CHAPTER 10

3D and Spatiotemporal Interpolation in Object and Surface Formation

Philip J. Kellman

Patrick Garrigan

Evan M. Palmer

CONTENTS

INTRODUCTION

As David Marr observed in his classic book *Vision* (1982), understanding visual perception involves issues at multiple levels of analysis. This observation applies not only to conceptual differences in the kinds of questions researchers must ask, but also to different levels of visual processing. Vision researchers have made great progress in understanding early cortical filtering. At the opposite end, research has revealed some areas in which high-level representations reside, such as those for objects or faces. Between these levels, however, there is a considerable gap. This gap in "middle vision" involves all of Marr's levels: the understanding of information for computing representations of contours, surfaces, and objects; the representations and processes involved; and the sites and roles of cortical areas. Although Marr emphasized that these levels have substantial independence in terms of the questions they pose, the lag in understanding what goes on "in the middle" is also related to interactions among these levels. Understanding the task and information paves the way for process descriptions. Similarly, detailed hypotheses about processes and representations guide meaningful neurophysiological investigations.

Fundamental to the middle game in vision are three-dimensional (3D) representations. What is the shape of a surface? How do we represent the shapes of 3D objects and obtain these representations from incomplete and fragmentary projections of an object to the eyes? How do we obtain descriptions of objects and surfaces in ordinary environments, where the views of most objects are partly obstructed by other objects, and visible areas change in complex ways as objects and observers move?

Although many traditional approaches to vision have sought to discover how meaningful perceptual representations can be gotten by inferences from static, two-dimensional (2D) images, it has become increasingly clear that human vision both utilizes complex 3D and spatiotemporal information as inputs and constructs 3D surface representations as outputs. Although human vision may exploit shortcuts for some tasks, 3D surface representations play many important roles both in our comprehension of the world and our ability to interact with it.

In this chapter, we consider several lines of research aimed at improving our understanding of 3D and spatiotemporal surface and object formation. Specifically, we are concerned with the achievement of surface or object representations when the visual system must interpolate across spatial and spatiotemporal gaps in the input. The human visual system possesses remarkable mechanisms for recovering coherent objects and surface representations from fragmentary input. Specifically, object and surface perception depend on interpolation processes that overcome gaps in contours and surfaces in 2D, 3D, and spatiotemporal displays. Recent research suggests that the mechanisms for doing so are deeply related in that they exploit common geometric regularities.

SOME PHENOMENA OF VISUAL INTERPOLATION

In ordinary perception, partial occlusion of objects and surfaces is pervasive. Panels (a)–(d) on the left side of Figure 10.1, for example, show views of a house occluded by a fence. Even in a single static view, we are able to get some representation of the scene behind the fence. If the several views were seen in sequence by a walking observer, we would get a remarkably complete representation, as suggested by Figure 10.1c.

Perceiving whole objects and continuous surfaces requires perceptual processes that connect visible regions across gaps in the input to achieve accurate representations of unity and shape. These have most often been studied for static 2D representations. Yet perception grapples with a 3D world and produces, in part, truly 3D representations of object contours and surfaces. Furthermore, when objects or observers move, the visible regions of objects change over time, complicating the requirements of object formation. The system deals with fragmentation, not only in space, but across time as well. Thus, we may think of contour and surface perception in the real world as a mapping from information arrayed across four dimensions (three spatial dimensions and time) into 3D spatial representations. If motion is represented, visual object and surface formation is a mapping from fragmented four-dimensional (4D) inputs into coherent, functionally meaningful, 4D representations.

These phenomena are formally similar in that the same physically specified contours of the central figure are given in Figure 10a, 10b, and 10c, and the completed object in each case is defined by the same collection of physically specified and interpolated contours. (Figure 10d includes only the corresponding interpolations in the middle part of the figure.)

Categories of Interpolation Phenomena

A number of phenomena involve connecting visible contours and surfaces across gaps (Figure 10.2). Figure 10.2a shows partial occlusion. Six

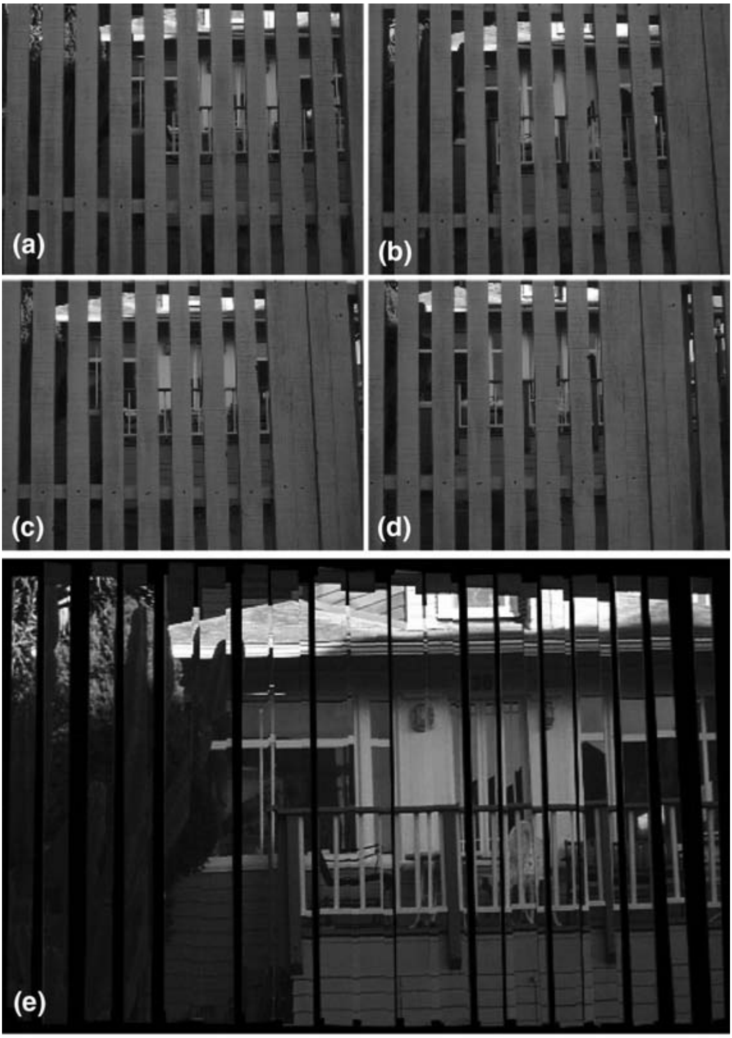

FIGURE 10.1 Real-world interpolation requires integration over time and space. Frames (a,b,c,d): Several images of an occluded real-world scene. The porch of this house is visible between the fence posts and is perceived as a series of connected visual units despite the fact that shape information is fragmented in the retinal projection. Frame (e): When motion and three-dimensional contour and surface interpolation operate, the visual system can generate a far more complete representation of the scene, of the sort depicted here.

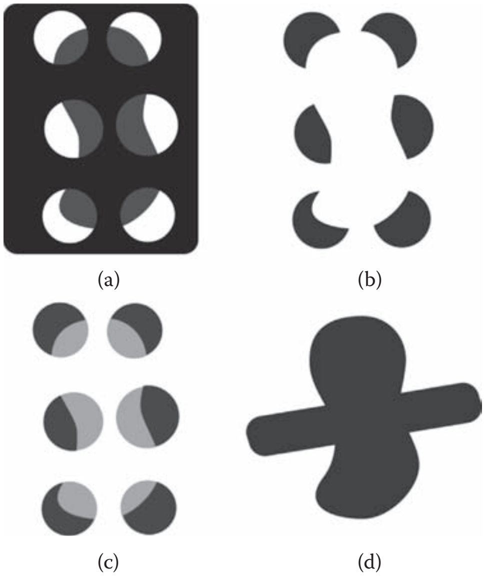

FIGURE 10.2 **(See color insert following p. 94.)** Four perceptual phenomena that can be explained by the same contour interpolation process. (a) A partially occluded object. The blue fragments are spatially disconnected, but we perceive them as part of the same object. (b) The same shape appears as an illusory figure and is defined by six circles with regions removed. (c) A bistable figure that can appear either as a transparent blue surface in front of six circles or an opaque blue surface seen through six circular windows. (d) A self-splitting object. The homogenous black region is divided into two shapes. This figure is bistable because the two shapes appear to reverse depth ordering over time.

noncontiguous blue regions appear; yet, your visual system connects them into a single object extending behind the black occluder. The object's overall shape is apparent. Perceptual organization of this scene also leads to the perception of circular apertures in the black surface, through which the blue object and a more distant white surface are seen. Figure 10.2b illustrates the related phenomenon of *illusory contours* or *illusory objects*. Here, the visual system connects contours across gaps to create the central white figure that appears in front of other surfaces in the array. Figure 10.2c shows a transparency version of an illusory figure; the figure is created but one can also see through it. Finally, in Figure 10.2d, a uniform black region is seen to split into two visible figures, a phenomenon that has been called *self-splitting objects.*

These phenomena are formally similar in that the same physically specified contours of the central figure are given in each case, and the completed object in each case is defined by the same collection of physically specified and interpolated contours.

Contour and Surface Processes

Evidence suggests two kinds of mechanisms for connecting visible areas across gaps: contour and surface interpolation. These processes can be distinguished because they operate in different circumstances and depend on different variables (See Figure 10.3). Contour interpolation depends

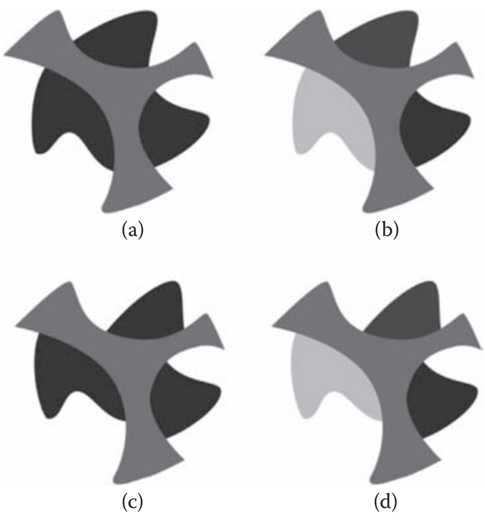

FIGURE 10.3 **(See color insert following p. 94.)** Contour and surface interpolation. (a) The three black regions appear as one object behind the gray occluder. Both contour and surface interpolation processes are engaged by this display. (b) Contour interpolation alone. By changing the surface colors of visible regions, surface interpolation is blocked. However, the relations of contours still engage contour interpolation, leading to some perceived unity of the object. (c) Surface interpolation alone. By disrupting contour relatability, contour interpolation is blocked. Due to surface interpolation, there is still some impression that the three fragments connect behind the occluder. (d) With both contour and surface interpolation disrupted, blue, yellow, and black regions appear as three separate objects.

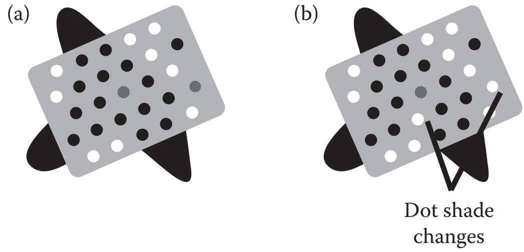

FIGURE 10.4 **(See color insert following p. 94.)** Illustration of two-dimensional surface interpolation. The circular areas in the display do not trigger contour processes, due to the absence of tangent discontinuities. Surface interpolation causes some circular areas to appear as holes in the occluder rather than as spots in front. Two dots in (a) are changed in color in (b), causing a difference in their appearance (e.g., the yellow spot in (a) when turned white becomes a hole due to its relation with the color of the surround). Relations of contour and surface interpolation are shown by blue spots appearing as holes if they fall within interpolated (or extrapolated) contours of the blue display.

on geometric relations of visible contour segments that lead into contour junctions. Surface interpolation in 2D displays can occur in the absence of contour segments or junctions; it depends on the similarity of lightness, color, or texture of visible surface patches.

Figure 10.4 illustrates the action of surface interpolation. Some of the circles in the display, such as the yellow ones, appear as spots on the surface. In contrast, most of the blue circles appear to be part of a single, occluded, blue figure, visible through holes. The white spots also appear to be holes rather than spots; through them, the white background surface is seen. These perceptual experiences arise from the surface interpolation process. Visible regions are connected across gaps in the input based on the similarity of their surface qualities (e.g., lightness, color, and texture). These connections cannot be given by contour interpolation, as the circles have no contour junctions. Certain rules govern surface interpolation; for example, it is confined by real and interpolated edges (Yin, Kellman, and Shipley, 2000). In the figure, note that the rightmost circle does not link up with the occluded object. This result occurs because that dot does not fall within real or interpolated contours of the blue object. Whereas contour interpolation processes are relatively insensitive to relations of lightness or color, the surface process depends crucially on these. Notice that the yellow dot on the lower left does *not* appear as part of the occluded object, despite being within the interpolated and real contours of the blue object.

This phenomenon of surface interpolation under occlusion appears to be one of a family of surface spreading or “filling-in” phenomena, such as the color-spreading phenomena studied by Yarbus (1967) and filling-in across the blind spot.

A MODEL OF CONTOUR INTERPOLATION IN STATIC 2D SCENES

Complementary processes of contour and surface interpolation work in concert to connect object fragments across gaps in the retinal image and recover the shape of occluded objects (e.g., Grossberg and Mingolla, 1985; Kellman and Shipley, 1991). Interpolated boundaries of objects, whether occluded or illusory, constrain spreading of surfaces across unspecified regions in the image, even if the interpolated boundaries are not connected to others (Yin, Kellman, and Shipley, 1997, 2000). Here, we briefly review the context for developing a 4D model of contour interpolation and surface perception.

The Geometry of Visual Interpolation

A primary question in understanding visual object and surface formation is what stimulus relationships cause it to occur? Answering this question is fundamental in several respects. It allows us to understand the nature of visual interpolation. Some visible fragments get connected, whereas others do not. Discovering the geometric relations and related stimulus conditions that lead to object formation is analogous to understanding the grammar of a language (e.g., what constitutes a well-formed sentence). Understanding at this level is also crucial for appreciating the deepest links between the physical world and our mental representations of it. Characterizing the stimulus relations leading to object formation is at first descriptive, but as unifying principles are revealed, they help us to relate the information used by the visual system to the physical laws governing the projection of surfaces to the eyes, whether these are deep constraints about the way the world works (e.g., Gibson, 1979; Marr, 1982) or scene statistics (e.g., Geisler et al., 2001).

Initiating Conditions for Interpolation

An important fact about contour interpolation is that the locations of interpolated contours are highly restricted in visual scenes. In general, interpolated contours begin and end at junctions or corners in visible contours (tangent discontinuities)—locations at which contours have no unique orientation (Shipley and Kellman, 1990; Rubin, 2001). Some have suggested that second-order discontinuities (points that are first-order continuous but mark a change in curvature) might also weakly trigger interpolation (Shipley and Kellman, 1990; Albert and Hoffman, 2000; Albert and Tse, 2000; Albert, 2001; for recent discussion see Kellman, Garrigan, and Shipley, 2005). Tangent discontinuities arise from the optics of how occluded objects project to the eyes: it can be proven that the optical projection of one object occluding another will contain these image features (Kellman & Shipley, 1991). Shipley and Kellman (1990) observed that, in general, interpolated contours begin and end at tangent discontinuities and showed that their removal eliminated or markedly reduced contour interpolation. Heitger et al. (1992) called tangent discontinuities "key points" and proposed a neurally plausible model for their extraction from images. The presence or absence of tangent discontinuities can be manipulated in illusory contour images by rounding the corners of inducing elements, which weakens contour interpolation (e.g., Albert and Hoffman, 2000; Kellman et al., 2005; Shipley and Kellman, 1990; Palmer, Kellman, and Shipley, 2006).

Contour Relatability

What determines which visible contour fragments get connected to form objects? Although tangent discontinuities are ordinarily necessary conditions for contour interpolation, they are not sufficient. After all, many corners in images are corners of objects, not points at which some contour passes behind an intervening surface (or in front, as in illusory contours).

Contour interpolation depends crucially on geometric relations of visible contour fragments, specifically the relative positions and orientations of pairs of edges leading into points of tangent discontinuity. These relations have been described formally in terms of *contour relatability* (Kellman and Shipley, 1991; Singh and Hoffman, 1999a). Relatability is a mathematical notion that defines a categorical distinction between edges that can connect by interpolation and those that cannot (see Kellman and Shipley, 1991, 175–177). The key idea in contour relatability is smoothness (e.g., interpolated contours are differentiable at least once), but it also incorporates monotonicity (interpolated contours

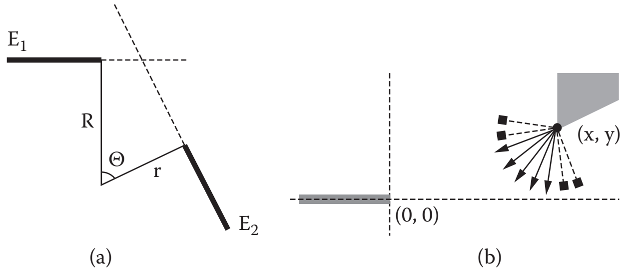

FIGURE 10.5 Contour relatability describes formally a categorical distinction between edges that can be connected by visual interpolation and those that cannot. (a) Geometric construction defining contour relatability (see text). (b) Alternative expression of relatability. Given one visible contour fragment terminating in a contour junction at (0,0) and having orientation 0°, those orientations θ that satisfy the equation $\tan^{-1}(y/x) \leq \theta \leq \Phi/2$ are relatable. In the diagram, these are shown with solid lines, whereas nonrelatable orientations are shown with dotted lines.

bend in only one direction), and a 90° limit (interpolated contours bend through no more than 90°). Figure 10.5 shows a construction that is useful in defining contour relatability. Formally, if E_1 and E_2 are surface edges, and R and r are perpendicular to these edges at points of tangent discontinuity, then E_1 and E_2 are relatable if and only if

$$0 \leq R\cos\theta \leq r \tag{10.1}$$

Although the precise shape of interpolated contours is a matter of some disagreement, there are two properties of relatability that cohere naturally with a particular class of contour shapes. First, it can be shown that interpolated edges meeting the relatability criteria can always be comprised of one constant curvature segment and one zero curvature segment. Second, it appears that this shape of interpolated edges has the property of being a minimum curvature solution in that it has lowest maximum curvature: any other first-order continuous curve will have at least one point of greater curvature (see Skeath, 1991, in Kellman and Shipley, 1991). This is a slightly different minimum curvature notion than minimum energy.

One Object or Two?

Relatability defines a categorical distinction—which relative positions and orientations allow edges to be connected by contour interpolation. Such a

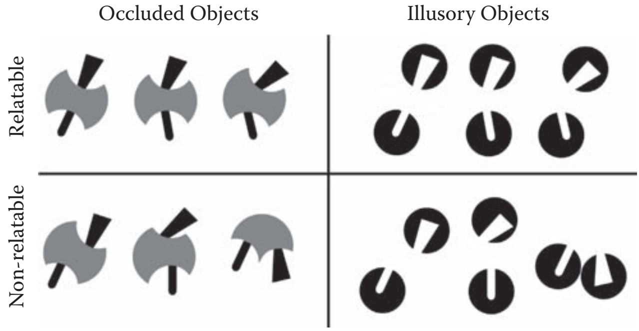

FIGURE 10.6 Examples of relatable and nonrelatable contours.

distinction is important, as object perception often involves a discrete determination of whether two visible fragments are part of the same object or not. Figure 10.6 shows examples of relatable and nonrelatable edges, in both perception of partly occluded objects and perception of illusory objects. Complete objects are formed in the top row but not in the bottom row. Object formation has profound effects on further processing, such as generation of a representation of missing areas, generation of an overall shape description, and comparison with items or categories in memory. Research indicates that the representation of visual areas as part of a single object or different objects has many important effects on information processing (Baylis and Driver, 1993; Zemel et al., 2002; Kellman, Garrigan, and Shipley, 2005).

Quantitative Variation

Although the discrete classification of visible areas as connected or separate is important, there is also reason to believe that quantitative variation exists within the category of relatable edges (Kellman and Shipley, 1991; Banton and Levi, 1992; Shipley and Kellman, 1992a, 1992b; Field, Hayes, and Hess, 1993; Singh and Hoffman, 1999b). For example, experiments indicating a decline to a limit around 90° were reported by Field, Hayes, and Hess (1993). Singh and Hoffman (1999a) proposed an expression for quantitative decline of relatability with angular change.

Ecological Foundations

The notion of relatability is sometimes described as a formalization of the Gestalt principle of good continuation (Wertheimer, 1923/1938). Recent work suggests that good continuation and relatability are separate but related principles of perceptual organization (Kellman et al., 2003). Both embody underlying assumptions about contour smoothness (Marr, 1982), but they take different inputs and have different constraints. The smoothness assumptions related to both of these principles reflect important aspects of the physical world as it projects to the eyes. Studies of image statistics suggest that these principles approach optimality in matching the structure of actual contours in the world. Through an analysis of contour relationships in natural images, Geisler et al. (2001) found that the statistical regularities governing the probability of two edge elements cooccurring correlate highly with the geometry of relatability. Two visible edge segments associated with the same contour meet the mathematical relatability criterion far more often than not.

3D CONTOUR INTERPOLATION

Object formation processes are three-dimensional. Figure 10.7 gives an example—a stereogram that may be free-fused by crossing the eyes. One sees a vivid transparent surface with a definite 3D shape. Object formation takes as inputs 3D positions and orientations of edges and produces as outputs 3D structures (Kellman and Shipley, 1991; Carman and Welch, 1992; Kellman et al., 2005; Kellman, Garrigan, and Shipley, 2005).

Until recently, there has been no account of the stimulus conditions that produce 3D interpolation. Kellman et al. (2005) proposed that 3D interpolation might be governed by a straightforward 3D generalization of

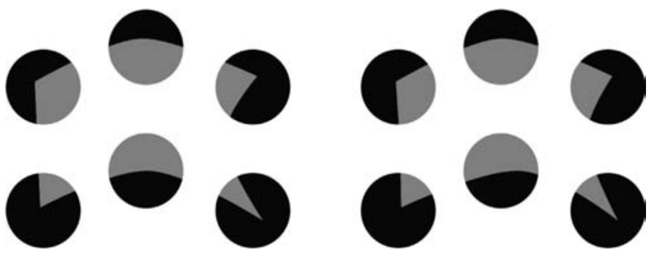

FIGURE 10.7 **(See color insert following p. 94.)** Three-dimensional (3D) interpolation. The display is a stereogram that may be free-fused by crossing the eyes. Specification of input edges' positions and orientations in 3D space (here given by stereoscopic disparity) leads to creation of a vivid, connected, transparent surface bending in depth.

2D relatability. As in the 2D case, interpolated contours between 3D edges must be smooth, monotonic, and bend no more than 90°. Similarly, where 3D interpolated contours meet physically given edges, the orientations of the physically given part and the interpolated part must match.

Formally, we define, for a given edge and any arbitrary point, the range of orientations that fall within the limits of relatability at that point. In the Cartesian coordinate system, let Θ be an angle in the x-y plane, and φ an angle in the x-z plane (for simplicity, in both cases zero degrees is the orientation parallel to the x-axis). Positioning one edge with orientation $\Theta = \varphi = 0$ and ending at the point (0, 0, 0), and positioning a second edge at (x,y,z) somewhere in the volume with $x > 0$, the range of possible orientations (θ, φ) for 3D-relatable edges terminating at that point are given by

$$\tan^{-1}\left(\frac{y}{x}\right) \le \theta \le \frac{\pi}{2} \tag{10.2}$$

and

$$\tan^{-1}\left(\frac{z}{x}\right) \le \varphi \le \frac{\pi}{2} \tag{10.3}$$

As in the 2D case, we would expect quantitative variation in the strength of interpolation within these limits. The lower bounds of these equations express the absolute orientation difference (180° for two collinear edges ending in opposite directions) between the reference edge (edge at the origin) and an edge ending at the arbitrary point oriented so that its linear extension intersects the tip of the reference edge. The upper bounds incorporate the 90° constraint in three dimensions.

How might the categorical limits implied by the formal definition of 3D relatability be realized in neural architecture? In 2D cases, it has been suggested that interpolation occurs through lateral connections among contrast-sensitive oriented units having particular relations (Field, Hayes, and Hess, 1993; Yen and Finkel, 1998). Analogously, 3D relatability specifies a "relatability field" or volume within which relatable contour edges can be located. At every location in the volume, relatable contours must have a 3D orientation within a particular range specific to that location.

The interpolation field suggests that, contradictory to some 2D models of contour interpolation, early visual cortical areas that do not explicitly code 3D positions and contour orientations may be insufficient for the neural implementation of 3D contour interpolation (Kellman, Garrigan, and

Shipley, 2005). As we discuss below, there are interesting considerations regarding exactly where the neural locus of contour interpolation may be.

Experimental Studies of Contour Interpolation

An objective performance paradigm for testing 3D contour relatability was devised by Kellman, Garrigan, Shipley, Yin, and Machado (2005) and is illustrated in Figure 10.8. In their experiments, subjects were shown stereoscopically presented 3D planes whose edges were either relatable or not. Examples of relatable and nonrelatable pairs of planes are shown in the columns of Figure 10.8. Orthogonal to relatability are two classes of stimuli, converging and parallel planes, shown in the rows of Figure 10.8. In these experiments, subjects were asked to classify stimuli like the ones shown as either parallel or converging. The idea is that, to the extent that 3D relatability leads to object formation, judging the relative orientations of 3D relatable planes should be easier than judging the relative orientations of 3D nonrelatable planes.

Kellman et al. (2005) found that subjects could make this classification more accurately and quickly when the planes were 3D relatable. This result

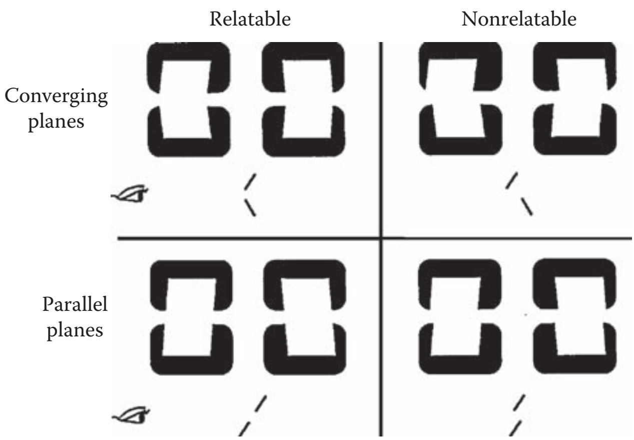

FIGURE 10.8 Experimental stimuli used to test three-dimensional (3D) object formation from 3D relatability. It was predicted that sensitivity and speed in classifying displays like these as either converging or parallel would be superior for displays in which unitary objects were formed across the gaps by contour interpolation, and that object formation would be constrained by 3D relatability. Both predictions were confirmed experimentally (Kellman et al., 2005).

is consistent with 3D relatability as a description of the geometric limits of 3D contour interpolation and object formation. A variety of other experiments indicated that the results depended on 3D interpolation, rather than some other variable, such as an advantage of certain geometric positions for making slant comparisons. (For details, see Kellman et al., 2005.)

3D SURFACE INTERPOLATION

Contour and surface processes often work in complementary fashion (Grossberg and Mingolla, 1985; Nakayama, Shimojo, and Silverman, 1989; Yin, Kellman, and Shipley, 1997, 2000; Kellman, Garrigan, and Shipley, 2005). Studies with 2D displays have shown that surface interpolation alone can link areas under occlusion based on similarity of surface quality. Surface similarity may be especially important in 2D, because all visible surface regions are confined to the same plane. In 3D, the situation is different. Here, geometric positions and orientations of visible surface patches may also be relevant.

We have recently been studying whether 3D surface interpolation depends on geometric constraints and, if so, how these relate to the constraints that determine contour interpolation. To study 3D surface interpolation apart from contour processes, we use visible surface patches that have no oriented edges. These are viewed through apertures (Figure 10.9).

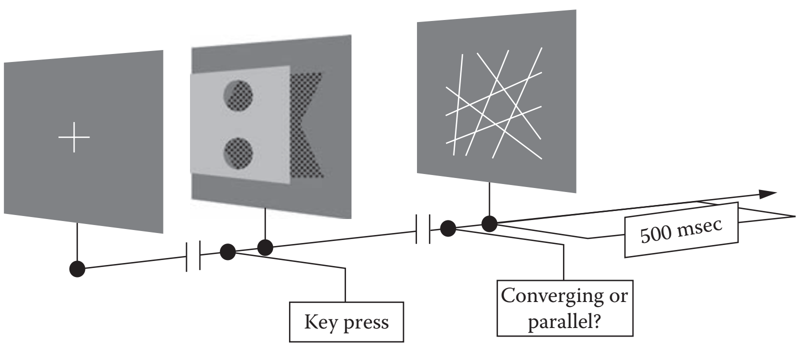

FIGURE 10.9 Use of the parallel/converging method for studying three-dimensional (3D) surface interpolation. A fixation point is followed by a display in which surface patches slanted in depth are viewed through two apertures. Participants make a forced choice as to whether the visible surface patches were in parallel or converging planes. (From Fantoni et al., 2008. With permission.)

We used a version of the parallel/converging method to study 3D surface interpolation. Displays were made of dot-texture surfaces; due to their lack of oriented edges, these surface patches could not support contour interpolation. Participants made a forced choice on each trial as to whether two surface patches, visible through apertures, lay in parallel or converging (intersecting) planes. As in 3D contour interpolation, we hypothesized that completion of a connected surface behind the occluder would facilitate accuracy and speed on this task. We also tested whether 3D relatability—applied to the orientations of surface patches rather than contours—might determine which patches were seen, and processed, as connected. 3D relatable patches were compared to displays in which one patch or the other was shifted to disrupt 3D relatability.

Figure 10.10 shows representative data on 3D surface interpolation (Fantoni et al., 2008). As predicted, 3D relatable surface patches showed sensitivity and speed advantages over nonrelatable surface patches. This effect was just as strong for vertically misaligned apertures as for vertically aligned ones. Consistent with a 90° constraint, the difference between 3D relatable and nonrelatable conditions decreased as the slant of each patch approached 45° (making their relative angle approach 90°). Many questions remain to be investigated, but these results suggest the fascinating possibility that both contour and surface interpolation in 3D share a common geometry (cf, Grimson, 1981). They may even be manifestations of some common process, although Kellman et al. (2005) showed that contour interpolation, not surface interpolation, was specifically implicated in their results.

The results of experiments on 3D surface interpolation support the notion that surface-based processes can operate independently of contour information, and that these processes are geometrically constrained by the 3D positions and orientations of visible surface patches. The pattern of results substantially replicates that of Kellman et al. (2005) for illusory contour displays, despite the lack of explicit bounding edges in the inducing surfaces. 3D relatability consistently affected speeded classification performance, by facilitating it for 3D relatable displays relative to displays in which 3D relatability was disrupted by both a depth shift of one surface relative to the other (violating the monotonicity constraint) and large values of relative stereo slant (violating the 90° constraint in converging displays).

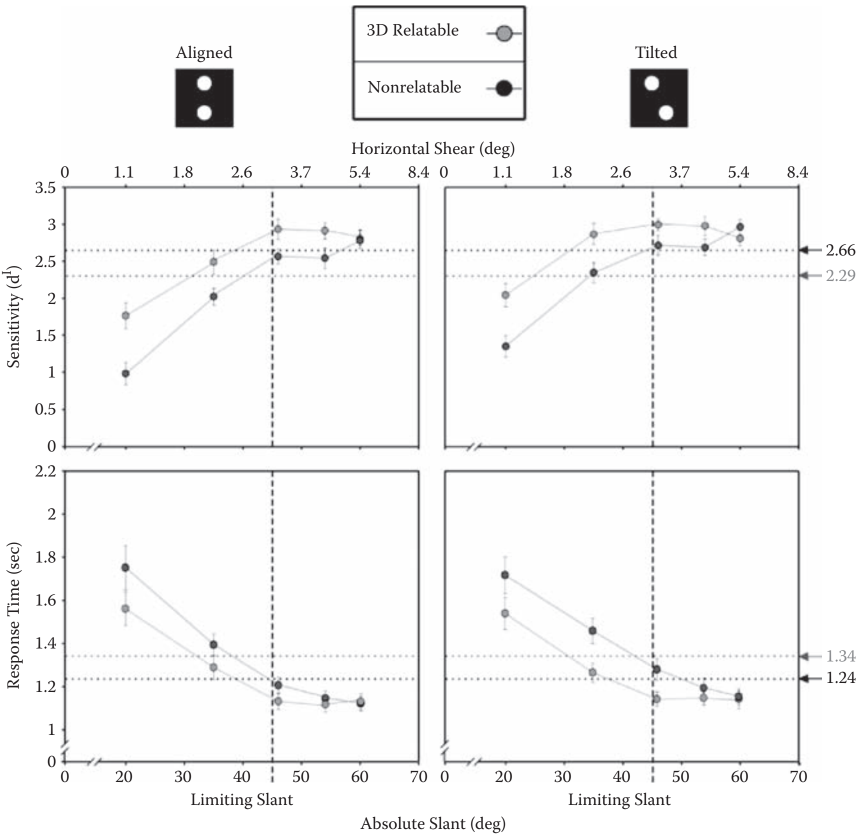

FIGURE 10.10 **(See color insert following p. 94.)** Three dimensional (3D) surface interpolation data. Sensitivity (upper panels) and response times (lower panels) for 3D relatable and 3D nonrelatable surface patches in aligned (right) and misaligned (left) aperture configurations. (From Fantoni et al., 2008. With permission.)

2D surface interpolation may constitute a special case of a more general 3D process. In 3D, the primary determinant of interpolation may be geometric relations, not similarity of surface quality. In our displays, position and orientation of surface patches seen through apertures were specified by binocular disparity, along with information from vergence. It appears that disparity provided sufficient information for the extraction of the 3D orientation of inducing patches necessary to constrain surface interpolation. The evidence suggests that contour and surface processes that surmount gaps in 3D are separable processes but rely on common geometric constraints.

SPATIOTEMPORAL INTERPOLATION

Despite the fact that perception in the laboratory is often studied with well-controlled, static, 2D images, ordinary perception usually involves diverse, moving, 3D objects. When an object is partially occluded and moves relative to the occluding surface, it is *dynamically occluded*. In such circumstances, shape information from the dynamically occluded object is discontinuous in both space and time. Regions of the object may become visible at different times and places in the visual field, and some regions may never project to the observer's eyes at all. Such cases are somewhat analogous to static, 2D occluded images, except that the partner edges on either side of an occluding boundary may appear at different times and be spatially misaligned.

For instance, imagine standing in a park and looking past a grove of trees toward a street in the distance. A car drives down the street from left to right and is visible beyond the grove. The car goes into and out of view as it passes behind the tree trunks and tiny bits and pieces of it twinkle through the gaps in the branches and leaves. You might see a bit of the fender at time 1 in the left visual field, a bit of the passenger door at time 2 in the middle of the visual field, and a bit of the trunk at time 3 in the right visual field. But what you perceive is not a collection of car bits flickering into and out of view. What you perceive is a *car*, whole and unified. In other words, your visual system naturally takes into account the constantly changing stimulation from a dynamically occluded object, collects it over time, compensates for its lateral displacement, and delivers a coherent percept of a whole object.

This feat of perception is rather amazing. Given that boundary interpolation for static images declines as a function of spatial misalignment, and given that the pieces of the car became visible at different places throughout the visual field, our perception of a coherent object is quite remarkable. The key is that the spatially misaligned pieces did not appear at the same time, but rather in an orderly temporal progression as the car moved. What unifies the spatial and temporal elements of this equation is, of course, motion. The motion vector of the car allows the visual system to correct for and anticipate the introduction of new shape information.

Palmer, Kellman, and Shipley (2006) considered the requirements for spatiotemporal object formation. One important requirement is *persistence*: In order to be connected with fragments not yet visible, a momentarily viewed fragment must be represented after it becomes occluded. (See Figure 10.11 at time t_0.) A second requirement is *position updating*. Not only must a previously viewed, moving fragment continue to be represented, its spatial position must be updated over time, in accordance with its previously observed velocity (Figure 10.11 at time t_1). Finally, previously viewed and currently viewed fragments are both utilized by processes of contour and surface interpolation, which connect regions across spatial gaps (Figure 10.11 at t_2). When new fragments of the object come into view, they are integrated with the persisting, position-updated

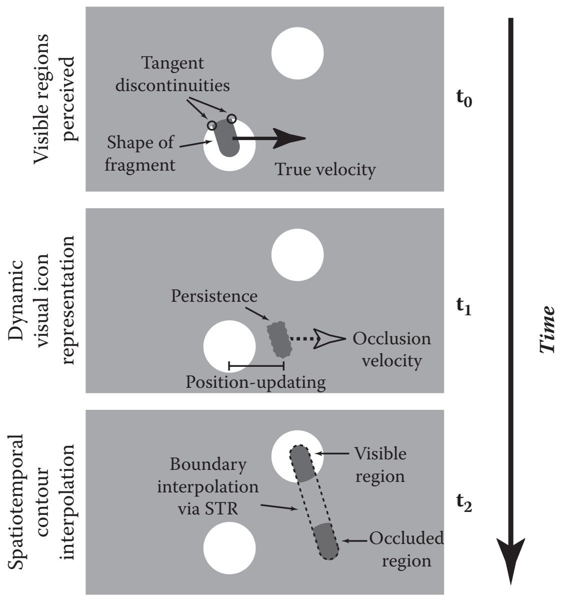

FIGURE 10.11 Spatiotemporal interpolation processes. An object moves from left to right behind an occluder with two circular windows. At t_0, shape and motion information about visible regions of the rod are perceived. At t_1, the shape and current position of the occluded region of the rod are represented in the dynamic visual icon representation. At t_2, another portion of the rod becomes visible, and contour interpolation occurs between the occluded and visible regions.

fragments via contour and surface interpolation processes that have previously been identified for static objects. As a result of persistence and updating processes, whether contours are interpolated is constrained by the same geometric relations of contour relatability (Kellman and Shipley, 1991; Kellman, Garrigan, and Shipley, 2005) that determine unit formation in static arrays.

As the spatial relationships that support interpolation are highly constrained, accurate representations of previously seen fragments are important for allowing spatiotemporal object formation to occur, and to operate accurately. Palmer, Kellman, and Shipley (2006) combined the requirements for object formation with proposals about visual mechanisms that represent, update, and connect object fragments over time in a model of spatiotemporal relatability (STR). The model provides an account for perception of dynamically occluded objects in situations such as that presented in Figure 10.1.

In a series of experiments, Palmer, Kellman, and Shipley (2006) found support for the persistence, position updating, and relatability notions of STR. The notion of persistence was supported because observers performed as if they had seen dynamically occluded objects for longer than their physical exposure durations. The notion of position updating was supported because observers were highly accurate at discriminating between two shape configurations that differed only in the horizontal alignment of the pieces. Because the dynamically occluded objects traveled horizontally and two partner edges on either side of an occluded region were not seen simultaneously, observers' accurate performance demonstrated that they had information about the locations of both edges despite the fact that, at all times, at least one was occluded. Finally, the notion that contour and surface interpolation processes operated in these displays was supported by a strong advantage in discrimination performance under conditions predicted to support object formation. Specifically, configurations that conformed to the geometric constraints of STR produced markedly better sensitivity than those that did not. The dependence of this effect on object formation was also shown in another condition, in which the removal of a mere 6% of pixels at the points of occlusion (rounding contour junctions) produced reliably poorer performance, despite the fact that the global configuration of the pieces was preserved. This last result was predicted from prior findings that rounding of contour junctions weakens contour completion processes (Albert, 2001; Shipley and Kellman, 1990; Rubin, 2001). When the contour interpolation process was compromised, the three projected fragments of the objects were less likely to be perceived as a single visual unit, and discrimination performance suffered.

Palmer, Kellman, and Shipley (2006) proposed the notion of a *dynamic visual icon*—a representation in which the persistence and position updating (and perhaps interpolation) functions of STR are carried out. The idea extends the notion of a visual icon representation, first discovered by Sperling (1960) and labeled by Neisser (1967), which accepts information over time and allows the perceiver to integrate visual information that is no longer physically visible. Perception is not an instantaneous process but rather is extended over time, and the visual icon is a representation that faithfully maintains visual information in a spatially accurate format for 100 ms or more after it disappears. The proposal of a dynamic visual icon adds the idea that represented information may be positionally updated based on previously acquired velocity information. It is not clear whether this feature is a previously unexplored aspect of known iconic visual representations or whether it implicates a special representation. What is clear is that visual information is not only accumulated over time and space, but that the underlying representation is geared toward the processing of ongoing events.

A yet unexplored aspect of the dynamic visual icon is whether it incorporates position change information in all three spatial dimensions. If so, this sort of representation might handle computations in a truly 4D spatiotemporal object formation process. Most studies to date, as well as the theory of STR proposed by Palmer, Kellman, and Shipley (2006), have focused on 2D contour completion processes and motion information. A more comprehensive idea of 3D interpolation that incorporates position change and integration over time has yet to be studied experimentally. Future work will address this issue and attempt to unify the 3D object formation work of Kellman, Garrigan, and Shipley (2005) with the STR theory and findings of Palmer, Kellman, and Shipley (2006).

FROM SUBSYMBOLIC TO SYMBOLIC REPRESENTATIONS

One way to further our understanding of contour interpolation processes is to build models of contour interpolation mechanisms and compare their performance to human perception. One such model takes grayscale images as input, and using simulated simple and complex cells, detects contour junctions (Heitger et al., 1992) and interpolates between them using geometric constraints much like contour relatability (Heitger et al., 1998).

The output of this model is an image of activations at pixel locations along interpolation paths. These activations appear at locations in images where people report perceiving illusory contours.

A more recent model (Kalar et al., 2010), generalizes the framework proposed in the model of Heitger et al. (1998) to interpolate both illusory and occluded contours. This model, which is a neural implementation of contour relatability and the identity hypothesis (that illusory and occluded objects share a common underlying interpolation mechanism), generates images of illusory and occluded contours consistent with human perception in a wide variety of contexts (e.g., Figure 10.12).

There is, however, an important shortcoming of models of this type. The inputs to these models are *images* where pixel values represent luminance. The outputs of the models are *images* of illusory and occluded (and real) contours. That is, these models take images that represent luminance at each pixel location and return images that represent illusory and occluded interpolation *activity* at each pixel location. There is nothing in the outputs to indicate that different pixels are connected to each other, that they form part of a contour, and so on. Nothing describes the areas that form a complete object, much less provides a description of its shape. In this sense, such models can be easily misinterpreted, as the observer viewing the outputs provides all of these additional descriptions.

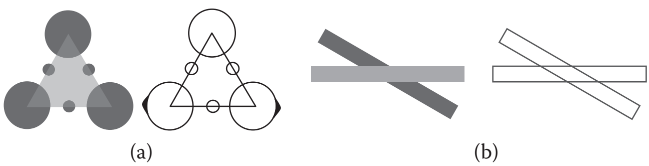

FIGURE 10.12 Displays and outputs from a filtering model that uses a unified operator to handle illusory and occluded interpolations. The model (Kalar et al., 2010) draws heavily on Heitger et al. (1998) but replaces their "ortho" and "para" grouping processes by a single operator sensitive to either L or T junction inputs. (a) Kanizsa-style transparency display on the left produces output on the right. The illusory contours would not be interpolated by a model sensitive to L junctions only (e.g., Heitger et al., 1998). Except for triangle vertices, all junctions in this display are anomalous T junctions. (b) Occlusion display on the left produces output on the right. This output differs from the Heitger et al. model, which is intended to interpolate only illusory contours.

To account for human perception, the important filtering information provided by these models must feed into mechanisms that produce higher-level, symbolic descriptions. Beyond representations of contours (both physically defined and interpolated) as sets of unbound pixel values must be a more holistic description, with properties like shape, extent, and their geometric relationships to other contours in the scene.

This is a very general point about research into object and surface perception. Vision models using relatively local filters describe important basic aspects of processing. At a higher level, some object recognition models assume tokens such as contour shapes, aspect ratios, or volumetric primitives as descriptions. The difficult problem is in the middle: How do we get from local filter responses to higher-level symbolic descriptions? This question is an especially high priority for understanding 3D shape and surface perception. Higher-level shape descriptions are needed to account for our use of shape in object recognition and our perceptions of similarity. As the Gestalt psychologists observed almost a century ago, two shapes may be seen as similar despite being composed of very different elements. Shape cannot be the sum of local filter activations. Understanding processes that bind local responses into unitary objects and achieve more abstract descriptions of these objects are crucial challenges for future research.

3D PERCEPTION IN THE BRAIN

Research in 3D and spatiotemporal perception also raises important issues for understanding the cortical processes of vision. A great deal of research and modeling has focused on early cortical areas, V1 and V2, as likely sites of interpolation processes (Mendola et al., 1999; Sugita, 1999; Bakin, Nakayama, and Gilbert, 2000; Seghier et al., 2000) for both illusory and occluded contours. On the basis of their results on 3D interpolation, Kellman, Garrigan, and Shipley (2005) suggested that interpolation processes involve all three spatial dimensions and are unlikely to be accomplished in these early areas. There are several reasons for this suggestion. First, orientation-sensitive units in V1 and V2 encode 2D orientation characteristics, which are not sufficient to account for 3D interpolation. Second, one could hypothesize that 2D orientations combined with outputs of disparity-sensitive cells might somehow provide a basis for 3D interpolation. Evidence indicates, however, that the type of disparity information available in these early areas is insufficient: whereas relative disparities are needed for depth computations, V1 neurons with disparity sensitivity appear to be sensitive to absolute disparities, which vary with fixation (Cumming and Parker, 1999). Third, even relative disparities do not directly produce perception of depth intervals in the world. There are two problems. One is that obtaining a depth interval from disparity involves a constancy problem; a given depth interval produces different disparity differences depending on viewing distance (Wallach and Zuckerman, 1963). To obtain a depth interval, disparity information must be combined with egocentric distance information, obtained from some other source, to at least one point. Figure 10.13 illustrates this problem, along with a second one. The experimental data of Kellman et al. (2005) suggest that edge segments at particular slants provide the inputs to 3D interpolation processes. Slant, however, depends not only on the depth interval between points, but also on their separation. Moreover, 3D slant may be specified from a variety of sources. The likely substrate of 3D interpolation is some cortical area in which actual slant information, computed from a variety of contributing cues, is available. These requirements go well beyond computations that are suspected to occur in V1 or V2.

Where might such computations take place? Although no definite neural locus has been identified, research using a single-cell recording in the

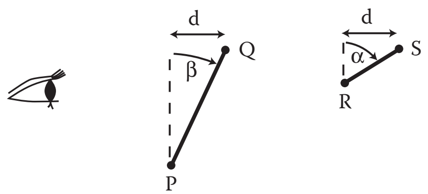

FIGURE 10.13 Relations between disparity, edge length, and slant. *Depth and disparity*: A given depth interval *d* in the world gives rise to decreasing disparity as viewing distance increases. Interval d given by points R and S produces smaller disparity differences than P and Q, if R and S are farther away. *Slant*: Obtaining slant from disparity depends not only on recovering the depth interval but also depends on the vertical separation of the points defining that depth interval ($\alpha > \beta$).

caudal intraparietal sulcus (cIPS) indicates the presence of units tuned to 3D slants (Sakata et al., 1997). It is notable that these units appear to respond to a particular 3D orientation regardless of whether that orientation is specified by stereoscopic information or texture information. These findings indicate where the kinds of input units required for 3D relatability—namely, units signaling 3D orientation and position—may exist in the nervous system.

In addition to the location of 3D processing, much remains to be learned about the nature of its mechanisms. Are there areas of cortex in which units sensitive to 3D positions and orientations of contour or surface fragments interact in a network that represents 3D relations? At present, we do not know of such a network, but the psychophysical results suggest that it is worth looking for.

All of these same sorts of questions apply as well to spatiotemporal object formation. We have impressive capabilities to construct coherent objects and scenes from fragments accumulated across gaps in space and time. Where in the cortex are these capabilities realized? And what mechanisms carry out the storage of previously visible fragments and their positional updating, based on velocity information, after they have gone out of sight? A striking possibility is that the same cortical areas are involved as those in 3D interpolation. At least, such an outcome would be consistent with a grand unification of processes that create objects from fragments. Although individual experiments have usually addressed 2D, 3D, and spatiotemporal interpolation separately, they may be part of a more comprehensive 4D process. Understanding both the computations involved and their neural substrates are fundamental and exciting issues for future research.