Journal of Experimental Psychology: General
1984, Vol. 113, No. 2, 221–224

Reflections on Surfaces: A Cross-Disciplinary Reply to Stevens

James E. Cutting
ATARI Sunnyvale Research Laboratory, Sunnyvale, California, and Cornell University

New interest in surface perception can be found in both psychological and in artificial intelligence communities. Both disciplines serve to gain through understanding, even adopting, the other's methodology and terminology. This reply is an attempt to affirm certain working principles common to the different methodologies and to discuss and compare certain terms that are used.

Researchers in psychology and in artificial intelligence agree on the theoretical importance of the perception of surfaces: Surface perception is crucial to any understanding of environments and to how we act within them. The two fields also agree that the area of research is complex but tractable. Moreover, there is a collective excitement that genuine progress is being made. Globally, the approaches disagree in two ways: methodology and terminology. Both of these are, I hope, temporary hindrances. Both disciplines have much to gain in sharing methodological tools and in understanding each other's concepts. This response is directed toward that end.

On Methodology: Experiments and Simulations

Psychologists interested in surface perception have traditionally run experiments. Experiments, of course, divide many ways, but those of Cutting and Millard (1984) were exploratory, using scaling and regression techniques. Researchers in artificial intelligence, on the other hand, traditionally run simulations. Simulations, like experiments, divide many ways, but the successful implementation of an algorithm for accomplishing a task in a relatively wide set of circumstances speaks well for the concreteness and workableness of the computational theory. It bodes well for the field of robotics. Successful implementation of one's algorithm on a machine, of course, does not guarantee that a human observer would accomplish the same task in the same manner. It is true that appended arguments—such as those suggesting that the physiology of the nervous system works along similar principles, and that the constraints of the geometry of the situation afford few alternatives (Marr, 1982)—add logical force to successful simulations. Nevertheless, no conjunction of logical arguments and simulation can ever serve as well as human data in establishing the algorithms of human perception.

Psychologists have a strong belief that no matter how rigorous the analytic methods have been in support of a theoretical position, they can always be overturned by data. The only guarantee that a theory has any basis in human perception is the tie that theory has with human data, which are our only measures of the phenomenal world. But often the only guarantee that the theory is concrete enough to be meaningful is through simulation. Thus, researchers in both fields would do well to borrow the tools of the other discipline. Given that Stevens (1983) reported several experiments and Cutting and Millard (1984) simulated an environment to study, certain methodological crossovers are already occurring.

Two Principles

Regardless of one's field, be it psychology or machine vision, two useful ideas can be applied to understanding perception. First, a given perceptual problem could be solved by the perceptual system in any number of ways, especially in a domain as rich as vision. Second, if the problem is an important one for the organism, one should not be surprised if the perceptual system does solve it in a number of different ways depending on the quality of

---

This research was supported, in part, by National Institute of Mental Health Grant MH37467.

Requests for reprints should be sent to James E. Cutting, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853.

the data that it is given. We (Cutting & Millard, 1984) explicitly embraced the former idea when discussing gradients alone; Stevens (1984) explicitly embraced both when discussing gradients in conjunction with other sources of information. I think there is no disagreement between us, because we (Cutting & Millard, 1984) implicitly espoused both principles. Let us consider, in reverse order, how these principles unfold for surface perception.

Multiple sources of information. The overriding question is: How do we know the attributes of a surface—such as its relative shape, orientation, reflectance, and even resiliency—from the information presented to us in the visual array? Obvious places to begin looking are in (a) stereopsis (the differential locations of corresponding elements in the optic arrays of the two eyes), (b) motion (the differential motions of elements across the optic array when one travels through an environment), and (c) the photometric, geometric, and statistical properties of surface elements. When it can, the human visual system almost certainly uses all three sources of information. It is likely that in the future the best machine vision systems will as well. It is an admirable goal to attempt to merge these three sources into the same system (or systems), as Stevens (1984; see also Marr, 1982) wishes to do, but it is unclear at present how much the economics of design in the nervous system demands this parsimony. Regardless, the work under discussion (Cutting & Millard, 1984) concerns only certain aspects of the last source of information: geometric and statistical properties of texture elements.

Multiple algorithms. Stevens (1984, p. 218) has suggested that the "radically different" solutions posed by us (Cutting & Millard, 1984) and by the computational approach may be a case in point. I am not so sure, although I certainly espouse this principle of multiplicity in general. Suppose that the approaches differ not in potential algorithmic instantiation, but in terminology. Dissolve the differences in terminology and perhaps few substantive differences in approach remain.

On Terminology: Gradients, Representations, and Maps

A texture gradient is a measure of change in the projection (image) of a projected environment or object along a chosen stimulus dimension for a chosen image axis. The middle panel of Cutting and Millard's (1984) Figure 3 shows three orthogonal gradients, and hence, three different representations of a flat surface; the bottom panel of Figure 3 shows the corresponding three for a particular curved surface. Each gradient is a measure of change in the image corresponding to (relatively) constant properties of the surface: texture size, texture flatness, and intertexture distance. Collectively, these gradients—perspective, compression, and density—are texture gradients. There is nothing really "raw" about them, contrary to the view of Stevens (1984); they are encoded versions of different information in the scene.

These gradients are external representations of the image, and indirectly of the environment. In much of psychology and in artificial intelligence, however, the term *representation* is used to stand for internal structures. It is not clear to me that these internal representations have (or even should have) ontological status beyond convenient fictions for discussing perceptual process. Regardless, let me adopt this terminology to see where it leads.

Couched in this manner, the theoretical question of import is whether these three representations of an image—perspective, compression, and density—have corresponding representations in the visual system of the perceiver. Our results suggest that (a) there is one psychological representation corresponding to the perspective gradient; (b) there is another corresponding to the compression gradient; (c) these two internal representations are, like their external counterparts, independent of one another; and (d) there may be only a relatively weak internal representation of density as applied to the perception of surfaces. One way to restate the goal of our study is that we sought to corroborate the existence of these internal representations. Because the computational approach has already assumed what seem to be analogous constructs—orientation maps and depth maps for internalized versions of compression and perspective gradients, respectively—let us pursue the analogy, leaving density aside.

Stevens (1984) suggested that local orientation (a point on the compression gradient) is represented internally on an orientation map, and that local scale (a point on the perspective gradient) is represented internally on a depth map. Both of these maps, obviously, are internal representations. Maps by their nature are two dimensional. Gradients, however, can also be extended into maps (or manifolds) in two dimensions. Since the stimuli of Cutting and Millard (1984) had no horizontal variation, these manifolds would have no characteristics not seen in the middle and bottom panels of Figure 3. From Stevens's (1984) discussion it would appear that, from a computational perspective, coarse orientation maps and coarse depth maps can be derived from images like our stimuli. Coarseness, of course, is a relative term. With Stevens, I suspect that considerably better depth maps could be achieved through motion and that somewhat better depth maps could be achieved through stereopsis. Relative strengths of these channels aside, however, I believe that sufficient maps can be built from a static and singular image. Moreover, I take the relative compellingness of the stimuli to our subjects as indicating just that: Because there was no other (global) information besides these gradients available to the observer, and because the best of the displays did indeed look like surfaces receding in depth, I conclude that, given this terminology, sufficient internal maps (internal representations, internal gradients) were constructed to perform the task.

What does the perceptual system do with these maps (internally represented gradients)? With Stevens, I am not sure that the visual system "calculates" the second derivative of the internally represented compression gradient (orientation map). I suspect that, in any literal sense, it does not. Instead, the visual system probably performs some approximation to calculus that yields the same answer. But because taking second derivatives may be a property of processing in the early visual system, at least the form of computing zero crossings as proposed by Marr (1982), I see no reason not to propose that they are also used later on internally represented gradients. Moreover, although the maps themselves may be constructed from zeroth-order measures, as suggested by Stevens (1984), that part of the visual system that uses the maps to derive shape must perform some operation analogous to taking second derivatives.

In his conclusion, Stevens (1984) noted two problems with our (Cutting & Millard, 1984) conclusion: He regarded it as undecided as to (a) whether perspective simply provides local distance information or whether the perspective gradient is used to determine flatness, and (b) whether compression provides local orientation information or whether a compression gradient is used to determine curvature.

These quandaries might be dealt with in several ways. First, if the perspective and compression gradients correspond to depth and orientation maps, respectively, which seems likely given the impossibility of inputs from stereopsis and motion in our experiments, then part of the difference between our positions dissolves. We are agreed on both the importance and the separation of two representations that serve the visual system. Second, according to our (Cutting & Millard, 1984) view, what is done with these maps determines what the observer will perceive.

Our view is that if the visual system can determine the presence of a nonmonotonicity in the orientation map, then the surface will appear curved. Calculus operations need not literally be performed; only approximations are necessary. If, on the other hand, the visual system can determine, within some tolerance, that $P \cdot \tan\alpha$ is a constant, then the surface will appear flat. The value P is the perspective (scale) value on the depth map that Stevens and we agree the visual system can derive for a given texture, and α is the angle between the texture under consideration, the eye, and the point on the ground beneath the eye, about whose computation Stevens and we may disagree. Again, this algorithm is only an approximation to an invariant: Given a reasonable texture size and eye height, one can fix the value of $P \cdot \tan\alpha$ at 89.9°, or very near the horizon (600 eye heights away from one's feet). In comparison, values at 85° (11.4 eye heights), 80° (5.7 eye heights), 75° (3.7 eye heights), 70° (2.7 eye heights), and 65° (2.1 eye heights) are 99.7%, 98%, 96%, 93%, and 89% of that value, respectively. Thus, from a distance of about 3 to many hundreds of eye heights away, the value of $P \cdot \tan\alpha$ remains within 5% of being constant.

We can agree, then, that perceptual information is "most successfully extracted by processes that access derived visible surface representations" (Stevens, 1984, p. 219) if we agree that (a) external representations of stimuli (gradients) map onto internal representations of them, (b) internalized versions of gradients are such maplike representations discussed from the computational perspective, (c) *derived* simply means the output of early visual processing where information concerning layout has been transformed into neural form, and (d) *extracting* simply means the manipulation of information already present in the map (internal representation), not the addition of new information.

It is possible that the terminology of Cutting and Millard (1984) and Stevens (1984) in particular, and that of psychology and machine vision in general, cannot be mapped in this manner, but the effort seems worth making. If such mapping of terms across disciplines is successful, then we have converging lines of argument that both fields are on the right track. Even if not successful, the attempt will reveal areas of disagreement and areas in need of empirical research.

References

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113,* 198–216.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Stevens, K. A. (1983). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics, 33,* 241–250.

Stevens, K. A. (1984). On gradients and texture "gradients." *Journal of Experimental Psychology: General, 113,* 217–220.

Received October 3, 1983
Revision received October 14, 1983 ■