SCIENCE BRIEFS

From Chaos to Coherence

How the Visual System Recovers Objects

By Philip J. Kellman, PhD

Philip J. Kellman, PhD, is Professor of Psychology at the University of California, Los Angeles (UCLA). He is also Co-Director of the UCLA Cognitive Science Research Program. Dr. Kellman's primary research interests are in perceptual and cognitive processes, especially visual perception of objects, motion, and space; perceptual development and perceptual learning; visual cognition and attention; and human factors applications of perceptual and cognitive research.

Dr. Kellman is the recipient of several honors, including the William Chase Memorial Award, the Wolf Aviation Prize, the Fulbright Senior Scholar Research Award, and the APA Boyd R. McCandless Young Scientist Award. Dr. Kellman received his PhD in psychology from the University of Pennsylvania in 1981.

Glancing out of a window, we may see the branches of a tree and a building behind them. We seldom stop to ponder that seeing the building as a single connected object requires stitching together hundreds of small, visible patches of building separated from each other by occluding branches. Peering into more dense foliage, we may see only meaningless fragments of color. If we begin walking, however, the objects and spatial layout of the scene may now be revealed. We do not often notice that the objects and surfaces we perceive are partly occluded by other objects. To perceive objects visually, the human visual system incorporates remarkable abilities to overcome fragmentation of the optical input that occurs over both space and time.

Understanding these abilities has proved to be a complex task. A generation ago, the Gestalt psychologists proposed principles of perceptual organization relevant to these problems. Those principles contained important insights, yet they were too vague to make precise predictions or to be implemented in artificial vision systems. Much of the current work on object perception involves deriving clear computational principles from earlier intuitions.

To perceive objects in ordinary scenes, the visual system must solve several different problems. Segmenting the world begins with work on detecting and classifying visible edges, based on discontinuities in luminance, depth, and motion. Our research focuses on the steps that come after basic edge detection: How does the visual system determine which visible edges and surfaces form parts of a single object, and how does it fill in gaps where edges and surfaces are occluded? An account of the stimulus relationships and processing routines used by our visual system to perceive objects would answer some of the most basic questions of human vision and would have enormous consequences for machine vision systems as well. At present, no artificial vision system can recover the unity and boundaries of three-dimensional (3-D) objects that are partly occluded in natural scenes.

Over the last several years, our research efforts, along with those of others, have revealed some of the key components of object perception in natural scenes. One of them is that constructing occluded edges and filling in surface qualities within edges are complementary processes in object perception. In general, surfaces "spread" within the confines of edges given by local information and those constructed by completion across gaps. Both edge completion and surface spreading can bind together separated visible regions according to their own rules.

One advance in understanding the edge-completion process has been the identification of image features that mark the beginning and endpoints of all visually filled-in (interpolated) edges. In thinking about the ecology of perception, we come across the following useful fact from projective geometry: Whenever one object partly occludes another, the optic projection will contain an important feature, namely sharp corners at the points of overlap. These corners are first-order or tangent discontinuities (TDs) in the orientations of projected edges. Our experiments suggest that the visual system uses TDs as the starting points for edge connections: All interpolated edges start and end at TDs.

Relatability

The key to our model of object completion involves a second ecological fact. Objects tend to have relatively smooth boundaries. Moreover, in determining where object parts go under occlusion, departures from smoothness would be difficult to anticipate. The visual system incorporates this constraint by connecting together edges that satisfy a certain mathematical criterion for smoothness. We have been able to show that a particular relationship—which we call *relatability*—accounts for most cases of edge interpolation in visual perception. Relatability, which makes precise the notion of good continuation from Gestalt psychology, is defined mathematically. Intuitively, it constrains interpolated boundaries to be smooth (differentiable at least once) and monotonic (bending in only one direction) connections between TDs in the image.

A unifying idea in our work on edge interpolation is that a single process underlies several visual phenomena that were thought to depend on different mechanisms. Perception of partly occluded contours and surfaces is one; another is the phenomenon of illusory contours (and illusory objects). Evidence suggests that a single boundary-interpolation process governs visual construction of boundaries across gaps in these cases (and some others). Some of this evidence consists of similar patterns of speed and accuracy found in perceptual tasks involving occluded and illusory contours having the same spatial relations. Other evidence involves displays in which the illusory and occluded edges join, as well as displays in which edge interpolation must occur before the final appearance (as occluded or illusory) is determined. Differences in the phenomenal appearance of illusory and occluded edges derive from their depth relations to other surfaces in the scene.

Several lines of recent research suggest that the edge-interpolation process I have described is carried out by neural units at surprisingly early levels of visual processing. These include evidence from other laboratories involving single-cell recording in monkeys performing psychophysical tasks, which demonstrated that sensitivity on detection and classification tasks is highly related to spatial interactions among orientation-sensitive units in the visual cortex. The formal notion of relatability accounts well for the relevant spatial relations in some of these tasks.

Object Perception in Three-Dimensional Space

A limitation of most existing models of boundary and surface interpolation is their focus on relationships in static, two-dimensional (2-D) arrays. The real world is 3-D, and researchers have suspected for some time that 3-D as well as 2-D edge and surface relationships affect object perception. Recently, we developed a new method for assessing the role of 3-D relations. The method requires subjects to make a quick classification of the relative orientations in 3-D of two surface parts separated by a gap. Speed and accuracy on this task are enhanced when the two pieces are connected by visual-interpolation processes into a single object. We found that a simple generalization of the notion of relatability to three dimensions accounted well for the conditions under which parts were seen as connected. These data provide the first evidence from an objective performance paradigm that the object completion process is truly 3-D.

Philip J. Kellman, PhD

Object Interpolation Across Space and Time

When objects or observers move, the visible fragments of objects change over time. How can visual processes produce a representation of objects and surfaces when visible surface and edge fragments are not only separated in both space, but are changing in time? The problem is daunting. Recently, we proposed a straightforward extension of our relatability model that might explain much of dynamic object perception. Activations caused by parts of edges that are visible at different times may be accumulated in some sort of storage buffer (spatiotemporal integration). But what relationships among stored fragments lead to object completion (spatiotemporal interpolation)? Consider a simple case in which an occluding object with two vertically and horizontally separated windows moved horizontally in front of another object. Different edge fragments from the occluded object would appear at different times. If early-appearing fragments persisted in some visual representation until the later fragments appeared, these could be inputs into a spatial relatability computation. Now consider a more difficult case. If the object further away moved, again, different parts would appear at different times and in different spatial relations than in the first case. Suppose, however, that when the first fragments appeared, they were not only stored in a buffer, but a velocity signal extracted from their motion allowed their spatial position to be extrapolated. Then, when later parts became visible, the updated positions of the previously visible parts, along with the currently visible parts, can enter into the standard spatial relatability computation. The positional updating of moving parts combined with new visible parts comprises spatiotemporal relatability.

Recently, we carried out the first systematic experiments testing whether spatiotemporal relatability predicts object completion in dynamic displays. Subjects viewed objects that moved behind an occluding panel containing several narrow slits, displaced so that some of the object never projected to the eyes. After the target presentation, subjects made a forced choice of which of two displays had been presented. We found much better accuracy at this task for object parts that were spatiotemporally relatable than for the same object parts scrambled so as to disrupt relatability. Further work is focusing on the effects of velocity in these displays. Based on prior research, we believe that spatiotemporal integration and interpolation require the object parts to be presented within a limited temporal window. We are currently studying this possibility.

Conclusion

Because of occlusion, seeing the objects in our environment often requires processes that construct coherent 3-D representations from separated fragments that change over time. Luckily, the human visual system seems to be equipped with sophisticated routines for accomplishing this task. A general principle that appears to describe the stimulus relationships that lead to object completion is the notion of *relatability*. Recent research indicates that with straightforward extensions, this concept may account for phenomena of object perception involving 3-D relationships and information given over time by motion. The formal notion of relatability in turn gives us clues to the character of process and mechanism. The picture emerging is that a number of object-perception phenomena can be explained in terms of a small set of spatial and temporal relationships and a relatively local and autonomous process that uses interactions of orientation-sensitive units to construct smooth connections that bridge gaps in the input. The process allows perceived objects to correspond to the objects in the world because these visual processes exploit ecological regularities about objects and the laws determining their projection to the observer.

Note: *I would like to thank my collaborators on this work: Thomas F. Shipley, PhD, Carol Yin, PhD, and Evan Palmer, PhD. I gratefully acknowledge support from the National Science Foundation.*