N90-22924

THE PERCEPTION OF THREE-DIMENSIONALITY ACROSS CONTINUOUS SURFACES*

Kent A. Stevens
Department of Computer Science
University of Oregon
Eugene, Oregon

ABSTRACT

The apparent three-dimensionality of a viewed surface presumably corresponds to several internal perceptual quantities, such as surface curvature, local surface orientation, and depth. These quantities are mathematically related for points within the silhouette bounds of a smooth, continuous surface. For instance, surface curvature is related to the rate of change of local surface orientation, and surface orientation is related to the local gradient of distance. It is not clear to what extent these 3D quantities are determined directly from image information rather than indirectly from mathematically related forms, by differentiation or by integration within boundary constraints. An open empirical question, for example, is to what extent surface curvature is perceived directly, and to what extent it is quantitative rather than qualitative. In addition to surface orientation and curvature, one derives an impression of depth, i.e., variations in apparent egocentric distance. A static orthographic image is essentially devoid of depth information, and any quantitative depth impression must be inferred from surface orientation and other sources. Such conversion of orientation to depth does appear to occur, and even to prevail over stereoscopic depth information under some circumstances.

INTRODUCTION

One can derive a compelling impression of three-dimensionality from even static, monocular surface displays. Figure 1, for example, suggests an undulating surface. The three-dimensionality of this figure can be dramatically enhanced when one removes the visual evidence about the surface on which the figure is printed. If, say, the pattern is viewed on a graphics display, in a darkened room, monocularly and without head movements, the apparent three-dimensionality is particularly vivid, sufficiently so that one could replicate the apparent surface by curving a ruled sheet of paper and holding it in a particular attitude.

On reflection, it is actually quite curious that a pattern of lines such as those in figure 1 provides so fixed and stable a percept. There is, after all, an infinity of possible 3D surfaces containing lines that would project to that 2D pattern. To posit that the pattern corresponds to a *particular* surface requires certain, specific, strongly constraining assumptions. A theory has been developed of the geometric constraints that support such inferential 3D percepts, one that explains how a range of 3D qualities, such as local surface orientation and curvature might be derived in principle (Stevens 1981a, 1983b, 1986). But it is difficult to extend such theories to explain more precisely *what* 3D information is extracted and internally represented in the process of deriving apparent three-dimensionality from such a 2D stimulus. It is one thing to discuss perception in terms of

* Supported by Office of Naval Research Contract N00014-87-K-0321.

"affordances," "cues," or other characterizations of incident information, and quite another thing to determine the specific course of processing that takes incident information into explicitly represented perceptual quantities.

The remarkable ability to derive surface information from simple monocular configurations has been quite difficult to explain adequately within any of the traditional psychological paradigms. The difficulty stems, I believe, from the lack of basic understanding about what constitutes "apparent three-dimensionality." Depth perception is an often-used term that refers to the perception of surfaces and points in 3D. What differentiates the perception of mere 2D patterns of stimulation from 3D arrangements, seemingly, is perception of the third dimension, namely depth or distance from the viewer to points in space. Gibson insightfully proposed that "visual space perception is reducible to the perception of visual surfaces, and that distance, depth, and orientation...may be derived from the properties of surfaces" (Gibson 1950). To Gibson, the term "apparent three-dimensionality" refers to the perception of more than merely the "third dimension." Visual perception clearly developed to operate in the richly redundant visual world. But the very little 3D information in figure 1 hardly compares to the redundant and seemingly unambiguous wealth of incident information afforded by a natural scene. It might justifiably be relegated to the domain of so-called "picture perception."

Approaches toward understanding surface perception that attempt to isolate the contribution provided by a particular cue, such as texture or contours, or motion or stereopsis, have often been criticized as failing to address enough of the problem. By not embracing the complexity of natural scenes, it is argued, one fails to examine the system in the environment for which it was designed. But while one might well fail to observe important phenomena when only examining components in isolation or in simple combination, by *not* doing so one might equally fail to observe effects central to the strategies that allow the system to effectively deal with complexity and redundancy.

If vision is regarded computationally as the construction of internal descriptions of the visual world, there is no particularly compelling reason to expect qualitatively different modes of visual processing depending on whether the retinal image derives from a picture or a real scene. If one does not expect a different mode for "picture perception," one must then explain how an ambiguous and obviously underspecified 2D stimulus can result in a definite and stable 3D percept.

The challenge, then, is to understand our seeming ability to perceive more specifically than is objectively specified by the stimulus. To Helmholtz, Gregory, and others, this ability stems from the basic perceptual strategy of "unconscious inference." To mix terminology from traditionally antagonistic schools of thought on this matter: higher-order variables in the incident optical array are cues that afford particular 3D inferences. After a while such word play is seen for what it is, and we should go on to more constructive explorations. Substantial progress will likely come only with understanding of the nature of the 3D percept, something that has been given remarkably little attention over the entire history of perceptual studies.

As will be discussed, this task is difficult in theory, because of various mathematical equivalences among different representational forms, and difficult in practice, because of the robustness of the visual observer in performing psychophysical judgments. Despite the intrinsic difficulty, however, there is some evidence that surface perception is sufficiently modular and restricted in its ability to extract and combine 3D information as to be amenable to study using traditional psychophysical methods.

QUANTIFYING APPARENT THREE-DIMENSIONALITY

Following the usage by Foley (1980), *absolute distance* will refer to the egocentric range from an observer to a specific 3D point, which might be a point on a visible surface. *Relative distance* refers to a ratio of absolute distances (without knowing the absolute distances, one might know that one distance is twice another). In this usage *depth* refers specifically to the *difference* of absolute distances to a given point and a reference point. (Hence the depth of a given point relative to a reference point might be known in absolute units without knowing the overall absolute distances involved. Also, if the depth at a point were known and the absolute distance to the reference point were known, their algebraic sum would specify the absolute distance to the given point.)

In addition to scalar distance information at a point, derivatives of distance information specify the orientation of the tangent plane and about curvature of the surface in the vicinity of a point. Surface orientation has two degrees of freedom, and is readily described as a vector quantity related to the normal to the tangent plane (Stevens 1983c). The psychological literature has long used the magnitude quantity *slant* to refer to the angle between the line of sight and the local surface normal (slant varies from 0 to 90°). The other degree of freedom, the *tilt* of the surface, specifies the direction of slant, which is the direction to which the normal projects onto the image plane, and also the direction of the gradient of distance (Stevens 1983a). Since the slant-tilt form aligns with the direction and magnitude of the local depth gradient, it provides many advantages for encoding surface orientation, such as allowing for simultaneous representation of precise tilt and imprecise slant, being closely related to various monocular cues such as shading, texture foreshortening, motion parallax, and perspectivity, and providing for (Necker-type) ambiguity in local surface orientation as reversals in tilt direction (see Stevens, 1983c).

Derivatives of surface orientation, or higher derivatives of distance, are related to surface curvature (across a continuous, twice-differentiable region). Surface curvature also has two degrees of freedom in the neighborhood of a surface point, which might be encoded as principle curvatures, or their image projections.

The central problem, which I will illustrate momentarily, is that across a continuous surface it is possible to convert among these different forms by differentiation (in one direction) and integration (in the other). One source of information about local slant might be used to infer both surface curvature and depth, and another might indicate curvature information directly. With sufficient boundary constraints the information provided by any source might be converted to a form comparable with another across a continuous surface. In general, then, it is difficult to determine whether a given 3D quantity M is derived directly from the image or indirectly from derivatives or integrals of M.

The mathematical equivalences among these various forms of 3D information leave quite open the empirical question of to what extent surface curvature is registered directly versus converted internally (Stevens 1981b; Cutting and Millard 1984; Stevens 1984), and furthermore, the question of the extent to which this information is represented quantitatively rather than qualitatively (Stevens 1981a, 1983b, 1986).

THE 3D INFORMATION CONTENT OF A SIMPLE STIMULUS

Returning to figure 1, what sorts of 3D information can be extracted feasibly? Observe that it consists merely of a family of parallel curves, interpreted as the orthographic projection of parallel curves across a continuous surface. Given the nature of orthographic projection, this pattern is devoid of information about the third dimension (distance). And yet, one sees measurable depth as well as slant in monocular stimuli consisting of line-drawing renditions of continuous ruled (developable) surfaces (Stevens and Brookes, 1984a). Both orthographic (as in figure 1) and perspective projection were used. Using a randomized-staircase forced-choice paradigm, apparent slant was measured by varying the aspect ratio of an ellipse that was briefly superimposed on the monocular surface stimulus. Observers readily interpreted the ellipse as a foreshortened circle slanted in depth, and by adjusting the aspect ratio it could be made to appear flush on the surface. The resulting slant judgments were in close correspondence to the predicted geometric slant of the stimuli.

The apparent depth in these stimuli was then tested by superimposing a stereo depth probe over the monocular surface. Apparent depth was probed stereoscopically using a device similar to Gregory's (1968, 1970) "Pandora's Box." A Wheatstone-style stereoscope provided near-field (38 cm) convergence and accommodation, well within the range of acute stereopsis. After first fixating a binocular point on an empty field, the monocular stimulus was presented briefly (for as little as 100 msec) to the dominant eye only, after which a binocular probe was superimposed at a given stereo disparity over the monocular stimulus for an additional brief interval. Subjects performed a randomized-staircase forced-choice experiment in which the depth of the stereo probe was compared with that of the monocular surface at various locations. Just as Gregory (1970) found measurable apparent depth in a variety of illusion figures, minimal renditions of monocular surfaces, such as figure 1, are also perceived quite measurably in the third dimension.

The experiments suggest that in orthographic projection the visual system can compute from local surface orientation a depth quantity that is commensurate with the relative depth derived from stereo disparity. Apparent slant is a measure of the local gradient of depth, i.e., the rate of change of depth (and being the derivative of distance, slant is independent of the absolute distance to the surface). Depth might be integrated from slant across the surface, but only up to a constant of integration. How, then, are monocular and stereo depth coupled so that they can be compared? The perceptual assumption used to link these two spaces, apparently, is that the absolute distance of the monocular surface at the given fixation point equals that of the stereoscopic horopter at that point. This hypothesis seems sound in that whatever surface location is fixated in sharp focus is likely to lie at zero disparity, since in the near field at least, there is close coupling between vergence and accommodation that brings into sharp focus the (zero disparity) fixation point. The fixated point (seen monocularly in our stimuli but binocularly in normal vision) is thus assumed to be at the absolute distance of the horopter. With the two depth measures sharing a common zero intercept, monocular depth from slant, appropriately scaled by the reference distance, could then be compared to depth from stereo disparity. This conjecture remains to be confirmed empirically.

DEPTH FROM GRADIENT, CURVATURE, AND DISCONTINUITY INFORMATION

In addition to demonstrating the perception of three-dimensionality from highly underspecified stimuli, these observations suggest to us that the visual system has a robust ability to internally convert one form of 3D information into another mathematically equivalent form. The perception of depth from the various so-called monocular "depth cues" (such as shading, contours, and texture gradients) may well provide "direct" information about surface curvature and shape, and only indirect information about depth.

More generally, we propose that shape properties associated with derivatives of distance, specifically surface orientation, curvature, and loci of discontinuity, both in depth (edge boundaries) and tangent plane (creases), are the primary percepts, and that smoothly varying depth across continuous regions is recovered subsequently and indirectly (Stevens and Brookes, 1987b,c).

This proposal explains various phenomena concerning apparent depth from stereopsis. The apparent depth of an isolated bar or point is predicted quite well by the geometry of the binocular system, with depth a straightforward function of stereo disparity and a reference binocular convergence signal (Foley, 1980). But various depth phenomena have been reported recently in the perception of more complicated surface-like stimuli that are not predicted by such a direct functional relationship (Gilliam et al., 1984; Mitchison and Westheimer, 1984). Gilliam et al. (1984) argue that depth derives most readily from disparity discontinuities, and Mitchison and Westheimer (1984) show that coplanar arrangements of lines result in elevated thresholds for depth detection. In a series of experiments in which binocular stimuli presented contradictory monocular and stereo information, we found instances where the stereo information was dramatically ineffective in influencing the 3D percept (Stevens and Brookes, 1987c). The patterns were line-drawn stereo depictions of planar surfaces, rendered orthographically and in perspective, and devoid of disparity discontinuities and disparity contrast (e.g., with a surrounding frame or background). Constant gradients of stereo disparity, consistent with slanted planes, were introduced that were orthogonal to or opposite to the monocularly suggested depth gradients. The monocular interpretation dominated in judgments of apparent surface slant and tilt and in 2-point relative depth ordering. Figure 2, for example, is a stereogram of coplanar lines, with disparities varying linearly in accordance with a slanted plane. The dominant depth impression is the monocular interpretation of a perspective view of a corridor extended in depth.

We hypothesize that stereo disparity influences the monocular 3D interpretation primarily where the distribution of disparities indicates surface curvature and depth discontinuities (i.e., where disparity varies discontinuously or has nonzero second spatial derivatives). Stereo depth across surfaces is substantially a reconstruction from disparity contrast, analogous to brightness from luminance contrast. Consistent with this conclusion are a variety of depth-contrast effects in stereopsis, such as a brightness-contrast analogue in depth (Stevens and Brookes, 1987b), a Craik-O'Brien-Cornsweet analog (Anstis et al., 1978), and various depth induction effects (e.g., Werner, 1938).

REFERENCES

Anstits, S.; Howard, I. P.; and Rogers, B.: A Craik-O'Brien-Cornsweet Illusion for Visual Depth. Vision Research, vol. 18, 1978, pp. 213-217.

Cutting, J. E.; and Millard, R. T.: Three Gradients and the Perception of Flat and Curved Surfaces. Journal of Experimental Psychology: General, vol. 113, 1984, pp. 217-220.

Foley, J. M.: Binocular Distance Perception. Psychological Review, vol. 87, 1980, pp. 411-434.

Gibson, J. J.: The Perception of the Visual World. Boston: Houghton Mifflin, 1950.

Gilliam, B. J.; Flagg, T.; and Finlay, D.: Evidence for disparity change as the primary stimulus for stereoscopic processing. Perception and Psychophysics, vol. 36, 1984, pp. 559-564.

Gregory, R.: Visual Illusions. Scientific American, vol. 218, no. 11, 1968, pp. 66-76.

Gregory, R.: The Intelligent Eye. London: Weidenfeld and Nicolson, 1970.

Mitchison, G. J.; and Westheimer, G.: The Perception of Depth in Simple Figures. Vision Research, vol. 24, 1984, pp. 1063-1073.

Stevens, K. A.: The Visual Interpretation of Surface Contours. Artificial Intelligence, vol. 217, Special Issue on Computer Vision, 1981a, pp. 47-74.

Stevens, K. A.: The Information Content of Texture Gradients. Biological Cybernetics, vol. 42, 1981b, pp. 95-105.

Stevens, K. A.: Surface Tilt (the Direction of Surface Slant): A Neglected Psychophysical Variable. Perception and Psychophysics, vol. 33, 1983a, pp. 241-250.

Stevens, K. A.: The Line of Curvature Constraint and the Interpretation of 3-D Shape from Parallel Surface Contours. Eighth International Joint Conference on Artificial Intelligence, August 1983b.

Stevens, K. A.: Slant-Tilt: The Visual Encoding of Surface Orientation. Biological Cybernetics, vol. 46, 1983c, pp. 183-195.

Stevens, K. A.: On Gradients and Texture "Gradients." Commentary on: Cutting and Millard, 1984. Three Gradients and the Perception of Flat and Curved Surfaces. Journal of Experimental Psychology: General, vol. 113, 1984, pp. 217-220.

Stevens, K. A.: Inferring Shape from Contours Across Surfaces. In: From Pixels to Predicates: Recent Advances in Computational Vision. A. P. Pentland, ed., Norwood, NJ: Ablex, 1986, pp. 93-110.

Stevens, K. A.; and Brookes, A.: Probing Depth in Monocular Images. Biological Cybernetics, vol. 56, 1987a, pp. 355-366.

Stevens, K. A.; and Brookes, A.: Depth Reconstruction in Stereopsis. Proceeding of the First IEEE International Conference on Computer Vision, London, June 1987b.

Stevens, K. A.; and Brookes, A.: Integrating Stereopsis with Monocular Interpretations of Planar Surfaces. Vision Research, vol. 28, no. 3, 1988, pp. 371-386.

Werner, H.: Binocular Depth Contrast and the Conditions of the Binocular Field. American Journal of Psychology, vol. 51, 1938, pp. 489-407.

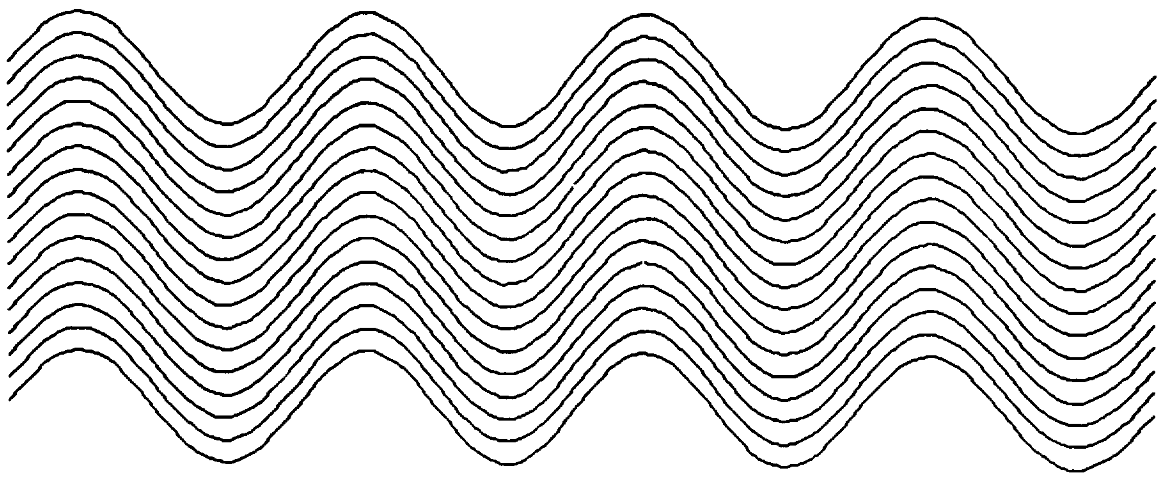

Figure 1.– Undulating lines.

Figure 2.– Stereogram of coplanar lines.