Theoretical and Biological Limitations on the Visual Perception of 3D Structure from Motion

James T. Todd

Department of Psychology* and The Center for Cognitive Science^
The Ohio State University

Todd, J. T. (1998) Theoretical and biological limitations on the visual perception of three-dimensional structure from motion. In T. Watanabe (Ed.) High-level motion processing- computational, neurophysiological and psychophysical perspectives. Cambridge MA: MIT Press, pp 359-380.

Correspondence Address : James T. Todd
Department of Psychology
142 Townshend Hall
The Ohio State University
Columbus, OH 43210

Theoretical and Biological Limitations on the Visual Perception of 3D Structure from Motion

One of the most powerful sources of visual information about the 3-dimensional (3D) layout of the environment is provided by the systematic transformations of optical stimulation that occur when objects are observed in motion. The importance of motion for 3D shape perception was demonstrated over 40 years ago in a classic series of experiments by Wallach & O'Connell (1953). They showed observers the projected shadows of wireframe figures that were placed on a turntable between a point light source and a translucent display screen. When the turntable was stationary, the shadows appeared as 2-dimensional patterns, but as soon as it was set in motion, the shadows could appear suddenly to pop out in depth as solid objects rotating rigidly in 3-dimensional space.

One of the earliest theoretical analyses of this phenomenon was developed by Ullman (1977, 1979, 1883, 1984). He noted that any rotary displacement of a set of points can be decomposed into a rotation about an axis in the image plane followed by a rotation about the line of sight, and he was able to prove that for any 2-frame motion sequence under orthographic projection, it is possible to remove the rotation about the line of sight to produce a pattern of parallel image trajectories (see Figure 1). This provides a potentially useful test for distinguishing rigid from nonrigid motion. Although it is mathematically possible for physically nonrigid motions to produce parallel image trajectories following a rotation about the line of sight, the probability of encountering such an event in natural vision is vanishingly small.

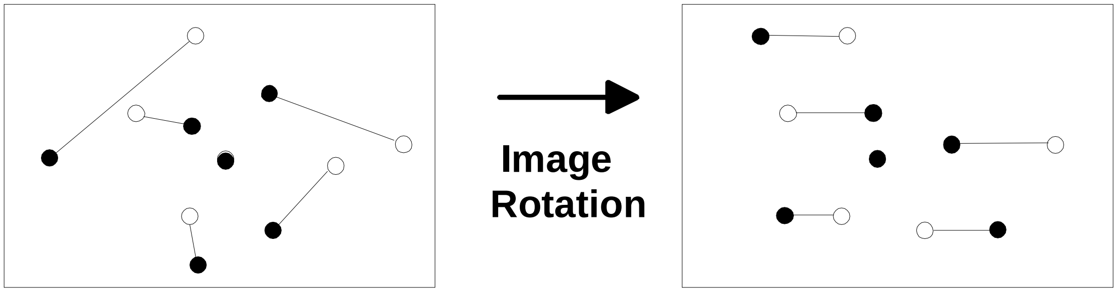

Figure 1 -- For any pair of images of a rigid object rotating in depth under orthographic projection, it is possible to produce a pattern of parallel trajectories by rotating one image with respect to the other. The connected pairs of open and filled circles represent corresponding points in different views of an apparent motion sequence

Whereas the first order relations between two distinct views provide sufficient information to distinguish rigid from nonrigid motion, they are inherently ambiguous with respect to an object's 3-dimensional structure. Ullman (1977, 1983) proved that an arbitrary 2-frame motion sequence under orthographic projection has an infinite 1-parameter family of possible 3D interpretations. In order to obtain a unique computation of Euclidean metric structure, the motion sequence must contain a minimum of three distinct views of at least four points. These theoretical limits define an absolute upper bound on what can be computed from pure motion information -- even for an ideal observer who can measure the projected position of each point and perform necessary mathematical operations with perfect accuracy.

It should also be noted within this context that the human visual system does not have infinite precision, and it is reasonable to expect that the performance of actual human observers might fall somewhat short of what is theoretically possible if all sources of measurement error could be eliminated. There is one aspect of visual sensitivity that is particularly relevant in this regard. It is important to keep in mind that a unique interpretation of 3D structure from motion requires the detection of higher order relations among three or more views of an apparent motion sequence, but there is a considerable body of evidence to indicate that our visual sensitivity to these higher order relations is extremely imprecise. For example, typical Weber fractions for the detection of acceleration are in the range of .2 to .3 (see Todd, 1981; Snowden & Braddick, 1991; Werkoven, Snippe & Toet, 1992).

One likely implication of such findings is that the perception of 3D structure from motion may be primarily based on first order spatio-temporal relations to which the human visual system is much more sensitive. Weber fractions for velocity or displacement are typically below .05 -- almost an order of magnitude lower than those obtained for the detection of acceleration (McKee & Welch, 1985; Snowden & Braddick, 1991). There are several strong predictions that follow from this hypothesis about which tasks are theoretically possible based solely on first order relations, and which ones are not. In the present article, we will examine the empirical support for these predictions among the numerous response tasks that have been employed in the literature to measure observers' perceptions of 3D form.

The Perception of Rigidity

When Ullman's structure from motion theorem was first published in 1979, it was widely believed that a perceptually compelling kinetic depth effect would be theoretically impossible from the first two frames of an apparent motion sequence, and that it must be built up gradually as an object's motion is observed over an extended period of time (e.g., see Ullman, 1984). Subsequent research has shown that that the first of these beliefs is unequivocally incorrect. A 2-frame apparent motion sequence can appear quite clearly as a rigid object rotating in depth. This was first demonstrated by Lappin, Doner and Kottas (1980) for objects viewed with an exaggerated polar perspective, though similar effects have also been reported for objects viewed under orthographic projection (e.g., see Braunstein, Hoffman, Shapiro, Andersen & Bennett, 1987; Todd, Akerstrom, Reichel & Hayes, 1988).

There are a number of important stimulus parameters that can influence the perceived rigidity of a 2-frame apparent motion sequence. The most compelling impressions of structure from motion are obtained when the two frames are presented in continuous alternation, so that observers can take as much time as necessary to process the available information. Two-frame sequences work best when the stimulus onset asynchrony (SOA) between frames is 200 msec or more, and there is a limited range of image displacements for each visible point. The displacements must be large enough to be detected, but not so large that it becomes difficult to identify the correspondence relations across successive views (see Todd et. al. 1988). Another important factor that can influence the perception of structure from motion is the structural configuration of the depicted pattern. Todd et. al. (1988) found that the most perceptually coherent 2-frame sequences were those that depicted patterns of connected lines or dense configurations of dots on smoothly curved opaque surfaces, and that coherence was significantly decreased for moving patterns composed of random dots in a volume.

Given the ability of human observers to perceive rigid rotation in depth from 2-frame motion sequences, it is reasonable to question the extent to which they are able to make use of additional information provided by the higher order spatiotemporal relations among three or more views. Several researchers have attempted to address this issue by asking observers to detect the presence of nonrigid deformations in apparent motion sequences composed of varying numbers of discrete frames. The results of this research have been somewhat contradictory, in that some investigators have reported systematic improvements with increasing numbers of views (Todd, 1982; Braunstein, Hoffman & Pollick, 1990) whereas others have not (Todd & Bressan, 1990).

There are a number of subtleties that need to be considered in interpreting the results of such experiments. It is important to keep in mind that the nonrigid deformations employed these studies were all theoretically detectable from 2-frame sequences using the technique described in Figure 1. In order to determine the expected discrimination performance of a strictly 2-view analysis, it is necessary to consider the potential impact of all possible pairs of views in the entire apparent motion sequence. If the probability of detecting the deformation in a 2-frame display is p, then the probability of detecting it in n independent pairs would be $\{1 - (1-p)^n\}$, which increases rapidly with the magnitude of n. Thus, one could only conclude that observers are exploiting higher order spatiotemporal relations if performance increased with the number of views at a rate that exceeds the predicted result based on simple probability summation.

An alternative procedure for addressing this issue has recently been developed by Norman and Todd (1993) and Perotti, Todd and Norman (1996). They investigated a special class of nonrigid deformations in which all points move in parallel image trajectories. The advantage of this approach is that displays can be presented with arbitrarily long apparent motion sequences, yet the structural deformations of the depicted objects are inherently undetectable using a 2-view analysis of structure from motion applied to any arbitrary pair of views in the sequence.

In order to demonstrate the extent of nonrigidity in any display where the points all move in parallel directions, it is useful to employ a tolerance analysis developed by Hogervorst, Kappers and Koenderink (1996). The analysis begins by measuring the horizontal positions of two points over a sequence of successive frames relative to a third point. These are then plotted in phase space, such that the relative positions of each point are represented along orthogonal axes. If the trajectory in this phase space is anything but an ellipse centered on the origin, then the motion has no possible rigid interpretation. The left panel of Figure 2 shows an elliptical phase space trajectory for a set of points that are rotated rigidly about a vertical axis in the image plane. The right panel, in contrast, shows the phase space trajectory for a set of points that rotate about the same axis at different frequencies. Note in that case that the trajectory deviates markedly from an ellipse, thus indicating that the depicted pattern of motion is nonrigid.

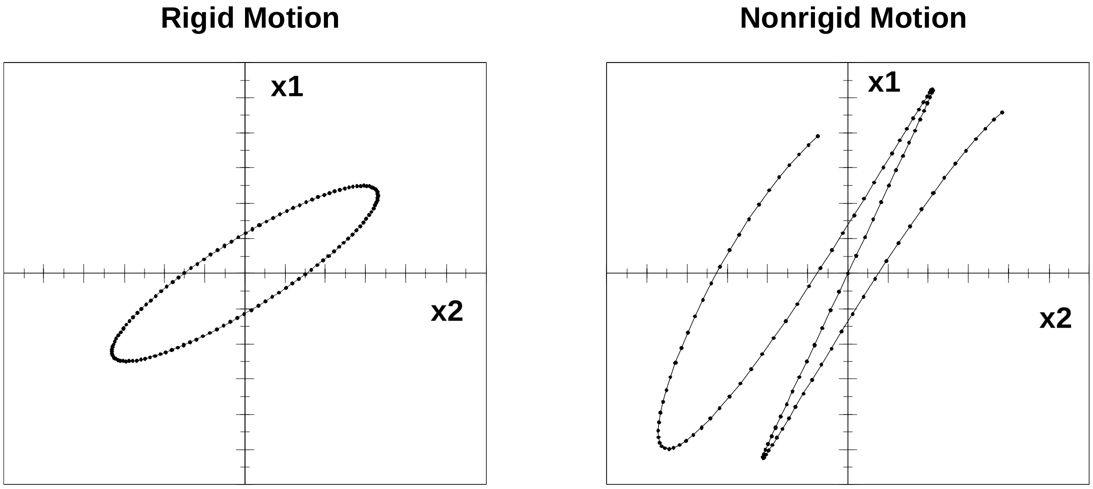

Figure 2 -- Example phase space trajectories for two different conditions investigated by Perotti, Todd & Norman (1996). Each plot represents the horizontal positions of two points x1 and x2 relative to a third point. The pattern of projected motion will only have a mathematically possible rigid interpretation if its phase space trajectories are elliptical and centered on the origin. Note that the diagram on the left exhibits these characteristics, while the one on the right does not. The right panel depicts the phase space trajectories for a nonrigid configuration of points that all rotate at different angular velocities.

Norman and Todd (1993) and Perotti, Todd and Norman (1996) have examined a variety of procedures for generating nonrigid motion patterns in which the projected displacements of all points are parallel to one another. When presented to human observers, some of these patterns can indeed be identified as nonrigid for long apparent motion sequences if the structural deformation is sufficiently large. Since nonrigid motions with parallel image displacements cannot be detected using a 2-frame analysis of structure from motion, this finding provides clear evidence that human observers are able to make use of higher order spatiotemporal relations among three or more views. It is also important to note, however, that for any given amount of structural deformation, nonrigid deformations that are detectable with a 2-frame analysis appear much more nonrigid than those that are not. Moreover, there are other displays with parallel image displacements that appear perfectly rigid, even though the depicted structural deformations are quite large. When considered as a whole, these findings provide strong evidence that human observers have only a limited sensitivity to higher order spatiotemporal relations, and that judgments of rigidity are based primarily on first order relations between pairs of views.

First-Order Information about 3D Structure

In general, if a 2-frame motion sequence passes Ullman's (1977) test for rigidity, it will have an infinity of possible rigid interpretations (see also Bennett & Hoffman, 1986; Hoffman & Bennett, 1985, 1986; Huang & Lee, 1989; Koenderink & van Doorn, 1991). Nevertheless, there is considerable information in the first-order spatiotemporal relations of a moving pattern that constrains its possible interpretations to a 1-parameter family of structures. The simplest way of demonstrating this is to begin with the analysis of Ullman (1977), in which any rotary displacement under orthographic projection can be decomposed into a rotation about an axis in the image plane followed by a rotation about the line of sight. By rotating each image appropriately with respect to the other, the latter of these components can be removed to produce a pattern of horizontal image displacements as shown in Figure 1. In other words, it is possible through image rotation to reduce the analysis of any 3D rotary displacement to the special case of rotation about a vertical axis in the image plane.

Let us now examine this special case in a bit more detail. Consider a point p(x,y,z) and its orthographic projection p'(x',y') in the image plane. If p is rotated in depth by an angle θ about a vertical axis, then p' will be displaced horizontally by a magnitude d'. For any given value of θ, this results in a mapping of R^3 onto R^3, such that (x,y,z) is transformed into (x',y',d') by the following set of equations:

$$x' = x$$

$$y' = y$$

$$d' = x\,(1 - \cos(\theta)) - z\,\sin(\theta)$$

Note in particular from the form of these equations that the pattern of image displacements (x',y',d') is an affine transformation of the true physical structure (x,y,z). It follows, therefore, that any aspect of an object's 3D structure that is affine invariant will be optically specified within the first-order pattern of image displacements under orthographic projection.

The exact nature of this affine transformation between an object's 3D structure and its pattern of image displacements depends on a variety of factors. If the angle of rotation θ is sufficiently small, then the image structure (x',y',d') will be related to the 3D structure (x,y,z) of the depicted object by a stretching transformation along the line of sight. Consider, for example, a rotating ellipse that is aligned parallel to the z-axis as shown in the left panel of Figure 3. For small values of θ, the image displacement d' plotted as a function of horizontal position x' will form an ellipse parallel to the d'-axis as represented in the middle panel of Figure 3. The overall range of image displacements in that case will vary proportionally with an object's extension in depth, but it will also vary with the magnitude of θ, so that changes in depth on the first order pattern of image motion cannot be distinguished from changes in the angle of rotation.

For large values of θ, the mapping from (x,y,z) to (x',y',d') involves a shearing transformation in addition to a stretch along the line of sight. The right panel of Figure 3 shows image displacement plotted as a function of position for the same rotating ellipse as in the previous example, but with a large angle of rotation. Note in this case that the elliptical pattern of displacements is slanted relative to the d'-axis. That is the effect of the shearing transformation. It is important to keep in mind, however, that in psychophysical experiments on 3D structure from motion the angular rotation between successive frames of a motion sequence is almost never more than a few degrees. Because there is only negligible shear under those conditions, the pattern of image displacement closely approximates the 3D structure of the depicted object up to an affine stretching transformation along the line of sight.

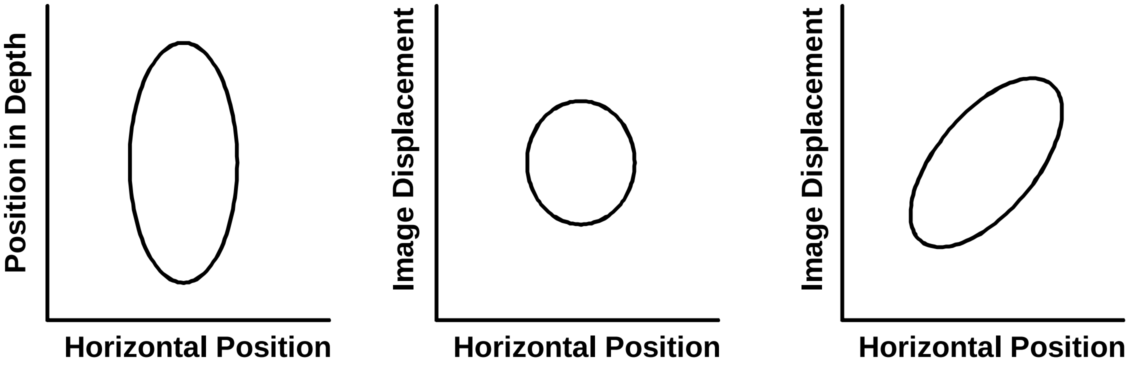

Figure 3 -- When an object rotates in depth under orthographic projection, its pattern of image displacement is related to its 3D structure by an affine transformation. If the angle of rotation between successive views of a motion sequence is sufficiently small, then the pattern of displacement will be related to the 3D structure by a stretching transformation along the line of sight. However, for large angles of rotation this relation will also involve a shearing transformation. The left panel depicts the instantaneous structure of a rotating ellipse, whose major axis is parallel to the line of sight. The middle and right panels show possible patterns of image displacement for small and large angles of rotation, respectively.

Discrimination of 3D Structure

It follows from this analysis that some aspects of 3D structure are discriminable from the first-order pattern of image motion, whereas others are not. If we restrict the range of possible interpretations to small angle displacements as are typically used in most psychophysical investigations of 3D structure from motion, then an object's shape can be determined up to an affine stretching transformation along the line of sight. When observing objects in natural vision, the probability of them being related in this manner is vanishingly small. Thus, it should be possible to discriminate almost any pair of 3D structures from their first-order patterns of image displacement, unless they are intentionally constructed to be perceptually ambiguous.

Todd and Norman (1991) have performed an experiment that was designed specifically to test this prediction. The simulated viewing situation was similar to that shown in Figure 4. Each display depicted an ellipsoid surface rotating in depth, and observers were required to discriminate whether it appeared expanded or compressed relative to a standard sphere. The surfaces were also occluded by a circular aperture so that the judgments could not be based on the changing outer boundary of its projection. For some of the displays, the objects were initially oriented so that the axis of compression or expansion was parallel to the line of sight, and they were rotated back and forth in an apparent motion sequence composed of eight distinct frames with a frame-to-frame angular displacement of 4° or 6°. For other displays, the motion sequences were limited to only two frames -- either the first two or the last two from the 8-frame sequences. When the depicted axis of compression or expansion in these 2-frame displays, was slanted relative to the line of sight, observers' discrimination thresholds were comparable to those obtained for the 8-frame displays. Performance deteriorated dramatically, however, when the axis of compression or expansion in a 2-frame sequence was parallel to the line of sight.

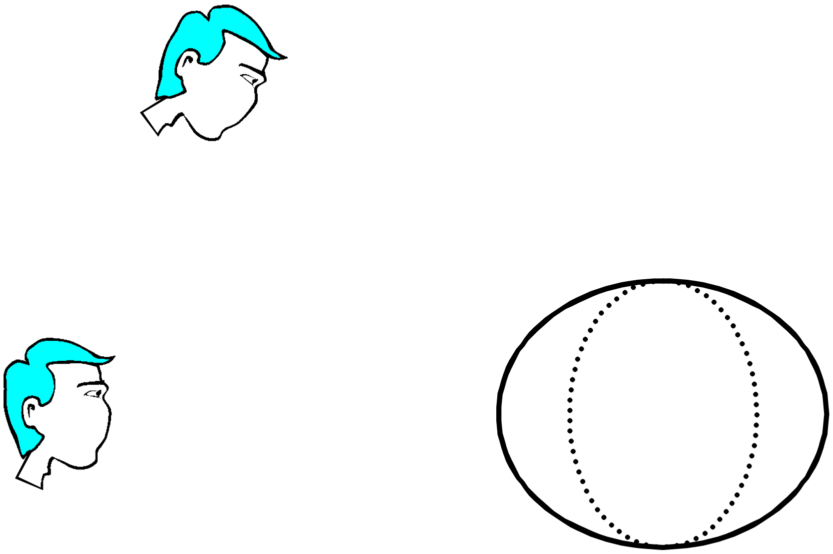

Figure 4 -- Two rotating ellipses viewed from different orientations. When seen from the lower position, the two ellipses are related by a stretching transformation along the line of sight. From the first order pattern of image displacements, the difference in their structure in this case cannot be distinguished from two identical ellipses rotated with different angular velocities. This ambiguity is eliminated, however, when the objects are viewed from the upper position.

This is exactly the result that would be expected if the perceptual analysis of structure from motion were limited to first-order relations in the overall pattern of image displacements. When the ellipsoids were aligned with the viewing direction as shown in Figure 4, the changes in image structure produced by compression or expansion were mathematically indistinguishable from those produced by the variations in angular displacement. When the objects were slanted with respect to the viewing direction, on the other hand, then the optical effects of these two manipulations were quite different.

It is also interesting to note in this regard that the significant effects of object orientation obtained by Todd and Norman (1991) could easily have been mistaken for an effect of sequence length if the axes of compression or expansion in the 2-frame displays had all been parallel to the line of sight (e.g., see Loomis & Eby, 1988; Johnston, Cumming & Landy, 1994). As the objects depicted in 8-frame sequences rotate farther and farther from this degenerate orientation, the variations in their structure become more and more discriminable over time. However, in order to demonstrate a true effect of sequence length , it is necessary to show that a multiple frame display is capable of producing more accurate performance than would be possible for any single pair of images presented in isolation. Although the eight frame condition of Todd and Norman (1991) produced much higher performance than did the 2-frame displays whose axes of compression or expansion were parallel to the direction of view, this difference was eliminated when the depicted objects were presented at a more slanted orientation.

Of the many different 3D discrimination tasks that have been employed in the literature, there are relatively few for which it is theoretically necessary to detect higher-order spatiotemporal relations in the pattern of projected motion to achieve accurate performance. Some paradigms that have been designed specifically for that purpose include discriminating the relative 3D lengths of nonparallel line segments (Todd & Bressan, 1990) or the magnitude of a rotating dihedral angle (Hogervorst, Kappers and Koenderink, 1993; Eagle and Blake, 1994). The Weber fractions for performing these tasks range from 30% to 90%, which is more than an order of magnitude higher than those typically found for many other low level visual discriminations. Thus, these findings provide additional evidence that human observers have only a limited sensitivity to higher order spatiotemporal relations among three or more views of an apparent motion sequence.

It is important to keep in mind, however, that the vast majority of 3D discrimination tasks that have been used to investigate the perception of structure from motion are all theoretically possible to perform based solely on first-order spatiotemporal relations (e.g., see Braunstein, Hoffman, Shapiro, Andersen & Bennett, 1987; Dosher, Landy & Sperling, 1990; Hildredth, Grzywacz, Adelsen & Inada, 1990; Sperling, Landy, Dosher & Perkins,1990; Treue, Husain & Andersen, 1991). Observers’ judgments are quite accurate on some of these tasks, but there are others that appear to be intrinsically more difficult. Consider, for example, the ability of observers to discriminate whether or not a moving configuration of dots is coplanar (see Todd & Bressan, 1990). The available psychophysical evidence shows clearly that small amounts of surface curvature that are parallel to the axis of rotation are much easier to detect than curvature in a direction that is perpendicular to the axis of rotation (Cornilleau-Peres & Droulez, 1989; Norman & Lappin, 1992). Other research by Werkoven & van Veen (1995) has shown that observers have difficulty discriminating which of two moving dots is closest to a rotating planar surface. Because relative distance to a plane is invariant under affine transformations, this task ought to be possible if observers were capable of exploiting all of the information that is potentially available within 2-frame motion sequences.

Depth Scaling

Although there is a growing amount of evidence that the visual perception of structure from motion is based primarily on the first-order pattern of image displacements, there are other aspects of the psychophysical data that this hypothesis cannot explain. A fundamental computational limitation of any 2-frame analysis of structure from motion is that an arbitrary configuration of points under parallel projection does not have a unique rigid interpretation. It will either have no possible rigid interpretation at all, or an infinite one parameter family of possible interpretations. This would explain why observers typically exhibit large errors in judgments of Euclidean metric structure from motion, and why they are unable to discriminate different structures within the one parameter family even when a motion sequence contains more than two distinct frames (e.g., see Todd & Bressan, 1990; Todd & Norman, 1991; Liter, Braunstein & Hoffman, 1994). The aspect of the data that is hard to explain, however, is the existence of systematic biases in observers' magnitude estimations of perceived depth.

There have been numerous experiments reported in the literature in which observers were required to make magnitude estimations or matching judgments for various aspect so 3D structure such as depth, orientation or curvature (e.g., see Braunstein & Andersen, 1984; Braunstein, Liter & Title, 1993; Liter, Braunstein & Hoffman, 1994; Loomis & Eby, 1988, 1989; Todd, 1984, 1985). For example, Todd and Norman (1991) had observers estimate the amplitudes of rotating sinusoidal surfaces relative to their periods. The results revealed that judged amplitudes increased linearly with the simulated amplitude, but with a slope greater than one, such that the perceived depths were systematically overestimated.

If the available information is infinitely ambiguous, then why should an object appear to have any specific depth at all? To the extent that it does, there would have to be some other constraint or heuristic at work to restrict the set of possible perceptual interpretations. One possible hypothesis that has been considered by several investigators is that perceived depth is determined by the overall range of projected displacements times some arbitrary scaling constant (e.g., see Liter, Braunstein & Hoffman, 1993; Todd, 1984, Todd & Norman, 1991). A strong prediction of this hypothesis is that perceived depth should increase proportionally with the angular velocity of an object's depicted motion as well as its extension in depth, since the pattern of projected displacements is determined by both of these factors. In one recent study by Liter et. al. (1993), in which observers judged the apparent depth of random dots in a volume, this prediction was confirmed. However, in a similar study by Todd & Norman (1991), in which observers judged the apparent amplitudes of sinusoidally corrugated surfaces, the magnitude of perceived depth remained relatively invariant over large changes in the depicted angular velocity. This latter finding suggests that the overall range of projected displacements cannot be the only source of information for perceptually specifying an object's extension in depth

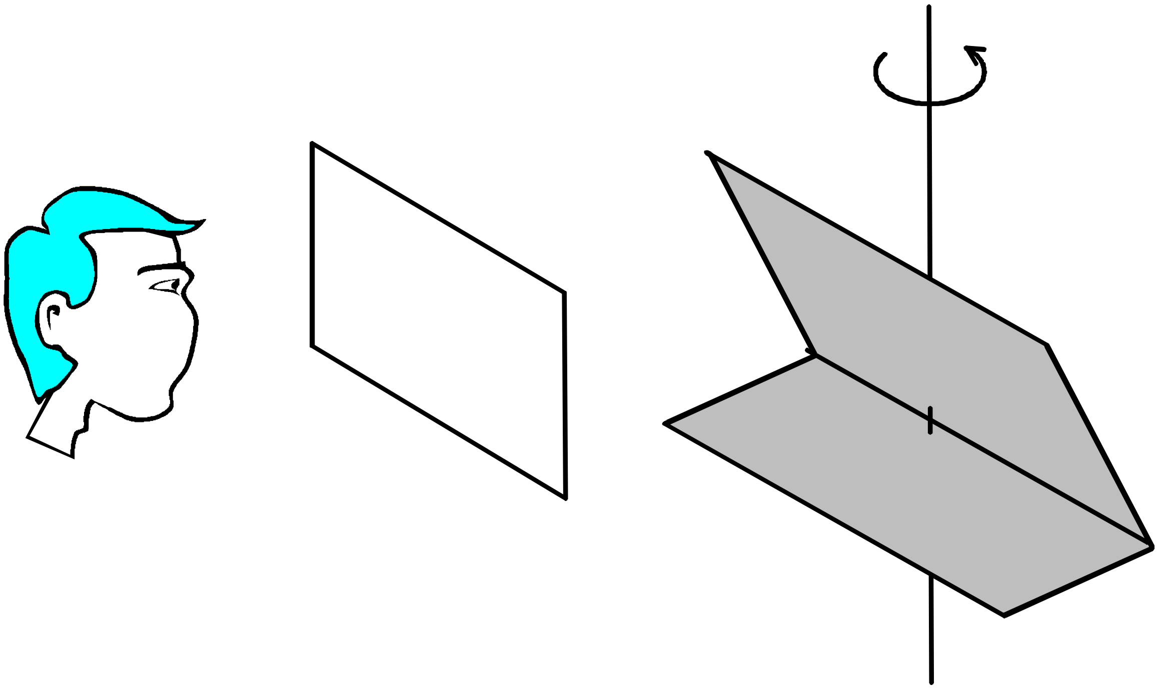

Figure 5 -- A schematic view of a simulated dihedral angle rotating in depth similar to the displays used by Braunstein, Liter and Tittle (1993) and by Tittle, Todd, Perotti and Norman (1995). When asked to judge the magnitude of the depicted angle, observers' estimates are systematically distorted.

Additional evidence to support this conclusion has also been provided by Braunstein, et. al. (1993) for observers magnitude estimations of rotating dihedral angles (see

Figure 5). They found the same systematic overestimations of depth as did Todd and Norman (1991), but they also discovered that this effect could be attenuated by adding a compression transformation to the pattern of projected motion in a direction perpendicular to the axis of rotation (see also the similar finding of Tittle, Todd, Perotti & Norman, 1995). They speculated that perceived depth is determined primarily by the vertical gradients of velocity in their displays, but that this information is scaled by the magnitude of compression in the horizontal direction.

To better understand why this suggestion is theoretically reasonable it is useful to consider how various stimulus factors can influence the pattern of projected motion. The optical flow field for a horizontally oriented dihedral angle rotating in depth about a vertical axis can be described analytically by the following equation:

$$v = \omega(\frac{|y'|\tan(\theta)}{\cos(\alpha)} - x'\tan(\alpha))$$

where v is the projected image velocity in the horizontal direction, ω is the angular velocity of rotation, θ is half the magnitude of the depicted dihedral angle, α is the slant of the dihedral edge relative to the frontoparallel plane, and x and y are the horizontal and vertical image coordinates. The gradients of velocity over space can be obtained by the first partial derivatives of this equation in the horizontal and vertical directions, which are expressed as follows:

Compression: $$\frac{\partial v}{\partial x'} = -\omega \tan(\alpha)$$

Shear: $$\frac{\partial v}{\partial y'} = \frac{w \tan(\theta)}{\cos(\alpha)}$$

Note in these equations that the vertical gradient (labeled shear) varies systematically with the magnitude θ of the rotating dihedral angle, but that it is also influenced by the angular velocity ω and the slant of the dihedral edge α. Because the horizontal gradient (labeled compression) provides a potential source of information about these latter two parameters, it is reasonable to suppose that it could be used as a scaling factor to provide some degree of orientation invariance when estimating the size of a dihedral angle from the magnitude of shear.

Planar Motion and Perspective

In all of the discussion presented thus far, we have considered the available information for configurations of points rotating in depth under orthographic projection. Such information would be mathematically insufficient to compute the Euclidean metric structure of arbitrary configurations, but it would make it possible -- at least in principle -- to obtain a unique rigid interpretation of an object's 3-dimensional form in certain special-case situations. One such special case occurs for object motions that are confined to a fixed plane (Hoffman & Flinchbaugh, 1982; Lappin, 1990). Lappin & Love (1993) and Lappin & Ahlstrom (1994) have recently argued that human observers can indeed discriminate Euclidean distance relations in this situation, though this result has been challenged by Pizlo and Salach-Golyska (1994) as arising from artifactual sources of information.

A second special case to consider includes objects viewed under strong polar perspective. Longuet-Higgens (1981) and Longuet-Higgens and Prazdny (1984) have shown that 2-frame sequences under polar perspective provide sufficient in formation to determine an objects 3D structure up to a homogeneous scaling transformation. Note that this is a different 1-parameter family of possible interpretations than the one described earlier for orthographic projections. For 2-frame sequences under orthographic projection, a small angular displacement of an object with a large extension in depth cannot be distinguished from a larger angular displacement of an object with a smaller extension in depth. For 2-frame displays under polar projection, in contrast, a large object at a far distance cannot be distinguished from a small object at a near distance. If, however, there is other information such as convergence to specify viewing distance, then an object's Euclidean metric structure can be specified uniquely from the first order pattern of image displacements.

Another important aspect of polar projections, which differs from the orthographic analysis described earlier, is that the pattern of projected displacements cannot be related to the 3D structure of an object by an affine transformation. Nevertheless, there is some evidence to suggest that they do provide sufficient information to perceptually specify at least some affine properties. Lappin and Fuqua (1983) have shown that observers are quite accurate at bisecting the distance between two moving dots at different depths , even when they are displayed with an exaggerated polar perspective. Discrimination performance is quite poor, however, when observers are asked to make judgments about Euclidean metric properties. For example, Norman, Todd, Perotti and Tittle (1996) obtained Weber fractions on the order of 0.3 for discriminating the relative lengths of line segments that were oriented in different directions.

If observers could exploit all of the information that is potentially available from patterns of projected motion with strong polar perspective, then they ought to be capable of perceiving an object's shape up to a homogeneous scaling transformation. There is some evidence to suggest, however, that this is not the case. Todd (1984) found that the perceived depth of a rotating cylinder relative to its width can be systematically expanded or compressed even when presented with strong polar perspective. A similar result has also been obtained by Tittle, Todd, Perotti and Norman (1995) for rotating dihedral angles. They also found a depth scaling effect of image compression similar to the findings of Braunstein et. al. (1993) for objects viewed under orthographic projection. Such findings suggest that the human visual system may not distinguish between polar and orthographic projection in its perceptual analysis of 3D structure from motion.

Different Types of Optical Deformations

In some respects, the mathematical models for computing structure from motion developed for machine vision may appear to be superior to human perception, because they can exploit the effects of polar perspective and the higher order spatiotemporal relations among many different views of an apparent motion sequence. However, there are some types of optical motion encountered in natural vision that cannot be analyzed by existing computational models, yet are interpreted correctly by human observers. A fundamental assumption of virtually all of these models is that it is possible to identify points in different views of an apparent motion sequence that all projectively correspond to the same physical point in 3-dimensional space. The displays used in most psychophysical investigations of perceived structure from motion are designed specifically to satisfy this assumption. Simulated objects are generally composed of small dots or lines, whose optical projections can be tracked over time in each successive frame of an apparent motion sequence.

In natural vision, however, the overall pattern of optical stimulation can contain a variety of other structures such as occlusion contours, cast shadows, and smooth variations of surface shading, which do not projectively correspond over time to identifiable features in the physical environment. This can have important theoretical implications for the analysis of 3-dimensional structure from motion. When objects are observed in motion, these different aspects of optical structure do not always change in the same way, and analyses that are designed to be used with one type of optical deformation will not in general be appropriate for others.

Consider, for example, the occlusion contour that forms the silhouette of a human head. If the head rotates in depth about a vertical axis, the optical contour that bounds its projection will be systematically deformed, but the locus of surface points to which it corresponds will also be continuously changing -- i.e., for a frontal view the occlusion contour will pass through the ears, and for a profile view it will pass through the nose. Analyses that assume projective correspondence will be of little use with this type of optical deformation, even as a local approximation. Indeed, it is often the case that the optical motion of the bounding contour will be in one direction while the projected motion of any identifiable point on that contour is in the opposite direction. (see Todd, 1985).

There are other types of image structure for which motions of the observer and motions of the observed object produce different patterns of optical deformation. When an observer moves in an otherwise rigid environment, visible objects will all maintain a constant relationship with their sources of illumination. Because shadow borders and gradients of Lambertian shading in this context remain bound to fixed positions in 3-dimensional space, their resulting patterns of optical deformation will satisfy the condition of projective correspondence, and can therefor be analyzed using conventional techniques for determining structure from motion. When an object moves relative to its light source, however, shadow borders and gradients of shading will move over its surface. Because this violates the condition of projective correspondence, existing computational models would be unable to generate a correct rigid interpretation.

There have been several demonstrations reported in the literature that human observers can obtain compelling kinetic depth effects from the optical deformations of smooth occlusion contours (Todd, 1985; Cortese & Andersen, 1991; Pollick, Giblin, Rycroft & Wilson, 1992, Norman & Todd, 1994, Norman, Todd & Phillips, 1995) and there have also been a few mathematical analyses of how this might be theoretically possible (Koenderink & van Doorn, 1977; Giblin & Weiss, 1987; Cipolla & Blake, 1990). There is some evidence to suggest that the optical deformations of shadows and shading may provide useful information as well (Todd, 1985; Norman & Todd, 1994; Norman, Todd & Phillips, 1995), but the generality of this evidence remains to be determined. One important factor that has limited research on these topics is the difficulty of creating controlled laboratory displays of moving shaded images. This difficulty is quickly diminishing, however, with the continuing advance of computer graphics technology, so that this is likely to be a more active area of research within the next several years.

Conclusions

The research reviewed in the present article provides clear evidence that the visual perception of structure from motion by human observers is quite different from the computational algorithms that have been developed for machine vision. One important reason for this difference is the poor sensitivity of the human visual system to higher order spatiotemporal relations among three or more views of a motion sequence (e.g., see Todd, 1981; Snowden & Braddick, 1991; Werkoven, Snippe & Toet, 1992). Because observers' perceptions must be based primarily on the first order relations between individual pairs of views, their performance is severely limited by the fact that the pattern of projected image displacements can have an infinite number of rigid interpretations.

The available evidence indicates that observers do reasonably well on tasks that are theoretically possible to perform from the available information within 2-frame motion sequences. That is to say, they are able to detect most types of nonrigid deformations, and to accurately discriminate structural properties that are invariant over affine transformations. They have considerably more difficulty, however, on tasks that require an accurate perception of Euclidean metric properties. The Weber fractions for discriminating 3D lengths or angles are an order of magnitude higher than those obtained for most other low level visual properties, and observers' magnitude estimates of a moving object's depth relative to its width tend to be systematically distorted.

The fact that observers can misperceive an object's extension in depth while correctly identifying that it is undergoing rigid rotation leads to an interesting conundrum. Suppose, for example, that an observer overestimates the extension in depth of a rotating object by 20%, as has been reported by Todd and Norman (1991) and Braunstein et. al. (1993). If such an observer were to view a rotating ellipsoid whose extension in depth is 20% smaller than its width at a particular moment in time, it should appear at that moment as a sphere. At a later point in its rotation cycle, however, its width would be 20% larger than its depth, and it should appear as an elongated ellipsoid. Why wouldn't this change in shape be perceived as a nonrigid deformation? This puzzle was first noted by Helmholtz in considering the systematic distortions of stereoscopic space, but it is also applicable to the visual perception of structure from motion.

One possible resolution of this conundrum, first suggested by Gibson (1979), is that Euclidean metric distances in 3-dimensional space are not a primary component of an observer's perceptual experience. This hypothesis has been developed more fully in a recent series of papers by Todd & Reichel (1989), Todd & Bressan (1990), Todd & Norman (1991), Norman & Todd (1992, 1993) and Tittle et. al. (1995). These authors have presented evidence that an observer's knowledge of 3-dimensional form may involve a hierarchy of different perceptual representations. Their findings indicate that observers are quite accurate and reliable at judging an object's topological, ordinal, or affine properties, and that perception of rigid motion occurs when these properties remain invariant over time. Although observers can exhibit a conceptual understanding of Euclidean metric structure, this knowledge may be more cognitive than perceptual. The available psychophysical evidence suggests that if observers are required to make judgments about lengths or angles of visible objects in 3-dimensional space, they will resort to using ad hoc heuristics, which typically produce low levels of accuracy and reliability, and which can vary unpredictably among different individuals or for different stimulus configurations.

References

Bennett, B. & Hoffman, D. (1986) The computation of structure from fixed axis motion: Nonrigid structures. Biological Cybernetics, 51, 293-300.

Bennett, B., Hoffman, D., Nicola, J., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. Journal of the Optical Society of America, 6, 1052-1069.

Braunstein, M.L., & Andersen, G.J. (1984). Shape and depth perception from parallel projections of three dimensional motion. Journal of Experimental Psychology: Human Perception and Performance, 10, 749-760.

Braunstein, M.L., Hoffman, D.D., & Pollick, F.E. (1990). Discriminating rigid from nonrigid motion. Perception & Psychophysics, 47, 205-214.

Braunstein, M.L., Hoffman, D.D., Shapiro, L.R., Andrsen, G.J., & Bennett, B.M. (1987). Minimum points and views for the recovery of three-dimensional structure. Journal of Experimental Psychology: Human Perception and Performance, 13, 335-343.

Braunstein, M. L., Liter, J. C., & Hoffman, D. D. (1994) Inferring structure from two-view and multi-view displays. Perception, in press.

Braunstein, M. L. Liter, J. C. & Tittle, J. S. (1993) Recovering three-dimensional shape from perspective translations and orthographic rotations. Journal of Experimental Psychology: Human Perception and Performance, 19, 598-614.

Cipolla, R., and Blake, A. (1990). The dynamic analysis of apparent contours. Proceedings of the Third International Conference of Computer Vision, 616-623.

Cortese, J. M., and Andersen, G. J. (1991). Recovery of 3-D shape from deforming contours. Perception and Psychophysics, 49, 315-327.

Cornilleau-Peres, V., & Droulez, J. (1989). Visual perception of curvature; Psychophysics of curvature detection induced by motion parallax. Perception & Psychophysics, 46, 351-364.

Doner, J., Lappin, J.S., & Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. Journal of Experimental Psychology: Human Perception and Performance, 10, 1-11.

Dosher, B.A., Landy, M.S., & Sperling, G. (1989). Ratings of kinetic depth in multidot displays. Journal of Experimental Psychology: Human Perception and Performance, 15, 816-825.

Dosher, B.A., Landy, M.S., & Sperling, G. (1990). Kinetic depth effect and optic flow -- I. 3D shape from Fourier motion. Vision Research, 29, 1789-1814.

Eagle, R. A., & blake, A. (1995) Two-dimensional constraints on three-dimensional structure from motion tasks. Vision Research, 35, 2927-2941.

Giblin, P., and Weiss, R. (1987). Reconstruction of surfaces from profiles. Proceedings of the IEEE First International Conference on Computer Vision, 136-144.

Gibson, J. J. (1979) The ecological approach to visual perception. Boston: Haughton Mifflin.

Grzywacz, N., & Hildreth, E. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. Journal of the Optical Society of America, A4, 503-518.

Hildreth, E.C., Grzywacz, N.M., Adelson, E.H., & Inada, V.K. (1990). The perceptual buildup of three-dimensional structure from motion. Perception & Psychophysics, 48, 19-36.

Hoffman, D. & Bennett, B. (1985) Inferring the relative three-dimensional positions of two moving points. Journal of the Optical Society of America A, 2, 242-249.

Hoffman, D. & Bennett, B. (1986) The computation of structure from fixed axis motion: Rigid structures. Biological Cybernetics, 54, 1-13.

Hoffman, D.D., & Flinchbaugh, B.E. (1982). The interpretation of biological motion. Biological Cybernetics, 42, 195-204.

Hogervorst, M., kappers, A. M. L. & Koenderink, J. J. (1993) Perception of metric depth from motion parallax. Perception, 22, supplement, 101.

Huang, T., & Lee, C. (1989). Motion and structure from orthographic projections. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 536-540.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994) Integration of stereopsis and motion shape cues. Vision Research, 34, 2259-2275.

Koenderink, J.J., & van Doorn, A.J. (1977). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual flow. In G. Hauske & F. Butenandt (Eds.), Kybernetik (pp. 224-247). Munich: Oldenberg.

Koenderink, J.J., & van Doorn, A.J. (1991). Affine structure from motion. Journal of the Optical Society of America A, 8, 377-385.

Lappin, J.S. (1990). Perceiving metric structure of environmental objects from motion, self-motion and stereopsis. In R. Warren and A.H. Wertheim (Eds.), The perception and control of self-motion (pp. 541-576). Hillsdale, NJ: Lawrence Erlbaum.

Lappin, J. S. & Ahlstrom, U. B. (1994) On the scaling of visual space from motion: In response to Pizlo and Salach-Golyska. Perception & Psychophysics, 55, in press.

Lappin, J.S., Doner, J.F., & Kottas, B.L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. Science, 209, 717-719.

Lappin, J.S., & Fuqua, M.A. (1983). Accurate visual measurement of three-dimensional moving patterns. Science, 221, 480-482.

Lappin, J.S., & Love, S.R. (1993). Metric structure of stereoscopic form from congruence under motion. Perception & Psychophysics, 51, 86-102.

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1994) Inferring structure from motion in two-view and multi-view displays. Perception, in press.

Longuet-Higgins, H.C. (1981). A computer algorithm for reconstructing a scene from two projections. Nature, 293, 133-135.

Longuet-Higgins, H.C., & Prazdny, K. (1984). The interpretation of a moving retinal image. Proceedings of the Royal Society of London B, 208, 385-397.

Loomis, J.M., & Eby, D.W. (1988). Perceiving structure from motion: Failure of shape constancy. In Proceedings from the second international conference on computer vision (pp. 383-391). Washington, D.C.: IEEE.

Loomis, J.M., & Eby, D.W. (1989). Relative motion parallax and the perception of structure from motion. In Proceedings from the workshop on visual motion (pp. 204-211). Washington, D.C.: IEEE.

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. Journal of the Optical Society of America A, 2, 243-251.

Norman, J. F. & Lappin, J. S. (1992) The detection of surfaces defined by optical motion. Perception & Psychophysics, 51, 386-396.

Norman, J.F., & Todd, J.T. (1992). The visual perception of 3-dimensional form. In G.A. Carpenter & S. Grossberg (Eds.), Neural networks for vision and image processing. Cambridge, MA: MIT press. pp. 93-110.

Norman, J.F., & Todd, J.T. (1993) The Perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. Perception & Psychophysics, 3, 279-291.

Norman, J. F. & Todd, J. T. (1994) The Perception of rigid motion in depth from the optical deformations of shadows and occlusion boundaries. Journal of Experimental Psychology: Human Perception and Performance, 20, 343-356.

Norman, J. F., Todd, J. T., Perotti, V. J. & Tittle, J. S. (1996) The visual perception of 3D length. Journal of Experimental Psychology: Human Perception and Performance, in press.

Norman, J. F., Todd, J. T., & Phillips, F. (1995) The perception of surface orientation from multiple sources of optical information. Perception & Psychophysics, 57, 629-636.

Perotti, V. J., Todd, J. T. & Norman, J. F. (1996) The visual perception of rigid motion from constant flow fields. Perception & Psychophysics, in press.

Perotti, V. J. & Todd, J. T., Tittle, J. S., & Norman, J. F. (1994) Perception of 3D form from instantaneous flow components. Investigative Ophthalmology & Visual Science, 35, 1317.

Pollick, F. E., Giblin, P. J., Rycroft, J., and Wilson, L. L. (1992). Human recovery of shape from profiles. Behaviormetrika, 19, 65-79.

Pizlo, Z. & Salach-Golyska, M. (1994) Is vision metric: Comment on Lappin and Love (1992). Perception & Psychophysics, 55, in press

Sperling, G., Landy, M.S., Dosher, B.A., & Perkins, M.E. (1989). Kinetic depth effect and identification of shape. Journal of Experimental Psychology: Human Perception and Performance, 15, 826-840.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995) The systematic distortion of perceived 3D structure from motion and binocular stereopsis. Journal of Experimental Psychology: Human Perception and Performance, 21, 663-678.

Todd, J. T. (1981). Visual information about moving objects. Journal of Experimental Psychology: Human Perception and Performance, 7, 795-810.

Todd, J. T. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. Journal of Experimental Psychology: Human Perception and Performance, 8, 238-251.

Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. Perception and Psychophysics, 36, 97-103.

Todd, J. T. (1985). The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? Journal of Experimental Psychology: Human Perception and Performance, 11, 689-710.

Todd, J.T., Akerstrom, R.A., Reichel, F.D., & Hayes, W. (1988). Apparent rotation in 3-dimensional space: Effects of temporal, spatial and structural factors. Perception & Psychophysics, 43, 179-188.

Todd, J.T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. Perception & Psychophysics, 48, 419-430.

Todd, J.T., & Norman, J.F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. Perception & Psychophysics, 50, 509-523.

Todd, J.T., & Reichel, F.D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. Psychological Review, 96, 643-657.

Treue, S., Husain, M., & Andersen, R.A. (1991). Human perception of structure from motion. Vision Research, 31, 59-76.

Ullman, S. (1977). The interpretation of visual motion. Ph.D. Thesis, Massachusetts Institute of Technology.

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

Ullman, S. (1983). Recent computational studies in the interpretation of structure from motion. In J. Beck & A. Rosenfeld (Eds.) Human and machine vision (pp. 459-480). New York: Academic Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. Perception, 13, 255-274.

Wallach, H., & O'Connell, D.N. (1953). The kinetic depth effect. Journal of Experimental Psychology, 45, 205-217.

Werkhoven, P., Snipe, H. P. & Toet, A. (1992) Visual processing of optic acceleration. Vision research, 32, 2313-2329.

Werkoven, P. & van Veen, H. A. (1995) Extraction of relief from visual motion. Perception & Psychophysics, 57, 645-656.