Geometry of shadows

David C. Knill

Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104

Pascal Mamassian

Department of Psychology, New York University, Washington Square, New York, New York 10012

Daniel Kersten

Department of Psychology, University of Minnesota, 75 E. River Road, Minneapolis, Minnesota 55455

Received December 5, 1996; accepted June 9, 1997

Shadows provide a strong source of information about the shapes of surfaces. We analyze the local geometric structure of shadow contours on piecewise smooth surfaces. Particular attention is paid to intrinsic shadows on a surface: that is, shadows created on a surface by the surface's own shape and placement relative to a light source. Intrinsic shadow contours provide useful information about the direction of the light source and the qualitative shape of the underlying surface. We analyze the invariants relating surface shape and light-source direction to the shapes and singularities of intrinsic shadow contours. The results suggest that intrinsic shadows can be used to directly infer illuminant tilt, qualitative global surface structure, and, at intersections with surface creases, the concavity/convexity of a surface. We show that the results obtained for point sources of light generalize in a straightforward way to extended light sources, under the assumption that light sources are convex. © 1997 Optical Society of America [S0740-3232(97)02212-6]

1. INTRODUCTION

Artists have long understood the importance of shadows for generating an impression of three-dimensionality in paintings.[1] Figure 1 shows two examples of how shadows can be used to depict surface shape or spatial displacement in scenes. Although both images are clearly two-dimensional depictions, they nevertheless induce strong percepts of three-dimensional structure. Figure 1(a) is an example of intrinsic shadows: shadows formed by an object on itself. Intrinsic shadows, with which we are primarily concerned in this paper, provide perceptually salient information about the shapes of objects and the direction of illumination in a scene.[2] Figure 1(b) is an example of extrinsic shadows: shadows cast on one object by another. Extrinsic shadows provide particularly salient cues to the relative positions and orientations of objects.[3,4] The information provided by extrinsic shadow contours about surface geometry has been extensively analyzed by Shafer and Kanade.[5]

One view of intrinsic shadows is that they provide a coarse description of the shading pattern on a surface. The information provided by shadows, however, is qualitatively different from that provided by shading, in that the former is fundamentally contained in the geometric relations between the curves formed by shadow boundaries and surface and lighting geometry. Furthermore, shadow contours in natural viewing conditions are not simply a special case of isophotes. Isophotes (curves of constant luminance in an image) are determined by multiple factors, including internal reflections on a surface and between surfaces. To the extent that one can model the geometric relations between isophotes and surface geometry for simplified, local reflectance models,[6] these relations are not likely to generalize to real scenes. Shadows, on the other hand, contain multiple cues for segmentation and thus, at least in theory, may be accurately measured in natural images. As a first step in studying the perceptual interpretation of intrinsic shadows, therefore, we have undertaken a theoretical analysis of their information content. In this paper we present the results of this analysis and suggest a number of implications they could have for human visual perception.

A. Approach

One approach to studying the information content of shadows would be to formulate a "shape-from-shadows" problem within the classical inverse-optics framework so often applied to problems in computer vision.[7–10] In this context, points along intrinsic shadow boundaries provide constraints on surface shape that can be used as boundary conditions for surface interpolation within shadow regions, much as occluding contours can be used to constrain shape from shading. The main constraints provided by shadow boundaries are that (1) the surface normals at points along an attached shadow contour are perpendicular to the lighting direction and (2) attached and cast shadow points that lie on the same illuminant ray in the image lie on the same illuminant ray in three dimensions (they have the same height in a light-source-centered coordinate system).

Although it is potentially useful in shape reconstruction algorithms, the approach described above does not make explicit the nature of the information provided by intrinsic shadows about surface shape (as well as about

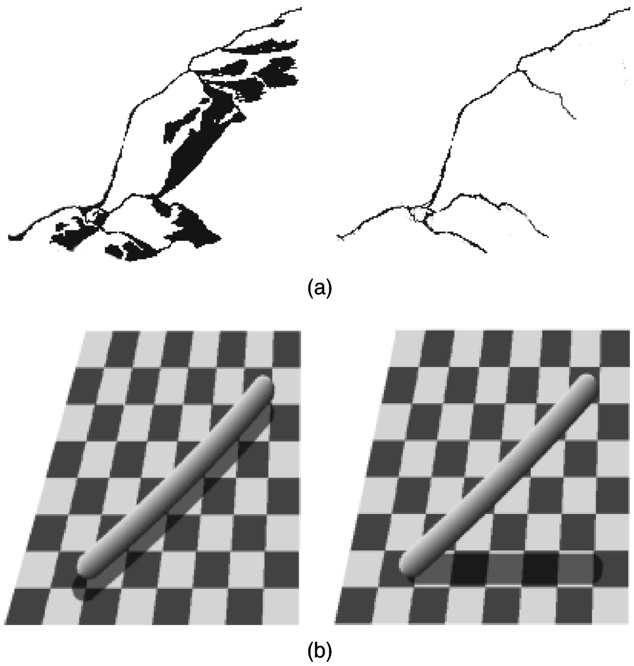


Fig. 1. Examples that clearly demonstrate the role of shadows for the perception of three-dimensional surface geometry. (a) Artists commonly use intrinsic shadows in pencil and ink sketches to induce a sense of three-dimensional surface shape—much more surface structure is evident in the left image than in the right, in which the shadows have been removed. (b) The shadow cast by one object on another provides perceptually salient information about the relative placement and orientations of the objects in three dimensions. The only difference between the two images in (b) is the orientation of the shadow cast by the pole, yet the pole in the right image appears more upright relative to the background surface than does the pole in the left image.

scene illumination). Our intuition tells us that it is the shapes of intrinsic shadow boundaries that directly provide information about surface shape and illumination. In this paper, therefore, we take a direct approach to specifying the information content of intrinsic shadows by analyzing the geometric constraints on the shapes of intrinsic shadows and the relationship between the behavior of shadow contours and surface shape and illumination. This is akin to the approach taken by Shafer and Kanade[5] to understanding the geometry of extrinsic shadow contours, though the requisite mathematics is more in the spirit of Koenderink's differential geometric approach to contour interpretation. The goals of the paper are twofold. First, we hope to provide a complete characterization of the local geometric and global topological behavior of intrinsic shadow contours. An understanding of this behavior is, we feel, a prerequisite for studying many problems involving intrinsic shadows. For a theoretical discourse of this type to be useful to the visual scientist, however, we must at least point to its potential implications for human vision; therefore a second goal of the paper is to outline these implications and to suggest possible predictions for human perception.

B. Outline

We will analyze five aspects of intrinsic shadow contour geometry:

- The global properties of intrinsic shadow contours and their relationship with surface shape, including the evolution of intrinsic shadows as an object is moved relative to a light source (creation, destruction, merging, and splitting of shadows).
- The relationship between the local shape of intrinsic shadow contours and surface shape and illuminant direction.
- The types and natures of singularities in intrinsic shadow contours and their relationship to surface structure.
- The behavior of intrinsic shadow contours at intersections with surface creases.
- The behavior of intrinsic shadow contours at intersections with smooth occluding contours.

Since we are targeting both computational and psychophysical audiences, the theoretical discussion will be at times technical and at times informal, and we will include both rigorous mathematical proofs and qualitative summaries of results. The paper is organized so that a reader who does not wish to wade through the details of the definitions and proofs can do so without missing the core of the discussion. We will explicitly indicate sections containing technical proofs that can be skipped by the more casual reader.

In Section 2 we provide a qualitative characterization of how intrinsic shadows are formed on smooth surfaces, for both point and extended sources of light. In Section 3 we describe the geometry of intrinsic shadows on smooth surfaces illuminated by a point source of light. The analysis in this section is, for the most part, a straightforward extrapolation of earlier analyses of smooth occluding contours[11,12] and of singularities in surface shading patterns[6,13]; therefore we derive most of our results by analogy with these earlier analyses and do not reconstruct the corresponding mathematical derivations. In Section 4 we extend the analysis to smooth surfaces illuminated by extended light sources. Since this section requires a different mathematical analysis from what has been presented before, it will be substantially more technical than previous sections. We will, however, summarize the major results at the end of the section. Section 5 will present an analysis of the behavior of shadow contours at or near creases of piecewise smooth surfaces. In Section 6 we will discuss the implications of the theoretical results for perception and suggest possible ways in which intrinsic shadow information might be used by the visual system for the perception of surface shape, illumination direction, and contour labeling.

2. SHADOW FORMATION

We begin our discussion with a qualitative description of how intrinsic shadows are created on smooth objects (Fig. 2). Two types of shadow regions may be contained in an intrinsic shadow: attached and cast. Attached shadow regions contain points on a surface that face away from a light source. If the light source is extended in space (e.g., a light bulb), the constraint must hold for the entire sheaf of rays connecting the point on the surface to points on the surface of the light source. A cast shadow region contains points that face a light source but are occluded from it by a distal part of the surface. Cast shadows on a surface illuminated by an extended light source have penumbra; that is, fuzzy boundaries formed by the gradual transition between fully illuminated regions of a surface and fully shadowed regions. For purposes of definition, we do not consider penumbra to be part of cast shadow regions; that is, we define cast shadows to contain surface points that are occluded from all points on a light source. The boundaries of cast shadows, then, are the interior edges of penumbra.

For the sake of analysis and discussion, we will distinguish between visible and invisible shadow boundaries. As we have defined them, shadow regions may be nested within one another or may abut one another along some segment of their boundaries. A large hill, for example, may cast a shadow over small hills, yet by our definition of shadow regions, the smaller hill might still have an attached shadow on it, though that shadow would be invisible. The boundaries of the small hill's attached shadow would then be invisible (regardless of viewpoint), as it would be hidden by the larger shadow. Cast shadow regions within intrinsic shadows are "created" by attached shadow regions on a surface (Fig. 2) and always abut attached shadow regions. We will refer to the attached shadow in such a pair as the parent of the cast shadow. Not all intrinsic shadows have cast shadow regions, as illustrated by the dark side of a ball, whose shadow is entirely attached; thus intrinsic shadows on surfaces are either entirely attached or contain both attached and cast shadow regions, but they are never entirely cast.

We will define shadow regions by their boundary curves and characterize the shadow formation process by the relationship between light sources and these boundary curves on surfaces. Figure 2 shows the geometric constructions that allow one to trace out the shadow boundaries on surfaces. For a light source at infinity, the light rays are parallel and attached shadow boundaries are formed anywhere that a light ray is tangent to a surface; that is, where the surface normal is perpendicular to the direction of illumination. The sheaf of rays passing through points on an attached shadow boundary form an imaginary surface, which, using the terminology of Shafer and Kanade,[5] we will henceforward refer to as the illumi-

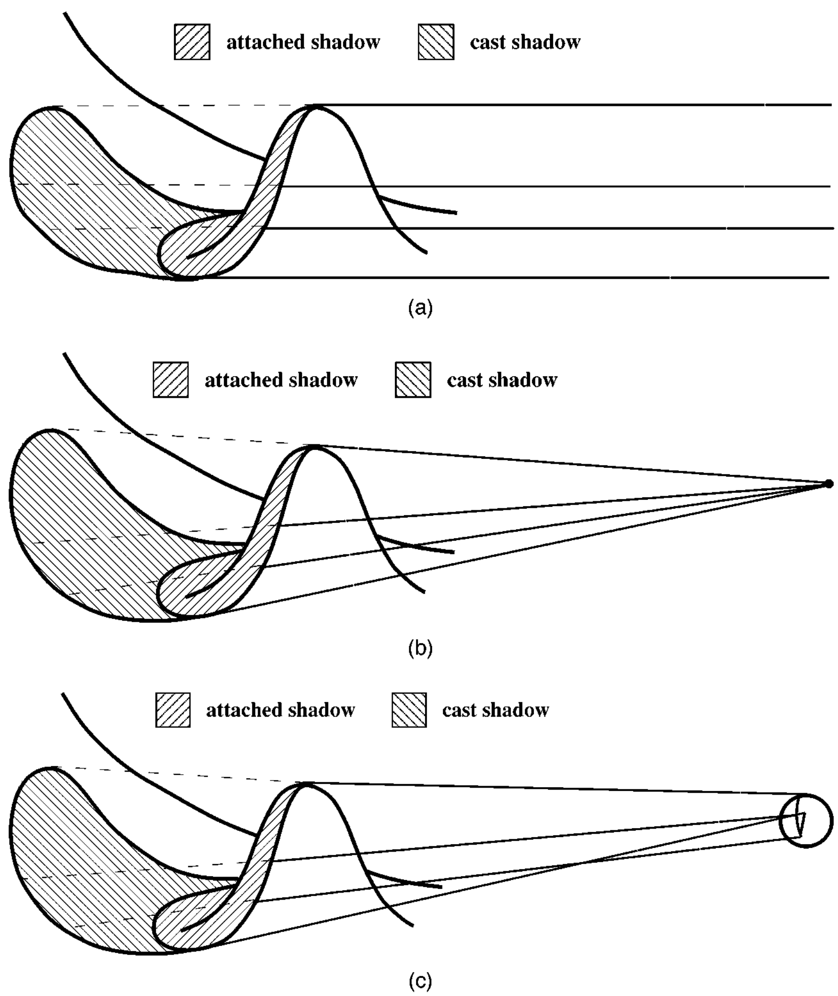


Fig. 2. Regions of a surface facing away from a light source are said to be in attached shadow. For point sources [(a) and (b)], the boundaries of attached shadows are formed where light rays from the source are cotangent to the surface. For extended sources [(c)], attached shadow boundaries are formed at points on a surface that share a tangent plane with points on the light source. The rays connecting attached shadow boundaries to points on the light source that share a common tangent plane form an envelope over the surface and the light source that is cotangent to both. When a surface region in attached shadow occludes part of the surface from the light source, a cast shadow is formed that is contiguous with the attached shadow.

nation surface. For a point source at infinity, the illumination surface is cylindrical. For a point source a finite distance from a surface [as in Fig. 2(b)], the light rays may be visualized as radiating from the source, and, as before, attached shadow boundaries are formed where any of these rays lie in the tangent plane of the surface. In this situation imaginary illumination surfaces are conical, being formed by the sheafs of rays connecting the source to points on attached shadow boundaries. The first points of intersection (away from the attached shadow boundary) between the illumination surface and the physical surface form the boundary of an attached shadow's child cast shadow.

For extended light sources, one can intuitively see that a surface point will be visible to a point on the surface of the light source if the two surfaces face each other at that point [Fig. 2(c)]. As we have defined it, a surface point is in an attached shadow if it faces away from all points on an extended light source. The boundary of such a region is formed at surface points where one can draw a ray between the surface point and a point on the light source that is in the local tangent planes of both the surface and the light source. The imaginary illumination surface passing through an attached shadow boundary is therefore a surface that is simultaneously tangent to both the illuminated surface and the light source (Fig. 3).

In general, objects that are bounded in space (having compact surfaces) have at least one intrinsic shadow whose boundary is entirely attached. Such a shadow will migrate over a surface as the surface is moved relative to a light source, but it will never be destroyed. We therefore refer to this shadow as an object's basic shadow. Other shadows on an object may be created or destroyed as the object is moved relative to a light source. To picture this, imagine taking a series of aerial photographs of a desert over the course of a day. At night, the entire desert is in shadow, but as the Earth rotates and the Sun crosses over the sky from dawn to dusk, the shadow breaks into successively more and smaller shadows on the dunes. As noon approaches, the shadows shift and many disappear. More new ones appear as the afternoon progresses and eventually merge back into one at nightfall. Similar events occur on a smaller scale as a surface rotates relative to a light source.

The singular events in the dynamic evolution of shadows come in two basic types: the creation or destruction of new shadow regions and the merging or splitting of shadows. To understand the nature of these events and where they occur on surfaces, we draw on an analogy between shadow boundaries and smooth occluding contours. For a point source of light, intrinsic shadow boundaries are formed at the loci of points on a surface that project to occluding contours from the viewpoint of the light source. We can therefore generalize results obtained by Koenderink and van Doorn[11] and Koenderink[12] on the singular events in the evolution of smooth occluding contours to similar events in the evolution of shadow boundaries. The results also hold for extended light sources, which can be treated as compact collections of point sources.

- Shadow creation/destruction (row 1 of Fig. 4). New shadows are always created at parabolic points of a surface. Similarly, shadows are destroyed at parabolic points. When a new shadow is created at a parabolic point, the shadow spreads to either side of the parabolic line; thus it contains both elliptic and hyperbolic regions of a surface. It also necessarily has both attached and cast shadow regions.

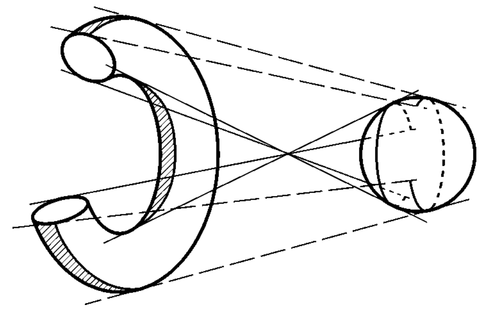


Fig. 3. The attached shadow on a toroid has two disjoint boundaries: one formed by the exterior of the donut, the other formed by the interior. The illumination surface associated with the interior boundary has the interesting property that it is self-intersecting. This is always true for that part of an illumination surface associated with an attached shadow boundary in a hyperbolic surface patch.

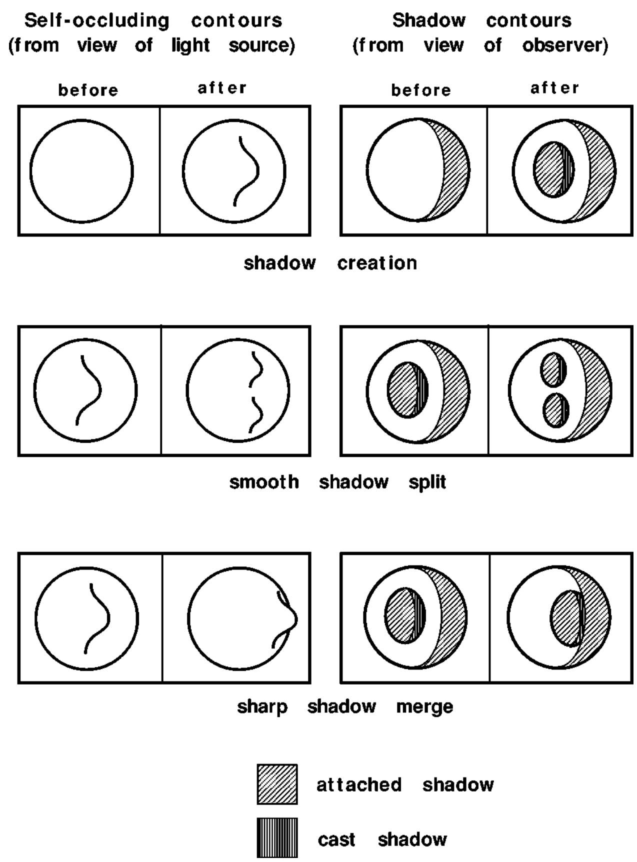


Fig. 4. The evolutionary events in shadow boundaries that occur as a light source is moved relative to a surface are qualitatively similar to events in the evolution of occluding contours that occur with motion of an observer relative to a surface. The first row of the figure shows an event corresponding, in this case, to the creation of a shadow at a bump on a surface, although in general it might occur at a dimple. The second row shows the splitting of a shadow into two, in this case, corresponding to two bumps on a surface. The last row shows an example of two intrinsic shadows merging as a result of occlusion of one attached shadow by another.

• Shadow merge/split (rows 2 and 3 of Fig. 4). There are two ways that shadows can merge together (or split apart) on a surface. From the point of view of the light source, these correspond to merging two smooth occluding contours into one smooth occluding contour (Fig. 4, row 2) or forming a T junction between smooth occluding contours (Fig. 4, row 3). The first type of merge occurs at parabolic points on surfaces. It results in the merging of both attached shadow boundaries and their child cast shadow boundaries. The intrinsic shadow boundaries resulting from such a merge are smooth. The second type of merge creates a junction between unrelated attached and cast shadow boundaries and results in a concave L junction in the visible boundary, one arm of which is attached, the other arm of which is cast.

3. SHADOWS ON SMOOTH SURFACES: POINT LIGHT SOURCES

Shadows formed on smooth surfaces by point sources of light provide the easiest case for analysis, since most of the results follow naturally from previous analyses of smooth occluding contours.[11,12] Attached shadow boundaries are curves on surfaces that would appear as self-occlusion boundaries from the viewpoint of the light source. Figure 5 shows the geometry relating the two types of boundaries. In formal terms, we say that attached boundaries are the preimages on a surface of self-occluding contours from the point of view of a light source. These curves have a number of interesting properties:

• They are everywhere smooth.

• The surface is convex in the direction of the light source over the visible extent of an attached shadow boundary. Visible attached shadow boundaries therefore traverse regions of a surface which are either hyperbolic, parabolic, or convex elliptic, but *not* concave elliptic.

• At points where they change from being visible to being invisible (end points of occluding contours), they are tangent to asymptotic directions on a surface (point b in Fig. 5) and to the light-source direction; thus surfaces are hyperbolic at these points.

The point at which an attached shadow boundary becomes invisible is a junction between the attached shadow boundary and a cast shadow boundary. We refer to such a junction as a parent–child junction, because it connects the boundary of a parent attached shadow to the boundary of its child cast shadow (as defined in Section 2). The other type of junction that can occur between attached and cast shadow boundaries corresponds to what would appear to the light source as a T junction in an occluding contour. Such a junction maps to two concave L junctions between unrelated cast and attached shadow boundaries (point a in Fig. 6).

What information, then, is provided by the shadow contours that result from projecting intrinsic shadow boundaries into an image? We answer this question individually for each of the qualitatively different points along such a contour.

A. Regular Points of Attached Shadow Contours (Point *a* in Fig. 5)

As mentioned above, a surface is convex in the direction of the light source at visible points of an attached shadow [excepting the end points, where attached shadow contours join with their children cast shadow contours, see Subsection 3.B]. Although this is useful information in its own right, the shape of attached shadow contours potentially provides more quantitative information about surface shape. We might hope to extract information about surface shape similar to that provided by smooth occluding contours, for example, in the relationship between contour curvature and surface curvature (e.g., the sign of curvature of an occluding contour indicates whether a surface is convex elliptic or hyperbolic). Unfortunately, since we view shadow boundaries from arbitrary viewpoints, no such simple invariants hold at regular points on attached shadow contours. In Appendix A we derive the quantitative relationships between attached shadow contour shape and surface shape. The local orientation of a shadow contour is related to the local curvature and orientation of the underlying surface. The curvature is determined by these plus the local third-order structure of a surface (derivatives of curvature). Although local attached shadow contour shape seems to be a weak form of static information about local shape, one could potentially use controlled viewer motion to extract reliable information from dynamic changes in contour orientation and curvature (much as this strategy has been applied to occluding contours[14]).

B. Parent–Child Intersection between Attached and Cast Shadow Contours (Point *b* in Fig. 5)

The parent–child junction between attached and cast shadows is a particularly interesting point on an intrinsic

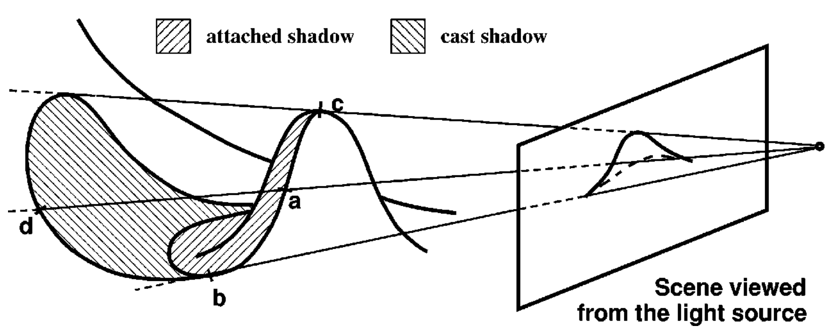


Fig. 5. We can distinguish several qualitatively distinct points on an intrinsic shadow contour (see text for discussion): a, regular points along attached shadow contours; b, points of intersection between cast-shadow contours and their associated attached-shadow contours; c, points of intersection between attached shadow contours and occluding contours; and d, regular points of cast shadow contours.

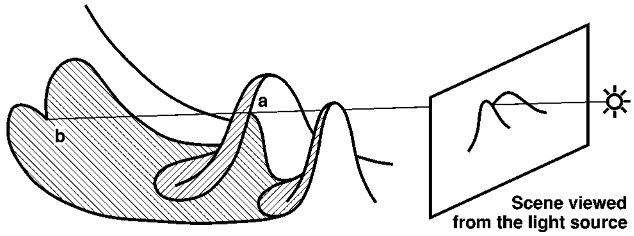


Fig. 6. When the light source's view of a surface contains a T junction in the occluding contour, an L junction is formed between one hill's cast shadow and another's attached shadow. Similarly, an L junction is formed at the junction of the two hills' cast shadows. All of these singular points lie along the same ray from the light source.

shadow contour, because at this point we know two things: (1) that the asymptotic direction of the surface is in the direction of the shadow contour and (2) that the light source tilt, i.e., its direction in the image plane, is in the direction of the shadow contour (leaving the slant of the light source away from the line of sight of a viewer undefined). Suppose that the visual system can detect junctions between attached and cast shadows, for example, by using luminance information across the contours for contour labeling.[15] It would clearly be useful to tell if it is a parent–child intersection or an intersection between unrelated shadow contours. A study of the behavior of intrinsic shadow boundaries at parent–child intersections reveals an interesting fact (and one which we have not found analogs to anywhere in the literature on occluding contours): the boundary is smooth at such an intersection, in the sense that the tangent directions of a parent attached shadow boundary and its child cast shadow boundary are the same at the intersection (as drawn in Fig. 5; see Appendix B for a proof of this result). Generically, unrelated attached and cast shadows will form an L at their intersection; thus the geometric behavior of the contours at an intersection indicates the nature of the intersection. We should also point out that knowledge of the light-source tilt (for a point source at infinity) would allow one to detect parent–child intersections by finding points on an intrinsic shadow contour that are tangent to that direction.

C. Intersections between Attached Shadow Contours and Smooth Occluding Contours (Points *c* and *d* in Fig. 5)

Two types of intersection between attached shadow contours and occluding contours may occur, corresponding to whether the surface on which the visible attached shadow boundary sits belongs to the occluding contour (point *c*). It is only the former point of intersection that is of special interest. At such an intersection, shadow contours are generically cotangent to smooth occluding contours (like other contours that project from markings on a surface).[16] Unlike at regular points along an attached shadow contour, the curvature of the contour is determined entirely by the orientation and curvature of the surface at the intersection point and not by the derivatives of curvature. Since the surface orientation is given by the normal to the occluding contour, the curvatures of the shadow contour and the occluding contour together provide two constraints on the three unknowns needed to specify local surface curvature at the point (see Appendix A). Although this point is theoretically true, we suspect that obtaining reliable measures of these curvatures would be difficult, making the information provided rather weak. We therefore do not dwell on this. On the other hand, the point of intersection does provide very reliable information about light-source direction. For a point source at infinity, it gives the tilt of the light source in the image plane. For a point source a finite distance from the surface, a ray drawn along the tangent direction passes through the projection of the light source into the image. Localization of two points of intersection between attached shadows and occluding contours thus determines the projected position of the light source in the image, which would be given by the intersection of rays drawn along the tangents of both points.

A corollary to the above result is that the tangent directions of attached shadow contours at intersections with occluding contours and with child cast shadow contours are parallel, for point sources at infinity, or intersect at a common point, for finite point sources. This provides a direct way to detect child–parent shadow contour intersections without relying on the mediating variable of light-source direction.

D. Regular Points on Cast Shadow Contours (Point *d* in Fig. 5)

The local shape of a cast shadow contour is determined by a large number of factors. These include the direction of light source, the local shape and orientation of a surface, and the shape and orientation of the surface along the attached shadow boundary from which the cast shadow was formed (see Ref. 5 for a full analysis). The local shape of the cast shadow contour component of an intrinsic shadow contour would therefore seem to be particularly uninformative about surface shape. As it turns out, however, we can derive a useful invariant relating the behavior of cast shadow contours at intersections with surface creases to both light-source direction and surface shape. We leave this to Section 5, which concerns the behavior of shadows on surfaces with creases.

4. SHADOWS ON SMOOTH SURFACES (EXTENDED LIGHT SOURCES)

Analyses of shading information and models of shape from shading generally rely on the assumption that the directional component of illumination comes from a point light source.[17,18] Many environments in which humans operate violate this assumption, containing as they do extended light sources. It is therefore important to ask how analyses and models using shading information generalize to extended light sources, if at all. For our purposes, we need to ask whether the behavior of shadow contours changes substantially when the illuminant is changed from a point source to an extended source. The answer is clearly no for light sources that are extremely small relative to the size of an object, in which case the point-source assumption holds approximately. We are concerned, however, with cases in which the spatial extent of light sources (when projected onto a surface) is on a scale not much smaller than that of the objects. Many objects with which humans are concerned—for example, the Sun—cannot be treated as simple point sources of light.

The mathematical tools needed to analyze the geometry of intrinsic shadows created by extended light sources are qualitatively different from those used to derive results for point sources of light (which follow simply from results on the singularities of projective mappings of smooth surfaces,[19] as applied to occluding contours). The mathematical tools we will use for our analysis are drawn from the elementary differential geometry of curves and surfaces. Because of the novelty of the analysis, we will describe it in detail in this section. The basic result concerns an abstract geometrical property of attached shadow boundaries, from which the more concrete consequences described above for point sources derive. The reader not interested in a detailed geometrical analysis of extended light sources may skip sections 4.A and 4.B, which contain the bulk of the mathematics. The end result is that the general characteristics described for intrinsic shadows with point light sources also hold for simply convex extended light sources (such as a ball or ellipsoid), with the caveat that light-source direction varies over the extent of a shadow boundary. Some of the results do not hold for nonsimply convex light sources (light sources with hyperbolic and possibly concave-elliptic regions), and these will be pointed out where they occur.

A. Formal Characterization of Intrinsic Shadow Boundaries Created by Extended Light Sources

In order to perform our analysis, we need to rigorously define intrinsic shadow boundaries on surfaces. A first step toward this end is to define the conditions that hold for points in shadow. In particular, we are concerned with points within attached shadows. Let Σ represent a smooth surface and Λ the surface of an extended light source. A point on the surface, $x_\Sigma \in \Sigma$, is said to be in an attached shadow if and only if it faces away from all points on the light source, $x_\Lambda \in \Lambda$, a condition that we state more formally as the condition

$$(\forall x_\Lambda \in \Lambda)\langle \mathbf{N}_\Sigma(\mathbf{x}_\Sigma), \mathbf{x}_\Lambda - \mathbf{x}_\Sigma\rangle < 0; \qquad (1)$$

that is, for each and every point on the surface of the light source, the angles between the rays from that point to points within an attached shadow are all in the range [90°, 270°].

The boundary of an attached shadow on a surface is clearly a curve. For surfaces with holes, this curve may be disjoint (i.e., the boundary may consist of two curves, for example, on the inside and outside of a donut). For simplicity, we will consider attached shadows that have a single connected curve for a boundary, though the analysis holds for each of a set of disjoint curves that may constitute an attached boundary on a surface with holes. Without loss of generality, we will assume this curve to be parameterized by its arc length normalized by the total length of the curve. Thus an attached shadow boundary is a curve on the surface, $\alpha : \alpha(t); 0 \leqslant t < 1$. Since illuminated points on a surface are defined by the condition that there exist points on a light source for which the sign of inequality in Eq. (1) is reversed, this curve is defined by the condition that at every point on the curve, the surface normal either points away from or is perpendicular to every point on the light source [satisfies Eq. (1)], *and* is perpendicular to at least one such point. That is, α is defined by the condition

$$\begin{aligned} &(\forall t \in [0,1)) \quad (\exists \mathbf{x}_\Lambda \in \Lambda) \quad \langle N_\Sigma[\alpha(t)], \quad \alpha(t) - \mathbf{x}_\Lambda\rangle = 0 \\ &\qquad \textit{and} \qquad (\forall \mathbf{x}_\Lambda \in \Lambda) \quad \langle N_\Sigma[\alpha(t)], \quad \alpha(t) - \mathbf{x}_\Lambda\rangle \geqslant 0 \end{aligned} \qquad (2)$$

One can easily show that condition 2 holds for points on the surface that share a tangent plane with the light source and that have parallel surface normals $\{\mathbf{N}_\Sigma[\alpha(t)] = \mathbf{N}_\Lambda(x_\Lambda)\}$ (see Fig. 7). We can therefore rewrite (2) as

$$(\forall t \in [0,1)) \quad (\exists \mathbf{x}_\Lambda \in \Lambda) \quad \langle N_\Sigma[\alpha(t)], \ \alpha(t) - \mathbf{x}_\Lambda\rangle = 0 \quad \textit{and} \quad N_\Sigma[\alpha(t)] = N_\Lambda(\mathbf{x}_\Lambda). \qquad (3)$$

The illumination surface connecting an extended light source to points on an attached shadow boundary is tangent to both the light source and the surface. This surface is necessarily developable (i.e., it can be unfolded, without stretching or compression, into a plane), a fact that will prove useful later in the paper.[20] Each point on an attached shadow has a corresponding point on the surface of the light source that matches condition 3. We will refer to this point as the light source's image of the corresponding point on the attached shadow boundary. The light source's image of an attached shadow boundary is itself a curve on the light source, which we will write as $\lambda : \lambda(t); 0 \leqslant t < 1$. We will assume λ to be parameterized so that $\lambda(t)$ is the light source's image of $\alpha(t)$. Rewriting condition 3 in terms of the attached shadow boundary and its light source image, we obtain

$$(\forall t \in [0,1)) \quad \langle N_\Sigma[\alpha(t)], \ \alpha(t) - \lambda(t)\rangle = 0 \quad \textit{and} \quad N_\Sigma[\alpha(t)] = N_\Lambda[\lambda(t)]. \qquad (4)$$

A cast shadow boundary is formed by the intersection of an attached shadow boundary's associated illumination surface with the surface being illuminated. By this definition, cast shadow boundaries are the internal edges of

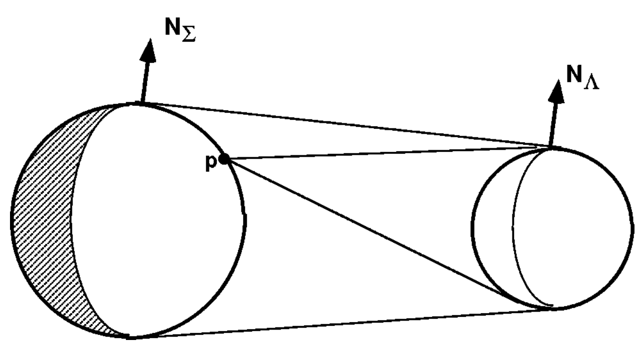


Fig. 7. Points on a surface, such as **p**, that face toward at least some parts of an extended light source, are said to be illuminated. Points on a surface that face away from all points on a light source are in shadow. The boundary between the two regions is formed where the envelope of the surface and the light source intersects the surface. The intersection between this envelope and the light source forms a curve on the light source analogous to the attached shadow boundary on the surface. We refer to this curve in the text as the light source's image of the attached shadow boundary.

penumbra; that is, they demarcate regions on a surface that receive at least some illumination from a light source from regions which do not receive any illumination from the source. For an attached shadow that casts a shadow on its surface, there exists for each point on the attached shadow boundary a corresponding point on the cast shadow boundary that lies along the same ray from the light source. We will write the collection of points on the cast shadow boundary, therefore, as a curve, $\gamma : \gamma(t);\ 0 \leqslant t < 1$, defined so that $\gamma(t)$ lies along the same light ray as $\alpha(t)$.

B. Geometry of Attached Shadow Boundaries for Extended Light Sources

In this subsection we use the geometrical relations between attached shadow boundaries and their images on an extended light source to derive the fundamental geometric properties of attached shadow boundaries. We will show that they behave qualitatively like attached shadow boundaries for point sources. In particular, we will show that such boundaries are everywhere smooth (their tangents are everywhere well defined) and are everywhere conjugate to the direction of illumination; that is, the derivative of the surface normal taken along the tangent to the shadow boundary at each point is perpendicular to the illumination ray at that point. Although the latter property may seem somewhat abstract, it underlies the constraints relating shadow boundary shape and surface shape. As an application of the result, we show that the behavior of intrinsic shadow boundaries at intersections of parent–child attached and cast shadow boundaries is the same for extended sources as it is for point sources. We state the result more formally as a proposition.

Proposition 1. Let Σ be a smooth surface illuminated by a convex light source, Λ. Assume that Σ contains no planar points. Let $\alpha : \alpha(t);\ 0 \leqslant t < 1$ be an attached shadow boundary on Σ that matches condition (4). Assuming general position of the light source, α is a smooth curve; that is, $\alpha'(t) \neq 0;\ 0 \leqslant t < 1$. Furthermore, the tangent of α is everywhere conjugate to the direction of illumination.

Proof. We will first show that if $\alpha'(t) \neq 0$, then $\alpha'(t)$ is conjugate to the direction of illumination. We want to show that for nonzero $\alpha'(t)$

$$\langle N'_\Sigma[\alpha'(t)],\ \mathbf{L}(t)\rangle = 0, \tag{5}$$

where $N'_\Sigma[\alpha'(t)]$ is the derivative of the surface normal computed in the direction $\alpha'(t)$ on the surface.

The direction of illumination at $\alpha(t)$ is given by $\mathbf{L}(t) = [\alpha(t) - \lambda(t)]/\|\alpha(t) - \lambda(t)\|$, where $\lambda(t)$ is the illuminant's image of $\alpha(t)$, as defined above. By condition (4), we have

$$\langle N_\Sigma[\alpha(t)],\ \mathbf{L}(t)\rangle = \langle N_\Sigma[\alpha(t)],\ \alpha(t) - \lambda(t)\rangle = 0. \tag{6}$$

Taking derivatives, we obtain

$$\frac{\partial}{\partial t}\langle N_\Sigma[\alpha(t)],\ \alpha(t) - \lambda(t)\rangle = \langle N'_\Sigma[\alpha'(t)],\ \alpha(t) - \lambda(t)\rangle + \langle N_\Sigma[\alpha(t)],\ \mathbf{L}'(t)\rangle = 0 \tag{7}$$

$$\langle N'_\Sigma[\alpha'(t)],\ \alpha(t) - \lambda(t)\rangle + \langle N_\Sigma[\alpha,\ (t)],\ \alpha'(t) - \lambda'(t)\rangle = 0 \tag{8}$$

$$\langle N'_\Sigma[\alpha'(t)], \alpha(t) - \lambda(t)\rangle + \langle N_\Sigma[\alpha(t)], \alpha'(t)\rangle - \langle N_\Sigma[\alpha(t)],\ \lambda'(t)\rangle = 0. \tag{9}$$

Noting that $N_\Sigma[\alpha(t)] = N_\Lambda[\lambda(t)]$ and that $\langle N_\Lambda[\lambda(t)],\ \lambda'(t)\rangle = 0$ (which implies that $\langle N_\Sigma[\alpha(t)],\ \lambda'(t)\rangle = 0$), we obtain finally

$$\langle N'_\Sigma[\alpha'(t)],\ \mathbf{L}(t)\rangle = 0, \tag{10}$$

the desired result.

In order to prove the first part of the proposition, that α is everywhere smooth, we need to show that at each point $\alpha(t)$, a unique, nonzero vector v exists in the tangent plane of Σ at $\alpha(t)$ such that (a) $\langle N'_\Sigma(v),\ \mathbf{L}(t)\rangle = 0$, and (b) $N'_\Sigma(v) = N'_\Lambda(w)$ for some vector, w (possibly 0), in the tangent plane of Λ at $\lambda(t)$. We then have $\alpha'(t) = v$ and $\lambda'(t) = w$.

Condition (a) is immediate under an assumption of general light-source position, since for nonplanar surface points the conjugate direction of any given vector is unique. This determines the direction of v. It merely remains to show that a vector, w, may be found such that condition (b) holds. We consider two cases.

Case 1: a nonparabolic point of Σ. Choose the vector $\mathbf{v}$ that matches condition (a) and has unit length. $\mathbf{N}'_\Sigma(\mathbf{v})$ is a vector in the tangent plane of $\alpha(t)$. Since Λ is assumed to be simply convex [the point $\lambda(t)$ is elliptic], a vector $\mathbf{w}$ exists for which $\mathbf{N}'_\Lambda(\mathbf{w})$ is equal to any other vector in the tangent plane of the surface at $\lambda(t)$. Since $\alpha(t)$ and $\lambda(t)$ share the same tangent plane, $\mathbf{N}'_\Sigma(\mathbf{v})$ is in the tangent plane of Λ at $\lambda(t)$, and it is clearly possible to define a vector $\mathbf{w}$ for which $\mathbf{N}'_\Lambda(\mathbf{w}) = \mathbf{N}'_\Sigma(\mathbf{v})$.

Case 2: a parabolic point of Σ. Assuming a general position for the lighting, the vector $\mathbf{v}$ matching condition (a) is in the asymptotic direction of the surface and is perpendicular to the parabolic line.[21] Thus we have $N'_\Sigma(v) = 0$. Since the light source is assumed to be elliptic at every point, no nonzero vector $\mathbf{w}$ exists in the tangent plane of Λ for which $N'_\Lambda(w) = 0$; therefore, we must choose $\mathbf{w} = 0$. With this choice, condition (b) is met and $\alpha'(t)$ is in an asymptotic direction of the surface. Since $\mathbf{w} = 0$, we have $\lambda'(t) = 0$, and the image of an attached shadow boundary on a convex light source has a cusp singularity where the boundary point is parabolic on the surface.

Since cases 1 and 2 include all points on smooth surfaces, α is everywhere smooth, proving part 1 of the proposition. Q.E.D.■

One of the important invariants we derived for point light sources that relates the behavior of intrinsic shadows and surface shape is that parent–child intersections between attached and cast shadow boundaries occur where the tangent to the boundary is in an asymptotic direction of a surface (point b in Fig. 5). This type of intersection is distinguished from other attached–cast shadow intersections by the fact that the visible intrinsic boundary is smooth at the intersection (up to first order), where other types of intersection form corners. We can use the result obtained above to show that these invariants hold for intrinsic shadows formed by extended light sources.

A parent–child intersection between attached and cast shadow boundaries occurs where an attached shadow boundary goes from being potentially visible to the light source to being invisible to the light source, since at this point the attached shadow boundary curves interior to the apparent shadow. The potential visibility of an attached shadow boundary to the light source is determined by the curvature of the surface in the direction of the illuminating ray: if the surface is convex in that direction, the boundary is potentially visible; if it is concave, it is invisible. Parent–child intersections between attached and cast shadow boundaries, therefore, occur at points of attached shadows at which the normal curvature of the surface in the direction of the illuminating ray changes from positive to negative. We show that at these points, the attached shadow boundary is in an asymptotic direction of the surface.

Proposition 2. Let Σ be a smooth surface illuminated by an extended light source, Λ. Assume that Σ contains no planar points. Let $\alpha : \alpha(t);\ 0 \leqslant t < 1$ be an attached shadow contour on Σ that matches condition (4), and let $\mathbf{L}(t)$ be a unit vector in the direction of the illuminating ray at α. Assuming general position for the light source, a necessary condition for the surface curvature in the direction of the illuminating ray $\{\kappa_n[\mathbf{L}(t)]\}$ to change sign at a point $\alpha(t)$ is that $\alpha'(t)$ be an asymptotic direction of Σ.

Proof. A necessary condition for the point at which $\kappa_n[\mathbf{L}(t)]$ changes sign is that $\kappa_n[\mathbf{L}(t)] = 0$; that is, $\mathbf{L}(t)$ must be an asymptotic direction of Σ. But where $\mathbf{L}(t)$ is an asymptotic direction of Σ, $\alpha'(t)$ is in the same direction as $\mathbf{L}(t)$, since $\alpha'(t)$ is everywhere conjugate to $\mathbf{L}(t)$ and an asymptotic direction is self-conjugate. The condition that $\mathbf{L}(t)$ be an asymptotic direction of Σ is therefore equivalent to the condition that $\alpha'(t)$ be an asymptotic direction. Q.E.D.■

C. Summary of Results

For simply convex light sources, the major results found for point sources generalize to extended sources. For light sources that are not simply convex, the situation is not nearly so happy. We have noted above that the image of an attached shadow boundary on an extended light source may have cusp singularities, precisely at points whose corresponding attached shadow boundary point is a parabolic point on a surface. This may seem a pedantic point until we consider the symmetry of the definitions of attached shadow boundaries and their images on a light source (see Fig. 7). The symmetry suggests that if a light source were hyperbolic, cusps could appear in attached shadow boundaries of an illuminated surface. The analysis of intrinsic shadow geometry for light sources that are not simply convex requires looking in more detail at the behavior of the illumination surfaces for such light sources. Because we feel that nonconvex light sources are the exception rather than the rule, we will not present this type of analysis here, though we will point out one result that is interesting; namely, that such cusps can create nonsmooth (i.e., corner) parent–child intersections between attached and cast shadow boundaries (effectively created by cusps in the attached shadow boundary). One lesson from this is that one has to be careful about generalizing results from constrained domains such as point-source illuminants to more general domains without first testing them by rigorous analysis.

5. SHADOWS AT SURFACE CREASES

Surface creases are edges on a surface formed by discontinuous changes in surface orientation. Attached shadows may intersect surface creases or may be formed by surface creases. Cast shadows may also intersect surface creases. In this section we will analyze the geometric behavior of intrinsic surface shadows as they intersect or are formed by surface creases. As it turns out, significant information about surface geometry and light source direction can be gleaned from the geometry of intrinsic shadows at surface creases. It can determine whether a crease is convex or concave, what the tilt of the light source is, and whether an edge in the image is, in fact, a crease between connected regions on a surface or is the occluding contour of one object placed in front of another; that is, it can be used to infer contact between surfaces.

Three significant shadow events can occur at a surface crease:

- Intersection between an attached shadow and a surface crease [Fig. 8(a)].
- Formation of an attached shadow along a crease [Fig. 8(b)].
- Intersection between a cast shadow and a surface crease [Fig. 8(c)].

In the following sections we will analyze the qualitative structure of shadow contours at such events and charac-

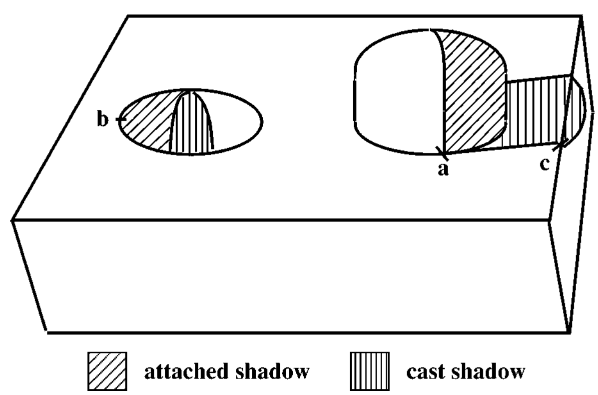


Fig. 8. Three categorically different shadow events at surface creases. a, An attached shadow boundary may intersect a crease. If the crease is concave *and* the surface apposite to the attached shadow faces the light source (as in the figure), the side of the crease with the attached shadow boundary will cast a shadow on the other side. The resulting cast shadow boundary will intersect the surface crease at the same point as the attached shadow boundary and will be cotangent to the crease (see text). The crease may, of course, be convex. b, The crease may itself form an attached shadow boundary. c, A cast shadow boundary may intersect a surface crease, which will generically form a tangent discontinuity in the cast shadow boundary.

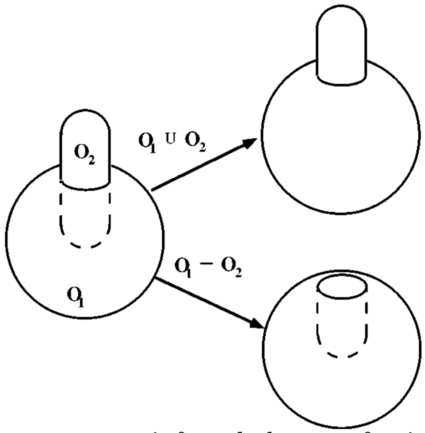


Fig. 9. A concave crease is formed when a surface is created by the solid union of two objects. A convex crease is formed when one solid object is subtracted from another.

terize the information contained therein about both the geometry of the surface crease and the light-source direction.

A. Definitions

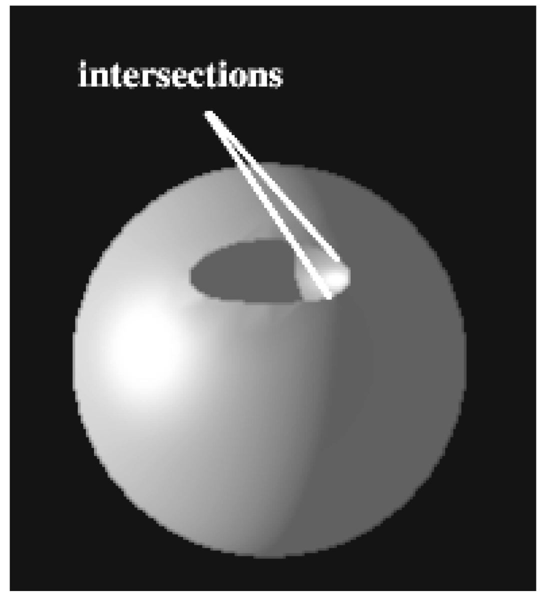
We will formally model surface creases as having resulted from the solid union or difference of two objects,[22] as demonstrated in Fig. 9. In either case, the local shape of a crease at a point of intersection between two surfaces is determined by the surface normals of the two surfaces. In the case of solid union, the crease is concave, and the surface normals of the creased surface at either side of the crease are simply the surface normals of the two intersecting surfaces. In the case of solid difference, the crease is convex. The surface normal on one side of the crease is equal to the surface normal of the subtracted-from surface, while the surface normal on the other side is the negative of the surface normal of the subtracted surface.

According to the synthetic model just described, surface creases are formed at the curves of intersection between two surfaces, $\eta = \Sigma_1 \cap \Sigma_2$. We will assume η to be parameterized as $\eta : \eta(t);\ 0 \leqslant t < 1$. The surface shape along η is given by pair of surface normals, $\mathbf{N}_\Sigma^+(t)$ and $\mathbf{N}_\Sigma^-(t)$, immediately to either side of the crease, and the unit tangent of η is defined as $\mathbf{t}_c = (\mathbf{N}_\Sigma^+ \wedge \mathbf{N}_\Sigma^-)/|\mathbf{N}_\Sigma^+ \wedge \mathbf{N}_\Sigma^-|$.

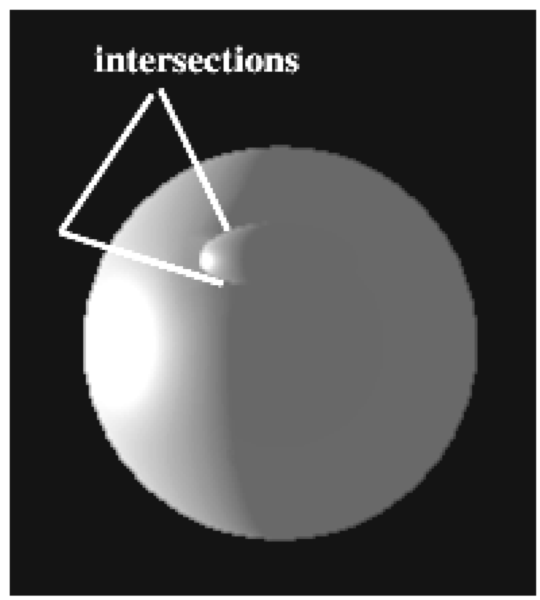
B. Intersections between Attached Shadows and Surface Creases

Figure 10 shows two examples of situations in which an attached shadow boundary intersects a surface crease. In general the attached shadow boundary will not extend continuously on either side of a surface crease. Rather, the surface on the side of the crease apposite to the attached shadow boundary will be either illuminated or in shadow. Intuitively, one would think that knowledge of whether the surface on the side of a crease apposite to an intersecting attached shadow boundary is illuminated or in shadow could determine the direction in which the surface bends at the crease, that is, whether the crease is concave or convex. This turns out to be true. We will describe here the invariant structure that can be used to make such a determination.

Figure 11 illustrates the local information available at a point of intersection between an attached shadow contour and a crease contour. This includes the direction of the crease contour, the light-source direction, and a label specifying whether the surface apposite to the attached shadow contour is in shadow or is illuminated (the direction of the attached shadow boundary is largely irrelevant). Depending on what the orientation of the surface crease is relative to the light-source tilt, illumination of the apposite surface may reflect either convexity or con-



(a)



(b)

Fig. 10. Two examples of attached shadows intersecting a surface crease. (a) The side of the surface apposite to the attached shadow boundaries is illuminated, and the crease is convex. (b) The side of the surface apposite to the attached shadow boundaries is illuminated, and the crease is concave. The inference of crease convexity/concavity from shading across the crease depends critically on the direction of the surface crease relative to the light-source direction.

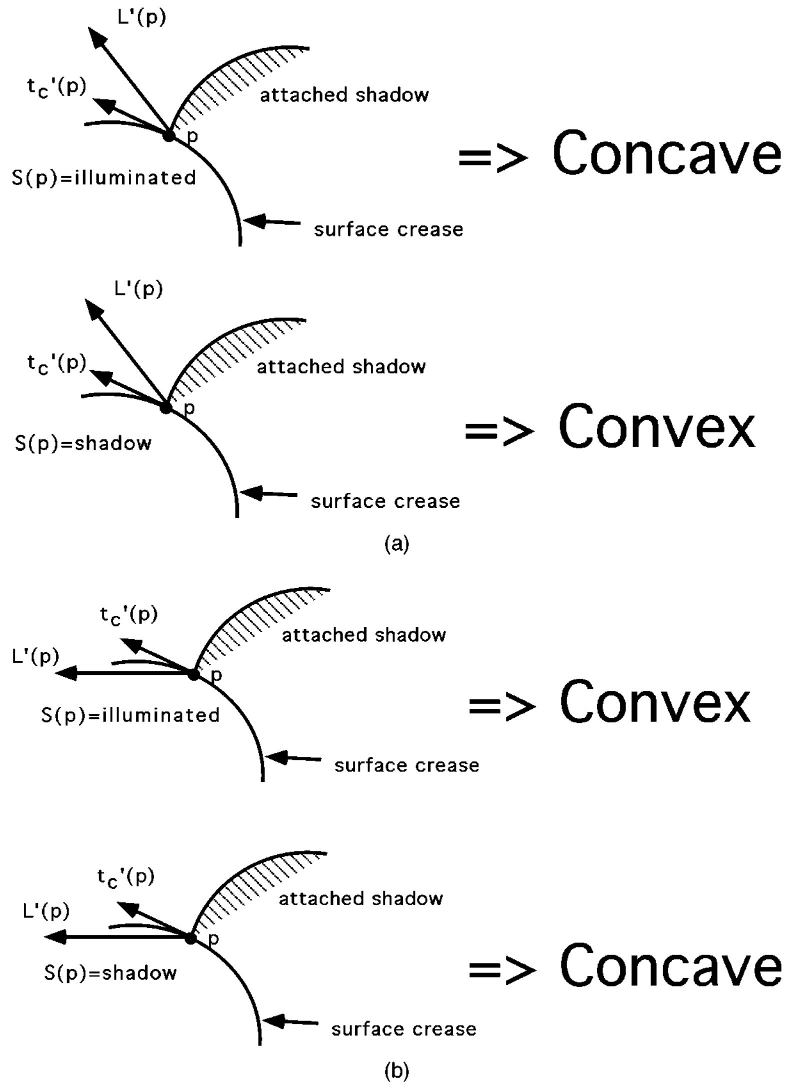


Fig. 11. Information available at an intersection between an attached shadow and a surface crease (point **p**): $\mathbf{t}_c'(p)$, the tangent direction, in the image, of the surface crease; $\mathbf{L}'(p)$, the light-source direction, as projected into the image (i.e., the light-source tilt); and $\mathbf{S}(p)$, a binary labeling of the shading on the side of the crease apposite to the attached shadow boundary—that is, indicating whether it is illuminated or in shadow. The interpretation rule described in the text is illustrated here. (a) The tangent to the surface crease is on the side of the light-source vector opposite to the attached shadow. (b) The tangent to the surface crease is on the same side of the light source as the attached shadow.

cavity. If the tangent to the crease, oriented in the same direction as the light source, is on the side of the light-source vector "opposite" to the attached shadow boundary, then the following inference holds: apposite surface illuminated⇒concave crease; apposite surface in shadow⇒convex crease. If the tangent to the surface crease is on the same side of the light-source vector as the attached shadow, then the opposite inference holds; that is, apposite surface illuminated⇒convex crease; apposite surface in shadow⇒concave crease.

We will only sketch the proof of the interpretation rule here. To do this, we consider the relationships between crease direction and surface shading in the tangent plane of the surface. The surface normal on the attached shadow side of the point of intersection is perpendicular to the light-source direction, since it is on the attached shadow boundary; therefore the tangent plane of the surface on this side of the crease contains the local illuminating ray. By definition, it also contains the tangent of the crease at the point of intersection. We can treat the surface discontinuity at the crease as being formed (locally) by folding the tangent plane along the tangent of the surface crease. If it is folded up, the crease is concave; if it is folded down, the crease is convex. It is easy to show that if the crease tangent is parallel to the light-source direction, the attached shadow boundary will extend continuously (with a tangent discontinuity) across the crease regardless of how the surface is folded at the crease; that is, the surface immediately adjacent to both sides of the crease will be on attached shadow boundaries. This requires nongeneric positioning of the surface, however, and in general the surface apposite to the attached-shadow boundary will either be illuminated or in shadow, as stated above.

For convenience, we can define a coordinate system in the tangent plane of the surface (on the attached shadow side of the crease) in which the light source direction is taken to be vertical. Similarly, we can define the orientation of the crease tangent so that it points in a positive y direction in this coordinate system (i.e., it points toward the light source). With these definitions, consider the case that the crease tangent points to the side of the vertical that contains the attached shadow boundary [Fig. 11(b)]. In this case, it is easy to show that folding the surface up on the side of the crease apposite to the attached shadow boundary will put that side of the surface in shadow, whereas folding it down will expose that side of the surface to the light source. If the crease tangent points toward the opposite side of the vertical, the opposite relations hold [Fig. 11(a)]. Since the qualitative relations between light-source direction and the crease tangent are maintained under perspective projection, the same rules apply in the image, taking the light-source tilt (the projected direction of illumination) as the vertical. This results in the interpretation rule illustrated in Fig. 11.

If the crease is concave and the surface apposite to the attached shadow boundary is illuminated [as in Fig. 8(a)], the surface on the attached shadow boundary side of the crease will, in general, cast a shadow on the other side of the crease. The cast shadow boundary intersects the surface crease at the same point as the attached shadow and is cotangent to the surface crease at the point of intersection (see Appendix C for a proof). The cotangency of surface crease and cast shadow boundary is a useful property for contour labeling, a point we elaborate on in Subsection 6.C below.

C. Formation of Attached Shadows at Surface Creases

Surface creases, when convex, can themselves form attached shadow boundaries when the surface on one side of the crease is illuminated and the surface on the other side is in shadow. Holes in surfaces are classical examples of such events. Attached shadows of this sort provide no information about surface geometry that is not provided by the shape of the crease contour itself, although a determination that an attached shadow boundary is also a surface crease does imply that the surface crease is convex.

D. Intersections between Cast Shadows and Surface Creases

A final type of intersection between shadow contours and surface creases is formed where a cast shadow boundary intersects an "unrelated" surface crease, as in Fig. 8(c). Shafer and Kanade[5] have shown that the behavior of cast shadow contours at such intersections provides strong quantitative constraints on surface interpretation when the casting shadow boundary is a straight surface crease and is visible in the image. These constraints are, however, by no means enough to uniquely determine surface shape at a crease. In this section we report two related qualititative constraints on scene interpretation provided by the shapes of cast shadow boundaries at intersections with surface creases. These constraints do not require any information about the shape of the casting edge and hence may be applied locally.

Figure 12 shows an example of a shadow cast over a surface crease. The shape of the cast shadow contour provides information about both the light-source direction and the shape of the underlying surface. This results from what we will refer to as a uniqueness constraint on the correspondence between attached shadow boundaries and cast shadow boundaries. Each point on an attached shadow boundary may project to one and only one point on a cast shadow boundary along the illuminating ray. Similarly, only one point on a cast shadow contour may project along the image of an illuminating ray. In the figure, $\mathbf{L}^*$ is not a physically possible light source, regardless of crease convexity/concavity.

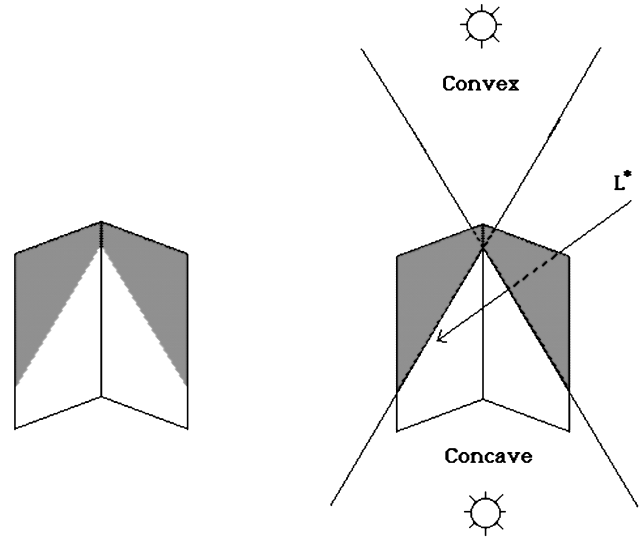


Fig. 12. The uniqueness constraint limits the range of plausible light-source directions for a cast shadow making a particular angle in the image at an intersection with a surface crease. The physically realizable light sources lie in quadrants of the image demarcated by the tangent directions of the cast shadow contour immediately to either side of a crease, as shown.

A qualitative constraint on surface shape interpretation follows naturally from the light-source constraint. If the image of the light source lies in the concave side of a cast shadow contour at a crease, the crease should be interpreted as being concave; otherwise, it should be interpreted as being convex. This follows from the facts that the image of the casting edge must be on the same side of the contour as the light source, and, as Shafer and Kanade[5] showed, the convexity/concavity of a crease can be determined by whether the shadow contour at the crease bends away from the casting edge or toward it. Note that the image of the casting edge is not needed to make the inference, only whether the illuminant is from one side of the contour or the other. Thus in Fig. 12 an assumption that the light source is from above should disambiguate the qualitative shape of the crease to be convex.

6. IMPLICATIONS

In this paper we have exhaustively analyzed the qualitative geometric structure of shadow contours, for both smooth surfaces and piecewise smooth surfaces (surfaces with creases) and for both point light sources and extended light sources. Along the way, we have highlighted features of the geometry that can provide useful information about scene structure. Since these observations were scattered through the different sections of the paper, we will summarize them here in a concise form. We have not attempted to derive techniques for inferring surface or light-source geometry from shadows for the simple reason that the information provided by shadows only loosely constrains such inferences. Using only shadow information to interpret surface geometry, for example, would require imposing strong prior constraints on surfaces. Shadows do, however, provide salient information that would be useful in conjunction with other cues, and it is in this spirit that we present our summary of results.

A. Light-Source Direction

Light sources located a large distance away from a surface may be approximated as point sources at infinity and characterized by the global slant and tilt of the rays illuminating the surface. Intrinsic shadows provide two local sources of information that directly determine the tilt of surface illumination: the tangent direction of attached shadows at intersections with smooth occluding contours (equivalently, the tangent directions of occluding contours at such intersections) and the tangent direction of attached shadow contours at intersections with their corresponding cast shadows. Moreover, global information is provided by the implicit lines connecting L junctions between cast and attached shadow contours and corresponding L junctions between cast shadow contours (see Fig. 6). Taken together, these cues overdetermine the illuminant tilt, itself an important parameter in the estimation of shape from shading.

B. Surface Shape

The cues provided about surface shape by intrinsic shadows may be broken into two classes: smooth surface shape and surface shape at creases. The local information provided by intrinsic shadows about smooth surface shape is sparse. The local orientation of an attached shadow contour provides some constraint on the curvature structure of surfaces (see Appendix A), but given the multidimensional character of local curvature, the constraint is weak. Somewhat counterintuitively, the local curvature of attached shadow contours, except at intersections with occluding contours, depends not only on local surface curvature but also on the third-order structure of local surface shape. Unlike smooth occluding contours, inflection points in attached shadow contours do not have any useful invariant relationship to local surface shape. The only such invariant relationship that exists is that at points of intersection between related attached and cast shadow contours, the tangent direction of the contours is an asymptotic direction of a surface. A more promising application of shadow information for smooth-shape interpretation may be to the global structure of surfaces; that is, as a source of information about the presence of hills, dimples, etc., on surfaces. In this context, qualitative shadow information could well be incorporated into qualitative shape-reasoning systems proposed for occluding contours.[23] An example of such an approach would be to extend the notion of an aspect graph[24] to include the topological structure of shadow contours.

Unlike for regular (smooth) points on a surface, intrinsic shadow contours, both attached and cast, provide strong constraints on qualitative surface geometry at surface creases. The local structure of intersections between attached shadow contours and surface creases, in conjunction with the qualitative shading at the point of intersection, reliably determines the convexity and concavity of a crease [Subsection 5.B]. The shape of cast shadow contours at intersections with surface creases also determines crease convexity/concavity [Subsection 5.D].

C. Contour Labeling

Shadow contours provide salient cues to contour labeling and attachment (to which side of an image a contour belongs). In particular, the geometry of shadow contours at intersections with other contours can provide a useful cue to the nature of the intersecting contour. If an attached shadow contour is cotangent to another contour at a point of intersection, one can reliably infer that the contour is a smooth occluding contour and that the contour is attached to the side with the attached shadow contour (this derives from previous analyses of extrinsic and intrinsic surface markings[16]). The constraints on shadow geometry at surface creases also allow one to draw inferences about contour labeling and attachment. When a surface to one side of a concave crease casts a shadow on the other side (see Fig. 8), the cast shadow contour intersects the crease at the same point as its parent attached shadow contour and is cotangent to the crease at the point of intersection. Violations of either of these constraints for cast shadow contours provide information that the intersecting contour is not, in fact, a surface crease but rather a contour occluding the surface containing the cast shadow. In a naturalistic setting, such information can be used to determine the contact between objects (e.g., between a cylinder and a tabletop).

APPENDIX A

We derive here the quantitative relationship between the shape of an attached shadow contour and the shape of the underlying surface, assuming orthographic projection into the image plane. Because the analysis depends only on the fact that an attached shadow boundary is conjugate to the illumination direction, it holds for all three light sources considered in the text: infinite point sources, point sources at a finite distance from a surface, and extended light sources. The resulting relationships, however, increase greatly in complexity for extended sources, since the variables describing light source direction at a point are local.

For simplicity, we will consider a point source at infinity for the analysis and comment on how it generalizes to the other types of light sources. Two cases are of interest to us: the local behavior of an attached shadow contour at a regular point and the local behavior of the contour at an intersection with a self-occluding contour. For both cases we assume the surface to be locally parameterized $[\Sigma : \mathbf{X} = \mathbf{X}(u, v)]$ so that $\mathbf{X}_u = \mathbf{e}_1 = \mathbf{L} \wedge \mathbf{N}$ and $\mathbf{X}_v = \mathbf{e}_2$

$= \mathbf{L}$, where $\mathbf{L}$ is a unit vector in the direction of the light source and $\mathbf{N}$ is the unit normal vector of the surface at the point of interest. The unit vector that is conjugate to $\mathbf{L}$, as expressed in terms of the coordinate system $\{\mathbf{e}_1, \mathbf{e}_2\}$, is given by

$$\mathbf{t} = \left[\frac{g}{(f^2 + g^2)^{1/2}}, \frac{-f}{(f^2 + g^2)^{1/2}}\right]^T, \tag{A1}$$

where f and g are the last two coefficients of the second fundamental form of the surface.[25,26] Since the tangent vector of the attached shadow boundary is conjugate to $\mathbf{L}$, we may consider $\mathbf{t}$ to be the unit tangent to the attached shadow boundary.

The tangent vector $\mathbf{t}$ of the attached shadow boundary may be related to the tangent vector of its image (the tangent of the attached shadow contour) $\tilde{\mathbf{t}}$ by

$$\mathbf{t} = \frac{(\tilde{\mathbf{t}} \wedge \mathbf{V}) \wedge \mathbf{N}}{|(\tilde{\mathbf{t}} \wedge \mathbf{V}) \wedge \mathbf{N}|}, \tag{A2}$$

where $\mathbf{V}$ is the unit normal vector of the image plane directed toward the surface (the viewing direction). Using the relation

$$\langle \mathbf{t}, \mathbf{e}_1 \rangle = \langle \mathbf{t}, \mathbf{L} \wedge \mathbf{N} \rangle = \frac{g}{(f^2 + g^2)^{1/2}}, \tag{A3}$$

we have

$$\frac{1}{|(\tilde{\mathbf{t}} \wedge \mathbf{V}) \wedge \mathbf{N}|} \langle (\tilde{\mathbf{t}} \wedge \mathbf{V}) \wedge \mathbf{N}, \mathbf{L} \wedge \mathbf{N} \rangle = \frac{g}{(f^2 + g^2)^{1/2}}, \tag{A4}$$

$$\frac{1}{|(\tilde{\mathbf{t}} \wedge \mathbf{V}) \wedge \mathbf{N}|} (\langle \tilde{\mathbf{t}} \wedge \mathbf{V}, \mathbf{L} \rangle - \langle \tilde{\mathbf{t}} \wedge \mathbf{V}, \mathbf{N} \rangle \langle \mathbf{N}, \mathbf{L} \rangle) = \frac{g}{(f^2 + g^2)^{1/2}}, \tag{A5}$$

and, since $\langle \mathbf{N}, \mathbf{L} \rangle = 0$, we have, finally,

$$\frac{1}{|(\tilde{\mathbf{t}} \wedge \mathbf{V}) \wedge \mathbf{N}|} \langle \tilde{\mathbf{t}} \wedge \mathbf{V}, \mathbf{L} \rangle = \frac{g}{(f^2 + g^2)^{1/2}}. \tag{A6}$$

We see that at regular points of an attached shadow contour, its orientation, as expressed by $\tilde{\mathbf{t}} \wedge \mathbf{V}$, is directly related to the local curvature structure of a surface. The curvature of the contour is related to the third-order differential structure of the surface.

At a point at which an attached shadow contour intersects a self-occluding contour, the previous analysis does not hold, as the normal of the surface is orthogonal to the line of sight. We note that the curvature vector of a curve may be decomposed into orthogonal components; one in the direction of the surface normal and one in the tangent plane of the surface. In the case that the surface normal is orthogonal to the line of sight, only the component of a curve's curvature in the direction of the surface normal affects the curvature of the contour to which the curve projects. Since the length of the curvature component in the normal direction of the surface is the normal curvature of the surface, we can relate the curvature of the contour, κ_s, to the normal curvature of the surface. This is given by

$$\kappa_s = \frac{1}{\sin^2 \phi} \kappa_n(\mathbf{t}), \tag{A7}$$

where ϕ is the angle made by the tangent vector of the curve and the viewing direction and $\kappa_n(\mathbf{t})$ is the normal curvature of the surface in the tangent direction, $\mathbf{t}$. In the coordinate system defined above, we have for the normal curvature in the tangent direction of an attached shadow boundary (based on the fact that it is conjugate to the lighting direction)

$$\kappa_n(\mathbf{t}) = \frac{g(eg - f^2)}{f^2 + g^2}, \tag{A8}$$

where e, f, and g are the three coefficients of the second fundamental form of the surface. We therefore have

$$\kappa_s = \frac{1}{\sin^2 \phi} \left[\frac{g(eg - f^2)}{f^2 + g^2}\right]. \tag{A9}$$

It remains to express $\sin^2 \phi$ in terms of e, f, and g and the direction of illumination. Letting $\phi = \theta - \sigma$, where θ is the angle made by the tangent of the attached shadow boundary and the illuminant direction, and σ is the angle made by the viewing direction and the illuminant direction, we have

$$\kappa_s = \frac{1}{(\sin \theta \cos \sigma - \cos \theta \sin \sigma)^2} \left[\frac{g(eg - f^2)}{f^2 + g^2}\right]. \tag{A10}$$

Noting that the tangent direction of the attached shadow boundary is $(g/\sqrt{f^2 + g^2}, -f/\sqrt{f^2 + g^2})^T$ in the coordinate system $\{(\mathbf{L} \wedge \mathbf{N}), \mathbf{L}\}$, we have

$$\kappa_s = \frac{g(eg - f^2)}{(f \sin \sigma + g \cos \sigma)^2}. \tag{A11}$$

For a sphere, we have $e = g = 1/R$, where R is the radius of the sphere and $f = 0$, so that

$$\kappa_s = \frac{1}{R \cos^2 \sigma}, \tag{A12}$$

and since for a sphere the curvature of the occluding contour, κ_c, is given by $\kappa_c = 1/R$, we have

$$\cos^2 \sigma = \frac{\kappa_c}{\kappa_s}, \tag{A13}$$

so that in this special case, the light-source slant can be determined by the ratio of curvatures of a self-occluding contour and an attached shadow.

The above analysis generalizes to point sources a finite distance from the illuminated surface and to extended light sources. The only difference is that the orientation of the local coordinate system used to define e, f, and g varies more in the latter two cases as an attached shadow contour is traversed, since $\mathbf{L}$ changes. Moreover, the illuminant slant becomes a local variable defining the slant of the local illuminating ray.

APPENDIX B

In Section 2 we claimed that a shadow boundary is smooth up to its first derivative at a junction where it changes from being attached to cast. We offer a simple proof of this here. Recall the definition of a cast shadow boundary given in Section 2: that as the curve γ on the surface Σ formed by the intersection of the sheaf of light rays passing through α (the illumination surface) and Σ at points away from α. More formally, we chose the particular parameterization of γ that associates each point on γ with the point on α that lies on the same light ray; thus we have as the defining condition for γ:

$$\frac{\gamma(t)-\alpha(t)}{|\gamma(t)-\alpha(t)|}=\mathbf{L}(t);\quad \gamma(t)\neq\alpha(t),$$

$$\gamma(t)\in\Sigma,\quad \kappa_n[\alpha'(t)]>0, \tag{B1}$$

where $\mathbf{L}(t)$ is a unit vector in the direction of the light ray illuminating $\alpha(t)$. The last condition captures the fact that only potentially visible parts of an attached shadow (where the surface is convex in the direction of the illuminating ray) cast shadows on the surface.

Since γ is defined only for potentially visible portions of α, we are interested in the limiting behavior of $\gamma'(t)$ as we approach the point, $t=\tau$, at which γ joins with α (at this point α changes from being potentially visible to being invisible, hence γ is undefined at τ). We will assume that α is oriented so that t is increasing as $\alpha(\tau)$ is approached from the potentially visible part of α; thus we are interested in the one-sided limit, $\lim_{t\to\tau^-}\gamma'(t)$. The tangent vector of α at $t=\tau$, $\alpha'(\tau)$, is in the direction of the light ray illuminating the point (see Subsections 2.C and 3.B); therefore we can express the proposition in the relation

$$\lim_{t\to\tau^-}\frac{\gamma'(t)}{|\gamma'(t)|}=\mathbf{L}(\tau). \tag{B2}$$

We will prove this assertion for the three cases considered in the text: a point source at infinity, a point source at a finite distance from the surface, and a convex extended source.

Case 1: point source at infinity.

Differentiating both sides of Eq. (B1), we obtain

$$\frac{\gamma'(t)-\alpha'(t)}{|\gamma(t)-\alpha(t)|}-\frac{[\gamma(t)-\alpha(t)]\langle\gamma'(t)-\alpha'(t),\gamma(t)-\alpha(t)\rangle}{|\gamma(t)-\alpha(t)|^3}=0, \tag{B3}$$

since $\mathbf{L}(t)=\mathbf{L}$ is a constant. Substituting from Eq. (B1) and simplifying gives

$$\gamma'(t)=\alpha'(t)+\mathbf{L}[\langle\gamma'(t),\mathbf{L}\rangle-\langle\alpha'(t),\mathbf{L}\rangle]. \tag{B4}$$

Normalizing and taking the limit as $t\to\tau^-$, we obtain for the unit tangent vector of γ

$$\lim_{t\to\tau^-}\frac{\gamma'(t)}{|\gamma'(t)|}=\lim_{t\to\tau^-}\frac{\alpha'(t)+\mathbf{L}(\langle\gamma'(t),\mathbf{L}\rangle-\langle\alpha'(t),\mathbf{L}\rangle)}{|\alpha'(t)+\mathbf{L}(\langle\gamma'(t),\mathbf{L}\rangle-\langle\alpha'(t),\mathbf{L}\rangle)|}, \tag{B5}$$

and since $\alpha'(\tau)/|\alpha'(\tau)|=\mathbf{L}$, we obtain

$$\lim_{t\to\tau^-}\frac{\gamma'(t)}{|\gamma'(t)|}=\frac{\mathbf{L}\langle\lim_{t\to\tau^-}\gamma'(t),\mathbf{L}\rangle}{\langle\lim_{t\to\tau^-}\gamma'(t),\mathbf{L}\rangle}=\mathbf{L}. \tag{B6}$$

This completes the proof for a point source at infinity.

Case 2: point source at a finite distance from the surface.

The situation differs from the previous one in that $\mathbf{L}(t)$ is not a constant but rather is given by $\mathbf{L}(t)=(\alpha(t)-\lambda)/|\alpha(t)-\lambda|$, where λ is the position of the light source in space. Since the left-hand side of Eq. (B1) does not involve $\mathbf{L}(t)$, we need merely show that the right-hand side, when differentiated, is equal to 0 at $t=\tau$, as it was for a point source at infinity. We have

$$\begin{aligned}\mathbf{L}'(t)&=\frac{\alpha'(t)}{|\alpha(t)-\lambda|}-\frac{(\alpha(t)-\lambda)\langle\alpha'(t),\alpha(t)-\lambda\rangle}{|\alpha(t)-\lambda|^3}\\&=\frac{\alpha'(t)-\mathbf{L}(t)\langle\alpha'(t),\mathbf{L}(t)\rangle}{|\alpha(t)-\lambda|}\end{aligned} \tag{B7}$$

Setting $t=\tau$ and noting that $\alpha'(\tau)/|\alpha'(\tau)|=\mathbf{L}(\tau)$, we obtain

$$\mathbf{L}'(\tau)=\frac{\alpha'(\tau)-\alpha'(\tau)\langle\alpha'(\tau),\alpha'(\tau)\rangle/|\alpha'(\tau)^2}{|\alpha(\tau)-\lambda|}=0, \tag{B8}$$

completing the proof.

Case 3: convex, extended source.

The situation is similar to case 2, differing only in that $\mathbf{L}(t)$ is given by $\mathbf{L}(t)=[\alpha(t)-\lambda(t)]/|\alpha(t)-\lambda(t)|$, where $\lambda(t)$ is not a constant but rather is the illuminant's image of $\alpha(t)$, as defined in Section 3. We will find it convenient to express the cast shadow boundary constraint in a somewhat different form, namely,

$$\frac{\gamma(t)-\lambda(t)}{|\gamma(t)-\lambda(t)|}=\frac{\alpha(t)-\lambda(t)}{|\alpha(t)-\lambda(t)|};\qquad \gamma(t)\neq\alpha(t). \tag{B9}$$

Differentiating both sides, we obtain, with some simplifications,

$$\begin{aligned}&\frac{\gamma'(t)-\lambda'(t)}{|\gamma(t)-\lambda(t)|}-\frac{\mathbf{L}(t)\langle\gamma'(t)-\lambda'(t),\mathbf{L}(t)\rangle}{|\gamma(t)-\lambda(t)|}\\&\quad=\frac{\alpha'(t)-\lambda'(t)}{|\alpha(t)-\lambda(t)|}-\frac{\mathbf{L}(t)\langle\alpha'(t)-\lambda'(t),\mathbf{L}(t)\rangle}{|\alpha(t)-\lambda(t)|}.\end{aligned} \tag{B10}$$

Taking the limit as $t\to\tau^-$ and noting that $\lim_{t\to\tau^-}\gamma(t)=\alpha(\tau)$, we obtain

$$\begin{aligned}\lim_{t\to\tau^-}\gamma'(t)&=\alpha'(\tau)-\lambda'(\tau)-\mathbf{L}(\tau)\langle\alpha'(\tau)-\lambda'(\tau),\mathbf{L}(\tau)\rangle\\&\quad+\lambda'(\tau)+\mathbf{L}(\tau)\left\langle\lim_{t\to\tau^-}\gamma'(t)-\lambda'(\tau),\mathbf{L}(\tau)\right\rangle,\end{aligned} \tag{B11}$$

which, when simplified, gives

$$\lim_{t \to \tau^-} \gamma'(t) = \mathbf{L}(\tau) \left\langle \lim_{t \to \tau^-} \gamma'(t), \mathbf{L}(\tau) \right\rangle, \tag{B12}$$

and we have, as before,

$$\lim_{t \to \tau^-} \frac{\gamma'(t)}{|\gamma'(t)|} = \frac{\lim\limits_{t \to \tau^-} \gamma'(t)}{|\lim\limits_{t \to \tau^-} \gamma'(t)|} = \mathbf{L}(\tau). \tag{B13}$$

This completes the proof.

APPENDIX C

In Section 5(b), we claimed that when the surface on one side of a concave surface crease casts a shadow on the surface on the other side, the resulting cast-shadow boundary is cotangent to the surface crease at the point of intersection between the two (Fig. 8). We prove this assertion here for an extended light source.

The cast-shadow boundary formed by a smooth attached boundary is the intersection of the illumination surface corresponding to the attached boundary and the illuminated surface. The tangent direction for a point on the cast-shadow boundary is then given by

$$\mathbf{t}_\gamma(t) = \frac{\mathbf{N}_\alpha(t) \wedge \mathbf{N}_\gamma(t)}{|\mathbf{N}_\alpha(t) \wedge \mathbf{N}_\gamma(t)|}, \tag{C1}$$

where $\mathbf{N}_\alpha(t)$ is the surface normal at a point along the smooth attached-shadow boundary (and thus of the illumination surface along the corresponding ray), and $\mathbf{N}_\gamma(t)$ is the surface normal at the corresponding point of the cast-shadow boundary. Let us represent the point of intersection between the smooth attached-shadow boundary and a concave crease as $\eta(\tau)$. As $\eta(tau)$ is approached, $\mathbf{N}_\alpha$ approaches $\mathbf{N}_\Sigma^-(\tau)$ (the surface normal on the casting side of the surface crease) and $\mathbf{N}_\gamma$ approaches $\mathbf{N}_\Sigma^+(\tau)$ (the surface normal on the other side of the surface crease); thus we have at the point of intersection

$$\mathbf{t}_\gamma(t) = \lim_{t \to \tau} \frac{\mathbf{N}_\alpha(t) \wedge \mathbf{N}_\gamma(t)}{|\mathbf{N}_\alpha(t) \wedge \mathbf{N}_\gamma(t)|} = \frac{\mathbf{N}_\Sigma^-(\tau) \wedge \mathbf{N}_\Sigma^+(\tau)}{|\mathbf{N}_\Sigma^-(\tau) \wedge \mathbf{N}_\Sigma^+(\tau)|}. \tag{C2}$$

This is the definition of the tangent direction of the crease; thus, $\mathbf{t}_\gamma(\tau) = \mathbf{t}_\eta(\tau)$. The tangent directions of the cast-shadow boundary and surface crease are equal at the point of intersection. Cotangent curves in the world project to cotangent contours in the image. Q.E.D.■

This work was supported by U.S. Air Force grant AFOSR 90-2074 and National Institutes of Health grant NEI-EY09383. The authors thank an anonymous reviewer of an earlier version of the manuscript for suggestions that greatly improved the paper.

David C. Knill's e-mail address is knill @cattell.psych.upenn.edu.

REFERENCES AND NOTES

1. L. Da Vinci, *The Notebooks of Leonardo da Vinci, Vol. 1* (Dover, New York, 1970).
2. P. Cavanagh and Y. G. Leclerc, "Shape from shadows," J. Exp. Psychol. **15**, 3–27 (1989).
3. A. Yonas, L. T. Goldsmith, and J. L. Hallstrom, "Development of sensitivity to information provided by cast shadows in pictures," Perception **7**, 333–341 (1978).
4. D. Kersten, D. C. Knill, P. Mamassian, and I. Bulthoff, "Illusory motion from shadows," Nature (London) **379**, 31 (1996).
5. S. A. Shafer and T. Kanade, "Using shadows in finding surface orientations," Comput. Vision Graphics Image Process. **22**, 145–176 (1983).
6. J. J. Koenderink and A. J. van Doorn, "The singularities of the visual mapping," Biol. Cybern. **24**, 51–59 (1976).
7. M. Hatzitheodorou and J. R. Kender, "An optimal algorithm for the derivation of shape from shadows," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Washington, D.C., 1988), pp. 486–491.
8. J. R. Kender and E. M. Smith, "Shape from darkness: deriving surface information from dynamic shadows," *Proceedings of the IEEE 1st International Conference on Computer Vision* (IEEE Computer Society, Washington, D.C. 1987), pp. 539–546.
9. M. Hatzitheodorou, "The derivation of 3-D surface shape from shadows," in *Proceedings of the DARPA Image Understanding Workshop* (Morgan Kaufman, Palo Alto, Calif. 1989), pp. 1012–1020.
10. D. G. Lowe and T. O. Binford, "The interpretation of geometric structure from image boundaries," in *Proceedings of the DARPA Image Understanding Workshop* (Morgan Kaufman, Palo Alto, Calif., 1981), 39–46.
11. J. J. Koenderink and A. J. van Doorn, "The shape of smooth objects and the way contours end," Perception **11**, 129–137 (1982).
12. J. J. Koenderink, "What does occluding contour tell us about solid shape?" Perception **13**, 321–330 (1984).
13. M. Ferraro, "Local geometry of surfaces from shading analysis," J. Opt. Soc. Am. A **11**, 1575–1579 (1994).
14. K. N. Kutulakos and C. R. Dyer, "Recovering shape by purposive viewpoint adjustment," Int. J. Comput. Vis. **12**, 113–36 (1994).
15. One local luminance cue that would distinguish between attached and cast shadow contours is the behavior of the field of isophotes in the neighborhood of the contours. On surfaces with locally constant albedo, isophotes are cotangent with attached shadow contours at their junction but intersect cast shadow contours (whether regarded as the interior or exterior of penumbra at sharp angles).
16. V. S. Nalwa, "Line-drawing interpretation: a mathematical framework," Int. J. Comput. Vis. **6**, 103–124 (1988).
17. Langer has developed a model of shape from shading that assumes a hemispherical sky as a light source. This is qualitatively different from spatially compact extended sources, which are the subject of our analysis.
18. M. S. Langer and S. W. Zucker, "Shape-from-shading on a cloudy day," J. Opt. Soc. Am. A **11**, 467–478 (1994).
19. H. Whitney, "On singularities of mappings of Euclidean spaces I: mappings of the plane into the plane," Ann. Math. **62**, 374–410 (1955).
20. A developable surface is a surface with zero Gaussian curvature: all developable surfaces can be generated by folding, without stretching, a flat surface. That an illumination surface from an extended source is developable follows from two facts: first, since it is formed by a sheaf of light rays, it is clearly a ruled surface; second, the surface normals of the illumination surface along a ruling (light ray) at the two distinct points where the ruling intersects the light source and the illuminated surface are equivalent. The condition that the surface normal along a ruling of a ruled surface be equivalent at two distinct points is uniquely met by developable surfaces.
21. At a parabolic point, the conjugate direction of the asymptotic direction is undefined, but the assumption of a general position for the light source makes the case of an illumination ray at a parabolic point being along the asymptotic direction a zero-probability event.
22. J. M. H. Beusmans, D. D. Hoffman, and B. M. Bennett, "De-

scription of solid shape and its inference from occluding contours," J. Opt. Soc. Am. A **4**, 1155–1167 (1987).

23. W. A. Richards, J. J. Koenderink, and D. D. Hoffman, "Inferring three-dimensional shapes from two-dimensional silhouettes," J. Opt. Soc. Am. A **4**, 1168–1175 (1987).
24. D. Kriegman and J. Ponce, "Computing exact aspect ratio graphs of curved objects: solids of revolution," Int. J. Comput. Vis. **5**, 119–135 (1990).
25. The second fundamental form reflects the local curvature structure of a surface. This can be seen most clearly by noting that the coefficients are those of the quadratic terms of a Taylor series expansion of the surface for z in the direction of the surface normal.[26]
26. M. P. Do Carmo, *Differential Geometry of Curves and Surfaces* (Prentice-Hall, Englewood Cliffs, N.J., 1976).