THE VISUAL PERCEPTION OF 3-DIMENSIONAL FORM

J. Farley Norman and James T. Todd[1]
Brandeis University

1. Introduction

One of the most perplexing phenomena in the study of vision is the ability of observers to determine an object's 3-dimensional structure from patterns of light that project onto the retina. Indeed, were it not for the facts of our day-to-day experience, it would be tempting to conclude that the perception of 3-dimensional form is computationally impossible, since the properties of optical stimulation have so little in common with properties of real objects encountered in nature. Whereas real objects exist in 3-dimensional space and are composed of tangible substances such as earth, metal, or flesh, an optical image of an object is confined to a 2-dimensional projection surface and consists of nothing more than flickering patterns of light. Nevertheless, for many animals including humans, these seemingly uninterpretable patterns of light are the primary source of sensory information about the layout of objects and surfaces in the surrounding environment.

Previous research has identified several different properties of optical structure from which an object's 3-dimensional form can be perceptually specified. Some of these properties -- the so-called pictorial depth cues -- are available within individual static images. Consider, for example, the patterns of image contours presented in Figure 1. The upper left panel of this figure shows a pattern of converging line segments, which is perceived as a ground plane receding in depth; the upper right panel shows a pattern of connectivity among parallel contours, which is perceived as two solid rectangular objects, one resting on top of the other; and the lower panel shows a pattern of curved contours, which is perceived as a smooth surface. Other perceptually informative properties of individual images that have been studied extensively include gradients of shading or texture (e.g., see Todd and Mingolla, 1983; Mingolla and Todd, 1986; Todd and Akerstrom, 1987).

Additional information about an object's 3-dimensional form can also be provided by the systematic transformations among a sequence of multiple images. For example, when an observer moves within a cluttered environment the texture elements on visible surfaces move projectively at different velocities (i.e., motion parallax) due to their different depths relative to the observer. Binocular

[1] The research described in this manuscript was supported by grants #89-0016 from the Air Force Office of Scientific Research and #BNS-8908426 from the National Science Foundation, the Office of Naval Research and the Air Force Office of Scientific Research to James Todd.

parallax occurs whenever an observer views an object with both eyes simultaneously. Because each eye views the world from a slightly different vantage point, texture elements at different depths project to different relative positions (binocular disparity) within each retinal image. Other optical transformations that can provide information about 3-dimensional form include accretions or deletions of texture when one surface occludes another, or the deformations of shading that occur when an object moves relative to the direction of illumination.

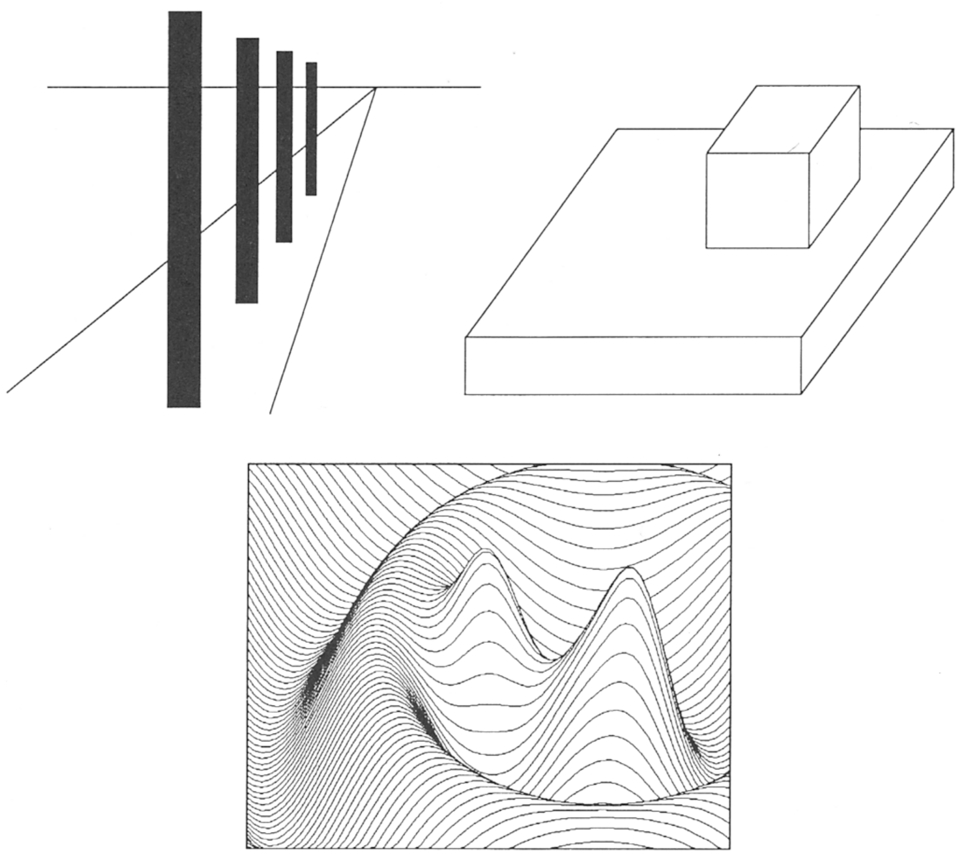

Figure 1. Three examples of how static configurations of image contours can provide perceptually compelling information about an object's 3-dimensional structure.

All of these different properties of optical structure have been studied extensively within controlled laboratory experiments, and the results show clearly that they can all produce compelling perceptions of 3-dimensional form. What is not clear, however, is how these perceptions relate to the actual structure of the physical environment. The mere fact that observers are able to perceive 3-dimensional form does not reveal the specific parameters by which visible objects are perceptually represented. Thus, in an effort to shed new light on this issue, the present article will examine some alternative geometric frameworks for representing shape, and it will review the available psychophysical evidence to see which of these frameworks are most similar to the properties of human perception.

2. What is shape ?

As it has evolved from ancient times, the concept of space in modern science has become increasingly abstract and farther removed from our intuitive beliefs derived from experience. In speculating about the properties of physical space, for example, some theorists have argued that the fabric of space can be deformed by the presence of large masses, and that these deformations are responsible for the phenomenon of gravity. Imagine taking a glass marble and rolling it across a smooth surface. When the rolling marble encounters a hill or a valley, its movements will be constrained by the overall shape of the surrounding terrain. In an analogous fashion, the large mass of the sun deforms its surrounding space, which constrains the earth to orbit in an elliptical trajectory.

The nature of visual space is at present not adequately known. We perceive a compelling 3-dimensional environment filled with many differently-shaped objects, despite the obvious fact that the spatio-temporal optical patterns that give rise to this perception are only 2-dimensional. A more complete understanding of the nature of visual space is critical for the study of solid shape, because the structure of this space determines the measurable properties of objects embedded within it. Let us now consider some of the possible alternatives.

Throughout the literature on human perception, classical euclidean geometry is by far the most common framework for describing the structure of the environment. The defining characteristic of a euclidean space is that it possesses a distance metric, such that the absolute distance between two points $\mathbf{a}$ and $\mathbf{b}$ having Cartesian coordinates $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ and $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ is $\sqrt{(\mathbf{a}_1-\mathbf{b}_1)^2 + (\mathbf{a}_2-\mathbf{b}_2)^2 + (\mathbf{a}_3-\mathbf{b}_3)^2}$. Since a euclidean space is isotropic, we can measure distances between any pair of arbitrary points. Thus, if human observers are able to perceive the euclidean metric structure of the environment, then they ought to be able to accurately compare distances (or angles) of line segments oriented in different directions according to the pythagorean theorem.

It is important to recognize that the euclidean distance metric is not universally applicable in all contexts. Consider, for example, the structure of a simple triangle. One of the fundamental theorems of euclidean geometry is that the sum of the three angles of a triangle must be exactly 180 degrees. This will always be true if the triangle is drawn on a flat surface. However, if a triangle is drawn on a sphere, then the sum of its three angles will always be greater than 180 degrees, and if a triangle is drawn on a saddle then the sum of its angles will always be less than 180 degrees. Beginning in the nineteenth century, results such as these led Gauss and others to conclude that the constraints imposed by euclidean geometry might not always be appropriate for describing the structure of the environment. In order to deal with curved surfaces, they discovered, it is necessary to adopt alternative noneuclidean distance metrics for which the Pythagorean theorem is invalid.

It is also possible to define even more abstract spaces. For example, in affine geometry the distance metric is allowed to vary in different directions -- i.e., it is anisotropic. Because distance intervals can only be compared when they are in the same direction, one cannot determine either distances or angles between arbitrary points located within an affine space. Surprisingly, however, this limitation does not seriously handicap the study of shape. As described by Snapper and Troyer (1971) : "affine geometry is what remains after practically all ability to measure length, area, angles, etc., has been removed from Euclidean geometry. One might think that affine geometry is a poverty-stricken subject. On the contrary, affine geometry is quite rich".

Other types of representation are possible which have no distance metric at all. For example, many of the important properties of a smoothly curved surface can be adequately described by the depth order relations among neighboring points, without including any information about depth magnitudes. Objects can also be described by their patterns of connectivity, or as a collection of categorically distinct parts.

There are many possible geometries for describing an object's 3-dimensional structure. As was first noticed by the German mathematician Felix Klein in 1872, these geometries can be organized in a hierarchy, based on the different transformations they allow, and the structural properties that remain invariant under those transformations. Euclidean geometry allows arbitrary translations and rotations, which preserve the distance between any pair of points on an object. Affine geometry allows arbitrary stretching transformations, which do not leave distance invariant, but do preserve a wide variety of other properties, such as the sign of Gaussian curvature of a surface or the parallelism of a pair of line segments. Similarly, with more general projective transformations, a conic section will remain a conic section and a collinear set of points will remain collinear. Even when an object is subjected to arbitrary smooth deformations, some of its properties such as the pattern of connectivity among neighboring points will still remain invariant.

3. The geometry of human vision

Which of these geometries is most relevant to the visual perception of

3-dimensional form? In the discussion that follows we shall consider several possible aspects of an object's structure, and we shall review the available psychophysical evidence about their relative perceptual salience.

A. Euclidean Structure

In order to obtain an accurate euclidean representation of the environment, it would first be necessary to somehow determine the relative depths of every visible point from the available patterns of optical stimulation (e.g., see Marr and Poggio, 1979; Marr, 1982). As we shall see, some types of optical information are theoretically sufficient to obtain an accurate depth map while others are not.

Consider, for example, the phenomenon of stereopsis. It is well known that human observers can perceive vivid and compelling 3-dimensional shapes defined by variations in binocular disparity. It is important to keep in mind, however, that it is not possible, even in principle, to uniquely determine euclidean structure from binocular optical patterns based solely on variations in horizontal disparity. In order to recover the 3-dimensional distance between a pair of points on the basis of their disparity, one must have accurate information about the distance to the observer's fixation point. This information about viewing distance could come from some optical source other than binocular horizontal disparity or from some non-optical, extra-retinal source like the state of convergence of the two eyes.

Figure 2 illustrates Panum's limiting case, the simplest possible stereoscopic situation. An observer fixates F and views two points separated in depth. In the left eye, the two points project to the same retinal location; while in the right eye, they project to different retinal locations (disparity) as a result of the points' differing depths. Since the binocular disparity δ is an angle, the actual depth difference D cannot be determined on the basis of the optical patterns alone. In order to recover D, one must know the viewing distance from the observer to F. The key point is that there is no one-to-one relationship between horizontal retinal disparity and depth. Any given disparity δ could result from any physical depth difference depending on the viewing distance. In our example, both viewing situations have the same disparity δ, yet the physical depth interval corresponding to that disparity is much larger for the viewing situation illustrated in the left half of the figure due to the larger viewing distance.

Longuet-Higgins (1982) and Mayhew and Longuet-Higgins (1982) have shown that the *vertical* disparity that exists for all environmental points lying off the horizontal meridian could be used to obtain the necessary information about viewing distance. However, Fox, Cormack, and Norman (1987) manipulated the magnitude of vertical disparity within line-element stereograms and found that variations in vertical disparity had no effect on the perceived depth intervals resulting from a given horizontal disparity.

The convergence angle of the two eyes could also be used as a possible source of information about viewing distance. Although Foley (1980) has shown that manipulations of convergence can have measurable effects on stereoscopically defined figures, it is important to keep in mind that the eyes are essentially parallel for viewing distances over two meters while stereopsis occurs over much larger viewing distances involving hundreds of meters (Cormack, 1984). Given that a pattern of binocular disparities covaries with both an object's physical form and its position relative to an observer, stereopsis appears to be ill-suited for delivering information about euclidean relations in 3-dimensional space. Indeed, the invariance of perceived binocular shape under changes in viewing position and distance led Julesz (1971, p. 290) to conclude that : "for stereopsis one must generalize the metric of space from a rigid Euclidean one to a less rigid affine or topological space" (see also Luneberg, 1947, 1950).

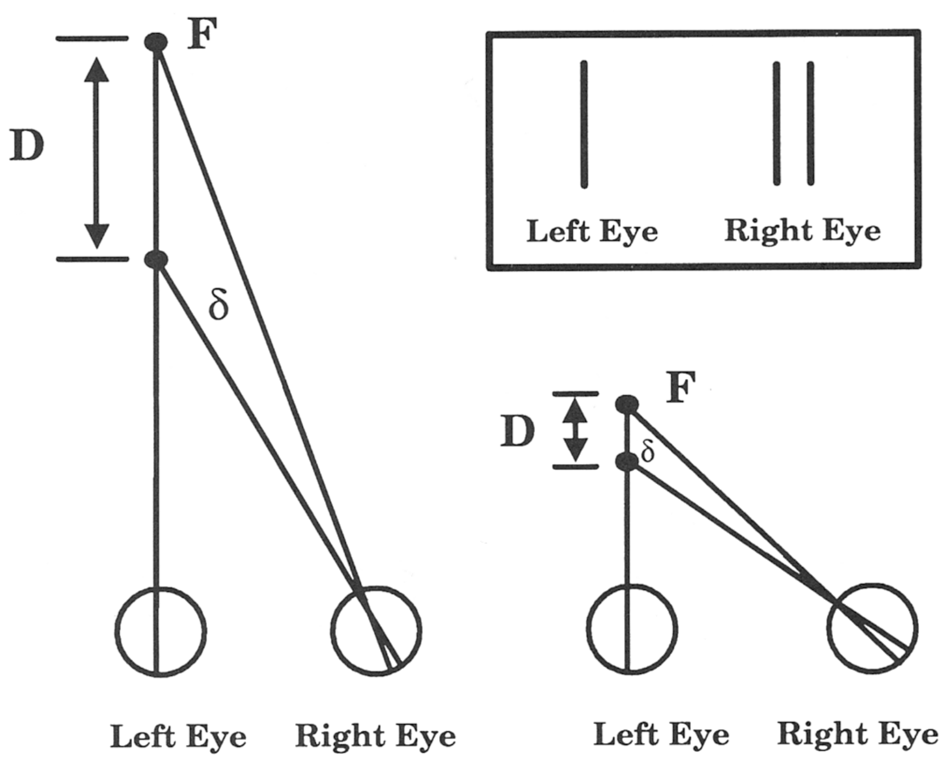

Figure 2. A schematic illustration of Panum's limiting case used to show that any given binocular disparity δ can be produced by any physical depth interval depending on the observer's viewing distance to the fixation point.

Although perception of 3-dimensional form from stereopsis may be inherently limited, perhaps there are other sources of information that are potentially more reliable. Of all the different aspects of optical stimulation that are known to influence observers' perceptions of 3-dimensional form, motion is the one that is most likely to provide perceptually useful information about euclidean metric structure. During the past decade, there have been numerous theoretical analyses of how it is possible to compute an object's structure from motion, provided that certain minimal conditions are satisfied. Most of these analyses are designed to be used with a discrete sequence of orthographic projections of an arbitrary configuration of points rotating in depth about an arbitrary axis. Within this context, it can be proven mathematically that there will always be a unique rigid interpretation for any apparent motion sequence that contains at least three views of four or more noncoplanar points (see Bennett, Hoffman, Nicola, and Prakash, 1989; Huang and Lee, 1989; Ullman, 1979). These conditions are both necessary and sufficient. For arbitrary configurations that contain fewer than three views or fewer than four points, the 3-dimensional structure will be mathematically ambiguous with an infinity of possible rigid interpretations.

During the past several years, however, there has been a growing amount of evidence that these theoretical limits may have surprisingly little relevance to actual human vision. Of particular importance in this regard are the recent findings from several different laboratories that 2-frame apparent motion sequences presented in alternation provide sufficient information to obtain compelling kinetic depth effects and to accurately discriminate between different 3-dimensional structures (Braunstein, Hoffman, and Pollick, 1990; Braunstein, Hoffman, Shapiro, Andersen, and Bennett, 1987; Doner, Lappin, and Perfetto, 1984; Lappin, Doner, and Kottas, 1980; Todd, Akerstrom, Reichel and Hayes, 1988; Todd and Bressan, 1990). Similar results can also be obtained using longer length sequences of scintillating random dot surfaces for which no dot is allowed to survive for more than two successive frames (Dosher, Landy, and Sperling, 1990; Norman, 1990; Todd, 1985).

Since euclidean depths and orientations cannot in principle be determined from 2-frame apparent motion sequences under orthographic projection, and human observers can perform accurately on tasks where the motion sequences are limited to two views, it seems reasonable to conclude that observers performance on these tasks cannot be based on a computational analysis of euclidean structure from motion. It would appear from this finding that observers are able to make use of other, more abstract forms of perceptual representation when presented with minimal amounts of information, but are they capable of perceiving euclidean structure when sufficient information is available to support such an analysis? It is important to keep in mind when considering this issue that the defining characteristic of euclidean structure that distinguishes it from other possible geometries is the existence of an isotropic distance metric. Thus, if there are any conditions in which observers can accurately discriminate lengths and angles of line segments oriented in different directions, then, by definition, their knowledge of 3-dimensional structure in those conditions must be euclidean.

Surprisingly, even though this is the defining characteristic of euclidean geometry, there have been relatively few experiments in which observers were required to make explicit judgements about isotropic metric structure. One such experiment has recently been performed by Todd and Bressan (1990). Observers in their study were asked to discriminate the relative 3-dimensional lengths or angles between moving line segments, whose relative orientations were carefully controlled so that above chance performance could not be achieved based solely on the projected lengths or projected angles depicted in each display. Performance on these tasks was extremely poor relative to other types of sensory discrimination. Weber fractions for the length and angle judgements were 25 and 50 percent, respectively. Moreover, although the overall level of performance was above chance, there were no significant improvements as the number of distinct frames in an apparent motion sequence was increased from two to eight. Thus, whatever information was used for performing these tasks, it was fully available within 2-frame displays for which an accurate analysis of euclidean metric structure was computationally impossible.

B. Affine Structure

If not euclidean structure, then what can observers perceive from the minimal amounts of information provided by 2-frame apparent motion sequences or stereograms under orthographic projection? Recent analyses by Koenderink and van Doorn (1991) and Todd and Bressan (1990) have shown that this information is mathematically sufficient to determine an object's structure up to an affine stretching transformation along the line of sight (see also Bennett et al, 1989). Although an object's euclidean structure cannot be uniquely specified from such minimal amounts of information, it is nonetheless severely constrained.

There are a wide variety of object properties that can be reliably detected based solely on an analysis of affine structure. For example, it is possible with this analysis to determine the metric length ratio between any pair of parallel line segments; to perform various nominal categorizations, such as distinguishing between planar and nonplanar configurations; and to accurately discriminate structural differences between any pair of objects that cannot be made congruent by an affine stretching transformation along the line of sight. It is also interesting to note in this regard that an analysis of affine structure from 2-frame displays is sufficiently powerful to perform most of the existing psychophysical tasks that have been employed previously to study observers' perceptions of structure from motion or stereopsis, including judgements of rigidity, or coherence, discriminations of rigid from nonrigid motion, judgements of ordinal depth relations, and the discrimination or identification of complex 3-dimensional forms.

In a recent series of experiments, Todd and Bressan (1990) and Todd and Norman (1991) have examined the accuracy of observer's judgements for several different aspects of a moving object's 3-dimensional form. The results reveal that performance is quite poor for tasks that require an analysis of euclidean metric structure, but that observers' judgements can be extremely accurate for tasks that are mathematically possible based on an analysis of affine structure. In addition, there is little or no improvement on any of these tasks as the number of distinct frames in an apparent motion sequence is increased beyond two.

It is important to keep in mind that a 2-frame apparent motion sequence or stereogram under orthographic projection can only provide sufficient information to determine an object's structure up to an affine stretching transformation along the line of sight. Thus, this latter finding suggests a surprising prediction : Consider an extended apparent motion sequence of an object rotating in depth that is stretched or compressed along the line of sight at each frame transition. The cumulative effects of these stretching transformations would be carried along by the rotation, resulting in potentially large deformations of the object's structure. However, if the human visual system is restricted to an analysis of first order displacements between 2-frame sequences, as we have suggested above, then this particular type of deformation should be perceptually undetectable, since every successive 2-frame sequence would have a possible rigid interpretation.

This prediction has been confirmed empirically in a recent series of experiments by Norman and Todd (1991). When a rotating object is stretched or compressed along the line of sight, it appears indistinguishable from a perfectly rigid object whose rate of rotation is accelerated or decelerated. Norman and Todd have also performed computer simulations to demonstrate that these different transformations can be distinguished by analyses of euclidean structure from motion that are able to integrate information over three or more views. A typical pattern of results from one such algorithm by Hoffman and Bennett (1986) is shown in Figure 3. From the projected positions of a set of points rotating in depth about a fixed axis in the image plane, this algorithm computes the radius of each point relative to the axis of rotation. The upper curve in Figure 3 shows the computed radius over a 100 frame sequence for a single point within a rigid configuration whose angular velocity varies sinusoidally over time. The computed radius in this case remains perfectly constant, indicating that the object's motion is rigid. The lower curve in this figure shows the output from the same algorithm for a nonrigid configuration that rotates at a constant velocity, but is sinusoidally stretched along the line of sight as it rotates. Note in this case that the computed radius varies over time indicating the object's motion is nonrigid. It is clear that these different transformations could easily be distinguished using existing algorithms for computing euclidean structure from motion. Thus, the fact that they are perceptually identical provides especially strong evidence that human observers may be restricted to a more abstract analysis of affine structure.

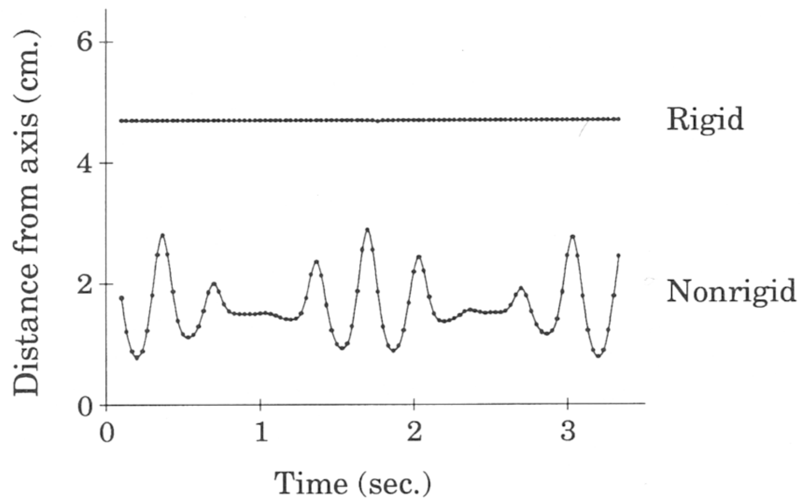

Figure 3. The distance of two points from their axes of rotation as a function of time. Each trajectory was computed from the optical projection of a moving configuration using an algorithm developed by Hoffman and Bennett (1986). The point represented by the upper curve was part of a rigid configuration, whose angular velocity varied sinusoidally over time. The point represented by the lower curve, in contrast, was part of a nonrigid configuration that was stretched sinusoidally along the line of sight as it rotated. Although the differences between these two trajectories are easily detectable using an analysis of euclidean structure from motion, they cannot be detected by actual human observers.

C. Ordinal Structure

Whereas an affine representation of 3-dimensional form retains some rudimentary information about ratiometric distances in any given direction, it is also possible to describe many of the essential properties of an object's structure without any distance metric at all. Gibson (1950) argued that much of our perceptual awareness of the environment is based on simple order relations that can be described in terms of "greater than" or "less than". More recently, Todd and Reichel (1989) have suggested that an observer's knowledge of smoothly curved surfaces can often involve a form of ordinal representation, in which neighboring surface regions are labeled in terms of which region is closer to the point of observation without specifying how much closer.

It is important to recognize that the order relations in this proposed representation are only defined for adjacent regions within an arbitrarily small neighborhood. This has some important consequences. Suppose, for example, that we wish to determine the ordinal depth relation between two visible surface regions R_1 and R_n that are not locally adjacent to one another. Using an ordinal representation, the relative depth of these regions can only be determined if there is a continuous chain of intervening regions that are ordinally transitive (i.e., if $R_1 < R_2 < R_3 \ldots < R_n$, then $R_1 < R_n$). If this restriction is violated (i.e., $R_1 < R_2 < R_3 \ldots > R_n$), then the relative depths of R_1 and R_n cannot be determined from an ordinal representation without providing additional information. Todd and Reichel (1989) have demonstrated psychophysically that ordinal transitivity is important for the perception of smoothly curved surfaces. That is to say, when observers are required to judge the relative depths of ordinally transitive surface regions, their responses are more accurate and they have faster reaction times, than when similar judgements are performed for ordinally intransitive surface regions.

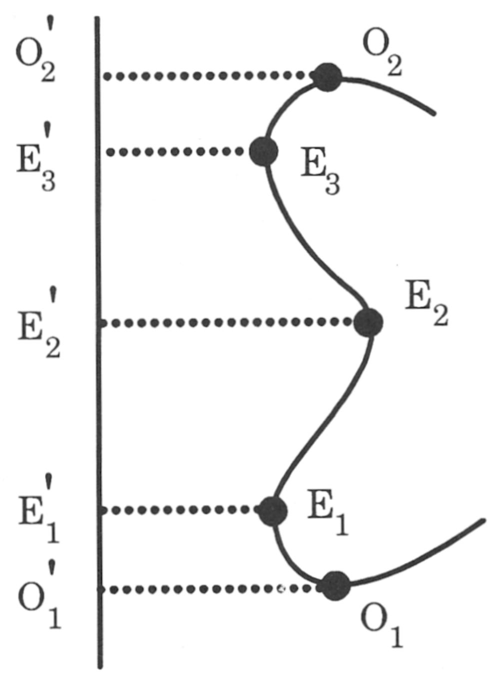

Figure 4. The ordinal structure of any surface is completely determined by the optical projections of its occlusion points and depth extrema. In this particular example there are two occlusion points O_1 and O_2, and three depth extrema E_1, E_2, and E_3. Their corresponding optical projections in the image plane are labeled O_1', O_2', E_1', E_2', and E_3', respectively.

To better appreciate the precise nature of ordinal representations it is useful to consider a planar cross-section of a smoothly curved surface as shown in Figure 4. Note in the figure that there are two occlusion points O_1 and O_2. It can be demonstrated theoretically (see Koenderink and van Doorn, 1976, 1982) that as we move from an occlusion point in an attached region, the ordinal depth of a surface relative to the image plane must decrease monotonically until a depth minimum is reached. Thus, occlusion contours provide potential information about the ordinal structure of attached surface regions in their immediate local neighborhood, and there is considerable evidence to suggest that human observers rely heavily on this information for the visual perception of 3-dimensional form (see Reichel and Todd, 1990; Todd and Reichel, 1989). A complete ordinal representation cannot be achieved, however, without also identifying the depth extrema (i.e., the depth maxima and minima), which define the boundaries of ordinal transitivity where a monotonic depth change switches from positive to negative or vice versa. The optical projections of these depth extrema together with those of the occlusion points are both necessary and sufficient for visually specifying the complete ordinal structure of any surface cross-section.

D. Topological Structure

Some aspects of an object's structure can be adequately characterized using even more abstract representations, in which the concept of distance is abandoned altogether. Consider the topological structure formed by the pattern of connectivity among the vertices of a polyhedron or the neighborhood relations among identifiable points on a continuous surface. If we allow objects to be smoothly deformed without tearing by arbitrary combinations of bending and stretching transformations, these connectivity and neighborhood relations will remain invariant. Solid objects can be distinguished topologically by the number of holes they contain. Within this framework, a doughnut and a coffee mug are topologically equivalent because they have the same number of holes and can therefore be deformed into one another without tearing. A doughnut and a potato, on the other hand, are topologically different because they do not have the same number of holes.

Is topological structure of any relevance to human perception? Consider the sequence of projected silhouettes of an object rotating in 3-dimensional space depicted in Figure 5. Note in the first frame of this sequence that the object's silhouette is bounded by a single connected contour. Between frames two and three, however, the pattern undergoes a qualitative change in which the emergence of a hole causes the silhouette to be bounded by two different contours that are not connected to one another. Once this hole is revealed, it provides potentially useful information about the depicted object's 3-dimensional structure -- i.e., if an object's silhouette contains a hole, then the object itself must also contain a hole.

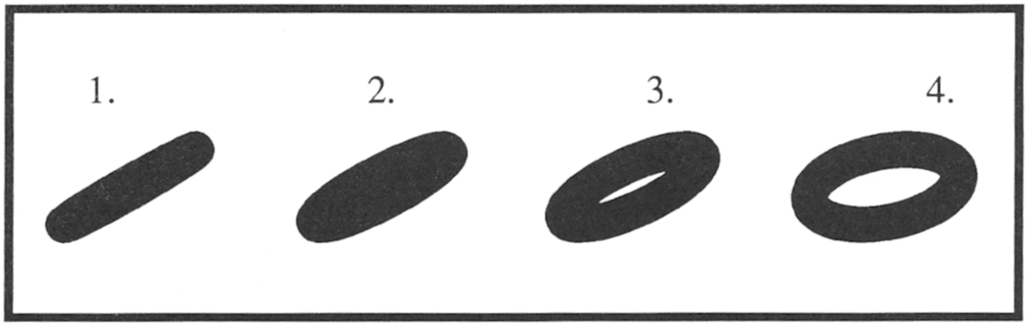

Figure 5. Four successive views of the projected silhouette of a rotating torus. Note the catastrophe between views two and three, which indicate the presence of a hole.

There have been several recent psychophysical experiments designed to demonstrate the importance of topological structure in human vision. For example, Chen (1982) has shown that topologically equivalent forms, such as a filled circle and a filled triangle, are more difficult to discriminate at brief exposure durations than are topologically different forms, such as a filled circle and a circular annulus, whose bounding contours have identical shapes. Topologically equivalent forms also produce stronger perceptions of apparent motion when presented sequentially over time (Chen, 1985; Prazdny, 1986). These findings provide strong evidence that human observers are indeed sensitive to topological relations at a relatively early stage of visual processing.

E. Nominal or Categorical Structure

Another way of representing 3-dimensional form that does not require the concept of distance is to decompose an object into a relatively small set of categorically distinct parts. There have been several variations of this approach described in the literature. For example, one possible strategy proposed by Koenderink and van Doorn (1976, 1980, 1982) is to decompose a surface into bounded regions of positive (elliptic), negative (hyperbolic), or zero (parabolic) Gaussian curvature (see also Koenderink, 1984; Richards, Koenderink, and Hoffman, 1987). These authors have demonstrated mathematically that the local Gaussian curvature of a surface can be optically specified by certain types of image features such as smooth occlusion contours and singular points within the field of image intensities. Objects can also be divided into parts at loci of negative minima along lines of principle curvature (Beusmans, Hoffman, and Bennett, 1987; Hoffman and Richards, 1984) or by decomposing their boundaries into the largest convex surface patches (Vaina and Zlateva, 1990).

With respect to the study of actual human perception, there is a growing amount of evidence that certain types of object recognition may be primarily dependent on part-based representations. Biederman (1987) has argued that most objects in the environment can be adequately represented to achieve recognition using a limited set of volumetric primitives, called geons, which are connected to one another in simple combinations, in much the same way that words can be composed from a relatively small alphabet of phonemes. The optical information from which these geons are perceptually specified is assumed to be based on easily measurable properties of image contours, such as the presence or absence of curvature, parallelism or symmetry, and the cotermination of contours at vertices. Biederman and his colleagues have conducted numerous empirical studies of object recognition which have generally supported the psychophysical validity of this analysis. The results have revealed that only two or three geons are usually sufficient to allow rapid and accurate object recognition, and that performance remains surprisingly unimpaired even when a test image is systematically transformed by changing its scale, deleting a large portion of its contours, or by rotating the depicted object in depth.

4. Discussion

Based on the available evidence, we believe it is the case that there is no one type of visual representation that can adequately account for observers' perceptions of 3-dimensional form. Consider, for example, the apparent usefulness of part-based representations for object recognition. The primary selling point for a nominal description of 3-dimensional structure is its stability over changes in viewing position. When an object moves relative to the observer -- even when the motion is not perfectly rigid as in a human gait -- the decomposition of its structure into parts remains largely invariant. The primary disadvantage of a nominal representation is its lack of precision. Part-based descriptions are invariant over change, because they typically do not encode the size, shape, or orientation of each part. This makes them incapable, however, of describing any variations along these dimensions, unless they are supplemented with some additional form of metric representation.

In some respects, the perceptual performance of human observers in judging various aspects of an object's 3-dimensional form seems to have much in common with the hierarchy of geometries proposed by Felix Klein. In a speech at Erlangen University in 1872, Klein argued that different geometries can be stratified by the properties of objects that are preserved by different types of geometric transformations. There is a growing amount of evidence to suggest that human vision involves a similar type of stratification in which the most perceptually salient aspects of an object's structure are those that remain invariant over the largest number of possible transformations.

Such findings indicate that visual knowledge of 3-dimensional form may exist at multiple levels of description, and that the specific type of representation required for any given task is dependent on the particular judgement an observer is asked to perform. The most difficult tasks are those that require a knowledge of metric structure. Ordinal judgements, in contrast, are performed significantly faster and significantly more accurately (Todd and Reichel, 1989), and those measures of performance can be improved still further for tasks such as object recognition involving topological or nominal judgements (e.g., see Biederman, 1987).

Ironically, although many theorists assume that the primary goal of human perception is to achieve an accurate euclidean representation of visible objects in the surrounding environment, there are good reasons to question whether human observers are able to perceive euclidean metric structure under any circumstances. For example, recent research by Todd and Bressan (1990) and Todd and Norman (1991) has shown that most tasks employed to measure observers perceptions of 3-dimensional form can be performed reliably based solely on an analysis of affine structure, and that tasks which specifically require a knowledge of euclidean structure typically result in dramatically impaired performance. All of this suggests that the geometry of perceived 3-dimensional form may be much more abstract than is generally taken for granted, and that our common intuitions about the importance of metric structure may have surprisingly little relevance to the processes of human perception.

REFERENCES

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

Beusmans, J. M. H., Hoffman, D. D., and Bennett, B. M. (1987). Description of solid shape and its inference from occluding contours. *Journal of the Optical Society of America A*, **4**, 1155-1167.

Biederman, I. (1987). Recognition-by-Components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Andersen, G. J., and 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., Hoffman, D. D. and Pollick, F. E. (1990). Discriminating rigid from nonrigid motion. *Perception and Psychophysics*, **47**, 205-214.

Chen, L. (1982). Topological structure in visual perception. *Science*, **218**, 699-700.

Chen, L. (1985). Topological structure in the perception of apparent motion. *Perception*, **14**, 197-208.

Cormack, R. H. (1984). Stereoscopic depth perception at far viewing distances. *Perception and Psychophysics*, **35**, 423-428.

Doner, J., Lappin, J. S., and 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. and Sperling, G. (1990). Kinetic depth effect and optic flow - I. 3-D shape from fourier motion. *Vision Research*, **29**, 1789-1814.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, **87**, 411-434.

Fox, R., Cormack, L. and Norman, F. (1987). The effect of vertical disparity on depth scaling. *Investigative Ophthalmology and Visual Science (Suppl.)*, **28**, 293.

Gibson, J. J. (1950). **The perception of the visual world**. Boston, MA : Houghton Mifflin.

Hoffman, D. D., and Richards, W. A. (1984). Parts of recognition. *Cognition*, **18**, 65-96.

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

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

Julesz, B. (1971). **Foundations of cyclopean perception.** Chicago, IL : University of Chicago Press.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, **13**, 321-330.

Koenderink, J. J., and van Doorn, A. J. (1976). The singularities of the visual mapping. *Biological Cybernetics*, **24**, 51-59.

Koenderink, J. J., and van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta*, **27**, 981-996.

Koenderink, J. J., and van Doorn, A. J. (1982). The shape of smooth objects and the way contours end. *Perception*, **11**, 129-137.

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

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

Longuet-Higgins, H. C. (1982). The role of the vertical dimension in stereoscopic vision. *Perception*, **11**, 377-386.

Luneberg, R. K. (1947). **Mathematical analysis of binocular vision.** Princeton, New Jersey : Princeton University Press.

Luneberg, R. K. (1950). The metric of binocular visual space.*Journal of the Optical Society of America*, **40**, 627-642.

Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, **204**, 301-328.

Marr, D. (1982). **Vision**. San Francisco: W. H. Freeman and Co.

Mayhew, J. E. W. and Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, **297**, 376-378.

Mingolla, E., and Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, **53**, 137-151.

Norman, J. F. (1990). The perception of curved surfaces defined by optical motion. Ph.D. Dissertation, Vanderbilt University.

Norman, J. F., and Todd, J. T. (1991) The perception of rigid motion for affine distortions in depth. *Investigative Ophthalmology and Visual Science (Suppl.)*, **32**, 958.

Prazdny, K. (1986). What variables control (long-range) apparent motion? *Perception*, **15**, 37-40.

Reichel, F. D., and Todd, J. T. (1990). Perceived depth inversion of smoothly curved surfaces due to image orientation. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 653-664.

Richards, W. A., Koenderink, J. J., and Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America A*, **4**, 1168-1175.

Snapper, E. and Troyer, R. (1971). **Metric affine geometry**. New York : Academic Press.

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., and Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 583-595.

Todd, J. T., and Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 242-255.

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

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

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

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

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London B*, **203**, 405-426.

Vaina, L. M., and Zlateva, S. D. (1990). The largest convex patches: A boundary-based method for obtaining object parts. *Biological Cybernetics*, **62**, 225-236.