Technical Report 512

Surface Perception from Local Analysis of Texture and Contour

Kent A. Stevens

MIT Artificial Intelligence Laboratory

SURFACE PERCEPTION FROM LOCAL ANALYSIS OF TEXTURE AND CONTOUR

by

Kent A. Stevens

Massachusetts Institute of Technology
February, 1980

This report is a revised version of a dissertation submitted to the Department of Electrical Engineering and Computer Science on February 2, 1979 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The visual analysis of surface shape from texture and surface contour is treated within a computational framework. The aim of this study is to determine valid constraints that are sufficient to allow surface orientation and distance (up to a multiplicative constant) to be computed from the image of surface texture and of surface contours. The report is in three parts.

Part I consists of a review of major theories of surface perception, a discussion of vision as computation and of the nature in which three-dimensional information is manifest in the image, and a study of the representation of local surface orientation. A polar form of representation is proposed which makes explicit surface *tilt* ("which way") and surface *slant* ("how much").

Part II reconsiders the familiar "texture gradient". The perspective transformation is described as two independent transformations that take a patch of surface texture into a patch of image texture: scaling inversely by the distance to the surface and foreshortening according to surface orientation. A measure of texture that varies only with scaling is described (called the *characteristic dimension*) whose reciprocal gives distance information. Evidence for uniformity of the physical texture (requisite for computing the depth map by this method) is provided by local regularity and global similarity of the image texture. A measure of texture that varies only with foreshortening may, in principle, be used to compute surface orientation, but it would be difficult to interpret without knowledge of the physical texture.

Part III examines our perception of surface contours, an ability that has received almost no theoretical attention. It is shown that surface contours are strong sources of information about local surface shape. Plausible constraints are given that would allow surface orientation to be computed from the image of surface contours. The problem of inferring surface shape from the image of a surface contour has two aspects: constraining the shape of the curve in three dimensions on the basis of its image, and constraining the relationship between the surface contour and the underlying surface. Computational constraints for both aspects of the problem are demonstrated, and their plausibility is discussed. Implications for the analysis of specular reflections and shading are noted.

Thesis Supervisor: David Marr

ACKNOWLEDGMENTS

I am deeply indebted to David Marr for supervising this research.

Shimon Ullman and Whitman Richards also provided considerable advice and insightful criticism.

Ellen Hildreth and Eric Grimson read portions of the thesis. I thank them for their contributions.

The environment at the M.I.T. Artificial Intelligence Laboratory has been most valuable.

The thesis is dedicated to Maike.

CONTENTS

PART I. THE COMPUTATIONAL BASIS

PART II. TEXTURE ANALYSIS

PART III. SURFACE CONTOUR ANALYSIS

PART I
THE COMPUTATIONAL BASIS

1. INTRODUCTION

Texture and surface contours are two sources of information about the 3-D shape of visible surfaces which is available in a single image. This report examines the computational basis for deriving an explicit description of surface shape from texture and from surface contours. In each case, the computation cannot be achieved solely on the basis of the image information -- additional constraints must be introduced. Identifying some of these constraints is the primary goal this report. Summaries of the three sections of the report are given in the following.

1.1 Summary of part I

A review of current theories of surface perception is provided which leads to (a) a discussion of how 3-D information is preserved in the image and (b) a discussion of the representation of surfaces.

1. 3-D information is present in the image, in part, as geometrical configurations such as parallelism, inflection points, and regularity. While often described as invariants, they do not have unique inverses back into three dimensions -- very different 3-D configurations may project to the same image configuration. So their 3-D interpretation must be further constrained.

2. Surface orientation is probably represented in a polar form which makes explicit the orientation of surface *tilt* ("which way") and the magnitude of surface *slant* ("how much") rather than the well-known Cartesian form based on Gradient space. The reasons are:

(a) Surface orientation (up to a reflection in slant) is naturally represented in a polar form. The ambiguity in the direction of surface tilt is implicit when tilt is specified only as orientation ($0 \leq \tau \leq \pi$). This ambiguity would have to be expressed explicitly in a Cartesian form.

(b) The computations of slant and of tilt may then be performed independently.

(c) It is observed that imprecision in apparent slant, when present, is not necessarily accompanied by imprecision in tilt. This is more easily attributed to a polar form which orthogonalizes slant and tilt, than to a Cartesian form (each of whose components necessarily are functions of slant and tilt).

(d) Since information about the orientation of surface tilt is often more reliable than information about the magnitude of the slant, discontinuities in surface orientation are more reliably detected when those components are independent. Furthermore, the detection of discontinuities in surface orientation can then be treated as two distinct "subproblems": detecting tilt discontinuities and detecting slant discontinuities.

3. Slant is probably not represented by either the tangent or the cosine of the slant angle (those being two natural choices). On the other hand, slant represented directly in terms of slant angle would require an internal precision of no more than than one part in one hundred to account for the experimental data.

1.2 Summary of part II

The second part of the report re-examines the problems of extracting surface shape information from the familiar "texture gradient". The results are summarized in the following:

1. The perspective projection may be usefully thought of as comprising two independent transformations to any patch of surface texture: scaling and foreshortening. Scaling is due to distance, foreshortening is due to surface orientation. An orthogonal decomposition of the problems of computing distance and surface orientation is therefore suggested: When computing distance, the texture measure should vary only with scaling; when computing surface orientation, the measure should vary only with foreshortening.

2. Texture density is not a useful measure for computing distance or surface orientation, since it varies with both scaling and foreshortening.

3. Distance up to a scale factor may be computed from the reciprocals of characteristic dimensions, which correspond to nonforeshortened dimensions on the surface. Characteristic dimensions may be defined geometrically by the following: (a) they are locally parallel, (b) they are oriented perpendicular to the texture gradient, and (c) they are parallel to the orientation of greatest texture regularity. The computation requires that the surface texture be uniform.

4. Evidence for uniformity of the actual surface texture is both global and local. Locally the texture must project as regular; globally the texture must be qualitatively similar. The assumption that allows one to deduce uniformity is as follows: if the surface texture has small size variance (which may be detected locally), the mean size is assumed constant regardless of where the texture is placed on the surface. Justification for this assumption stems from the following: constraints on the texture size that cause it to be roughly constant (and therefore of small variance) often occur independent of position on the surface.

5. Surface orientation may be computed from the depth map (by computing the gradient of distance) when significant scaling variation is present in the image, otherwise the depth map indicates a flat surface despite the foreshortening gradient (this occurs with curved surfaces in orthographic projection). But measures of foreshortening that do not vary with scaling (such as aspect ratio) are difficult to interpret unless the particular foreshortening function is known which relates the measure to surface slant. Furthermore, successive occlusion associated with viewing texture which lies in relief relative to the mean surface level acts to confound the apparent foreshortening. Slant is therefore difficult to accurately compute. However the tilt may be computed as the orientation of the characteristic dimensions.

1.3 Summary of part III

The third part of the report examines our perception of surface contours, (e.g., the edges of shadows cast on a surface, gloss contours on specular surfaces, wrinkles, seams, and pigmentation markings). Generally the contours interior to the silhouette of an object have been regarded as merely contributing to texture, or to making the surface appear solid, or to simply increasing the complexity of the image. In fact, surface contours provide information about surface shape, given certain restrictions on their interpretation.

1. The analysis of the shape of a surface from surface contours may be decomposed into two problems: reconstructing the corresponding 3-D curves (the *contour generators*) and determining their relation to the surface. This decomposition separates the problem of determining the projective geometry from that of determining the intrinsic geometry.

2. The first problem is constrained by the following restrictions: general position, planarity, symmetry, and minimum curvature variation.

3. The second problem is reduced by assuming the angle between the surface and the plane containing the contour generator is constant. Then if that angle is a right angle, the contour generator is geodesic; if the angle is zero, the contour generator is asymptotic. In either case the contour generator is also a line of curvature. Since it is also planar, the surface is locally a cylinder.

4. We also arrive at the cylinder restriction in the case of parallel surface contours, given two forms of the principle of general position (that of viewpoint and of contour generator placement on the surface). The opacity restriction is also useful, given the planarity and geodesic restrictions, in understanding how an opaque surface lies under a contour generator.

5. Surface markings on synthetic and biological objects and the edges of cast shadows are often geodesic and planar. Gloss contours are asymptotic and planar, at least in the case of orthographic projection and distant light sources. Hence if the contour generator can be reconstructed as a 3-D curve, the surface orientation along the curve can be computed subject to either the geodesic or asymptotic interpretations.

6. Constraints on the intrinsic geometry are also provided by surface contours even if the contour generator is not well determined in space: Gloss contours, highlights, and shading edges tell us of the local Gaussian curvature in some cases.

2. CURRENT THEORIES OF SURFACE PERCEPTION

Surface perception is usually considered to be a process of reconstructing three-dimensional scenes from two-dimensional images. The dimension that is missing in the image is the distance from the eye to points in the environment. That dimension appears to be recovered somehow and its recovery has often been taken as the primary goal of surface perception. While controversy has arisen regarding the source of the distance information (e.g., whether it is derived exclusively from the image or in part from previous experience) it appears irrefutible that we gain a sense of depth from a single monocular image, such as a commonplace photograph. It would therefore seem natural to assume that the visual system internally expresses the three-dimensionality in terms of perceived distance (at least, distance specified up to a scale factor).[1]

But a single image is not what is usually presented to the visual system, for we move through the environment with both eyes open and the environment often contains objects engaged in independent motion. This has lead some investigators to treat single images as special, and to expect that their interpretation, distinguished as "picture perception", is either some derivative of our ability to interpret the dynamic environment [Gibson, 1971; Kennedy, 1974] or a learned skill of interpretation analogous to reading, subject to cultural convention (e.g., [Arnheim, 1954]). Nonetheless, the visual system is often presented with input that is effectively a single image, due to various combinations of monocular presentation, stationary observer, and motionless or distant subjects. An effectively single image also occurs with binocular vision at distances where the stereo disparities are negligible and there is no relative motion. It is reasonable to expect that the visual system has developed means to derive useful information about the environment in these commonly occurring instances.[2]

The single image does not have a unique 3-D interpretation, for the projection that produces the image is a many-to-one mapping, and therefore does not have a unique inverse.[3] Regardless, we usually derive a definite and accurate 3-D interpretation from a given image. So unless we choose to disregard this paradox, we are faced with explaining how we analyze a single image despite its ambiguity. The problem is to understand the source of additional information that allows the unique interpretation to be chosen from the infinity of possible interpretations.

As traditionally understood, there is a perceptual process that recovers distance from the retinal image (or images). Alternatives to recovering distance, such as recovering surface orientation relative to the viewer (slant) or some qualitiative description of surface shape, have also been investigated. But by and large, distance is usually regarded as the primary consequence of the 3-D interpretation, as evidenced in terms such as "depth cues".

Several controversial issues have emerged which have become focal points for the three major theories that

---

1. The orientation of patches of the visible surfaces is a complementary means for describing three-dimensional scenes. Surface orientation will be discussed in section 4.

2. As we attend to details in a scene, the lens accommodates to bring into focus points at different distances. We probe in depth as we vary the accomodation. But the contribution of focus to our perception of distance is weak [Ogle, 1962, p. 266; Graham, 1965, p. 519]. We have no other direct way to "extract" or "recover" 3-D information from the single image.

3. This was actually demonstrated, e.g., by the well-known Ames room [Ittelson, 1960].

will be reviewed momentarily. These issues are:

(a) *the information content of the image.* This issue is emphasized by Gibson. He proposes that complete 3-D information is available in the images presented as one moves through the environment with binocular vision. Similar claims are made about the information carried by texture in the single image.

(b) *the need for interpretation and assumptions in order to process that information.* This issue is emphasized by the depth cue theory (due largely to Helmholtz) which proposes that the image is interpreted on the basis of prior experience.

(c) *the strategies for efficient processing.* This is emphasized by the *Praegnanz* theory (derived from the Gestaltists) which attributes the apparent immediacy of the 3-D interpretation to the application of rules embedded in a representation which is an analog of 3-D space.

These three theories of surface perception will be discussed in the following.

2.1 Gibson's theory

Gibson was the first to suggest that space perception is reducible to the perception of visual surfaces, and that the fundamental sensations of space are the impressions of surface and edge [Gibson, 1950a]. These statements contrasted with the notion of the time that *space* was the object of perception. While not specific as to how surfaces might be represented, his hypothesis led to a shift in research from attempting to understand how the visual system might recover distance for all points in the visual field (as proposed by Helmholtz [1925]) to studying how the various spatial properties of the visible surfaces are perceived.

Gibson's theory of surface perception [1950a, 1950b, 1966] may be viewed as an hypothesis concerning the information content of the visual input, and an hypothesis on how that information is extracted.

First, concerning the information content, it is claimed that there are "variables in the stimulation" sufficient to specify "the essential properties or qualities of a surface" including hardness, color, illumination, slant, and distance [Gibson, 1950b]. For instance,

> *The distance at any point on a receding surface may be given by the relative density of the texture, the finer the density the greater being the distance.*
>
> *The slant of a surface to the line of regard at any point may be given by the rate of increase of elements at the corresponding point in the image. The direction of the slant would correspond to the direction of the gradient* [Gibson, 1950b].

Initially the theory stated that image texture carries sufficient information to perceive these surface qualitites. This conjecture was later dropped; instead the dynamic and binocular images that occur when moving through the environment were expected to provide the complete 3-D information. But the later conjecture is also wrong. Our perception of visual motion from successive images and of depth from stereo pairs of images must embody assumptions (c.f., [Ullman, 1979; Marr & Poggio, 1978]). Simply stated, the visual input does not specify a unique 3-D scene.

Little is said of contours in this theory. In particular, the contours that comprise the boundary of an object's silhouette are distrusted as a source of 3-D information since a given image curve may arise from infinitely many 3-D curves. And surface contours in general are considered only to the extent that they comprise texture (e.g., the furrows of a plowed field).

Let us now discuss how 3-D information is extracted according to this theory. Given the evident richness of visual information provided by natural scenes, Gibson proposes the "generalized psychophysical hypothesis" [Gibson, 1959]:

> *... for every aspect or property of the phenomenal world of an individual in contact with his environment, however subtle, there is a variable of the energy flux at his receptors, however complex, with which the phenomenal property would correspond if a psychophysical experiment could be performed* [p. 465].

The major implication of this hypothesis is that the 3-D information impinging on the retina need only be "registered" in a manner perhaps analogous to a touch sensor registering physical contact. There are two points of contention here: whether there is, in fact, sufficient information in the (possibly dynamic) image to specify a unique 3-D reconstruction, and secondly, whether the computational problems of extracting that information are trivial. First, we consider the sufficiency issue.

Gibson predicted that there is a one-to-one correspondence between the subjective qualities (e.g., apparent slant) of a perceived surface and the actual qualities of the actual surface. Considerable effort has been spent attempting to empirically verify this claim. The following conclusion was drawn in a review by Epstein and Park [1964]:

> *Concerning the psychophysical hypothesis it can be said that Gibson has not proved his case. The experimental data simply do not support the hypothesis of perfect psychophysical correspondence. Nor does the evidence support the contention that perception is "in contact with the environment," that is, veridical, in cases of psychophysical correspondence* [p. 362].

Furthermore they quote Boring [1951]:

> *What Gibson calls a "theory" is thus only a description of a correlation, a theory which tells* how *but skimps on* why *... eventually science must go deeper into the means of correlation, must show in psychology why a gradient of texture produces a perceived depth, not merely that it does* [p. 362].

By and large, Gibson believes that the laws governing light insure that complete 3-D information *must* be present in the image especially in the dynamic case of moving through the environment. The difficulty experienced by others in empirically demonstrating this fact has been attributed to the experimental methodology which attempts to isolate the contributions of a particular source of 3-D information, often termed "reduction conditions". Such experiments are criticized as not "ecological", hence not necessarily involving the processes that govern everyday visual perception:

> *But the research reviewed by Epstein and Park may not be appropriate to test psychophysical hypotheses ... it seems unlikely that our perception of objects in space is based on the processing of only one or a few cues, but rather depends on the generation of a scale of space from which all references are made. Since in the natural environment all of the information about space is consistent, we probably make use of it all in an integrated fashion, rather than separately, cue by cue. What*

> *seems most unlikely is that cues are processed individually and then added together in some manner* [Haber & Hershenson, 1973, p. 302].

It is interesting to observe that Gibson is essentially advocating a scheme for integrating multiple sources of visual information although he does not believe that vision involves "intermediate variables", i.e., representations (section 4). It should be noted, however, that the refusal to expect that the individual sources of information (or "cues") are separately analyzed is quite contrary to the viewpoint taken by this study. Incidentally, Haber and Hershenson's deduction (above) that the visual processing is not modular simply does not follow from the observation that the various cues are consistent. The visual system may make use of the 3-D information in an integrated fashion and also be modular; these two concepts are not mutually exclusive.

This raises a final point. Gibson postulated that our perception is "immediate". But the apparent immediacy of visual perception -- the subjective ease of seeing -- which Gibson cites belies the complexity of the underlying processing. Immediacy suggests rapid computation, but cannot be taken as evidence for trivial, "direct registration". The complexity is recognized by attempting to formulate the problem that is being solved, regardless of how effortlessly we seem to solve it. In that light, it appears doubtful that the various sources of information (e.g., stereo disparity, motion, texture gradients, shading) may be made use of in an "integrated fashion", as suggested. Deriving 3-D structure from visual motion, stereopsis, shading, and texture gradients are all fundamentally different tasks -- the computations are based on different principles and therefore differ fundamentally.

2.2 Depth cue theory

The single image has been understood to be ambiguous, in that infinitely many 3-D scenes could have produced any given image. Helmholtz [1925] described the 3-D interpretation of the image as a problem of determining the radial distance from the viewer to the physical surface along every line of sight. Thinking of the problem in terms of distance, Helmholtz proposed that the visual system interprets *depth cues* by "unconscious inference" drawing on previous visual experiences (c.f. [Helmholtz, 1925; Ittelson, 1960, 1968]).[1] Therefore familiarity with the visual world is central to this theory.[2] Helmholtz is explicit about this in the following:

> *Knowing the size of an object, a human being, for instance, we can estimate the distance from us by means of the visual angle subtended, or what amounts to the same thing, by means of the size of the image on the retina. ... Houses, trees, plants, etc., may be used for the same purpose, but they are less satisfactory, because, not being so regular in size, such objects are sometimes responsible for bad mistakes* [Helmholtz, 1925, p. 283].

Seven depth cues in a single image are given in the following. These are commonly believed to be the sources

---

1. Gregory [1973] draws an analogy between unconscious inference and the process of scientific hypothesis formation, wherein illusions would be attributed to inappropriate assumptions.

2. The emphasis on the role of prior experience appears to address a developmental issue. The approach adopted by this study is to first determine the nature of the computations performed in surface perception, without concern for the nature-nurture issue.

of 3-D in single images.

1. *Occlusion*, if correctly interpreted, constrains the relative depth in the locality of the occlusion. That is, the occluding edge is nearer than that which is occluded. Occlusion has been studied primarily in relation to subjective contours (e.g., [Coren, 1972; Stevens, 1976]).

2. *Retinal size*, from which absolute distance can be inferred, given that the object is recognizable and its actual size is known. However, retinal size has been found to be only a weak source of distance information [Rock & McDermott, 1964]. The relation between perceived physical size, retinal size, and perceived absolute distance is sometimes called the size-distance invariance. Attempts to demonstrate this invariance have produced equivocal results [Epstein & Landauer, 1969; Gogel, 1971].

3. *Aerial perspective*, a subtle cue known to artists that might also be used by the visual system: the tendency for atmospheric haze to reduce contrast and to give a blue tint to distant surfaces.[1] This effect cannot be of general importance to surface perception, particularly in cases of nearby surfaces. And its contribution to the impression of large distances is doubted by Gibson and Flock [1962].

4. *The position of an object in the visual field*. Since we usually see objects that rest on the ground, distance tends to vary monotonically with height in the visual field. Evidence for our sensitivity to this has been found [Weinstein, 1957; Smith, 1958]. Also, the *equidistance tendency*: objects that are adjacent in the visual field tend to appear at similar depth [Gogel, 1965].

5. *Linear perspective*, the projection of parallel lines on a surface into convergent lines in an image; the notion of a vanishing point, and distortions of proximal objects. Usually the effectiveness of perspective is measured by the subjective slant of planar surfaces (e.g., [Attneave & Frost, 1969]), however Jernigan and Eden [1976] have also demonstrated our ability to make accurate distance judgements on the basis of the perspective projection of a cube.

6. *Texture gradients*, e.g., the systematic variation in projected texture (primarily attributed to variations in distance). While usually quantified as the gradient of texture density, other texture measures are proposed [Purdy, 1960].

7. *Shading and shadows*, illumination effects that cause surfaces to appear in relief. These effects are well utilized by artists.

The last three cues are generally termed "depth cues" even though they will be shown to more naturally give surface orientation. In fact, the hypothesis by Helmholtz that the visual system recovers distance information for all points in the image has lead to theoretical difficulties, especially with regard to the information carried by shading and shadows. The addition of shading and shadows to a line drawing strongly enhances the three-dimensionality, therefore, within the Helmholtz framework, these illumination effects are depth cues. But shading is more directly useful as a source of information about surface orientation than about depth. In fact, Ittelson recognized the difficulty in considering shading as a depth cue:

---

1. Depth can also be suggested by brightness, where nearer means brighter. If this is found to be actually contrast, and not brightness, then it could be partially subsumed by aerial perspective.

> *It seems intuitively obvious, and consistent with the evidence, that illumination, color, and shading do serve as cues to apparent depth. However, the exact manner in which they function seems to be qualitatively different from all the other cues. In all other cases, there is some impingment characteristic which, for a given object, varies in some predictable way with the distance of the object. ... It seems most reasonable to consider these cues as contributing to the integration of a complex situation. The observer organizes the total experience in such a way as to make the best "sense" out of it, that is, to make it correspond to the most highly probable condition* [Ittelson, 1960, p. 102].

Shading can be caused by variations in illumination, reflectivity, or surface orientation. When shading is due solely to variations in surface orientation (and not to illumination or reflectivity), the local surface orientation may be determined [Horn, 1975]. With regard to cast shadows, their role in specifying surface shape has not been examined (part III, section 3.3.1).

In contrast to the many depth cues, few cues specific to surface orientation have been proposed. Texture gradients have been related to slant [Purdy, 1960], as has foreshortening (usually described in terms of the height/width ratio of a simple form such as an ellipse [Nelson & Bartley, 1956; Flock, 1964a]). Also, the perspective projections of rectangles as trapezoids have been studied for cues to slant [Freeman, 1966; Braunstein & Payne, 1969; Olson, 1974]. One of the most discussed slant cues is the image of a right trihedral vertex, such as the corner of a cube. There is sufficient information preserved in its image to uniquely specify the 3-D orientation of each of its face. In the general case of the corner projecting as a "Y" configuration, the slant σ of each face of the vertex is related to the opposite obtuse angles α and β by:

$$\sin\sigma = (\cot\alpha \cot\beta)^{1/2}.$$

The apparent three-dimensionality we see in drawings of objects with square corners (as commonly occur in our "carpentered world") might be attributed, in part, to the above relation.

In summary, the 3-D interpretation of depth cues requires additional knowledge, which is usually attributed to prior visual experiences. Depth cue theory expects some form of information processing (in contrast to the direct perception proposed in Gibson's theory), but does not consider how information from distinct depth cues might be integrated into a consistent "depth map". That issue is directly addressed by the following theory.

2.3 Praegnanz theory

The *Gestalt* psychologists observed that we tend to choose visual interpretations that result in things appearing to have minimum complexity. Koffka [1935] then proposed the principle of *Praegnanz*, that "psychological organization will always be as good as the prevailing conditions allow". So rather than have to explain this tendency as a side effect of certain visual processes, it is made integral to a theory of vision:

> *A Praegnanz principle assumes a teleological system (as Koffka [1935] explicitly recognized) in which simplicity has the status of a final cause, or goal-state. It assumes that the rules of perspective (or some approximation thereto) are implicit in an analog medium representing physical space, within which the representation of an object moves toward a stable state characterized by figural goodness or minimum complexity"* [Attneave & Frost, 1969].

This theory, although addressing vision in general, concentrates on simple line drawings where the visual interpretation may vary from simply two-dimensional and lying parallel to the image plane to strongly three-dimensional (c.f., [Attneave & Frost, 1969]). By studying these simple images, they hope to uncover the perceptual rules[1] governing surface perception.

The *Praegnanz* theory directly addresses our ability to combine potentially contradictory information (a point that Gibson dismisses as irrelevant to real situations [Attneave, 1972 p. 284]). Rather than expect that the visual system explicitly resolves this conflict (e.g., by disregarding the lesser reliable information), it is proposed that all contributions meld together to reconstruct a 3-D model within a continuous "analog medium".[2] That representation would preserve the information most essential for survival: the invariants corresponding to the inherent properties of an object as well as its spatial relation to the viewer. The internal representation and its implicit "rules of formation and transformation"[3] are presumed to be in some way complementary to the corresponding external objects and to the "rules of projection and transformation in three-dimensional space" [Shepard, 1979]. Hence the *Praegnanz* theory, like Gibson's, emphasizes the importance of extracting invariant properties, e.g., of size and shape from the variable and shifting patterns of light. To be efficient in this task, the 3-D structure of an object is determined from its image by "rules of formation" which reflect these invariant properties -- the visual system has evolved to take advantage of the constraints imposed by the nature of physical objects and the image-forming process.

Attneave and Frost [1969] take issue with both Gibson and the depth cue theory concerning interpreting geometrical configurations in the image:

> *A cue theory, as we understand it, would have to assume the neural equivalent of a massive table listing correspondences between particular combinations of angles, for examples, and particular slants. With all due allowance for approximation, interpolation, etc., this would require a formidable number of associations.* [With respect to Gibson:] *We have, in fact, employed a "higher-order stimulus variable"* [slant expressed by an trigonometric expression] ... *as a rather successful basis for predicting slant judgements. To suppose that the visual system likewise solves this equation to abstract such a variable strains one's credulity, the more so as one considers in detail the operations involved in the transformation* [p. 395].

Instead, the analysis is believed to be most economically implemented within the analog medium by essentially pulling the image into three-dimensions where the particular 3-D shape would be the result of the simultaneous application of various rules of interpretation; an analogy is drawn to the static equilibrium achieved in a mechanical structure to which various forces are applied. Presumably the visual system converges towards a stable perceptual solution by maximizing some measure of simplicity with a

---

1. The distinction between "cue" and "rule" -- if any distinction may be made -- lies in the manner by which the information is utilized. Cues would be analyzed separately and explicitly; rules would be implicit in some process that imposes them in an integrated manner.

2. The notion of "analog" in this regard has been recognized to be problematic. Probably the intended distinction is that during a perceptual process such as rigid rotation or the determination of a 3-D shape, the stored values representing some perceptual quantity (such as slant, perhaps) would pass through an effectively continuous range of values before settling on the final percept. This is contrasted to a process by which the final value is arrived at directly.

3. E.g., to interpret angles as right angles, shapes as symmetrical, lines as straight and parallel, and to assume that objects are in "general position", i.e., slight changes in viewpoint do not qualitatively change the image [Shepard, 1979]. General position has been recognized as important in studies of machine vision, e.g., [Waltz, 1975], and arises in the analysis of surface contours in part III.

"hill-climbing" procedure [Attneave, 1972]. This measure would include homogeneity of angles, lengths, and surface orientations in the model, coplanarity or equidistance of components, simplicity of spatial relationships, and goodness-of-match between the model and stored schemata [Attneave, 1972].

The analog medium would also serve object recognition by allowing the 3-D structure to be rigidly rotated in order to bring the perceived structure from its initial spatial orientation (relative to the viewer) into some orientation more useful for recognition. Experimental data showing the time to perform mental rotation to vary linearly with the required angle of rotation has been interpreted as evidence for the visual system performing continuous 3-D transformations [Shepard & Metzler, 1971]. Three-dimensional reconstructions would be made from the image within this medium by the implicit application of "rules of formation". But a set of rules has yet to be proposed that would be sufficient to account for our perceptions in natural situations, not simply those involving geometrically simple and symmetric objects. Furthermore, explicit geometrical analysis of the image is regarded as infeasible by the *Praegnanz* theory. Instead, the transformation from image to three dimensions is the implicit consequence of some process that seeks to minimize the complexity of the percept. The theory even proposes a particular mechanism, hill climbing, to perform the minimization. But a computation characterized as a minimization has other equivalent descriptions -- the choice of description is primarily a matter of convenience [Ullman, 1979].

The central hypothesis of the *Praegnanz* theory is probably not minimization, but the feasibility of determining 3-D shape directly from images in general. By "directly" I mean computing a representation of 3-D shapes from a representation of the retinal image without the intermediate construction of a representation of the visible surfaces. This intermediate level is proposed by Marr [1977b] and Marr & Nishihara [1978]. Briefly stated, there is too large a gap between image and object to be bridged by a single "stage" of processing, as it were. That is because features of an image (intensity edges and gradients of intensity, for instance) are not easily related to volumetric, or object, features -- in fact, the whole notion of "object" is difficult to define in terms of its image [Marr, 1977b]. On the other hand, a surface representation is feasibly constructed on the basis of image information since discontinuities and gradients in the image are related to surface features (physical edges, and surface curvature). The surface description would then serve as a natural basis for constructing a volumetric description.

The previous discussions of Gibson, depth cues, and *Praegnanz* have shown the prominent schools of thought on surface perception. In the following section I shall briefly review the computational approach introduced by Marr.

3. COMPUTATIONAL ASPECTS OF VISION

From one point of view, vision provides the organism with useful descriptions of the visible environment [Marr, 1976; Marr & Poggio, 1977; Marr, 1977b]. Early in the course of visual processing the image itself is described in terms of edges, blobs and other intensity variations [Marr, 1976; Marr & Hildreth, 1979]. Subsequently the visible surfaces in the scene are described in terms of distance, surface orientation, and apparent physical edges -- using information from the image description [Marr, 1977b]. And later 3-D shapes are described in terms of volumetric primitives -- using information from the surface description [Marr & Nishihara, 1978].

We may then focus on either of two complementary aspects of vision: understanding the descriptions themselves (e.g., what are the primitives of the description?) and understanding the processes that construct the descriptions.

Visual processes are most feasibly understood when approached at several levels of abstraction [Marr & Poggio, 1977]. At first, a process is understood as an abstract computation -- as a method for applying a set of constraints to a problem. Basic understanding of a visual process comes from recognizing the computational problem that must be solved and determining the set of constraints that allow its solution. More specific understanding of the process comes from determining the algorithm that incorporates those constraints. At the level of algorithm, one addresses such aspects as intermediate constructs (e.g., *place tokens* and *virtual lines* [Marr, 1976; Stevens, 1978]), and computational operations that are biologically feasible [Ullman, 1979]. Finally, to understand the actual mechanisms that implement the algorithm involves neurophysiology.

Since much of this report concerns constraints, it is important to discuss some basic issues concerning them.

3.1 A discussion of constraints

The ambiguity of the image requires that its interpretation be additionally constrained. Stereopsis, motion, shape-from-shading, shape-from-texture, and other processes must incorporate assumptions that further constrain their respective problems. But actually, the degree of ambiguity facing a given visual process depends on *when* it is tackled by the visual system. For example, the false-targets ambiguity in stereopsis does not exist if stereopsis is deferred until after the objects in each of the two images have been recognized (apple in the left image matches apple in right image, etc.). Similarly, motion correspondence would be easier if each image were analyzed to the point of recognized objects prior to determining the correspondence between frames (the rabbit in the first frame matches the rabbit in the second frame). However Julesz [1971] has shown that stereopsis precedes the perception of objects, and Ternus [1926] demonstrated that motion correspondence can be established between simple elements (e.g., edges and points) in successive images without requiring objects recognition.

With regard to texture and surface contours, when are their analyses attempted? In determining that, we fix the sort of information that is available to solve the associated information processing problems -- and thereby determine the sort of constraints that must be applied. In particular, is surface shape described after objects are recognized? If deferred until after objects are recognized then knowledge of the 3-D shape could be brought to bear on interpreting the surface shape from a particular view of that object. On the other hand, if performed prior to recognition, the only information that is available is the geometry of the texture and contours. What, in fact, is the earlies point at which the human visual system can feasibly solve this problem?

First, we know that some aspects of surface perception do not require object recognition. Random dot stereograms, texture gradients, and various abstract art provide example in which surfaces are perceived independent of any understanding of what object might be portrayed. Furthermore, it is infeasible to attempt object recognition without having previously analyzed the image to the point of describing the visible surfaces, in general [Marr, 1977b]. That is to say, surfaces are feasibly described prior to object recognition (as easily demonstrated), and object recognition without previously describing their visible surfaces is probably infeasible in general.

But do all processes of surface perception strictly precede object recognition? That would imply that recognition could not effect the perceived surface shape. This is not the case, as has been demonstrated by the *Gestalt* completion tests [Street, 1931]. Object recognition does contribute to surface perception, however the relative importance of this contribution is not known.

What sort of constraint is provided us for solving the surface shape from texture and surface contours? Primarily they will be geometrical. To illustrate, consider *planarity*, i.e., restricting a 3-D curve which lies across a surface to be planar. The shape of the curve is more feasibly deduced from its projection in the image if it is planar than if it has torsion (twists in space). Hence planarity may be considered as a constraint. But is planarity a reasonable property to assume? How often are curves on surfaces (such as cracks, scratches, pigmentation markings) actually planar? Probably few cu:ves are globally planar, but many can be reasonably approximated as planar for sizeable portions of their length. We might assume that segments of a curve are planar (but certain criteria are needed to delimit the extent of a curve that may be treated as planar).

It follows that constraints that need be valid only locally are more useful to the visual system, as those have a higher likelihood of be valid. A further advantage for local contraint is apparent when actual algorithms are considered that would apply the constraint: If a *local* constraint is sufficient to solve the problem, then the algorithm can be local -- the computation may be performed wholly on the basis of input from some prescribed region of the image.[1] Local algorithms provide an advantage to a biological implementation, both in terms of actual neural connectivity and simplicity of design [Ullman, 1979]. Finally, it would be advantageous to use the results of local surface analysis to constrain subsequent global analysis.

But local constraints whose validity cannot be verified might result in global inconsistency. Do we check for global consistency? The persistent bafflement that we experience in the artwork of M.C. Escher suggests that global consistency testing is not incorporated in our visual system.

Nonetheless, visual analysis based on constraints that are not invariably valid must deal with potentially inconsistent information. The inconsistency might be of the sort just mentioned (i.e., a locally consistent but

1. That region need not be fixed, e.g., in terms of visual angle: The region of visual input may be determined by some local measures in the image. An example of this is given by the description of local parallelism in dot patterns [Stevens, 1978]. The neighborhood size is determined by the local dot density so that a relatively constant number of dots is included. The computation is therefore scale independent (over at least an order of magnitude range of dot density).

globally impossible 3-D configuration) or inconsistency between the independent solutions of either surface orientation or distance provided by independent processes.

This study will not consider the problem of integrating multiple sources of information. The computational problems that arise are probably best studied after the processes that deliver the information are better understood.

One final introductory point regarding constraints should be made: While it is important to understand the particular constraints that are brought to bear in solving a given problem in vision, understanding the constraints alone does not constitute a theory. It is also necessary to understand how the constraints are applied to the visual input -- i.e., the computational method must be determined. This study, however, only attempts to understand some of the constraints themselves.

3.2 Constraints or invariants?

There is widespread agreement that the visual system must utilize "invariants" in the image, where the term "invariant" is intended in its mathematical sense, i.e., when some property or relation is unchanged by a given transformation (see e.g., [Gibson, 1971; Shepard, 1979]). The use of the term stems from the expectation that, in order to "recover" three dimensions, there must be 3-D information preserved by the projection transformation that leads from three to two dimensions. How do these invariants differ from the constraints that I just discussed? This will be examined in the following.

To postulate that the visual system is sensitive to invariant relations is appealing, however one point will be stressed in the following: few properties in the 3-D scene are in fact invariant over the perspective projection onto the image. Of those that are, few have the necessary feature of having an invariant inverse. That is to say, the presence of the relation or property in the image does not necessarily imply the corresponding scene property. For instance, simply because two edges are parallel in the image, their 3-D counterparts needn't be parallel.

We shall see that there is unlikely a sufficient set of invariants with invariant inverses on which to base rules for vision. On the other hand, there are geometrical relations in the image that do have this useful feature, but not invariably. The following is not intended to pan the term "invariant", but to emphasize the necessity for assuming physical properties in order to take advantage of the constraint afforded by these image properties and relations that generally, but not invariably, hold.

First of all, few spatial relations and properties are invariant over projection. Angles and lengths are not preserved, therefore the important properties of perpendicularity, size, and extrema of length are not invariant. Neither are points of maximum or minimum curvature on a curve. Due to obscuration, neither the continuity of a curve and nor its closure are necessarily preserved. Some invariant properties and relations are:

> *collinearity*: If two physical edges are exactly collinear, they will appear so in the image. (This forms the basis for the *Gestalt* rule of "good continuation" across an obscuration.)

cross ratio: If A, B, C, and D are four distinct collinear 3-D points, then the following ratio is preserved in any perspective projection: the quotient of the ratio in which C divides AB and the ratio in which D divides AB.

inflection points on planar curves: An inflection point (of curvature) along a planar curve is preserved in the orthographic image of that curve.

parallelism: Parallel 3-D edges appear (in orthographic projection only) as parallel edges in the image.

proximity: If two 3-D points are proximate, their projections will be proximate in the image.

smoothness: If a physical edge is smooth, its projection will be smooth, when visible.

spatial order: The order of places along a straight line in 3-D is preserved in the image of the places along the image of the line.

straightness: If a 3-D edge is straight, it will appear so in the image.

For most of the above properties and relations their inverse is not invariant, i.e., the presence of the property in the image does not guarantee the presence of that property in 3-D. Consider the invariant relation of proximity: if two 3-D points are proximate, they invariably appear so in the image. The inverse is not guaranteed -- two adjacent points in an image do not always correspond to adjacent points in 3-D.[1] The fact that a given relation or property is invariant does not guarantee that it would be useful for visual processing: the inverse also must be invariant or at least generally[2] valid: invariance alone is not sufficient.

So let us turn the problem around and ask what properties or relations, when present in an image, are necessarily present in the 3-D scene. Consider first the invariances whose inverses are always valid:

cross ratio, inflection points on planar curves, and *spatial order.*

To these we add the invariances for which the inverses are often valid:

collinearity, parallelism, proximity, smoothness, and *straightness.*

To those we add geometrical properties that, when present in the image, imply the corresponding 3-D property. But note that these properties are not invariant over projection.

perpendicularity: If two image contours are perpendicular, they are probably perpendicular in three dimensions.

---

1. However, the inverse is often true, as may be demonstrated by selecting a closely-spaced pair of points at random on a photograph of a 3-D scene. The points usually correspond to physical locations that are nearby in space. This is because, by and large, the world is comprised of smooth surfaces. This relation, phased in terms of continuity, forms one of the basic constraints on stereopsis [Marr & Poggio, 1976].

2. This is the issue of "ecological validity" discussed by Gibson, Brunswick, and others (c.f., [Gibson, 1950a; Postman & Tolman, 1959]).

occlusion: If the termination of a contour lies along another contour, that termination might be due to occlusion, and, if so, implies an ordinal relation between the distances to the two corresponding physical edges.

regularity: Various measures of regularity (e.g., regularity of spacing, density, length, or size) when present in the image reflect 3-D regularity and do not result from a coincidental viewpoint of an irregular surface. Regularity will be discussed further in part II.

symmetry: If a symmetrical configuration is present in the image, it is almost always due to some symmetrical 3-D configuration, and not coincidental. Symmetry will be discussed further in part III.

The above properties, while useful to the visual system as sources of 3-D information, are not strictly invariant.

The basic point regarding these relations is that, to be applied to vision, there is necessarily an assumption that their inverses are invariant. Consider the parallelism relation. While parallel edges in the image do not invariably correspond to parallel 3-D edges, in order for the parallelism to be misleading (i.e, for the 3-D edges to not be parallel) there must be a particular arrangement between the viewer and the 3-D edges. If the *a priori* probability is low for this to occur, then image parallelism would be useful for inferring 3-D structure. There remains the problem of what to do when the situation is misleading, however. With independent information which reveals this fact (e.g., from stereopsis or motion) the analysis might be recognized as incorrect. Clearly, without independent information, the analysis would be incorrect and a "visual illusion" would result.

3.3 One representation, many contributing processes

We will be examining the constraints on the analysis of texture and of surface contours, but in so doing, we implicitly assume that these analyzes are distinct. Is there a single perceptual process, or is the percept the consequence of relatively independent contributions that are combined in some manner? Introspection has often suggested the former (see section 2.1); computational arguments now suggest the latter. This question will be discussed a bit further, since it is important to the rest of the work.

If one introspects on the percept, i.e., the three-dimensionality, there is a unity or homogeneity that some investigators find difficult to explain by separately analyzed cues (e.g., Haber, see section 2.1). Consider the following progression: observe a scene binocularly as you walk about. Then stand still and stare. The absence of motion subtly diminishes the three-dimensionality. Then close one eye (no steropsis) and the sense of depth is further diminished. Next, substitute a photograph taken from the same vantage point (no accommodation), then an architectural rendering (contours, shading, but no texture), then finally a line drawning (no shading). Observe that each successive step weakens the three-dimensionality. This has been interpreted as evidence for a single monolithic process whose performance is progressively degraded under these "reduction conditions".

The subjective homogeneity may also be explained by there being a common surface representation that is developed by relatively independent perceptual processes. The 3-D impression common to the above situations stems from the visual system combining the information from various sources (stereopsis, texture gradients, etc.) into a common representation, from which subsequent analysis and spatial judgments are made. But why should each source be separately processed? There are computational arguments for expecting a modular design [Marr, 1976].

A natural, modular decomposition of visual processing is suggested by the distinct computational problems that must be solved. This is because the sources of information are fundamentally distinct: for instance, *occlusion* is very different from *shading* both in terms of the nature of the information and the assumptions that must be made to utilize that information. It is reasonable to treat occlusion as distinct from shading and to expect that any implementation, biological or otherwise, will reflect that distinction -- there would be no advantage in having interactions between these processes except after their computations are performed and the results are to be combined in some consistent manner.

4. REPRESENTING VISIBLE SURFACES

This section reviews the framework for describing visible surfaces and 3-D shapes proposed by Marr and Nishihara [1978] and gives a computational argument for a specific form in which to represent surface orientation.

4.1 The 2 1/2-D Sketch

Ultimately, the visual system constructs descriptions of 3-D shapes for such purposes as recognition and manipulation. Some of these descriptions are object-centered, i.e., independent of the viewpoint. But an earlier -- and probably prerequisite -- visual description is of the shape and arrangement of surfaces relative to the viewer. This description is viewer-centered. Surfaces are described in terms of surface orientation, distance, and the contours along which surface orientation or distance are discontinuous. Physical boundaries of surfaces are made explicit, but not necessarily those of 3-D objects (whose boundaries are not so easily defined). Hence two distinct representations are proposed: the surface description, called the *2 ½-D Sketch*[1] and the 3-D shape description, called the *3-D Model* [Marr & Nishihara, 1978].

The 2 ½-D Sketch is envisioned as a field of thousands of individual primitive descriptors, each describing the surface orientation or distance at the associated point in the visual field. It would allow information about surfaces derived from stereopsis, motion, shading, and other analyses to be integrated and maintained in a consistent manner. The information in the sketch would then be accessible to later processes, e.g., those that derive volumetric descriptions such as the 3-D Model.

Each representation should be of a form which is easily computed by early visual processes, and also of a form that is useful for the later processes that access the representation. The 2 ½-D Sketch describes surfaces locally and relative to the given viewpoint -- this is a form which is naturally delivered from the image and which may be directly interpreted by subsequent processes. On the other hand, the 3-D Model describes 3-D shapes relative to their prominent axes of elongation (for instance) hence largely independent of viewpoint -- this is a form which is useful for recognition.

We now focus on representing visible surfaces within the 2 ½-D Sketch. This representation probably makes both distance and surface orientation explicit. This would serve three purposes:

> Each type of information, being explicit, would be immediately available for efficient use by later visual processes.
>
> It makes feasible the independent acquisition of each type of information by processes which, by their nature, provide information in one type or the other.
>
> At times information of one type may be more precisely known than the other. Since they would be represented independently, the more precise information would not be degraded by the less precise.

1. So named as it represents 3-D information, but only of the surfaces in the scene that are visible to the viewer.

Surface orientation and distance are roughly equivalent in the following sense: Surface orientation is computable from distance by taking the gradient of distance; the relative distance of two points may be computed by integrating surface orientation along a path connecting those points. The visual system probably takes advantage of this equivalence and explicitly computes surface orientation from distance in one direction, and distance from surface orientation in the other.

We may illustrate one direction by means of stereopsis, which provides distance information in the form of stereo disparity. But we also perceive surface orientation in the random-dot stereogram. It seems most reasonable to expect that the apparent surface orientation stems from analyzing the variations in perceived depth, e.g., by the gradient of the depth map. Another example of our deriving surface orientation from distance is given by figure 1. In this figure occlusion is the only source of 3-D information -- hence most likely a depth map is computed first, and from this we subsequently infer slant. Note that the apparent slant varies with the degree to which successive rows are obsured -- the slant varies according to whether the figure is interpreted as three coins lying on a table, three coins standing on end, or as three billiard balls. In each case the slant is a consequence of the depth interpretation.

In the other direction, distance is derived from surface orientation. Figure 2, which is borrowed from part III of this report, suggests an undulating surface seen in orthographic projection. One may argue that surface orientation is more directly analyzable than distance in this case (part III, section 1.1). On this basis, I suggest that the visual system first computes a surface orientation description from the contours, and subsequently computes a depth map from that description. The following psychological observation also supports this clam: the impression of depth is less definite than the impression of surface orientation. If figure 2 were analyzed in terms of distance, one would then have to explain how surface orientation would be computed from distance with better precision in orientation than in distance. Finally, the "depth reversals" of the familiar Necker cube (see [Gregory, 1970]) is another example of distance being derived from surface orientation, for the cube is usually drawn in orthographic projection. There is only surface orientation information preserved in the orthographic projection of the cube.

In light of these examples of our deriving distance from surface orientation, and *vice versa*, it seems likely that representations of both surface orientation and distance exist and that they are probably coupled. We now will turn to the problem of representing surface orientation.

4.2 Surface orientation

The most direct approach for expressing surface orientation is in terms of the normal to the surface at a point. However there are several ways to describe the surface normal, as will be demonstrated, so criteria will be introduced for judging the likelihood that a given form of surface orientation representation is incorporated in the human visual system. First, we will consider various natural forms for representing surface orientation, then discuss one form that meets these criteria.

4.2.1 Slant, tilt, and gradient space

Since the description of local surface orientation will be relative to a particular line of sight, it is sufficient to treat the optical geometry locally as a spherical projection (the radius at each point on the sphere defines a

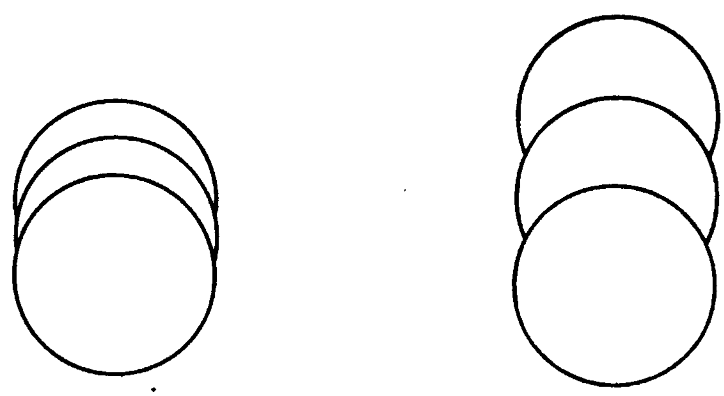


Figure 1. Surface slant can be inferred from distance information: The only source of distance information above is occlusion, for the illustration is in orthographic projection (the circles are equal-sized). The circular figures appear to lie on some supporting plane, the slant of which varies as the figures are interpreted as three coins lying on a table, three coins standing on end, or as three billiard balls. The slant is a function of the degree to which successive figures are occluded, and the radial distance assumed to separate the figures.

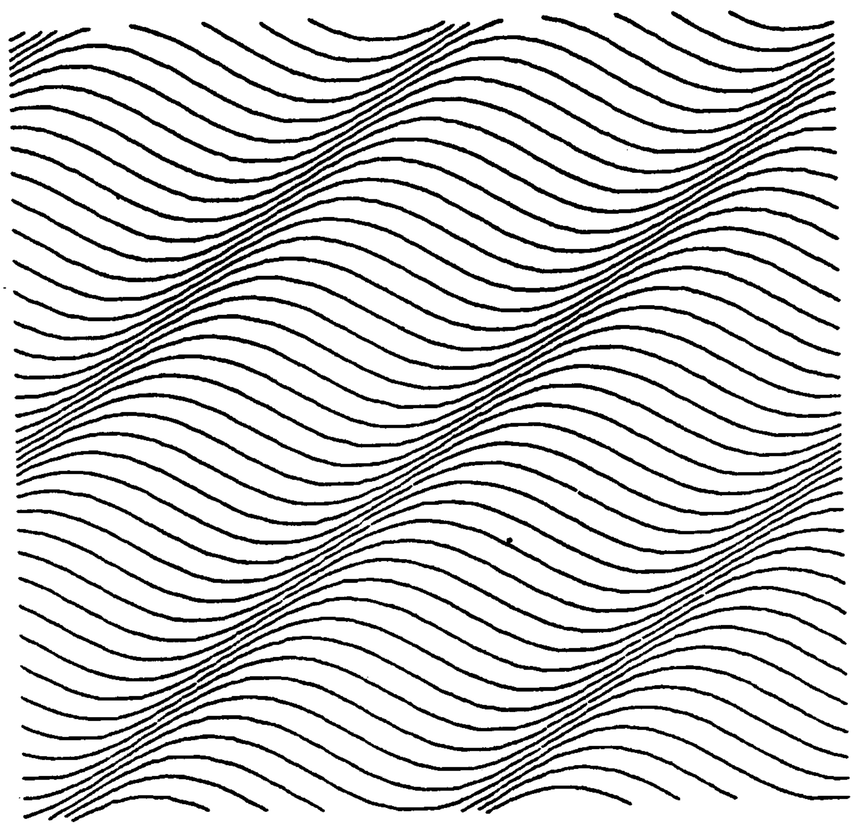


Figure 2. Distance can be inferred from surface orientation: The undulating surface is suggested simply by translated sine waves. In the absence of occlusion and other depth cues, the visual system probably interprets the local surface orientation, and from this derives a sense of depth. Note that just as the local surface orientation is ambiguous in orthographic projection, the depth may be seen to reverse (especially along the diagonal strip where the contours are closely spaced).

particular line of sight). The image in the immediate vicinity of a point on the sphere would project normally onto the tangent plane at that point. Since the image plane is always perpendicular to the line of sight, the projection is locally orthographic. It is important to recognize that the "image plane" notion is an approximation which is valid only locally.

Now we impose a local Cartesian coordinate system on the image plane in order to address nearby image points. We will label the axes of the local system as x and y, remembering that they measure angular displacements about a given image point. Then distance z along the line of sight to points on a surface is given by $z = f(x, y)$. The surface normal N can be expressed as grad f:

$$\mathbf{N} = f_x\,\mathbf{i} + f_y\,\mathbf{j} - \mathbf{k}$$

where f_x and f_y are the first partial derivatives with respect to x and y. The orthographic projection of N is the two-dimensional vector **n**:

$$\mathbf{n} = f_x\,\mathbf{i} + f_y\,\mathbf{j}.$$

Local surface orientation therefore has two degrees of freedom, and the pair (f_x, f_y) would constitute one form of description. That is, surface orientation can be expressed by the rate of change of radial distance in two perpendicular image directions (but the orientation of that coordinate system is arbitrary).

The rate of change of radial distance in an arbitrary image orientation α is given by the directional derivative in the direction α, equivalently the dot product of the unit radial vector of that direction and grad f:

$$dz/dr = f_x \cos\alpha + f_y \sin\alpha. \qquad (1)$$

The image orientation in which this rate is maximized (actually maximum in one direction and minimum in the opposite direction) is given by differentiating (1) with respect to α and equating the result to zero:

$$-f_x \sin\alpha + f_y \cos\alpha = 0$$

which gives

$$\alpha = \tan^{-1}(f_y/f_x) = \tau.$$

This orientation τ indicates the orientation in which radial distance to the surface changes most rapidly. That orientation will be termed *tilt*, where $0 \leq \tau \leq \pi$. Figure 3 illustrates surface tilt by an ellipse, the familiar image of a circular disk in orthographic projection. The orientation of the minor axis coincides with the tilt orientation. Note that specifying only the orientation $(0 \leq \tau \leq \pi)$ and not the direction $(0 \leq \tau \leq 2\pi)$ of surface tilt allows two surface orientations that differ by a reflection about the image plane. This is precisely the amount to which surface orientation can be specified in orthographic projection in general (section 4.2.3).

The slant angle, measured between the line of sight and the normal, is given by:

$$\sigma = \tan^{-1}(f_x^2 + f_y^2)^{1/2}.$$

In short, tilt specifies "which way" and slant specifies "how much".

The tilt orientation was seen to correspond to the orientation of the gradient of distance from the viewer. The orientation in which the distance is locally constant is given by setting (1) to zero, which gives

$$\alpha = \tan^{-1}(f_y/f_x) + \pi/2$$

that is,

$$\alpha = \tau + \pi/2.$$

Thus distance to nearby surface points varies most rapidly in the tilt orientation and is locally constant along the perpendicular orientation. Hence a local Cartesian coordinate system with the y-axis aligned with τ

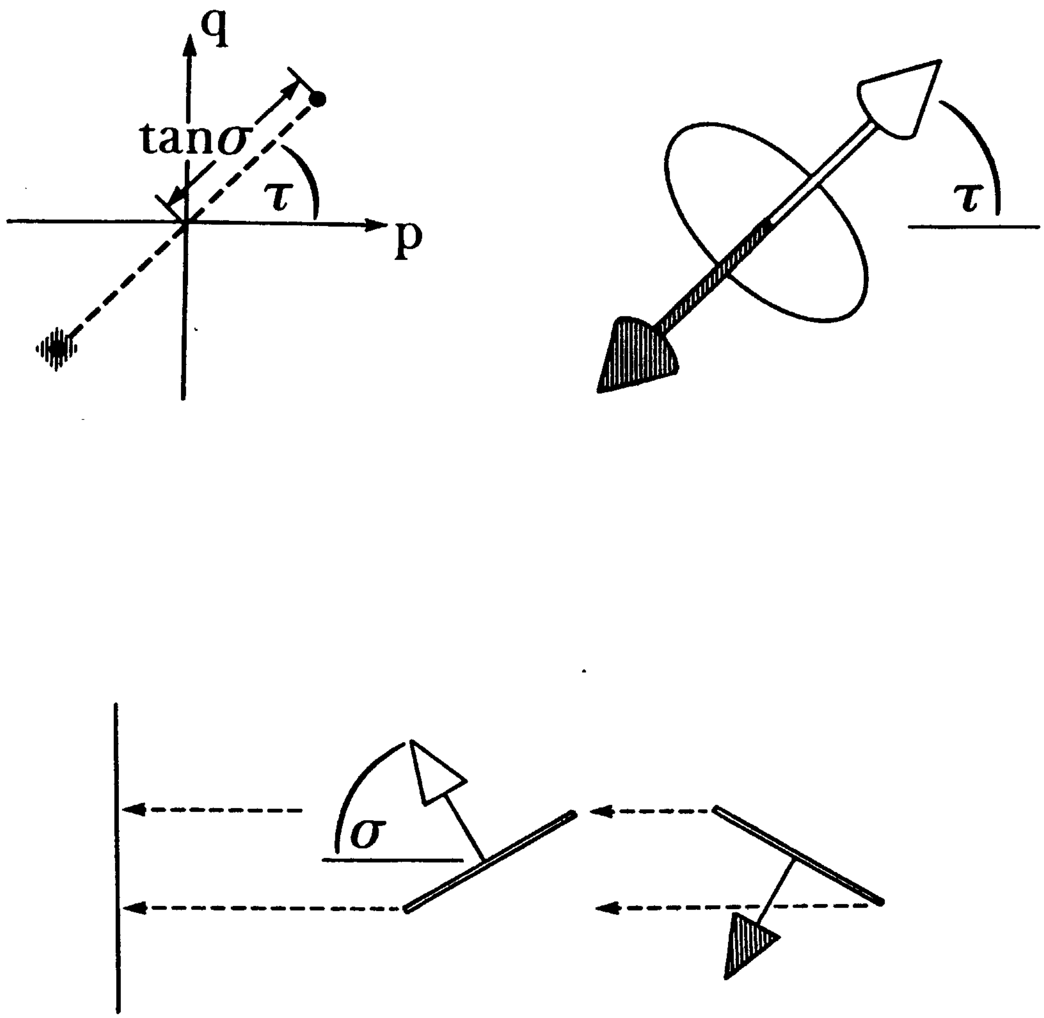


Figure 3. The two degrees of freedom of local surface orientation can be described as the coordinates of a point in gradient space, either as Cartesian coordinates (p,q) or as polar coordinates $(\tan\sigma, \tau)$. The angle σ between the line of regard is termed the angle of *surface slant*, and the orientation τ is termed *surface tilt*. If τ specifies only the orientation $(0 \leq \tau \leq \pi)$ and not the particular direction of surface tilt, then the surface orientation is determined only up to a reversal about the image plane. This ambiguity matches the degree to which surface orientation can be determined from orthographic projection. The slant ambiguity is demonstrated above, with the two interpretations indicated with 3-D arrows. To observe the two interpretations, alternately cover one of the arrows.

provides a convenient way for describing variations in distance in the vicinity of a point on a surface. This will have application in the analysis of texture gradients (part II).

It is common to refer to f_x and f_y as p and q. Then the pair (p,q) may be thought of as the Cartesian coordinates of a point on a plane called *gradient space*.[1] The surface orientation at any point on an smooth surface maps to some point in gradient space. The origin of gradient space corresponds to a surface is parallel to the image plane (zero slant angle).

A natural alternative to addressing a point in Cartesian coordinates is to use polar coordinates. The straightforward conversion gives us $(\tan\sigma,\tau)$ where

$$\tau = \tan^{-1}(q/p) \qquad (2)$$
$$\tan\sigma = (p^2 + q^2)^{1/2}.$$

From this we see that the two degrees of freedom of surface orientation can be expressed as either (p,q) or $(\tan\sigma,\tau)$. However, the representation of surfaces whose slant angle approaches $\pi/2$ would require approximation with both of these forms. (All surface orientations with slant of $\pi/2$ correspond in gradient space to points infinitely far from the origin.) This suggests a second polar form for the primitive descriptor of surface orientation: the pair (σ,τ) where the slant angle, and not its tangent is used. This form will be referred to as *slant-tilt*. Attneave [1972] proposes a third polar form for representing local surface orientation in terms of small ellipses whose orientation corresponds to surface tilt τ, and whose ratio of minor to major axes corresponds to the cosine of the slant angle. That form would be equivalent to $(\cos\sigma,\tau)$.

To summarize, the two degrees of freedom of surface orientation are naturally described in Cartesian form as (p,q), or in various polar forms:

$$(\tan\sigma,\tau)$$
$$(\sigma,\tau)$$
$$(\cos\sigma,\tau).$$

We now consider some criteria for judging the likelihood that a given form would be useful for describing surface orientation within the 2 ½-D sketch. I will use these criteria to argue that a polar form of surface orientation is more likely incorporated in the human visual system than a Cartesian form. But the criteria distinguish primarily between Cartesian and polar forms. They do not distinguish among the various polar forms just listed. The representation of slant was studied experimentally, and it is concluded that slant is probably represented directly in terms of slant angle. That is to say, the representation is probably equivalent to (σ,τ).

4.2.2 Criteria for a representation of surface orientation

The criteria are given in the following, and discussed subsequently. The first two are the most basic:

---

1. Representing local surface orientation by the pair (p,q) has been useful in machine vision (c.f. [Huffman, 1971; Mackworth, 1973; Horn, 1975; Woodham, 1977]). Gradient space is convenient for applying constraints imposed by object geometry and by reflectance properties. A typical use of the space is to represent the allowable range of surface orientations that are consistent with a given illumination situation. When the surface reflectance properties and the position of the light source are known, then the locus of possible surface orientations that might give rise to a particular image intensity can be neatly characterized as a curve in gradient space. Successive application of constraints may further restrict the solution until a small arc, or perhaps a point in gradient space remains [Woodham, 1977].

C1: Is residual ambiguity implicit in this representation? That is, does the ambiguity in the primitive descriptor of the representation reflect the extent to which that information can be known locally?

C2: Is the form compatable with that in which the information can be inferred from the image? In particular, can each component of the primitive descriptor be computed separately?

While it is parsimonious to store information in the same form as it is computed, that form of representation must also be useful to subsequent processes that access the information. So:

C3: Are discontinuities in surface orientation efficiently derived from this form?

C4: Can distance be computed from this form efficiently?

Finally, two phenomena are associated with surface perception that probably bear on the form of the representation of surface orientation:

C5: There is often a disparity in precision between surface slant and tilt judgements. Disregarding the cause of this disparity, does the given form of representation allow slant and tilt to be represented with differing precision?

C6: Can reversals in surface orientation that are associated with depth reversals be attributed to this form of representation?

4.2.3 Residual ambiguity and reversals (criteria C1 and C6)

Surface orientation can be determined in orthographic projection only up to a reflection about the image plane, which I shall term a *slant reversal*.[1] The ambiguity is illustrated in figure 3. How does the visual system handle this ambiguity? One possiblity is that, in fact, the ambiguity does not get carried beyond the analysis of surface orientation. That is to say, the ambiguity is resolved immediately by some means, and so at any one instant only one of the two slant interpretations is taken. The other possibility is that surface orientation is first determined only up to a slant reflection, and that the ambiguity is preserved until it can later be resolved by some subsequent process. This alternative seems more feasible, and is consonant with the hypothesis that the visual system follows the principle of least commitment [Marr, 1976b].

A natural means for preserving the slant ambiguity is by representing surface orientation in a polar form where τ specifies only tilt orientation ($0 \leq \tau \leq \pi$) and not tilt direction ($0 \leq \tau \leq 2\pi$). Hence surface orientation is made explicit only up up to a slant reflection. Subjective depth reversals may then be explained in terms of the slant ambiguity in the surface orientation representation, not to reversals in represented depth, *per se*. Distance may be computed up to a constant from surface orientation, but surface orientation can be determined in orthographic projection only up to a slant reversal. Therefore distance can be computed from this information only up to a sign.

In contrast, a Cartesian form is not as naturally suited to the task of keeping slant ambiguity implicit. The

1. Figures projected in perspective also reverse, whereupon the figure looks distorted [Gregory, 1970].

form (p, q) overspecifies the surface orientation, but if we take the absolute values of each component (|p|, |q|) now there is four-way ambiguity. Since reversals in slant are constrained to either quadrants 1 and 3 or quadrants 2 and 4; one more bit of information is needed which specifies which pair of quadrants are involved. A Cartesian form can be made to specify slant only up to a reversal, but only explicitly.

4.2.4 Computing the primitive descriptor (Criteria C2 and C5)

Criterion C2 states that the form of the representation should match the form in which the information can be naturally computed. The polar form of representation allows a decomposition of the problem of computing surface orientation into two distinct subproblems: determining the orientation in which the surface tilts, and the amount of slant. This decomposition is valuable, for different techniques exist for determining these two quantities. Also, the computation would be robust, for cues to tilt might be present even when the magnitude of slant cannot be determined to any precision. On the other hand, the Cartesian form does not as readily decompose into distinct computations of its two components. In short, the problem of computing surface orientation is naturally solved by determining "which way" and "how much" and a polar form is better suited to that task.

Criterion C5 addresses the problem of accounting for the difference in precision with which two aspects of local surface orientation are judged, the *slant*, or how much the surface orientation differs from the image plane, and *tilt*, the orientation in which the surface normal faces. Slant is often significantly underestimated ("regression to the frontal plane") in monocular and binocular presentation of either perspective and orthographic projections.[1] Furthermore, the perceived slant is strongly affected by the length of presentation time [Smith, 1965]. Apparent slant may even vanish under prolonged observation (this may be observed in figure 2). In marked contrast, judgements of surface tilt are usually more precise, stable, and accurate (appendix A). So although the slant of a surface may or may not be known with precision, the orientation in which it is slanted is usually obvious.

Discussion of the imprecision in judging slant ("regression to the frontal plane", large variance, or U-shaped effect) has usually centered on *explaining* the effect, e.g., as a consequence of a competing tendency to perceive the surface as lying in the frontal plane [Attneave & Frost, 1969]. Of importance to this study is not the cause of the imprecision, but the fact that the imprecision in slant, when present, is not necessarily accompanied with imprecision in tilt.

A polar form would allow the independent computation of tilt and slant. In part II, for instance, we will discuss methods for performing these two computations from texture. The methods for computing tilt are fundamentally different than those for computing slant, and therefore are expected to provide solutions with differing precision. The differing precision is preserved in polar form.

One might argue that surface orientation is, in fact, represented in Cartesian form and therfore the

1. For evidence of underestimation of slant judgments from texture gradients see [Gibson, 1950b; Clark, Smith, & Rabe, 1956; Bergman & Gibson, 1959; Purdy, 1960; Kraft & Winnick, 1967]; in the case of rectangles projected as trapezoids see [Flock, 1965; Flock, et. al., 1967; Kaiser, 1967; Olson, 1974]. Underestimation of slant in orthographic projections is demonstrated in [Attneave & Frost, 1969; Attneave, 1972]. The underestimation may occur even with binocular presentation [Smith, 1965; Kaiser, 1967; Youngs, 1976]. (Note that under excellent binocular viewing conditions the underestimation is not significant, as shown in appendix B and [Olson, 1974].)

experimental design unnaturally imposes slant and tilt judgments on that representation.[1] By this argument, the differing precision in slant and tilt may be an artifact of the experiment. However this argument does not explain the following. The variance and underestimation in slant is dependent on the quality of the visual input: With orthographic projection, the slant judgments are poor and variable while the tilt judgments are more accurate and less variable. And yet, under excellent binocular viewing, both slant and tilt can be judged with precision and accuracy. A Cartesian form is not well suited to the task of simultaneously representing surface orientation known to precision in tilt but imprecisely in slant. But with a polar form, imprecise slant can be represented simultaneously with precise tilt.

4.2.5 Discontinuities (Criterion C3)

A representation of surface orientation would be useful for detecting discontinuities in surface orientation. Some evidence for surface orientation discontinuities are readily extracted by local operators designed specifically to operate on a symbolic description of the image (such as the Primal Sketch [Marr, 1976b]). For example, a discontinuity in tangent along a contour is evidence for a discontinuity in surface orientation, since that would be the most common cause for a contour to remain continuous but suddenly change direction (especially when several such discontinuities align [Marr, personal communication]).

Other evidence for surface orientation discontinuities are not so directly evident in the image, but may be detected after local surface orientation is computed (figure 5). As these discontinutities are more subtle, it would be economical to defer their detection until the 2 ½-D Sketch rather than attempt their detection directly from the image.

Consider the situation where surface orientation is known more precisely in tilt than in slant. This introduces the point of Criterion C3. The detection of a discontinuity would then decompose into two subproblems: finding discontinuities in tilt independent of those in slant. Then the computation becomes straightforward: rather than compute some difference measure that involves both components of surface orientation, the discontinuity would be detected by independent comparisons of slant components and of tilt components. Then a small difference in the tilt components would be significant evidence if the tilt were known with precision.[2]

4.2.6 Distance from surface orientation (Criterion C4)

Distance can be computed from surface orientation, as mentioned. Since surface orientation is the derivative of distance, the difference in radial distance between two points on a smooth surface can be computed up to a constant by integrating surface orientation along a path between the two image points. This computation is straightforward when surface orientation is represented by the Cartesian coordinates (p,q) of Gradient space, for those coordinates are the partial derivatives of radial distance with respect to the image axes.

---

1. If, as is postulated, the visual system represents surface orientation in a polar form, it would be unnatural to judging the components of surface orientation projected along two orthogonal image axes (e.g., horizontal and vertical).

2. The detection of discontinuities in surface tilt then closely resembles the problem of detecting discontinuities in parallelism in an image [Stevens, 1978]. A texture consisting of locally parallel edges can be represented by a field of short oriented elements (virtual lines) which are everywhere locally oriented in the same manner. Analogously, the 2 1/2-D Sketch of a smooth surface would have locally parallel tilt components.

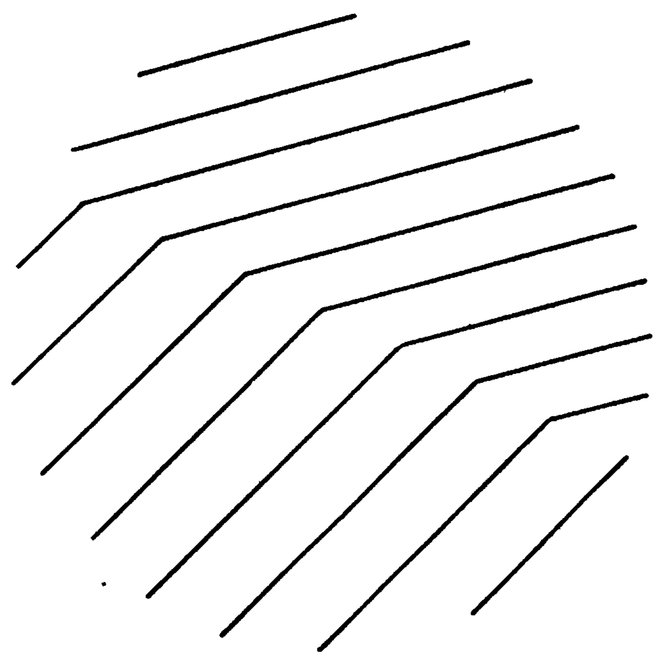


Figure 4. A discontinuity in surface orientation is usually accompanied by a contrast edge in the image, but not necessarily. Other evidence for a discontinuity in surface orientation would be an abrupt change in the slope of continuous image contours. The discontinuity in tangent is strong evidence, since that would be the most common cause for a contour to remain continuous but suddenly change direction, especially when several such discontinuities align. Such evidence can be detected by simple local operators which only signal the presence of a discontinuity without solving the surface orientation on either side of the discontinuity.

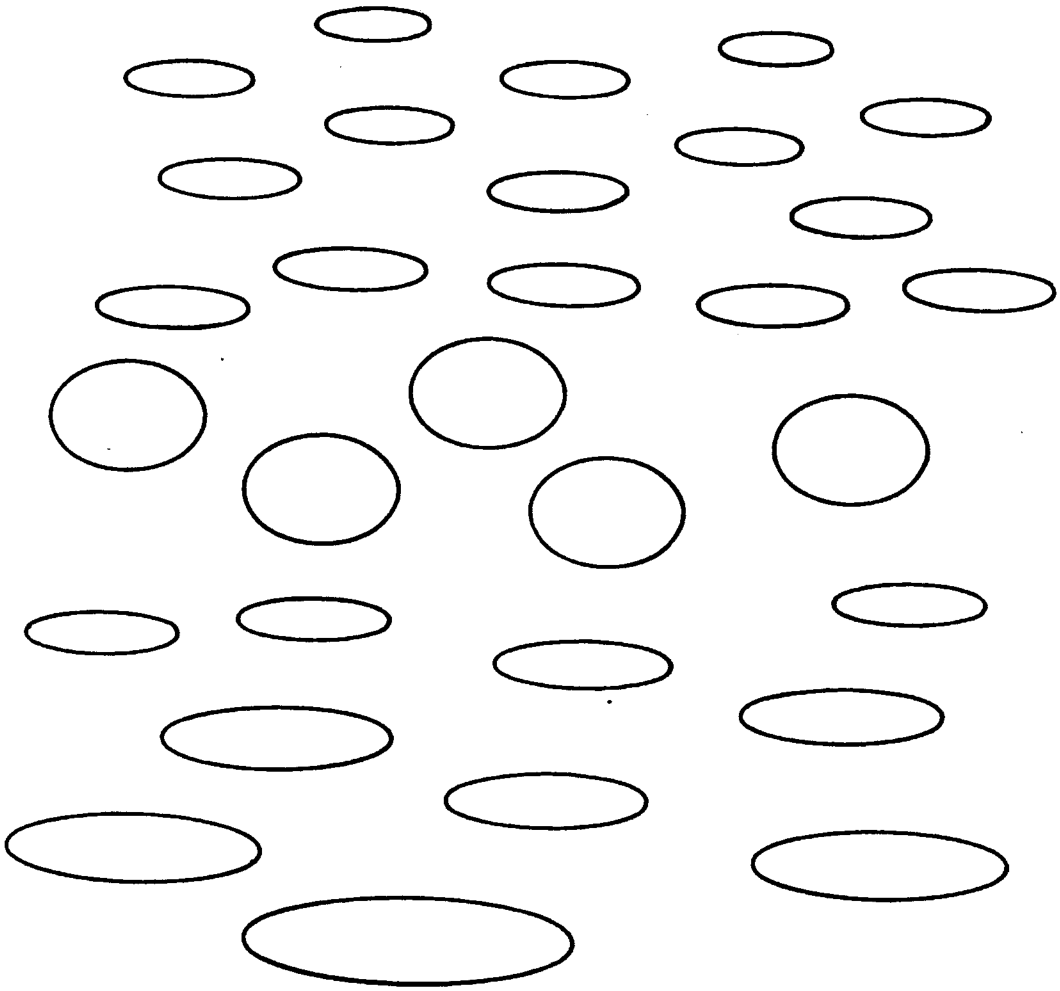


Figure 5. Some discontinuities in surface orientation are probably best detected after the local surface orientation is solved. In the above example, the discontinuity is not evidenced by contrast edges or discontinuities in tangent to contours, but only by a local measure of texture whose value is proportional to the slant (discussed in part II). The detection of discontinuities would be performed economically if deferred until a representation of the local surface orientation is developed. Then discontinuities could be found by examining the representation regardless of the source of the information (e.g., stereopsis, motion, texture gradients). (Note that this and subsequent figures depicting texture are drawn somewhat schematically with ellipses. The discontinuity effect occurs with more natural textures, as well.)

The discussion thus far has favored a polar form for representing local surface orientation, hence it is important to ask whether distance is feasibly computed from a polar form. That computation can be performed by a summation along the path between the two points in question. If the orientation of the path between those points is θ, and the surface orientation of a nearby point along that path is (σ,τ), then the contribution to the summation at that point would be

$$|\tan\sigma\,[\cos(\tau-\theta)]\,|.$$

Since surface orientation can be known only up to a slant reversal in orthographic projection, scaled distance can be computed only up to a sign. Hence the computation of distance information does not have to wait until the surface orientation ambiguity is resolved -- the distance can be computed up to a sign, i.e., to the same specificity to which surface orientation can be known locally. Then other knowledge can either specify the sign and simultaneously the slant direction is resolved, or the slant direction can be determined hence the direction in which distance increases is resolved.

4.2.7 Representing slant

The form in which slant is represented has not been discussed. The range of slants from 0 to 90 degrees is assumed to be represented within the visual system as a set of n resolvable values. That is to say, n distinguishable slants are represented. For any n, there is a grain of resolution that corresponds to an uncertainty in slant. Three natural forms for representing slant would be to store the slant angle σ directly, or either $\tan\sigma$ or $\cos\sigma$. The tangent of the slant angle is suggested, for (a) it is the straightforward polar component taken from gradient space hence the computation of distance from surface orientation would be simplified (section 4.2.6), and (b) a normalized texture gradient provides surface slant directly in that form (part II, section 4). The cosine form has been suggested (e.g., by Attneave [1972]) as a natural expression of slant, in part because it is simply related to the eccentricity of the foreshortened image of a radially symmetric form (e.g., a slanted circle images as an ellipse).

An experiment was performed to determine between these possible forms for representing slant (see appendix B). The result is that slant can be resolved with a precision of better than two degrees over the entire range of slant angle. To represent slant by the cosine of slant angle to this precision would require that the cosine of zero and the cosine of two degrees be resolvable. Consequently, roughly 10^4 resolvable values would be required, which is unlikely, given that slant judgments are precise to only a few degrees out of ninety. Similarly, the tangent form would require considerably finer grain of resolution than is exhibited by our ability to resolve slant angle. If, however, slant were represented directly by angle, the slant representation would not require resolution greater than one part in one hundred.

5. SUMMARY

1. 3-D information is present in the image, in part, as geometrical configurations such as parallelism, inflection points, and regularity. While often described as invariants, they do not have unique inverses back into three dimensions -- very different 3-D configurations may project to the same image configuration. So their 3-D interpretation must be further constrained. The central issue of this report is examining the needed constraints.

2. Surface orientation is probably represented in a polar form which makes explicit the orientation of surface *tilt* ("which way") and the magnitude of surface *slant* ("how much") rather than the well-known Cartesian form based on Gradient space. The reasons are:

> (a) Surface orientation (up to a reflection in slant) is naturally represented in a polar form. The ambiguity in the direction of surface tilt is implicit when tilt is specified only as orientation ($0 \leq \tau \leq \pi$). This ambiguity would have to be expressed explicitly in a Cartesian form.
>
> (b) The computations of slant and of tilt may then be performed independently.
>
> (c) Imprecision in apparent slant, when present, is not necessarily accompanied by imprecision in tilt. This is more easily attributed to a polar form which orthogonalizes slant and tilt, than to a Cartesian form (each of whose components necessarily are functions of slant and tilt).
>
> (d) Since information about the orientation of surface tilt is often more reliable than information about the magnitude of the slant, discontinuities in surface orientation are more reliably detected when those components are independent. Furthermore, the detection of discontinuities in surface orientation can then be treated as two distinct "subproblems": detecting tilt discontinuities and detecting slant discontinuities.

3. Slant is probably not represented by either the tangent or the cosine of the slant angle (those being two natural choices). On the other hand, slant represented directly in terms of slant angle would require an internal precision of no more than than one part in one hundred to account for the experimental data.

PART II
TEXTURE ANALYSIS

1. INTRODUCTION

The image of a textured surface (refer to figure 6) contains 3-D information about the shape and distance of the surface relative to the viewer, and information about the texture itself such as its detailed structure and physical composition. It seems natural to expect that 3-D information can be extracted independently of information about the physical texture. But what about the various types of 3-D information -- can surface orientation and distance information be extracted by distinct computations? The feasibility of such computations is the subject of this part of the report.

The 3-D information is often attributed to the "texture gradient", an informal term referring to the systematic variation in image texture associated with projections of smooth surfaces. There are two assumptions:

(a) that quantitative measurements of image texture such as density are mathematically related to 3-D quantities such as distance, and

(b) that the human visual system somehow capitalizes on these relations in order to derive or extract those 3-D quantities.

It is probably fair to say that neither assumption has been adequately substantiated, as the following discussion will show.

The first assumption concerns the mathematical basis for extracting 3-D information. Several mathematical relationships have been proposed which express either the slant of a patch of surface, or its distance from the viewer, in terms of various "image variables", which I shall term *texture measures*, such as density, size, and foreshortening. Let us consider first the proposed slant relations.

The slant angle was shown to be related to the gradient of various texture measures [Purdy, 1960; Stevens, 1979]. For example, $\tan \sigma = \nabla\rho/3\rho$, where σ is the slant angle, ρ is the texture density at a given region in the image, and ∇ is the "grad" operator. These relations are mathematically correct, but most are probably not useful since they embody assumptions which are seldom satisified in natural scenes. Those assumptions will be discussed in detail later in the article.

The other 3-D quantity which has been related to the texture gradient is distance. Two forms of distance information have been proposed. First, Gibson [1950a, 1950b] claimed that the relative texture density at two regions of the image equals the relative distance of the corresponding surface points. This is not correct. Density is a function of the foreshortening as well as the distance to a give surface point, as will be discussed later. The other form of distance information is not merely a ratio of distances, but some linear distance determined up to a multiplicative constant. Unfortunately, instead of measuring distance radially from the eye to the surface, the distance is measured "on the ground" from the observer's feet, as it were [Purdy, 1960; Bajcsy, 1972; Bajcsy & Lieberman, 1976]. A recent example is found in Rosinski [1974], citing [Purdy, 1960], in which distance D is related to the gradient of texture density ρ by $D = H\nabla\rho/3\rho$, where H is the height of

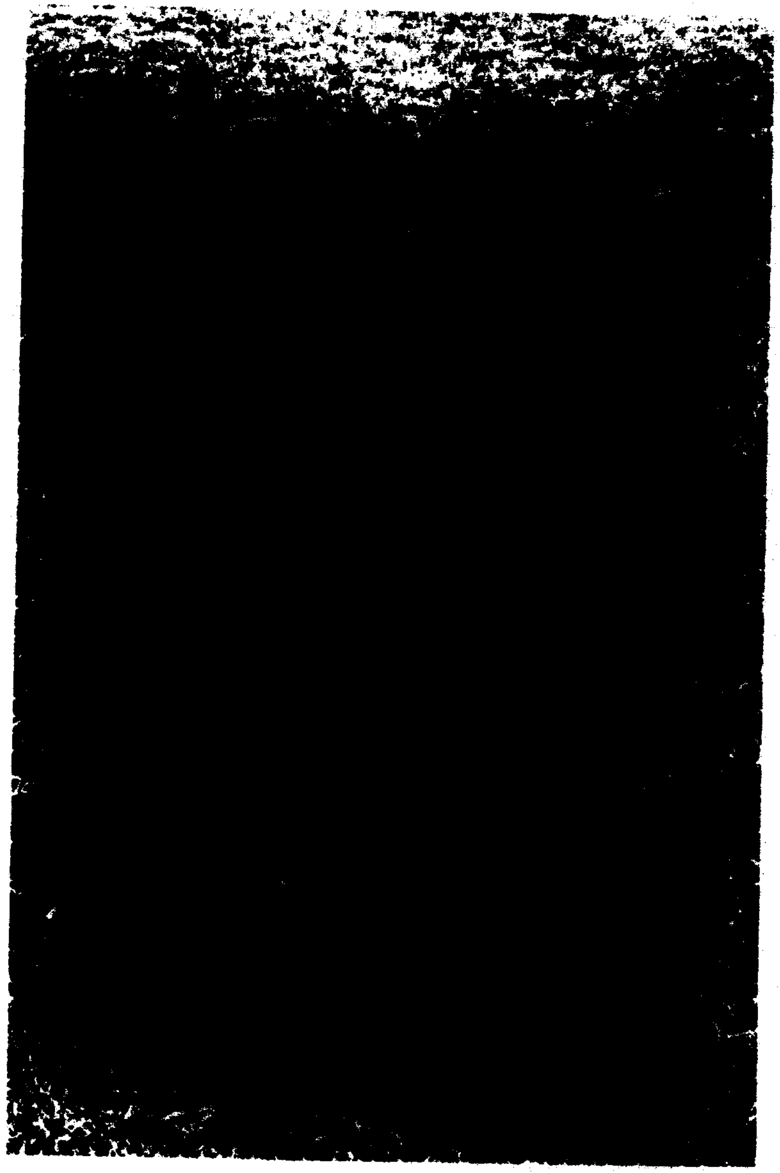


Figure 6. An image of surface texture. The apparent "texture gradient", the smooth variation in image texture, is a consequence of perspective projection. How do we derive the 3-D interpretation of this image? What is computed -- distance, or surface orientation, or both? What constraints underlie the computation?

the eye above the surface. The appealing simplicity of this relation notwithstanding, there are several problems with the underlying definition of distance, D. That definition does not extend reasonably to surfaces other than the horizontal ground (two surface points that are radially equidistant from the viewer but differ in slant would lie at different distances according to that definition). Also it seems not to correspond to the psychological notion of visual distance.

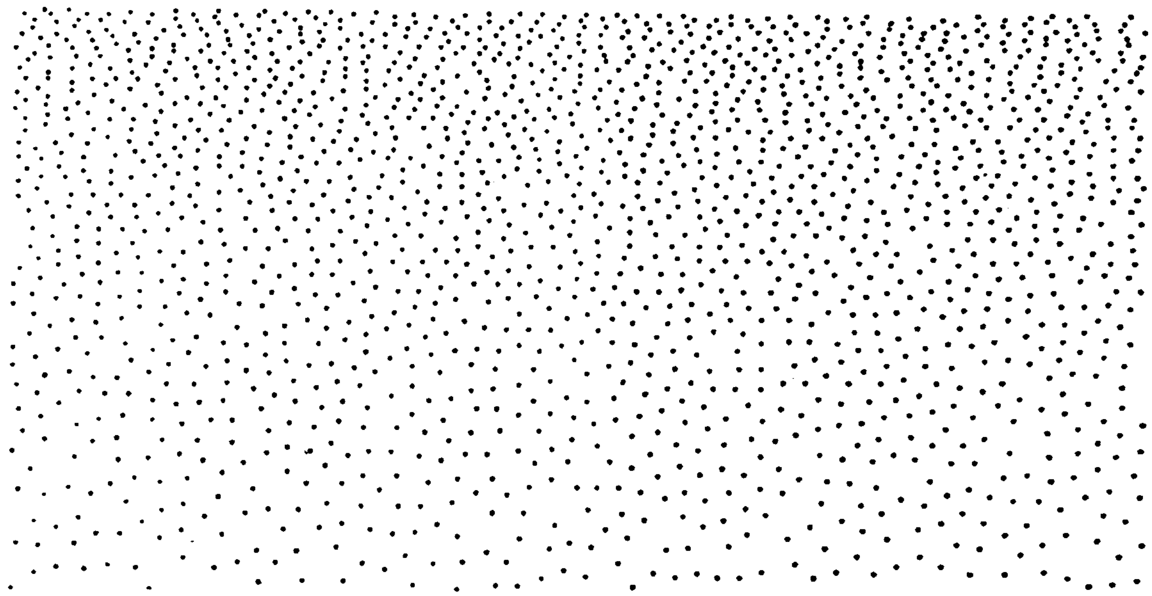
A texture gradient does carry information about the radial distance to points on a surface, however. Distant features on a surface project to a smaller size than those that are closer. A smooth surface of uniform texture therefore presents a continuously varying scale from which distance up to a multiplicative constant might be recovered. (see Gibson's "law of visual angle" [Gibson 1950a] and the discussion of "scale" by Haber and Hershenson [1973]). What remains to be made precise is the notion of "size" or "scale" in terms of real images. That would lead to a simple and elegant mathematical relationship between distance (radial distance specified up to a multiplicative constant) and the texture measure corresponding to "size". It is somewhat surprising that so little attention has been paid to this almost obvious source of distance information. Instead, the mathematical treatment of texture gradients has usually involved rates of change of texture measures.

To summarize this discussion, texture gradients do carry useful 3-D information, but not in the way that it is usually formulated. We now turn to discuss the second assumption, the psychological reality of the proposed mathematical relations, an aspect of the texture gradient problem which has actually received more attention than the theoretical aspect just discussed.

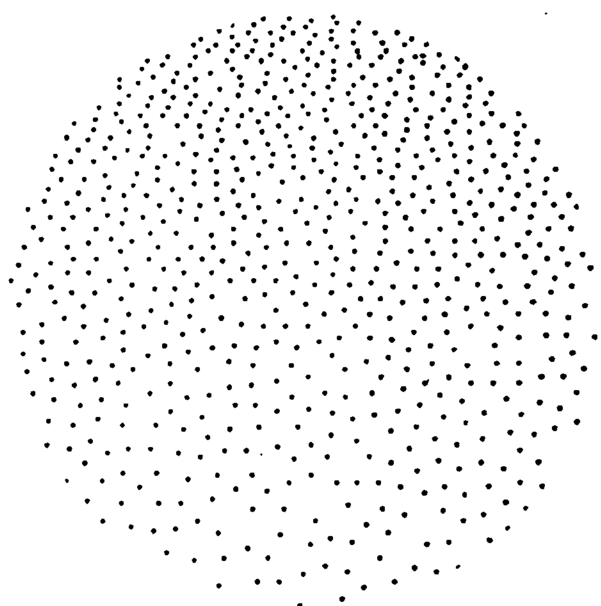
Even if we derive a mathematical expression relating some measure of texture and some 3-D quantity, and this relation is founded on reasonable computational restrictions, it remains to be determined whether the visual system actually uses the given texture measure. For example, one would like to determine, by experiment, whether the visual system derives slant information from the variations in texture density. Unfortunately there is not a sufficiently close correlation between slant judgments and those predicted mathematically to do so -- the experimental evidence is inconclusive (see [Epstein & Park 1964] for a review).

A good example of the difficulty inherent in demonstrating whether a given texture measure is used by the visual system concerns the density measure. Although Gibson [1950a, 1950b] argues the importance of the density gradient, a density gradient of dots does not suggest a surface of definite slant [Smith & Smith, 1957; Braunstein, 1968; Braunstein & Payne, 1969]. To pursue this point a bit further, note that the dot pattern in figure *7a* may seem to be a counterexample -- the impression of a slanted surface is strong. But figure *7b* shows that the impression is due to the apparent horizon. (Figure *7a* viewed with a field-limiting mask similarly fails to suggest a definite surface so long as the "horizon" is not visible).

The ineffectiveness of the density gradient in the case of dot patterns needs explanation. Is it the case that the density gradient is used as a source of 3-D information, but not for dot patterns? (If so, why are dot patterns ineffective -- they provide excellent density information.) Alternatively, is it because the density gradient is not used as a source of 3-D information, and a dot pattern presents no other information such as a gradient of texture size? Later in this article we shall see a strong reason for not using the density gradient. Hence the later alternative is currently favored. The primary point I which to make is the following: there is experimental evidence against the density measure being used as a source of 3-D information, but little



A



B

Figure 7. The density gradient in *a* seems to suggest a surface, but the impression is largely due to the apparent horizon. In *b* the upper boundary is no longer interpreted as an horizon and the pattern no longer suggests a definite surface. There are computational reasons to expect that a density gradient would not be useful for computing shape from texture.

evidence of what measure *is* used.

Another, surprisingly difficult, problem is to determine what sort of 3-D information is computed -- whether it is distance, or surface orientation, or whether both are computed independently. (Other, more qualitative, descriptions of surface shape are also a possibility.) We simply do not know *what* is computed. This point must be settled in addition to the issues of which texture measures and which mathematical relations form the basis of the computation.

Empirical study of texture gradients has been difficult for several reasons. First of all, the slant judgment is a difficult quantity to interpret. The apparent slant is usually underestimated, a phenomenon called "regression to the frontal plane" which varies with time [Gibson, 1950b; Smith & Smith, 1957; Beck, 1960; Purdy, 1960; Freeman, 1965]. The variability and underestimation in slant may be due to several factors, not the least of which is the effectiveness of the given texture in suggesting a cohesive and continuous surface. This confounds any attempt at studying texture gradients with synthesized (e.g., line drawing) textures. For instance, the apparent slant may be increased and the variance of slant judgments reduced simply by increasing the overall texture density while holding the image geometry constant (corresponding to a fixed viewing position relative to a surface whose texture density has been increased). Phenomena such as this make it difficult to postulate differences in visual mechanism on the basis of differences in slant judgment, as attempted in the following.

Figure 8 appears to be a perspective projection of a planar surface with parallel equally spaced rulings, like a plowed field. In fact, a texture gradient comprised of converging linear contours usually produces a more compelling 3-D effect than does a texture gradient of individual elements (figure 9) [Clark, Smith, & Rabe, 1956]. The gradient of spacing between contours has been distinguished from other texture gradients and termed "linear perspective" [Gibson, 1950b; Purdy, 1960; Freeman, 1965]. It has been suggested that linear perspective is analyzed by a distinct perceptual processes, primarily on the basis of the superiority of linear perspective over a gradient of discrete texture elements in suggesting a slanted surface [Gibson, 1950b; Purdy, 1960; Freeman, 1965]. But we shall see later that the computational problems presented by these figures are equivalent and therefore may be solved by the same method. There is no computational reason to postulate separate mechanisms. Furthermore, the noted difference in apparent slant may have other causes -- one need not postulate separate mechanisms to explain that observation.

Also, a texture gradient is difficult to present "in isolation" of other sources of 3-D information. One must first present the texture monocularly, preferably with a synthetic aperature to remove accomodation cues to distance and a chin rest to restrict motion. (A photograph of a textured surface presented in this manner usually provides a satisfactory 3-D impression.) The difficulty occurs in further "dissecting" the texture gradient, for instance, to understand whether the 3-D inpression is due to a gradient of density, or of element size, or of height-to-width ratio, or some combination of the gradients of these and other measures. In a natural scene all measures of texture vary together: as the density increases the elements get smaller, etc. So a computer display seems an appropriate tool, for one may generate synthesized texture gradients where this does not necessarily occur. By controlling the dimensions of the individual texture constituents of the display, one may vary one measure at a time, it would seem. But isolating the contribution of one texture measure is difficult when the "texture elements" have measureable size. (Recall that texture gradients of mere dots do

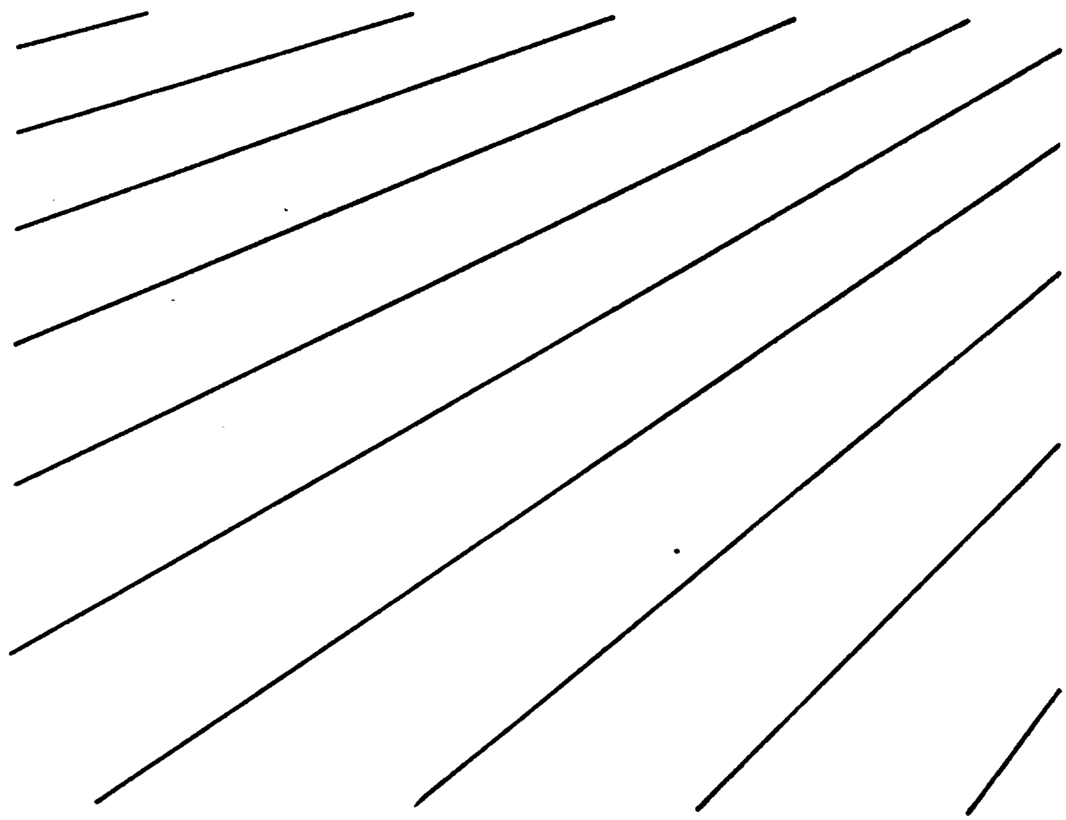


Figure 8. The texture gradient in this figure depicts a planar surface ruled with parallel and equally spaced straight lines. The figure should be viewed monocularly from a distance of roughly 10 inches. This gradient of spacing between contours has been termed "linear perspective" and distinguished from other texture gradients (e.g., figure 9).

Figure 9. This photograph shows a texture gradient which is qualitatively different from the "linear perspective" in figure 8. While these two figures appear different, the 3-D information that they carry may be extracted by a common method. There is no computational reason to postulate separate perceptual mechanisms.

not effectively suggest 3-D surfaces. We are pretty much forced to use textures composed of finite elements.)

For example, suppose one wishes to examine the contribution of density gradients to the 3-D effect. How should the texture elements themselves project? In true perspective the texture elements should be scaled according to their distance. But that would introduce an unwanted gradient of texture size in addition to the desired gradient of texture density. On the other hand, one might attempt to vary texture density while holding the element dimensions constant (this is easily achieved using computer displays, one merely increases the element density appropriately but keeps the element dimensions fixed). But that too is unsatisfactory -- the lack of scaling with distance is distracting and acts to decrease the apparent slant. This problem occurs in attempting to isolate other forms of texture gradients as well.

We will leave the difficult problem of psychological verification just reviewed in order to concentrate on the theoretical problem of relating variables in the image texture to distance and to surface orientation. The first step will be to consider the transformations that occur in projecting surface texture onto the image.

2. SCALING AND FORESHORTENING

When a patch of textured surface projects in perspective onto the image plane, two geometrical transformations occur: scaling and (in general) foreshortening:

> *Scaling* occurs because the surface patch subtends a visual angle that varies inversely with its distance from the viewer.
>
> *Foreshortening* occurs when the surface patch projects obliquely onto the image plane, and so causes the texture to appear compressed in the direction that it slants away from the viewer.

Scaling is actually a function of two variables: the scale of the actual surface texture (whether it is sand or sea waves) and the absolute distance of the surface from the viewer, but if we want to recover distance only up to a scale factor the surface scale is irrelevant. Scaling is an isotropic transformation -- linear dimensions in all orientations are equally scaled. Foreshortening, on the other hand, is an anisotropic transformation -- surface dimensions that lie parallel to the image plane are not foreshortened, all others are foreshortened according to the angles they make to the image plane.

To visualize the commonplace foreshortening function, consider all the diameters of a circle drawn on a slanted surface. The circle projects orthographically to an ellipse; its various diameters are differently foreshortened except for that diameter which lies parallel to the image plane (and which projects to the major axis of the ellipse). The greatest foreshortening occurring to that diameter which projects to the minor axis.

This decomposition of perspective projection into scaling and foreshortening lets us explicitly address the two effects of the projection that are directly related to surface shape. It is from these effects that one may infer distance and surface orientation.

Each small region of image texture may be thought of as the projection of a patch of the physical texture, where the transformation is completely determined by the distance and orientation of the corresponding patch on the physical surface. Can we recover the distance and orientation by somehow measuring the effect of this transformation, without having *a priori* knowledge of the physical texture? (If the transformation has a unique[1] inverse, perspective would be invertible and this would be possible.) The crucial point is to choose the right measure of the image texture. We shall see, for instance, that texture density does not lead to a unique inverse -- the perspective projection is not invertible when described in terms of density.

In general surface texture projects nonuniformly. But what might we infer if the texture is uniform across the image? One interpretation is that the surface texture is uniform and both the scaling and foreshortening are constant. In that case, all points on the surface would be equidistant from the viewer and would present the same surface orientation. On the other hand, the surface texture might not have been uniform; it was only the viewpoint that caused the texture to appear uniform. This is not usually the case, simply because of the rarity of combinations of irregular surface texture and viewpoint that would mislead us this way.

Image texture that varies systematically has been informally termed a "texture gradient". I will continue

1. The inverse phrased in terms of distance need only be specified up to a scale factor.

this use of the term. There are three contributions to the texture gradient, i.e., three causes for the variation in texture:

> (a) variation in distance to points across the surface. The result of distance variation on texture will be termed a *scaling gradient.*
>
> (b) variation in surface orientation across the surface relative to the viewer. The result of variation in surface orientation on texture will be termed a *foreshortening gradient.*
>
> (c) variation in the physical texture across the surface. Nonuniformity of the surface texture may produce a texture gradient that is indistinguishable from that due to scaling and foreshortening. So it is probably necessary to assume that the surface texture is uniform so that the nonuniformity may be attributed to changing distance and surface orientation. (However we shall see that positive evidence may be found in the image that would support this assumption, and also indicate when the surface texture is probably not uniform.)

The foreshortening gradient may be isolated from the scaling gradient by viewing a curved surface from a distance that is large enough so that variations in distance to points on the surface is small compared to their absolute distances, i.e., the surface is viewed in orthographic projection.[1] Bear in mind that the physical texture is assumed uniform. In this situation the scaling is effectively constant across the image of the surface -- there is no gradient of scaling, only a gradient of foreshortening.

But if the same surface is viewed from nearer by, there would be significant variation in the distance to points on the surface. The farther patches of surface project with a smaller scale, so a scaling gradient would also be apparent.

(Note that there will also be a gradient of foreshortening due to variation in the surface orientation relative to the viewer. Hence even a plane surface seen in perspective presents a gradient of foreshortening -- as the line of sight approaches the horizon the slant approaches $\pi/2$ and the foreshortening increases accordingly. Thus it is relative, viewer-centered curvature and not intrinsic surface curvature that causes the variable foreshortening.)

Scaling and foreshortening must be described quantitatively in terms of some measures of texture. By judicious choice of the measure, we can attend to that component of the texture gradient that encodes surface orientation or that which encodes distance. What measurements should be made? Candidates that have been proposed are density, size (the linear dimensions of distinct "texture elements"), area, and height/width ratio (or "aspect ratio"). To preserve the orthogonal decomposition that we have been seeking, the following criteria should be met:

---

1. If the surface subtends a relatively small visual angle one may treat the projection as the conventional orthographic projection (also called parallel projection) onto a planar image. Otherwise it is more appropriate to treat the projection as polar orthographic onto a spherical image.

> When computing distance, the texture measure should be independent of foreshortening.
>
> When computing surface orientation, the texture measure should be independent of scaling.

At this point we understand why density is not a useful measure for computing either distance or surface orientation: Texture density ρ is a function of both the surface slant σ and the radial distance d from the viewer:

$$\rho = \frac{\rho_s d^2}{\cos\sigma}$$

where ρ_s is the surface texture density. Density does not meet either of these criteria, hence does not lead to a simple computation of either distance or surface orientation. This may provide an explanation for the ineffectiveness noted earlier of density gradients suggesting 3-D surfaces.

The next section will introduce a measure of texture that does meet the first of the two criteria, hence would be appropriate for computing distance.

3. COMPUTING DISTANCE FROM TEXTURE

A direct method for computing a *depth map* (a visible surface representation whose values specify the radial distance to the surface up to some scale factor) will be introduced which is based on measurements of texture that vary only with scale, not with foreshortening. Simply stated, we wish to extract a quantitative measure of the local texture that varies only with the distance to the surface, not with the orientation of the surface relative to the viewer. The reciprocal of this measure would be proportional to the radial distance to the surface. The computation itself, therefore, is very simple. The effort lies in extracting the appropriate measures from the image.

A natural measure is provided by what I shall term *characteristic dimensions* which correspond to dimensions on the surface that are not foreshortened, i.e., dimensions that lie parallel to the image plane. One can easily gain intuition for characteristic dimensions by means of a surface texture of circles (figure 10). Each circle foreshortens into an ellipse, with eccentricity that varies by the cosine of the slant angle. The major and minor axes, being well defined in the image, present natural lengths to measure. Of these, the major axis length is the characteristic dimension for this idealized texture -- its reciprocal would constitute scaled distance. (Note however that a real texture would not present as simple an image geometry from which to choose the characteristic dimensions.)

The distance computation based on the reciprocals of characteristic dimensions is valid for any smooth surface, but there is a fundamental restriction: To derive a consistent depth map the measured characteristic dimensions must all correspond to equal surface dimensions -- the surface texture must be uniform. This restriction is probably unavoidable in any method for computing distance from texture, as will be discussed later.

To summarize, the depth map may be computed by:

(a) determining the local characteristic dimensions,

(b) taking their reciprocals as specifying distance up to a single multiplicative scale factor, assuming that they correspond to equal length surface dimensions.

The two steps present the following two problems, both of which are to be solved without *a priori* knowledge of the surface texture. The first will be referred to as the *characteristic dimensions problem*: which of the dimensions definable in the image correspond to nonforeshortened physical dimensions? Secondly, the characteristic dimensions must correspond to equal length surface dimensions for their reciprocals to define a consistent depth map. When is this assumption of global surface uniformity justified? Solutions to these two problems will now be discussed.

3.1 The characteristic dimensions problem

The difficulty of this problem depends on when its solution is attempted. If deferred until the physical units of texture are recognized (as individual rocks, waves, or blades of grass) then their characteristic dimensions may be extracted with assurance. (Also the problem of justifying the equal surface dimension assumption is simplified.) But this texture analysis is probably attempted prior to recognizing the physical causes of the

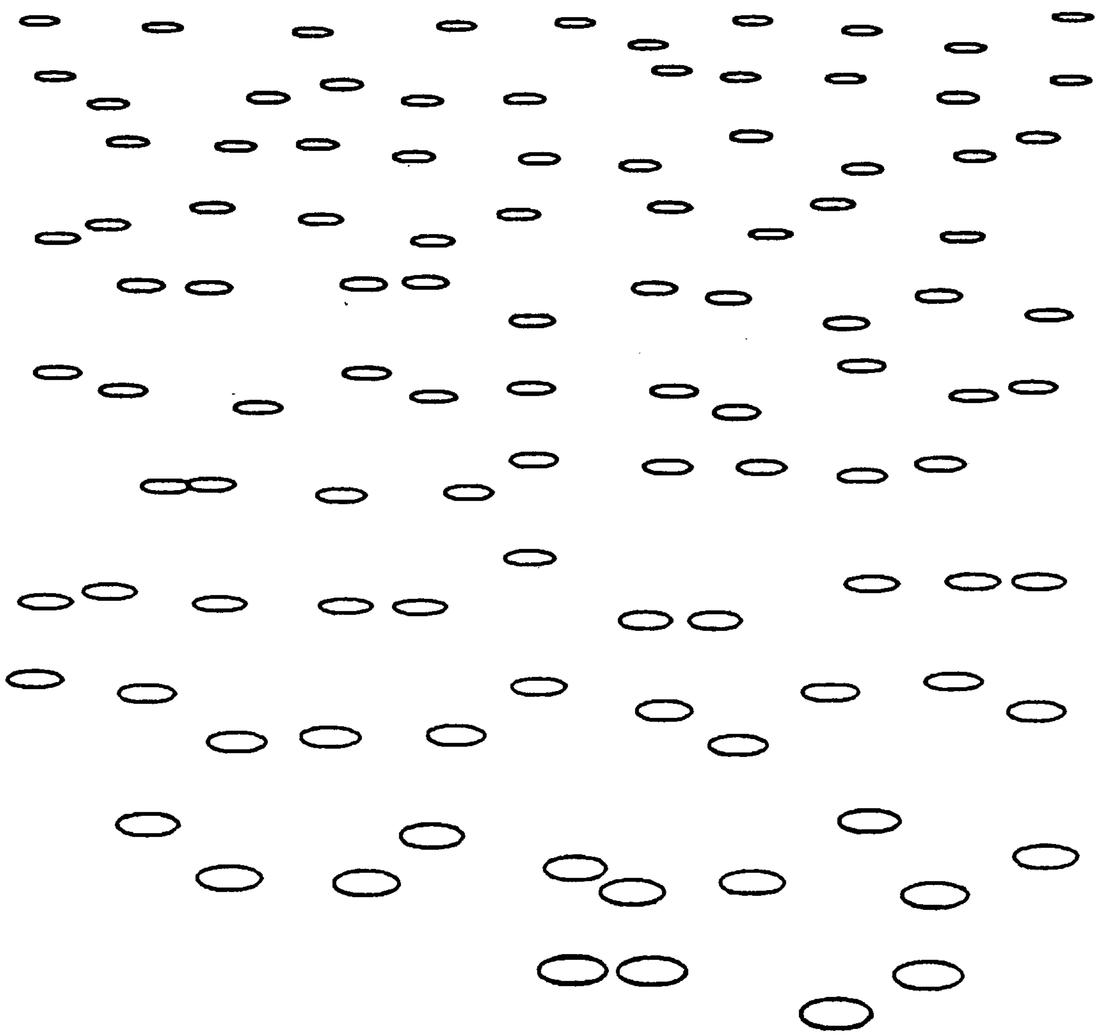


Figure 10. A texture of circles is useful for introducing characteristic dimensions. In this instance, the major axes of the individual ellipses are nonforeshortened and thus may serve as characteristic dimensions. Assuming that the circles are all of equal diameter, the reciprocals of these lengths would provide values for a depth map. A basic visual problem is to determine these dimensions from real images without *a priori* knowledge of the physical surface texture.

image texture, so all that is available to determine the characteristic dimensions is the arrangement of intensity variations in the image. Consequently we seek a geometrical solution.

3.1.1 Characteristic dimensions and intensity variations in real images

Figure 11 shows images of real surface textures where examples of characteristic dimensions are indicated by line segments. These were drawn by intuition, and in questioning how to consciously choose them in these figures we recognize a fundamental computational problem in their extraction: on the one hand, the measurements should depend solely on the viewing geometry and the geometry of the physical texture, but on the other hand, these measurements are to be extracted from intensity information which is intimately tied to the particular illumination and reflectance properties of the surface.

Using the metaphor of applying a ruler to the image -- what should we measure? Perhaps the dimensions of patches of roughly constant image intensity? Or the separations between edges that are intersected by the ruler along its length? Or the dimensions of closed zero-crossing contours available in the computation of the primal sketch [Marr & Hildreth, 1979]. This ruler metaphor suggests methods for extracting quantative descriptions based on explicit measurement of discrete image "features". Alternatively, should we distinguish peaks in the Fourier power spectra [Bajscy, 1972; Bajcsy & Lieberman, 1976]) as signifying the prominent dimension of the texture in any vicinity? This method would use spatial frequency as an image "feature" which seems more continuous than discrete.

How characteristic dimensions are actually measured is not easily settled, since one cannot point to any one method as being intrinsically "correct" -- it is inevitable that any method of solution to this problem will only be heuristic if attempted on the basis of insufficient information, as is the case in attempting to compute a depth map without *a priori* knowledge of the surface texture. The solution is probably based on detectable geometrical properties of the texture which indicate the appropriate lengths to serve as characteristic dimensions. In the following we shall examine these geometrical properties. The distinct issue of how the lengths are actually extracted will not be addressed in this study.

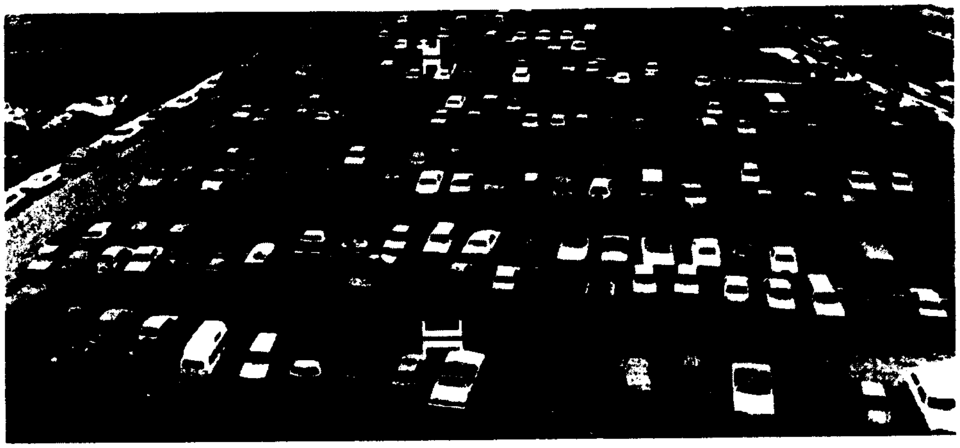
3.1.2 Characteristic dimensions may be defined geometrically

Characteristic dimensions correspond to nonforeshortened surface dimensions, therefore each is the projection of a length lying in the tangent plane of the surface, oriented such that it lies parallel to the image plane. For a smooth surface that means that the characteristic dimensions are locally parallel (and also globally parallel if the surface is planar). Local parallelism is the first of several geometrical properties of characteristic dimensions that may be used as the basis for their selection.

Secondly, the characteristic dimensions are oriented perpendicular to the local surface tilt (this fact was observed in part I, section 4.2.1). What remains to be shown in order to use this property is that the local tilt can be determined on the basis of the texture. But that is straightforward:

For any smooth surface the scaling and perspective gradients coincide -- the orientation of greatest change in foreshortening and the orientation in which scaling varies most rapidly both align with the surface tilt. Consequently the gradient of any measure of texture that is sensitive to either foreshortening or scale, or both, may be used to indicate the tilt orientation.

This second property may be rephrased in the the following way, which although mathematically



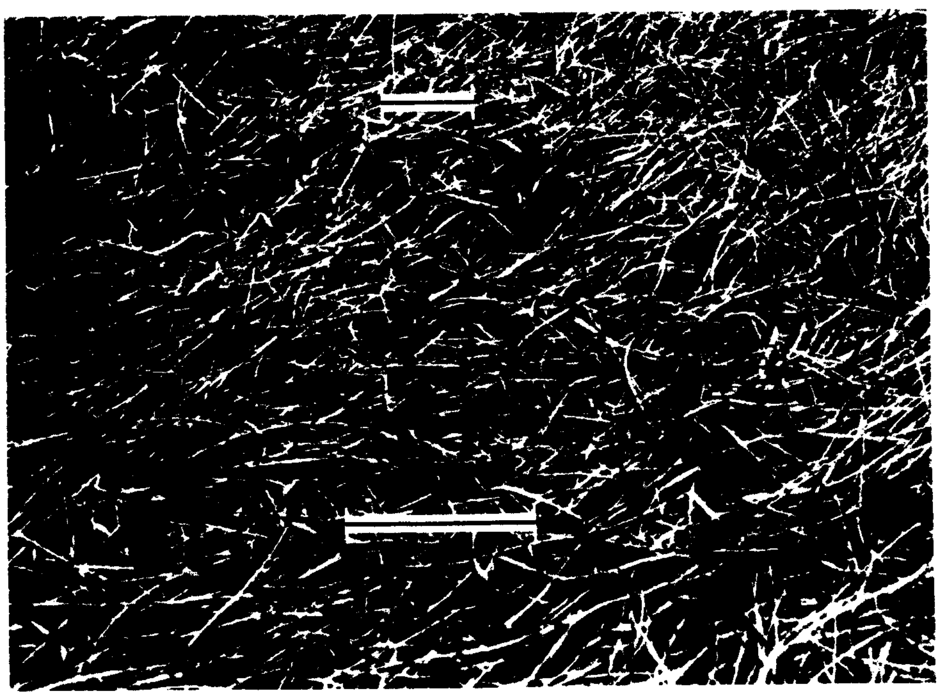


Figure 11. Intuitive choices for characteristic dimensions are indicated by line segments in these instances of textures. In questioning how to consciously choose the characteristic dimensions we recognize a fundamental computational problem in texture analysis: the extraction of quantatative descriptions from intensity information.

equivalent suggests a different algorithm: The orientation of the characteristic dimensions is everywhere equal to the orientation in which measures of texture (that are sensitive to foreshortening or scale variations) exhibit the least variability. That is, the characteristic dimensions are locally aligned with the orientation of greatest regularity. Note that computing this orientation is distinct from computing the orientation of the gradient.

In sum, the characteristic dimensions are locally parallel, oriented perpendicular to the texture gradient, and aligned with the orientation of least texture variability.

3.1.3 An example

In the introduction, the converging lines pattern in figure 8 was given as an example of "linear perspective" and I suggested that there is no computational reason for treating this sort of figure as a special case distinct from textures composed of small discrete features. We will now pursue this point and at the same time provide an example of how characteristic dimensions might be defined in an image.

Consider the texture in figure 12*a*, which when viewed monocularly from the appropriate distance is interpreted as a slanted surface receding in depth. The "texture elements", as it were, are straight lines which, in and of themselves, do not provide useful dimensions (especially when viewed through an occluding mask, as the circular boundary in figure 12 is meant to suggest). One useful texture measure is the *separation* between the lines, which diminishes with increasing distance to the surface. However the term "separation" must be made precise, and towards this end the geometric properties of characteristic dimensions just introduced are useful: An imaginary ruler placed across the image will intersect successive lines at increasing or decreasing intervals along its length, in general. At one orientation, however, successive lines are intersected at regular intervals -- this orientation corresponds to that of the characteristic dimensions (figure 12*b*). The reciprocals of these intervals between lines would give us the depth map. Two observations may be made from this.

First, the characteristic dimensions are locally parallel and oriented with the greatest regularity. But it is difficult to determine the orientation of the *gradient* of spacings between successive lines -- it is not well defined locally. This is particularly true when few lines are presented. Three divergent lines are sufficient for precisely computing the tilt orientation in terms of regularity but not in terms of the gradient. So, despite their mathematical equivalence, the orientation with greatest regularity (or least variability) is easier to compute than the orientation with the texture gradient.

Second, the relevant texture measure does not correspond to the dimensions of discrete "texture elements". Instead, the measurements correspond to laying down a ruler, as it were, and determining the local statistic (such as the separation between successive contours) that is most regular. Importantly, this approach which is exemplified by the "linear perspective" case, extends as well to the more natural case of discrete blob-like textures.

3.2 Uniformity and regularity of surface texture

As discussed earlier, the surface texture is assumed uniform when inferring distance from the reciprocals of the characteristic dimensions. By "uniform" we mean that the physical dimensions corresponding to the

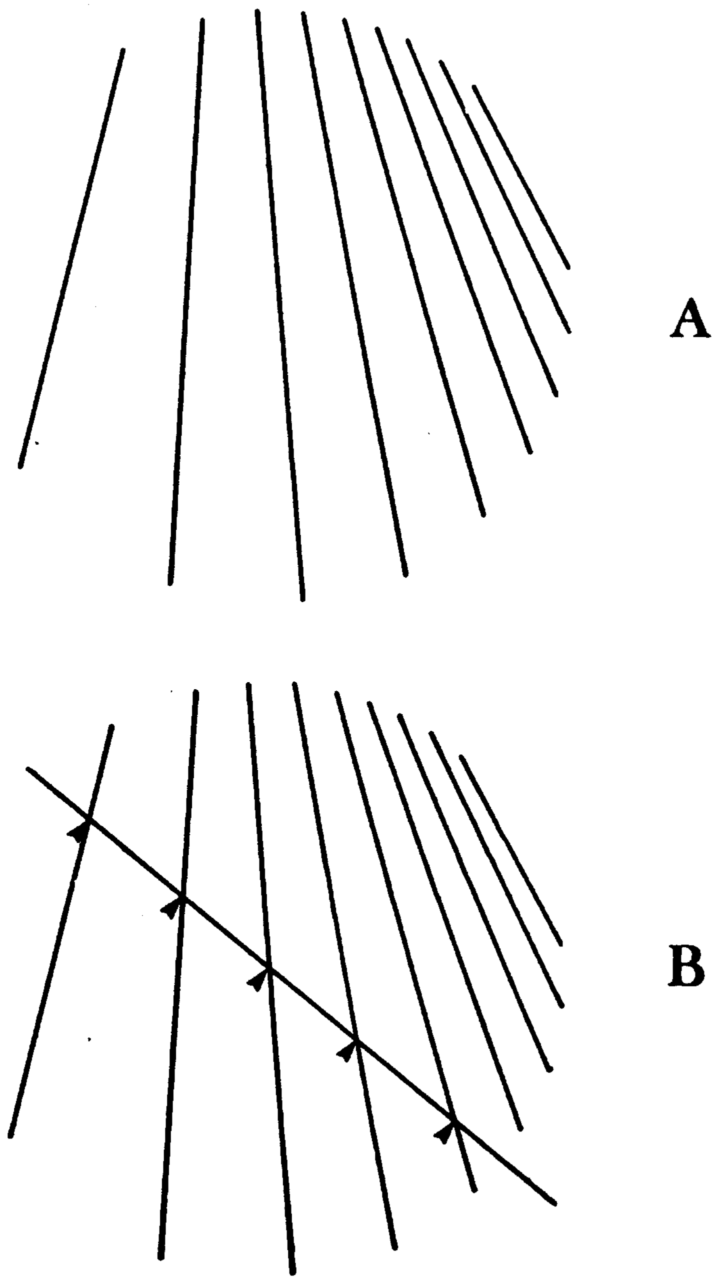


Figure 12. The texture in *a* poses an interesting question regarding the extraction of characteristic dimensions from an image -- how are they defined when the dimensions of the individual "texture elements" are not relevant? The appropriate texture measurement seems to involve the separation between lines. In these terms, we find that the orientation of the gradient is not easily determined, but the perpendicular orientation is. The orientation in which successive lines are intersected with the most regular intervals may be accurately determined by a simple local process. This orientation is shown in *b*, and corresponds to the orientation of the characteristic dimensions. The reciprocals of these intervals, would give us the depth map.

characteristic dimensions are equal across the surface. Is there visual evidence in the image that would support the uniformity assumption? That evidence would allow the distance computation to be restricted to only those instances where the results would likely be accurate.

There are two basic issues that must be addressed. The first is local regularity, as measured by the variation in physical size of the texture markings in any sufficiently small locality. The second is global uniformity, whether the local properties are constant across the surface. The four extremes that might occur are as follows:

> 1. *Locally regular and globally uniform.* Examples would be a field of poppies, cars in a parking lot, leaves on the ground. In each instance the individual elements are restricted to a small range of sizes, and the mean size is constant across the texture. That is, the variance is small and the mean is constant.
>
> 2. *Locally regular but globally varying.* An example would be waves on a lake, where the waves in any vicinity are of similar size but that size varies gradually across the lake according to the wind strength in each region. Another example would be a rocky beach where the surf acts to sort the pebbles according to size. While the variance is small the mean is not constant. I suspect that this case is less frequent than case (1) for reasons that will be discussed shortly.
>
> 3. *Locally irregular but globally uniform.* An example would be a field of rocks where in any vicinity small pebbles might be found beside large boulders, but the distribution of sizes is constant across the field. Another example would be sea waves, where there is a large range of wave sizes in any vicinity, with small waves superimposed on larger. While the variance is large the mean is constant. This is probably a common situation.
>
> 4. *Locally irregular and globally varying.* Any case where the variance is large and the mean is not constant would be useless for the depth computation.

These extremes were presented in the order of decreasing usefulness for the depth computation. Physical texture of type 1 is the best for our purposes. The small variance and constant mean across the surface results in a depth map that is accurate and precise. If the mean varyies slowly (type 2) the depth map would falsely indicate greater distance where the surface texture diminishes in actual size, and *vice versa*. The depth map would be precise but not accurate. If the local size statistics are not tightly distributed, as in types 3 and 4, a different problem occurs: The depth map would be imprecise due to uncertainty in the local characteristic dimensions. For example, with the field of rocks a small pebble might lie adjacent to a large boulder. The characteristic dimensions must therefore be locally averaged in order to estimate the corresponding distance to the surface. In the case of sea waves, however, the distribution of sizes may be broad: small proximate waves may be as plentiful as large distant waves and all intermediate wave sizes may be equally plentiful. In that case it is difficult to compute a useful estimate of the local mean, and depth computation on the characteristic dimensions would require more complexity. (One possibility is to select only qualitatively similar waves, in effect ignoring the small superimposed waves in order to attend to sea waves of common size.)

Reflecting on these four extreme cases, it is apparent that an estimate of the local variance in characteristic dimensions is important. If the variance is low, we have either type 1 or 2 texture and the depth map accuracy is limited by the constancy of the physical mean size across the surface. If the variance is larger (type 3), but the local mean may still be estimated, the depth map may be computed, but to less precision.

The local variance of characteristic dimensions provides an indication of the precision of the depth map, but no indication of its accuracy. Evidence for the accuracy is global, and is based on qualitative similarity of properties that would be invariant over perspective projection. Examples of possible similarity measures are color and intensity statistics, qualitative shape descriptions of the individual markings, and other measures which allow one to determine whether the physical surface texture is qualitatively constant across the surface. That is, global similarity indicates *qualitative* uniformity. The two criteria that we will use, then, are (a) local regularity and (b) global similarity. From these we may infer global texture uniformity in the following manner.

Local regularity indicates the physical surface is either type 1 or 2. Global similarity indicates the surface is more likely type 1, since any physical texture so constrained is probably produced identically across the surface. For example, oak leaves strewn across a yard are qualitatively similar and have similar sizes. The global uniformity in leaf size is a consequence of how leaves develop and is independent of how they are distributed across the ground. In short, type 1 is probably more likely than type 2. If this is true, then in the presence of global similarity:

> the mean physical texture size is assumed constant across the surface if the local variance in image texture is small.

We have discussed the case where the texture has small variance locally. What about types 3 and 4? Can they be distinguished? Without the tight constraint on texture size the constraint on mean size cannot be as readily assumed. Nonetheless, if the texture is qualitatively similar on various dimensions we can assume that the mean, despite the large variance, is roughly constant. That is to say, significant global similarity indicates the surface is likely type 3 rather than type 4.

It must be stressed that these justifications for assuming texture uniformity are heuristic, and that their utility stems from the overall tendency for surface textures that are strongly constrained in their qualitative properties to be constrained in size as well. It easy to find counterexamples to this, nonetheless, it seems unlikely that better evidence may be found in the image.

4. COMPUTING SURFACE ORIENTATION

In perspective projection where significant scaling variation occurs across the image, we have two ways to compute the local surface orientation. The orientation may be computed from the gradient of distance values in the depth map. Also, the orientation may be computed in the image, by the gradient of the characteristic dimension δ:

$$\tan\sigma = \frac{\nabla\delta}{\delta}$$

where σ is the slant angle. In fact, this computation has the benefit over the depth computation in requiring only that the surface texture be locally uniform. But the computation of either distance or surface orientation from characteristic dimensions is ineffective when the surface is in orthographic projection. Despite the foreshortening gradient in the image due to surface curvature, the depth map would be constant, falsely indicating a flat surface. How then might surface orientation be computed?

4.1 Aspect ratio: dependent on foreshortening, independent of scaling

To take advantage of the foreshortening gradient as a source of information about surface orientation, it would be necessary to have the computation valid not only when the projection is orthographic but also when the scaling gradient is significant. This may be achieved by having the texture measure sensitive only to foreshortening, as suggested earlier. A texture measure that has this property is the "height/width" ratio, also called "aspect ratio". This measure is the ratio of the projected dimensions of individual surface markings taken in the direction of the gradient and perpendicular to the gradient (the latter being the characteristic dimension). In the special case of roughly circular surface markings (which project as roughly elliptical) the aspect ratio ε directly indicates the local surface orientation:

$$\cos\sigma = \varepsilon. \qquad (1)$$

But if we are not going to restrict ourselves to circular markings on the surface, the normalized gradient is useful:

$$\tan\sigma = \frac{\nabla\varepsilon}{\varepsilon} \qquad (2)$$

where the particular aspect ratio of the actual surface markings need not be known; they only must be locally constant. The difficulty that arises from this measure ε is as follows: how do we know that the aspect ratio (which we define on blobs in the image, for instance) is a valid measure of foreshortening of markings on the surface?

4.2 The difficulty in computing slant from foreshortening

Surface texture is foreshortened according to the cosine (1) if it lies flat on the surface, as is the case with pigmentation markings and patches of differing physical composition. Examples would be fallen leaves, lichen on a rock, water lillies on a pond, and patterns of mottled light on the ground below a tree. But surfaces are usually textured "in relief" -- the elements that comprise the texture extend above and below the mean surface level. Consider the crests and troughs of waves, rocks strewn across the ground, and blades of grass. When viewed other than at zero slant, the texture is foreshortened, but not simply by the cosine. The relation between ε measured in the image and surface slant σ is not as easily determined without knowledge of the physical texture.

In one extreme, if the surface elements are roughly spherical (e.g., pebbles on a beach) their dimensions would be roughly constant regardless of viewpoint, hence there would not be a foreshortening gradient -- if measured in terms of aspect ratio ε. Nonetheless, there would be a texture gradient due to foreshortening because the surface patch is foreshortened regardless of whether the individual markings on the surface are foreshortened. This would be apparent in terms of texture density, but unfortunately density is confounded by a scaling gradient as well.

In the other extreme, the surface elements might be grass blades which extend normal to the surface, whose foreshortening (measured by the eccentricity ε) would vary according to the sine, not the cosine, of the slant angle. Then we would have that

$$\cot\sigma = \frac{\nabla\varepsilon}{\varepsilon}.$$

Consequently, we have three well-defined foreshortening functions, cosine, sine, and no foreshortening. To choose among these cases in order to infer slant σ from ε measured in the image we must know whether ε derives from texture that lies flat on the surface or from texture that extends above the surface -- and if the texture is in relief, whether it is foreshortened by the cosine or not at all. (Most physical textures do extend in relief and therefore fall intermediate between the extremes of sine foreshortening and no foreshortening.)

Furthermore, if the surface markings are closely packed (as is the case with water waves, tree bark, and pebbles on a beach) there is a succession of occlusion -- of waves occluding waves, for instance. The occlusion is relatively greater with increasing slant and thus affects the apparent aspect ratio as measured by ε. Hence *successive occlusion* amounts to another, confounding, foreshortening effect. For example, the amount of occlusion of successive waves is a complex function of the viewing angle. As this depends critically on the particular surface geometry (it is quite different for tree bark, for instance) we are left with two difficult problems when attempting to infer slant from aspect ratio ε:

> Distinguishing the foreshortening due to oblique projection from that due to successive occlusion. The measure ε would confound the two effects.
>
> Inferring the particular foreshortening function for this texture. What is the relation between ε and σ?

Aspect ratio ε was proposed as an appropriate texture measure for computing surface orientation because it is related to foreshortening but is independent of scaling. But the relationship between ε and σ depends on the particular surface texture, and any choice appropriate for a given situation will often be inappropriate for another. For instance, if the slant computation is correct for flat surface textures it will be incorrect for surface textures in relief. Thus the usefulness of aspect ratio would appear slight.

There is probably no alternative texture measure that is independent of scaling but varies in a predictable manner with foreshortening. Consequently we might turn to a special case approach: using some measure such as texture density, which does vary with both scaling and foreshortening, but only use it when it is known that the scaling contribution to the density gradient is negligible. If the depth map (computed by the reciprocals of characteristic dimensions) is flat, we know the scaling is constant so the gradient of texture density is solely a consequence of foreshortening. Thus we may compute surface orientation from a texture measure that varies with both scaling and foreshortening when the scaling is constant.

We have discovered the difficulty in computing surface slant from measures of foreshortening -- the foreshortening function depends on the particular relation between the surface texture and the surface, which cannot be known *a priori*. Alternatively, the computation may be based not on the foreshortening of the individual surface markings (as measured by ϵ) but on the cosine foreshortening of patches of the surface (as measured by density, for instance). Relative to the computation of a depth map, the computation of local surface orientation appears difficult -- at least the computation of slant does. But the other component of surface orientation, tilt, is readily computed.

The characteristic dimension δ was given a geometrical definition in section 3.1.2: in any small region, they are locally parallel, oriented perpendicular to the texture gradient, and parallel to the orientation of least texture variability (where one may use any measure of texture that is sensitive to foreshortening, or scaling, or both). This definition also suggests a way to computing the surface tilt τ, since tilt is perpendicular to δ. That is, the tilt corresponds to the orientation of the gradient, and is perpendicular to the orientation of least texture variability. (Again I give both definitions because they suggest different computations although they are mathematically equivalent.) Hence one should expect to compute from texture the tilt of the surface more readily and more precisely than its slant.[1]

1. This point supports the argument made earlier (section 4.2 in part I) in favor of decomposing the two degrees of freedom of surface orientation into slant and tilt.

5. SUMMARY

1. The perspective projection may be usefully thought of as comprising two independent transformations to any patch of surface texture: scaling and foreshortening. Scaling is due to distance, foreshortening is due to surface orientation. A decomposition of the problems of computing distance and surface orientation from texture measures is therefore suggested: When computing distance, the texture measure should vary only with scaling; when computing surface orientation, the measure should vary only with foreshortening.

2. Texture density is not a useful measure for computing distance or surface orientation, since it varies with both scaling and foreshortening.

3. Distance up to a scale factor may be computed from the reciprocals of characteristic dimensions, which correspond to nonforeshortened dimensions on the surface. Characteristic dimensions may be defined in the image by the following geometrical properties: they are locally parallel, oriented perpendicular to the texture gradient, and are parallel to the orientation of greatest texture regularity. The computation requires that the surface texture be uniform.

4. Evidence for uniformity of the actual surface texture is both global and local. Locally the texture must project as regular; globally the texture must be qualitatively similar. The assumption that allows one to deduce uniformity is as follows: if the surface texture has small size variance (which may be detected locally), the mean size is assumed constant regardless of where the texture is placed on the surface. Justification for this assumption stems from the following: constraints on the texture size that cause it to be roughly constant (and therefore of small variance) often occur independent of position on the surface.

5. Surface orientation may be computed from the depth map, by computing the gradient of distance, when significant scaling variation is present in the image. However the depth computation fails for curved surfaces in orthographic projection, hence surface orientation cannot be computed from the depth map in those cases -- the depth map would falsely indicate a flat surface. In attempting to compute surface orientation from the image, the texture measure should vary with foreshortening but not vary with scaling. However such measures are difficult to interpret unless the particular foreshortening function is known which relates the measure to surface slant. Furthermore, successive occlusion associated with viewing texture which lies in relief relative to the mean surface level acts to confound the apparent foreshortening. Slant is therefore difficult to compute. However the tilt may be computed as the orientation of the characteristic dimensions.

PART III
SURFACE CONTOUR ANALYSIS

1. INTRODUCTION

This part describes geometrical constraints that may govern the way in which we perceive surface shape from surface contours in an image. In figure 13, for example, the smooth curves are seen in 3-D as lying on an undulating surface. We appreciate not only the shape of the surface, but also its spatial orientation relative to us, and to some extent we perceive the overall surface as receding in depth. The difficulty we face in interpreting figure 13 as merely a two-dimensional family of sinusoids (which it is) shows that we impose constraints in the form of *a priori* assumptions. Some of these assumptions lead us to interpret certain curves in the image as being *surface contours* (which correspond to actual curves across 3-D surfaces); others constrain the inferred surface shape that we derive by analysis of the surface contours. For the surface percept to be both definite and accurate, such constraints must define a unique surface, and must generally be valid.

Although many have considered our perception of the shape of contours (e.g., [Koffka, 1935]), the problem of inferring *surface shape* from surface contours has received virtually no attention. The primary intentions of this part of the report are

(a) to formalize the computational problem,

(b) to introduce useful and valid constraints towards its solution, and

(c) to describe why those constraints are useful.

1.1 What information is carried by surface contours?

The contours in figure 13 are in orthographic[1] projection; hence we cannot derive distance information from perspectivity in the image. But the shape of the contours does provide surface shape information in two forms. In the vicinity of the surface contour one may deduce either:

> *surface orientation.* The relative surface orientation may be solved uniquely (i.e., up to a slant reflection since the projection is orthographic) or only to within a restricted range of slant and tilt.
>
> *qualitative surface shape.* The intrinsic geometry of the surface may be deduced from the shape of the surface contours. The primitive descriptors might include "flat", "singly curved", "cylindrical", "doubly curved" and so forth. This sort of shape information is independent of the viewpoint.

1. Orthographic projection is equivalent to a parallel projection, as opposed to a perspective projection. Figure 13 demonstrates that we may perceive shape from surface contours in orthographic projection. Later we will see that assuming that the projection is orthographic (and not perspective from some unknown viewing geometry) is probably necessary in the analysis.

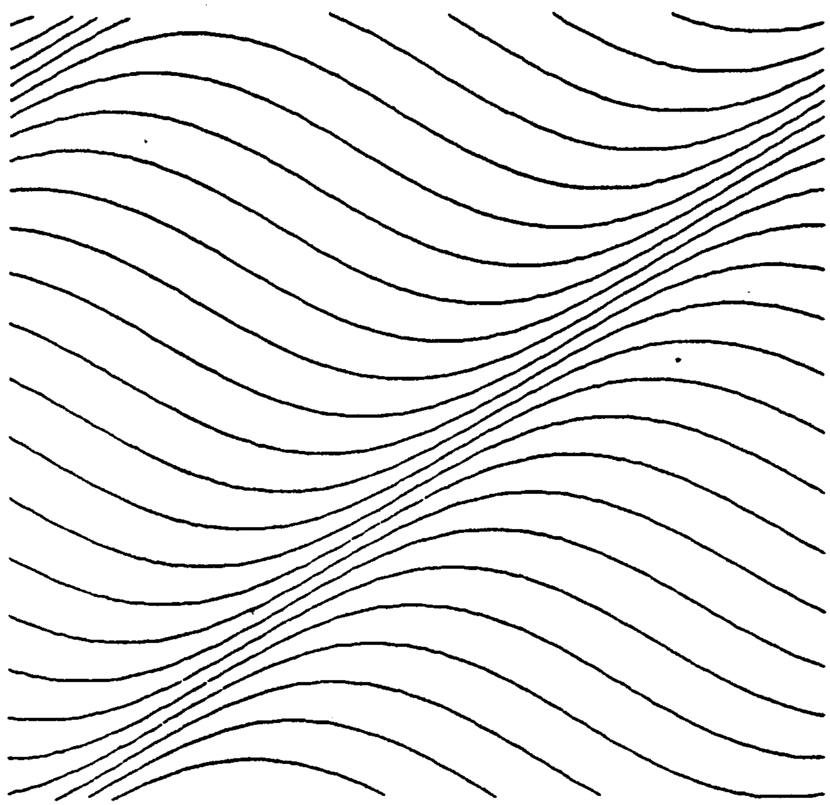


Figure 13. The undulating surface is suggested by a family of sinusoids. (This figure is adapted from Bridget Riley's *Kataraki 3*.) The curves are naturally interpreted as surface contours, i.e., the images of markings on a physical surface. What constraints can be brought to bear in making this 3-D interpretation?

This is not to say that a depth map may not be computed from the image, but that the geometry of contours in an orthographic image more directly constrains surface orientation and intrinsic geometry than distance -- the computation of a depth map would effectively require the intermediate computation of surface orientation.

Note that information about intrinsic surface shape serves two useful purposes: (a) it constitutes a primitive, coordinate-free shape descriptor, and (b) it constrains the values in any representation of surface orientation or distance. Suppose that it can be determined from the image that a surface region must be singly curved, then this restriction can be imposed on any independently computed distance or surface orientation representation -- the distance or surface orientation must vary in a manner consistent with a singly curved surface. Later we shall see the contribution of this qualitative shape constraint on the computation of "shape from shading" (c.f., [Horn, 1975]).

1.2 Contours and contour generators

It is valuable to distinguish between a *contour* in an image and the corresponding curve in 3-D, called the *contour generator*, that projects to that contour (see [Marr, 1977a]). The contour generator is a physical curve which lies across a surface, such as a boundary between patches of differing reflectance (e.g., a pigmentation marking), a discontinuity in illumination (e.g., a shadow edge cast across the surface) or a discontinuity in surface orientation (e.g., a crease). The contour generator may also correspond to the boundary of the surface from the given viewpoint.

So on the one hand, we have the contours in the image; on the other hand, their corresponding physical curves in 3-D, the contour generators. To make 3-D interpretations from the image contours we often need to understand what causes them -- whether they correspond to object boundaries, shadow edges, or what.

One basic distinction that is often proposed is between *object outlines* (also termed *bounding contours* or *occluding contours*) which correspond to the edge of an object's silhouette from the given viewpoint, and those contours that lie internal to the silhouette (which Gibson has called "inlines"). A slight variant would be to distinguish only those bounding contours that correspond to the silhouettes of *smooth objects*. This distinction is probably fundamental for reasons that will be given in the following.

1.3 Tangential contours and surface contours

Physical objects are often smooth, and their silhouettes alone provide a strong source of information about the overall shape [Marr, 1977a]. For instance, consider a vase. Its silhouette projected onto the retinal image might appear like the outline shown in figure 14*a*. In this case, the contour that comprises the outline wil be termed a *tangential contour*. The name stems from the important fact that the line of sight just grazes the surface (i.e., lies tangential to the surface) along the corresponding contour generator. This is a direct consequence of the smoothness of the object. An important class of outlines are those that exhibit qualitative symmetry across an axis (e.g., figure 14*a*). If is assumed that the corresponding surface is smooth then the silhouette is that of a generalized cone whose 3-D shape is recoverable (given some other restrictions, see [Marr, 1977a]). In this case, the silhouette boundary is comprised of tangential contours. Note that the surface orientation is known along a tangential contour: the slant is $\pi/2$ and the tilt is perpendicular to the contour.

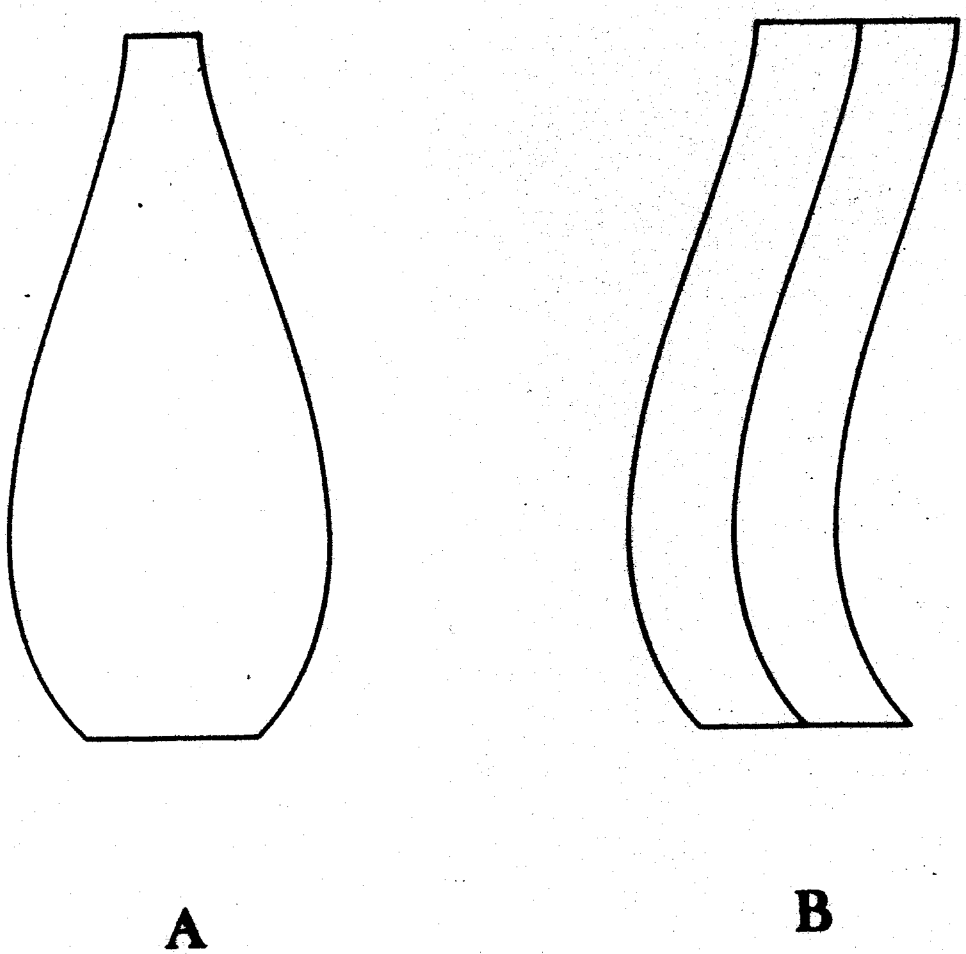


Figure 14. The curves in *a* are interpreted as tangential contours and the underlying surface is seen as a generalized cone, in this case, a vase-like object. Those in *b* are interpreted as surface contours and the surface appears like a gently curved flag or a ruled sheet of paper.

In the previous discussion the object was assumed smooth, whereupon its outline is comprised of tangential contours. But this is not the case for objects with angular faces (as do many man-made objects), or objects that are basically 2-D surfaces (e.g., a leaf). For such objects the surface orientation is discontinuous along the contour generator which corresponds to the outline. Since the line of sight does not graze the surface along the edge, the silhouette boundary is not a tangential contour. Observe that the contours in figure 14*b*, which we interpret as the outline of a gently curved sheet, present a fundamentally different problem than the contours in figure 14*a*. Neither do we assume that the surface is smooth nor that the contours are tangential contours.

The distinction that I propose is therefore not between "outlines" and "inlines" -- not whether the contour is along the boundary of the silhouette or interior to the bounary. Instead, the distinction is between the special case of outline contours, the tangential contours, and all other contours regardless whether they are outlines or lie interior to the object's projection. This means that the outlines of objects that are not smooth will be treated as surface contours for our purposes. The reason for this is the following. The fact that a given contour is part of an object outline does not constrain the shape of the underlying surface, expect when the surface is smooth. Otherwise, the contours merely delimit the visual extent of a object from the given viewpoint. The rest of this section will address the problem of using surface contours. In general, it will not concern us whether the surface contour is a outline contour as well.

1.4 Surface contours: structural and illumination

Thus far, we have only distinguished between tangential contours which correspond to the outlines of smooth objects, and all other contours (those being collectively termed surface contours). But there are various, distinct physical causes of these surface contours. In particular, we can distinguish two broad categories of surface contours, roughly speaking by whether the associated contour generator corresponds to a physical feature on the surface or merely due to illumination. The first category will be termed *structural contours*, the latter, *illumination contours*.

Structural contours are the projections of contour generators which mark some discontinuity on the surface, e.g. of reflectance or of surface orientation. Examples that occur in nature are given by the images of pigmentation markings on a zebra, wrinkles on skin, parallel ridges on leaves, rings on bamboo stalks, and cracks on wood or rock. Images of synthetic objects commonly present structural contours corresponding to seams, sharp edges, groves, and pigmentation markings.

Illumination contours are of three types: (a) the projections of glossy reflections, such as those that appear on metallic or wet surfaces, (b) the projections of shadow edges that have been cast upon a surface, and (c) the images of self-shadows, or "terminators" on surfaces. These three types have been grouped together as illumination contours because their presence is strongly dependent on the particular illumination and may shift their position relative to the surface as the viewpoint or light source geometry changes. They are all potentially useful sources of information about the shape of the surface, as we shall see, but since they depend on particular arrangements of illumination and viewing geometry, they may be considered as fortuitous.

It is noteworthy that we derive such strong 3-D impressions from line drawings. It suggests that we do not restrict the 3-D analysis of surface contours to contours of known physical interpretation. The curves in figure

13 are given strong geometrical interpretations without evidence as to whether they are structural or illumination.

It will therefore be useful to the subsequent discussions to present a few examples of line drawings and to comment on their 3-D interpretations. Later I shall refer back to these figures in order to illustrate particular constraints.

1.5 Examples of 3-D interpretations

Perhaps contrary to intuition, individual line drawn curves may be given stable and definite 3-D interpretations. That is to say, the curve appears to have a definite contour generator fixed in space relative to the viewer. Admitedly, the impression one gains from casual observation of these figures may be weak; if so, view them monocularly with a field-limiting tube to help suppress the fact that the figures are merely drawn on paper. Slant reversals will be disregarded in this discussion since they are expected with orthographic projection.

An ellipse is a familiar example of a simple curve that appears in 3-D. There are actually two interpretations: the curve may be treated as a surface contour whose contour generator is a circle, or the curve may be treated as a tangential contour and the figure is seen as the silhouette of a smooth object (an ellipsoid). We will only consider the case where the curve is interpreted as a surface contour. If an ellipse is deformed, a "potato chip" surface is visualized (figure 15*a*). That is to say, the surface appears singly curved. The following observation is consistent with that interpretation: the dashed lines in figure 15*b*, which connect parallel tangents, appear to lie entirely on the surface.

A few observations may be made about the 3-D interpretations of individual curves in general. First, if the contour is smooth and not self-intersecting (as in figure 16*a*) it tends to appear planar. That is to say, the contour generator is planar. Note that we may confidently judge the spatial orientation of the planes containing the contour generators. (Again, disregard the reversals in apparent slant of those planes.) Our tendency to assume planarity is strong; it is difficult to draw a smooth curve (that is not self-intersecting) which appears to twist in space; it almost invariably appears planar.

Secondly, if the contour has a sharp discontinuity in tangent, as in figure 16*b*, the corresponding corner in 3-D appears to be a right angle. In other words, figure 16*b* appears to be the corner of a sheet of paper.

Finally, if the curve is self-intersecting (figure 16*c*) it is given either of two spatial interpretations. In one interpretation, the contour generator is seen to twist in space so that it does not actually intersect itself. In the other interpretation, the contour generator is self-intersecting, and the intersection is a right angle. In general, we tend to assume that obtuse angles (formed either by discontinuities in tangent or intersections) are foreshortened images of right angles. Figure 17 shows various examples of intersecting straight lines, each of which appears to be a right angle in space. First, note that a simple intersection (figure 17*a*) is quite effective in defining a plane. This effect was observed by Wundt and Herring (see [Luckiesh, 1965; Robinson, 1972]). The parallelograms in figures 17*b* and 17*c* are constructed with the same obtuse angles of intersection and line lengths as the corresponding intersections in figure 17*a*. Their spatial orientations are very similar. (Appendix A examines our perception of surface orientation with these figures.)

Figure 18 demonstrates both tendencies, i.e., for planarity and for right angles. The smooth curve in figure

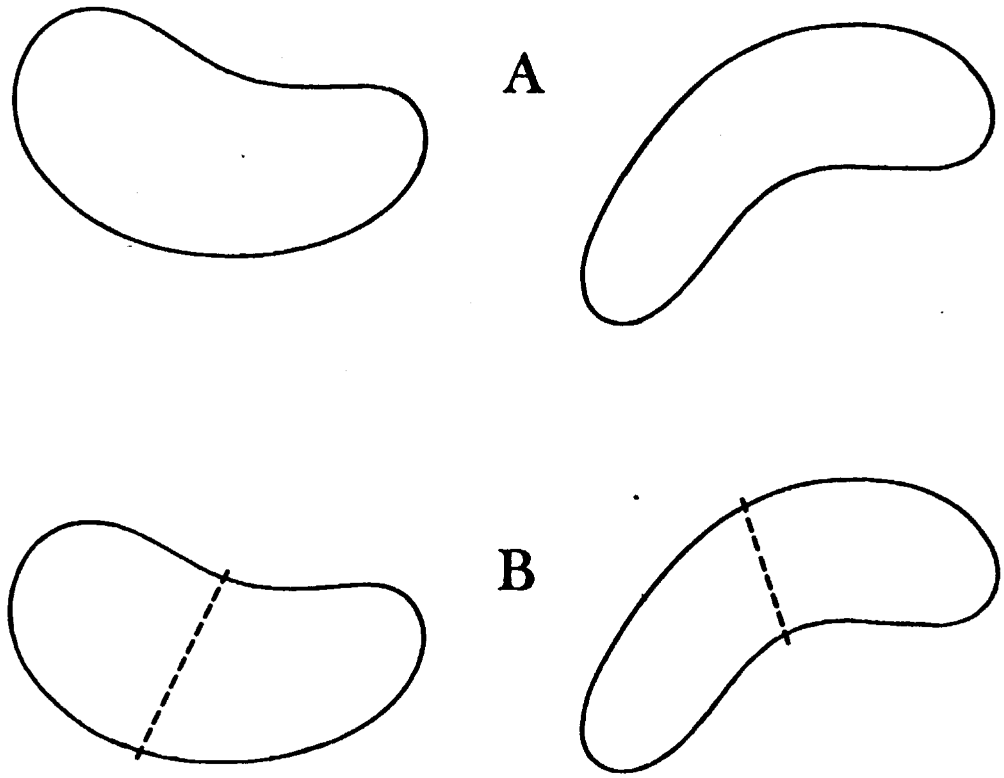


Figure 15. The curves in *a* are seen either as the silhouettes of smooth objects (tangential contour interpretation) or as the image of potato chips (surface contour interpretation). In the latter case, the surface is seen as singly curved, and the dashed lines in *b* appear to lie entirely on the surface.

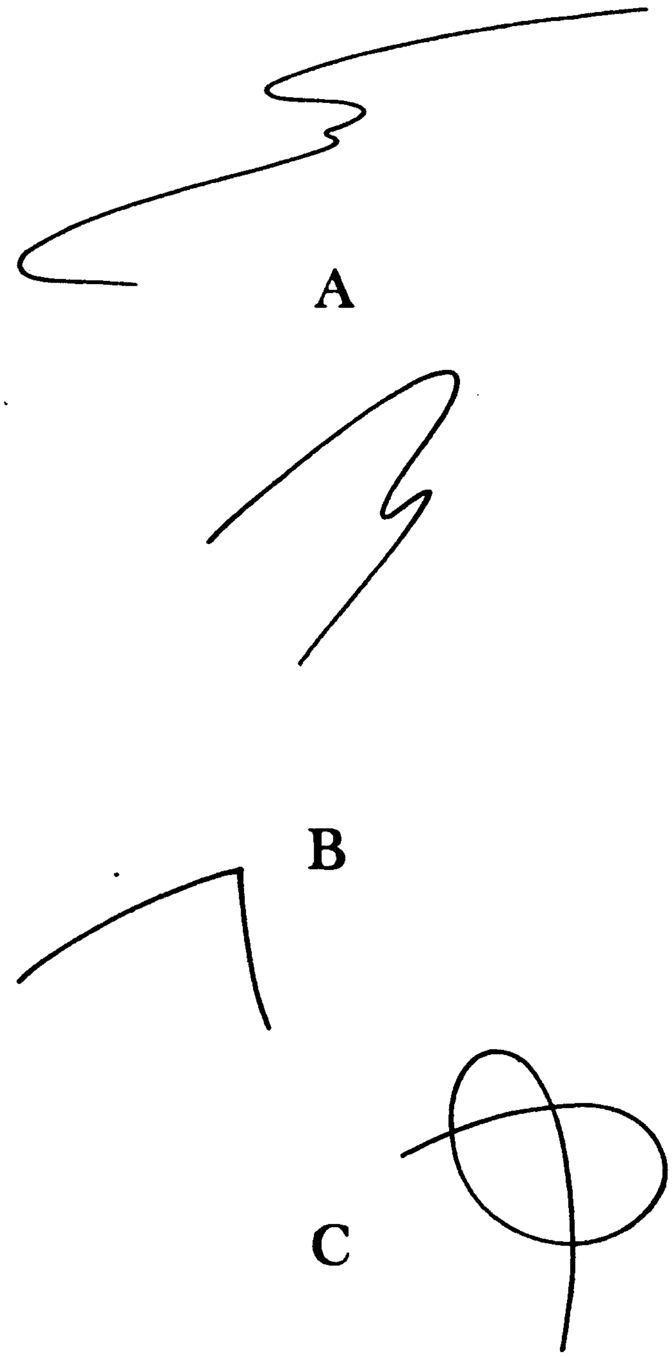


Figure 16. In *a* smooth contours that do not intersect tend to appear planar and to assume definite spatial orientations. In *b* sharp discontinuities in tangent in the contour are interpreted as the images of right angles. The self-intersecting contours in *c* are seen either to twist in space (so that the contour generator does not actually intersect itself) or as the image of a self-intersecting contour generator, where the intersection is a right angle.

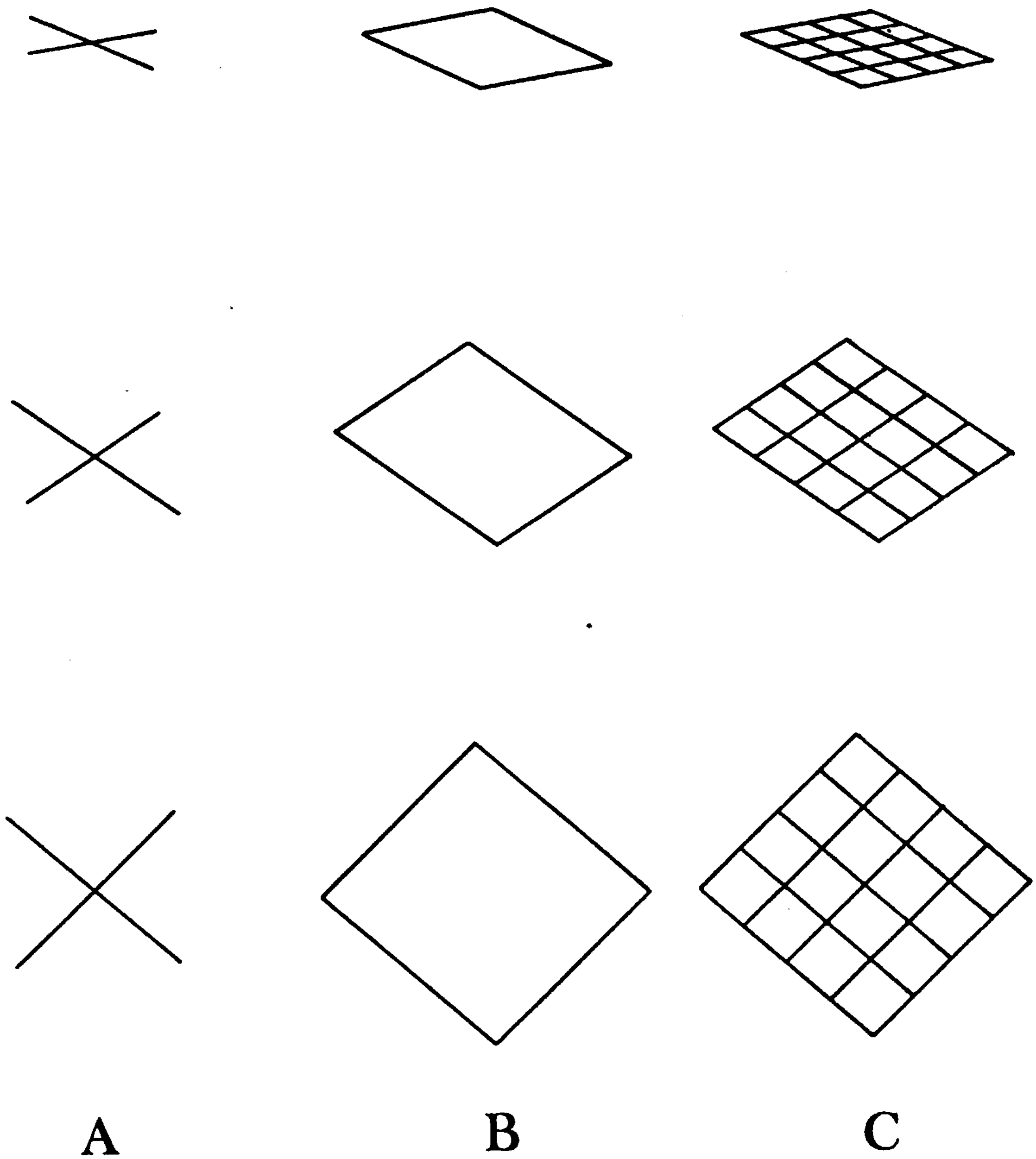


Figure 17. Each intersection in *a* has a definite spatial orientation and appears to be a right angle in 3-D. The spatial orientations in each row of this figure appear very similar. Note that the figures in *b* and *c* are constructed with the same obtuse angles of intersection and line lengths as those in *a*.

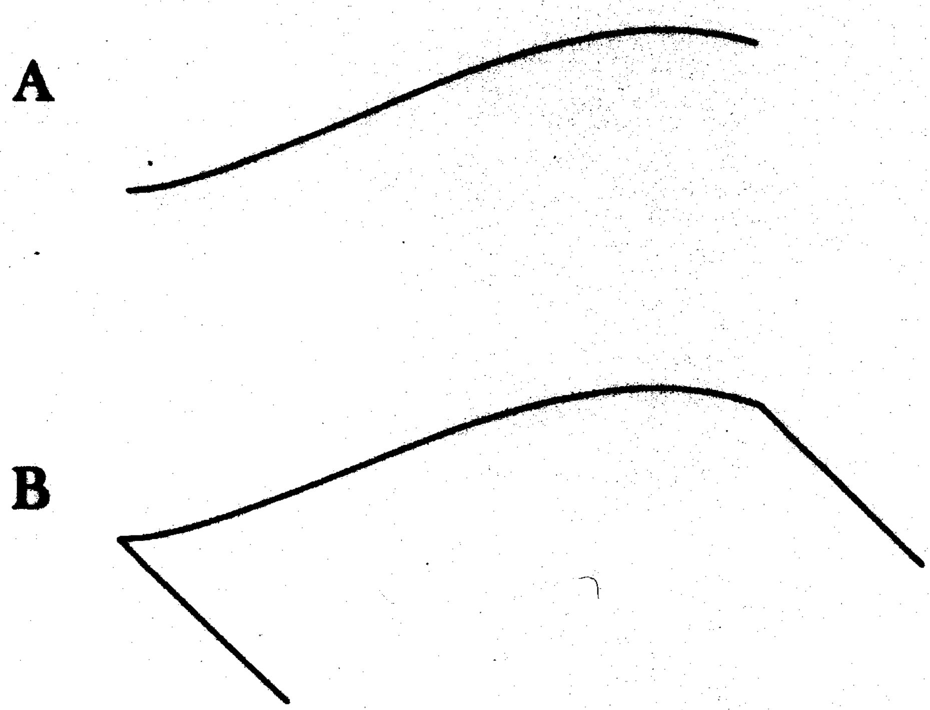


Figure 18. When the curve in *a* is intersected by a few parallel straight line segments, as shown in *b*, a surface like a gently curved piece of paper emerges. Each intersection appears to be a right angle in space, and the curve itself appears planar. As in figure 15*b*, the surface seems to be singly curved.

18*a* presents little 3-D effect. But when the curve is intersected by a few parallel straight line segments (figure 18*b*) a surface like a gently curved piece of paper emerges. Each intersection appears to be a right angle in space, and the curve itself appears planar. As in figure 15*b*, the surface seems to be singly curved, apparently because of the parallelism of the added lines. If those lines are not parallel, two interpretations result. First, one may interpret the figure in perspective, as if the surface were very near the viewer, thus explaining the divergence of the two lines. Secondly, the surface may be seen to twist in space, as a helicoid, i.e., a spiraling piece of paper. It is worth sketching similar curves in order to observe these effects.

Keeping in mind our tendency for planarity and right angle interpretations, let us examine a few more simple configurations of curves. In figure 19*a* the sinusoid does not appear in 3-D, but if a linear component is added ($y = \sin ax + bx$) the curve appears to recede in depth (figure 19*b*). The mouse hole in figure 19*c* also appears in 3-D. These figures are examples of our sensitivity to projections of bilateral symmetry. That is to say, if a surface contour may be given a 3-D interpretation for which the contour generator would be symmetric, that interpretation is taken.

The examples thus far have involved either single curves or simple intersections of curves. In general, multiple curves (treated as surface contours) are not particularly useful in suggesting a surface unless they are parallel, or they comprise a familiar arrangement. (The latter case is not of interest to this study.) An example of parallel contours of which we are seldom aware is provided by *hatchures*, the regular parallel markings used by engravers. Examine the bust of Washington on a dollar bill. The engraver varies the spacing of the hatchures in order to shade the depicted surface, but also, the hatchures follow the surface relief "appropriately". Observe that the undulations in the hatchures suggest surface features such as ridges and depressions. Another instance in which parallel contours suggest a surface is shown in figure 20, a graphical depiction of a function of two variables. A function $z = f(x,y)$ is often displayed by a family of curves produced by holding either x or y constant for various values, and continuously varying the other parameter. These curves are othographically projected (usually from an oblique viewpoint) to present a display of the function surface as if it were intersected by a set of parallel planes.

There are complicating factors in our perception of this figure. Both assumptions of viewpoint and of occlusion are involved, as readily demonstrated by inverting the figure. A paradoxical depth impression may arise by these assumptions being brought into conflict. If the viewpoint is assumed to be such that distance to the surface increases as one scans from bottom to top (as is almost always true in outdoor scenes) then the top of the inverted figure should be farther than the bottom, contrary to that which is indicated by occlusion (the central peak appears occluded by the upper portion, and to occlude the lower portion, thereby implying that the top of the figure is near than the bottom). The paradox may be resolved by imaginging that the top is farther (as if the surface hangs downward from the ceiling) whereupon the figure is seen as consistent in depth.

In addition to the influences of viewpoint assumptions and of occlusion, our interpretation of contours may involve assumptions of perspective. Figure 21*a* appears to be a tunnel in perspective projection, wherein the circles are seemingly taken to be of equal diameter in 3-D. Figure 21*b* has two interpretations, a flattened tunnel (again a perspective interpretation) or a flat disk such as a phonograph record (an orthographic interpretation).

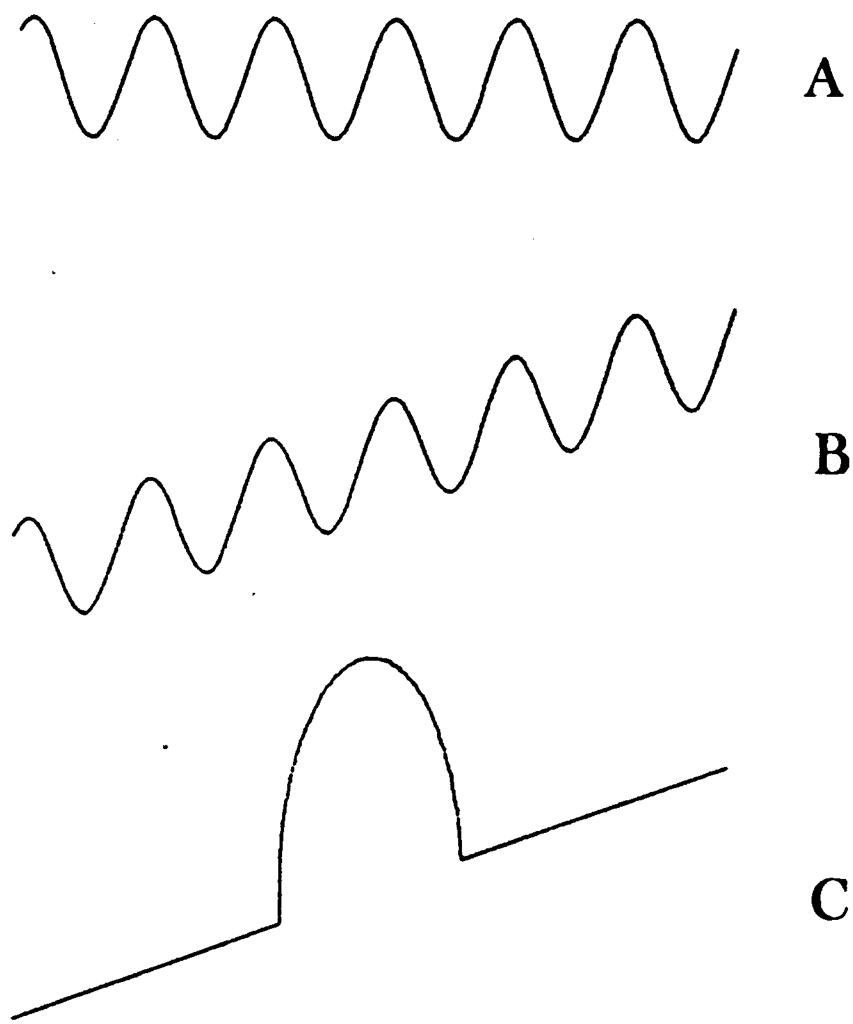


Figure 19. In *a* the sinusoid does not appear in 3-D, but if a linear component is added ($y = \sin ax + bx$) the curve appears to recede in depth, as shown in *b*. The mouse hole in *c* also appears in 3-D. These figures demonstrate our sensitivity to projections of bilateral symmetry.

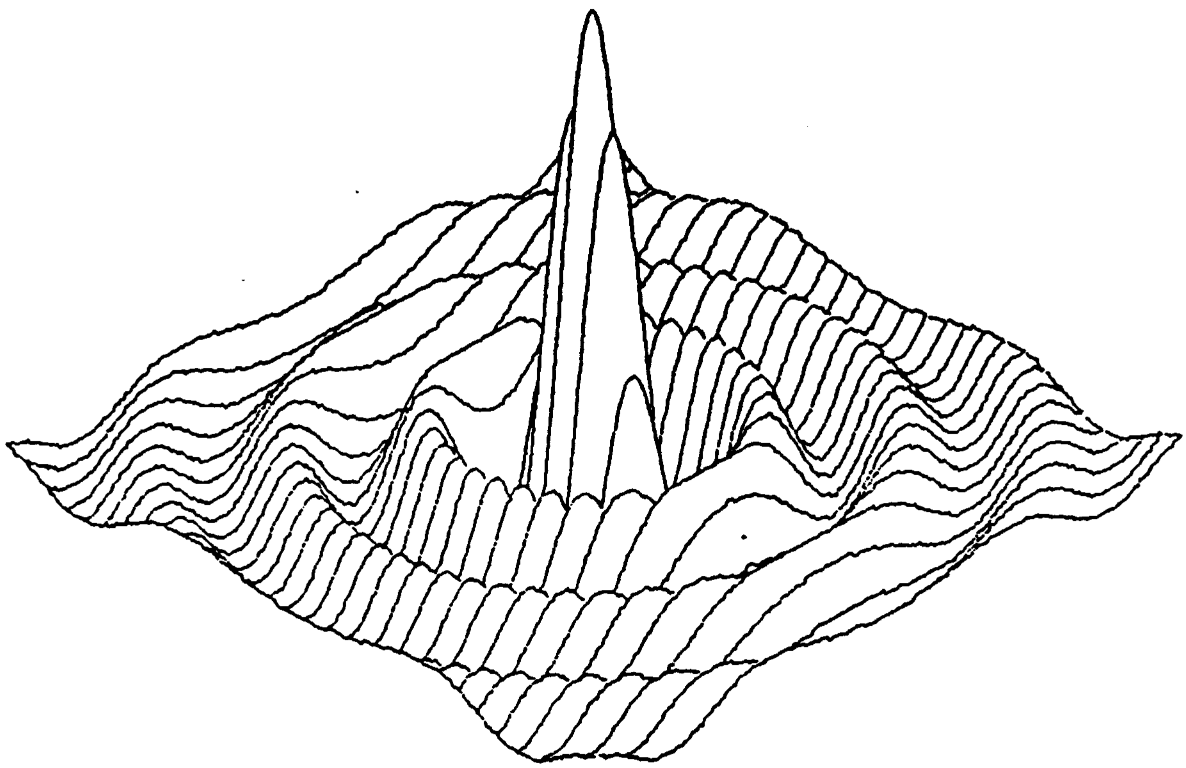
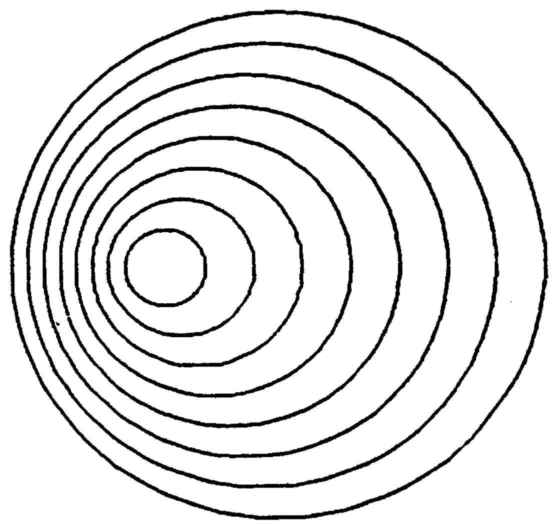
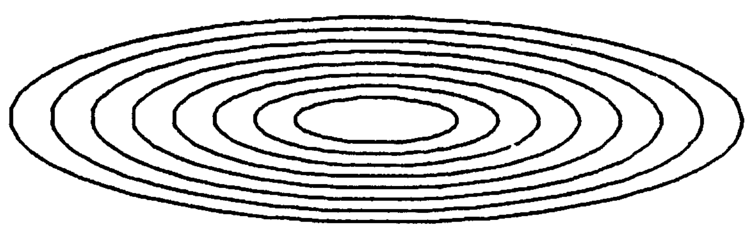


Figure 20. An example of the familiar depiction of a function of two variables $z = f(x,y)$ as the orthographic projection of the curves defined by by holding either x or y constant for various values, and continuously varying the other variable. There are complicating factors in our perception of this figure. Assumptions of viewpoint and of occlusion are involved, as readily demonstrated by inverting the figure. A paradoxical depth impression may arise by these assumptions being brought into conflict. If the viewpoint is assumed to be such that distance to the surface increases as one scans from bottom to top (as is almost always true in outdoor scenes) then the top of the inverted figure should be farther than the bottom, contrary to that which is indicated by occlusion (the central peak appears occluded by the upper portion, and to occlude the lower portion, thereby implying that the top of the figure is near than the bottom). The paradox may be resolved by imagining that the top is farther (as if the surface hangs downward from the ceiling) whereupon the figure is seen as consistent in depth.

Given these examples of our 3-D interpretation of surface contours we now turn to address the problem of constraining their interpretation. First, we will examine a decomposition of the problem into two steps, each of which must be constrained. Constraints for each step are then introduced, and their validity discussed. Discussion of how these constraints are computationally useful is given in section 4.



A



B

Figure 21. In *a*, which we interpret as a tunnel in perspective projection, the circles are apparently assumed to be of equal diameter in 3-D. (A reversal causes the figure to appear as a cone protruding from the page.) In *b* there are two interpretations, a flattened tunnel or a flat disk such as a phonograph record.

2. THE CONSTRAINTS

In the following discussion a surface will be denoted by Σ, a contour generator by Γ, and the projection of Γ from viewpoint V will be the contour Cv (see figure 22). (When the viewpoint is not discussed, the contour will be referred to simply as C.)

A surface contour in the image is the projection[1] of a contour generator Γ lying on a surface Σ; neither the shape of Γ nor Σ is known *a priori.* Note that the surface contour C is completely determined by the 3-D locus of its generator Γ in space relative to the viewer, regardless of the orientation of the surface on which Γ lies so long as the surface allows Γ to be continuously visible along its length. This is an important point. We want to infer the shape of the surface Σ from the shape of the surface contour C, but in fact C is not a function of the shape Σ; C is only a function of Γ. In order to infer the shape of Σ, the relationship between Γ and Σ must be constrained. Likewise, to infer Γ from C, the relationship between Γ and C must be constrained. The decomposition that is suggested, therefore, involves two stages:

(a) inferring the shape of the contour generator in 3-space ($C \Rightarrow \Gamma$) then

(b) determining how the surface lies under the contour generator ($\Gamma \Rightarrow \Sigma$).

This can be thought of as (a) bending a wire in 3-space so that it appears to the viewer as does the contour in the image, then (b) gluing a ribbon along the wire to represent the strip of surface that lies directly under the contour generator. In these terms, we see that infinitely many bendings are possible that would appear identical from the given viewpoint, and the ribbon may twist arbitrarily along the wire. These two aspects of the problem are distinct.

This characterization applies equally to the problem of inferring surface shape from multiple surface contours $\{C_i\}$ in the image, such as those in figure 13. The geometrical arrangement of $\{C_i\}$, particularly if they are parallel, may constrain both stages I and II (section 4.2.2). Note that the appearance of figure 13 may lead one to suspect that parallelism uniquely constrains the surface, but the image is in orthographic projection and significantly different surfaces may project to the same image -- the separation in depth between the contour generators on the surface is not restricted.[2] Thus even in the case of multiple parallel contours, the surface interpretation process must be constrained, and that constraint is naturally described in terms of the above two stages.

This decomposition provides a framework for applying constraints to the problem of inferring Σ from C. The constraints necessary for stage I involve projective geometry, for the problem is naturally one of "deprojecting" from the image curve to the curve in space. The constraints necessary for stage II do not involve projective geometry -- they do not depend on the particular viewpoint. Rather they involve intrinsic

1. The projection is assumed orthographic, i.e., the contour generator is assumed small compared to its viewing distance. The perspective distortions otherwise induced in its projection would be infeasible to differentiate from those induced by slight twisting along its length. Note further that the informal term "image plane" will be used, although the retinal projection is more closely approximated by spherical projection.

2. In fact, one consistent surface solution is given immediately by the sheet of paper on which figure 13 is printed -- the parallel contour generators would be the ink on the page.

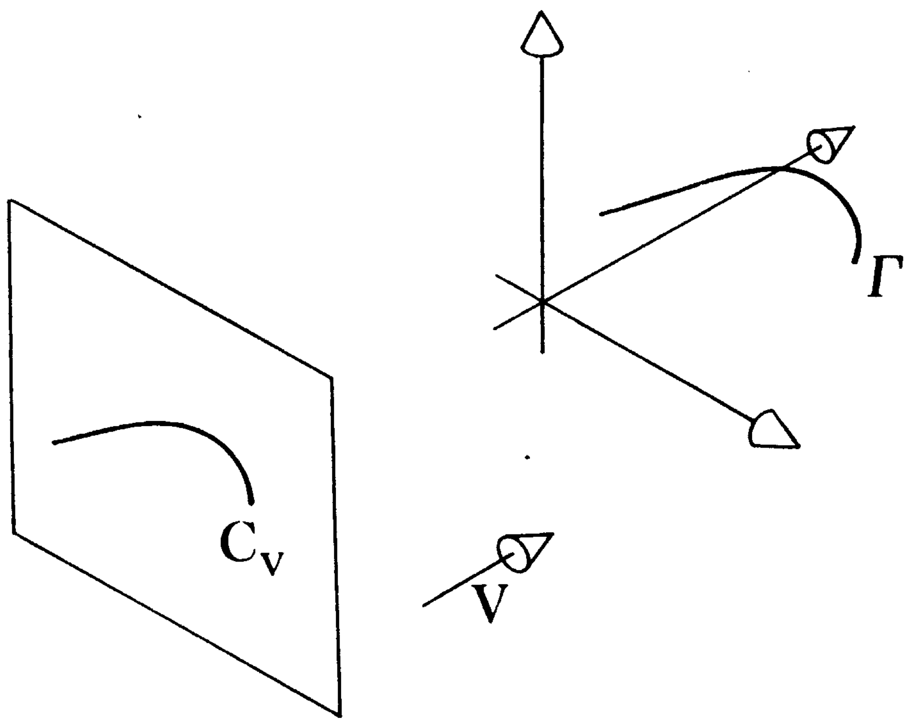


Figure 22. The orthographic projection of *contour generator* Γ from viewpoint V is C_v. The curve C_v is termed an *occluding contour* if it is an edge of the silhouette of an object from viewpoint V. In particular, if the line of sight just grazes the surface along Γ then the curve C_v is also a *tangential contour*. The image curve C_v is termed a *surface contour* if it is not a tangential contour.

geometry, specifically the relationship between the curve on the surface and the surface itself.

2.1 Some geometrical concepts

This section reviews some concepts that are necessary for discussing the relation between a curve on a surface and the underlying surface itself. I shall review the notions of Gaussian curvature, lines of curvature, developable surfaces and cylinders, asymptotic curves, and geodesics (c.f. [Hilbert & Cohn-Vossen, 1952]).

To introduce Gaussian curvature, consider the family of normal sections at some point of a smooth surface, i.e., the contours that result from sections that contain the surface normal at that point. The various section contours through that point usually vary in curvature, with greatest and least curvature occurring at two principal directions (except when the curvature is constant for all directions, as with a sphere). An important property of the two principal directions is that they are mutually orthogonal at every point on the smooth surface.

The *Gaussian curvature* at a point is the product of the greatest and least curvatures. The Gaussian curvature may be positive, negative, or zero, and for an arbitrary surface may vary continuously across the surface. For example, the curvature is positive on a smooth pebble, negative on a saddle surface, and zero on a cylinder (defined momentarily).

A *line of greatest* (or *least*) *curvature* is a curve whose tangent everywhere coincides with one of the two principal directions. Important examples are the cross sections and meridians of surfaces of revolution (which of these is the line of greatest curvature depends on the surface shape).

A *developable surface* is a surface with zero Gaussian curvature everywhere (i.e., the curvature in at least one of the principal directions vanishes). Thus the lines of least curvature are straight lines on a developable surface. Examples of developable surfaces are planes, cylinders, and helicoids. Informally, they correspond to the class of surfaces that may be made by twisting and curling a sheet of paper.

A *cylinder* is a developable surface where the lines of least curvature are parallel. Cylinders may be formed by curling a sheet without torsion -- it may be rolled into a tube or be rippled like a hanging curtain. It is useful to think of a cylinder as a one-dimensional surface.

An *asymptotic curve* is a locus of points on the surface where the Gaussian curvature is zero. By definition, all curves on developable surfaces are asymptotic. On the other hand, surfaces with everywhere positive Gaussian curvature (such as a sphere) have no asymptotic curves. And surfaces of negative Gaussian curvature must have asymptotic curves, since the principle curvatures are of opposite sign and for some direction between the principle directions at each point on the surface the curvature must vanish.

Finally, a *geodesic*, usually defined as the shortest path between two points on a surface, is also a curve whose principal normal[1] everywhere coincides with the surface normal. Importantly, the lines of greatest and least curvature on a cylinder are geodesics.

1. The principal normal to a planar curve is the perpendicular to the tangent to the curve and lies in the plane of the curve. The principal normal to a curve with torsion, similarly, is perpendicular to the tangent but lies in the osculating plane of the curve at that point (where the osculating plane is defined by two successive tangents at the given point). Note that we will often restict curves to be planar, so fixing the plane of a geodesic immediately fixes the normal to the surface.

2.2 What constraints might be useful?

We now introduce some constraints that allow solutions to steps I and II. They are provided by restricting the geometrical properties of the contour generators, and restricting the relationship between the contour generators and the surface on which they lie. This section only tabulates the various geometric restrictions. Next, in section 3 we will discuss the validity of assuming that these restrictions hold in natural situations involving actual contour generators on physical surfaces and, in section 4, we will describe how the restrictions constrain the shape-from-contour analysis.

2.2.1 Constraints on the contour generator

With regard to step I, the 3-D shape of a contour generator Γ (corresponding to a given surface contour C) may be recovered if restrictions are imposed on Γ and on the viewing position. Some of these restrictions are listed below.

> (a) *general position*, the viewpoint is not misleading. This allows one to infer properties of the contour generator Γ on the basis of the properties of its image, the surface contour C. For instance, if C is smooth then Γ is smooth; if $\{C_i\}$ are parallel then $\{\Gamma_i\}$ are parallel.
>
> (b) *planarity*, Γ is planar. This reduces the problem of determining Γ to that of determining the orientation of the plane Π containing Γ. The plane Π is constrained by the following.
>
> (c) *symmetry*. Given planarity and general position, if C presents evidence of symmetry then Γ is symmetric, and the orientation of Π must be consistent with Γ being symmetric.
>
> (d) *minimum curvature variation*. Given planarity and general position, if the curvature of Γ is roughly constant then the variations in curvature apparent in C may be attributed to foreshortening. Consequently that plane Π that minimizes the variation in curvature of Γ would solve Γ.

2.2.2 Constraints on the relation between contour generator and surface

Given the contour generator Γ, the surface Σ may be solved if the relationship between Γ and Σ is restricted. If Γ is planar and lies on some plane Π then the relationship between the contour generator and the surface is naturally described in terms of the angle between Π and the tangent plane to Σ for points along Γ. The relation between the surface and the contour generator is quite simple if we make the strong restriction that this angle is constant along the length of Γ. That is to say, the plane containing the contour generator meets the surface at a constant angle. The two cases we will consider is when the angle is $\pi/2$ and zero.

If the angle between Π and the tangent plane to Σ is $\pi/2$, then:

> Γ is *geodesic*. The surface normal coincides with the principal normal to Γ for points along Γ.

If the angle between Π and the tangent plane to Σ is zero, then:

> Γ is *asymptotic.* The surface normal coincides with the normal to Π for points along Γ, and furthermore, the Gaussian curvature of Σ for points along Γ is zero.

These two solutions, geodesic and asymptotic, form the basis for constraining the relation between the contour generator and the surface. Given general position and planarity, we also have an important restriction on Σ in the case of parallel surface contours $\{C_i\}$:

> $\{\Gamma_i\}$ are *parallel lines of curvature* and Σ is a *cylinder.* Furthermore, if the contour generators are geodesics, they are lines of greatest curvature; if asymptotics, the surface degenerates to be planar.

And finally, a derivative of the cylinder restriction may apply in the case of a single surface contour, if the corresponding contour generator is a line of greatest curvature and the surface is cylindrical, by the following restriction:

> Σ is *opaque.* The image of an individual line of greatest curvature on a cylinder allows some restriction on the shape of the surface.

Surface contours are often weak sources of information about the surface shape when analyzed individually, primarily because it is difficult to deduce the shape of the contour generators on an individual basis. The more important case probably involves the geodesic restriction on a collection of parallel contours taken together. Then the parallelism may be used to advantage in constraining the shape of both the contour generators and the surface on which they lie. Before pursuing the utility of these constraints any further, it is important to gain some insight into their validity.

3. WHEN ARE THE CONSTRAINTS VALID?

Do the contour generators in the real world meet these restrictions? In some situations it is valid to assume that a contour generator is, say, planar and geodesic, as we shall see. But there are also instances where the same assumptions are not valid -- the real world does not necessarily constrain the curves on surfaces to comply with any of the various ideal geometries. How often are the restrictions met in actuality? This is the issue of "ecological validity" discussed by Gibson, Brunswick, and others (c.f. [Gibson, 1950; Postman & Tolman, 1959]). We start with considering the validity of assuming general position.

3.1 General position

General position implies that the viewpoint is representative -- that the image taken from this position does not mislead us by accidental alignments. Two examples of viewpoints that are not general position may be imagined for a cube: In one instance the cube is positioned so that its silhouette is a regular hexigon. Equally misleading would be a cube positioned so that its silhouette is a perfect square.

When the assumption of general position is correct we may make valid deductions, in particular, deductions about contour generators. Two examples of these deductions which we shall pursue are the following: If a surface contour is smooth, the corresponding contour generator is smooth, and if surface contours are parallel, their contour generators are also parallel.

The contour generator need not be smooth simply because its projection is smooth: a discontinuity in tangent along a contour generator might be hidden from the given viewpoint -- the plane containing the discontinuity might also contain the line of sight so that the discontinuity would not be apparent. But if the distribution of spatial orientations of planes relative to the viewer is uniform, the likelihood of such an accidental alignment would be insignificant. Similarly, some non-parallel curves may be constructed such that they appear parallel from certain viewpoints, but the probability of achieving a viewing position that allows this alignment becomes insignificant as the curves diverge from parallelism in 3-space. [1]

3.2 Geometrical properties of structural contours

In general, the geometry of structural contours is not strongly constrained because the processes that cause them are varied and often random. There are, however, some types of physical markings that are well constrained.

The clearest examples, perhaps, involve synthetic objects. With reference to the objects about you, observe that the smooth surfaces of man-made objects are usually comprised of either (a) planar surfaces, (b) singly curved surfaces, in particular cylinders, or (c) surfaces of revolution. In general, the boundaries between surfaces are planar, primarily for reasons of fabrication. Again, because of convenience in manufacturing as well as utility, curved surfaces are usually sliced by normal sections. Thus joints between surfaces of an object

1. Implicit in the above argument is the reasonable expectation that the instances of actual parallelism, straightness, and so forth, are more probable than accidental alignments.

comprise geodesics on one or the other of the joining surfaces. The end of a "tin can" would be an example. Surface markings other than seams or joints are often geodesics as well, particular when the markings are on cylinders. When the markings are also planar, they additionally constitute lines of curvature. This combination of properties, planarity and geodesic, is particularly common.

Markings on surfaces of revolution usually follow either the axis or some cross section. Hence these seams, edges, ridges, and pigmentation markings are lines of curvature, geodesic, and planar. (A notable exception can be found in the spiral seams on cardboard tubes. They are geodesic but nonplanar.)

Flexible surfaces, both natural and synthetic, tend to be noncompressible hence developable, and are therefore cylinders when not subjected to torsion. Wrinkles produced by compression tend to be lines of curvature.

Many biological forms may be approximated as being composed of generalized cones [Marr, 1977a]. These surfaces often have markings that follow cross sections and meridians on the surface, and therefore are also lines of curvature, geodesic, and planar. Biological objects are often bilaterally symmetric, such as leaves. Their axes of symmetry are often evidenced by physical markings, and symmetric patterns are usually arranged across that axis. The symmetry may be used to advantage to restrict the possible orientations that would be consistent with the 3-D form being symmetric.

3.3 Geometrical properties of illumination contours

3.3.1 Cast shadows

The edge of a shadow cast across a surface is a fortuitous source of information about surface shape. We are familiar with the effectiveness of the shadow a fence post cast upon snow in indicating the undulations in the surface. But to accurately analyze the surface from the image of the cast shadow, a number of variables must be known. There are essentially two projections involved: the projection of the shadow onto the surface (the edge of which becomes the contour generator Γ) and the subsequent projection of Γ onto the image plane (as contour C). Thus the contour C in the image depends on (a) the shape of the physical shadow-casting edge, (b) the position of the light source -- together they specify the bundle of rays that will be cast upon the surface -- and (c) the position of the shadow-casting edge relative to the surface, and finally (d) the shape of the surface itself.

To appreciate the complexity of shadow interpretation in the general case, consider again the image of a tree trunk shadow cast on snow. Suppose there is a kink along the shadow edge. Is that due to a sharp depression in the snow (for instance, is the shadow falling across a footprint) or is it due to a kink in the tree (and the snow itself is flat)? If analyzing the shape of the surface is attempted prior to knowing the above factors, some assumptions are necessary. In the approach suggested here, the assumptions are two:

the contour generator is *planar* and *geodesic*.

In terms of this example, the above translate into assuming the edge casting the shadow is straight and that its profile (determined by the sun position and the trunk) intersects the ground at a right angle. Then if there is an apparent kink in the shadow edge it will be attributed to the surface, not to the tree. (Incidentally, it is informative to observe the shadow cast on the flat ground by a young tree which has a crooked trunk. The ground often appears to undulate according to the curves in the cast shadow.)

So we should discuss how the planarity and geodesic restrictions help the shape analysis. First note that if the shadow-casting edge is straight the contour generator (the shadow edge cast across the surface) constitutes a planar section of that surface. That is, the contour generator lies in the plane defined by the straight shadow-casting edge and the point light source. In this case, we may already determine qualitative information about the surface shape. Given general position, if the contour in the image corresponding to the shadow edge is straight, the surface is flat; if it is curved, the surface is curved. To determine more quantitative shape information requires that (a) the relation between the contour generator Γ and the surface be known, and (b) the orientation of the plane of Γ be known. Hence we introduce the geodesic assumption. That is to say, the shadow edge across the surface is assumed to be a normal section of the surface. Weak justification for this assumption derives from considering shadows cast on the ground: Since shadow-casting edges are usually vertical (e.g., tree trunks, building edges, telephone poles, fences), the edge of the shadow amounts to a normal section, i.e., the shadow edge is roughly geodesic.

When do multiple, parallel sections occur in real situations? We may disregard the shadow of a picket fence as being artificial, but notice that two parallel sections would result from the shadow edges cast on some surface by a relatively narrow object such as a tree trunk. Another possibility concerns motion: successive views of a moving shadow edge. Successive positions of a shadow edge that sweeps across a surface in translatory motion would constitute parallel sections of the surface. Does the visual system take advantage of this fact? Is our ability to analyze parallel surface contours a derivative of an ability to analyze moving shadows? This hypothesis would be supported if we could perceive a surface defined only by a single moving contour that scans across an otherwise invisible surface. In fact, this ability may be demonstrated by a motion sequence of a single contour on a CRT, where each frame presents only a single curve. Note that the moving curve might be interpreted simply as a flexible wire that bends as it translates, or more literally, as a curve in the plane of the screen that changes shape as it moves. But, in fact, there are instances when we interpret the moving contour as a shadow edge sweeping across a 3-D surface (e.g., when the individual curves in figure 13 are presented in succession).

3.3.2 Specular reflections: gloss contours and highlights

Gloss contours, like shadows, are fortuitous, i.e., useful but not necessarily present. They are present only under directional lighting conditions on specular surfaces, when the surface normal lies in the plane defined by the point light source, surface point, and viewer and bisects the angle defined by that configuration. This configuration (the *specularity condition*) is rarely met with planar surfaces but is commonplace for curved surfaces, especially when viewed indoors with multiple lights illuminating the surface. The specularity condition may be met only at an isolated point, causing a *highlight*, or met along a curve, causing a *gloss contour.*

For a doubly curved patch of surface the specularity condition is met at only a point, if at all, and would only produce a highlight in the image. A gloss contour cannot occur on a surface with nonzero Gaussian curvature in orthographic projection given a point light source.[1] For a gloss contour to occur -- for the specularity to appear not as a point but as a curve -- the specularity condition must be met along a continuous curve on the surface. With orthographic projection and distant light source it is necessary that the contour generator (the locus along which the specularity condition is met) be planar. That plane corresponds to the tangent plane to the surface along the contour generator. Now two results in differential geometry are useful:

> A curve is asymptotic if it lies in a plane everywhere tangent to the surface along the curve.
>
> If the angle between a planar curve and the tangent plane of the surface is constant, then that curve is a line of curvature.

Using the above, we may conclude that the curve across the surface that corresponds to the gloss contour is asymptotic and a line of (least) curvature. Since the asymptotic curve follows a path of zero Gaussian curvature, we have information about the intrinsic geometry in the vicinity. Of importance is the following:

> If the gloss contour is curved, the surface is *planar*. This is true in orthographic projection with distant light source. (With nearby objects and perhaps nearby illumination, the surface would not be strictly planar. But in general the surface curvature measured along the contour generator will be small, much less than that measured across the contour generator.)
>
> If the gloss contour is straight, the surface is *cylindrical* when either (a) gloss contours from successive viewpoints are parallel, or (b) if there are multiple light sources (as is common in interior scenes) and multiple gloss contours are parallel.

These deductions hold subject to general position, of course.

Thus the specular reflections in the image can tell us not only something of the reflectance properties of the surface, that the surface is specular [Beck, 1972], but also something about the surface shape, namely, that the Gaussian curvature is nonzero in the vicinity of a highlight and zero in the vicinity of a gloss contour. The shape of the gloss contour also specifies the intrinsic shape of the developable surface.[2] This does not strictly hold when the surfaces or light sources are near by, and especially when the light comes from an extended, rather than a point, source. Nonetheless, it is instructive to observe the gloss contours on specular surfaces -- they almost invariably follow the least curvature paths on actual surfaces.

3.3.3 Shading contours and terminators

The previous discussion assumes bright, directional light sources. However the specular surface not only reflects the light sources as a highlight or gloss contour, but also acts as a mirror -- the various glossy

---

1. In real situations we have two ways in which gloss contours may arise. First, extended light sources (such as fluorescent lights, bright windows) will extend point reflections into images of the light sources, which appear as gloss contours if compressed because the two principle curvatures are very different. Secondly, in perspective projection we may have that as the line of sight sweeps across the surface (the projection is not parallel) the angle between the line of sight and the surface stays relatively constant due to curvature of the surface, such as when viewing the inside surface of a cup from nearby. Then if the specularity condition is met at one point in that vicinity, it would be met along a locus. Thus in perspective projection highlights may spread into gloss contours as well.

2. Furthermore, the surface normal coincides with the normal to the plane containing the gloss contour, but to utilize that fact the 3-D curve corresponding to the gloss contour must be determined. That is the topic of section 4.1.

reflections comprise an image of the surrounds distorted by the geometry of the surface. This is the extreme case of mutual illumination which makes "shape from shading" difficult. The incident illumination is an intractably complex function of the surrounds. But without understanding this illumination, the shape of the surface cannot be solved from the shading.

With the addition of a matte component, the fine details in the reflections are lost, and the gloss contours become less definite. In the limit case of a Lambertian surface there is no specular component and the shading is only a function of the surface orientation relative to the various sources of illumination. For this reason one would expect that the surface orientation would be computed from shading most feasibly, however the illumination is still determined by the surrounds and is still quite unconstrained. Consequently, the computation of shape from shading (where "shape" means local surface orientation) is quite difficult.

Most surfaces are neither totally matte nor glossy so their images present weak highlights and gloss contours -- the distinction between shading and gloss becomes vague. One may postulate, therefore, that shading only constrains the local surface geometry in the manner just described -- the local surface orientation is not computed directly from the shading. Instead, the local surface orientation would be smoothly interpolated between those tangential contours and surface contours along which surface orientation can be solved. The interpolation would be subject to the constraint on intrinsic surface geometry provided by the gloss and shading contours. This constraint is naturally described in terms of Gaussian curvature: A highlight indicates positive Gaussian curvature in the vicinity. Similarly, a gloss contour indicates a locus of zero Gaussian curvature.

Constraint on intrinsic geometry is also provided by the shading contours known as *terminators*, surface contours which correspond to paths on the surface along which the light grazes the surface so that points on one side of the contour are illuminated, points on the other side are in shadow. (A terminator is analogous to a tangential contour seen from the light source position.) A strong restriction on the surface shape is provided wherever the terminator is straight in the image: the surface is locally developable (again, assuming general position) and therefore the terminator indicates a locus of zero Gaussian curvature.

4. HOW THE CONSTRAINTS ARE USEFUL

Thus far we have discussed a number of geometrical properties that may be useful in constraining the analysis of shape from surface contours. Instances in which these properties hold in real scenes were described. What remains is to become more specific about why these properties are computationally useful.

4.1 The relation between a surface contour and its contour generator

The current problem is to determine the contour generator Γ in 3-space on the basis of its projection, the surface contour C. The projection will be restricted to be orthographic. This restriction would hold whenever the dimensions of the curve in space are small relative to the distance from the curve to the viewer. Orthographic projection is linear, hence some useful geometrical properties are preserved, notably parallelism.

Now, in determining the shape of contour generators in 3-space we are confronted with a problem wherever the tangent to the contour (its slope) is discontinuous: Is that discontinuity the projection of a discontinuity in tangent along the contour generator, or is the discontinuity due to the adjoining of distinct contour generators on the surface? Since this cannot be answered locally without *a priori* knowledge of the specific surface, we follow the principle of least commitment [Marr, 1977a] and partition the surface contours in an image into their smooth segments.

4.1.1 General position

A number of constraints will be consequences of assuming general position -- that the viewpoint is such that images from nearby viewpoints would not present significant differences in the geometry of the projected contours. By this we rule out viewpoints that cause accidental alignments which mislead. For instance, if a contour C is straight from viewpoint V, then assuming general position, it would be straight from a similar viewpoint -- it is not the case that the contour generator Γ is curved in a plane but that plane is viewed "edge on" so that the image of Γ is foreshortened into a straight line. General position allows one to infer properties of contour generators on the basis of their images, such as smoothness, continuity, and parallelism.

Our first application of general position is as follows. Since the contour C is smooth and continuous, Γ is smooth and continuous.[1] Furthermore, in general position, nearby and distinct points on Γ project to nearby and distinct point on C. That is, there are no kinks or loops in Γ hidden by the particular viewpoint. In short, assuming general position allows us to consider Γ as a smooth wire in 3-space. Now we consider additional constraints which allow us to determine its shape.

4.1.2 The planarity restriction

If the contour generator Γ is constrained to be planar, the shape of Γ would be completely determined by the equation of the plane containing the curve given its orthographic projection C. Hence the planarity

1. We would like to say something about the smoothness of the surface directly under the contour generator on the basis of the surface contour being smooth, but unfortunately that does not follow from general position as stated. The smooth contour generator may lie along a sharp ridge, for instance.

restriction reduces the problem of determining Γ to that of finding the spatial orientation of the plane Π containing Γ.

Since the contour generator Γ is determined once Π is specified, one approach is to impose an *a priori* choice of Π, then examine the shape of Γ that results. That is, one assumes a particular spatial orientation for the plane containing the contour generator. But there do not appear to be any reasonable choices for Π, except for the ground plane, i.e., the horizontal plane defined by gravity. However it is not feasible to assume that all surface contours are projections of horizontal contour generators.

Alternatively, one may make *a priori* assumptions about the shape of Γ in the same spirit as assuming that Γ is planar. Then Π would be a consequence of C and those restrictions on Γ. What restrictions can be reasonably placed on Γ, and how are those restrictions to be phrased? I shall consider two -- symmetry and minimum curvature variation.

4.1.3 Symmetry

Bilateral symmetry is commonly found in nature and usually preserved, at least indirectly, in orthographic projection. We are interested in symmetry, for evidence of symmetry in an image will provide constraint on the shape of Γ. We start with the usual definition of a bilaterally symmetric, planar curve as comprising two loci of points that are reflections of each other across a straight line, the axis of symmetry (figure 23*a*). The symmetric points are equidistant across the axis, the line connecting any two symmetric points is perpendicular to the axis, and all such lines are therefore parallel.

In any orthographic projection of this curve, the image of symmetric points are equidistant across the image of the axis, the *correspondence lines* connecting those points are parallel, but the correspondence lines are no longer perpendicular to the image of the axis in general (figure 23*b*). This configuration has been aptly termed "skewed symmetry" by Kanade and Kender [1979]. If a unique line can be found that behaves, in this sense, as the image of an axis of symmetry, then by general position we will assume that the planar curve in space is bilaterally symmetric. (Refer back to figure 19.) That is, we have criteria for detecting bilateral symmetry. When these criteria are satisfied in an image we may assume that it is not coincidental, that it would also be satisfied in an image taken from a different viewpoint -- hence due to actual symmetry. The problem that remains is to detect the images of symmetric pairs of points.

Orthographic projection is linear, hence a number of properties are preserved by the transformation including midpoints, points of inflection, and convexity and concavity [Marr, 1977a]. Marr has shown, in the context of finding the axes of generalized cones, that axial symmetry can be efficiently detected by the *qualitative symmetry* between convex and concave segments, rather than on a point-by-point basis. This extends to the detection of bilateral symmetry, where the correspondence lines between qualitatively symmetric segments would be parallel. The line defined by the midpoints of the correspondence lines would be the image of the axis of symmetry.

Returning to the problem of constraining the shape of the contour generator, the symmetry detected in C constrains Γ to be symmetric and this in turn constrains the orientation of the plane Π containing Γ. Specifically, Π must be oriented relative to the viewer such that, given C, Γ would be symmetric if lying on Π.

This constraint is simply expressed in terms of the *correspondence angle*, the angle in the image between the correspondence line and the projected axis of symmetry (figure 23*b*). Since the correspondence angle is

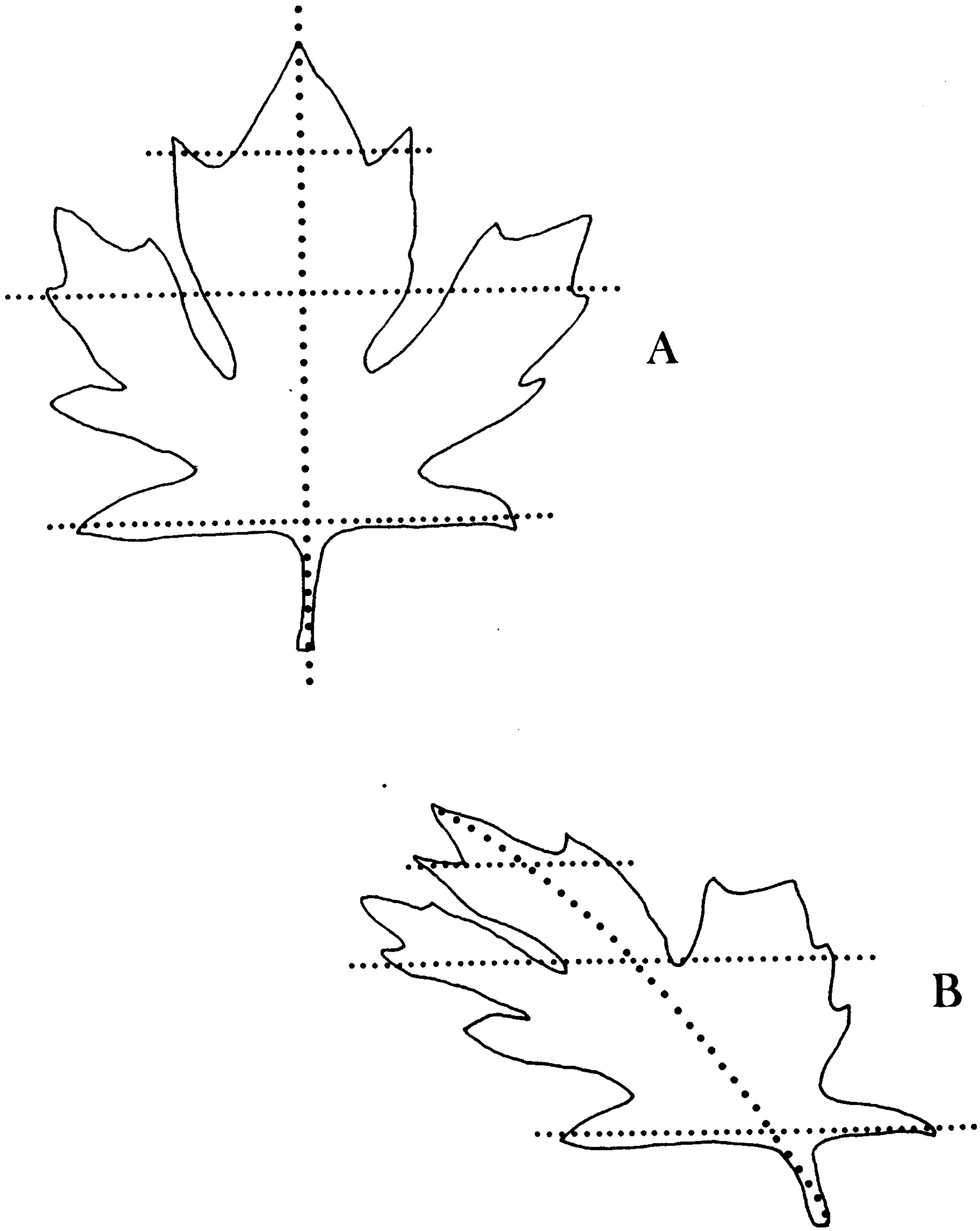


Figure 23. The bilateral symmetry in *a* can be described in terms of correspondence lines which connect symmetric points lying equidistant from a straight line, the axis of symmetry. The parallel correspondence lines are perpendicular to the axis of symmetry. In *b* the correspondence lines connecting qualitatively symmetric segments of the curve are also parallel but make an oblique angle β with the axis of symmetry.

the image of a right angle on the surface, the magnitude of the correspondence angle β constrains the possible spatial orientations for the tangent plane at that point (see figure 24).

In short, Γ is presumed symmetric if an axis of symmetry can be reconstructed from the midpoints of parallel correspondence lines, where the correspondence lines are constructed between qualitatively symmetric segments of C. The correspondence angle then constrains the spatial orientation of the plane containing Γ.

4.1.4 Minimum curvature variation

The curvature of C encodes information about the orientation in space of the contour generator Γ, if Γ is planar and some other restrictions hold. Witkin [1979] has shown that the orientation of the plane Π containing Γ may be estimated on the basis of the curvature along C if we assume that systematic variations in the curvature that resemble foreshortening are due to foreshortening. Then one may choose that plane Π that maximally accounts for the variation in curvature in terms of foreshortening. The following assumptions are sufficient to allow this analysis:

(a) the possible surface orientations of Π are equally likely,

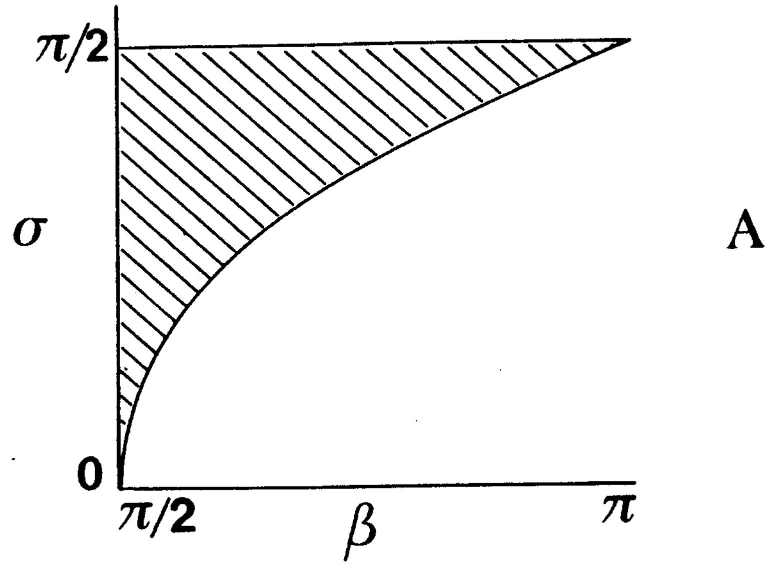
(b) the tangents to the contour generator are arbitrarily aligned relative to the viewer (they are independent of slant σ and tilt τ), and

(c) the curvature along the contour generator is independent of σ, τ, and the orientation relative to the viewer of the tangent to the contour generator Γ.

The constraint on Γ that results is roughly equivalent to assuming that the variation in curvature along Γ is minimum [Witkin, 1979]. Then the variation in curvature along its projection C may be attributed primarily to foreshortening, whereupon the degree of foreshortening -- hence the orientation of the plane Π containing Γ -- may be estimated. To introduce this, consider the case when Γ is a circle, a planar curve with constant curvature. The orthographic projection C is an ellipse; the curvature along the ellipse varies according to the foreshortening of the corresponding segment of the circle. One may derive from the variance in curvature an estimate of the orientation of the plane containing Γ.

This constraint has been phrased in terms of minimum curvature variation, but Witkin describes it more generally as a problem of signal detection. The "waveform" that we consider is the contour in the image (parameterized in terms of contour curvature). The curvature at any point on the contour consists of two components, one being the curvature of the contour generator at each corresponding point, the other being a "projective component" which increases or decreases the apparent curvature according to the orientation of the given segment of the contour generator relative to the viewer (in the circle example, where the tangent lies parallel to the image plane, the curvature on the ellipse is minimum; where the tangent to the circle is oriented away from the viewer the curvature is greatest). The curvature of the contour generator is treated as noise; the projective component is the signal. Since the projection is orthographic and the contour generator is planar, the projective component will be regular.

The problem of determining the orientation of the plane containing Γ may be recast as that of estimating the amplitude and phase of a signal of known waveform (the projective component) in the presence of noise



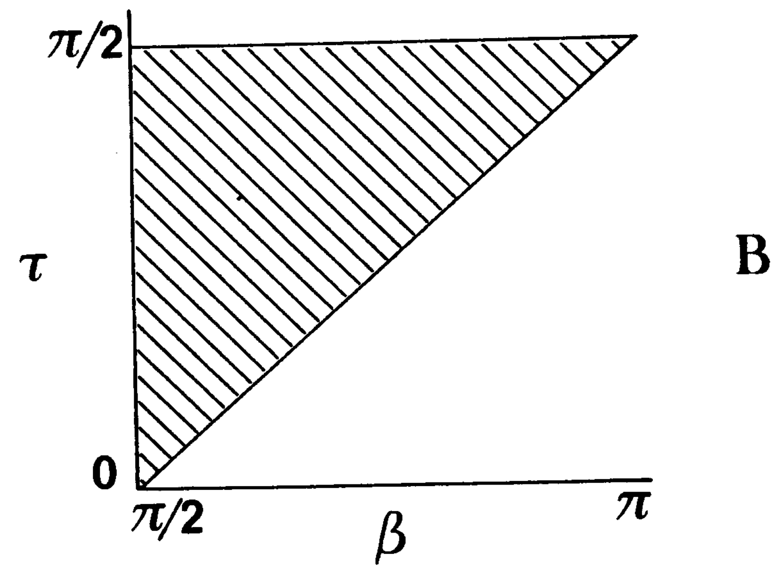


Figure 24. The oblique angle β formed by the projection of a right angle provides some constraint on both the slant σ and tilt τ components of surface orientation relative to the viewer. The possible values of slant and tilt are shown as cross-hatched for correspondence angle β varying from $\pi/2$ to π. Tilt τ is measured relative to one of the contours in the image, and varies from parallel ($\tau = 0$) to perpendicular ($\tau = \pi/2$).

(the unknown shape of Γ). The problem can then be solved by seeking to account for as much as possible of the variance in the surface contour in terms of the projective component. The constraint stems from the fact that the processes that determine the shape of contour generators on actual surfaces usually do not impose the same kind of systematic regularity as that imposed by orthographic projection.

4.2 The relationship between a contour generator and the surface

Given the contour generator Γ is a planar 3-D curve, how does the surface Σ lie under Γ? In terms of the wire and ribbon, a primary question concerns whether the ribbon may twist along the wire. More formally, if the plane containing Γ is Π, does the angle between Σ and Π vary along Γ?

A result in differential geometry is that given a curve Γ defined by the intersection of a plane Π and a surface Σ, if the angle between Σ and Π is constant along Γ, Γ is a line of curvature (see, e.g., [O'Neill, 1966, p. 224]). Thus if the contour generator is planar, and that plane intersects the surface with a constant angle, the contour generator is a line of curvature. The next issue is to determine the angle between Π and Σ.

4.2.1 The geodesic and asymptotic restrictions

If the plane Π containing the contour generator Γ is perpendicular to Σ, i.e., Γ is a normal section, then Γ is geodesic. Consequently the surface normal along Γ everywhere coincides with the principal normal to Γ. In essence, the contour generator follows a path on the surface which locally indicates where the greatest curvature occurs. The binormal to the contour generator, being perpendicular to both the principal normal and the tangent, coincides with the direction of least curvature. However all such binormals are parallel, for the tangent and normal along Γ only rotate in the plane Π. Consequently all lines of least curvature are parallel; equivalently, the strip of surface under the contour generator is a cylinder.

The previous discussion considered the case where the contour generator is geodesic; where the angle between Π and Σ is $\pi/2$. If that angle is everywhere zero, then Π coincides with the tangent plane of Σ and the surface normal along Γ coincides with the normal to Π. As mentioned earlier if a curve lies in a plane everywhere tangent to the surface along the curve, that curve is asymptotic, i.e., a locus of points of zero Gaussian curvature. The importance of the asymptotic restriction is found in gloss contours. The contour generators corresponding to gloss contours in the image correspond to asymptotic curves on the surface. Hence where gloss contours appear we know that the surface is locally developable (likewise, where point specularities occur we also know that the surface must be doubly curved). To some extent we may further understand the surface geometry simply on the basis of the shape of the contour in the image without determining the particular 3-D shape of its contour generator. If the contour is a straight line in the image we cannot tell much, for the surface may be either cylindrical or twisting (like a spiraling piece of paper). But if it is any smooth curve in the image the surface is roughly planar since the contour generator is restricted to be planar and asymptotic.

4.2.2 Parallelism

The discussion thus far has concerned the analysis of surface shape from a single surface contour. This analysis requires that the contour generator Γ may be determined from its image, however the constraint afforded by planarity, general position, symmetry, and constant curvature will not always allow a strong determination of Γ. It is perhaps not coincidental that, in fact, our perception of surface shape from a single, unfamiliar contour is weak when compared to the vivid impression afforded by multiple, parallel contours (figures 13 and 20). The basis for the apparently greater constraint from parallel contours will now be discussed.

If surface contours are parallel in the image, then by the of general position, their contour generators are parallel. The fundamental issue now concerns the behavior of the surface between the contour generators. In the absence of independent sources of information about the surface such as shading or texture we must make some *a priori* assumption about the nature of the surface between the contour generators. A conservative assumption would be that the surface extends in a "simple manner" between them. This can be formalized by a second form of general position: that the particular positions of the contour generators on the surface are not critical, that if shifted slightly, the contour generators would project qualitatively the same. This is equivalent to assuming that the surface is a cylinder between the contour generators.

We now use the geodesic-asymptotic restrictions from the previous section, and consider two interpretations for the cylindrical surface: Either the surface is (a) curved and the contour generators are parallel geodesics, or (b) flat and the contour generators are asymptotic curves. To aid in visualizing these two cases, compare figure 13 (geodesic interpretation) and figure 25 (asymptotic interpretation). Note that in the latter case of asymptotic curves, the parallelism does not provide additional constraint on the surface solution -- the contour generators lie in the same plane. Nor does the shape of each contour generator in the plane; it is as if the curves are merely arrayed on a flat surface. The interpretation of parallel contour generators as geodesics, however, constrains both the local surface orientation and the shape of the contour generators.

4.2.3 Computing parallel correspondence

Recall that the angle between the plane containing the contour generator and the surface is restricted to be constant, hence the contour generator is a line of (greatest) curvature. Also, the lines of least curvature on a cylinder are straight, parallel, and perpendicular to the lines greatest curvature. If a line of least curvature were reconstructed in the image, the angle of intersection that it would make with a surface contour (a line of greatest curvature) would be the projection of a right angle. This angle constrains the local surface orientation, as already demonstrated with regard to bilateral symmetry. In fact, the lines of least curvature can be reconstructed.

In the orthographic image of a cylinder the lines of least curvature would project as straight and parallel, and each would intersect successive surface contours at a constant angle (since the contour generators are parallel). This is illustrated in figure 26 (where the lines of least curvature are superimposed on figure 13). Note that we attempt to reconstruct only the *projections* of the lines of least curvature. This may be achieved by identifying points on adjacent contours whose tangents are parallel and connecting those points by straight lines that are parallel. This may be thought of as bringing points on adjacent contours into *parallel correspondence*. The constructed line representing the image of a line of least curvature will be termed a *correspondence line*. Note that if the surface contours are straight for a portion of their length (figure 27*a*) the tangent to a point P on one contour may be parallel to various tangents on the adjacent contour, however only one choice would result in a correspondence line that is parallel to the other correspondence lines between

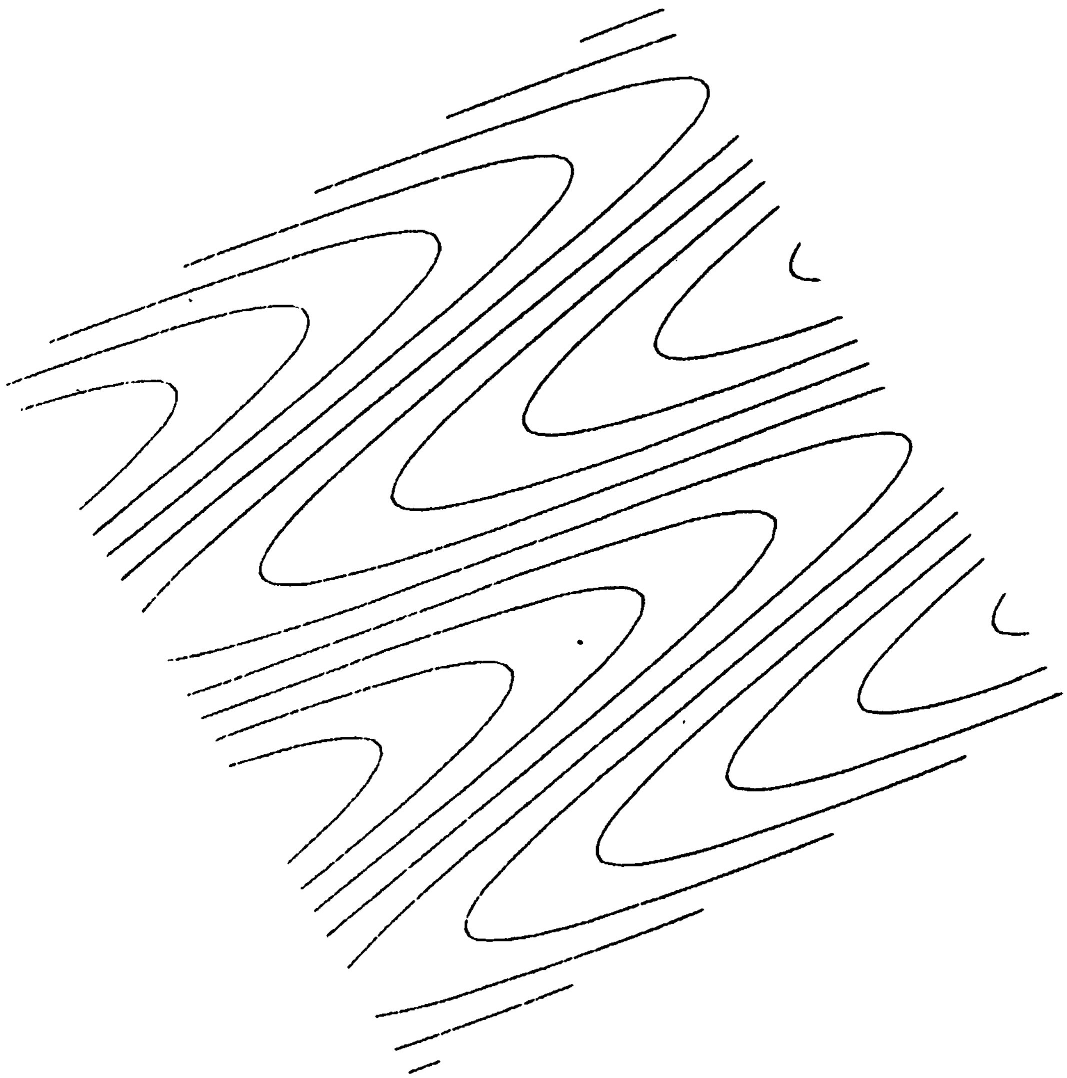


Figure 25. The contours seem to be interpreted as the image of asymptotic curves on a planar surface. Note that the surface appears flat in given this interpretation.

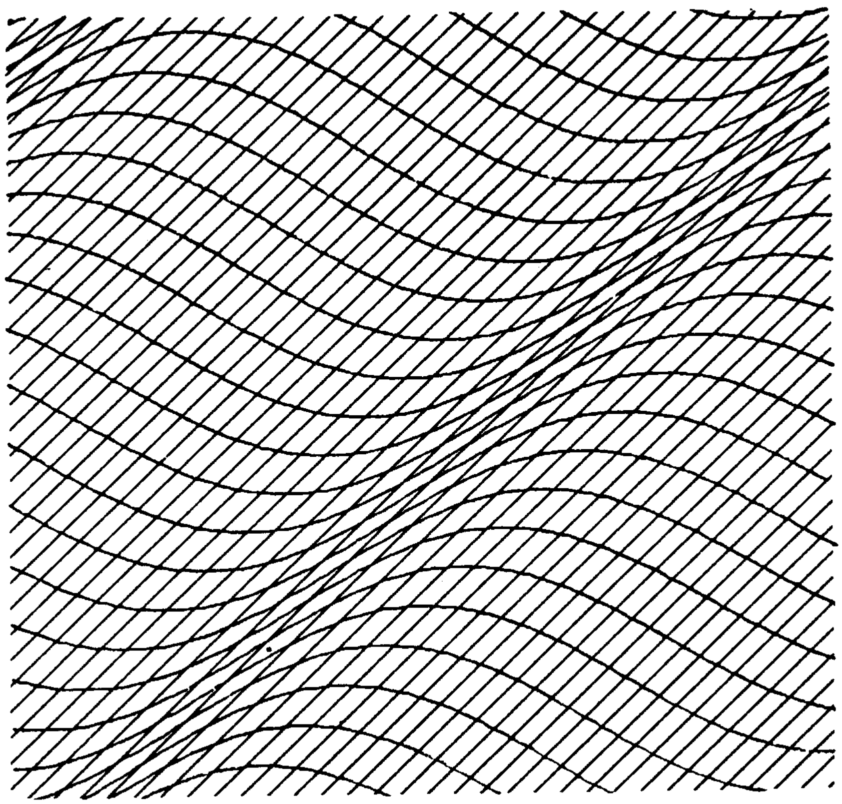


Figure 26. In the orthographic image of a cylindrical surface the lines of least curvature project as straight and parallel, and each intersect successive surface contours at a constant angle. Identifying points on adjacent contours whose tangents are parallel and connecting those points with lines that are parallel establishes *parallel correspondence*, one basis for postulating that the underlying surface is a cylinder (subject to general position).

curved portions of adjacent surface contours (figure 27*b*).[1]

This correspondence is unique in general, and therefore may be used as a constructive criterion for detecting parallelism between surface contours and for postulating that the surface is a cylinder.[2]

An important consequence of the parallel correspondence is that the surface orientation is necessarily constant along the lines of least curvature (in orthographic projection, as we have been assuming). Thus if the surface orientation were determined along the contour, it can be simply propagated along the correspondence lines to provide a complete, interpolated solution to the surface orientation across the cylindrical surface between parallel surface contours.

We have seen that assuming that the contour generator Γ is planar and that the angle between the plane containing Γ and the surface is constant along Γ restricts the surface under Γ to be a cylinder. Also, for parallel surface contours the two forms of general position together restict the surface to be a cylinder. Consequently, the curvature of the surface is attributed entirely to the curvature of the contour generator, that being a line of greatest curvature.

Note that the cylinder restriction is only local, for the parallel correspondence need only be established between adjacent surface contours, and the parallelism between reconstructed lines of least curvature is defined only locally. Consequently, the cylinder restriction may be applied, for example, to the surface contours in figures 20 and 28 where the surface may be approximated locally by patches of cylinders while the global surface is not cylindrical.

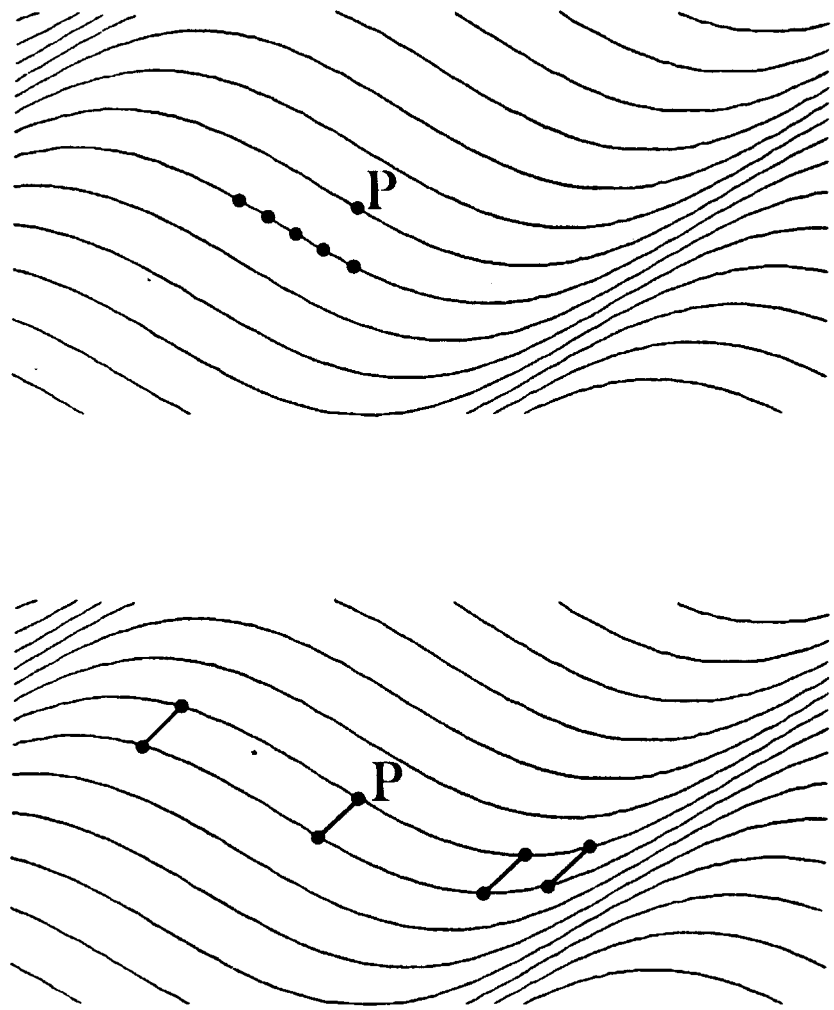
4.2.4 Opacity

We now consider the constraint afforded by restricting the surface to be opaque. In general, opacity does not significantly restrict the shape of the underlying surface. However the opacity restriction is important if, as before, the contour generator is assumed to be a line of greatest curvature and the surface under the contour generator is assumed cylindrical. In the following, a geometrical construction will be described that shows how these restrictions constrain the range of orientations to which the parallel lines of least curvature would project. The angle between those lines and the tangent to the surface contour is, again, the projection of a right angle. Thus the opacity restriction is useful in constraining local surface orientation in the same manner as skewed symmetry and parallel correspondence. The restriction imposed on slant and tilt as a function of this angle is shown in figure 24.

The constraint follows from the fact that if a line of curvature is continuously visible from a given viewpoint, so must an adjacent line of curvature. This can be described geometrically in the following way: The correspondence lines (the projections of lines of least curvature) that connect adjacent surface contours would make no intersections with the surface contours except at their terminations. That is, the situation in figure 29*a* would be disallowed. (Note that in figure 13, where this does not arise, the surface may be transparent nonetheless.) Now, given a single surface contour (the image of a line of greatest curvature on a

1. Selection of that choice may be accomplished by a local, parallel algorithm similar to that in [Stevens, 1978].

2. Note that the correspondence is not unique if, for instance, the parallel surface contours are periodic, as in figure 13. One solution in that case is to choose the parallel solution which results in the shortest correspondence lines.



A

B

Figure 27. If the surface contours are straight for a portion of their length, as in *a*, the tangent to a point P on one contour may be parallel to various tangents on the adjacent contour, however only one choice would result in a correspondence line that is parallel to the other correspondence lines between curved portions of adjacent contours, as in *b*.

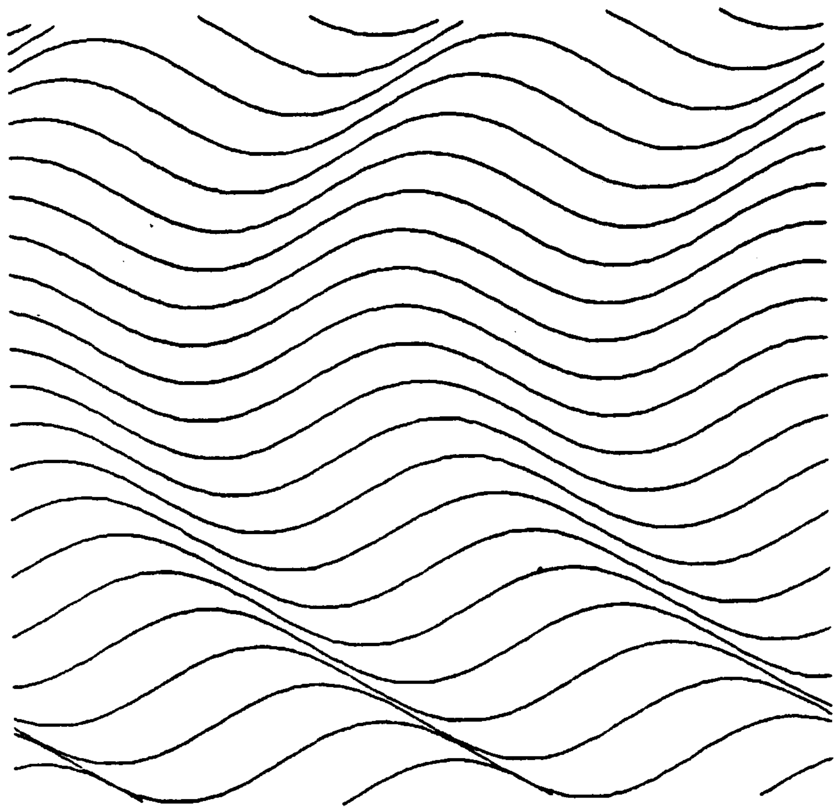


Figure 28. The cylinder restriction is only local, for the parallel correspondence need only be established between adjacent surface contours, and the parallelism between reconstructed lines of least curvature is defined only locally. Consequently the local cylinder restriction may be applied to the surface contours above although the global surface is not cylindrical.

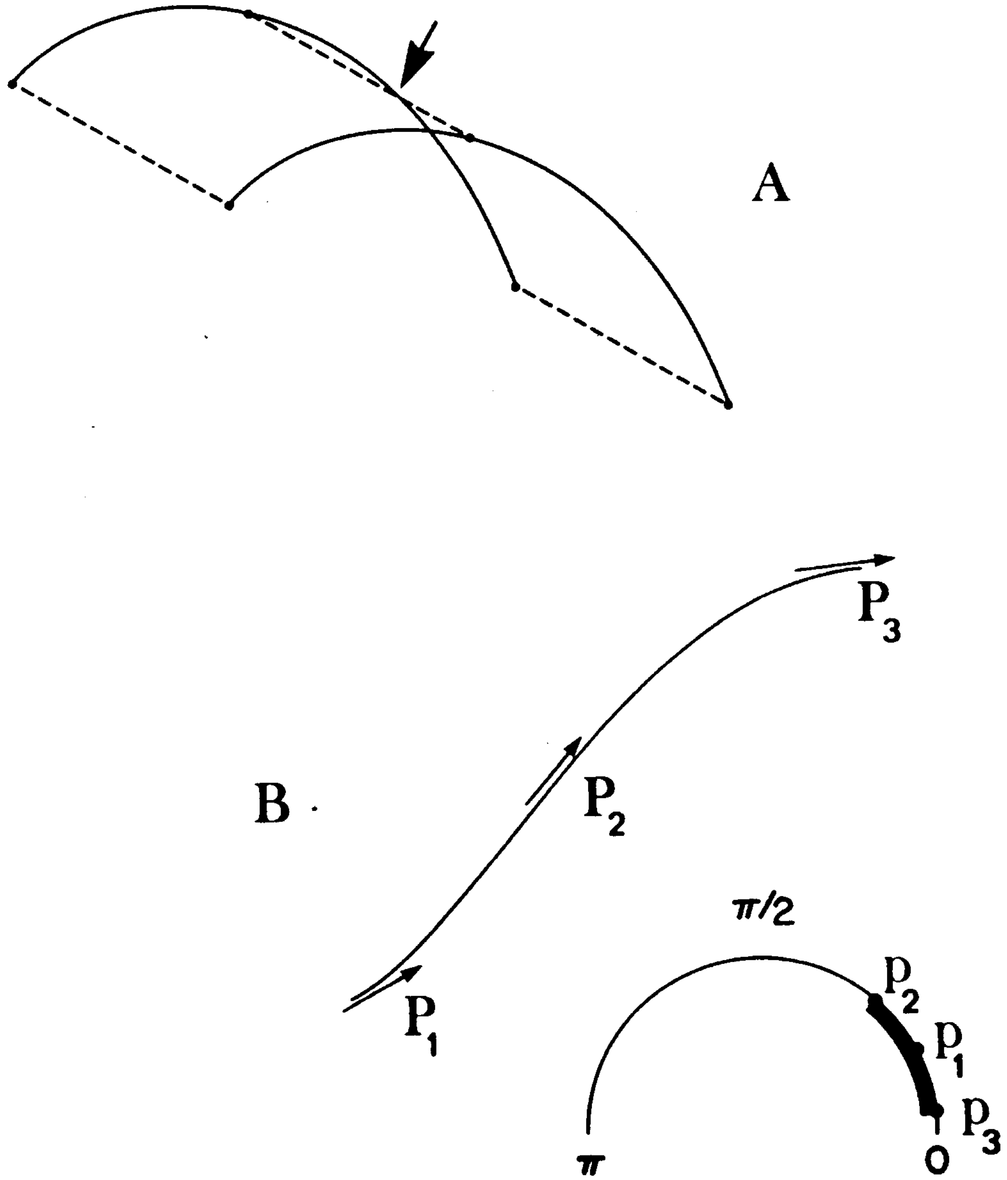


Figure 29. The opacity restriction disallows the correspondence lines (the projections of lines of least curvature) that connect adjacent surface contours to intersect the surface contours except at their terminations. That is, the situation in *a* is disallowed. Opacity provides some constraint on the relation between a contour generator and the underlying surface. Towards representing this constraint, we represent the surface contour by its Gauss map onto a semi-circle, as in *b*.

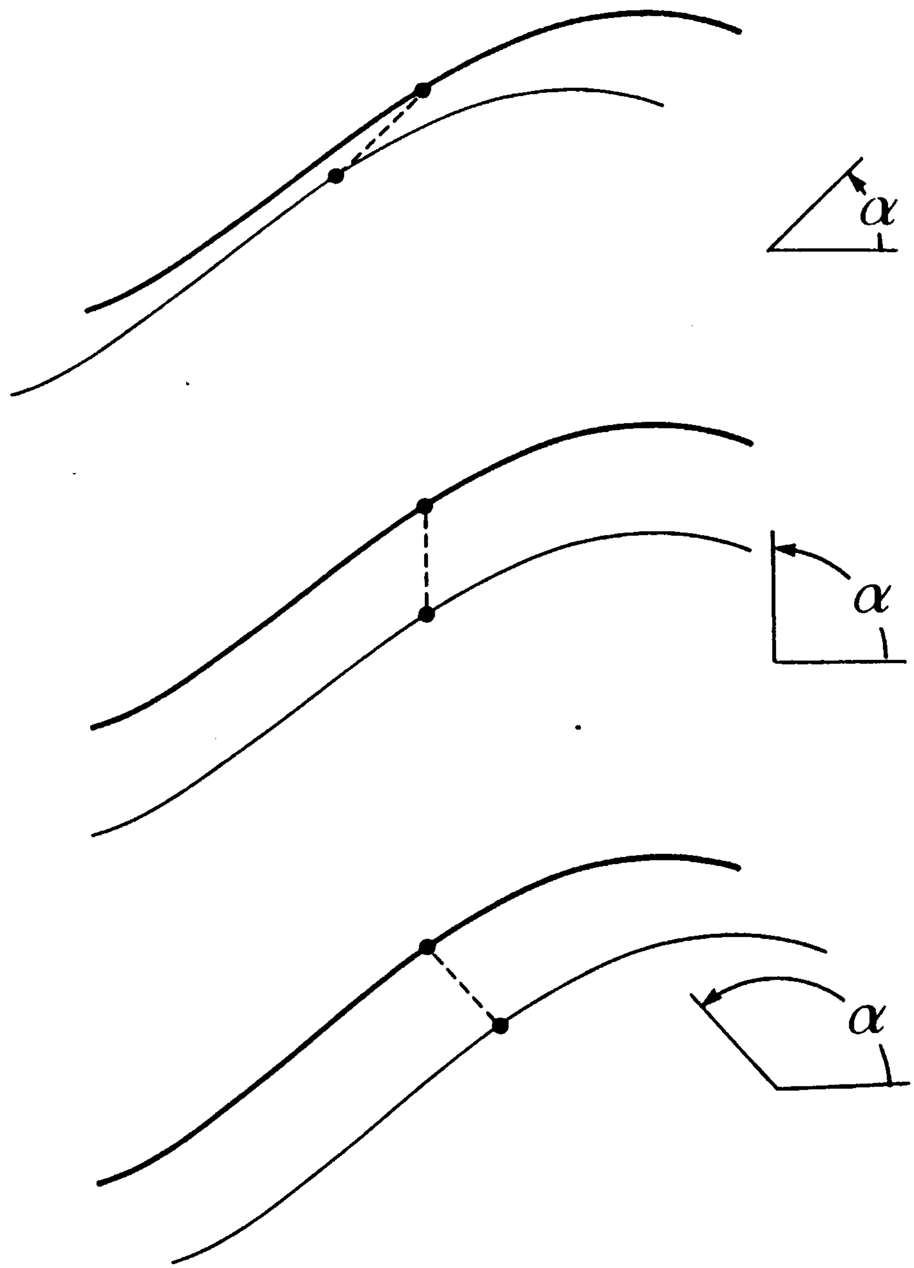


Figure 30. The surface underlying the contour (heavy line) is assumed to be a cylinder, and the problem is to determine the orientation α to which the lines of least curvature would project. Three examples of α are shown above. The opacity restriction places some constraint on α.

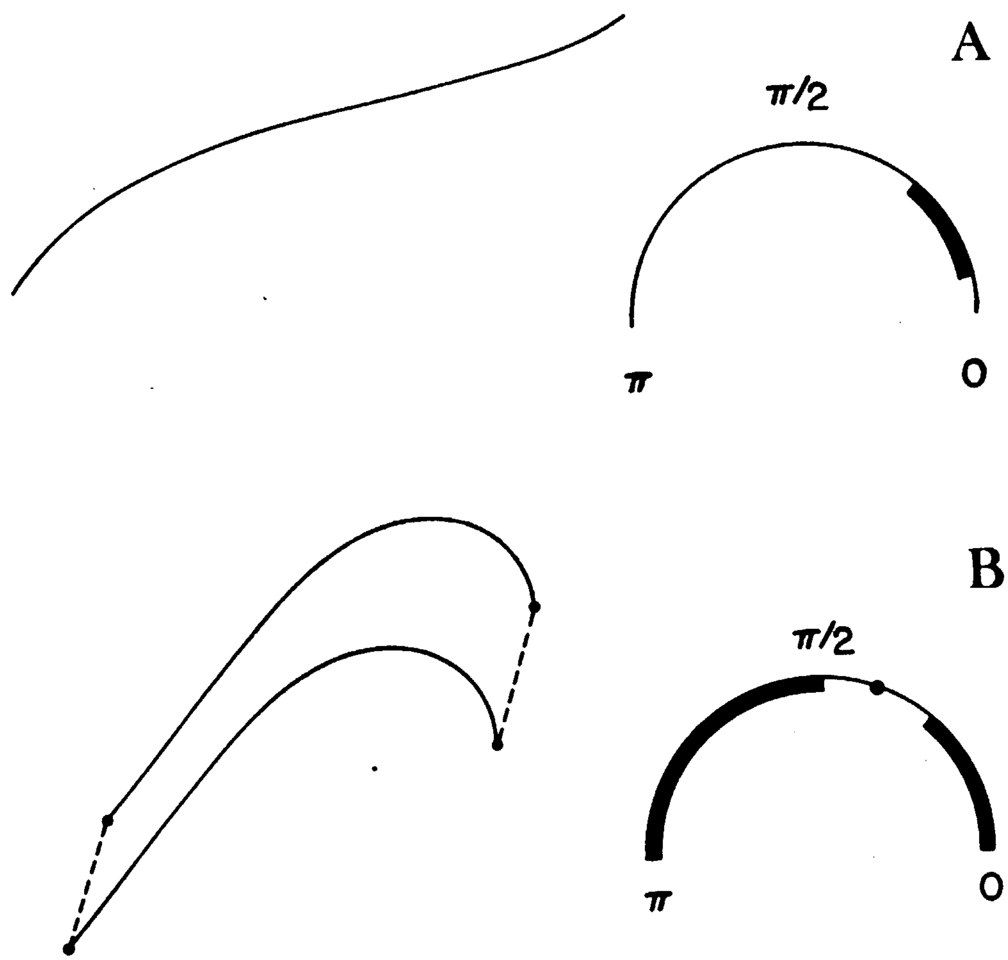


Figure 31. The image of the lines of least curvature map to a single point on the Gauss map. If opaque, that point cannot already be occupied by the mapping of the surface contour. In *a* the surface contour is a shallow curve which maps to a small arc on the Gauss map. This does not strongly constrain the possible orientations of the correspondence lines (the projected lines of least curvature). But in *b* the curve covers much of the Gauss map, hence the orientation of the lines of least curvature is strongly constrained. One choice of that orientation is shown, and the position of an adjacent, parallel surface contour is drawn. The opacity restriction then provides constraint on surface orientation by the oblique correspondence angle.

cylinder) we have some constraint on where an adjacent line of curvature would project, and this in turn constrains the local surface shape.

This constraint is conveniently represented by the Gauss map (see, for example, [Hilbert & Cohn-Vossen, 1952]). A Gauss map is a simple representation of the range of orientations of tangents along a curve. The given curve is mapped to an arc on a unit semi-circle where each point on the curve maps to the point on the semi-circle whose radius is parallel to the tangent to the curve. This is illustrated in figure 29*b*. Observe how tangents at various points P map to corresponding points on the semi-circle.

The next step is to use the Gauss map to represent the range of possible orientations of the correspondence lines. Let that orientation be α, which maps to a single point on the semi-circle (that point P whose radius has the orientation α). In figure 30 three choices for α are shown which are consistent with the surface being opaque. Now, the constraint that the correspondence lines not intersect the surface contours equates to the restriction that the point P not lie on the arc of the semi-circle already covered by the surface contour. The degree of constraint imposed by the opacity restriction depends on the surface contour. In figure 31*a* the shallow contour maps to only a short arc, and the correspondence lines could have a large range of orientations. But in figure 31*b* the correspondence lines are restricted to a narrow range of orientations.

Given that the correspondence lines are the projections of lines of least curvature which on a cylinder are identically the binormals to the plane containing the lines of greatest curvature, the orientation to which the correspondence lines projects provides us with the tilt component of surface orientation for the plane containing the given curve. It is worthwhile to refer back to figures 15*b*, 16*b*, and 18*b*, which seem to be patches of cylinders. The curves would be lines of greatest curvature, the straight lines would be lines of least curvature. Their mutual orthogonality would explain our interpretation of them as right angles in 3-D.

4.3 Criteria governing the tangential/surface contour decision

Earlier we discussed the distinction between tangential contours (silhouette boundaries along which the line of sight grazes the surface) and surface contours, noting that surface contours include silhouette boundaries that are not tangential contours. Marr [1977a] has delineated properties of the silhouettes of generalized cones (whose boundaries are tangential contours) -- surfaces whose shape can be recovered from their silhouettes. The silhouette of a generalized cone exhibits qualitative symmetry; where the correspondence lines connecting symmetric segments of the contour would be perpendicular to the axis of symmetry. For instance, the symmetric silhouette in figure 14*a* is generally interpreted as a vase-like object, and the contours are seen as tangential contours.

Similarly, geometrical criteria can be given which indicate that a contour is a surface contour. (Note that non-geometrical means also exist, e.g., determining that the corresponding contour generator is a shadow edge, or a gloss contour or a discontinuity in surface texture) Two geometrical criteria are suggested by the preceding discussion. First consider qualitative symmetry where the correspondence lines are not perpendicular to the axis of symmetry (as just discussed in the case of bilateral symmetry) but oblique to the axis (as in figure 23*b*). When achieved, this skewed symmetry suggests a surface contour, as opposed to a tangential contour, interpretation. Secondly, if parallel correspondence between contours can be achieved (as in figures 13, 14*b*, and 15*b*) those contours can be interpreted as surface contours.

5. SUMMARY

1. The analysis of the shape of a surface from surface contours may be decomposed into two problems: reconstructing the corresponding 3-D curves (the *contour generators*) and determining their relation to the surface. This decomposition separates the problem of determining the projective geometry from that of determining the intrinsic geometry.

2. The first problem is constrained by general position, planarity, symmetry, and minimum curvature variation.

3. The second problem is reduced by assuming the angle between the surface and the plane containing the contour generator is constant. Then if that angle is a right angle, the contour generator is geodesic; if the angle is zero, the contour generator is asymptotic. In either case the contour generator is also a line of curvature. Since it is also planar, the surface is locally a cylinder.

4. We also arrived at the cylinder restriction in the case of parallel surface contours, given the two forms of the principle of general position. The opacity restriction is also useful, given the planarity and geodesic restrictions, in understanding how the surface lies under a contour generator.

5. We have considered instances when the various constraints are valid. Surface markings on synthetic and biological objects and the edges of cast shadows are often geodesic and planar. Gloss contours are asymptotic and planar, at least in the case of distant light sources and orthographic projection. Hence if the contour generator can be reconstructed as a curve in 3-D, the surface orientation along the curve can be computed subject to either the geodesic or asymptotic interpretations.

6. Constraints on the intrinsic geometry are also provided by surface contours even if the contour generator is not well determined in space: Gloss contours, highlights, and shading edges tell us of the local Gaussian curvature in some cases.

REFERENCES

Arnheim, R. 1954 *Art and visual perception.* Berkeley: University of California Press.

Attneave, F. 1972 Representation of physical space. In *Coding processes in human memory.* Melton, A.W. and Martin, E., eds. New York: John Wiley.

Attneave, F. and Frost, R. 1969 The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics 6*, 391-396.

Bajcsy, R. 1972 Computer identification of textured visual scenes. Memo AIM-180, Stanford University.

Bajcsy, R. & Lieberman, L. 1976 Texture gradients as a depth cue. *Computer Graphics and Image Processing 5*, 52-67.

Beck, J. 1960 Texture-gradients and judgments of slant and recession. *American Journal of Psychology 73*, 411-416.

Beck, J. 1972 *Surface color perception.* Ithaca: Cornell University Press.

Bergman, R. and Gibson, J.J. 1959 The negative after-effect of the perception of a surface slanted in the third dimension. *American Journal of Psychology 72*, 364-374.

Boring, E.G. 1951 Review of J.J. Gibson, *The perception of the visual world. Psychological Bulletin 48*, 360-363.

Braunstein, M. 1968 Motion and texture as sources of slant information. *Journal of Experimental Psychology 78*, 247-253.

Braunstein, M. & Payne, J.W. 1969 Perspective and form ratio as determinants of relative slant judgements. *Journal of Experimental Psychology 81*, 584-590.

Clark, W.C., Smith, A.H., & Rabe, A. 1956 The interaction of surface texture, outline gradient, and in the perception of slant. *Canadian Journal of Psychology 10*, 1-8.

Coren, S. 1972 Subjective contours and apparent depth. *Psychological Review 79*, 359-367.

Epstein, W. and Landauer, A.A. 1969 Size and distance judgements under reduced conditions of viewing. *Perception and Psychophysics 6*, 269-272.

Epstein, W. and Park, J. 1964 Examination of Gibson's psychophysical hypothesis. *Psychological Bulletin 62*, 180-196.

Flock, H.R. 1964 A possible optical basis for monocular slant perception. *Psychological Review 71*, 380-391.

Flock, H.R. 1964 Three theoretical views of slant perception. *Psychological Bulletin 62*, 110-121.

Flock, H.R. 1965 Optical texture and linear perspective as stimuli for slant perception. *Psychological Review 72*, 505-514.

Flock, H.R., Graves, D., Tenney, J. and Stephenson, B. 1967 Slant judgments of single rectangles at a slant. *Psychonomic Science 7*, 57-58.

Freeman, R.B. 1965 Ecological optics and visual slant. *Psychological Review 72*, 501-504.

Freeman, R.B. 1966 The effect of size on visual slant. *Journal of Experimental Psychology 71*, 96-103.

Gibson, J.J. 1950 *The perception of the visual world.* Boston: Houghton Mifflin.

Gibson, J.J. 1950 The perception of visual surfaces. *American Journal of Psychology 63*, 367-384.

Gibson, J.J. 1959 Optical motions and transformations as stimuli for visual perception. *Psychological Review 64*, 288-295.

Gibson, J.J. 1966 *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J.J. 1971 The information available in pictures. *Leonardo 4*, 27-35.

Gibson, J.J. and Flock, H. 1962 The apparent distance of mountains. *American Journal of Psychology 75*, 501-503.

Gogel, W.C. 1965 Equidistance tendency and its consequences. *Psychological Bulletin 64*, 153-163.

Gogel, W.C. 1971 The validity of the size-distance invariance hypothesis with cue reduction. *Perception and Psychophysics 9*, 92-94.

Graham, C.H. 1965 *Vision and Visual Perception.* edited by C.H. Graham. New York: John Wiley.

Gregory, R.L. 1970 *The intelligent eye.* New York: McGraw-Hill.

Gregory, R.L. 1973 The confounded eye. In *Illusion in nature and art*, Gregory, R.L. & Gombrich, E.H., eds. New York: Charles Scribner's Sons.

Haber, R.N. & Hershenson, M. 1973 *The psychology of visual perception.* New York: Holt, Rinehart and Winston.

Helmholtz, H. 1925 *Physiological optics*, Vol. 3 (3rd edition translated by J.P. Southall). New York: Optical Society of America.

Hilbert D., & Cohn-Vossen, S. 1952 *Geometry and the Imagination.* Chelsea Publishing.

Horn, B.K.P. 1975 Obtaining shape from shading information. In *The psychology of computer vision*, P. H. Winston, ed. New York: McGraw-Hill.

Huffman, D.A. 1971 Impossible objects as nonsense sentences. In *Machine intelligence 6*, R. Meltzer and D. Michie, eds. 295-323. Edinburgh: The Edinburgh University Press.

Ittelson, W.H. 1960 *Visual space perception.* New York: Springer-Verlag.

Ittelson, W.H. 1968 *The Ames demonstration in perception.* New York: Hafner Publishing.

Jernigan, M.E. and Eden, M. 1976 Model for a three-dimensional optical illusion. *Perception and psychophysics 20*, 438-444.

Julesz, B. 1971 *Foundations of cylopean perception.* Chicago: Chicago Press.

Kaiser, P.K. 1967 Perceived shape and its dependency on perceived slant. *Journal of Experimental Psychology 75*, 345-353.

Kanade, T., and Kender, J.R. 1979 Skewed symmetry: Mapping image regularities into shape, Technical Report, Computer Science Department, Carnegie-Mellon University (forthcoming).

Kennedy, J.M. 1974 *A psychology of picture perception.* San Francisco: Jossey-Bass.

Koffka, K. 1935 *Principles of gestalt psychology.* New York: Harcourt Brace.

Kraft, A.L. and Winnick, W.A. 1967 The effect of pattern and texture gradient on slant and shape judgments. *Perception and Psychophysics 2*, 141-147.

Luckiesh, M. 1965 *Visual illusions - Their causes, characteristics and applications.* New York: Dover Publications.

Macworth, A.K. 1973 Interpreting pictures of polyhedral scenes. *Artifical Intelligence 4,* 121-137.

Marr, D. 1976 Early processsing of visual information. *Phil. Trans. Roy. Soc. B. 275,* 483-524.

Marr, D. 1977 Analysis of occluding contour. *Proc. R. Soc. Lond. B. 197,* 441-475. Also available as M.I.T. A.I. Lab. Memo 372.

Marr, D. 1977 Representing visual information. *AAAS 143rd Annual Meeting, Symposium on Some Mathematical Questions in Biology,* February. Also available as M.I.T. A.I. Lab. Memo 415.

Marr, D. & Hildreth, E. 1979 Theory of edge detection. *Proc. R. Soc. Lond. B.* in the press. Also available as M.I.T. A.I. Lab. Memo 518.

Marr, D. & Nishihara, K. 1978 Representation and recognition of the spatial organization of three-dimensional shapes. *Phil. Trans. Roy. Soc. B. 200,* 269-294.

Marr, D. and Poggio, T. 1976 Cooperative computation of stereo disparity. *Science 194,* 283-287.

Marr, D. & Poggio, T. 1977 From understanding computation to understanding neural circuitry. *Neuroscience Reseach Progress Bulletin 15,* 470-488. Also available as M.I.T. A.I. Lab. Memo 357.

Marr, D. and Poggio, T. 1978 A theory of human stereo vision. *Proc. Roy. Soc. Lond. B. 204,* 301-328. Also available as M.I.T. A.I. Lab. Memo 451.

Nelson, T.M. and Bartley, S.H. 1956 The perception of form in an unstructured field. *Journal of Experimental Psychology 54,* 57-63.

Ogle, K.N. 1962 Perception of distance and of size. In *The eye,* Vol. 4, Edited by H. Davson. New York: Academic Press.

O'Neill, B. 1966 *Elementary Differential Geometry.* New York: Academic Press.

Olson, R.K. 1974 Slant judgments from static and rotating trapezoids correspond to rules of perspective geometry. *Perception and Psychophysics 15,* 509-516.

Postman, L. & Tolman, E.C. 1959 Brunswick's Probabilistic Functionalism. In *Psychology: A study of a science.* Edited by S. Koch. New York: McGraw-Hill.

Purdy, W.C. 1960 The hypothesis of psychophysical correspondence in space perception. General Electric Technical Information Series, No. R60ELC56.

Robinson, J.O. 1972 *The psychology of visual illusion.* London: Hutchinson University Library.

Rock, I. and McDermott, W.P. 1964 The perception of visual angle. *Acta Psychologica 22,* 119-134.

Rosinski, R.R. 1974 On the ambiguity of visual stimulation: A reply to Eriksson. *Perception and Psychophysics 16,* 259-263.

Shepard, R.N. 1979 Psychophysical complementarity. To appear in *Perceptual organization* M. Kubovy and J.R. Pomerantz, eds. Hillsdale, N.J.: Lawerence Erlbaum Associates.

Shepard, R.N. and Metzler, J. 1971 Mental rotation of three-dimensional objects. *Science 171,* 701-703.

Smith, A.H. 1965 Interaction of form and exposure time in the perception of slant. *Perceptual and Motor Skills 20,* 481-490.

Smith, O.W. 1958 Judgements of size and distance in photographs. *Americal Journal of Psychology 71*, 529-538.

Smith, O.W. & Smith, P.C. 1957 Interaction of the effects of cues involved in judgments of curvature. *Americal Journal of Psychology 70*, 361-375.

Stevens, K.A. 1976 Occlusion clues and subjective contours. M.I.T. A.I. Lab Memo 363.

Stevens, K.A. 1978 Computation of locally parallel structure. *Biological Cybernetics 29*, 19-28. Also available as M.I.T. A.I. Lab Memo 392.

Stevens, K.A. 1979 Representing and analyzing surface orientation. In *Artificial Intelligence: An MIT Perspective*. P.H. Winston and R.H. Brown, eds. 104-125. Cambridge: MIT Press.

Street, R.F. 1931 A Gestalt completion test: A study of a cross-section of intellect. In: *Teachers College Contributions to Education, No. 481*. New York: Teachers College, Columbia University.

Ternus, J. 1926 Experimentelle Untersuchung uber phanomenale Identitat. *Psychol. Forsch. 7*, 81-136. Translated in Ellis, W.D. 1967 *A Source book of Gestalt Psychology*. New York: Humanities Press.

Ullman, S. 1979 *The interpretation of visual motion*. Cambridge, Ma.: M.I.T. Press.

Waltz, D. 1975 Understanding line drawings of scenes with shadows. In *The Psychology of computer vision*, P.H. Winston, ed. New York: McGraw-Hill.

Weinstein, S. 1957 The perception of depth in the absence of texture gradient. *American Journal of Psychology 70*, 611-615.

Witkin, A.P. 1980 Shape from contour. Ph.D. Thesis (Psychology), M.I.T., February.

Woodham, R.J. 1977 Reflectance map techniques for analyzing surface defects in metal castings. Ph.D. Thesis, M.I.T., September.

Youngs, W.M. 1976 The influence of perspective and disparity cues on the perception of slant. *Vision Research 16*, 79-82.

APPENDIX A
TILT EXPERIMENTS

Two experiments were performed concerning the judgment of surface tilt from configurations of intersecting straight lines. The first established that the tilt judgments are well defined relative to the geometry of the figure and independent of the orientation of the figure on the display screen. The second experiment demonstrated that the tilt judgment is dependent on the relative lengths of the two lines and on their angle of intersection. It is concluded that we probably solve the tilt by assuming that the lines are actually equal-length and that the angle of intersection is a right angle in three dimensions.

Judgements of surface slant were not made; the apparatus was designed to allow tilt to be decoupled from slant. While judgments of surface slant from line drawings are generally poor both in terms of underestimation ("regression to the frontal plane") and substantial variability, this study has discovered that surface tilt judgements can be considerably more accurate and precise. The two experiments shared a common design which is discussed in the following.

A.1 Experimental design

A.1.1 Apparatus

The subjects observed line-drawn figures on a Knight rasterscan CRT display. The lines were luminous against a dark background; the room was darkened. The figures were viewed monocularly through a 25 mm diameter circular aperature of an occluding mask positioned roughly 50 cm from the display.

In order to measure tilt, it was planned that the Ss would adjust an actual rod so that it appeared normal to the visualized surface. The rod was situated between the S and the CRT screen, attached to a transparent plate by a small universal joint which allowed the rod to be placed at any spatial orientation. When viewed monocularly the rod appeared to extend from the surface suggested by the figure towards the S. By grasping the free end, the S could place it so that it appeared normal. The tilt component was then projected onto the image plane (by displaying a vector with one end fixed so that it was coincident with the fixed end of the rod, and rotating it until it was occluded by the rod from the S's viewpoint). Measuring the tilt component in this manner avoided having the S adjust the tilt direct. However this precaution was unnecessary: Instead of this apparatus, the S merely rotated a displayed vector to appear normal to the imagined surface. Surprisingly, the Ss reported greater confidence when judging the projected tilt directly than when adjusting the rod. This was reflected in improved consistency between trials. Presumably the rod was more difficult to position due to the additional, implicit task of adjusting its slant.

In the first experiment of the first series, the length of the normal vector was roughly comparable to the dimensions of the stimulus figure. The Ss commented that the length seemed inappropriately long when the surface appeared nearly parallel to the image plane (slant roughly zero), and that the vector often appeared to change length as it was rotated in the image. It was suspected that the length of the normal vector was affecting the perceived surface orientation, therefore in subsequent experiments the vector was extended beyond the field afforded by the aperature. This enhanced the illusion of the vector being normal to the surface. With the vector continuously displayed, Ss stated that a range of orientations were equally acceptable, however if the vector were removed and redisplayed, the initial impression of the orientation of the vector could be used to make more critical judgements. Therefore, in later experiments, only the surface contours were continuously displayed, the normal vector would be flashed on the screen, providing the S with a glimpse of the vector to compare with the imagined normal.

The control of stimulus display, rotation of the vector, and data collection were all performed interactively by keyboard. Rotation was stepped clockwise and counterclockwise in five-degree and one-degree increments. The S would position the normal vector by a succession of keystrokes that first flash the vector then make incremental rotations.

A.1.2 Procedure

An attempt to measure the subjective tilt of an orthographically projected surface must contend with spontaneous reversals in depth which affect the *direction* of the tilt. (In the absence of perspective, the depth interpretation of a figure is ambiguous.) One factor that affects the interpretation is the orientation of the figure in the image plane. For example, an ellipse oriented with a horizontal major axis can either be seen as a disk with the lower edge nearer, or with the upper edge nearer. In general, when the perceived surface is roughly horizontal, there is a tendency to prefer the interpretation with an upward pointing normal. However, if the figure is oriented such that the surface is roughly vertical, the surface may be interpreted with the normal pointing to the left or the right with roughly equal preference. With the ellipse, therefore, if the figure were rotated in the image plane, at some point the observer may experience a reversal in depth. If the left edge of the disk were seen to lie further than the right, then the normal would point horizontally to the left, and vice versa.

Each S was given an introduction to the depth reversals. Given a figure, the S was asked to indicate the surface orientation (by orienting a piece of paper or the palm of the hand). Then the S was asked to see it "another way". The figures used in this study were oriented such that the tilt directions associated with the two depth interpretations were in the second and fourth quadrant. However, the Ss were generally to use the interpretation that placed the normal in the second quadrant. This restriction was not described to the Ss in terms of quadrants; the Ss would occasionally place the vector in the fourth quadrant, whereupon it was requested that the surface be seen "the other way". Reversals in interpretation were easy to achieve by all Ss. Before collecting data, each S was given a few trials on figures that were similar to those in the experiment. The vector was supposed to be seen as the normal to an opaque surface, hence projecting towards the S.

A.2 Experiment I

The goal of the first experiment was to simply show that tilt judgements can be made with precision from a simple intersection of two straight lines (see figure A-1*a*). The tilt was expected to be somehow determined by the contour geometry, independent of the orientation of the figure on the display screen, i.e., there was an expectation for a linear association between tilt judgements and image orientation (with unity slope).

A.2.1 Method

Stimuli: The intersection figure was described by the ratio R of the two line lengths, the obtuse angle of intersection β, and the orientation α of the figure on the screen (figure A-1*b*). The surface tilt was measured by the orientation τ of the normal vector. All angles were measured counterclockwise. In experiment I, R = 0.27 and β = 110 deg. The experimental variable was α. Since spontaneous reversals in depth interpretation were expected if the total rotation exceeded 90 deg, the various orientations in the image were restricted to within a range of 70 deg, i.e., α = 10, 20, 40, 60, and 80 deg. The figures subtended roughly seven deg of visual angle. During this experiment, data was also collected for a similar figure, a parallelogram. The parallelogram can also be described by the R, β, and α parameters. In this experiment, these parameters were the same as for the intersection figure.

Procedure: The experiment involved randomized presentations of the two types of figures at five orientations. Each of the 10 presentations were given once with unlimited viewing time. For each presentation, the S first viewed the figure, then the normal vector was displayed and positioned. Six unpaid, volunteer graduate students (five male, one female) were subjects.

A.2.2 Results

The data were tabulated separately for the intersection and parallelogram figures. In both cases, the linear association between τ and α was significant: for the intersection figures r = 0.98 (t = 27.736, $d.f.$ = 30, $p < 0.05$); for the parallelogram figures r = 0.94 (t = 14.473, $d.f.$ = 30, $p < 0.05$). The computed slopes of simple linear regression lines were: 0.96 (standard error = 0.035) for the intersection figures and 0.95 (standard error = 0.066) for the parallelograms. Neither slope was significantly different from 1.0: (t = 0.785, $d.f.$ = 30, $p > 0.2$) and (t = 1.126, $d.f.$ = 30, $p > 0.2$), respectively.

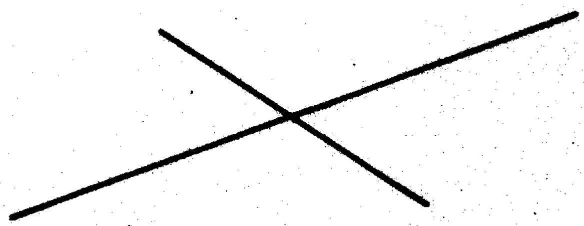
The data for both types of figure for each S were then analyzed individually, and the correlation coefficients were all significant: the least significant finding was r = 0.94 (t = 4.007, $d.f.$ = 3, $p < 0.05$). For the intersection figures, the slopes of the linear regression lines for each S ranged from 0.88 to 1.05. In comparing these slopes to 1.0, none of the differences reached significance ($p > 0.2$). For the parallelogram figures, only the slopes for two Ss were significantly different from 1.0.

The values of τ were reduced by the quantity (α-10.0) so that the judgements of tilt could be normalized to one image orientation, α = 10 deg. The resulting mean tilt for the intersection figures was 104.0 deg ($s.d.$ = 1.58 deg), and for the parallelogram was 101.4 deg ($s.d.$ = 3.36 deg). The difference between these two means did not reach significance (t = 1.57, $d.f.$ = 8, $p > 0.1$).

A.2.3 Discussion

We conclude that, at least for the surfaces suggested by a pair of intersecting lines or a parallelogram, the tilt is not functionally dependent on the particular orientation of the figure in the image plane. The low standard deviations of 1.58 and 3.36 deg demonstate that tilt judgements can be well defined. The parallelogram and intersection figures share the same contour geometry, described by the parameters R and β.

The basic finding given by this experiment was that on very simple configurations the surface orientation can be well defined. The intersection figure strongly suggests a surface, and the tilt component can be judged with precision. The intersection figure is further examined in experiment II.



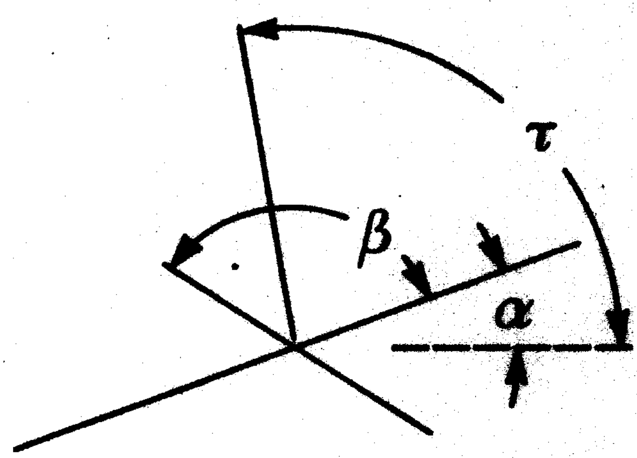


Figure A-1. The subjects observed a simple intersection figure (*a*) for varying values of orientation α and angle of intersection β, and adjusted the apparent orientation τ of surface tilt (*b*).

A.3 Experiment II

The goal of this experiment was to demonstrate that, for the intersection figure, tilt is dependent on the relative lengths of the two contours and on their angle of intersection. From experiment I we can discount the angle of orientation in the image as a functional parameter that governs the tilt.

A.3.1 Method

Stimuli: The intersection figures were presented with three values of angle of intersection β = 110, 130, and 170 deg, and three length ratios R = 0.272, 0.455, and 0.727. So that the presentations would appear varied, two image orientations α = 20 and 60 deg were used. In this experiment, the normal vector was extended beyond the field of view provided by the occluding mask.

Procedure: The total of 18 presentations were performed with successive presentations alternating between α = 20 and 60 deg. The sequence was randomized in terms of β and R. Each presentation was given once, however the data from the two image orientations would effectively provide two data points for each combination of β and R. Five unpaid, volunteer graduate students (four male, one female) were subjects. Only one subject (male) had participated in experiment I.

A.3.2 Results

The τ data collected at α = 60 were reduced by 40.0 in order to normalize to α = 20 deg. The values of τ for each image orientation were then tabulated for each of the nine combinations of β and R. The results of a two-way analysis of variance with equal replications are given in table A-1.

The data from α = 20 deg were compare to the adjusted data from α = 60 deg to further test whether there is a functional dependence of τ on the image orientation. The results are given in table A-2. The differences between the two sample means reached significance in three instances (β = 130, R = 0.27; β = 110, R = 0.40; and β = 110, R = 0.73) however the actual differences are 0.4, 2.4 and 7.4 deg, respectively. The mean tilt judgments are shown in figure A-2 as short line segments that extend from the intersection, much as presented to the Ss. However in the actual experimental situation, the line segment that was adjusted to appear normal to the intersection extended beyond the field of view and thus did not contribute a length to the local configuration. In observing figure A-2, the apparent 3-D length of the normal will appear inappropriate for the configurations near the lower right, especially for the case where R = 0.73 and β = 110. As a consequence, the line representing the image of the normal will probably appear overrotated counterclockwise in those cases. In the experiment, however, these choices of tilt orientation appeared appropriate.

A.3.3 Discussion

A strong functional dependence of τ on both β and R was found. (However the judgements of tilt also exhibited some dependence on the image orientation, as noted.) The values of τ were compared to the corresponding values that would be predicted if the lines were perpendicular and of equal length in 3-D. These values are given in the third column of table A-2. The judgment means did not differ significantly from those predictions, except where indicated with superscripts.

Source	s.s.	d.f.	M.S. (s.s./d.f.)	M.S.R.
Between β	1340.188	2	670.094	23.805
Between R	1351.438	2	675.719	24.005
β-R interaction	404.390	4	101.098	3.591
Residual	2280.047	81	28.149	---

Table A-1. Analysis of variance. Mean tilt (combined data from α = 20 and 60 deg) examined according to effects of obtuse angle β and length ratio R. All M.S.R.'s reach 0.05 significance.

β	R	Predicted τ	Mean τ for $\alpha=20$	Mean τ for $\alpha=60$	Comparison
170	0.27	110.68	110.73 (1.53)	111.13 (1.76)	($p>0.2$)
170	0.45	111.69	110.33 (3.06)	111.13 (3.69)	($p>0.2$)
170	0.73	113.45	112.73 (2.82)	113.13 (4.59)	($p>0.2$)
130	0.27	112.12	112.93 (2.00)	113.33 (6.86)	($p<0.05$)[4]
130	0.45	115.96	116.33 (4.60)	119.90 (4.09)[2]	($p>0.2$)
130	0.73	124.91	124.93 (6.92)	127.13 (6.53)	($p>0.2$)
110	0.27	111.45	111.53 (5.60)	117.13 (7.31)[1]	($p>0.2$)
110	0.45	114.48	117.73 (3.34)[1]	120.13 (10.86)	($p<0.05$)[4]
110	0.73	124.88	123.70 (5.66)	131.10 (4.27)[3]	($p<0.05$)

[1]($0.2<p<0.1$) [2]($0.05<p<0.1$) [3]($p<0.05$) [4]variances significantly different by F-test.

Table A-2. Values of mean tilt τ (with standard deviations in parentheses) for two image orientations, $\alpha=20$ and 60 deg, over nine combinations of obtuse angle β and length ratio R. The last column shows the results of comparison of the means at the two values of α. In comparing the two means, if the variances were not significant, then a t-test was performed. Each mean was also compared to the corresponding theoretic value, and except where superscripted, the differences did not reach significance ($p>0.2$).

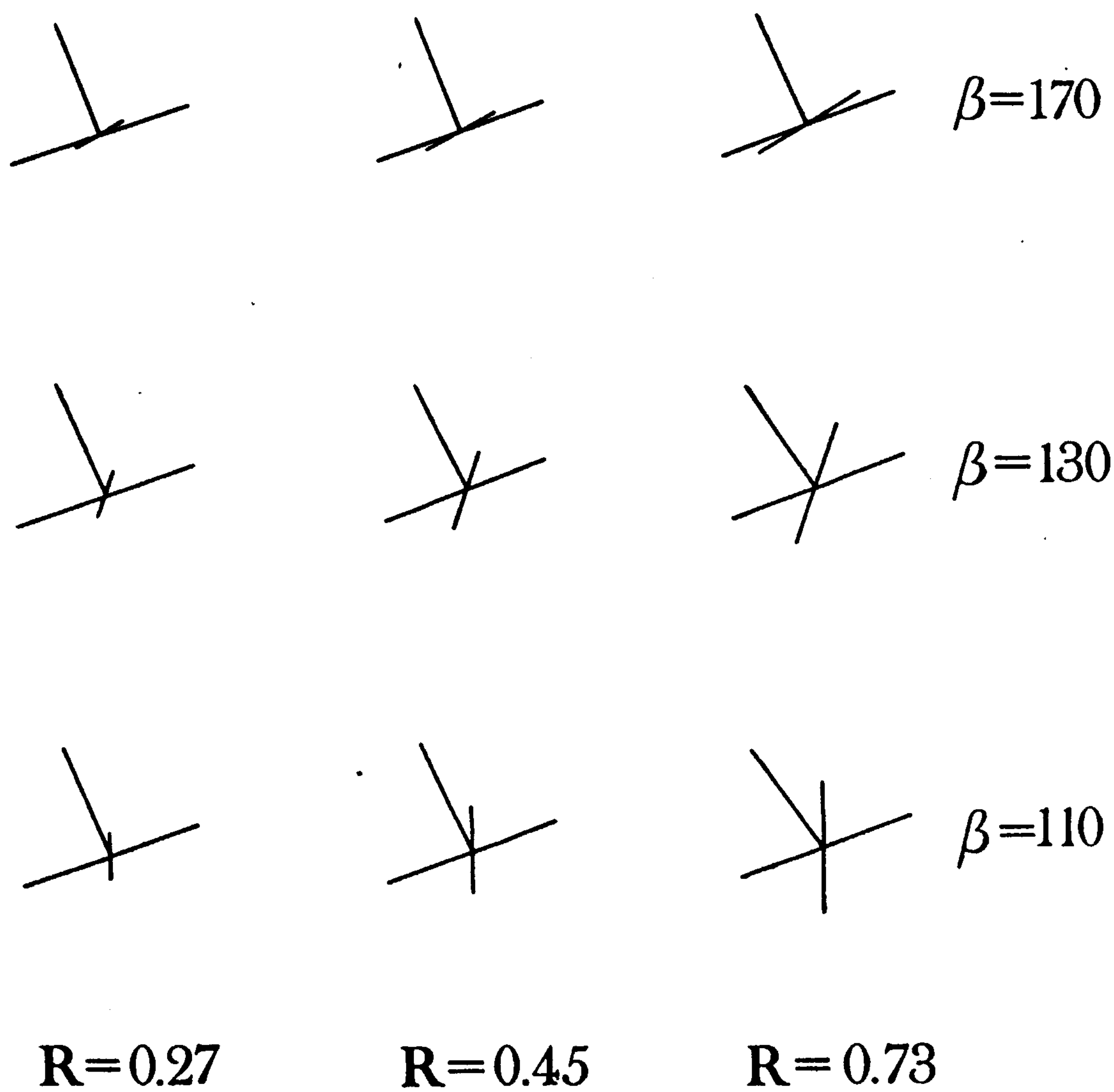


Figure A-2. These figures show the mean judgements of surface tilt as a function of relative line length R and angle of intersection β. Note that the apparent 3-D length of the normal will appear inappropriate for the configurations near the lower right. As a consequence, the line representing the image of the normal may appear overrotated counterclockwise in those cases. In the experiment, line representing the normal extended beyond the field of view, and these choices of tilt orientation appeared appropriate.

Consider the case where the vectors are assumed to be equal-length and orthogonal, however their actual lengths are unspecified. This case admits an exact solution to the surface orientation. Without loss of generality, have $u_x = 1$ and $u_y = 0$ (i.e., the image coordinate system is rotated so that the x axis is collinear with the image of the vector U, and the projected length is normalized to 1). Then the expression for the normal **N** is

$$\mathbf{N} = -u_z v_y \mathbf{i} + (u_z v_x - v_z)\mathbf{j} + v_y \mathbf{k} \qquad (A.1)$$

$$\mathbf{n} = -u_z v_y \mathbf{i} + (u_z v_x - v_z)\mathbf{j}. \qquad (A.2)$$

Since U and V are orthogonal, their dot product is zero

$$v_x + u_z v_z = 0. \qquad (A.3)$$

And since they are equal-length

$$1 + u_z^2 = v_x^2 + v_y^2 + v_z^2. \qquad (A.4)$$

Substituting v_z from (A.3) into (A.4)

$$1 + u_z^2 = v_x^2 + v_y^2 + v_x^2/u_z^2. \qquad (A.5)$$

Similarly, subsititute v_z from (A.3) into (A.2)

$$\mathbf{n} = -u_z v_y \mathbf{i} + (u_z v_x + v_x/u_z)\mathbf{j}$$

or

$$u_z \mathbf{n} = -u_z^2 v_y \mathbf{i} + (u_z^2 + 1)v_x \mathbf{j}. \qquad (A.6)$$

From (A.6) the tilt is expressed by

$$\tau = \tan^{-1}[(u_z^2 + 1)v_x / -u_z^2 v_y]. \qquad (A.7)$$

We have now to sol·e (A.5) for u_z^2. Note that this assumes that u_z is nonzero, i.e., that the vector **u** is foreshortened. If that were not the case, then trivially τ is 90 deg (perpendicular to **u**). Solving (A.5) for u_z^2 gives

$$u_z^2 = [(v_z^4 + v_x^2(2v_y^2 + 2) - 2v_y^2 + v_x^4 + 1)^{1/2} + v_x^2 + v_y^2 - 1]/2. \qquad (A.8)$$

Substituting (A.8) into (A.7) gives us the desired expression for the tilt τ.

Note further that from (A.3) we have that

$$v_z = -v_x/u_z.$$

Therefore u_z and v_z can be computed and therefore slant can also be computed from (A.1) by a similar process.

In conclusion, when the visual system is presented with well-defined lengths at a corner or intersection configuration, the angle of intersection is assumed to be a right angle, and the lengths are assumed equal. These two constraints are sufficient to admit a solution of local surface orientation up to a slant reflection, and, in fact, appear to be utilized by the human visual system.

APPENDIX B
SLANT RESOLUTION EXPERIMENTS

The internal form in which slant is represented was studied experimentally, by measuring lower-limit estimates of the internal precision to which slant is stored. While the resolution cannot be directly measured, the representation would have a grain of resolution no worse than the judgment variance. The apparatus should therefore provide the subject with excellent visual input, and yet the visual task must be solvable only by performing slant judgments. The magnitude of the variance as a function of slant angle was determined in order to argue the likelihood of a various forms for representing slant.

Three experiments were performed: The first examined various slants in the range $0 \leq \sigma \leq 44$ degrees, while holding tilt constant at 90 degrees (i.e., the surfaces were rotated about a horizontal axis). The second experiment examined the same range of slants, but with tilt held constant at 45 degrees. Finally, slant judgments for large slants ($60 \leq \sigma \leq 80$ degrees) were examined for constant tilt of 90 degrees. The conclusions of the three experiments are given in section B.5. The method was substantially the same in the three experiments, hence described in detail in the following

B.1 Experimental design

B.1.1 Apparatus

The experiment was designed to present a well illuminated and highly textured planar surface to a subject whose task was to match the slant of that surface by adjusting the slant of another surface. The two surfaces were placed so that they appeared adjacent in the visual field, however they differed considerably in distance. The distances to the fixation points of the two surfaces were 38 and 76 cm, the adjustable surface being the nearer. Both surfaces were viewed binocularly, however head movements were eliminated by using a chin rest. The Ss were instructed to compare the slants of the surfaces at fixation points marked on the surfaces. The line of sight to each fixation point was horizontal; the horizontal displacement required to shift gaze between the two fixation points was approximately 10 degrees.

Each surface rotated about a horizontal axis (i.e., the tilt was vertical), and the slant (angle between surface normal and the line of regard) was indicated by a protractor. The slant could be set and read with precision better than 1/2 degree. The adjustable surface was 15 cm (horizontal dimension) by 17 cm; the other surface was viewed through a 14 cm (horizontal dimension) by 9 cm opening in a barrier placed immediately in front of that surface. The opening served to occlude the boundaries of the surface being examined. The two surfaces had similar illumination.

The texture used in the first experiment was a gauze material with fine fibers, chosen to provide an excellent surface for stereo viewing. However a slight concern arose with that texture: The gauze provided linear markings oriented with the surface tilt that might have allowed judgments that did not require matching perceived slants, but simply the adjustment of the surface slant so that the linear markings on the two surfaces appeared parallel from various viewpoints. Although the chin rest prevented head movements, the separate monocular views from the two eyes might have been sufficient. Hence in the second and third experiments the surface texture had no linear markings: the surfaces were the commercially-available

Mecanormal "Normatone type 651" transfer pattern (a texture resembling the patterns on a giraffe).

B.1.2 Procedure

Each experiment consisted of multiple presentations of a randomized sequence of slants presented on the farther surface. The Ss were instructed to set the nearer, adjustable surface to the same slant as that presented, converging on their match by intentional over- and under-estimation. The Ss closed their eyes or averted their vision while the successive slant was adjusted for presentation. At the midpoint in the experiment the Ss were given a few-minute rest. The first sequence was used for training, and that data was not analyzed.

B.2 Experiment I

The first experiment measured slant judgments in three vicinities: near zero degrees, near ten degrees, and near forty degrees. Three slants were examined in each vicinity, differing by two degrees.

B.2.1 Method

Procedure: Four unpaid, volunteer, male subjects participated. Each had excellent vision, and found the task of matching slants to be natural and easy. The Ss were presented with nine slants: 0, 2, and 4 degrees, 10, 12 and 14, and 40, 42, and 44 degrees. The tilt was held constant at 90 degrees (the slants were achieved by rotations about a horizontal axis). The sequence of nine slants was presented seven times after the initial, trial sequence.

B.2.2 Results

The slant judgments for each S were analyzed separately. The means and standard deviations were computed for the seven trials at each slant (table B-1). The low standard deviations are notable. The slant judgments for similar slant angles, for each subject were compared to determine if the means for similar slants were significantly different, thereby providing another measure of our precision in performing slant judgments. For instance, the slant judgments at 10 and 12 degrees were compared to determine if their means differed significantly. It was found that for slants that differed by four degrees the means were significantly different ($p > 0.05$), except for subject KI where the difference in means at 40.0 and 44.0 degrees did not reach significance ($p > 0.10$, $t = 1.45$, d.f. $= 12$). The judgments of slants that differed by only two degrees differed significantly ($p > 0.05$) in roughly one third of the comparisons. For instance, the judgments for subject JH at 0.0 and 2.0 degrees of slant were not significantly different, but at 2.0 and 4.0 degrees the means differed significantly. Similarly, the judgments for subject SU between 12.0 and 14.0 degrees slant were significantly different, but those between 10.0 and 12.0 were not. There was a weak overall tendency for slants differing by two degrees to be less distinguishable at slant angles around 40 degrees than at smaller slant angles. The mean slant values and the means of the standard deviations are shown in table B-2.

B.3 Experiment II

This experiment was similar to the first experiment, but performed with the apparatus tilted 45 degrees (τ = 135 degrees).

Slant	Subject JH	Subject EM	Subject SU	Subject KI
0.0	1.21 (1.82)	-0.71 (1.15)	0.21 (1.38)	-0.43 (0.19)
2.0	2.93 (1.71)	1.89 (2.43)	2.40 (1.52)	0.18 (1.48)
4.0	4.83 (0.72)	3.61 (2.60)	4.14 (1.73)	2.93 (1.06)
10.0	11.46 (1.75)	9.07 (1.67)	12.43 (2.44)	8.83 (1.33)
12.0	11.21 (1.68)	9.76 (3.12)	14.64 (1.75)	10.14 (1.86)
14.0	15.57 (3.10)	13.37 (1.48)	16.79 (1.35)	11.11 (1.27)
40.0	37.79 (2.38)	37.87 (1.92)	39.93 (2.09)	41.79 (2.74)
42.0	38.86 (3.08)	37.76 (1.39)	41.11 (1.37)	42.64 (3.00)
44.0	41.11 (2.36)	39.57 (1.72)	42.43 (1.72)	43.50 (1.53)

Table B-1: Individual subject means (and standard deviations)

Slant	Mean (std. dev.)
0.0	0.07 (1.14)
2.0	1.85 (1.79)
4.0	3.88 (1.52)
10.0	10.45 (1.80)
12.0	11.44 (2.10)
14.0	14.21 (1.80)
40.0	39.34 (2.28)
42.0	40.09 (2.21)
44.0	41.65 (1.83)

Table B-2: Mean slant judgments, and mean subject standard deviations

B.3.1 Method

Procedure: Four unpaid, volunteer, male subjects participated (three of these participated in the first experiment also). The Ss were presented with randomized sequences of four slants: 0, 2, 42, and 44 degrees. Each S had a trial sequence followed by ten sequences for which data were collected.

B.3.2 Results

The means and standard deviations of slant judgments were computed separately for each S and each slant angle (table B-3). The slant judgments at a tilt of 45 degrees are not significantly different than those at tilt of 90 degrees from experiment I (neither the mean slant judgments, nor the means of the standard deviations of the judgments differed significantly by t-test). The second test was to determine for each S whether the mean judgments at zero and at two degrees slant were significantly different (similarly for 42 and 44 degrees slant). Only in two instances the means were not significantly different: for subject SU at 42 versus 44 degrees ($p > 0.1$, $t = 1.57$, d.f. $= 18$), and for subject DW between zero and two degrees ($p > 0.2$, $t = 1.17$, d.f. $= 18$). Otherwise, the judgments of slant differing only by two degrees were significantly different. The data collected at 45 degrees of tilt demonstrated no consistent underestimation or regression to the frontal plane.

B.4 Experiment III

The final experiment examined slants near 60 and 80 degrees. Tilt was 90 degrees.

B.4.1 Method

Procedure: Four unpaid, volunteer, male subjects participated (some were in the previous experiments). The slants were 60, 62, and 78, 80 degrees presented in seven trials in randomized sequence. The data from the first trial were not used.

B.4.2 Results

The data were analyzed in the same manner as in the previous two experiments, and presented in tables B-5 and B-6. Again there is no regression to the frontal plane; the judgments are accurate and have low variance. The standard deviations for slants near 80 degrees are slightly less than at 60 degrees, on the average: The most significant difference was between 60 and 78 degrees ($p < 0.10$, $t = 1.95$, d.f. $= 6$).

The individual judgments at 60 and 62 degrees were compared to see if the mean judgments were significantly different (similarly for 78 versus 80 degrees). Only for two subjects were the means insignificantly different (between 60 and 62 degrees: for subject KI ($p > 0.20$, $t = 1.34$, d.f. $= 10$) and for subject EM ($p > 0.05$, $t = 2.03$, d.f. $= 10$).

By now we have accumulated the standard deviations of slant judgments over a range of slants from zero to 80 degrees (see figure B-1). The mean value was 1.65 degrees.

B.5 Discussion

The experiments have demonstrated that slanted surfaces can be accurately aligned on the basis of visual information so that they are spatially parallel. The experimental design was such that the visual task of matching slant was probably achieved by comparing the perceived slants of the two surfaces, and matching

Slant	Subject DW	Subject EM	Subject SU	Subject KI
0.0	0.85 (0.91)	2.75 (1.32)	0.80 (1.01)	1.19 (1.60)
2.0	1.75 (2.26)	4.25 (1.53)	3.23 (1.25)	3.86 (1.53)
42.0	40.45 (2.79)	44.22 (2.91)	40.80 (1.23)	41.22 (1.56)
44.0	44.05 (1.77)	47.93 (2.41)	41.88 (1.78)	44.06 (2.11)

Table B-3: Individual subject means (and standard deviations)

Slant	Mean (std. dev.)
0.0	1.40 (1.21)
2.0	3.27 (1.64)
42.0	41.67 (2.12)
44.0	44.48 (2.02)

Table B-4: Mean slant judgments, and mean subject standard deviations

Slant	Subject DW	Subject EM	Subject MM	Subject KI
60.0	60.79 (1.49)	60.75 (1.86)	56.66 (0.75)	59.38 (2.12)
62.0	62.67 (0.52)	62.71 (1.44)	60.00 (1.52)	61.17 (2.48)
78.0	77.58 (0.74)	80.88 (1.00)	77.00 (0.84)	76.92 (1.20)
80.0	79.83 (0.61)	82.83 (1.08)	78.96 (1.31)	78.42 (1.07)

Table B-5: Individual subject means (and standard deviations)

Slant	Mean (std. dev.)
60.0	59.40 (1.56)
62.0	61.64 (1.49)
78.0	78.09 (0.94)
80.0	80.01 (1.02)

Table B-6: Mean slant judgments, and mean subject standard deviations

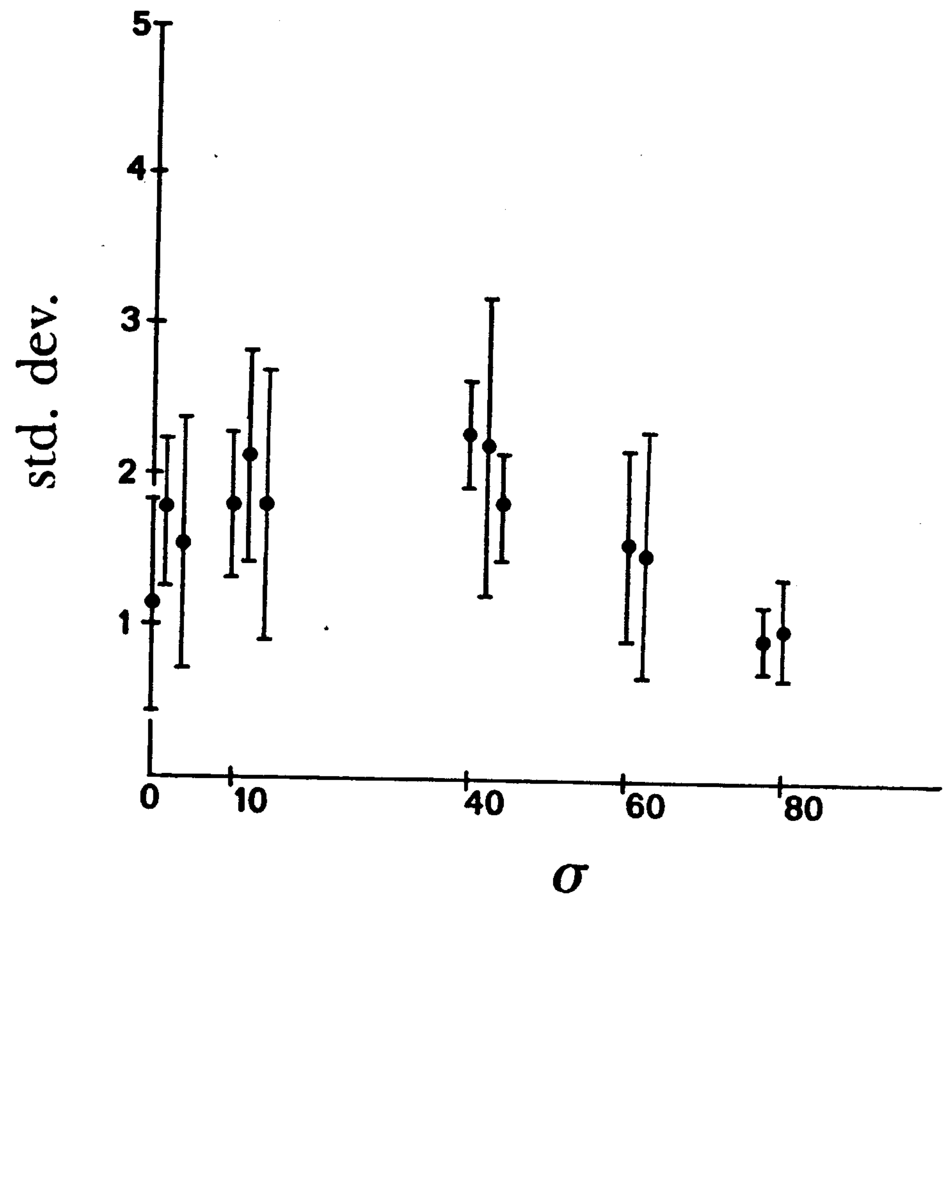


Figure B-1. The standard deviations of slant judgments were computed for each subject, for each slant angle. The averages across subjects are plotted above. Error bars show inter-subject variance (bar length = two standard deviations). The mean value was 1.65 degrees.

those values.

To reiterate, the two surfaces were adjacent in the visual field but differed considerably in distance. Head movement was not allowed, and the boundaries of the target surface were obscured (except for extreme slants where the top and bottom edges were visible but unlikely to be useful to the S since the dimensions of the two surfaces were different and the Ss never saw the overall dimensions of the surface whose slant was to be matched). The latter two experiments used surfaces that provided a rich texture for stereopsis but did not allow the simple aligning of texture edges so as to be parallel from both left and right eyes.

These experiments demonstrate that the visual system can match spatial orientations with precision, even when the distances to the surfaces are dissimilar. The average standard deviation is surprisingly small (1.65 degrees). Furthermore, for each S, the mean judgments of slant almost always differed significantly when the slants to be matched differed by only two degrees. These two results tell us something about the precision to which slant may be resolved, if the judgments indeed were based on comparing perceived slants: the grain of resolution in surface slant must at least as good as the precision in slant judgments, i.e., better than two degrees at all slants.

In what manner is slant represented (by angle σ, $\cos\sigma$, or $\tan\sigma$, for instance)? The cosine does not vary rapidly near zero degrees: cos (0 degrees) = 1.0000, cos (2 degrees) = 0.9994, cos (4 degrees) = 0.9976. Thus if slant were represented by $\cos\sigma$, an inordinately fine grain of resolution in the representation would be necessary to allow zero and four degrees of slant to be distinguished, let alone zero and two degrees of slant angle. On this basis, this form of representation is considered unlikely.

If the slant were represented by the tangent of the slant angle, then in order to resolve between slants around zero differing by a few degrees of slant angle (where tan (0 degrees) = 0.000, tan (2 degrees) = 0.0349, tan (4 degrees) = 0.0699) and simultaneously represent the range of slant angles from zero to 88 degrees (i.e., within two degrees resolution of 90 degrees slant), then the grain of resolution would have to be on the order of one part in eight hundred. Although this experiment does not resolve the question of how slant is represented, it probably allows us rule out the cosine and tangent forms. If slant angle were represented directly, the range of slants would be represented by less than one hundred resolvable values which (effectively) vary linearly with slant angle. The internal resolution would be commensurate with the measured j.n.d. of slant.

CS-TR Scanning Project
Document Control Form

Date : 3 / 14 / 96

Report # AI-TR-512

Each of the following should be identified by a checkmark:
Originating Department:

- [x] Artificial Intellegence Laboratory (AI)
- [ ] Laboratory for Computer Science (LCS)

Document Type:

- [x] Technical Report (TR)
- [ ] Technical Memo (TM)
- [ ] Other: ____________

Document Information

Number of pages: 122 (129-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
- [x] Single-sided or
- [ ] Double-sided

Intended to be printed as :
- [ ] Single-sided or
- [x] Double-sided

Print type:
- [ ] Typewriter
- [ ] Offset Press
- [ ] Laser Print
- [ ] InkJet Printer
- [ ] Unknown
- [x] Other: COPY MADE FROM XEROX

Check each if included with document:
- [x] DOD Form
- [x] Funding Agent Form
- [x] Cover Page
- [ ] Spine
- [ ] Printers Notes
- [ ] Photo negatives
- [ ] Other: ____________

Page Data:

Blank Pages (by page number): ____________

Photographs/Tonal Material (by page number): 38, 43, 51

Other (note description/page number):

Description : Page Number:

(A) IMAGE MAP! (1-122) UN#TITLE PAGE, 2-122
(123-129) SCANCONTROL, COVER, FUNDING AGENT, DOD, TRGT'S (3)

(B) SOME PAGES HAVE XEROXING MARKS.

Scanning Agent Signoff:

Date Received: 3 / 14 / 96 Date Scanned: 4 / 22 / 96 Date Returned: 4 / 25 / 96

Scanning Agent Signature: Michael W. Cook

Rev 9/94 DS/LCS Document Control Form cstrform.vsd

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AI-TR-512	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Surface Perception From Local Analysis of Texture and Contour		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kent A. Stevens		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0643
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE February 1980
		13. NUMBER OF PAGES 120
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Surface Perception
Texture
Texture Gradients
Depth Perception
Scene Analysis
Computer Vision
Surface Contours

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The visual analysis of surface shape from texture and surface contour is treated within a computational framework. The aim of this study is to determine valid constraints that are sufficient to allow surface orientation and distance (up to a multiplicative constant) to be computed from the image of surface texture and of surface contours. The report is in three parts.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)