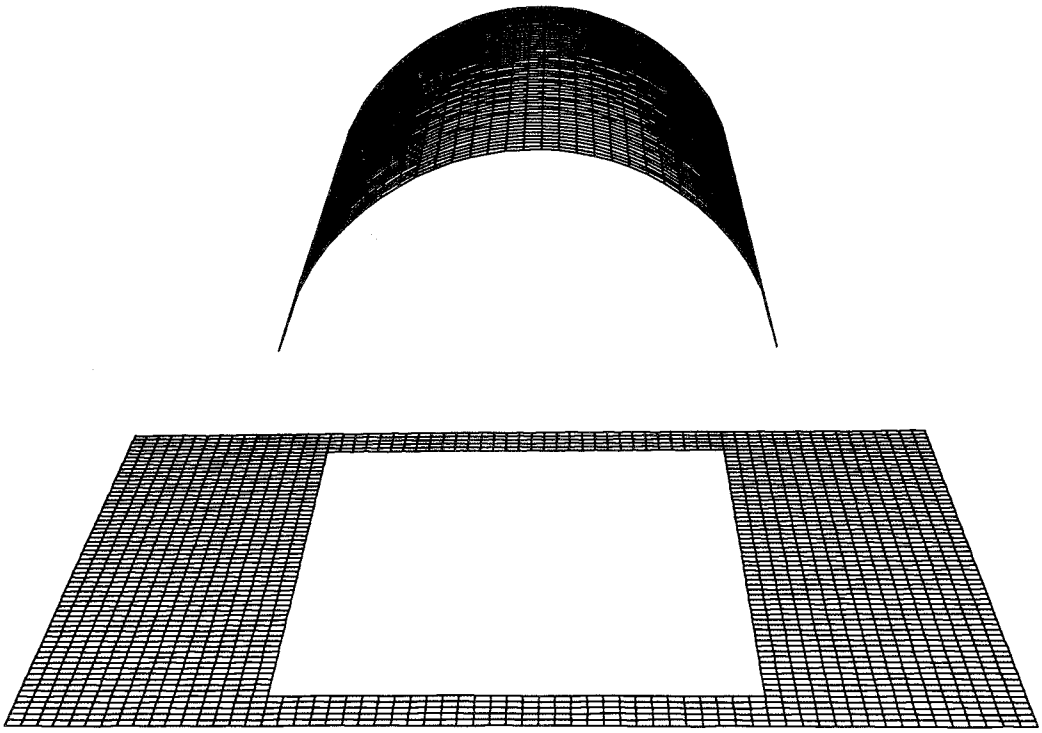


FIG. 8. Reconstruction of rotating cylinder surface bounded by a wide aperture.

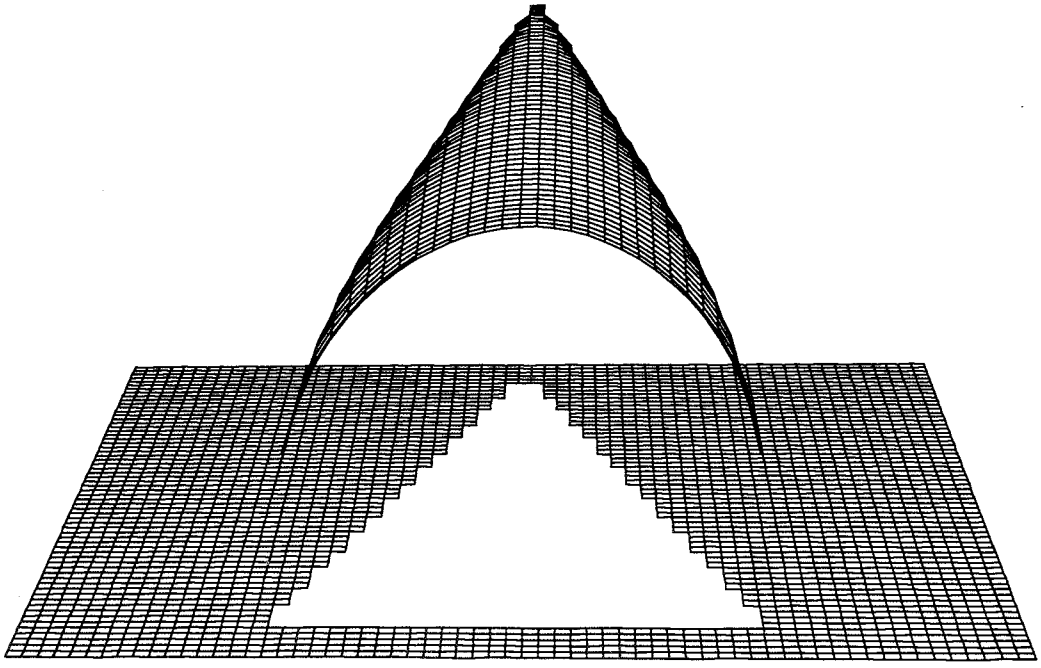


FIG. 9. Reconstruction of rotating cylinder surface viewed through a triangular aperture.

6. DISCUSSION

The above results provided an existence proof that very simple qualitative properties of edges in optical flow fields can be used to reconstruct the surface shape of moving objects. These qualitative constraints account for a number of illusions described in the perceptual psychology literature which involve a moving dot pattern being seen in a manner inconsistent with predictions made by standard structure-from-motion methods. Rather than being just a curiosity, we argue that these illusions are a consequence of a visual system evolved and adapted to be effective in more realistic circumstances. Flow patterns at edges involve large-magnitude effects that should be obtainable with greater reliability than the subtle spatial or temporal changes used by classical structure-from-motion algorithms. Thus, it make good engineering sense to anchor surface reconstruction on these data.

The method described here has as yet been tested only on simulated data. Our goal was to demonstrate the plausibility of the approach without confounding the evaluation with effects due to errors in detecting, localizing, and classifying flow boundaries. Efforts are now underway to experiment with real image sequences, though results will likely depend on improvements in the ways motion edges can be detected.

We make no claims that the particular surface reconstruction algorithm presented here is the best approach. It does deal with singularities in surface slope which are difficult for most methods based around functional approximation of depth (range). It is also able to mix constraints on relative depth, surface orientation, and surface curvature. Likewise, we make no claims that the methods described in this paper are sufficient for more complex motion analysis problems. Our use of the second directional derivative of flow as a surface curvature measure is obviously naive. The best approach to the recovery of surface shape and motion, we feel, will come from a combination of qualitative constraints such as we have described and traditional quantitative methods. In fact, such constraints may make feasible the use of otherwise noise sensitive methods such as [7] and a host of similar techniques.

REFERENCES

1. S. Ullman, *The Interpretation of Visual Motion,* MIT Press, Cambridge, MA, 1979.
2. G. Kaplan, Kinetic disruption of optical texture: The perception of depth at an edge, *Perception Psychophys.* **6**(4), 1969, 193–198.
3. C. Royden, J. Baker, J. Allman, Perceptions of depth elicited by occluded and shearing motions of random dots, *Perception,* **17,** 1988, 289–296.
4. W. Thompson, D. Kersten, and W. Knecht, Structure-from-motion based on information at surface boundaries, *Biol. Cybernet.,* 1992, 327–333.
5. V. Ramachandran, S. Cobb, and D. Rogers-Ramachandran, Perception of 3-d structure from motion: The role of velocity gradients and segmentation boundaries, *Perception Psychophys.* **44**(4) 1988, 390–393.
6. W. Thompson, K. Mutch, and V. Berzins, Dynamic occlusion analysis in optical flow fields, *IEEE Trans. Patterns Anal. Mach. Intelligence* **PAMI-7,** 1985, 374–383.
7. H. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image, *Proc. R. Soc. London B* **208,** 1980, 235–397.