AI-TR-589

SHAPE FROM CONTOUR

Andrew P. Witkin

NOVEMBER 1980

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

This blank page was inserted to preserve pagination.

SHAPE FROM CONTOUR

by

Andrew P. Witkin

Massachusetts Institute of Technology

November, 1980

Revised version of a dissertation submitted to the Department of Psychology on February 22, 1980 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

The problem of using image contours to infer the shapes and orientations of surfaces is treated as a problem of statistical estimation. The basis for solving this problem lies in an understanding of the geometry of contour formation, coupled with simple statistical models of the contour generating process. This approach is first applied to the special case of surfaces known to be planar. The distortion of contour shape imposed by projection is treated as a signal to be estimated, and variations of non-projective origin are treated as noise. The resulting method is then extended to the estimation of curved surfaces, and applied successfully to natural images. Next, the geometric treatment is further extended by relating contour curvature to surface curvature, using cast shadows as a model for contour generation. This geometric relation, combined with a statistical model, provides a measure of goodness-of-fit between a surface and an image contour. The goodnes-of-fit measure is applied to the problem of establishing registration between an image and a surface model. Finally, the statistical estimation strategy is experimentally compared to human perception of orientation: human observers' judgements of tilt correspond closely to the estimates produced by the planar strategy.

Contents

1. Introduction 5
   1.1 The problem 5
   1.2 Approach 6
   1.3 Outline 11
2. Estimating the orientation of planar surfaces 15
   2.1 Introduction 15
   2.2 Geometric model 16
   2.3 Statistical model 19
   2.4 Estimating surface orientation 21
   2.5 Implementation 25
   2.6 Avoiding failures of the estimation strategy 33
3. Extension to curved surfaces 42
   3.1 Introduction 42
   3.2 Extension of goodness-of-fit measures to curved surfaces 43
   3.3 Estimating orientation locally 44
   3.4 Orientation and scale 49
   3.5 Implementation and results 55
4. Using surface curvature 68
   4.1 Introduction 68
   4.2 Geometric model 69
   4.3 Statistical model 80
   4.4 Registration with a surface model: a demonstration 82
   4.5 Conclusions 87

References 89

A. Human perception of surface orientation 92

ACKNOWLEDGEMENTS—I gratefully acknowledge the support and advice of Whitman Richards, Shimon Ullman, Alex Pentland, among others, including my thesis committee, Berthold Horn and Ruzena Bajcsy. I also thank Eric Grimson and Mike Brady for their thorough analysis of some of my derivations. The work was supported by National Science Foundation and Air Force Office of Scientific Research under grant MCS79-23110, by the Advanced Research Projects Agency of the Department of Defense under the Office of Naval Research contract number N00014-75C-0643, and by a National Institutes of Health training grant 5T32GM07484.

CHAPTER 1

INTRODUCTION

1.1 The problem

The human perceiver is able to derive enormous amounts of information from the contours in an image, as evidenced by our ability to interpret line drawings. As part of this capacity, we are able to use the shapes of image contours to infer the shapes and dispositions in space of the surfaces they lie on. To the extent the inferences we draw are accurate, our strategies for drawing them must have some basis in the character of the visual world, just as the efficacy of stereopsis as a source of depth information has a basis in the geometry of projection and triangulation. The aim of the research reported in this thesis is (1) to discover constraints on the the visual world that allow surface shape to be reliably inferred from image contours; (2) to derive methods of inference from those constraints; and (3) to apply those methods to natural images. The inference of surface shape will be treated as a problem of statistical estimation, combining constraints from projective geometry with simple statistical models of the processes by which contours are formed.

The interpretation of contours falls into three sub-problems:

Locating contours. If contours are to be used to infer *anything*, they must be found. The human perceiver has little difficulty deciding what is and is not a contour, yet the automatic detection of edges has proved extremely difficult (see, e.g., Falk, 1972; Zucker, Hummel, & Rosenfeld, 1977). Perhaps this difficulty should not be surprising: the contours we see in natural images usually correspond to definite physical events, such as shadows or discontinuities of depth. Our ability to detect these events may say more about their significance for image interpretation than about their ease of detection. Why should we expect events that have simple descriptions in terms of the structure of the scene to have simple descriptions in terms of changing image intensity as well? If the physical significance of contours is taken as their primary feature, then at least we know *what* is being detected, even if we don't know *how*.

Labeling contours. If contours correspond to such diverse physical events as shadows, and discontinuities of depth or orientation, then an essential component of their interpretation must be to decide which contours denote which events, because each kind of contour imparts a different mean-

ing. The work of Clowes (1971) Huffman (1971), and Waltz (1975) have shown that strong structural constraints can be applied to distinguish one kind of contour from another. Horn (1977a) has related characteristic intensity profiles to physical contour types.

Interpreting contours. Even after contours have been found and labeled, not much is known about the physical structure of the scene. It is clear that contours play a role in the human perceiver's ability to decide where things are and what they're shaped like, apart from the application of specific "higher level" knowledge to objects of known shape. This ability must have some basis in real properties of the world, yet that basis is not known.

This proposal addresses the third problem, with emphasis on contours of reflectance and illumination, that is, surface markings and cast shadows. The problem, restated, is: given the contours in an image, infer the shapes and orientations of the surfaces in the scene.

1.2 Approach: image understanding as the application of knowledge

1.2.1 The perceiver and the world

The most interesting property of the human visual capacity is that it *works:* we rely on vision, more than any other sense, to inform us of our immediate surroundings; and the information it gives us is, with respect to our needs, remarkably complete and accurate. The magnitude of this accomplishment, because it is so familiar, often escapes us; but it is truly remarkable how well the small amount of light that our eyes capture informs us of the objects from which it has been reflected. The connection between the light we receive and the world we perceive is not obvious, as evidenced by the slow progress of attempts to mimic the human capacity; yet we are living proof that that connection exists—our visual capacity embodies it.

The accuracy of vision implies a special relation between our visual capacity and the world in which it functions: to the extent vision draws true conclusions about the world, its means of doing so must rest on true premises about the world. Biological vision provides an existence proof that such premises, and means of drawing conclusions from them, exist. But to know that they exist is not to know what they are or how they're used. To obtain such knowledge is an empirical task, and a difficult one.

To understand *why* a visual system works is to understand the relation between that system and its world. And part of understanding that relation is understanding the world in its relevant aspects. If we knew the workings of a visual system in full detail, we would know *what* that system did, and *how.* In a sense, we would know what conclusions the system drew from the evidence given it. But to understand *why* this particular "what" and "how" lead to *true conclusions about the world* we would also have to know something of the world — namely, the valid premises about the world from which those conclusions follow. It follows that understanding the behavior or construction of a visual system is not sufficient for understanding the system's ability to draw *valid* conclusions, to the extent it does,

about its world.

Conversely, the truth or falsehood of premises about the world, and the ability or inability of those premises to support arguments from evidence of a particular kind to conclusions of a particular kind, do not turn on those premises being embodied in any particular visual system: the truth or falsehood of premises about the world must be evaluated against the world, not against systems that might use those premises. Thus, understanding the behavior or construction of any particular visual system is not *necessary* to the discovery of valid premises, or to the evaluation of methods that follow from them. Of course, to understand the basis in the world for some particular system's effectiveness, it is necessary to understand both the system and its world.

The notion that perception has a basis in real properties of the world is an old one, evident for example in the "natural geometry" of Descartes (1637). More recently, this view is reflected in J.J. Gibson's *ecological optics* (1966). But Gibson insisted that the solutions to perceptual problems reside not in knowledge brought to bear on the image, but *in the image itself.* In consequence, the assumptions about the world on which his solutions rested were never explicitly stated nor critically examined.

The construction of artificial vision systems which emulate the human capacity, or aspects of it, entails solving the same problems that were solved in evolution, since biological and artificial systems must operate on the same world. Hence the study of the world from the standpoint of solving perceptual problems is relevant to both domains. This focus is most evident in the work of Horn (in Winston, 1975), Land & McCann (1971), Marr (1976, 1979), Ullman (1979), and Barrow & Tenenbaum (1978). In particular, the work of Marr and of Ullman explicitly treats both the biological and artificial domains.

1.2.2 Understanding image formation: two uncertainties

Of the sorts of knowledge that might pertain to the interpretation of images, one kind stands out by virtue of its accessibility and its obvious relevance to the problem: knowledge of the process of image formation itself. This knowledge, expressed in the equations that describe the transmission, reflection, and projection of light, is sufficient to *synthesize* an image from a model of a scene (see, e.g., Newman & Sproul, 1979), but is notoriously insufficient to recover a scene given its image. While these equations determine one unique image for each scene, they allow an infinity of scenes for each image. That is, the mapping they specify from scenes to images is many-to-one.[1]

The ambiguity in the mapping from images to scenes specified by the imaging equations reduces to two fundamental uncertainties. The first of these is *photometric:* the amount of light reaching a point in the image, having been reflected from a surface, depends on the light-reflecting properties of the surface, on its orientation in space with respect to the viewer, and on the light incident upon it (Horn, 1975). These three components may combine in an infinity of ways to produce any given image intensity, so given the intensity alone, the equations can't be solved. The second uncertainty is

[1]Since we tend to think of the laws of image formation as the "real, hard facts," and other constraints as somehow softer or less real, the ambiguity in the imaging equations are often regarded simply as the ambiguity of images. This is not strictly true: if the imaging equations were unknown, images would be even more ambiguous. If more were known, they might be less so. The ambiguity resides not in the image itself, but in what we *know* about the image.

geometric: in the process of projection, a line may project to a point; and knowing the position in the image of a point's projection only constrains that point to lie on a line in space.

It is these two uncertainties that make the recovery of surfaces' photometric and geometric properties so difficult: the imaging equations are known, and obviously relevant, but they are not sufficient. Something more is required, and finding that elusive "something more" is the heart of the problem.

1.2.3 Background

Using the photometric relation. Understanding the imaging process, while insufficient, is essential to interpreting images. A notable example of the value of this understanding is Horn's (1975) treatment of the inference of surface shape from shading information. In a constrained situation, where the direction of illumination, and the surface's light-reflecting properties are known, the photometric ambiguity can be overcome, and the shape of a smooth surface recovered, by integrating from a curve along which surface orientation is known.

The structure of Horn's solution is particularly instructive: its basis lies in the dependence of image intensity on surface orientation, as described by the photometric equation. This lawful dependence is not by itself sufficient to recover surface orientation from the image, because illumination and surface reflectance, which are also unknown, appear in the equation as well. To solve for surface orientation, additional constraints must be brought to bear, and these constraints must meet two conditions: they must be sufficiently powerful to determine a unique solution, and they must be true. The first condition is purely formal, but the second is empirical: if the assumptions from which a solution logically follows are wrong, one can hardly expect the solution to be right. It is not difficult to find assumptions that meet the formal condition, nor to find assumptions that meet the empirical condition; the hard part is meeting both simultaneously.

The constraints on which Horn's solution is based—known illumination and reflectivity, surface smoothness, a curve of known orientation—are not always met, but the solution is useful, first, because the constraints *are* met in various situations of practical interest, and second, because the solution is a starting point from which less stringent sufficient conditions can be sought (Ikeuchi & Horn, 1979).

Horn's use of the photometric equation provides the model on which the present work is based: that equation relates quantities that can be measured in the image—intensity, in this case—to quantities that are to be recovered—surface orientation—and quantities that may not be of direct interest, such as the direction of illumination. To solve the problem, these relations must be "untangled" using valid constraints, and the quantities of interest isolated.

constraints that are valid in their context must be brought to bear, that permit this relation to be "decomposed," and the quantities of interest isolated.

Using the geometric relation. Geometric properties of the image, like photometric ones, depend lawfully on properties of the scene: the distance between two points in space is related to the

distance between their images by the projective transformation, and all *metric* properties[2] are likewise related. One may further distinguish *foreshortening*, the effect on projected distance of inclination from the image plane; and *perspective*, the effect on projected distance of distance from the image plane. The latter is absent in *orthographic*, or parallel, projection. As in the photometric case, the projective relation alone is not sufficient to recover the metric properties of the scene from the metric properties of the image.

Of the many possible formulations of the projective relation, the most useful for the description of surfaces relate surface orientation with respect to the viewer, metric properties on the surface, and the corresponding metric properties in the image. The first application of such a formulation to the recovery of surface orientation from images was by J. J. Gibson (1950a, 1950b, 1966), who proposed that the texture gradient—the rate of change with respect to position in the image of the distance between adjacent texture elements—specifies the slant, or orientation, of the textured surface, by virtue of the diminution of projected size with increasing distance. This theme has since been pursued extensively (Purdy, 1960; Bajcsy, 1972; Haber & Hershenson, 1973; Rosinski, 1974; Bajcsy & Lieberman, 1976; Kender, 1978; Stevens, 1978).

While Gibson recognized the importance of geometric constraints for understanding surface perception, his argument is seriously flawed by the failure to make explicit assumptions on which his conclusions critically rest. It is readily shown that the projective relation alone is severely lacking as a justification for Gibson's claim that the texture gradient specifies the actual slant of the surface,[3] yet no other justification was offered to support that claim:

Having located a texture element at a point in the image, the corresponding surface point is constrained by projection to lie on the line that contains the image point and the optical focal point; and projection imposes no other constraint. One can imagine the line to be a straight wire in space, extending from the image, and the surface point to be a bead on that wire, in which case the projective constraint allows the bead to slide freely along the wire. By extension, a collection of texture elements may be imagined as a bundle of wires, each with a freely sliding bead. And as long as each bead remains on its respective wire, the projective constraint is guaranteed to be met. If all of the beads are presumed to lie on some surface, then it is clear that the projective constraint, by itself, tells us nothing about the surface's shape or orientation: for any surface we construct, as long as it intersects each wire, it is always possible to arrange the beads so they all lie on the surface.

In fact, the relation drawn by Gibson between the texture gradient and surface slant only holds if the textured surface is planar (Stevens, 1978), and the texture elements on the surface have exactly equal spacing. That is, the variation of texture observed in the image depends on the variation of the texture on the surface, and the curvature of the surface, as well as on slant. Yet Gibson has assumed that the first two contributions are absent, and that the observed variation derives entirely from slant. Subsequent work has largely accepted this premise. While it is obvious that this assumption will seldom hold exactly in the natural world, I am not at this point arguing for or against its validity. The point is that, since the assumptions were not made explicit, their validity, even as useful idealizations,

[2]Distance and properties that depend on distance, such as orientation and curvature.

[3]This must be distinguished from the claim that the slant specified by the texture gradient corresponds to the slant perceived by human observers.

was never even addressed. Since image understanding is an empirical endeavor, the value of any method turns on the validity of the assumptions it entails.

A quite different use of the geometric relation between image and scene is Ullman's (1979) treatment of the recovery of structure from motion: given the orthographic projections of a set of moving points, Ullman addressed the problem of determining whether the motions are consistent with an interpretation of the points as rigidly connected, and, if so, recovering their three-dimensional motion and spatial relations uniquely. Ullman found that three views of four non-coplanar points are in general sufficient to recover a unique rigid interpretation, if one exists, or else to determine that the points are not in rigid motion.

Of perhaps more interest than the purely geometric finding itself is the power it assumes, when linked to a very simple assumption about visual scenes, by Ullman's *"rigidity hypothesis:"* if a motion in the image can be given a unique interpretation as a rigid motion in space, that interpretation is correct. With rewording it becomes apparent that the basis for this assumption is ultimately a *statistical* claim about the world: rigid connection is sufficiently common that the existence of a unique rigid interpretation is far more likely to arise from actual rigid motion, than from the accidental alignment of independent motions. In other words, it is possible that the projection of a chance alignment will be indistinguishable from the projection of a rigid motion, within one's tolerance of measurement, but it is extremely unlikely. One could in principle reformulate the rigidity hypothesis in terms of the error distribution of the image measurements, and the likelihood of rigid connection, to specify a *most likely* rigid interpretation, and evaluate its likelihood. However, the rigidity hypothesis is so strong that those likelihoods are divided into near certainties and near impossibilities, so an explicit statistical treatment is superfluous. Although the statistical character of the rigidity hypothesis is perhaps obscured by its strength, we see in this instance a hint of the potential power of a coupling of geometric constraints with simple statistical assumptions.

1.2.4 Image interpretation as a statistical problem

The importance of understanding image formation is well established. A great gap separates the firm but limited and insufficient foundation provided by image formation from the astonishing level of performance attained by the human perceiver. But how should this gap be reduced? To solve substantial image understanding problems in simple and general ways, i.e. without recourse to very specific "higher level" knowledge, constraints must be discovered that are powerful enough to determine solutions, and valid enough, over a broad range of situations, to determine the *right* solutions. Yet the variability and irregularity of the world seems to preclude the existence of formally sufficient constraints that are valid in the lawful, exceptionless sense of the imaging equations. For example, an assumption that textures are uniformly spaced before projection may quite often approximate reality reasonably well, but it will never be strictly true, and will often err seriously. Any categorical assumption of this kind, although it might in the long run provide a reasonable idealization, may in any given instance be very, very wrong. In short, the basic uncertainties intrinsic to the imaging equations may be attenuated by additional constraints, but they cannot be eliminated. Whatever assumptions one adopts, one's interpretations will sometimes be wrong.

If uncertainty cannot be avoided, then it ought at least to be recognized, and understood as nearly as possible. If one's assumptions can't always be true, then the best one can hope to do is to somehow distinguish the cases in which they hold from those in which they fail, and at least avoid acting on false assumptions. Ullman's rigidity hypothesis clearly illustrates this point: if one assumes that an observed set of points are rigidly connected one can under appropriate conditions recover their structure and motion in three dimensions. The rigidity assumption is sometimes valid, sometimes not. But assuming rigidity incorrectly almost always leads to an identifiable contradiction, and the absence of contradiction almost surely means the rigidity assumption is valid. When contradiction is detected, the rigidity assumption is rejected as inconsistent, and nothing is inferred about the structure and motion of the points, but otherwise the recovered structure and motion are almost certain to be correct. Thus, the strategy almost always yields either a correct result or no result at all. While a categorical rigidity assumption would hardly be valid, the same assumption becomes extremely powerful when made conditional on the outcome of applying it. By conditioning an assumption on its consequences, the assumption's uncertainty can be made *internal* to the interpretation strategy.

Recall that the rigidity hypothesis, by which the uncertainty of a rigidity assumption was incorporated into the structure-from-motion strategy, may be cast in statistical terms. This ought not to be surprising, because statistics is exactly that branch of mathematics whose purpose in application is to deal with uncertainty intelligently, and to minimize its damaging effects Statistical methods offer the best tools that have been devised for this purpose. If uncertainty is intrinsic to image interpretation, then it stands to reason that the application of statistical methods to image interpretation bears investigation. Yet, surprisingly, with the exception of an early and isolated effort by Brunswik (1948), the recovery of the physical structure of scenes from images has never been attempted by statistical means.

The application of statistical assumptions figures prominently in the work to be reported in this thesis. While the resulting methods do not eliminate the fundamental uncertainty of image interpretation, they offer the important advantage of yielding not only interpretations, but also measures of the confidence to be attached to those interpretations. Although the interpretations are not always accurate, it is therefore usually possible to distinguish the accurate from the inaccurate ones. The methods to be presented demonstrate the value of this approach.

1.3 Outline

The primary geometric basis for the methods to be presented is the *foreshortening* relation—the dependence of metric properties in the image on surface orientation. In consequence, orthographic projection will be used throughout, although some perspective effects will be treated briefly. The *distortion* of metric properties by projection will be treated, quite literally, as a signal, and the metric properties themselves, as noise. Since that distortion *depends* on surface orientation, it in a sense *encodes* surface orientation. To estimate the parameters of the distortion is to estimate the orientation of the surface.

The entities to be examined in the image are *contours*, curves in the image that correspond at least roughly to significant physical events on the surface. Attention will be limited to contours cor-

responding to surface markings and cast shadows. The local metric properties of curves are naturally described by the tangent, and the direction of the tangents along the image contours will be the primary measure on the image. To use the observed tangents to estimate surface orientation requires a statistical model of the process by which the contour generator[4] was placed on the surface.

1.3.1 Estimating the orientation of planar surfaces

The estimation problem is first considered subject to the artificial restriction that the surface is known to be planar. While not realistic, this limited case provides the groundwork from which more general methods will be developed.

First, the relation between surface orientation, tangent direction on the surface, and tangent direction in the image, is expressed geometrically. This expression relates the image quantities that can be measured to the scene quantities that are to be recovered.

Second, the contour generating process is given a simple statistical characterization: surface orientation and tangent direction on the surface are isotropic and independent. That is, all surface orientations, and all tangent directions on the surface, are assumed equally likely.

Together, the geometric and statistical models specify a probability density function for surface orientation, given a set of image measurements. This function is derived. The surface orientation value at which this function assumes a maximum is the maximum likelihood estimate for surface orientation, given the model. And the integral of the function over a range of surface orientations is the probability that the actual orientation lies in that range.

The estimator is first applied to geographic contours: projections of coastlines drawn from a digitized world map. This choice of data circumvents the problem of contour detection, and allows the actual orientation to be precisely controlled. The overall accord between estimated and actual orientation is excellent, and, equally important, the confidence measures generated by the estimator effectively distinguish the accurate estimates from the inaccurate ones.

The same technique is then applied to natural images, using zero-crossing contours in the convolution of the image with a $\nabla^2 G$ function, as described in Marr & Poggio (1978) and Marr & Hildreth (1979). While the veridical orientations were not independently measured, the maximum likelihood estimates are in close accord with the perceived orientations.

Methods are then considered for avoiding failures of the estimation strategy that arise from failures of the premises on which it is based, as distinct from sampling errors. In particular, the dependence of the image measures poses a potential problem. This problem can be overcome in part by judicious sampling of the image data, but must in part resort for its solution to independent measures of orientation, as may come from perspective effects, or from photometric information.

1.3.2 Extension to curved surfaces

The planar method is then extended to the estimation of curved surfaces. First it is shown that the

[4]the curve on the surface, of which the contour is a projection

estimator can be applied globally, but only if strong prior restrictions are placed on the surface. In the general case, when such restrictions are not available, local estimation is more appropriate.

To apply the methods developed in the planar case to curved surfaces without additional assumptions, it would be necessary to obtain *at each point in the image* a measure of the distribution of tangent directions. But such a local measure is never available, because the density of the contour data is limited. On the other hand, a distribution can be taken at each point of the data in a surrounding region, as small as possible, but large enough to provide a reasonable sample. This spatially extended distribution may be represented as a three dimensional convolution of the image data with a summation function.

To understand how such a distribution should be applied to estimate surface disposition, the meaning of surface orientation is considered in some detail. It is argued that surface orientation is not a unique property of the surface, but must be regarded as a function of *scale*. The scale at which orientation is described corresponds to the spatial extent over which it is measured. Thus, by measuring orientation over a large extent, the surface is described at a coarse scale. We may thus expect the *scale* at which the surface is estimated to depend on the spatial extent over which the distribution is computed. Since that extent must be sufficiently large compared to the density of the data, the density effectively limits the spatial resolution of the estimate. It is shown that this strategy closely parallels Horn's (1975) method for inferring shape from shading. The local estimator for orientation is a geometric analogue to the photometric reflectivity function.

The strategy was implemented, and applied to natural images. Contours were extracted as in the planar case, and the spatially extended distribution approximated by a series of two-dimensional convolutions with a "pillbox" mask. The estimated surfaces were in close accord with those perceived by the human observer. The effect on the estimate of varying the mask size was investigated.

1.3.3 Using surface curvature

The above method estimates curved surfaces, but never treats surface curvature explicitly. But, just as the tangent direction of a contour encodes surface orientation, the curvature of a contour encodes surface curvature. This encoding is investigated for cast-shadow contours. The shape of the image contour is shown to depend on the shape of the shadowed surface, the shape of the shadowing object, and their geometric relation to the light source and the viewer. An expression is obtained relating the tangent and curvature of the image contour to the orientation and curvature of the shadowed surface. This relation also depends on the illuminant and the shadowing object.

This geometric relation may be applied to surface estimation by extension of the statistical logic applied to the previous cases: the distortion imposed by surface curvature and orientation is regarded as a signal whose parameters must be estimated from the image data. The problem is simplified because the curvature and orientation of the surface can be safely assumed to be independent of the properties of the light source and the shadowing object.

A limited estimation problem is addressed: using only the shapes of shadow contours, establish registration between an image and a surface model, when nothing is known of the object casting the shadow. Images were synthesized by casting irregular shadows on a digital terrain model (DTM). It is

shown that registration can be established, using the geometric relation and a simple cross-correlation technique, to within a few DTM pixels, even with sparse image data.

1.3.4 Relation to human perception

A psychophysical experiment is reported, whose aim is to examine the relation between human observers' judgments of the orientations of curves perceived as planar, and the estimates obtained from the estimation strategy outlined above.

A series of "random" curves were generated using a function with randomly chosen parameters. Although such curves have no "real" orientation outside the picture plane, they often appear slanted in space. Observer's judgments of orientation were obtained by matching to a simple probe shape. The judgments of tilt (direction of steepest descent from the viewer) were highly consistent across observers, while the slant judgments (rate of descent) were much more variable.

Orientation estimates for the same shapes were computed using the planar estimator, and these estimates proved to be in close accord with those of the human observers, although the shapes had no "real" orientation.

While no conclusion is drawn about the mechanism by which human observers judge orientation from contours, or about the measures they take on the image, this result provides evidence that the human strategy, and the one developed on geometric and statistical grounds, are at the least close computational relatives.

CHAPTER 2

ESTIMATING THE ORIENTATION OF PLANAR SURFACES

2.1 Introduction

This chapter addresses a relatively simple problem: given a collection of image contours, which are known to be projections of curves on a planar surface, estimate the orientation of the surface in space. Because it is relatively simple, this problem provides an appropriate introduction to the statistical approach. Moreover, the methods developed to solve the problem are useful: even though the restriction to planar surfaces is artificial, a more general solution will be presented later as a direct extension of the planar case. Following is a summary of the chapter:

Geometric model. Although projective geometry alone does not solve the problem, it is essential to understand the geometric relation between the quantities measured in the image, and the scene properties we wish to recover. As a first step, a geometric expression is obtained that relates the tangent angle along the image contours to the tangent along the corresponding curve in the scene, and the orientation of the surface on which that curve lies. Because the image tangent depends in part on surface orientation, this expression can be used to infer surface orientation, in conjunction with additional constraints.

Statistical model. Next, a statistical model of the scene parameters will be introduced. As a simple idealization, surface orientation and tangent direction in the scene will be assumed isotropic and independent. In practice, we might want to bias this joint distribution to reflect anisotropies imposed by gravity. In any event, while the results depend quantitatively on the form of the distribution, the method can be applied to any chosen distribution.

As a further simplification, the tangent measures taken on the image will be assumed independent. That is, we view the measured tangents as the projected orientations of a collection of needles thrown randomly on the surface. This last assumption is not realistic, and leads to unacceptable consequences in some situations. Problems arising from this assumption will be treated later in the chapter.

Estimation. Together, the geometric and statistical models determine a maximum likelihood estimator for surface orientation, because the geometric model allows the statistical assumptions about the scene to be carried into the image. A density function for the *image* tangent angle, given surface

orientation, will be derived. This last density function in turn determines a density function for surface orientation, given the tangents measured in the image. This function provides a maximum likelihood estimate for surface orientation, with confidence intervals.

Implementation. The theoretical analysis will next be moved into application. Since we know exactly what we want to compute, the implementation of the estimator is straightforward. The more substantial problem is the extraction of contours on which to apply the strategy. This complication was avoided in an initial test, by applying the strategy to geographic contours: coastlines of lakes and islands drawn from a digitized world-map. Then, the strategy was applied to natural images using zero-crossing contours in the convolution of the image with a $\nabla^2 G$ function (Marr & Hildreth, 1979). The performance of the strategy on these domains is evaluated, and shown to lead to useful estimates of surface orientation. Furthermore, the confidence information generated by the statistic effectively distinguishes reliable from unreliable estimates.

Analysis of failures. Finally, conditions under which the statistical model might fail are considered. The most serious problems arise from failures of the assumption that the image data are independent. Some of these failures can be avoided by appropriate sampling of the data; others cannot be detected except by reference to independent estimates of surface orientation.

2.2 Geometric model

The tangent to a curve at a given point is defined as the first derivative of position on the curve with respect to arc length. The tangent is a unit vector, and may be visualized as an arrow that just grazes the curve at the specified point. The problem, as defined, is to estimate the orientation of a surface, given the tangent along an image contour which is the projection of a curve on that surface. The task of a geometric model is to express the functional relationship between the quantities to be estimated and the quantities that are measured. In this section, the tangent direction at a point on an orthographically projected curve will be expressed as a function of the orientation of the plane in which the corresponding space curve lies, and of the tangent direction in that plane.

2.2.1 Notation and terminology

Vector quantities will be denoted in boldface (e.g. $\mathbf{X}$, $\mathbf{Y}$), and angles by lower case Greek letters (e.g. α, β.) The components of vectors will be given in brackets (e.g. $\mathbf{X} = [1, 0, 0]$.) Projected quantities will be denoted by the same symbol as their unprojected counterparts, with a "$*$" superscript, e.g. the projection of a vector $\mathbf{X}$ is denoted by $\mathbf{X}^*$.

The objects with which we will deal are: an image plane, $\mathbf{I}$; a surface, $\mathbf{S}$, in space, which we assume to be planar; a curve, $\mathbf{C}(s)$ on $\mathbf{S}$, which, following Marr (), we call a *contour generator;* a curve in the image, $\mathbf{C}^*(s)$, which is the orthographic projection of $\mathbf{C}(s)$ onto $\mathbf{I}$.

The orientation of $\mathbf{S}$ with respect to $\mathbf{I}$ may be denoted by two angles σ and τ (for *slant* and *tilt* respectively,) with σ the angle between $\mathbf{I}$ and $\mathbf{S}$, and τ the angle between the projection of $\mathbf{S}$'s normal

onto I, and the x-axis in I. That is, σ says how much S is slanted, while τ says which way (See fig. 1). The direction of a contour generator $\mathbf{C}(s)$'s tangent at a point s will be denoted by $\beta(s)$, where β is the angle between the tangent, and a fixed coordinate axis in S.

2.2.2 The projected tangent angle

If the orientation of the surface, S, with respect to the image normal, I is given by (σ, τ), then I may be taken into S by a rotation by (σ, τ). Therefore, the projection of a curve in S onto I may be obtained by placing the curve in I, rotating it by (σ, τ), and projecting it back onto I. The rotated coordinate axes of I, (x, y), can be taken as the axes for S, and the tangent angle β measured with respect to the rotated x-axis. Then the tangent to $\mathbf{C}(s)$ in those coordinates is $[\cos\beta, \sin\beta]$.

It will be convenient for the moment to let I's x-axis coincide with the tilt direction, so that $\tau = 0$. In that case, the equations for rotation by (σ, τ) of a point $(x, y, 0)$ into (x', y', z') reduce to

$$\begin{aligned} x' &= x\cos\sigma \\ y' &= y \\ z' &= x\sin\sigma \end{aligned}$$

and the orthographic projection of (x', y', z') onto I is just (x', y'). So the tangent vector $\mathbf{t} = [\cos\beta, \sin\beta]$ becomes, after rotation and projection

$$\mathbf{t}^* = [\cos\beta\cos\sigma, \sin\beta]$$

(which is not in general a unit vector.) The projected tangent angle β^* is the angle between this vector and the x-axis, whose tangent is given by

$$\tan\beta^* = \frac{\tan\beta}{\cos\sigma}$$

so that

$$\beta^* = \tan^{-1}\left(\frac{\tan\beta}{\cos\sigma}\right)$$

To reintroduce τ, suppose we now pick arbitrary coordinate axes for I, and define α^* the angle between the x-axis in the image, and the projected tangent. Since β^* is the angle between the projected tangent and the tilt direction, we have

$$\beta^* = \alpha^* - \tau$$

and

$$\alpha^* = \tan^{-1}\left(\frac{\tan\beta}{\cos\sigma}\right) + \tau \qquad (2\text{–}1)$$

where α^* is the projected tangent angle, β is the angle between the unprojected tangent and the tilt direction's projection onto S, and (σ, τ) is the orientation of the curve in space. This expression relates α^*, which can be measured in the image, to (σ, τ), which we wish to recover, which is what we sought.

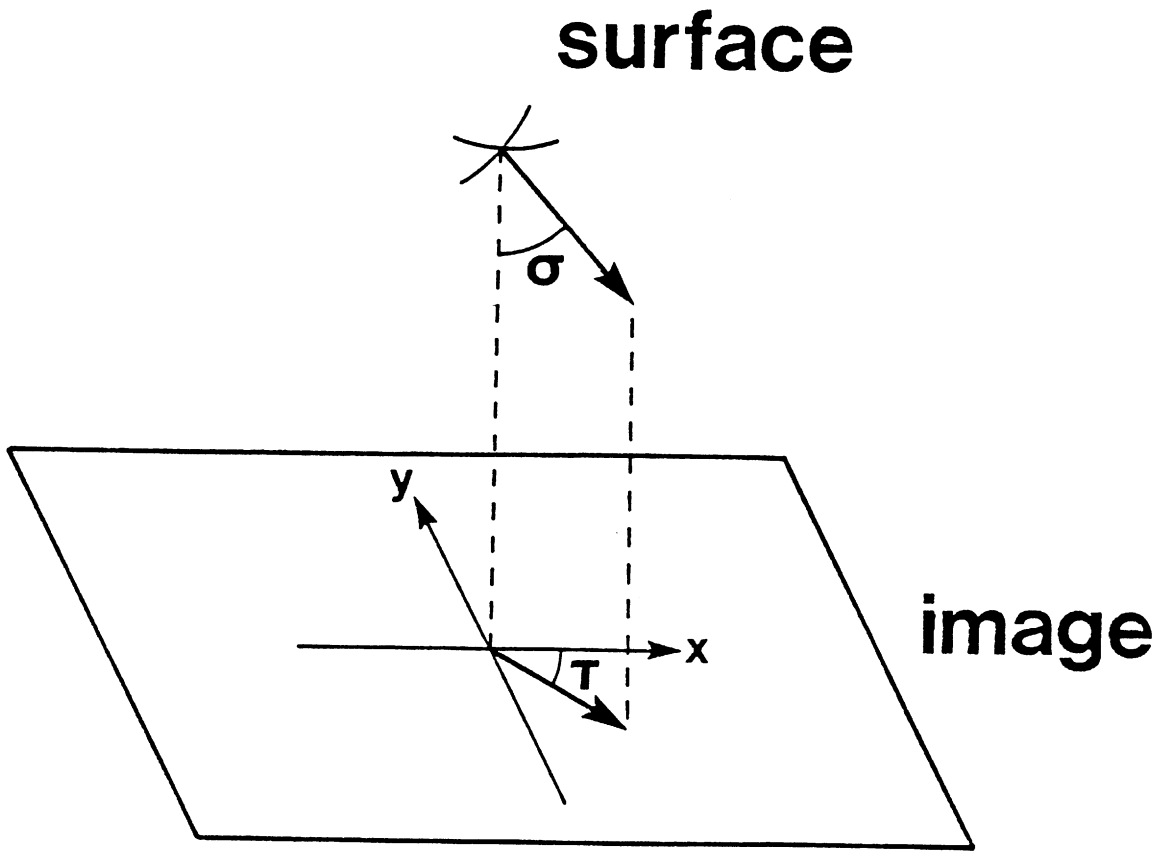

Figure 1. Representing surface orientation by slant (σ) and tilt (τ): Slant is the angle between a normal to the surface and a normal to the image plane. Tilt is the angle between the surface normal's projection in the image plane, and a fixed coordinate axis in that plane.

2.3 Statistical model: isotropy and independence

We now have a geometric relation between the measured tangents along an image contour, and their counterparts in space, in terms of the orientation (σ, τ) of the plane of the contour generator. But the measurable quantity α^* depends on three unknown parameters (β, σ, τ), so we can't solve for (σ, τ). As expected, the geometry alone does not admit a solution for surface orientation given only measurements of the tangents along image contours.

The problem, then, is to find a method for estimating the two parameters of surface orientation, (σ, τ), given only a set of measurements of α^* along the contours in an image.

2.3.1 Signal and noise

Although the projective relation alone is not sufficient to recover surface orientation from the image, the problem would be straightforward if the shape of an observed contour prior to projection were known. For example, if an ellipse in the image were known to be the projection of a circle, the orientation of the circle would be uniquely determined[1] by the shape of its projection, and could be easily computed. That is, the *distortion* imposed by projection on a known curve is usually sufficient to recover the curve's orientation. The shape of the image contour may be conceptually divided into the unprojected shape, and a projective distortion imposed on that shape. If one component is known, the other can usually be recovered.

Even if the shapes of the contour generators contributing to a particular image are unknown, surface orientation may be *estimated* if the shapes of contour generators in general are given a *statistical* characterization, because the projective distortion, which depends on surface orientation, has a regular and systematic effect on the image. The *irregularity* of natural contour generators may be pitted against the *regularity* of the projective distortion, by treating the distortion imposed on the contour generators as a signal whose parameters must be estimated, and the shapes of the contour generators themselves as noise from which that signal must be isolated.

In the last section, a geometric expression was derived relating the tangent direction on a contour to the tangent direction on the contour generator, and the orientation of the surface. In terms of that expression, (σ, τ) is the signal to be estimated, β is the noise to be discarded, and α^* is the combination of signal and noise that can be measured.

A number of measures of α^*, taken across the image, define a *distribution* of observed tangent directions, which might for example be represented as a histogram. For any hypothesized surface orientation, the geometric relation translates each value of α^* into a corresponding value of β, and so translates the observed distribution of α^* into a corresponding distribution of β; a possible distribution of β may be obtained for each value of (σ, τ). If something were known about the expected distribution of β, then the distribution could be chosen, from the set of possible ones, that most closely resembled the expected distribution. To the extent the expected distribution was a good description of the shapes of contour generators, the value of (σ, τ) that was used to obtain the "best fit" to that expected distribution would be a good estimate for the orientation of the surface.

[1]Except, of course, for the inevitable reflective ambiguity of orthographic projection

In this section, a simple statistical characterization of the distribution of β will be coupled with the geometric model, to obtain a maximum likelihood estimator for surface orientation, given the image contours. The mathematical basis for this estimator lies in basic statistical theory, which allows us to derive the density function for a function of random variables, when the density functions for those random variables are known. To make quantitative estimates, we must assume some form for the joint distribution of (β, σ, τ). What form this distribution ought to have is an empirical question, but a plausible idealization is the assumption that tangent direction and surface orientation are independent and isotropic.

2.3.2 A joint density function for (β, σ, τ)

If β, σ, and τ are treated as random variables, and a joint probability density function (j.p.d.f.) is assumed for those variables, then (σ, τ) may be estimated statistically. To the extent the assumed j.p.d.f. accurately describes the world, that estimate will be valid in some statistical sense.

A j.p.d.f. is intended to give us the relative likelihood, for each set of values of the variables, that that particular set of values will be observed together. Consider the meaning of this j.p.d.f. in concrete terms: suppose we walked around the world, looking around as we usually do. We might measure the orientation of every surface our eyes fell on. Whenever the surface had curves on it of the sort that project into contours on the retinal image, we might also measure the tangent direction of the curves on the surface. Each time we did this, we could record a triple of numbers, (β, σ, τ) for tangent direction, slant, and tilt, respectively. On the basis of a large number of records of this kind, we could develop an empirical picture of the joint distribution of these variables in the environment.

To literally gather these "statistics of the universe" is possible, and has actually been done in other contexts (Brunswik, 1948; Switkes *et al.*, 1978), but I believe this is not the most cost-effective way to proceed. I will argue that plausible idealizations for these distributions can be inferred, and their validity subjected to indirect test by evaluating the consequences of adopting them.

We might expect, in practice, to find anisotropies in the distribution of surface orientations, arising ultimately from the effects of gravity, and of our characteristic orientation with respect to gravity. That is, the ground tends to center on the horizontal, and we tend to stand above it. Suppose we ignore this effect for now, leaving open the possibility of reintroducing it as a bias in the distribution. In that case, what properties might the distribution be expected display, if we lived in free fall? Given these limitations, the following propositions provide a simple idealization of the distribution:

i. Over the long run, surfaces are as likely to appear at any one orientation in space as at any other.

ii. Over the long run, tangents to contour generators are as likely to appear at any one orientation in the surface they lie on as at any other, regardless of the orientation of the surface.

What these propositions say is that there is no reason *a priori* to prefer any surface orientation to any other, or any tangent direction to any other, and that surface orientation and tangent direction are statistically independent.

The statement that all surface orientations are equally likely requires clarification: the orientation

of a surface can be given by the unit normal, i.e. a "needle" of unit length perpendicular to the surface. The set of normals corresponding to all possible surface orientations define a unit sphere[2] which contains the points of the needles. When we say that all surface orientations are equally likely, we mean the needle is as likely to land at any one point on the sphere as any other.[3]

When surface orientation is represented by the slant and tilt angles, σ and τ, the isotropy assumption does *not* translate into the assumption that all values of σ and τ are equally likely. For each value of σ, the possible values of τ define a circle on the gaussian sphere, whose radius approaches zero as σ approaches zero, and one, as σ approaches $\pi/2$. The circumference of the circle is easily shown to vary with $\sin\sigma$. Because the likelihood of landing on each point on the sphere is equal, the likelihood of landing on a curve on the sphere is proportional to the length of the curve. Since each value of σ corresponds to a circle with circumference proportional to $\sin\sigma$, the relative likelihood of σ is proportional to $\sin\sigma$. Noting that all values of τ, the tilt, and β, the tangent angle are equally likely over the range range $[0, \pi]$, we have the density function

$$\text{p.d.f.}(\beta, \sigma, \tau) = \frac{1}{\pi}\frac{1}{\pi}\sin\sigma = \frac{\sin\sigma}{\pi^2} \tag{2-2}$$

which, it is easily shown, integrates to one over the ranges of the parameters. We now have a statistical model for the scene parameters, and a geometric model relating these parameters to the image measurements. Together these measures determine a maximum likelihood estimator for surface orientation, given measures on an image. This estimator will now be derived.

2.4 Estimating surface orientation

Given a geometric model, which expresses the projected tangent direction α^* as a function of (β, σ, τ), and a statistical model which gives a j.p.d.f for (β, σ, τ), we derive the maximum likelihood estimator for (σ, τ) that follows from these models.

The first step is to derive the conditional[4] p.d.f for $(\alpha^* \mid \sigma, \tau)$. From this function, we obtain the joint conditional p.d.f for $(A^* \mid \sigma, \tau)$ where A^* is a set of measures $A^* = \{\alpha^*_1, \alpha^*_2, \ldots, \alpha^*_n\}$. At this step we introduce the assumption that the projected tangent directions, α^*_i, are independently drawn from p.d.f.$(\alpha^* \mid \sigma, \tau)$. The implications of this assumption will be discussed at some length in a later section. Then, using Bayes' rule, the joint conditional p.d.f. for $(\sigma, \tau \mid A^*)$ is obtained. A maximum likelihood estimate for (σ, τ) is the value of (σ, τ) for which that function is maximized.

[2]Called the *gaussian sphere*

[3]Since we are only concerned with visible points on opaque surfaces, we know in advance that the unit normal is confined to the hemisphere of visible directions, but this makes no difference for the derivation that follows.

[4]The *conditional probability* $(A \mid B)$ is defined as the probability of an event A, given that event B has occurred. A *conditional p.d.f.*, $f(a \mid B)$ is the p.d.f. for a random variable a given that event B has occurred. The conditional p.d.f. for $(\alpha^* \mid \sigma, \tau)$ is simply the p.d.f. for α^* given that σ and τ assume specified values.

2.4.1 Density function for $(\alpha^* \mid \sigma, \tau)$

From the last section we have the density function

$$\text{p.d.f.}(\beta, \sigma, \tau) = \frac{1 \sin \sigma}{\pi^2}$$

and the geometric relation

$$\alpha^* = \tan^{-1}\left(\frac{\tan \beta}{\cos \sigma}\right) + \tau$$

Thus α^* is a function of random variables with known distributions. To obtain the p.d.f. for $(\alpha^* \mid \sigma, \tau)$ we treat α^* as a function of β, with σ and τ as parameters. We use the relation

$$\text{p.d.f.}(\phi(x)) = \text{p.d.f.}(x)\frac{dx}{d\phi(x)}$$

where $\phi(x)$ is a function of random variable x. From this relation we have

$$\text{p.d.f.}(\alpha^*(\beta) \mid \sigma, \tau) = \text{p.d.f.}(\beta \mid \sigma, \tau)\frac{\partial \beta}{\partial \alpha^*}$$

From (2-1) it follows that

$$\beta = \tan^{-1}(\cos \sigma \tan(\alpha^* - \tau))$$

Differentiating with respect to α^* gives

$$\frac{\partial \beta}{\partial \alpha^*} = \frac{\cos \sigma}{\cos^2(\alpha^* - \tau) + \sin^2(\alpha^* - \tau)\cos^2 \sigma}$$

and $\text{p.d.f.}(\beta \mid \sigma, \tau)$ is simply $1/\pi$. So

$$\text{p.d.f.}(\alpha^* \mid \sigma, \tau) = \frac{1}{\pi}\left(\frac{\cos \sigma}{\cos^2(\alpha^* - \tau) + \sin^2(\alpha^* - \tau)\cos^2 \sigma}\right)$$

This density function tells us, under the assumptions of isotropy and independence for (β, σ, τ) how the image tangent direction is distributed as a function of surface orientation. This distribution is graphed at several values of σ and τ in Fig. 2.

2.4.2 Joint density function for $(A^* = \{\alpha_1^*, \ldots, \alpha_n^*\} \mid \sigma, \tau)$

Suppose we have measured the image tangent direction, α^*, at a series of n positions along an image contour. A basic relation in probability theory states that the joint density of n *independent* measures, each with density function $f(x)$ is

$$\text{p.d.f.}(X = \{x_1, \ldots, x_n\}) = f(x_1)f(x_2)\ldots f(x_n)$$

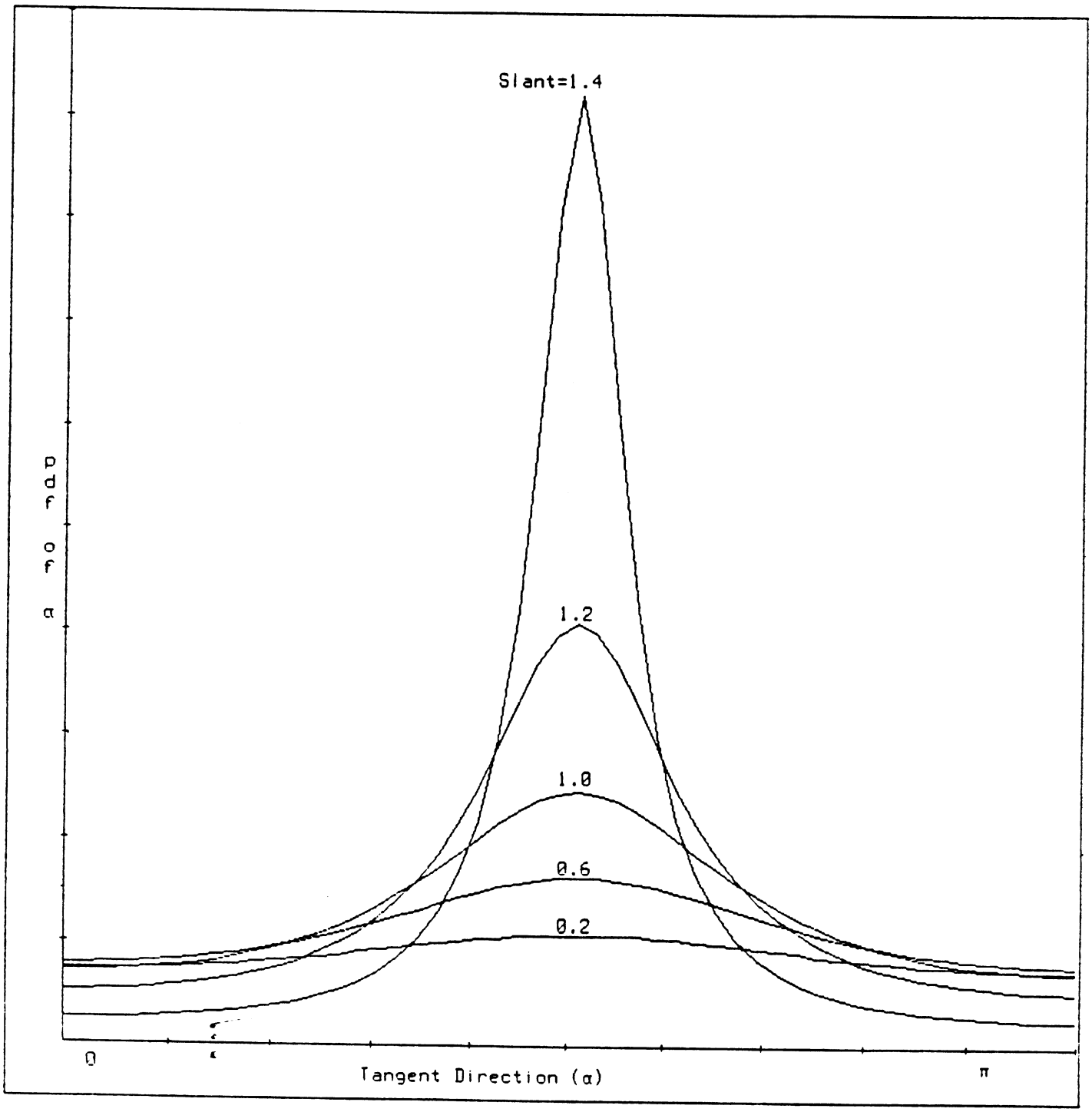

Figure 2. Curves in the function p.d.f.($\alpha^* \mid \sigma, \tau$), plotted against α^*, with $\tau = 0$, at several values of σ.

If we are willing to assume that a set of measures of tangent direction are independent, we have

$$\text{p.d.f.}\left(A^* = \{\alpha_1^*, \ldots, \alpha_n^*\} \mid \sigma, \tau\right) = \prod_{i=1,n} \text{p.d.f.}(\alpha_i^* \mid \sigma, \tau)$$

$$= \prod_{i=1,n} \frac{\pi^{-1} \cos\sigma}{\cos^2(\alpha_i^* - \tau) + \sin^2(\alpha_i^* - \tau)\cos^2\sigma}$$

where the symbol $\prod$ denotes an iterative product. This expression gives the relative likelihood for the set of observed image tangents, at each value of (σ, τ). By Bayes' rule, the density function for (σ, τ) given A^* is

$$\text{p.d.f.}(\sigma, \tau, \mid A^*) = \frac{\text{p.d.f.}(\sigma, \tau)\ \text{p.d.f.}(A^* \mid \sigma, \tau)}{\int\int \text{p.d.f.}(\sigma, \tau)\ \text{p.d.f.}(A^* \mid \sigma\tau)\, d\sigma d\tau}$$

where integration is performed over the ranges of σ and τ. Dividing by the integral simply normalizes the function to integrate to 1. The value of (σ, τ) for which this function assumes a maximum is the maximum likelihood estimate for surface orientation, and the integral of the function over a region gives the probability that the surface orientation lies inside that region.

Noting that

$$\text{p.d.f.}(\sigma, \tau) = \frac{1 \sin\sigma}{\pi}$$

the relative likelihood of $(\sigma, \tau \mid A^*)$ is

$$\text{p.d.f.}(\sigma, \tau)\ \text{p.d.f.}(A^* \mid \sigma, \tau) = \prod_{i=1,n} \frac{\pi^{-2} \sin\sigma \cos\sigma}{\cos^2(\alpha_i^* - \tau) + \sin^2(\alpha_i^* - \tau)\cos^2\sigma} \qquad (2\text{–}3)$$

We normalize this relative likelihood function to obtain the density function, by dividing by its integral, which can be approximated by summing values of the function taken at equal intervals of σ and τ.

2.4.3 Summary of the model

The geometric/statistical model from which this estimator follows constitutes a set of claims about the domain. The estimator is valid to the extent these claims are true of the domain. These, in summary, are the claims that comprise the model:

1. Geometric model. Each image tangent measure, α^*, is related to the scene parameters (β, σ, τ) by the expression

$$\alpha^* = \tan^{-1}\left(\frac{\tan\beta}{\cos\sigma}\right) + \tau$$

2. Planarity restriction. The surface orientation (σ, τ) is constant over position.

3. Statistical model. The joint distribution of (β, σ, τ) is given by

$$\text{p.d.f.}(\beta, \sigma, \tau) = \frac{1 \sin \sigma}{\pi^2}$$

and the image measures of α^* correspond to values of β independently drawn from this distribution for some value of (σ, τ).

4. Estimator. Derived from these assumptions is a density function for surface orientation, given the image data, given by

$$\prod_{i=1,n} \frac{\pi^{-2} \sin \sigma \cos \sigma}{\cos^2(\alpha_i^* - \tau) + \sin^2(\alpha_i^* - \tau) \cos^2 \sigma},$$

normalized by its integral with respect to (σ, τ). The value of (σ, τ) at which this function assumes a maximum is the maximum likelihood estimate for surface orientation, under the assumptions of the model; and the integral of the function over a region of (σ, τ) is the probability that surface orientation lies in that region.

2.5 Implementation

In this section, an implementation of the estimation strategy is reported and assessed. The strategy was applied to two natural domains: geographic contours, drawn from a digitized world map, and natural images, using zero-crossing contours in the $\nabla^2 G$ convolution, as described by Marr & Poggio (1979), and Marr & Hildreth (1979). The zero-crossings of this convolution are peaks in the first derivative of the band-passed image. While these zeros are regarded by Marr & Poggio as precursors of contours, they correspond closely enough to significant events on the surface to have the desired properties for estimation. Since the strategy is limited to estimating planar orientations, images of approximately planar surfaces were chosen. The key questions addressed in assessing the performance of the strategy are: how accurately does it estimate surface orientation, and how accurately does it estimate the *error* of its own estimates, i.e. the confidence regions for the estimates.

2.5.1 Computing the estimate

The aim of the computation is to determine the density function and maximum likelihood estimate for surface orientation, given a set of tangent measures. The data are conveniently represented in grouped form, by dividing the continuum of tangent direction, on the interval $[0, \pi)$, into a set of subintervals of equal length, and recording the number of measures that fall into each subinterval. Since the data are a collection of curves, this amounts to recording the total arc length that falls in each orientation band.

Let $A^* = \{a_1^*, \ldots, a_n^*\}$ be the data grouped into n orientation bands, with a_i^* the midpoint of the ith band. That is, each a_i^* gives the *number* of data points falling in the corresponding interval. Then, for the grouped data, the relative likelihood of $(\sigma, \tau \mid A^*)$ becomes, from (2–3),

$$L(\sigma, \tau \mid A^*) = \exp\left(\sum_{i=1,n} a_i^* \log \left(\frac{\pi^{-2} \sin\sigma \cos\sigma}{\cos^2(a_i^* - \tau) + \sin^2(a_i^* - \tau)\cos^2\sigma} \right) \right) \qquad (2\text{–}4)$$

And, if this function is computed at m equally spaced values of σ, and p equally spaced values of τ, the density function is approximated by

$$\text{p.d.f.}(\sigma, \tau \mid A^*) \approx \frac{L(\sigma, \tau \mid A^*)}{\sum_{i=1,m} \sum_{j=1,p} L(\sigma_i, \tau_j \mid A^*)} \qquad (2\text{–}5)$$

where $L(\sigma, \tau \mid A^*)$ is from (2–4).

The value of (σ, τ) at which this function assumes a maximum approximates the maximum likelihood estimate for surface orientation, and the sum of the function sampled at uniform intervals on a region of (σ, τ) approximates the probability that surface orientation lies inside the region. The computation was facilitated further by placing the values of $\log \text{p.d.f.}(a_i^* \mid \sigma_j, \tau_k)$ in a lookup-table.

2.5.2 Experiment I: geographic contours

Stimuli. The initial test of the strategy employed geographic contours drawn from a digitized world map, which obviously posed no extraction problem. Beyond this advantage, these contours provide a data-base of curves which were generated by physical processes, and, when taken small enough to neglect the curvature of the earth, are planar. Moreover, by subjecting the curves to rotation/projection transforms, stimuli are generated whose "real" orientation in space is known exactly. This degree of control is much more difficult to obtain using natural images.

The curves are land-water boundaries, represented in the data base as chains of points in latitude/longitude coordinates. These were converted to cartesian coordinates, and projected onto the earth's tangent plane in the neighborhood of the curve, giving a frontal-plane representation. Sufficiently small curves were selected that the curvature of the earth was negligible. The coastlines of islands and lakes were chosen as a class of closed curves of reasonable size. Several of the curves are shown in fig. 3.

Stimuli were generated from the frontal-plane curves by rotating them through a given (σ, τ), and orthographically projecting them to produce an image contour. These curves were converted to grouped-data form as follows: between each pair of vertices, α^* is given by $\tan^{-1}(\Delta y / \Delta x)$, and the arc length between the vertices by $\sqrt{\Delta x^2 + \Delta y^2}$. The arc length was summed into the appropriate orientation cell. Seven orientation cells were used, since it was found that finer divisions had little effect on the estimate.

Coastlines of islands and lakes were selected from the data base on the basis of size: chains of several hundred vertices each were chosen. The maximum likelihood estimate and p.d.f. were computed for each curve at 36 orientations, with orientations uniformly spaced on the gaussian sphere.

P14475
INES EL
ELLIS

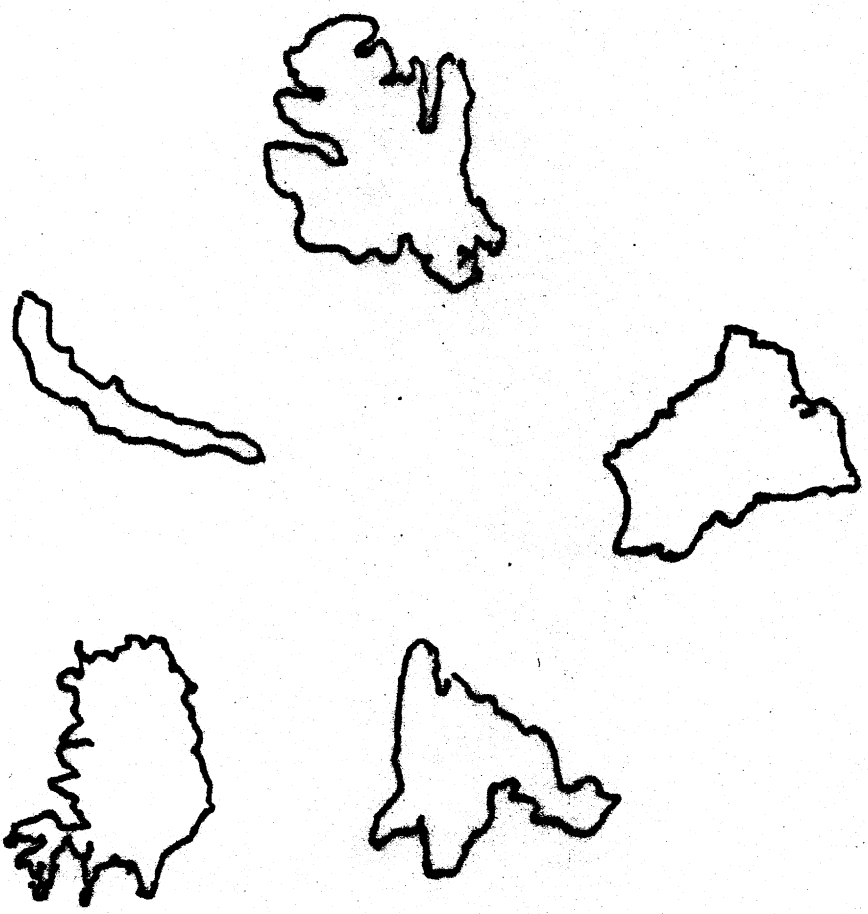

Figure 3. Some islands drawn from the geographic data base

Results. The results for one curve at a number of orientations are shown in detail in fig. 4. For each orientation, the appearance of the curve and a contour plot of the log p.d.f. are shown. In general, as slant increases, the accuracy of the estimate increases and the density function falls off more steeply around the estimate. This is to be expected, if the projective distortion is viewed as a signal, because σ is approximately the amplitude of the signal, so increasing σ increases the signal-to-noise ratio: that is, there is *more* projective distortion at larger slants.

Figs 5 and 6 summarize the results for seven curves: scatter plots are shown for estimated against actual σ and τ, as well as histograms of the observed error distributions. Clearly, for this class of shapes, the strategy makes good estimates.

Next we consider the effectiveness of the strategy at estimating its own error. A simple measure of confidence in the estimate is the maximum value of the p.d.f. Since the p.d.f. was computed at discrete points, this value may be viewed as the probability that (σ, τ) lies in a small region of fixed size around the maximum likelihood estimate. As shown in fig. 7, the mean error of estimation drops sharply as this value increases, for both σ and τ. Thus, the peak value of the p.d.f. can be used to reliably distinguish good estimates from bad ones. A more thorough gauge of confidence can be obtained by computing a confidence region, i.e. an iso-density contour within which the integral of the p.d.f. assumes a specified value.

2.5.3 Natural images

Extracting contours. The most substantial problem in applying the estimate to natural images is that fully adequate means of locating image contours do not yet exist. A promising basis for the location of image contours are *zero-crossing contours*, developed by Marr & Poggio (1979). The image is convolved with a circular $\nabla^2 G$ mask, the laplacian of a two-dimensional gaussian, and the zero-crossings of the convolution correspond to peaks in the first directional derivative of intensity in the band-passed image.

Zero-crossing contours are proposed by Marr & Poggio to be an effective description of the intensity changes in images at different spatial scales; they are regarded as precursors of perceptual contours. To provide appropriate data for the estimation strategy, the shapes of zero-crossing contours must bear a regular relation to processes acting on the surface, and they appear to possess this property.

Veridical orientation. A less serious problem is that, unless a scene was photographed under carefully controlled conditions, the orientations of surfaces are not precisely known. The geographic contours provided the opportunity to systematically compare the estimates to precisely known veridical orientations. For the present purpose, we can trust our own perceptions of the photographs; if the strategy and our perceptions agree, at worst they err in the same direction.

Selection of photographs. The estimation strategy under consideration is limited by the planarity restriction. In observance of this restriction, pictures of approximately planar surfaces were chosen. Several kinds of contour generating processes are represented, including surface markings and cast shadows. Also of interest are surfaces which are not planar, but have an "overall orientation,"

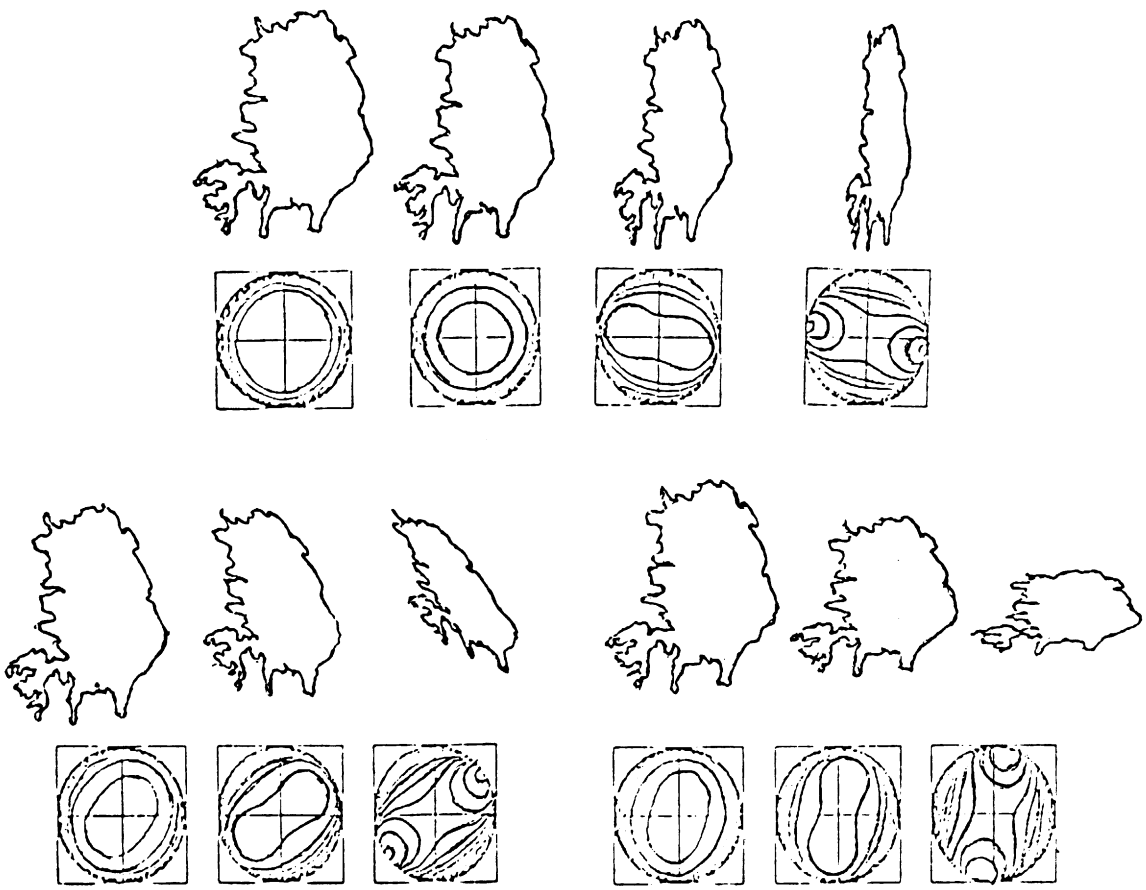

Figure 4. One of the geographic contours shown at various orientations, with the density function obtained at that orientation. The density function is plotted by iso-density contours, with (σ, τ) represented in polar form: σ is given by distance to the origin, τ by the angle. The radial symmetry of the plots reflects the symmetry of orthographic projection. The sharp, symmetric peaks clearly visible at higher slants are the maximum likelihood estimates for (σ, τ).

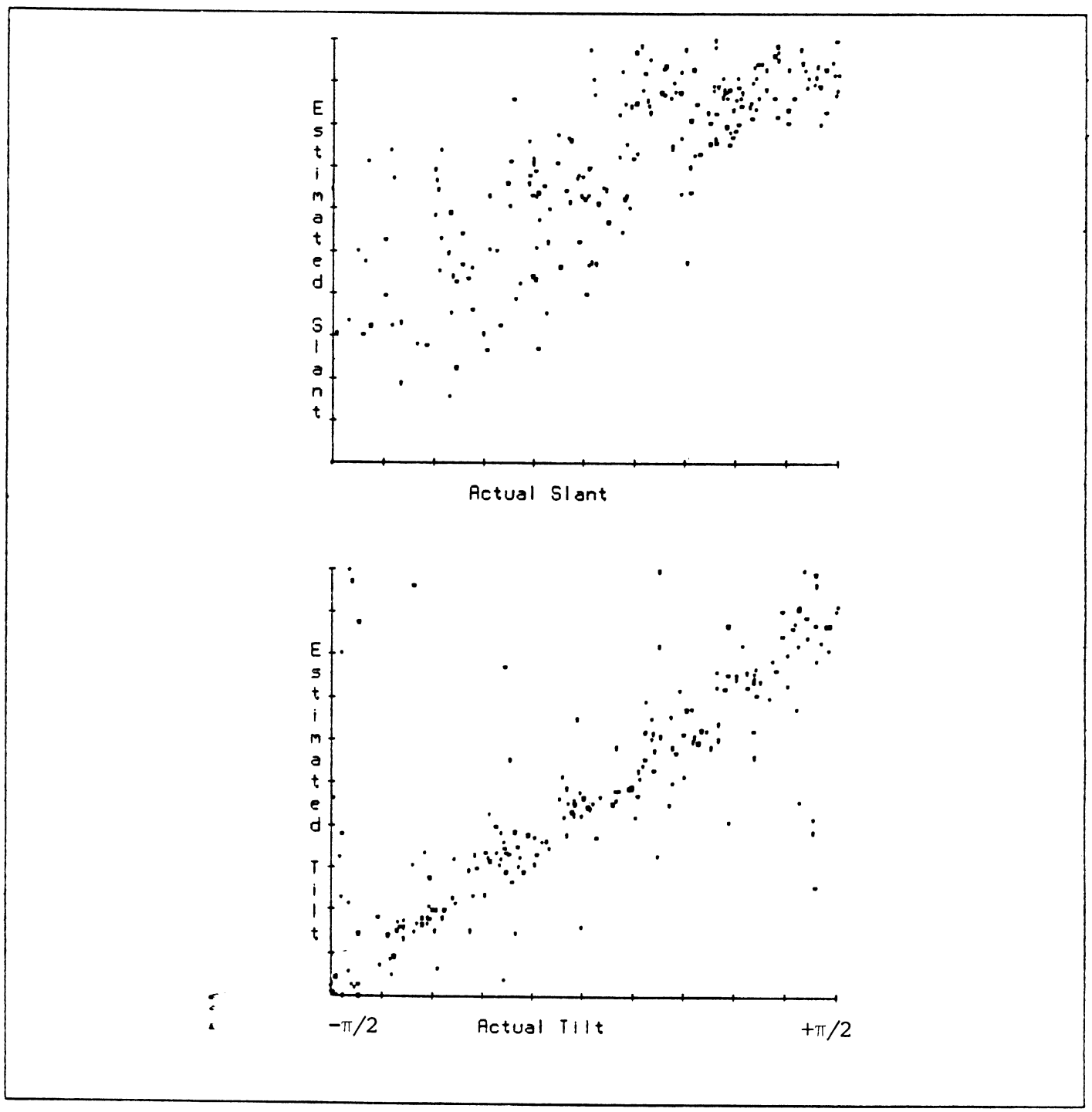

Figure 5. Scatter plots of actual vs. estimated σ (top) and τ (bottom) for the geographic contours.

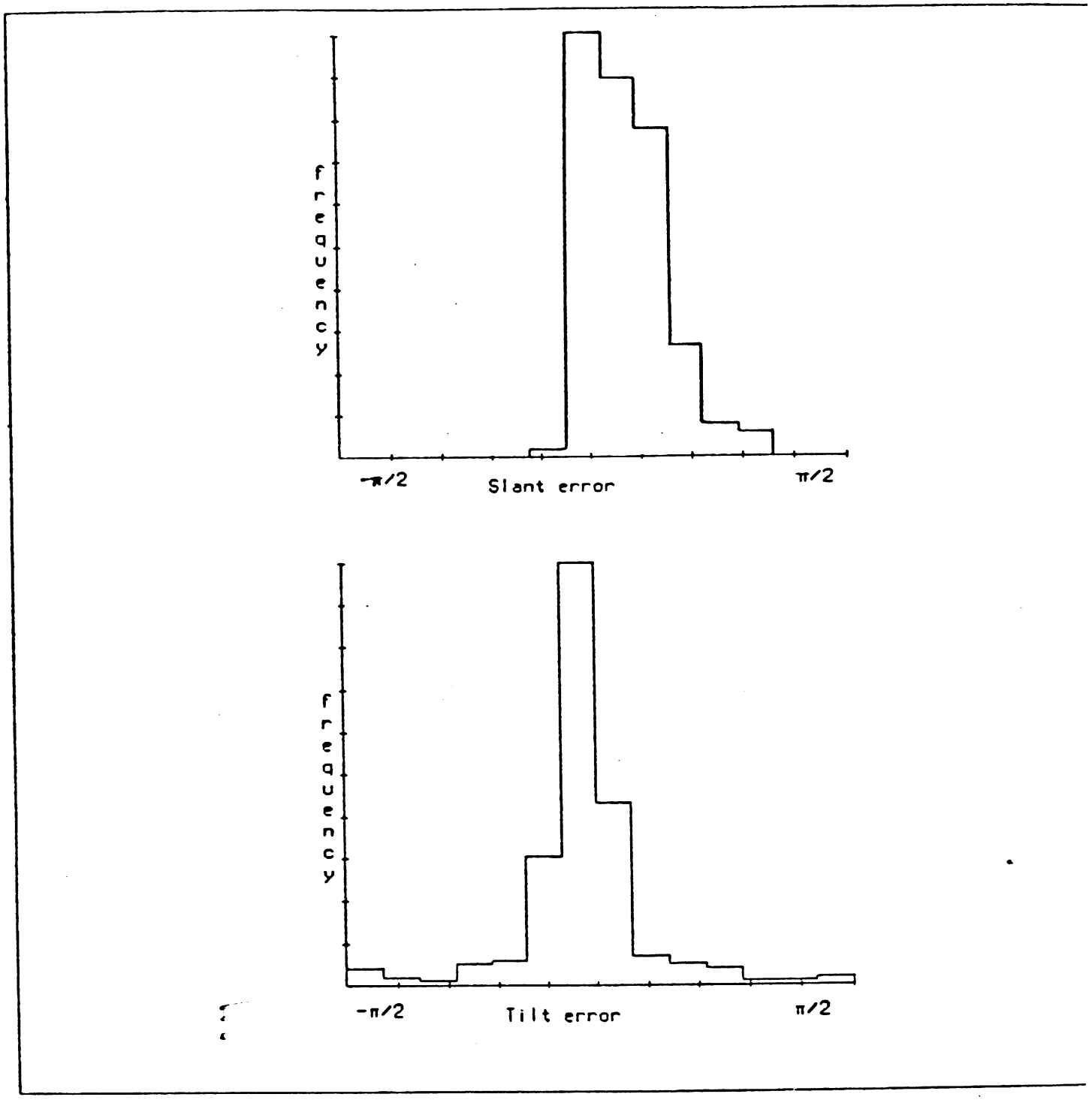

Figure 6. Error distributions for σ (top) and τ (bottom). For both σ and τ, the largest possible error is $\pi/2$.

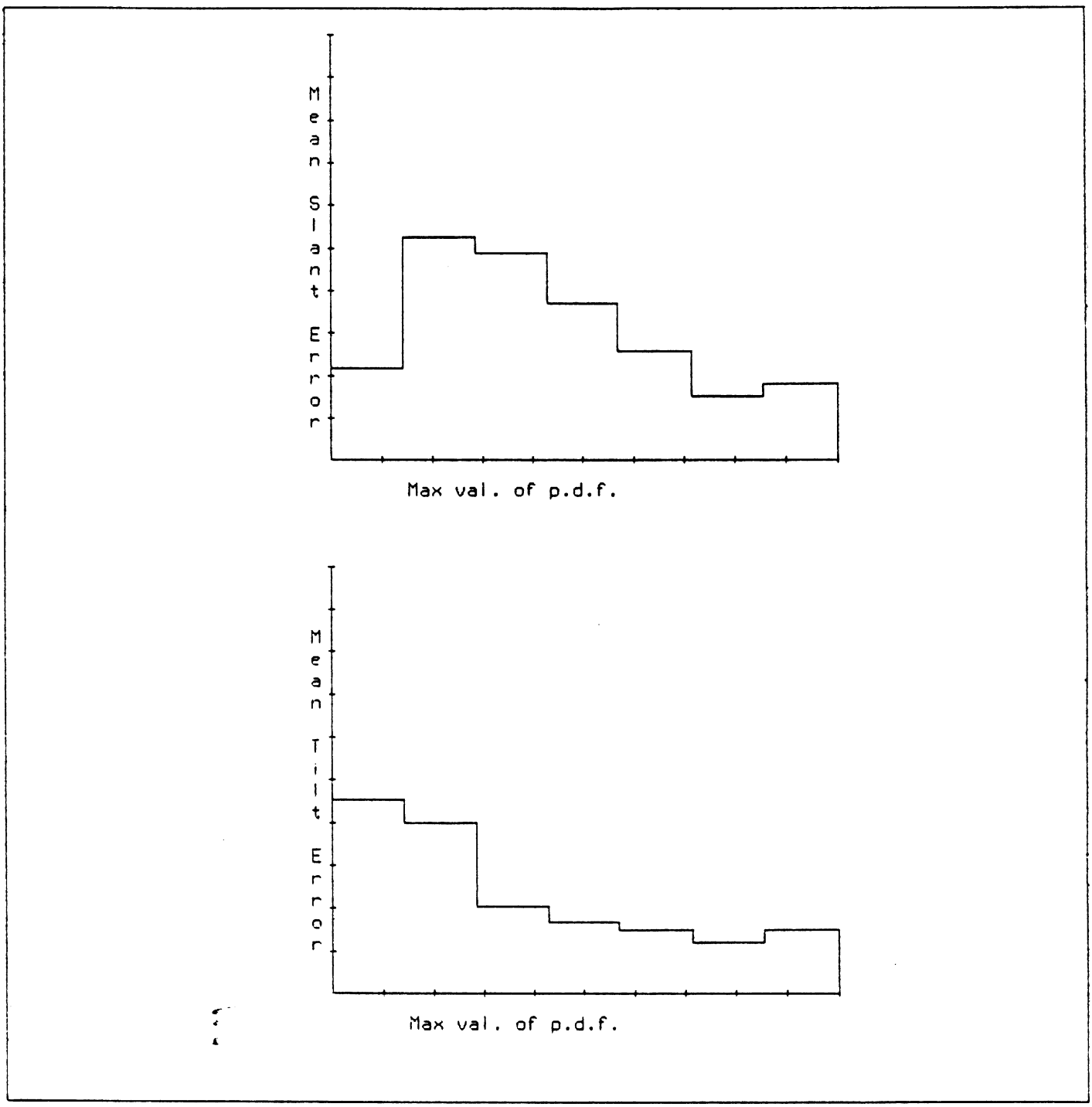

Figure 7. Mean error of estimation as a function of the maximum value of the p.d.f. for σ (top) and τ (bottom). The mean error drops sharply as this value increases, showing that the reliability of the estimates can be effectively gauged.

i.e. a substantial low-frequency component in the depth function. A potentially practical application of the planar strategy is the estimation of this component.

Digitization. The photographs were digitized on the Optronics Photoscanner at the MIT AI lab, an accurate, high-resolution digitizing device. The digitized images contained between three and four hundred pixels in each dimension, with intensity quantized to 256 grey levels.

Convolution. The digitized images were convolved with $\nabla^2 G$ masks, as described in Marr & Hildreth (1979). The convolutions were performed on a Lisp Machine at the MIT AI Lab, using specialized convolution hardware. A mask with a central radius of eighteen pixels was used; the total diameter of the mask was sixty pixels. Figure 8 shows a digitized image, its convolution with a $\nabla^2 G$ function, and the zeros of the convolution.

Extraction of tangent direction. Tangent direction was measured along the zero-crossing contours of the convolutions by first locating points on the contours, then measuring the gradient of the convolution at those points. The tangent to the contour is orthogonal to the gradient.

Grouping the data. The data were grouped by tangent direction, in the form described above, by sampling the contours at fixed increments of arc-length, measuring the tangent orientation, and summing into the appropriate orientation cell. From this point on, the estimate was computed as for the geographic contours.

Results. The photographs, together with the computed density functions for (σ, τ), are shown in Fig. 9. These should be compared with the apparent orientations of the pictured surfaces. Most observers' perceptions of these surfaces agree closely with the estimates.

2.6 Avoiding failures of the estimation strategy

2.6.1 Two kinds of error

It is necessary at the outset to distinguish two kinds of error that may arise in any statistical estimate or decision: first, those errors that the estimation strategy "knows about" and describes statistically, e.g. in terms of error bars around an estimate. Second, those errors that arise because the model on which the estimation strategy is based misrepresents the domain, e.g. when a population assumed to be normal is actually badly skewed.

The first of these, sampling errors, are not really a problem. Although it is of course desirable to minimize sampling error, its magnitude is statistically predictable, even when large. It was shown in the last section that, although the surface orientation estimation strategy sometimes makes large errors, the confidence information it generates can be used to distinguish the good estimates from the bad ones. While it would be better to avoid large errors altogether, if they must occur, it is important that they be recognized.

Figure 8. A digitized image, its convolution with a $\nabla^2 G$ function, and the zeros of the convolution.

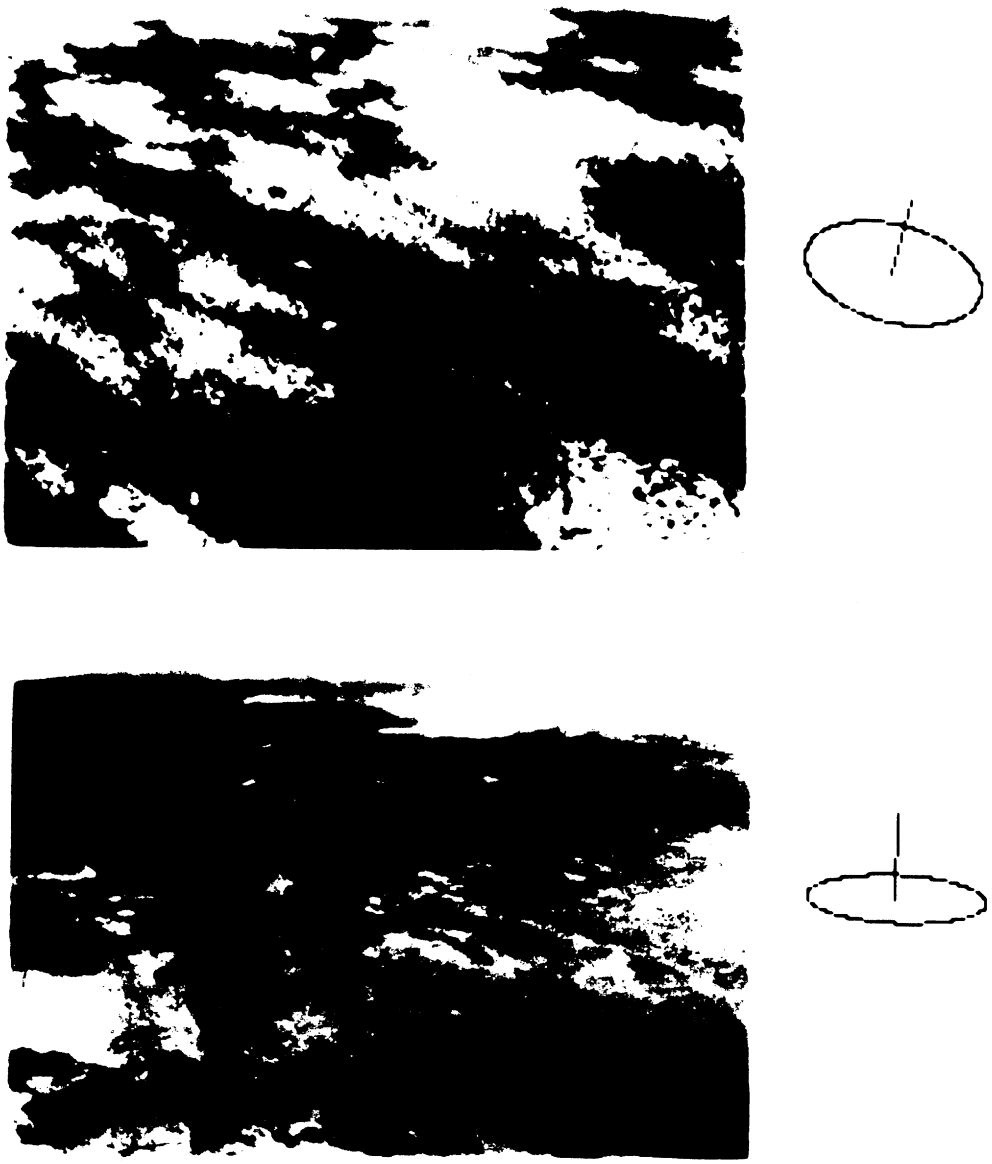

Figure 9. Surface orientation estimates from photographs. The estimated surface orientation is indicated by an ellipse, representing the projected appearance a circle lying on the surface would have, if the maximum likelihood estimate were correct.

Failures that follow from misrepresentations of the domain will thus be the principal subject of this section. Such failures cannot be predicted by the estimation strategy because they are failures of the premises on which the strategy is based. They must either be detected by means external to the strategy, incorporated into the strategy's model of the domain, or just tolerated.

We have seen that the planar estimation strategy *works* on a variety of naturally generated contours. The geometric/statistical model on which it is based therefore captures enough of the character of the domain to be useful. While this model can't be dismissed as just wrong, we are concerned with situations that arise often enough in natural scenes to be a problem, and in which the assumptions comprising the model are drastically and systematically violated. To the extent such situations arise, the strategy will fail to a degree its own confidence information can't predict. What assumptions comprising the model are likely to cause failure?

2.6.2 Violations of the model

The estimation strategy presented in this chapter is based on a model of the domain to which it applies. Aside from the planarity restriction, which is a domain restriction rather than a realistic assumption, the components of the model are a geometric expression that gives α^* as a function of β, σ, and τ; and a statistical model that gives a joint density function for (β, σ, τ). Together, these functions specify a density function for $\alpha^* \mid \sigma, \tau$. Given a set of measured values of α^*, we used the density function to estimate (σ, τ), making the additional assumption that the measured values of α^* were independently drawn from that density function.

Each component of this model is an assumption about the domain. If all the assumptions are true, then the statistical conclusions that follow from them are true, and all the estimation errors are sampling errors whose distribution can be computed. If any of these assumptions is violated, the model may fail in a manner not predicted by the error distribution. Of interest, therefore, are the conditions under which one or another of the assumptions is likely to fail. We will briefly consider each assumption in turn. Of course, planarity restriction is not expected to hold for real scenes, and will not be discussed.

Geometric model.

Given the idealization of orthographic projection, and given that the mathematical notions of surface orientation and tangent direction are idealizations when applied to physical surfaces, the geometric relation between α^*, β, σ, and τ is lawful and exceptionless. This assumption will not fail.

Joint density function for (β, σ, τ). This density function is intrinsically a statement about the domain *in the long run*, across the variations from one scene to the next, so it is not meaningful to look for exceptions to it in particular scenes. Of course, the simple density function following from the assumption of isotropy may be wrong *in the long run*. But were we to change that function, for example, to reflect the systematic effects of gravity, the estimation strategy would remain substantially the same, although the estimates it produced would change quantitatively. That is, if it turned out the density function had to be changed, the corresponding change to the estimation strategy would be straightforward, so this is not serious problem.

In fact, if error feedback were available to the estimation strategy, it would be quite possible for it to "learn" this distribution, or tune it by an error-reduction criterion.

Independence of image measures. This assumption amounts to treating the measured tangent directions as the projected orientations of a number of needles randomly and independently dropped on the surface. While this assumption greatly simplifies the computation of the estimate, it is not realistic unless the measurements really do correspond to independently thrown needles. The assumption can fail in at least two ways, each of which may have serious consequences.

First, the orientation along contours usually varies continuously, so two very nearby points are liable to have nearly the same orientation. In other words, the orientations at nearby points are correlated, and therefore dependent. Since the data were taken by sampling at fixed intervals of arc length, the consequence of this dependence is the artificial inflation of the sample size, if the sampling interval is chosen too small. This inflation does not, in general, substantially change the estimate obtained, but it does inflate the confidence of the estimate. This may not be so serious if the estimate happens to be accurate, but has drastic consequences if the estimation is performed on a small enough arc of contour that all the data are highly correlated. In that case, we may think we have a great many data points, when we really have only a few. And an estimate that probably has almost no meaning may be taken for a very reliable one.

A second source of dependence among the image measures is more global. Dependence may be imposed by the operation of a process that systematically influences the orientations of contour generators across the surface. An example of such a process is shadow casting: the geometry of the process that casts a shadow onto a surface is almost identical to the geometry of the shadow's projection into the image. That is, the contour generator, by the time it is formed, has already undergone a projective transformation. Needless to say, the estimation strategy may be badly fooled by cast-shadow contours, unless they can be distinguished from other kinds. Some surface-marking processes may also impose systematic projection-like distortions on contours.

Clearly, the assumption that the image measures are independent is the weak link in the chain. Next some means of avoiding the failures that come from this assumption will be considered.

2.6.3 Continuity of contours

On a smooth curve, the tangent directions at nearby points are highly correlated. But, unless the curve was generated in some regular way, the orientations at more widely separated points are likely to be uncorrelated. To legitimize the assumption that the measured tangent directions are independent, a large enough sampling interval must be chosen that the dependence among successive measures is small enough to be neglected. On the other hand, if the sampling interval is too large, data are lost. How can we decide which sampling interval is best? Clearly, this decision must depend on the shapes of the contours.

Analogous problems in signal detection. The problem of dependence among nearby data points is one that arises in signal processing applications, as when a signal must be detected in noise.

Since, for most kinds of noise, nearby values of the noise are correlated, signal detection strategies face much the same dependence problem that has arisen here. If the properties of the noise are well enough understood, this dependence may be explicitly taken into account in the detection procedure, or effectively eliminated by judicious sampling of the incoming signal.

The degree to which values of a function at some fixed separation are correlated is usually expressed by the *autocorrelation function*, which is just the convolution of the function with itself, or by its normalized equivalent, the autocovariance function. When that function is known *a priori*, the simplest procedure is to choose a large enough sampling interval that the dependence between successive data points is negligible. Considering the enormous variations in scale of the contours with which we might be concerned, this procedure is not applicable to the current problem.

An alternative is to use a sample estimate of the autocovariance function to choose a sampling interval. When nothing is known of the autocovariance *a priori*, this is probably the best that can be done to combat dependence; but it is an expensive computation, and probably much more elaborate than necessary. Rather than vastly complicate the estimation procedure, it would be desirable to find a rough, simple guide to the appropriate sampling interval. All such a measure really has to do is avoid the disastrous consequences of oversampling that arise when the contours are very smooth, hopefully without discarding much more data than necessary, and without introducing a new class of failures. And it need only work for the sort of cases that are likely to arise in natural images. There are many ways such a crude measure might be formulated, but a few simple constraints can be placed on the measure:

First, two sets of contours, identical except for size, ought to lead to the same estimate, and the same confidence in the estimate. In other words, when the data are transformed by uniform scaling, the sampling interval should transform in the same way.

Second, a sampling interval small enough that adjacent measures are highly correlated, is small enough that orientation undergoes little change from one measure to the next. So, in general, the more rapidly the direction of a curve changes along its length, the the smaller the sampling interval should be. That is, samples should be taken less often along a contour that is nearly a straight line, than along one that twists and turns rapidly. This may not be so if the curve twists rapidly but regularly, i.e. is periodic, but such curves are not liable to appear in natural images.

If we incorporate these constraints into a sampling strategy, we are not likely to be fooled into excessively close sampling of very smooth curves. A very simple way to do this is to measure the total change in orientation along the contour normalized by total arc length; which is a measure of the average change in orientation per unit arc length.[5] This can be expressed by

$$\frac{1}{l}\int_a^b |\kappa|\, ds$$

[5]In fact, if we were measuring anything but tangent direction, we might do quite well to *sample* by fixed steps of change in tangent direction. That is, each time the total change in tangent direction exceeds some fixed amount, take another data point. This, of course, is nonsensical as a way to sample tangent direction, but if an estimate could be based on some other measure on the curve, this extremely simple sampling strategy might substantially avoid the dependence problem. Such a measure is the *curvature* of the contour. An estimator based on contour curvature has also been applied successfully to planar curves.

where l is the length of the curve, a and b the endpoints, κ is the curvature, and s is a natural parameter. In practice, we would just sum the (unsigned) angles between points at fixed steps of arc length to approximate the integral.

Recall that the estimator was implemented using data grouped into orientation bands, effectively giving the a histogram of the total arc length in each band. All that need be done is to *scale* this histogram to reflect the degree of dependence.

2.6.4 Systematic effects on tangent direction

If the local effects of continuity can be reduced by judicious sampling of the data, a more difficult form of dependence is manifested in the sort of global systematic distortion of tangent direction on the surface epitomized by cast shadows. This kind of dependence cannot be overcome by proper sampling, because it is liable to apply everywhere in the image. Rather, two options are available: find ways to decide when such a process has operated, and, where possible, allow for it in the estimation strategy; or always leave open the possibility of such failure, but suspend final judgment until the estimate can be compared with independent ones, such as those derived from shading information.

Detecting and modeling distorting processes. The first option entails somehow detecting the offending process. If that can be done, and the process is understood, its properties can be incorporated into the domain model and the consequent estimation strategy. This is option surely leads to the best estimates, but is also the most difficult to realize. It might be possible to recognize, say, cast shadows by the intensity profiles across their edges. Then, the appropriate geometric model would be a model of shadow casting.[6] But how could we recognize some surface-marking process that systematically stretched or otherwise deformed the contour generators? And even if we could, how could we hope to model such processes sufficiently well to allow for the distortion? With the possible exception of cast shadows, which are common enough and regular enough in their geometry to perhaps warrant special treatment, detecting, no less modeling, all of the processes that might systematically alter the distribution of tangents on a surface seems to be a hopeless task.

Using independent estimates of the surface. Fortunately, a number of properties of the image potentially inform us of surface shape and orientation in the image; otherwise image interpretation would probably be impossible. It is important to remember that a method for estimating surfaces from image contours is intended ultimately to act in concert with methods based on other sources. Each method may, under some realistic conditions, fail badly. But, to the extent the causes of the various methods' failures are independent, failures can be detected in the discrepancies among the estimates. For example, even if the contours in a given image mislead, shading information and such are most unlikely to mislead in just the same way.

In many cases, the strategy of combining independent estimates simply means that the significance attached to any one estimate must be qualified, leaving open the possibility that some confounding

[6] In fact, for planar surfaces and orthographic projection, no estimate can be made using cast shadow contours, if the direction of illumination is unknown: the contour has undergone two projections, and there is no way to differentiate them in the image geometry. If the direction of illumination is known, then the first projection can be allowed for, and an estimate can be made. The geometry of cast shadows will be considered in detail in Chapter 4.

process has operated. For the planar estimation strategy, this means we take our estimate for (σ, τ) to be the best estimate of the *projection-like component* in the image data, but we leave open the possibility that the projection-like component is not due to projection, deferring a final decision until we have corroborating evidence.

Using perspective effects. The planar estimation strategy uses foreshortening distortion, and assumes orthographic projection. However, real images are usually subject to perspective distortions (i.e. change of projected size with distance to the viewer) as well. It may be possible to take perspective-based and foreshortening-based estimates, derived from the same image data, as independent in terms of confounding projection-like effects: while real foreshortening and perspective distortions are rigidly linked on smooth surfaces, it is very unlikely that projection-like distortions of the surface markings themselves will be linked in this way. Thus, a high correlation between perspective-based and foreshortening-based estimates provides strong evidence that the estimates are valid. This possibility will be explored more fully in the next chapter.

Extracting contours at different spatial scales. There is another interesting possibility for obtaining several approximately independent estimates of the surface from contour information alone. This possibility is suggested by Marr's observation (1979) that several independent processes may often be responsible for the markings on a surface, and that these processes often operate at different spatial scales. This observation in part motivated Marr & Poggio's () advocacy of the $\nabla^2 G$ convolution, and its zero-crossing contours, as primitive descriptions of the intensity changes in images. The zero-crossing contours are approximately zeros of the second derivative of intensity in a band-passed image, with the band pass depending on the size of the $\nabla^2 G$ mask. Hence, contours obtained with masks of different size encode properties of the image at different spatial scales; and the contours often depend on distinct, independent physical processes.

It is easy to find examples of independent processes acting at different spatial scales. For example, a lawn with overhanging trees may be marked by cast shadows at a large scale, leaves at a smaller scale, and still smaller grass blades. The shadows may be systematically stretched by low sun elevation, and the grass texture in the image may be strongly oriented, but it is most unlikely that such projection-like effects will err in the same way. The zero-crossings of $\nabla^2 G$ convolutions at different scales do tend to isolate such processes, and separate estimates can be obtained from each convolution. However, contours at different scales don't necessarily derive from independent processes, and there is no clear way to establish independence. How valuable the additional information at different scales will be depends on the frequency with which each scale captures independent processes. This empirical question has not yet been addressed.

In sum, any strategy that estimates surfaces — from contours, shading, texture, or whatever — is liable from time to time to encounter scenes that, for one reason or another, are systematically misleading. This is precisely why it is important to have as many independent means of estimation as possible, because independent estimates are most unlikely to fail in the same way at the same time. The integration of independent estimates into a final decision about the scene is a substantial and general problem in image interpretation that has not yet been solved.

2.7 Summary

A method for estimating the orientation of planar surfaces from contours was derived from a model of the relevant imaging geometry, and some simple statistical assumptions about visual scenes. The geometric model related surface orientation, and the tangent direction of a contour generator on the surface, to the projected direction of the tangent in the image. The statistical model postulated that surface orientation and tangent direction in the scene are isotropic and independent. Together, the geometric and statistical assumptions determine a maximum likelihood estimator for surface orientation, given a set of independent tangent measures in the image. This estimator was derived and implemented.

The estimation strategy was tested on geographic contours, whose orientations could be controlled exactly, and on natural images, using zero crossing contours in the $\nabla^2 G$ convolution. The strategy was shown to give reliable estimates, as well as estimates of reliability.

Next were considered circumstances under which the model, and hence the estimator, might fail. It was argued that the assumption that the image measures are independent is unrealistic, and can lead to serious failures, due to the continuity of contours, and to the existence of "projection-like" processes that impose a systematic distortion on the contour generators. The problem of continuity can be reduced by careful sampling, but failures due to systematic "projection-like" effects cannot be avoided unless those processes can be detected directly in the image, or indirectly by reference to independent estimates of surface orientation.

CHAPTER 3

EXTENSION TO CURVED SURFACES

3.1 Introduction

This chapter treats the estimation of curved surfaces, by extension of the planar technique. First it is shown that the relative likelihood statistic developed in the last chapter readily extends to curved surfaces, in the sense that the statistic can be evaluated given a set of image data, and an arbitrary hypothesized surface: each data point in the image may be judged against the surface orientation assigned to that point by the hypothesized surface, and a combined likelihood computed across the image. This global approach may be used to find a maximum likelihood surface, but only in very restricted situations. For example, if it were known in advance that one of a small set of surfaces were present, and a prior probability for each of those surfaces were also known, a probability could be computed for each surface, and a maximum likelihood surface selected. More generally, if the surface were known in advance to belong to a restricted family of a few parameters—for example, a surface with known shape but unknown position, size, or orientation—a maximum likelihood surface could be chosen from that family. But in the general case, such strong constraints are not available, and the unknown surface may really be any surface at all. Because the contour data are discrete, and their density limited, it is always possible to construct an infinity of "perfect fit" surfaces, just as any finite set of data can be perfectly fit to an infinity of polynomials of sufficiently high degree. Because these "solutions" have absolutely no meaning as estimates, the global surface-fitting approach cannot apply to the general case.

The remainder of the chapter presents an alternative strategy, more suited to the general case, that estimates surface orientation at each point using only the data in a local neighborhood surrounding the point. Because this procedure tends to average out variations in surface orientation occurring at a smaller scale than the region size, the result is a "coarse" estimate of the surface. Because choosing a larger region around each point eliminates surface features over a larger spatial extent, the spatial resolution of the estimate is determined by the size of the region, just as the resolution of a surface obtained by shading information is limited by the resolution of the image. As well as reducing resolution, a larger region incorporates more data, and hence reduces the variance of the estimate; so this local strategy trades off *resolution* and *accuracy*. Since the region must be large enough, compared to the density of the data, to incorporate a reasonable data sample, the density of the data effectively limits the resolution with which the surface can be described.

To implement the neighborhood estimation strategy, a spatially extended distribution is first computed from the image data, to obtain at each point the distribution of tangent directions observed in a surrounding region. Surface orientation at each point is then estimated using the planar estimator of the last chapter. In effect, a region surrounding each point is thus treated as a plane. But rather than assuming that surface orientation is constant within a region, variations in orientation at a small scale are treated as noise, and the estimated surface is smoothed accordingly. Estimates were computed on natural images by this procedure, and are shown to provide good "coarse" estimates of surface shape and orientation.

3.2 Extension of goodness-of-fit measures to curved surfaces

The planar estimation strategy provided a basis for assigning relative likelihoods to surface orientations, given the image data. The success of that strategy depended on the surface being known in advance to be planar. The set of planar surfaces is one particular family of surfaces; however, it will be shown in this section that the same estimation strategy applies identically to any several-parameter family of smooth surfaces.

A visible surface may be represented as a function $\mathbf{S}(x, y)$, where $\mathbf{S}$ is a surface orientation vector, $[\sigma, \tau]$, and (x, y) is the position in the image to which the surface point projects. In these terms, the set of planar surfaces is the set of surfaces for which $\mathbf{S}(x, y)$ is a constant function of (x, y).

Each image measure of tangent direction at a point on a contour consists of a triple (α^*, x, y), where α^* is the measured tangent angle, and (x, y) is the position in the image at which the measurement was made. The geometric/statistical model on which the planar estimator was based specifies a density function p.d.f.$(\alpha^* \mid \sigma, \tau)$. If (σ, τ) is given as a function $\mathbf{S}(x, y)$, this may be rewritten as p.d.f.$(\alpha^* \mid x, y, \mathbf{S}(x, y))$. Of course, for the set of planar surfaces, (σ, τ) does not depend on (x, y), so there is no reason to write the function this way, or to remember the position associated with a tangent measure, if $\mathbf{S}$ is known to be planar.

Suppose, though, that we were given a set of curved surfaces, $\{\mathbf{S}_1, \ldots, \mathbf{S}_n\}$ each with a known prior probability $p(\mathbf{S}_i)$, and were given the task of evaluating the probability $p(\mathbf{S}_i \mid A^*)$, where A^* is a set of measures of tangent direction and position in the image.

Let H_i be the hypothesis that $\mathbf{S}_i$ is present. Then under H_i, the surface orientation at a point (x_j, y_j) is given by $\mathbf{S}_i(x_j, y_j)$, and the density function for α^* *at that point* given H_i is

$$\text{p.d.f.}(\alpha^* \mid H_i, x_j, y_j) = \text{p.d.f.}\left(\alpha^* \mid \mathbf{S}_i(x_j, y_j)\right)$$

Now given a collection of image measures $A^* = \{(\alpha_1^*, x_1, y_1), \ldots, (\alpha_m^*, x_m, y_m)\}$ the likelihood of $A^* \mid H_i$ is

$$\prod_{j=1,m} \text{p.d.f.}(\alpha_j^* \mid H_i, x_j, y_j)$$

and the probability of $H_i \mid A^*$ is

$$p(H_i)\frac{p(A^* \mid H_i)}{\sum_j p(H_j)p(A^* \mid H_j)} = p(H_i)\frac{\prod_{j=1,m} \text{p.d.f.}(\alpha_j^* \mid H_i, x_j, y_j)}{\sum_{j=1,n} p(H_j) \prod_{k=1,m} \text{p.d.f.}(\alpha_k^* \mid H_j, x_k, y_k)}$$

This expression differs from the planar estimator in the nature of the hypothesis: instead of hypothesizing one or another surface orientation, which does not depend on position, we hypothesize one or another surface. Each hypothesis maps surface orientations onto positions, and the density function against which we judge each image measure is determined by the orientation of the hypothesized surface at the position of the measure. If the set of candidate surfaces is a family of a few parameters, instead of a discrete set, a density function corresponding to the probability function above is defined.

Thus, the problem of estimating the orientation of a planar surface using the tangent measure is just a special case of the estimation of a curved surface restricted to a family of several parameters: given the geometric/statistical model, an estimator for any such family is defined.

Some realistic problems entail the selection of a surface from such a family, for example the problem of establishing registration of an image with a surface model. This problem will be addressed in the next chapter, in the context of developing more powerful geometric models. But in the general case, this sort of global surface-fitting approach is not appropriate, because in the general case the surface cannot be restricted *a priori* to a known family of a few parameters—it might be *any surface.* Because the density of the image data is finite, a set of data has a finite number of degrees of freedom; and unless the set of hypotheses is far more restricted, meaningful estimation cannot be performed: just as a set of points can always be collocated by a polynomial of sufficiently high degree, an infinity of "perfect fit" surfaces can always be constructed by fixing the surface at the data points to optimize the statistic, and completing the surface arbitrarily. Clearly such surfaces have no meaning as estimates. The remainder of the chapter is devoted to the development of a local strategy more suited to the general case.

3.3 Estimating orientation locally

3.3.1 Why a local strategy is appropriate

The geometric/statistical model that was used to estimate the orientation of planar surfaces relates surface orientation at a point in the scene to the probability distribution for tangent direction at the corresponding point in the image; and *only* at the corresponding point. This relation, by itself, provides no link between a surface and its image, except at pairs of points that correspond projectively. It is in this sense a strictly local relation. In consequence, the geometric/statistical model, by itself, provides no justification for using data at a point in the image to infer surface orientation at any but the corresponding point on the surface.

Such justification could only come from additional knowledge or assumptions, that must somehow constrain the relation among disparate points on the surface. For example, a planarity restriction asserts that surface orientation is the same at every point, allowing the image data to be pooled without regard to their locations. Similarly, a prior restriction to a limited family of surfaces asserts that one of a specified set of global relations obtains among the points on the surface. Such restrictions make the task of estimation comparatively easy, but they simply aren't valid except in very unusual situations:

in the general case, the unknown surface may be any surface at all.

Natural surfaces are sufficiently varied and irregular that categorical assumptions about their shapes ought not to be made casually: it is hard to imagine any global restrictions that could describe natural surfaces well enough to be valid, and yet be powerful enough to be useful. And without such restrictions, an estimation strategy based on the methods already developed must be local, in the sense that the surface orientation obtained at one point is not influenced by image data at distant points.

3.3.2 The distribution at a point

Strictly, if no prior assumptions are made about the surface, then the estimated orientation at a point on the surface must derive entirely from measures on the corresponding image point. This requirement in turn determines the nature of the point measure that would have to be obtained to allow meaningful, strictly local, estimation to be performed: the geometric/statistical model gives a p.d.f. for image tangent direction as a function of surface orientation. The logic by which this relation can be used to infer surface orientation, given measures of image tangent direction, was explained in the last chapter: in essence, an observed *distribution* of tangent directions is compared to the theoretical distribution over the range of surface orientations, and the surface orientation for which the theoretical and observed distributions have the best fit is the maximum likelihood estimate. To apply this strategy using just a point in the image therefore requires that a *distribution* of tangent directions be available at that point.

3.3.3 An apparent dilemma

But in that case, meaningful estimation is clearly impossible, because contour data are available only at some points in the image: usually, no contour will coincide exactly with an arbitrary image point, hardly ever more than one. So an arbitrary image point will usually yield no measure of tangent direction; or one measure at most. And one such measure—no less none at all—is obviously insufficient to infer surface orientation. Thus arises a seeming dilemma: without making some assumptions about the surface, the estimate of surface orientation must be strictly local, yet a strictly local estimate cannot be made due to the limitations of the image data. In other words, non-local estimation appears to require some strong assumptions about surfaces, which are not justified, while local estimation requires some measure of the distribution of tangent directions at a point, which is not available.

3.3.4 The distribution *around* a point

If a strictly local measure of the distribution of image tangent directions is unavailable, and a non-local strategy requires apparently untenable assumptions about surfaces, then some compromise between strictly local and global strategies is indicated: if nothing can be inferred from an isolated image point, then a natural alternative is to examine a region around the point—as small as possible, but large enough compared to the density of the image contours to provide a reasonable data sample—and base the estimate on the surrounding data. Repeating this procedure across the image, in the

manner of a convolution, would yield an estimate of the surface.

This strategy, which is the one I will adopt, may be divided into two stages: first, computing the spatially extended distribution around each image point, and second, using the distribution to estimate surface orientation. While the first stage is strictly an operation on the image data, the second relates the image to the scene, and hence embeds an empirical claim about the relation between the distribution around a point in the image, and the orientation of the corresponding surface point. The first stage will now be characterized.

3.3.5 Computing the spatially extended distribution

The image data take the form of values of tangent direction, α^*, at various positions x, y in the image. That is, each measure is a triple (x, y, α^*). The distribution required at each point (x, y) is a function, $f(\alpha^*)$, giving the frequency with which each value of α^* has been observed in some region surrounding (x, y). Because this function must be given at every point, it may be denoted by $f(x, y, \alpha^*)$.

In its simplest form, this function might give, for each value of x, y, and α^*, the number of data points that lie within a circle of radius r in the image around (x, y) in the image, and whose tangent direction is separated by some amount δ or less from α^*. Since each data point maps into a point in the three-dimensional space (α^*, x, y), the function so defined gives at each point in that space the number of data points falling inside a volume surrounding that point. That volume is a circular cylinder of radius r and length δ, whose axis is normal to the (x, y) plane. The midpoint of the axis is (x, y, α^*) (see Fig. 1). In these terms, the function $f(x, y, \alpha^*)$ may be represented as a three dimensional convolution:

Let each image measure be represented as a function $u(\alpha^*, x, y)$, that assumes a unit value at the position of the measure, and is zero everywhere else. That is, for a given measure (α_i^*, x_i, y_i), we define a function

$$u_i(\alpha^*, x, y) = \begin{cases} 1, & \text{if } (\alpha^*, x, y) = (\alpha_i^*, x_i, y_i); \\ 0, & \text{otherwise,} \end{cases}$$

in which case a set of image measures can be represented as the sum

$$A^* = \sum_i u_i(\alpha^*, x, y).$$

Suppose that $f(x, y, \alpha^*)$ is to incorporate the data points that lie within a radius r of (x, y) in the image, and whose tangent direction is separated from α^* by a or less. Together, r and a define the volume within which the data points are to be counted. A function that assumes unit value inside that volume, and zero outside, is defined by

$$\xi(x, y, \alpha^*) = \begin{cases} 1, & \text{if } \Delta s \leq r \text{ and } \Delta\alpha^* \leq \delta; \\ 0, & \text{otherwise.} \end{cases}$$

where Δs is the distance between (x, y) and (x_i, y_i), and $\Delta\alpha^*$ is the distance between α^* and α_i^*. Then the desired frequency function is defined by

$$f(x, y, \alpha^*) = \xi(x, y, \alpha^*) * A^*(x, y, \alpha^*),$$

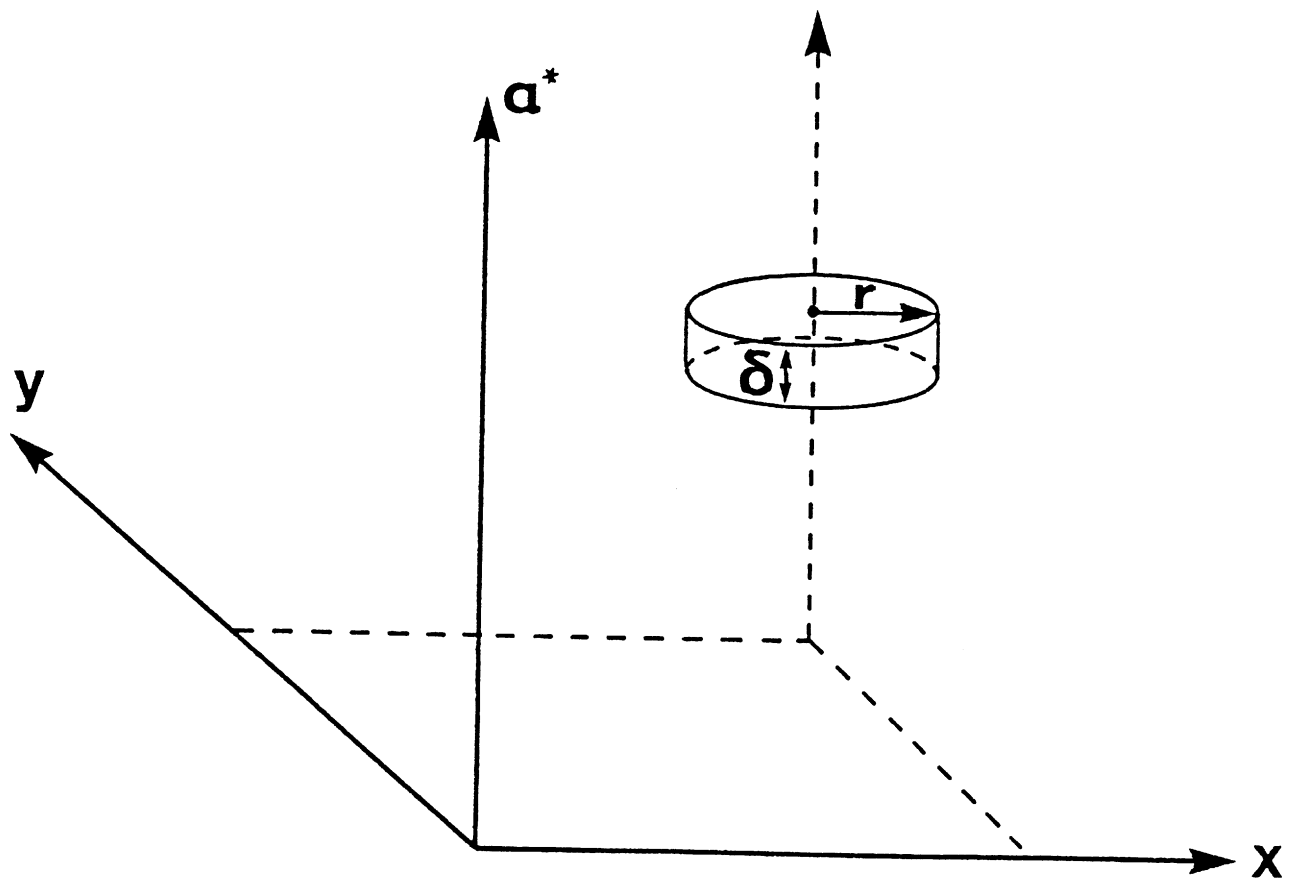

Figure 1. The image data map into points in the space (x, y, α^*). The data whose distance in the image to a point (x, y) is r or less, and whose tangent direction is within δ of some value of α^* lie within a cylinder in this space, with radius r and length δ. The axis, whose midpoint is (x, y, α^*), is normal to the (x, y) plane. The spatially extended distribution $f(x, y, \alpha^*)$ may be defined by counting the data points in such a cylinder around each point in the three-dimensional space.

where $*$ denotes convolution. The value of the convolution at each point (x, y, α^*) is the number of measurements whose distance in the image is r or less from (x, y), and whose tangent direction is separated by α or less from α^*. This function is not in general continuous, but, if desired, a continuous function may be obtained by replacing ξ with a weighted mask, whose value is a continuously decreasing function of Δs and $\Delta\alpha^*$.

3.3.6 Using the spatially extended distribution

The outcome of the computation outlined above is a representation of the image data that gives at each image point the distribution of tangent directions over a surrounding circular region. It remains to *use* that representation to estimate surface orientation at each point. The required estimator must take the distribution $f(x, y, \alpha^*)$ into a function $\mathbf{S}(x, y)$, where $\mathbf{S}$ is an estimated surface orientation vector. If the spatially extended distribution is regarded as the best available approximation to the unobservable distribution-at-a-point, then the natural candidate for this estimator, although not the only candidate, is the one developed in the last chapter; and it will be shown later that this estimator is in fact effective for a variety of natural images.

But, before proceeding to the strategy's implementation, it is necessary to arrive at a better understanding of the goal of the estimation strategy, and of the assumptions it entails: since the data in a region in the image derive from a corresponding region on the surface, a point estimate for orientation derived from those data, by whatever method, must depend in part on the behavior of the surface in that region; i.e., variations in surface orientation within the region will be reflected in the data, hence in the estimate. This means that the spatially extended distribution cannot be guaranteed to closely approximate the point distribution unless something is known about the shape of the surface in advance. For example, on a rough surface, the relation between the orientation at the center of a region, and that of the surround, may be unpredictable. Thus it seems that assumptions about the surface, albeit relatively local assumptions, are still required to relate a point to its neighborhood, if the goal of the strategy is to estimate orientation at a point.

Intuitively, one would expect variations of the surface within the radius r to tend to average out, and the estimate to reflect more nearly an *average* orientation in the region, than the exact orientation of the region's center. In that case, as the value of r increases, and the estimate at each point depends on a larger surrounding region, the estimate will reflect an average or overall orientation on a larger scale, and the estimated surface will lose detail, much as detail is lost when resolution is diminished by distance. And in that case, the estimation of this overall orientation, and not the orientation at a point, might best be taken explicitly as the goal of the strategy.

In the next section I will argue that the goal of the strategy *can't* be the estimation of orientation at a point, because there is no such thing as orientation at a point. Rather, the orientation of a physical surface is *always* an overall orientation over some area on the surface. The size of that area determines the *scale* at which the surface is described, but the orientation at a coarse scale is in no sense an approximation to the orientations at finer scales, and is no less accurate or correct than a finer orientation, except perhaps as determined by one's needs. On this view, physical surfaces don't have unique orientations at each point, but continua of orientations depending on scale. While this notion is at odds with the usual differential definition of surface orientation, that definition cannot be applied

coherently to physical surfaces. It is thus perfectly natural that the distribution be used to estimate orientation at a scale corresponding to its spatial extent.

3.4 Orientation and scale

This section will examine the notion of surface orientation as it applys to physical surfaces. The usual definition of orientation is differential, and by that definition, the orientation of a differentiable function is unique at each point, while that of a nondifferentiable function is undefined. We will see that this definition isn't quite right for physical surfaces: although we usually measure surface orientation in the same way that we approximate numerically the derivative of a function, we will find that the measure cannot be taken as an approximation to any unique "true" value. This view is supported by Mandelbrot's (1977) argument that many natural curves and surfaces behave like a class of nondifferentiable functions called *fractals*. Rather, to preserve our intuitive notion of surface orientation, we must regard orientation as a function of scale. That is, unlike a differentiable function, whose orientation is unique at each point, a physical surface has a continuum of orientations, each at its own scale, and none intrinsically more "correct" than another, except as dictated by one's needs. The orientation at a coarse scale is in no sense an approximation to those at finer scales; rather it is a property that only arises when orientation is measured over a sufficiently large spatial extent. The alternative to accepting a scale-dependent continuum of orientations, if we take the differential definition literally, is that physical surfaces have no orientation at all.[1]

The significance for the surface estimation problem of orientation's dependence on scale lies in the relation between the scale of a measured orientation, and the spatial extent over which it is measured: because the density of the image contour data is limited, an area of the image—which may have to be quite large—must be examined to characterize the "local" distribution of tangent directions. Unless a great deal is known in advance about the shape of the surface, it is hopeless to use this spatially extended distribution to recover surface orientation at a much finer scale than the size of the observed area—for example, there might be a small bump at the center of that area, whose orientation bears little relation to that of the surrounding surface. But corresponding to the observed image area is an area on the surface, that, however large, has an orientation of its own, apart from the orientations at smaller scales. The recovery of this overall orientation, discarding features at a smaller scale, is a far more plausible goal for the estimation strategy. In that case, the scale—or resolution—at which the surface is estimated ought to be determined by the spatial extent of the "local" distribution, which is in turn set by the density of the image data.

3.4.1 Surface orientation and differentiability

It is clear in common sense terms that the surfaces around us have orientations, that may be dis-

[1]Although local properties like orientation depend critically on scale, an interesting class of curves and surfaces show a more global scale invariance called *self-similarity:* the general shape of a coastline, or of the ocean's surface, appears the same over a wide range of scale. That is, a coastline comprises the same *kind* of bays and peninsulas at widely different scales, even though the particular bays and peninsulas differ. For curves and surfaces of this kind, the orientation assigned at a particular point depends on scale, but a common *statistical* description might apply over a wide range of scales.

covered by sight or manipulation, and that may be taken to be real and meaningful properties of those surfaces. If some definition of surface orientation denies that this is so, then the definition is surely wrong, at least for physical surfaces; rather than discard the useful notion of surface orientation, we ought in that case to discard the definition.

The usual definition of surface orientation comes from differential geometry. In differential geometry, a curve or surface is a function, a mathematical object. The orientation of a curve is defined in geometry by its tangent, that is, the first derivative with respect to arc length. The orientation of a surface may be defined by a unit normal to the tangent plane, and the tangent plane is also defined by the first derivative of the surface with respect to position. Thus, differentiation lies at the heart of the usual definition of orientation: wherever a curve or surface may be differentiated, its orientation is uniquely defined; elsewhere, it is undefined. Thus, a curve or surface that is nondifferentiable has no orientation.

Physical surfaces are are routinely described by differentiable functions. While any mathematical description of the physical world entails some idealization, such descriptions of surfaces are often useful and perfectly reasonable. In particular, the orientations defined by those descriptions usually correspond closely to the intuitive perceptual orientations of the surfaces being described. Thus, it might seem that defining the orientation of physical surfaces differentially poses no difficulty, so long as it is understood that the definition applies to differentiable descriptions of those surfaces.

However, Mandelbrot (1977) has argued that natural curves and surfaces in important respects behave more like nondifferentiable than differentiable functions, in particular with respect to arc length and surface area: to compute numerically the length of a curve, the curve may be approximated by a polygon consisting of straight lines of fixed length. For differentiable curves, the limit of this approximation as the length of the lines goes to zero is the exact length. When this approximation is computed, say, for a circle, using successively smaller steps to approximate the curve, the approximate length quickly levels off toward the exact length. But applying the same procedure to a natural curve, such as a coastline or river, gives a surprisingly different result, as shown by empirical data of Richardson (1961): as these curves are approximated in successively smaller steps, over a wide range, the "approximate" arc length increases continually, with no sign of leveling of, and apparently without bound, according to a function

$$L(\eta) = \lambda \eta^{1-D},$$

where L is the measured length, η is the length of the approximating line, and λ and D are constants.[2] Intuitively, as η decreases, the approximation incorporates ever smaller bays and peninsulas that add to the measured length. Since, for all practical purposes, this incorporation of ever smaller features continues without limit, Mandelbrot reaches the startling conclusion that coastlines are infinitely long!

To account for this curious behavior, which is shown to typify many natural phenomena, Mandelbrot proposes that coastlines, rivers, and terrain, be represented by a class of nondifferentiable functions called *fractals*. Many of the very interesting properties of these functions, and of the phenomena they describe, need not concern us here. Of particular importance is the picture they

[2]Since the empirical values of D fall between 1 and 2, the exponent is negative, and L increases without bound as η approaches zero.

give of natural curves and surfaces, as "bottomless pits" of structure at ever finer scales: each bay or peninsula, examined more closely, proves always to consist of a similar string of bays or peninsulas at a smaller scale. Practically any natural curve or surface displays this property when examined over a wide enough range of scales.

This view considerably complicates our common sense notion of measure: in a very real sense, coastlines and the like may be regarded as infinitely long. Yet we are accustomed to regard their lengths as definite finite values. For example, the lengths of international borders are often listed in atlases as if they were unique properties of those curves. In fact, Mandelbrot reports that the common border of Spain and Portugal is given radically different "official" values by those two countries; and he suggests that the difference reflects in part a different choice of η. If we accept Mandelbrot's argument, then such lengths reflect the arbitrary choice of the "yardstick" —the value of η—that was used to compute them. Crucially, such lengths have no meaning unless the associated value of η—the parameter of scale—is given; and therefore the closest we can come to preserving our common sense notion of measure is to allow that length and area are not fixed, unique properties of natural objects, but functions of a parameter of *scale.*

While Mandelbrot's argument is developed in terms of length and area, a corresponding argument applies to the tangent of a curve, and the orientation of a surface: the "yardstick" approximation applies equally to a curve's tangent as to its length. For differentiable curves, the direction of a line between two nearby points—the "yardstick" — approaches the true tangent direction as the length of the line decreases, rapidly stabilizing around that direction. But, on a natural curve like a coastline, the direction of the yardstick does not converge on a limiting value; rather, it flops around without apparent limit as it encounters smaller and smaller features of the curve. Once again, a measured tangent direction has no meaning, unless the length of the yardstick—the parameter of scale—is given. And tangent direction must be regarded not as a fixed, unique attribute of the curve, but as a function of scale; the same argument applies to surfaces when the measure is taken in two dimensions.

Thus, although the operation of measuring orientation on a physical surface resembles the operation of approximating a derivative numerically, the differential definition of orientation does not apply, because there is no limit—the "approximation" is all there is. Decreasing the scale of measurement does not give a *more accurate* value, as it would for a differentiable function, just a *different* value, at a smaller scale. No scale of description is categorically "best" or most accurate, except as dictated by the use to which the description will be put. Thus, an astronomer, a geologist, and a mountain climber each have a different idea of the "best" scale at which to describe the earth's surface.

To reconcile Mandelbrot's argument for the nondifferentiablity of natural surfaces with the usefulness of differentiable functions for describing them, we cite the observation of Perrin (1906), quoted by Mandelbrot:

"It must be borne in mind that, although closer observation of any object generally leads to the discovery of a highly irregular structure, we often can with advantage approximate its properties by continuous functions. Although wood may be indefinitely porous, it is useful to speak of a beam that has been sawed and planed as having a finite area. In other words, at certain scales and for certain methods of investigation, many phenomena may be represented by regular continuous functions, somewhat in the same way that a sheet of tinfoil may be wrapped round a sponge without following

accurately the latter's complicated contour."

Thus, the description of a physical surface by a differentiable function resembles the description of a curve by a particular polygonal approximation. In each case, the description is appropriate only at a certain scale; and a continuum of differentiable descriptions is defined as a function of scale.

3.4.2 Scale of description and area

The polygonal, or "yardstick" approximation, is just one way to introduce scale into the measurement of area or orientation: Mandelbrot gives four. But all of these measures behave in much the same way, sharing in particular the common feature that, the coarser the scale of description, the larger the area on the surface contributing to the measure "at a point." The analogue for surfaces of a polygonal approximation to tangent direction is a simple finite difference measure, taking the plane defined by three nearby points as the tangent plane. Other reasonable measures are the least-squares plane on a patch of the surface, and the derivatives of low-pass filtered or otherwise spatially averaged functions. In each case, the scale of description is determined by the spatial extent over which the measure is computed—the greater the extent, the coarser the scale. In this sense, surface orientation is a property not of the nominal point to which the orientation is assigned, but of an area surrounding the point, whose extent depends on scale.

While the measures listed above apply to collections of points on the surface, the data derivable from image contours are projected tangent directions. And the measure by which surface orientation is obtained must of course reflect the nature of the available data. But we may still expect to find the scale of description related to the spatial extent of the measurement: the spatially extended distribution of tangent directions reflects at each point the properties of an area on the surface. Since to any such area, however large, corresponds an orientation at a particular scale, it stands to reason that the spatially extended distribution is better suited to estimate the orientation at that scale than at others.

Next, these ideas will be applied to the related problem of inferring shape from shading information (Horn, 1975, 1977). It will be shown that the central features of the proposed strategy—computing a spatially extended measure on the image, and using that measure to obtain surface orientation at a corresponding scale—have already been applied successfully to the shape-from-shading problem: image intensity at a "point" is spatially extended by the imaging system itself, and the orientation at a "point" to be recovered is really an extended property, distinguished from surface "microstructure." The role of the estimator in the shape-from-contour strategy is precisely analogous to that of the reflectivity function in shape-from-shading.

3.4.3 An analogy to shape-from-shading

The basis for inferring surface shape from shading rests in the dependence of image intensity on surface orientation. The relation between them is expressed by the *reflectivity function*, which gives the intensity of light at a point in the image as a function of the orientation of the surface at the corresponding point, and the illumination incident on that point. But all imaging systems have finite resolution, so the image intensity at a "point" is really an integral over a small area around the point, and depends on the light incident on and reflected from a corresponding area on the surface. The size

of this area depends on the imaging resolution, and the distance from the surface to the viewer. So in this regard, shape-from-contour and shape-from-shading are identical: each begins with a measure on an area of the image, that derives from a corresponding area on the surface.

The reflectivity functions of surfaces have been modeled in terms of surface "microstructure," i.e. structure at a scale too small to resolve. For example, () modeled a small patch on a surface as a collection of facets whose orientations are distributed with radial symmetry around an overall orientation.[3] The gross reflecting properties of the surface were derived from the reflecting properties of the facets, and from the distribution of their orientations over the patch. The "real" orientation, i.e. the one entering into the reflectivity function, is the overall orientation, not the orientation of a single facet. As such, it is a property that arises at a scale much larger than the size of an individual facet. In these terms, the relation between image intensity at a "point" and surface orientation at a "point," as expressed in the reflectivity function, is really a relation between intensity taken over an area, and surface orientation taken over a corresponding area.

In this sort of treatment, the partitioning of the surface into gross shape and microstructure is determined by the limit of resolution, not by any intrinsic properties of the surface. When resolution changes, e.g. from a change in viewer-to-surface distance, features of the surface that cross the threshold of resolvability migrate between gross shape and microstructure; features that at high resolution contribute to a description of the surface's shape contribute at lower resolution to its reflectivity function.[4] Thus, both the "overall" orientation that describe the surface's shape, and the reflectivity function that relates its shape to the image, depend on the spatial extent over which image intensity is measured. Increasing that extent by reducing the resolution of the image causes smaller features of the surface to descend into microstructure, and the surface is recovered at a coarser scale.

The analogy to the shape-from-contour problem is direct: shape-from-shading begins with a measure of intensity on an area of the image, while shape-from-contour begins with a measure of the distribution of tangents on an area. The former measures the *amount* of light incident on the area, while the latter measures aspects of its *spatial distribution.* While in shape-from-shading the spatial extent of the measure is determined by the resolution of the image, in shape from contour it is determined by the density of the contour data. Thus characterizing the distribution "at a point" using data from a surrounding region is closely analogous to reducing the resolution of the image.

At the heart of the solution to the shape-from-shading problem is the reflectivity function, which we have seen relates spatially extended image intensity to surface orientation at a corresponding scale. Smaller features of the surface are relegated to microstructure, and are not considered part of the surface's "real" shape. The reflectivity function thus plays the exactly same role in shape-from-shading that the estimator must play in shape-from-contour. Only the nature of the image measure, and the constraints on its spatial extent are different. In fact, the estimator might be viewed as a "geometric reflectivity function."

[3]The assumption of radial symmetry might be taken as an implicit definition of the overall orientation as the mean orientation of the facets.

[4]This observation is due to David Marr (1978)

3.4.4 Choosing an estimator

Several lessons can be gained from this analogy. Most generally, surface orientation *can* be inferred using spatially extended measures on the image, at a scale determined by the spatial extent; i.e. shape-from-shading works. Of more specific value is the relation between the reflectivity function and the statistical estimator for shape-from-contours: reflectivity functions may in principle assume diverse and arbitrary forms, and the shape-from-shading problem cannot be solved unless the reflectivity function has been characterized reasonably well. And, as seen in the various analytic treatments of the subject (refs.), reflectivity functions may be modeled in extremely complicated ways, down to the level of physical optics. But happily, the potential complexity and diversity of reflectivity functions is simply not a problem in practice: Horn's treatment of the recovery of shape from shading, by far the most complete and successful, barely drew on the elaborate analytic treatments. Instead it turns out that a few simple functions comprise a good enough approximation to nature to solve the problem. After all, the goal of the enterprise is not to understand the reflecting properties of surfaces in the greatest possible detail, but to recover their shapes with reasonable precision.

Similar problems arise in principle in the shape-from-contour problem: the "geometric reflectivity function" that is needed to solve the problem must express statistically the relation between the distribution of tangents in the image and the orientation of the surface. Just as the photometric reflectivity function may depend in complicated ways on surface microstructure, so may its geometric counterpart. Each function might in principle assume nearly any form. We might be led to modeling the microstructures of surfaces and their markings in excruciating detail, and deriving from such models a maximum likelihood estimator for an average or other overall orientation, but the lesson from shape-from-shading is that this simply shouldn't be necessary.

Ultimately, the choice of an estimator, or collection of alternative estimators, is an empirical issue: the measured distributions of tangent directions in images, and the way those distributions transform with changing surface orientation, are subject to empirical investigation. It may turn out that different estimators are required to recover surfaces whose fine structures differ radically. For example, the image measures that derive from a boulder field in direct sunlight with low elevation would reflect a complicated mixture of occluding contours, terminators, and cast shadows, as well as surface markings. The relation between such measures and the overall orientation of the boulder field might be quite different than that for a smoother surface with markings in low relief. It may well be necessary to draw some crude distinctions of this kind, just as the reflectivity functions of glossy and matte surfaces must be distinguished.

The implementation to be reported in the next section adopts as an estimator the one applied to planar surfaces in the last chapter. Since this estimator was developed on the assumption that the measured tangents correspond to surface points with the same orientation, it is better suited to markings in low relief on comparatively smooth surfaces, than to surfaces that are deeply textured at the scale of the contours, like a boulder field. While it will be shown to yield good estimates for a variety of natural surfaces, it should to be regarded as a provisional choice, subject to empirical elaboration or improvement.

3.5 Implementation and results

The implementation parallels in many respects that of the planar method presented in the last chapter. Tangent data were obtained from the image by the method already described, taking the gradient along zero-crossing contours in the $\nabla^2 G$ convolution (Marr & Hildreth, 1979). The principal difference lies in the need to preserve position information. The spatially extended distribution was computed by approximating the three-dimensional convolution of (3.1) by a series of two-dimensional convolutions. The outcome, at each image point, is a histogram of the observed tangent angles in a circular region surrounding the point, identical in form to the grouped representation employed in the planar case. The surface is then estimated by repeating the planar estimation procedure at each point. A map of relative depth can in principle be obtained by integration, but this procedure has the undesirable property of error propagation, and so compromises the local character of the strategy.

An important aspect of the computation is the choice of r, the radius of the summation mask. Rather than attempt to set that parameter automatically, the computation was repeated for each image at several values, and the results are compared.

Estimation was performed on several natural images, and the results are compared to the perceived shapes and orientations of the surfaces. The strategy is shown to be capable of producing good "coarse" descriptions of natural surfaces.

3.5.1 Image digitization and Contour extraction

Tangent data were obtained by the methods described in the last chapter: photographs were digitized on an Optronics photoscanner. The digitized images were then convolved with a $\nabla^2 G$ function, and the zero-crossing contours of the convolution extracted, by the method of Marr & Hildreth (1979). These contours are peaks in the first derivative of intensity in the band-passed image. Tangent direction was sampled along the zero-crossing contours of the convolution by taking the normal to the gradient vector at intervals on the curves. For each measure, the tangent angle, and the position in the image at which the measure was taken, were recorded as a triple (α^*, x, y).

3.5.2 Computing the spatially extended distribution

Since computing a three dimensional convolution was not feasible, the function of (3.1) was approximated by a series of two-dimensional convolutions: the α^* dimension was broken into seven equal intervals between 0 and π, so that each measured tangent direction fell in one interval. For each interval was constructed a two-dimensional array, whose coordinates corresponded to position in the image—orientation planes. Thus, each data point maps to the plane determined by its tangent direction, and a position in the plane corresponding to its position in the image. For each data point, the corresponding cell, initially zero, was incremented by one.

Note that summation on an area of one plane then gives the number of data points in the corresponding area of the image, whose tangent direction lies on the interval specified by the plane. In consequence, the convolution of a plane with a circular, unit-value mask of radius r—a pillbox—gives

at each point the number of data points inside a cylinder of radius r and length $\pi/7$, which is a value of the convolution in (3.1). Convolving each plane with such a mask gives $f(x, y, \alpha^*)$ across the image, at the seven values of α^* corresponding to the midpoints of the intervals (Fig. 2) Also note that a column—the values from each plane at some position (x, y) gives a histogram of the frequency of α^* in a surrounding image region of radius r, and this histogram has the same form as the grouped representation of the planar data (Fig. 3).

3.5.3 Computing the estimate

The estimator to be used was given in (2.3):

$$\text{p.d.f.}(\sigma, \tau \mid A^*) = \text{p.d.f.}(\sigma, \tau)\ \text{p.d.f.}(A^* \mid \sigma, \tau)$$

$$= \prod_{i=1,n} \frac{\pi^{-2} \sin\sigma \cos\sigma}{\cos^2(\alpha_i^* - \tau) + \sin^2(\alpha_i^* - \tau)\cos^2\sigma},$$

normalized by its integral with respect to (σ, τ), and for data of the grouped form obtained from a column,

$$\text{p.d.f.}(\sigma, \tau \mid A^*) \approx \exp\left(\sum_{i=1,n} a_i \log\left(\frac{\pi^{-2} \sin\sigma \cos\sigma}{\cos^2(\alpha_i^* - \tau) + \sin^2(\alpha_i^* - \tau)\cos^2\sigma}\right)\right),$$

(where a_i is the number of measures in the ith tangent direction plane), also normalized by its approximate integral,

$$\sum_i \sum_j \text{p.d.f.}(\sigma_i, \tau_j \mid A^*).$$

To estimate the surface orientation, (σ, τ), at a point (x, y), the value of (σ, τ) for which this function is maximized is found. The surface is estimated by repeating this procedure across the image.

3.5.4 Computing a map of relative depth

Distance to the viewer may in principle be obtained up to an additive constant by integrating the gradient space representation of surface orientation, if the surface is continuous. However, this procedure has the undesirable effect of propagating local errors in the surface orientation estimate. This propagation can be attenuated by integrating from different starting points, and averaging the results, effectively smoothing the estimate. It is not clear that this step is necessary or desirable, because it compromises the local character of the method.

3.5.5 Results

Figures 4 and 4a illustrate the entire estimation process, starting with a digitized image, computing the $\nabla^2 G$ convolution, extracting the zero-crossing contours, convolving the tangent direction planes,

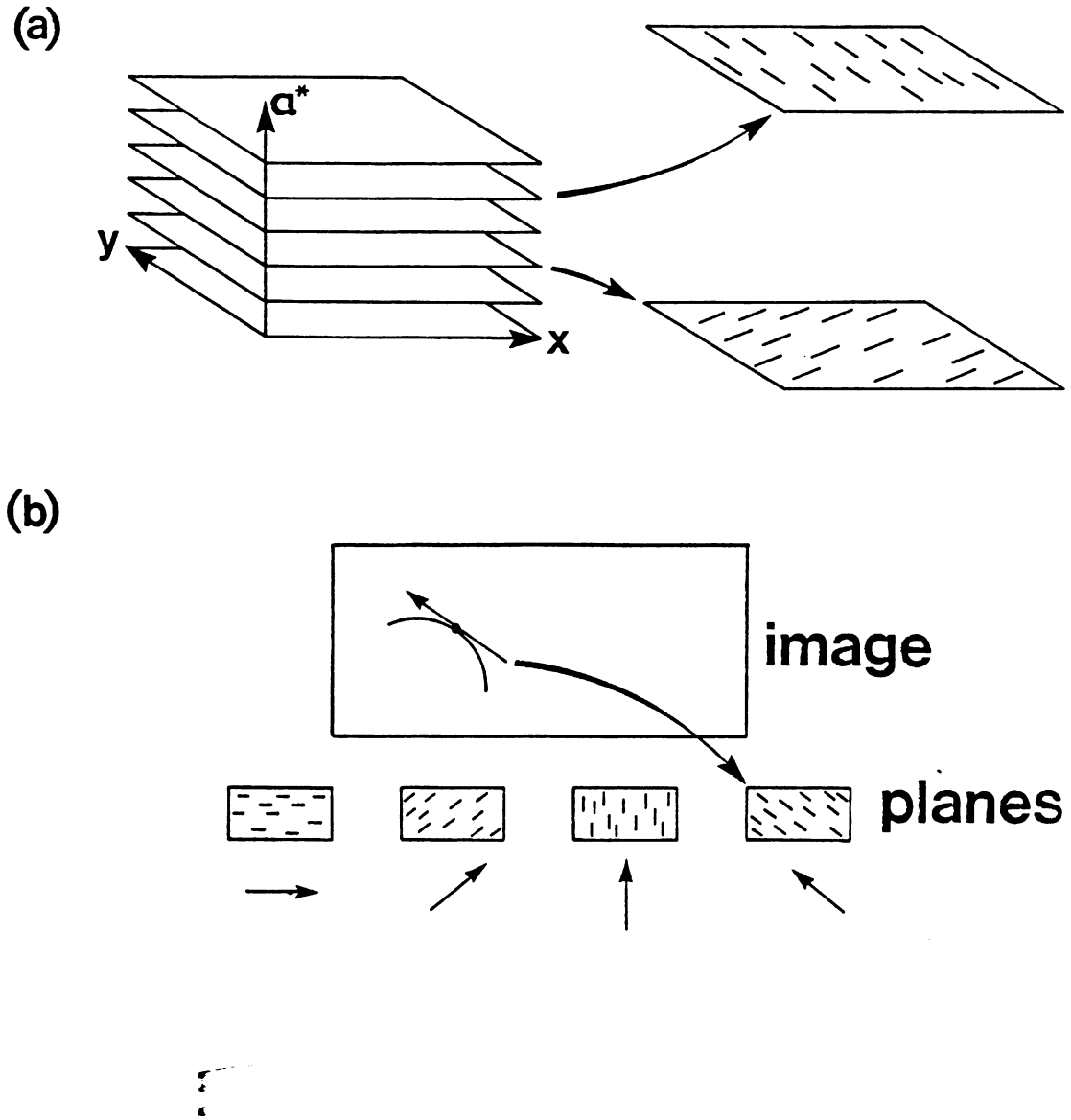

Figure 2. Orientation planes: the tangent direction dimension is divided into each interval. For each interval, a duplicate image is constructed. Thus, each data point maps into the plane determined by its tangent direction, and the cell in that plane corresponding to its image position. For each data point, the corresponding cell is incremented by one.

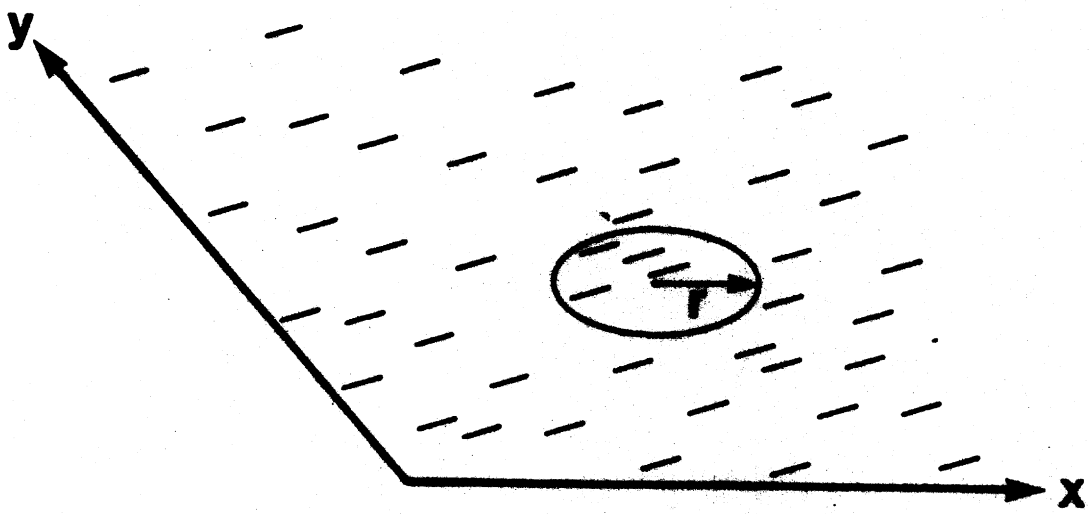

Figure 3. Summation of a plane over a circular area gives the number of data points that lie within a cylindrical volume in the space (x, y, α°). The cylinder has radius r and length $\pi/7$. Each such summation is a value of the convolution $f(x, y, \alpha^\circ)$.

and estimating the surface. The surface is represented by a collection of ellipses, as if the surface were covered by circles of constant size and uniform density. Perspective is added to the picture using a depth map obtained by integration. In this photograph, the contours derive primarily from the pattern of shadows cast through overhanging trees. The estimated surface corresponds closely to the shape perceived in the original photograph.

Figure 5 uses the same image to illustrate the effect of varying r, the radius of the summation mask. At one extreme, a single overall orientation is assigned to the entire surface. At the other, the amount of data contributing to each local estimate becomes so small that orientation varies erratically with position, bearing little relation to the actual shape of the surface. Over a wide intermediate range, the estimate portrays the surface reasonably well.

Figures 6 through 9 show several additional images, and the estimates obtained from them. Figure 7 shows a good estimate obtained from a more complicated picture. Figures 8 and 9 demonstrate two failures of the strategy: the first, a Viking picture of the Martian surface, fails because the contours arise from a high-relief texture of rocks. Such textures transform differently with projection than those in low relief, and so must be modeled differently. Figure 9 shows a systematically elongated texture of human hair. It should be noted that such textures, seen without disambiguating context, appear incorrectly as waving surfaces to the human observer as well.[5]

3.5.6 Using perspective information

The method just presented uses an observed distribution of tangent directions to estimate the foreshortening distortion, and hence surface orientation. The method fails when distortions of the surface markings themselves *mimic* projective distortion, as when the markings are systematically elongated.

While the estimates were computed assuming orthographic projection, the images actually include perspective distortions, that is, dependence of image metric properties on distance. While these distortions are generally much smaller than the foreshortening distortions on which the estimates were based, they offer a simple check of the estimates' reliability: actual foreshortening and perspective distortions on smooth surfaces are rigidly linked by projective geometry. On the other hand, systematic distortions of the surface markings that *mimic* foreshortening are most unlikely to co-occur with distortions that *mimic* perspective in just the same way: effects that elongate surface markings don't usually cause density to vary most rapidly orthogonal to the axis of elongation, and vice versa. In other words, *real* perspective and foreshortening are geometrically linked, while distortions of similar appearance, but not of projective origin, are likely to be independent.

It may thus be possible to distinguish projective from non-projective effects by comparing apparent foreshortening with apparent perspective: if the relation between them is roughly consistent with the relation predicted by projective geometry, then the observed effects are probably due to real projective distortion, and the interpretation based on that assumption is probably accurate. If not, the interpretation may well be wrong.[6]

[5] A striking example is the formation known as *landscape agate*.

[6] That some consistency test of this kind plays a role in human perception is suggested by Stevens' (1979) observation that it is difficult to induce the impression of slant unless perspective and foreshortening cues are consistent.

Figure 4. Computing the estimate: (a) a digitized photograph, (b) convolution of the image with a $\nabla^2 G$ function, (c) zeros of the convolution. The circle in (a) denotes the size of the summation mask used to compute the spatially extended distribution, i.e. the size of the area contributing to the estimate at each point.

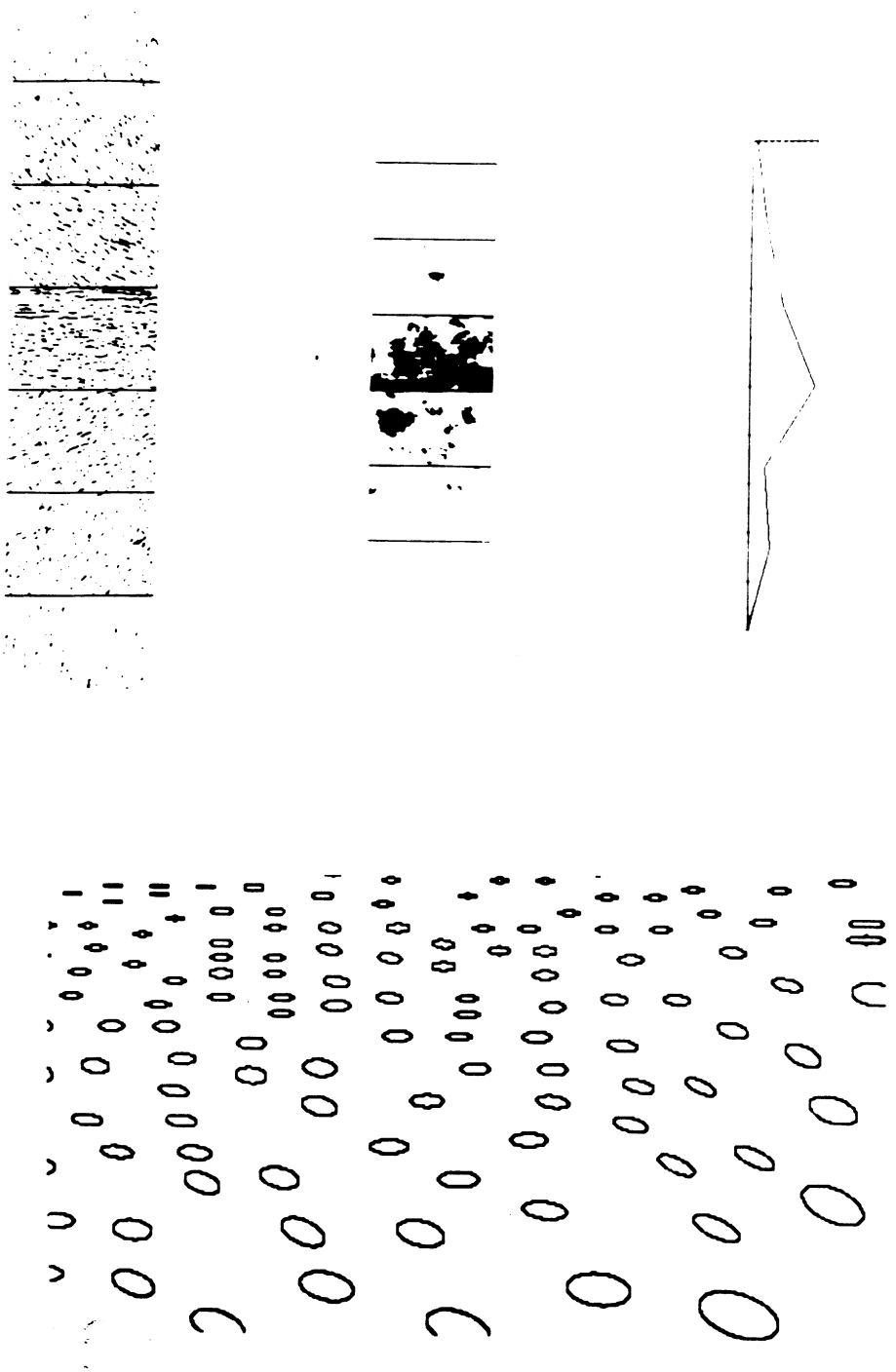

Figure 4a. Computing the estimate: (a) tangent direction planes, (b) pillbox convolutions of planes, (c) a column through the convolved planes at one image point, and (e) the estimated surface. Orientation is represented by ellipses, as if the surface had been covered with circles, and then projected. A perspective effect is added using a depth map obtained by integration. Note that the overall orientation coincides with that perceived in the original image, as does the increase in slant moving from foreground to background.

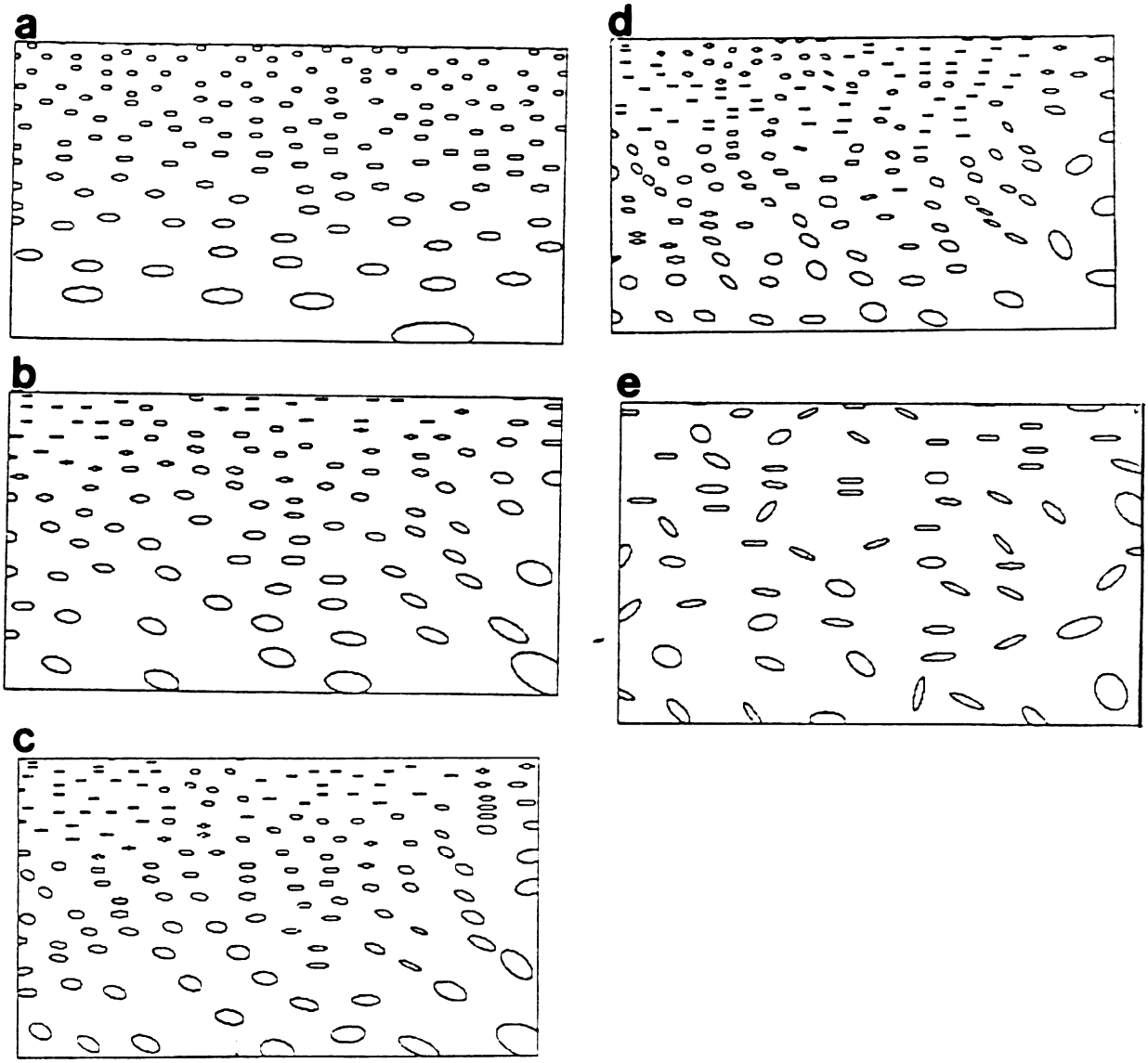

Figure 5. The effect of r, the summation mask radius, on the estimated surface. (a) the limiting case of a mask covering the entire image, obtaining a single overall orientation, (b) and (c) intermediate sizes that portray the surface reasonably well. (d) and (e) show the deterioration of the estimate when the averaging radius is to small compared to the density of the data.

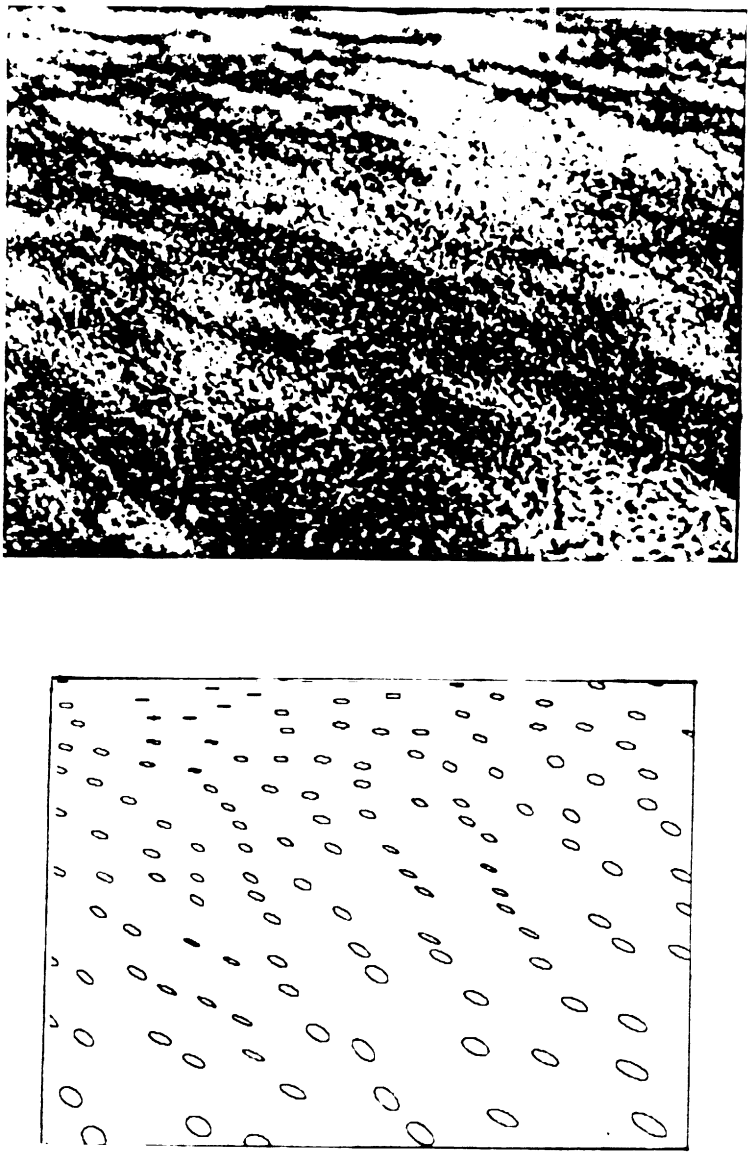

Figure 6. An additional image, and estimated surface, similar to that pictured in Fig. 5.

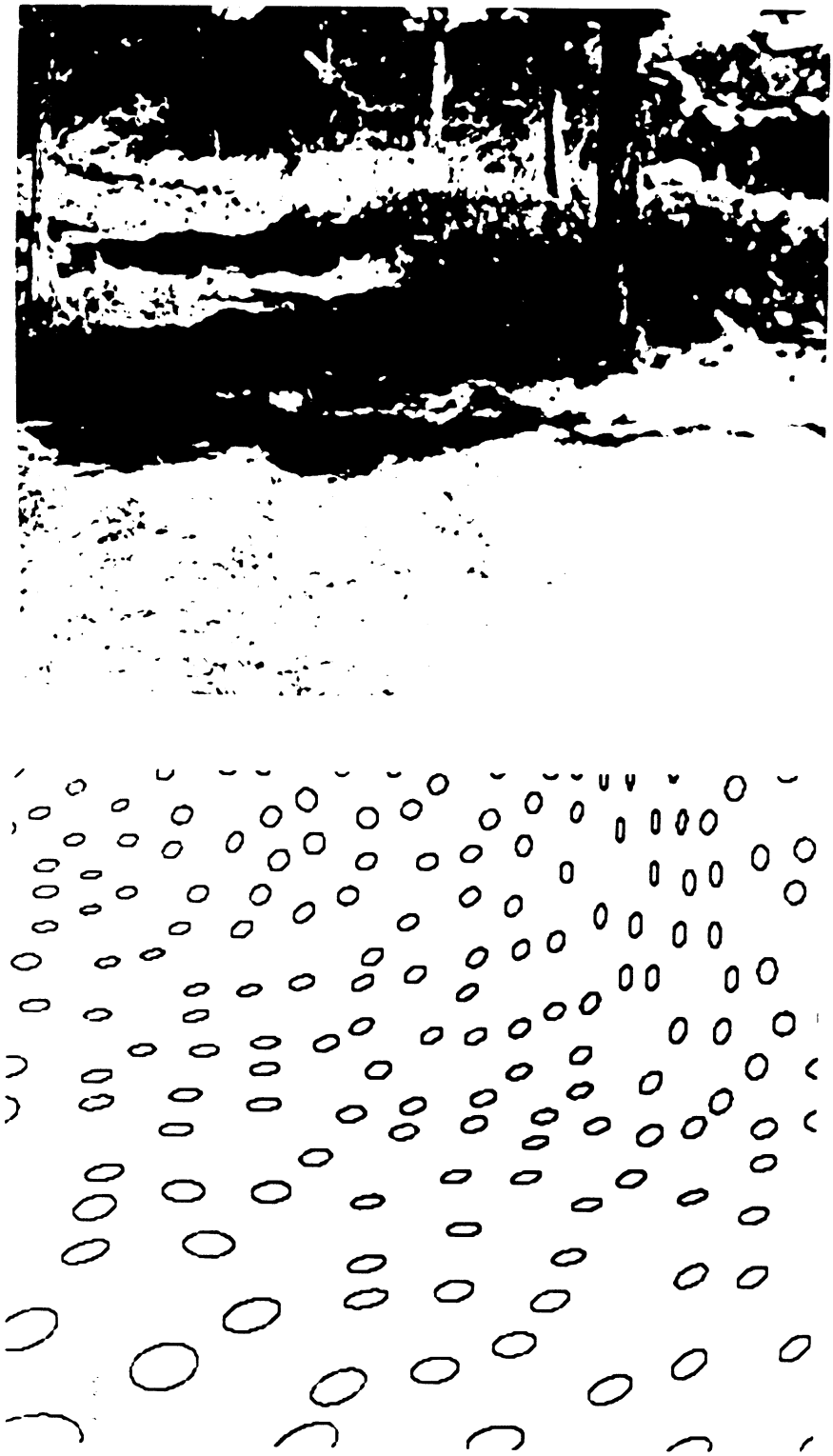

Figure 7. A more complicated image, and the estimated surface. Note that the estimate correctly distinguishes the highly slanted foreground from the more nearly frontal background. The upward pitch of the right foreground is also detected. Since the strategy doesn't know about discontinuities in depth, it is confused by the trees in the right background. Such marked local distortions might be used as evidence of a surface discontinuity.

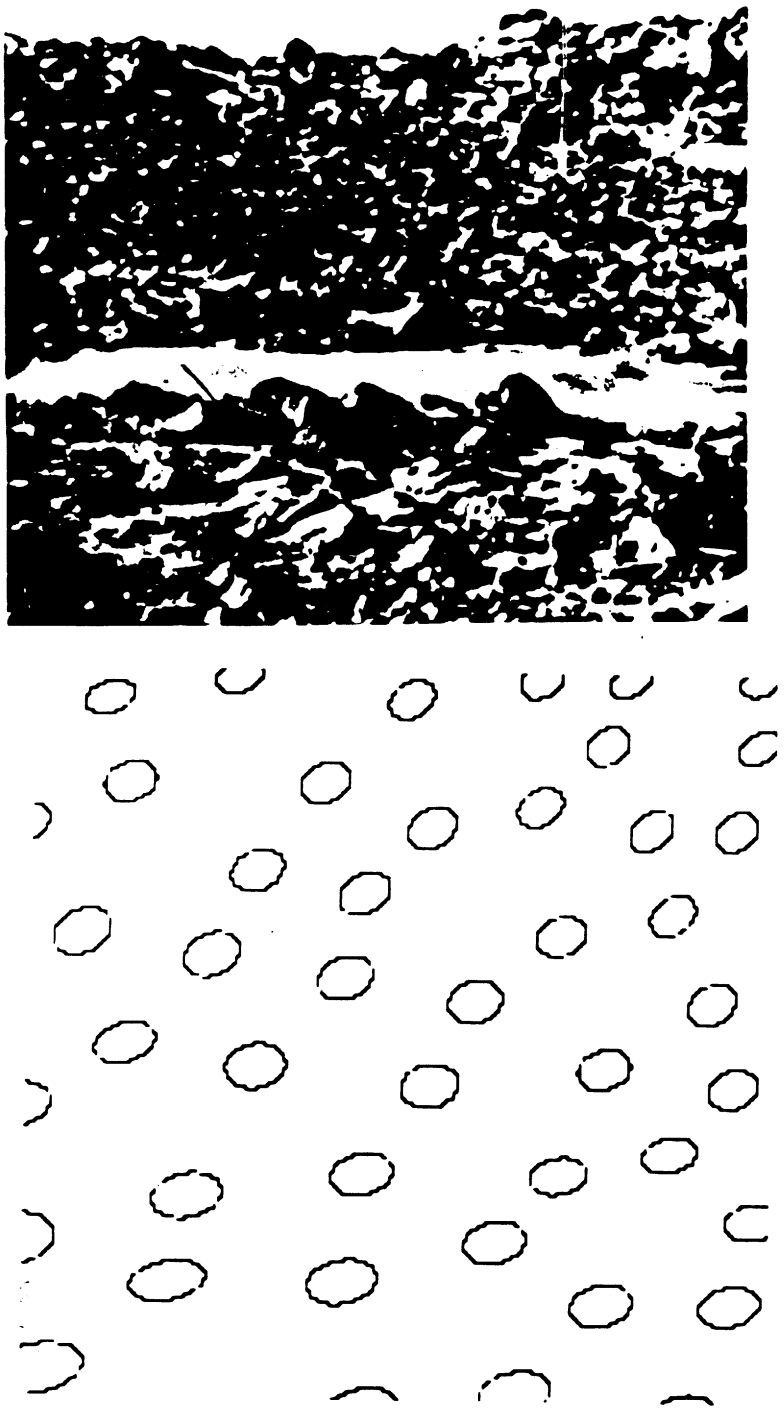

Figure 8. A Viking picture of the Martian surface. The high-relief texture of rocks does not transform with projection in the same way as those in low relief. The strategy therefore systematically underestimates the slant of the surface. High-relief surfaces must be modeled differently.

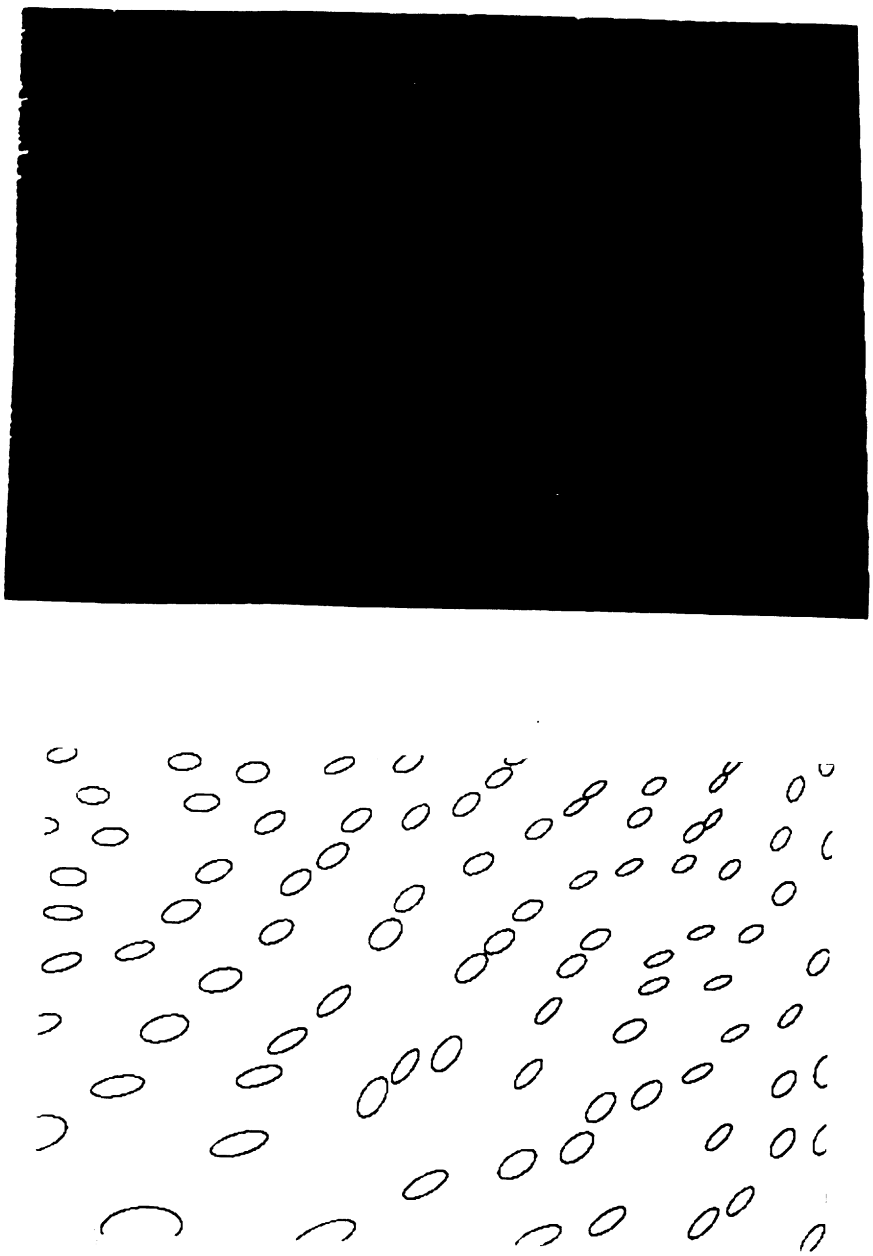

Figure 9. A hair texture, whose systematic elongation deceives the estimation strategy. But such surfaces, viewed without disambiguating context, deceive the human observer in much the same way.

A marked change in size with distance is evident in the images treated above, particularly those of figures 5 and 6. In those images, the change in size is consistent with the estimated surfaces. On the other hand, the foreshortening-like elongation of figure 9 is not accompanied by a consistent gradient of size. If the change in size could be measured in natural images, perspective could add substantial information to the estimate. A promising approach for this measurement is the change in spatial spectral content with position (Bajcsy, 1970). Since the present method entails convolution of the image with band-limiting masks of several sizes, the change in spectral content might be assessed by comparing the power convolutions with masks of different size.

CHAPTER 4

USING SURFACE CURVATURE

4.1 Introduction

Anyone who has flown on a sunny day has probably noticed the plane's shadow on the ground below. As it moves, especially over rugged ground, the shadow bends, twists, and wrinkles in conformity to the terrain. The phenomenon is particularly vivid with moving shadows, but it tells us something about any shadow: the shape of the shadow changes with the shape of the ground, and hence depends on the shape of the ground. In the sense I've used the term, the shape of the shadow thus *encodes* the shape of the ground. In this chapter, the geometry of the encoding process will be untangled, and it will be shown that the information implicit in the curvature of a cast-shadow contour can in fact be used to draw inferences about the surface onto which the shadow is cast.

4.1.1 Untangling surface curvature

The last chapter dealt with the estimation of curved surfaces, but only by the local estimation of surface *orientation;* surface *curvature* was never treated explicitly. The purpose of this chapter is to show that, just as the *orientation* of image contours provides information about the *orientation* of a surface, so the *curvature* of image contours provides information about the *curvature* of a surface. And by understanding the transformation, or "encoding" process, the information latent in image contours can be more fully used.

At the heart of the strategy based on tangent direction was the isolation of a projective component in the image measures. The same approach will be followed in the estimation of surface curvature: the measurable curvature of the image contour must be geometrically decomposed in terms of the scene properties it depends on, among which is surface curvature. Then that component must be isolated statistically.

Cast shadows provide the ideal contour type on which to perform this decomposition. Although the details of shadow geometry are rather complicated, the process by which shadow contours appear in the image is geometrically regular, and has a clear decomposition into causally independent components. Surface curvature naturally arises as one of those components. The corresponding decomposition for surface markings is more difficult, because, unlike shadows, their generation is less readily

described in uniform geometric terms.

4.1.2 An application to image registration

This chapter focuses on the geometry of cast-shadow contours, leading to a measure of "goodness-of-fit" between a hypothesized surface and the image data, using the curvature of the shadow contours. The usefulness of this measure will then be demonstrated by application to the restricted problem of establishing *registration* between an image and a surface model: registration will be established given only the shapes of shadow-contours, cast by objects of unknown shape, but assuming the direction of illumination to be known. The registration problem arises when an image shows a known surface, but the exact point of view from which the image was taken is not known. To map features of the image onto the corresponding locations of the surface, the image and the surface model must be brought into spatial register.

The registration problem has been treated successfully by Horn & Bachman (1977) in the domain of LANDSAT images and digital terrain models, by cross-correlating a synthetic image with the satellite image. While effective, this method is limited because the correlation is performed in the image domain. Thus anything appearing in the image to be registered, but not in the synthetic image (such as shadows cast by objects external to the terrain model) enters into the correlation as noise. The method to be presented in this chapter is also based on cross-correlation, but in a more abstract domain that compares expected with observed contour curvature, putting to good use features that would have to be dismissed as noise in the image domain. This method is not offered as a solution to practical registration problems, but as a demonstration that the shapes of shadow contours are meaningful, once they are understood.

4.2 Geometric model

A cast shadow arises when an opaque object is interposed between a light source and a reflecting surface. A cast-shadow contour is just the edge of the shadow's projection into the image. The purpose of this section is to express the geometric relation between the curvature of a cast-shadow contour, and its determinants in the scene, notably the curvature and orientation of the surface onto which the shadow is cast. Although the details of this geometry are somewhat complicated, the process it describes is easy to grasp intuitively. Before proceeding to the formal development, an intuitive account will be given.

4.2.1 Intuitive geometry of cast-shadow contours

We can begin by tracing a typical ray of light from its origin at the light source, to its eventual destination along a cast-shadow contour in the image. The light from a distant source, like the sun, can be idealized as a parallel bundle of rays, coming from the direction that the viewer points to when he points to the source. We call the common direction of these rays the *direction of illumination.*

When a light ray hits a diffusely reflecting surface, it will be reflected in all directions. If the point

of contact is within the viewer's line of sight, and unobstructed, some of that light will eventually reach the image.

If a solid, opaque object is interposed between the source and the reflecting surface, all of the rays that intersect that object will be stopped—reflected or absorbed, but the rays that don't intersect will reach the reflecting surface, and after reflection, the viewer. In that case, the interposed object (*shadowing object*) casts a shadow on the reflecting surface (*shadowed surface.*) The region on the shadowed surface whose light rays have been blocked by the shadowing object will be inside the the shadow, the rest will be outside (fig. 1).

To land on the *edge* of the shadow, a ray from the source has to just miss being blocked by the shadowing object—if we moved it over a little in one direction, it would be blocked; in the other direction, it would pass the object. That is, the rays that just *graze* the surface of the shadowing object will land on the shadowed surface just along the edge of the shadow. The projection into the image of the ray's point of contact with the shadowed surface will then be a point on the *shadow contour.*

So to trace a ray from the source to a point on the shadow contour in the image, we first draw a line, parallel to the direction of illumination, starting at the source, and just grazing the shadowing object. We continue that line until it hits the shadowed surface, to locate a point on the edge of the shadow. Then, to project that point into the image, we draw another line from the shadowed surface, through the optical focal point, until it hits the imaging surface. That point of contact is a point on the shadow contour in the image (fig. 2).

To construct the entire cast-shadow contour, we would have to repeat this procedure for all of the rays that graze the shadowing object. But this construction can be easily visualized by considering that the set of all the grazing rays together define a surface in space. All of these rays are straight lines, and they all pass through the (point) source, so that surface is a cone of general cross-section (fig. 3).[1] This surface will be called the *shadow cone,* and can be envisioned as dividing space into two regions: anything lying inside the cone will be in shadow, anything outside, in light. This is just like the cone of light from a movie projector, except it is a cone of *dark.* The shape of the shadow cone's cross-section depends on the shape of the shadowing object, and its position relative to the source. The direction of the cone's axis is the direction of illumination.

Recall that the grazing rays define the edge of the shadow, where they contact the shadowed surface. Since all the grazing rays also lie on the shadow cone, the edge of the shadow is actually *the curve of intersection of the shadow cone and the shadowed surface.* And the image contour is just the projection of that curve of intersection.

As a final step, the intersection of the shadow cone with the shadowed surface, and the projection into the image, must be expressed in terms of local geometry. The curve of intersection between two surfaces can be specified in terms of the curvatures and orientations of the two intersecting surfaces—and this is how the curvature and orientation of the shadowed surface finally enter the picture. Another function takes the curve thus specified into the image, specifying the curvature and tangent direction of the contour—the quantities we can measure in the image. We wind up with a function relating the image curvature to the curvature and orientation of the shadowed surface, the curvature of the shadow cone, and the direction of illumination. Next this function will be derived.

[1]In fact, since we'll deal with a source at infinity, all the rays are parallel, and the "cone" is really a cylinder.

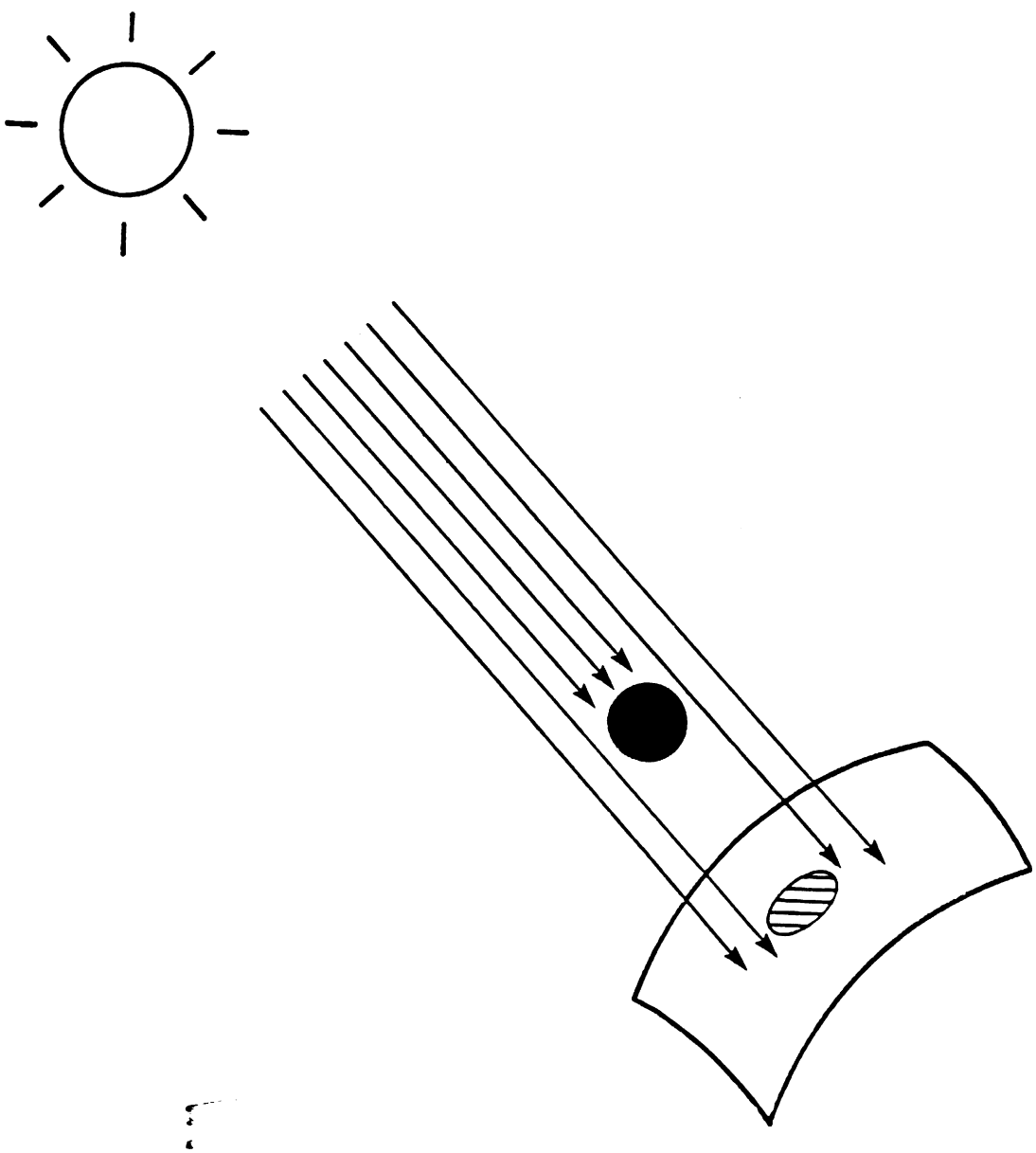

Figure 1. From the source to the shadowed surface. The light from a distant source may be visualized as a parallel bundle of rays. A cast shadow is demarked by the rays that intersect an opaque object, before they reach the shadowed surface.

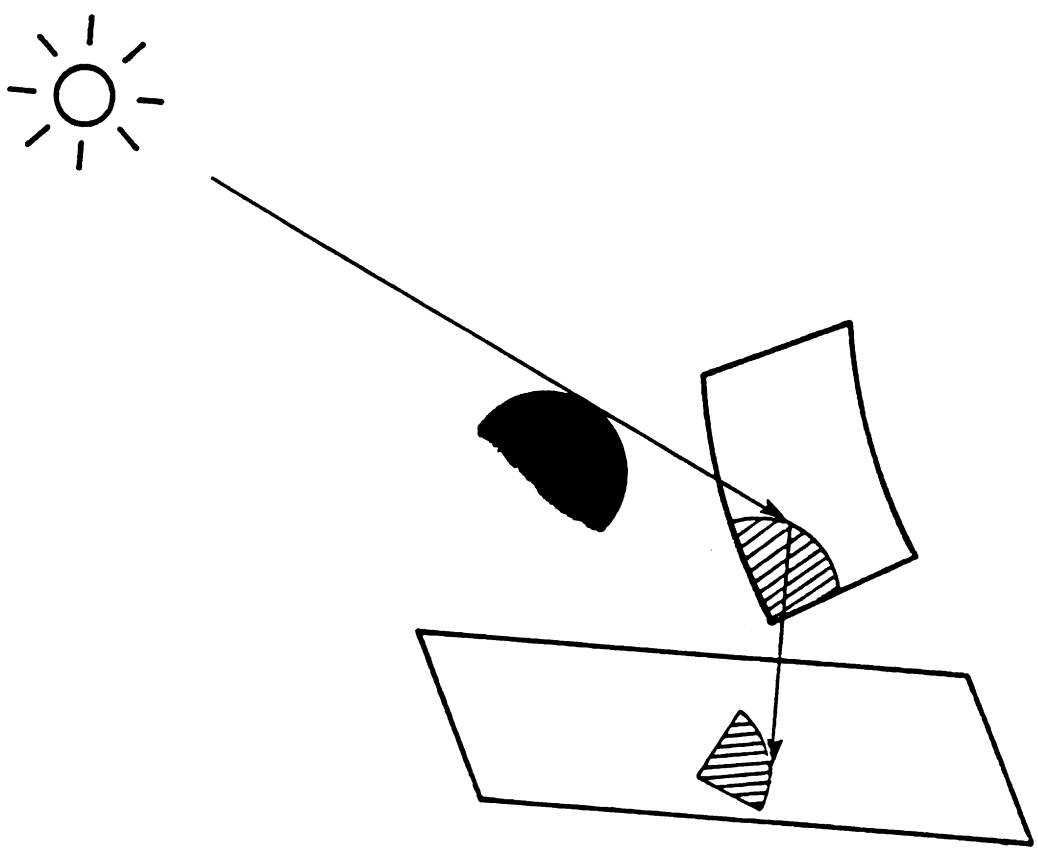

Figure 2. From the source to the image: a ray from the source that just grazes the shadowing object, contacts the shadowed surfaces, then projects to the image, giving a point on the shadow contour.

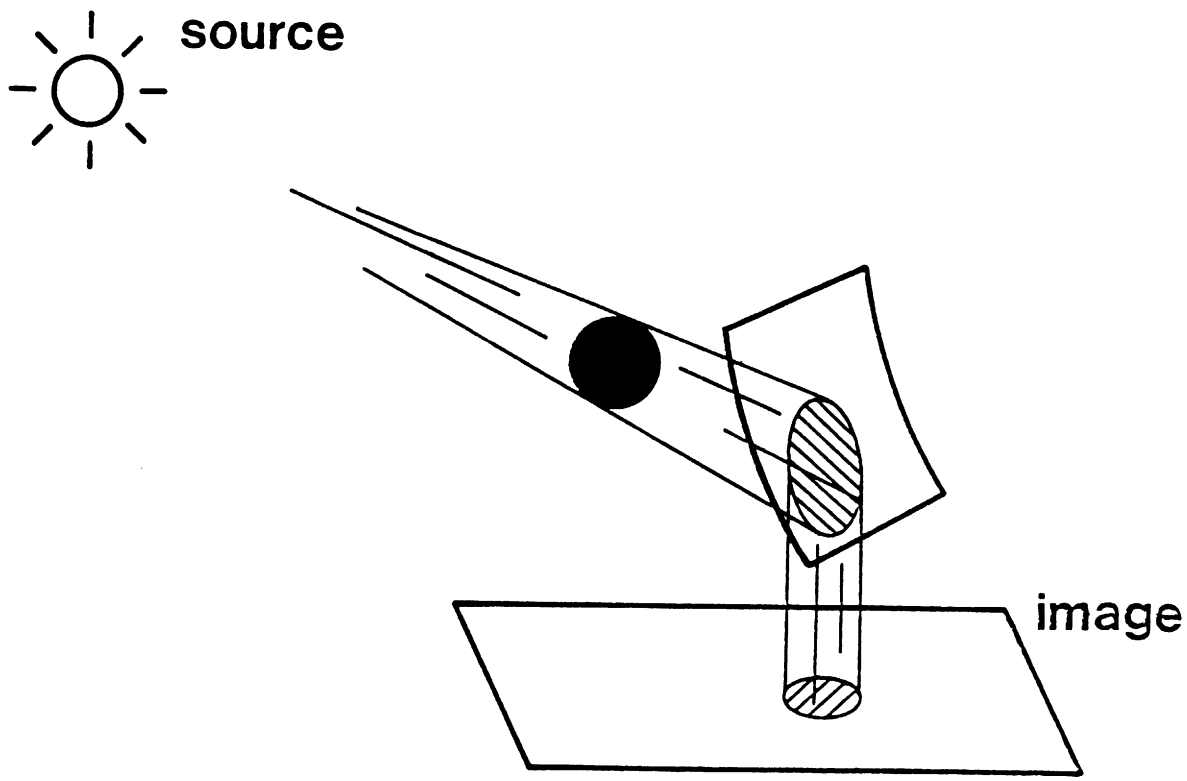

Figure 3. The shadow cone: a surface in space, defined by all the rays from the source that just graze the shadowing object. Anything inside this cone will be in shadow, anything outide, in light. The edge of the cast shadow is the curve along which the shadow cone intersects the shadowed surface; and the image contour is the projection into the image of that curve of intersection.

4.2.2 Notation and terminology

Unit vectors. We will assume that illumination is by a distant point source, and that projection into the image is orthographic. The direction of illumination will be given by a unit vector, **L** parallel to that direction. The image tangent angle, as before, will be denoted by α^*. The orientation of the shadowed surface will be given by its unit normal, $\mathbf{N}_s$. In addition, two vectors which can be derived from **L**, $\mathbf{N}_s$, and α^* will be introduced for convenience: **t**, the tangent to the contour generator, and $\mathbf{N}_c$, the normal to the shadow cone at the point of intersection with the shadowed surface. Their derivations will be given later on. The relation among these vectors is illustrated in Fig. 4.

Curvature. The curvature of a surface differs from the curvature of a curve in that surface curvature varies with direction on the surface. The curvature of a surface in a particular direction is given by the *normal curvature:* cutting the surface with a plane defines a curve of intersection. If the cutting plane contains the surface normal at a given point, then the curvature of the curve of intersection through that point is the normal curvature of the surface in the direction of the curve (fig. 5). A useful relation (which is sometimes used to define normal curvature) is

$$\kappa_n = \mathbf{k} \cdot \mathbf{N}_s \tag{4.1}$$

where κ_n is the normal curvature at a point **p**, **k** is the curvature vector[2] at **p** of a curve on the surface, and $\mathbf{N}_s$ is the surface normal. Thus a simple relation holds among the normal curvature of the surface, the angle between the plane of a curve on the surface and the surface normal, and the curvature of the curve.

At any point on any smooth surface, the normal curvature assumes a minimum and a maximum in two orthogonal directions. The directions are called *principal curvatures,* and the directions, *principal directions.* The normal curvature in *any* direction is determined by the principal curvatures and directions. The normal curvature in a direction **t** is given by the following relation (Euler's theorem):

$$\kappa_n = \kappa_1 \cos^2 \alpha + \kappa_2 \sin^2 \alpha \tag{4.2}$$

where κ_1 and κ_2 are the principal curvatures, and α is the angle between **t** and the principal direction corresponding to κ_1.

We will denote the normal curvature to the shadowed surface in the direction of **t** by κ_n, and the normal curvature on the shadow cone in the same direction by κ_c'. The principal curvatures of the shadow cone fall in directions parallel to and orthogonal to the cone's axis, **L**. The first of these is zero, and the second will be denoted by κ_c.

In these terms, the quantities that should appear in the final expression are α^* and κ^*, the image tangent and image curvature; **L** and $\mathbf{N}_s$, the light direction and the normal to the shadowed surface; κ_n, the normal curvature to the surface along the contour generator; and κ_c, the curvature of the shadow cone's cross-section. Intermediate quantities, derivable from these, are **t**, the tangent to the

[2]The *curvature vector* is the second derivative of the curve with respect to arc length. Its magnitude is simply called the *curvature.*

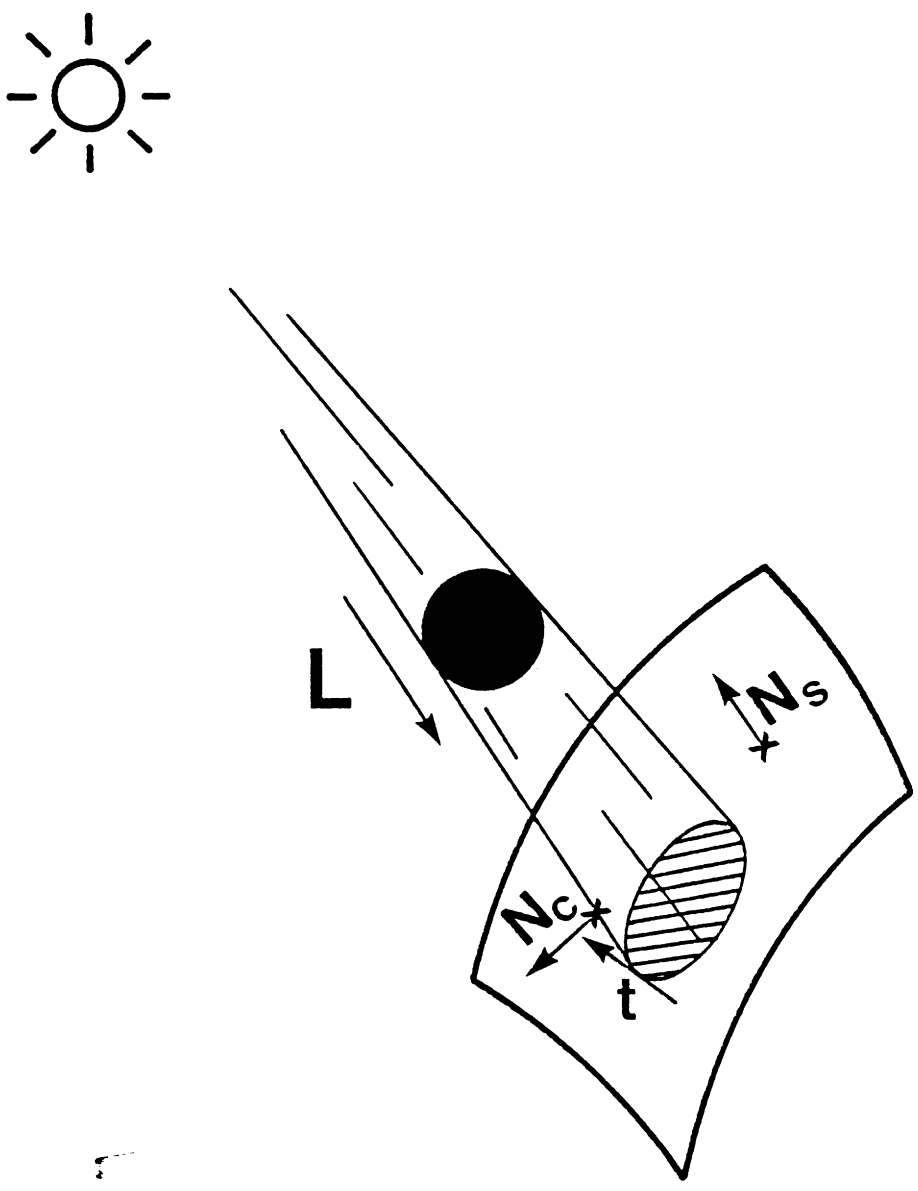

Figure 4. Unit vectors: **L** is the illumination vector, **t** is the tangent to the cast-shadow edge, $\mathbf{N}_s$ is the surface normal, and $\mathbf{N}_c$ is the normal to the shadow cone.

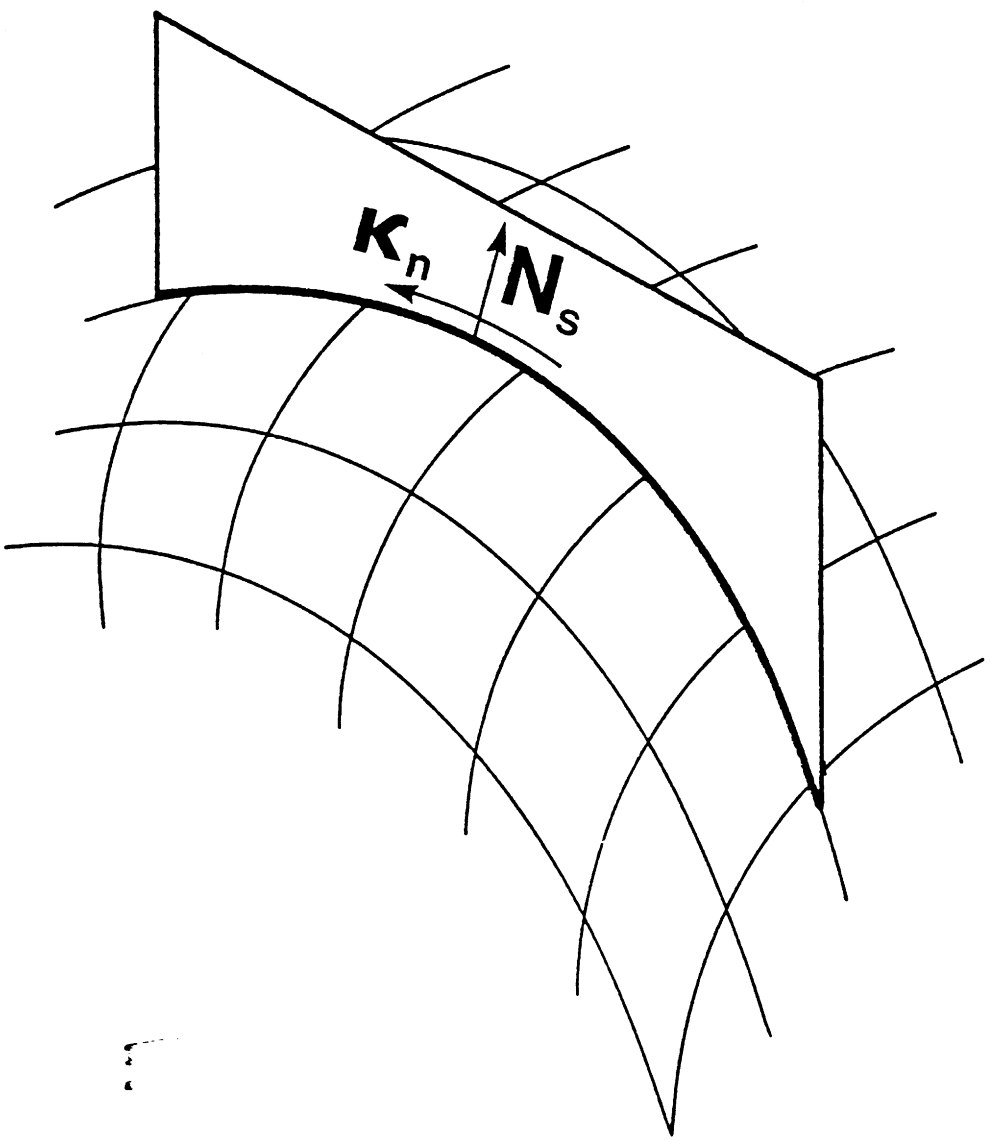

Figure 5. When a surface is cut by a plane containing the surface normal, the *normal curvature* is the curvature of the curve of intersection.

contour generator; $\mathbf{N}_c$, the normal to the shadow cone; and κ'_c, the normal curvature to the shadow cone in the direction of $\mathbf{t}$. These intermediate quantities will be used for convenience in the derivation, then removed by substitution. Also appearing along the way will be $\mathbf{k}$, the curvature vector of the contour generator.

4.2.3 Derivation

Recall that the contour generator is the intersection of the shadow cone with the shadowed surface, with tangent given by $\mathbf{t}$. Then the contour generator lies on both the shadow cone and the shadowed surface. Therefore, from eq. x, its curvature vector $\mathbf{k}$ must satisfy

$$\kappa_n = \mathbf{k} \cdot \mathbf{N}_s$$
$$\kappa'_c = \mathbf{k} \cdot \mathbf{N}_c,$$

(recalling that κ'_c is the normal curvature of the shadow cone in the direction of the contour generator, while κ_c is the principal curvature of the shadow cone in the direction normal to $\mathbf{L}$.) Since $\mathbf{t}$ must be orthogonal to $\mathbf{k}$,

$$0 = \mathbf{t} \cdot \mathbf{k}$$

and we have three equations for $\mathbf{k}$. Solving them gives

$$\mathbf{k} = \frac{\kappa_n(\mathbf{N}_c \times \mathbf{t}) - \kappa'_c(\mathbf{N}_s \times \mathbf{t})}{[\mathbf{N}_s\mathbf{N}_c\mathbf{t}]} \tag{4.3}$$

where $[\mathbf{N}_s\mathbf{N}_c\mathbf{t}]$ is the triple product $(\mathbf{N}_s \times \mathbf{N}_c) \cdot \mathbf{t}$.

Next we substitute κ_c for κ'_c, using equation (4.2), and the fact that the principal curvature of the shadow cone in the direction of $\mathbf{L}$ is zero. The relation then becomes

$$\kappa'_c = \kappa_c \sin^2 \alpha$$

where α is the angle between $\mathbf{L}$ and $\mathbf{t}$. Hence

$$\kappa'_c = \kappa_c |\mathbf{L} \times \mathbf{t}|^2$$

and substituting in (4.3) gives

$$\mathbf{k} = \frac{\kappa_n(\mathbf{N}_c \times \mathbf{t}) - \kappa_c|\mathbf{L} \times \mathbf{t}|^2(\mathbf{N}_s \times \mathbf{t})}{[\mathbf{N}_s\mathbf{N}_c\mathbf{t}]}$$

Next we remove $\mathbf{N}_c$, noting that it is a unit vector orthogonal to both $\mathbf{L}$ and $\mathbf{t}$, so that

$$\mathbf{N}_c = \frac{(\mathbf{L} \times \mathbf{t})}{|\mathbf{L} \times \mathbf{t}|}$$

then

$$\mathbf{k} = \left(\kappa_n \frac{(\mathbf{L}\times\mathbf{t})\times\mathbf{t}}{|\mathbf{L}\times\mathbf{t}|} - \kappa_c|\mathbf{L}\times\mathbf{t}|^2(\mathbf{N}_s\times\mathbf{t})\right)\frac{|\mathbf{L}\times\mathbf{t}|}{\mathbf{N}_s\cdot((\mathbf{L}\times\mathbf{t})\times\mathbf{t})}$$

$$= \left(\frac{-\kappa_n\mathbf{t}\times(\mathbf{L}\times\mathbf{t}) - \kappa_c|\mathbf{L}\times\mathbf{t}|^3(\mathbf{N}_s\times\mathbf{t})}{-\mathbf{N}_s\cdot((\mathbf{t}\cdot\mathbf{t})\mathbf{L} - (\mathbf{L}\cdot\mathbf{t})\mathbf{t})}\right)$$

$$= \frac{+\kappa_n\mathbf{L} - \kappa_n(\mathbf{L}\cdot\mathbf{t})\mathbf{t} + \kappa_c|\mathbf{L}\times\mathbf{t}|^3(\mathbf{N}_s\times\mathbf{t})}{\mathbf{N}_s\cdot\mathbf{L}}. \tag{4.4}$$

This expression gives the curvature vector of the shadow edge. It remains to obtain the projected curvature in the image. To simplify this derivation, we will take the "*" symbol as a projection *operator*, when applied to a vector quantity. For orthographic projection onto a plane, the * operation may be defined by

$$\mathbf{X}^* = \mathbf{X} - (\mathbf{X}\cdot\mathbf{V})\mathbf{V},$$

where $\mathbf{X}$ is any vector, and $\mathbf{V}$ is a unit normal to the image plane. It follows from this definition that

$$\frac{d\mathbf{X}^*}{du} = \left(\frac{d\mathbf{X}}{du}\right)^*,$$

and

$$(c\mathbf{X})^* = c(\mathbf{X}^*),$$

where c is a constant.

Next, we derive the projected curvature κ^*, for an arbitrary curve in space, given the tangent and curvature vectors on the curve. Given a curve $\mathbf{X}(s)$, where s is a natural parameter,[3] the curvature of $\mathbf{X}$ is defined by

$$\kappa = \left|\frac{d^2\mathbf{X}}{ds^2}\right|.$$

The projection of $\mathbf{X}(s)$ is $\mathbf{X}^*(s)$. Since s is not in general a natural parameter for $\mathbf{X}^*$, we must introduce a new parameter, s^*, which is a natural parameter for $\mathbf{X}^*$, noting that

$$\frac{ds}{ds^*} = \frac{1}{|\mathbf{t}^*|}.$$

In terms of s^*, the projected curvature, κ^*, is given by

$$\kappa^* = \left|\frac{d^2\mathbf{X}^*}{ds^{*2}}\right| = \left|\frac{d}{ds^*}\left(\frac{d\mathbf{X}^*}{ds^*}\right)\right| = \left|\frac{d}{ds}\left(\frac{d\mathbf{X}^*}{ds}\frac{ds}{ds^*}\right)\left(\frac{ds}{ds^*}\right)\right|$$

[3] A *natural parameter* of a curve is by definition proportional to arc length on the curve.

$$= \frac{1}{|\mathbf{t}^*|}\left|\frac{d}{ds}\left(\frac{\mathbf{t}^*}{|\mathbf{t}^*|}\right)\right|$$

$$= \frac{1}{|\mathbf{t}^*|^3}\left||\mathbf{t}^*|\,\mathbf{k}^* - \mathbf{t}^*\frac{d|\mathbf{t}^*|}{ds}\right|.$$

Noting that

$$\frac{d|\mathbf{t}^*|}{ds} = \frac{\mathbf{t}^* \cdot \mathbf{k}^*}{|\mathbf{t}^*|},$$

we find that

$$\kappa^* = \frac{1}{|\mathbf{t}^*|^3}\left||\mathbf{t}^*|\,\mathbf{k}^* - \frac{(\mathbf{t}^* \cdot \mathbf{k}^*)\mathbf{t}^*}{|\mathbf{t}^*|}\right|$$

$$= \frac{1}{|\mathbf{t}^*|^3}\left(|\mathbf{t}^*|^2|\mathbf{k}^*|^2 - (\mathbf{t}^* \cdot \mathbf{k}^*)^2\right)^{\frac{1}{2}}$$

$$= \frac{1}{|\mathbf{t}^*|^4}|\mathbf{t}^* \times (\mathbf{k}^* \times \mathbf{t}^*)| \tag{4.5}$$

Equation (4.5) gives the projected curvature of an arbitrary curve, in terms of the curvature and tangent vectors. Combining this result with (4.4) we have

$$\mathbf{k}^* = \frac{\kappa_n(\mathbf{L}^* - (\mathbf{t} \cdot \mathbf{L})\mathbf{t}^*) + \kappa_c|\mathbf{L} \times \mathbf{t}|^3(\mathbf{N}_s \times \mathbf{t})^*}{\mathbf{N}_s \cdot \mathbf{L}}$$

and

$$\mathbf{k}^* \times \mathbf{t}^* = \frac{\kappa_n(\mathbf{L}^* \times \mathbf{t}^*) + \kappa_c|\mathbf{L} \times \mathbf{t}|^3(\mathbf{N}_s \times \mathbf{t})^* \times \mathbf{t}^*}{\mathbf{N}_s \cdot \mathbf{L}}.$$

Therefore

$$\kappa^* = \frac{|\mathbf{t}^* \times (\mathbf{k}^* \times \mathbf{t}^*)|}{|\mathbf{t}^*|^4}$$

$$= \frac{|\kappa_n\mathbf{t}^* \times (\mathbf{L}^* \times \mathbf{t}^*) + \kappa_c|\mathbf{L} \times \mathbf{t}|^3\mathbf{t}^* \times ((\mathbf{N}_s \times \mathbf{t})^* \times \mathbf{t}^*)|}{|\mathbf{t}^*|^4\mathbf{N}_s \cdot \mathbf{L}} \tag{4.6}$$

which gives the curvature of the image contour in the desired terms.

An expansion of this expression substituting slant/tilt or gradient space representations for the unit vectors is unwieldy, but the function may be computed given either of those representations by first computing the corresponding unit vectors, then substituting into (4.6). For example, given gradient representations (p, q) for $\mathbf{N}_s$ and (p_s, q_s) for $\mathbf{L}$, we have

$$\mathbf{N}_s = \frac{[p, q, -1]}{\sqrt{p^2 + q^2 + 1}}$$

$$\mathbf{L} = \frac{[p_s, q_s, -1]}{\sqrt{(p_s^2 + q_s^2 + 1)}} \tag{4.7}$$

$$\mathbf{t} = \frac{[\cos\alpha^*, \sin\alpha^*, p\cos\alpha^* + q\sin\alpha^*]}{\sqrt{1 + (p\cos\alpha^* + q\sin\alpha^*)^2}}$$

and corresponding expressions are readily obtained given slant/tilt representations.

4.2.4 What it means

Having gone through this somewhat involved derivation, we review the meaning of the resulting expressions, and our reason for wanting to derive them. For a particular contour generating process, namely cast shadows, the expressions tell us how the measures we can take on the image contour *got there*, in terms of the various scene parameters that determine them. The curvature of the contour depends on the image tangent direction, α^*, which we can measure; the orientation and curvature of the shadowed surface, $\mathbf{N}_s$ and κ_n, which we would like to recover; the direction of illumination and the curvature of the shadow cone, which are not of direct interest.

In effect, the relation tells us that the curvature of the image contour has three components: one from the curvature of the shadowed surface, one from its orientation, and one from the properties of the shadow cone. Two of these components are interesting from the standpoint of describing the shadowed surface. In the next section we couple this geometric model with a statistical one, to determine a very simple "goodness of fit" statistic, i.e. a measure of an interpretation's ability to explain the image data.

4.3 Statistical model

In this section, a simple means of evaluating "goodness of fit" between a surface and the cast-shadow contour in an image is developed. Although it is only approximately valid, the method is computationally simple, and avoids making assumptions about the distribution of curvature; it will be shown to be sufficient for the registration problem.

4.3.1 Independence

The shadow-contour generating process includes the illuminant and shadowing surface (which together comprise the shadow cone,) the shadowed surface, and the viewer. The shape of the shadow contour depends on these entities and the geometric relations among them. In a realistic situation, the illuminant might be the sun, the shadowing object the branch of a tree, and the shadowed surface a rock on the ground. Unless the branch and the rock have been selected and positioned by a psychologist with the intention of generating unusual shadows, we can be quite confident that the shape and orientation of the branch bear no systematic causal relation to the shape and orientation of the rock, and neither is related to the direction of illumination or our direction of view. That is, knowing about any one of these components is of absolutely no value in predicting the properties of the others. And so, by definition, they are statistically independent.

This elementary observation is the most important one we can make, if we want to use the geometry of shadow generation to draw inferences about surfaces. In fact, it can in some instances, be used directly to decide that an interpretation is unreasonable: if you were shown a picture containing a straight-line edge, and were asked to believe that this was the projection of a shadow-edge

cast on a jagged surface, you would probably be skeptical. Your skepticism would have no basis in the geometry of shadow generation, because the geometric possibility exists that a jagged shadowed surface and a jagged shadowing surface assumed perfect complementary configurations to produce a straight shadow edge. Rather, the hypothesized jagged surface doesn't fit the evidence of a straight shadow edge because we know that such perfect complementarity is unlikely to occur by accident, and we know that the processes that align the components of the shadow generating process are accidental. Roughly speaking, the jagged-surface hypothesis would force us to the conclusion that a perfect negative correlation exists in a sample drawn from populations we know to be uncorrelated. While such an event is possible, it is hardly likely. So, knowing something of the geometry of the shadow generating process, an elementary observation about the statistical relation among some of the components of that process provides a basis for rejecting some hypotheses as implausible.

4.3.2 Evaluating likelihood

Given a prior density function for κ_c, the curvature of the shadow cone, a likelihood measure can be derived by exactly the same reasoning that led to the planar estimator. That is, given the surface orientation and normal curvature, and the direction of illumination, a density function can be computed for the image measures, κ^* and α^*. But we could not specify this function without assuming a specific distribution for κ_c. Moreover, the function would have to include the derivative $d\kappa_c/d\kappa^*$ obtained from the geometric model, and this is a most unwieldy expression.

For the comparatively simple registration problem, an approximate measure of likelihood can be obtained by correlation techniques. We can think of the unknown surface as a function that transforms shadow cones into image contours, in accordance with the geometric model. Given the image contours, and a set of candidate surfaces to choose from, the problem is to decide which of the set of transformations, corresponding to the set of surfaces, has operated on the unknown shadow cone to generate the given image contours.

A "map" of the transformation performed by a given surface can be generated by running a fixed value of κ_c through that surface, at all tangent directions and positions, using the geometric model to compute the resulting curvatures in the image. Since the starting values of κ_c were all the same, all the variations in the computed κ^* entirely reflect the distortion imposed by the curvature and orientation of the surface. Roughly, this map tells us how we would expect curvature in the image to vary with position and tangent direction, if the surface that generated it were present. This is only roughly so, because the geometric relation between κ_c and κ^* is non-linear. But, if we are willing to ignore the non-linearity, we can evaluate the goodness-of-fit by simply correlating the observed values of κ^* with the values drawn from the map at the same position and tangent direction. A high positive correlation indicates that the observed curvature is systematically varying with the curvature predicted for the hypothesized surface. And to choose a most likely surface from a given family, we simply choose the one that gives the highest correlation with the data.

While this method ignores the substantial nonlinearities in the geometric model, I will argue that it is appropriate for the registration problem because it is far simpler than an exact estimator, and because it *works*.

4.4 Registration with a surface model: a demonstration

A problem of practical interest is the registration of an image with a surface model, when the viewpoint of the image with respect to the model is known only approximately. Horn and Bachman (1977) have treated the registration problem by using the surface model to synthesize an image, then establishing registration between the synthesized and real images by maximizing the cross-correlation of the synthetic image with the real one. The principle advantage of this technique over matching with a real reference image is that changes in direction of illumination can be taken into account. They applied the technique to the registration of LANDSAT photos with digital terrain models.

Since the correlation is performed in the image domain, a limitation of the synthetic image approach is that anything which appears in the image, but not in the terrain model, enters into the correlation as noise. In this section, it will be shown that registration can be established using nothing *but* such "noise:" suppose that the image to be registered includes shadows cast by unknown objects outside the field of view.[4] The variations in image intensity introduced by the shadow cannot be predicted from the surface model, hence would not appear in a synthesized image. Therefore the shadows would only hinder a cross-correlation in the image domain, entering into the correlation as pure noise. If we extracted the edges of the shadowed regions, and threw the rest of the image away, we would be left with nothing *but* noise, in the image domain. But our geometric and statistical understanding of the shadow generating process gives the shapes of the shadow edges enough meaning that registration can be established using those edges and nothing else.

4.4.1 Method

In a simple instance of the registration problem, the direction of illumination is known in advance, and the unknown component in the viewpoint of the image is expressed by translation in a plane. That is, we know, for example, that we are looking straight down at the surface, but we don't know exactly where we are. The viewpoint of the image with respect to the surface model is known to lie in a specified region, and, for simplicity, all viewpoints within the region can be assumed equally likely *a priori*. This is the same problem to which Horn and Bachman applied synthetic image techniques.

In other words, we are free to slide the surface model with respect to the image by some specified amount in any direction, and we have to find the position of the surface model which corresponds to the actual position of the surface in the image. Each position $(\Delta x, \Delta y)$ relative to some reference point may be viewed as a hypothesis $H_{\Delta x, \Delta y}$, and we have a two-parameter family of candidate hypotheses, each equally likely, and constrained to a specified region of the (x, y) plane.

The surface model is a function $z(x, y)$ which gives the elevation of the modeled surface with respect to the image plane, at regular intervals of x and y. A hypothesis $H_{\Delta x, \Delta y}$ asserts that the actual surface is given by $z(x + \Delta x, y + \Delta y)$, and the problem is to find a density function for $H_{\Delta x, \Delta y}$ given the image data. The data are a set of measures of (κ^*, α^*) taken along the image contours.

[4]This supposition is clearly unrealistic if the image to be registered is a LANDSAT photo, but might be realistic for other domains in which the registration problem arises. For example, model-based recognition in industrial assembly applications can be hindered by unpredictable cast shadows. In any event, the point of treating the problem is not to arrive at instant applications to practical problems, but to focus on the information on surface shape implicit in a shadow contour.

For each image measure, each hypothesis specifies corresponding values for $\mathbf{N}_s$ and κ_n, measured by finite difference from the surface model. Given these values, and for any image position (x, y), and image tangent direction, α^*, the "expected" image curvature κ^* is determined, for a fixed value of κ_c, by (4.6). Rather than using (4.6) directly, the "expected" curvature was computed numerically by projecting a circle of radius $1/\kappa_c$ onto the DTM in the direction of $\mathbf{L}$, and projecting again into the image. The image curvature obtained for a given (x, y, α^*) is the expected curvature.

Since the contours are comparatively sparse, it is more efficient to compute these values "on line," only where they're needed, rather than in a pre-computed lookup table. Given a hypothesis $H(\Delta x, \Delta y)$, and a set of image measures of the form $(\kappa^*, \alpha^*, x, y)$, the "expected' curvature κ^*_{est} is computed for each (α^*, x, y), and the linear correlation coefficient is computed for the pairs of $(\kappa^*, \kappa^*_{est})$. The value of this coefficient is taken as the approximate relative likelihood of $H(\Delta x, \Delta y)$. By computing the coefficient at increments in the allowable region of the (x, y) plane, the best-fit value for $(\Delta x, \Delta y)$ is obtained. Since we are translating the map of κ^*_{est}, and correlating those values with the observed κ^*, this is a cross-correlation.

4.4.2 Stimuli

The surface models used were digital terrain models (DTM's) like those used by Horn and Bachman. The shadow contours were synthesized by projecting random shapes onto the DTM, then onto an image plane. The shapes represent the sillouhettes of shadowing objects; their projections onto the DTM, the edges of cast shadows; and the projections into the image, cast shadow contours. The use of synthesized stimuli was necessary on practical grounds, but is not a drawback because the modeled surfaces are natural surfaces, and because the shapes of the "shadowing objects" were unknown to the estimation strategy. A synthetic shaded images of a DTM, with a synthetic shadow, is shown in Fig. 6.

4.4.3 Results

Figure 7 shows side by side several "shadowing object" sillouhettes, and the image contours that were generated by projecting them onto the DTM, and again onto the image. The *difference* between each sillouhette and the corresponding contour represents the distortion imposed by the shape of the DTM, and by projection. Recall that registration is established not by comparing the contour to the sillouhette, which is unavailable to the registration algorithm, but to the DTM. While the distortion is in some cases not very great, it suffices to establish registration accurately.

The registration algorithm computes the cross-correlation, as a function of $(\Delta x, \Delta y)$, of observed contour curvature with predicted curvature. The peak value of the correlation gives the estimated offset of the image with respect to the DTM. Figure 8 shows contour plots of the cross-correlations obtained from the contours shown in the previous figure. In each case, the estimated offset differs from the correct value by less than five DTM pixels. Note that the exact peak of the cross-correlation is in each case surrounded by an elongated "ridge." These in fact coincide with ridges in the terrain model. Where a shadow bends across a ridge, sliding the shadow along the ridge maintains the goodness of fit much more than sliding it across the ridge.

Figure 6. A picture synthesized from a DTM, including a "fake" shadow. The edge of such a shadow is the input to the registration algorithm.

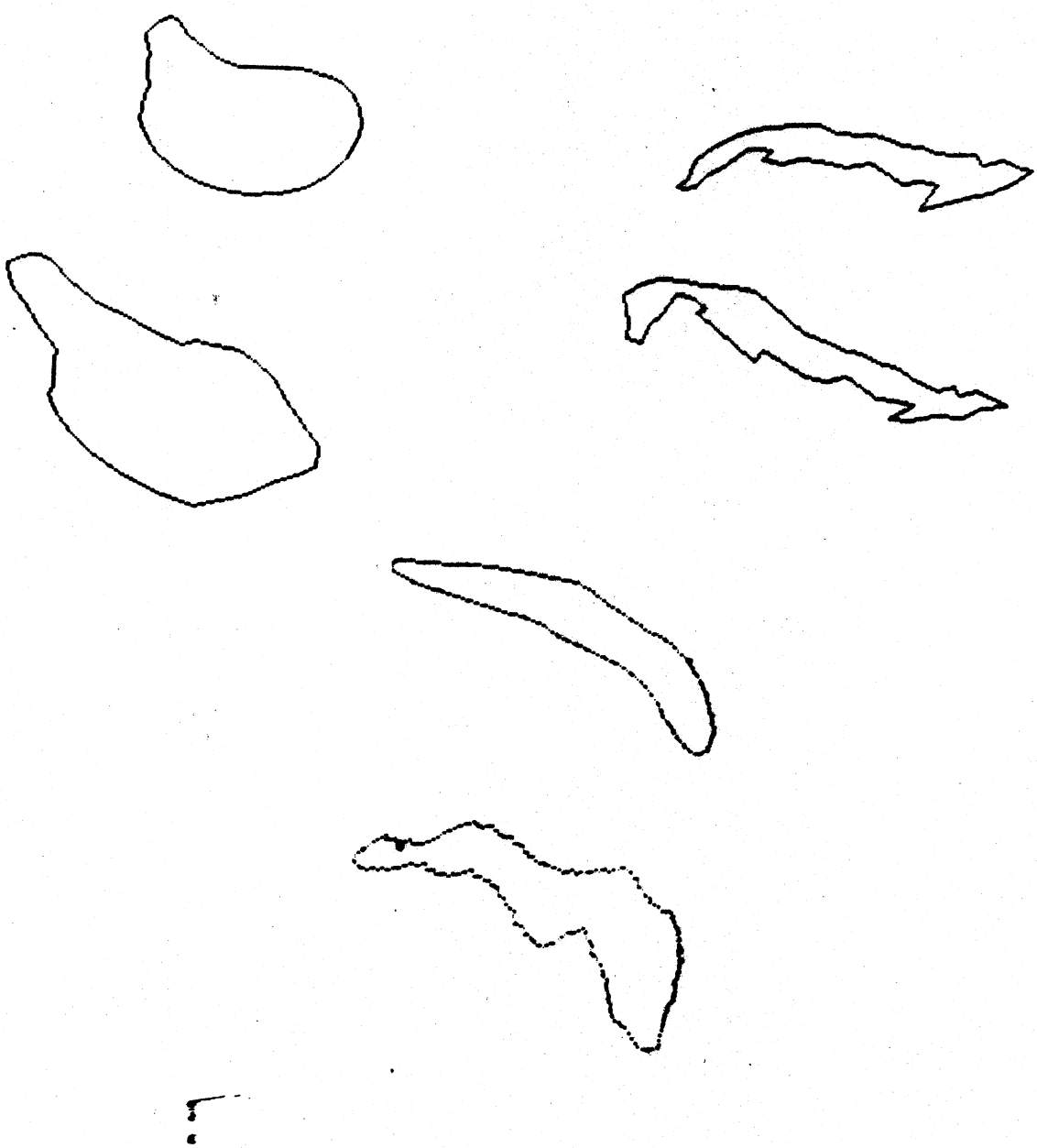

Figure 7. "Shadowing object" silhouettes, and the image contours that were generated by projecting them onto a DTM, and again onto the image. The difference between the silhouettes and the contours reflects the distortion imposed by the shape of the DTM and by projection. This distortion is the registration algorithm's "signal."

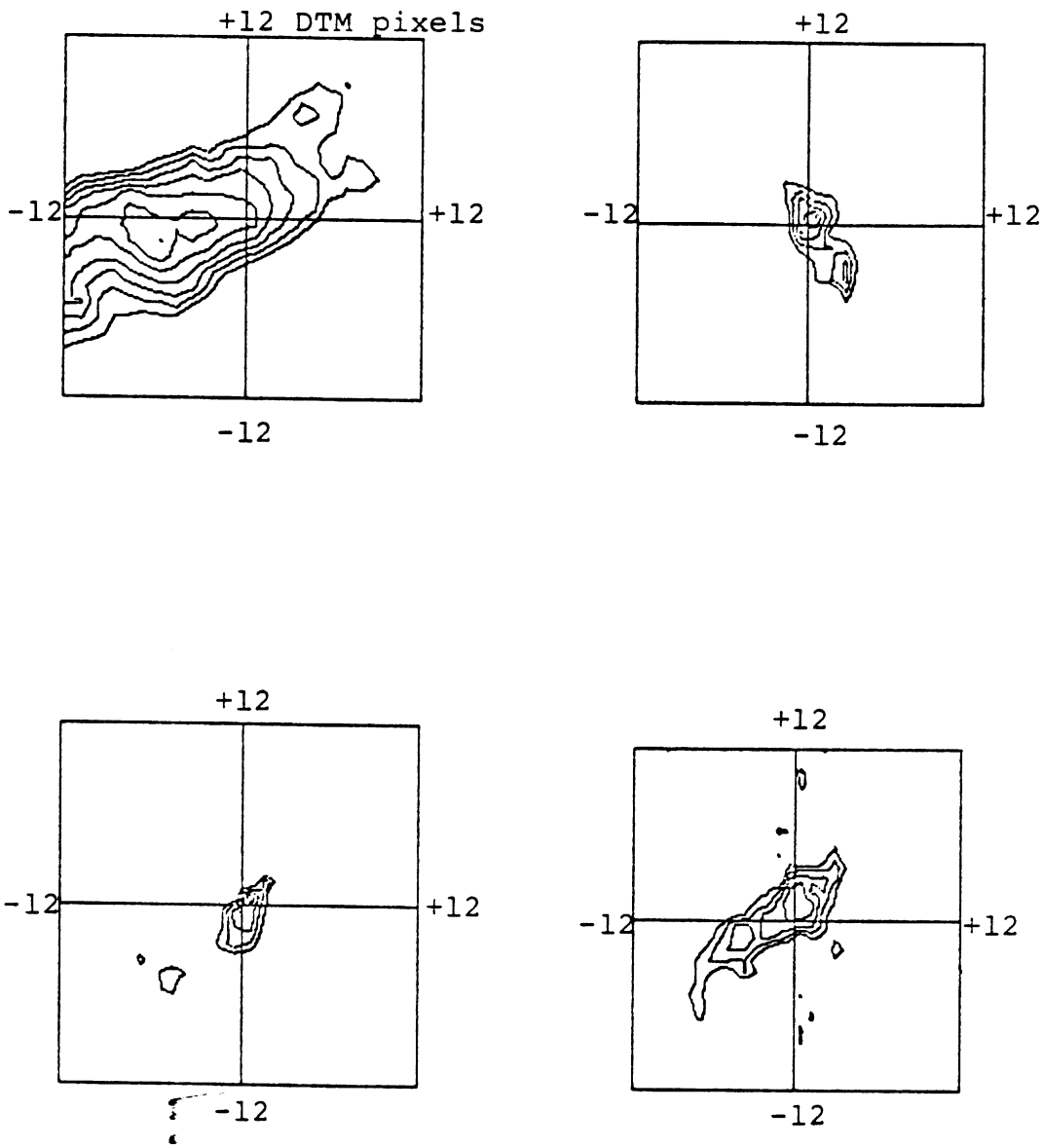

Figure 8. Contour plots of the cross-correlations. The correct offset is positioned at the center of the plot, so the line joining the center to the peak value of the cross-correlation shows the error of the estimated offset. The coordinates are given in DTM pixels.

4.5 Conclusions

It was shown that a detailed understanding of the local geometry by which cast-shadow contours are formed permits the curvature of a shadow contour to be decomposed into independent components, among which are surface curvature and orientation; and, in the restricted registration problem, that these components can be isolated statistically. The generality and usefulness of these results might be extended in two directions: first, by extension to less restricted problems, and second, by extension to more general contour types such as surface markings. In this section, the flavor of these extensions will be given.

4.5.1 Treating the general estimation problem

In extending the planar estimator to the estimation of curved surfaces, results from a restricted situation with strong prior constraints were extended to a less constrained, more general problem domain. This was done by applying the planar estimator locally, to a small region around each point, and estimating the overall orientation in the region. The effect was to discard changes in orientation that occurred on a scale that was small compared to the region size, obtaining a reliable estimate at the expense of resolution.

It is natural to try the same approach toward extending the curvature-based method, that is, examining a region around a point to characterize the surface at that point. But that local description, since it now includes the curvature of the surface, would have to become much more complicated. In fact, the description of a surface at a point, up to the second derivative, has five degrees of freedom: two for the orientation of the tangent plane, one for the orientation of the principal directions, and two for the principal curvatures. To construct a local strategy closely analogous to the extension of the planar method, we would need a way of evaluating the joint likelihoods of these parameters, given the data in a small region. This might be done by taking the point description of the surface as a second-order approximation to a patch around the point, and varying that description to optimize a goodness-of-fit measure on the surrounding data. The goodness-of-fit measure already applied to the registration problem might be adequate for this purpose.

A related, but more structured approach is the representation of the surface by a surface-patch function, that is, a patch-wise approximation representing each patch by a simple function, with constraints on continuity where the patches join. Such representations have been used extensively in computer graphics applications (see, e.g., Newman & Sproul, 1979). Since the whole surface is then specified by a vector of the parameters governing the patches, a best-fit surface can be found by hill climbing.

4.5.2 Extension to more general contour types

Cast shadows were selected for initial investigation because they are generated by a process that is geometrically uniform, and falls naturally into clearly independent components. That is, we can say very exactly that the curvature of the contour has a component from the curvature of the shadow cone, and a component from the curvature of the shadowed surface, and that the combination is the

curve along which those two surfaces intersect. We can also confidently assert that the shadow cone and the shadowed surface are almost always independent.

Intuitively, surface markings, such as ground cover or pigmentation, bend and twist with the surface they lie on, in much the same way cast shadows do. But the processes by which such markings are formed are varied and irregular; there is no exact geometric decomposition that can describe them. The only firm geometric constraint on the relation between surfaces and the markings on them is that the marking can never be *less* curved than the surface.

To use the curvature of contours arising from surface markings, a way must be found to partition them into a component from the curvature of the surface, and a component from the process that marked the surface. This decomposition must be realistic enough that the components it specifies are liable to be at least roughly independent. Otherwise, there is no hope of statistically isolating the components of interest.

One decomposition, though idealized, may be a good enough approximation to be of use: the decomposition of the curvature of the marking into a *geodesic* component and a *normal* component. Intuitively, a geodesic is the path you follow when you try to move in a straight line on a curved surface, for example driving a tractor on hilly terrain without turning the wheel. The *geodesic curvature*, again intuitively, is the deviation from that path. That is, if you turn the tractor's steering wheel a fixed amount, and keep it there, the path you follow will have constant geodesic curvature. The actual curvature of the tractor's path in space is nothing more than the vector sum of the geodesic curvature and the normal curvature of the surface. If you periodically decide to turn your tractor's steering wheel, and your decisions have nothing to do with the surface curvature, then the geodesic and normal curvatures along your trajectory will be independent.

The reason this decomposition seems reasonable is that many processes that mark surfaces, particularly processes of growth and propagation, act in very much this way. For example, a lichen grows at roughly a uniform rate on the surface, without regard to the curvature of the surface. So the result is usually a "circle" that bends with the surface. The same is roughly true of the growth of rust spots on your car, or ink spots on absorbent cloth, weeds on a lawn, or mould on a piece of bread.

Such a decomposition is also a good approximation for cast shadows, when the angle between the surface normal and the illumination vector is not too large. In fact, when the light is orthogonal to the surface, the curvature of the shadow cone becomes exactly the geodesic curvature of the shadow edge. Since the curvature of a curve on a surface is a vector sum of the normal and geodesic curvatures, this decomposition has the further advantage of simplicity.

The point of describing the geometry of contour formation, or any aspect of image formation, is to achieve a good enough model of the generating process to untangle the image. The important question about this, or any other, decomposition is how well it will work when applied to images of natural scenes. And that empirical question is not yet answered.

REFERENCES

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

Bajcsy, R., "Computer identification of textured visual scenes," *Memo AIM-180, Stanford University AI Lab* (1972).

Barrow, H.G. & Tenenbaum, J.M., "Recovering intrinsic characteristics from images," *S.R.I. Technical Report* (1978).

Brunswik, E., *Perception and the representative design of pyschological experiments.* University of California Press, Berkeley and Los Angeles, 1947.

Clowes, M.B., "On seeing things," *Artificial Intelligence* 2, 1 (1971), 79–112.

Descartes, R., *La Dioptrique,* Bobbs Merrill, New York, 1965.

Falk, G., "Interpretation of imperfect line data as a three-dimensional scene," *Artificial Intelligence* **4**, 2 (1972), 101–144.

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

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

Helstrom, C.W., *Statistical theory of signal detection,* Pergamon, Oxford, 1968.

Horn, B.K.P & Bachman, B.L., "Using synthetic images to register real images with surface models," *MIT AI Memo 437* (1977).

Horn, B.K.P., "Understanding image intensities," *Artificial Intelligence* **21**, 11 (1977), 201–231.

Horn, B.K.P., "Vision review," *MIT AI Working Paper 157* (1978).

Kender, S., *Ph.D Thesis, forthcoming,* 1979.

Land, E.H. & McCann, J.J, "Lightness and retinex theory," *J. Optical Society of America* **61**, 1 (1971), 1–11.

Mandelbrot, B.B., *Fractals—form, chance, and dimension,* W.H. Freeman and Co., San Fransisco, 1977.

Marr, D.C. & Hildreth, E., "Theory of edge detection," *MIT AI Memo 518* (1979).

Marr, D.C. & Poggio, T., "A computational theory of human stereo vision," *Proc. Roy. Soc. Lond.* **204** (1979), 301–328.

Marr, D.C., "Analysis of occluding contour," *Proc. Roy. Soc. Lond.* **197** (1977), 441-475.

Marr, D.C., "Artifical intelligence—a personal view," *Artificial Intelligence* **9** (1977), 37–48.

Marr, D.C., "Early processing of visual information," *Phil. Trans. Roy. Soc.* **275** (1976), 483–524.

Newman, W.M & Sproul, R.F., *Principals of Interactive Computer Graphics*, McGraw-Hill, New York, 1979.

Perrin, J., "La discontinuite de la matiere," *Revue du Mois* **1** (1906), 323–344.

Pentland, A.P., *personal communication*, 1979.

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

Richardson, L.F., "The problem of contiguity: an appendix of the statistics of deadly quarrels," *General Systems Yearbook* **6** (1961), 139–187.

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

Stevens, K., "Surface perception from local analysis of texture and contour," *Ph.D. Thesis, Dept. Elec. Engr. & Comp. Sci., MIT* (1979).

Switkes, E., Mayer, M.J. & Sloan, J.A., "Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis," *Vision Research* **18** (1978), 1393–1399.

Thompson, D'arcy W., *On growth and form*, Vol. II, Cambridge University Press, Cambridge, England, 1917.

Torrance, K.E. & Sparrow, E.M., "Theory for off-specular reflection from roughened surfaces," *Journal of the Optical Society of America* **57**, 9 (1967), 1105–1114.

Torrance, K.E., Sparrow, E.M. & Birkebak, R.C., "Polarization, directional distribution, and off-specular peak phenomena in light reflected from roughened surfaces," *Journal of the Optical Society of America* **56**, 7 (1966), 916–925.

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

Ullman, S., "Filling the gaps: The shape of subjective contours and a model for their generation," *Biological Cybernetics* 25 (1976), 1–6.

Winston, P.H. (ed.), *The psychology of computer vision*, McGraw Hill, New York, 1975.

Zucker, S.W., Hummel, R.A., & Rosenfeld, A., "An application of relaxation labeling to line and curve enhancement," *IEEE Transactions on computers* C-26, 4 (1977), 394–403.

APPENDIX A

HUMAN PERCEPTION OF SURFACE ORIENTATION: A COMPARISON

A.1 Introduction

This appendix reports an experimental comparison of the perceived orientations in space of unfamiliar curves, with the orientations assigned to the same curves by the planar estimation method of Chapter 2. That method was derived from consideration of the perceptual problem, without reference to its solution, if any, in biological vision, and cast in terms of assumptions about the world. The aim of the comparison is to evaluate those same assumptions as an abstract description of human performance.

The planar method is a solution to the problem of inferring orientation from contour shape. The human perceiver, to the extent he uses contour shape to infer surface orientation, has solved the same problem, but not necessarily by the same method: it is possible by experiment with natural images to show that a model of the world, like the geometric/statistical model of Chapter 2, is *sufficient* to solve the planar orientation problem, but never to show that the set of assumptions comprising that or any other model is *necessary* for a solution; because those assumptions are empirical assertions about the world, it is always possible that some other, undiscovered properties of the world would also suffice to solve the problem. In short, no claim of uniqueness can be attached to the theories that have been presented, or to any theories of their kind. And therefore, evidence that the methods work is not evidence that they are used by the human perceiver, or by any other system.

This point is clarified by a simple example: binocular disparity, accommodation, and vergence are all well known to be potential "depth cues" in the sense that each can be used to measure distance to the viewer. But that each of these measures could in principle be used to infer depth does not imply that any or all of them are used by the human perceiver. Whether the human visual system uses one, all, or none of these possible methods to infer depth can only be decided by observing that system.

The planar method is effective, and derives from very simple assumptions about the world. The method performs a *mapping* from image contours to surface orientations, and that mapping is concisely specified by the assumptions on which the method is based. The human perceiver, in using contours to judge orientation, performs the same kind of mapping. To the extent the two mappings are isomorphic, the assumptions that underlie the planar method also describe the mapping performed by the human perceiver. Such isomorphism says nothing about the mechanism by which the mapping is

performed (Marr, 1977), but if found, it succinctly describes the perceptual strategy's behavior, and accounts for its accuracy in terms of properties of the visual world.

All generally accurate methods for inferring orientation, even if they differ fundamentally, must by definition give substantially similar results overall: a method is only accurate to the extent that its results tend toward the correct result. Since there is in each case only one correct result, all methods, to be accurate, must share this tendency. That is, if each surface has a correct orientation, then there is a unique *correct* mapping from contours to orientations, and any method is accurate only to the extent it approximates that mapping. Two perfectly accurate methods are both isomorphic to the correct mapping, hence to each other. Rather, the divergence of fundamentally different but generally accurate methods appears only in their *failures*. In other words, different strategies generally lead to different illusions. For this reason, errors of the planar method will be compared to errors of the human strategy.

Human observers readily perceive simple drawn shapes as slanted in space, even though the surface they're drawn on lies in the frontal plane (fig. 1). A familiar example, the appearance of an ellipse as a tilted circle, might be explained on the basis of the circle's greater familiarity, but unfamiliar shapes also appear slanted in seemingly systematic ways. While such shapes often appear bent in space, as well as slanted, a substantial subset may be perceived as planar.

If the curves actually lie in the frontal plane, and if they have not in their construction undergone projective distortions, then they have no "real" orientations outside that plane, and any deviation from the frontal plane in assigned orientation is an error. Such curves are therefore suitable for the comparison.

A.2 Method

A.2.1 Stimuli

The stimuli were nineteen "random" shapes, i.e. shapes defined by a function with pseudorandom parameters. That function was an iterative product of polar sech functions, with pseudorandom translation, uniform scaling, and rotation at each iteration. This procedure ensures simple (i.e. non-self-intersecting) closed curves. The curves were then smoothed by replacing each point with the average of its near neighbors, to avoid discontinuities of orientation. Examples of the curves are shown in fig. 2. In all, twenty curves were used as experimental stimuli.

A.2.2 Orientation judgments

Observers' orientation judgments were measured by matching the orientations of the experimental shapes to that of a probe shape, consisting of the projection of three mutually orthogonal lines. The perceived orientations of these configurations have been shown by Stevens (1979) to be consistent with the orthogonal interpretation. The experimental and probe shapes were shown concurrently on a CRT screen, with the orientation of the probe controlled in real time by joystick operated by the

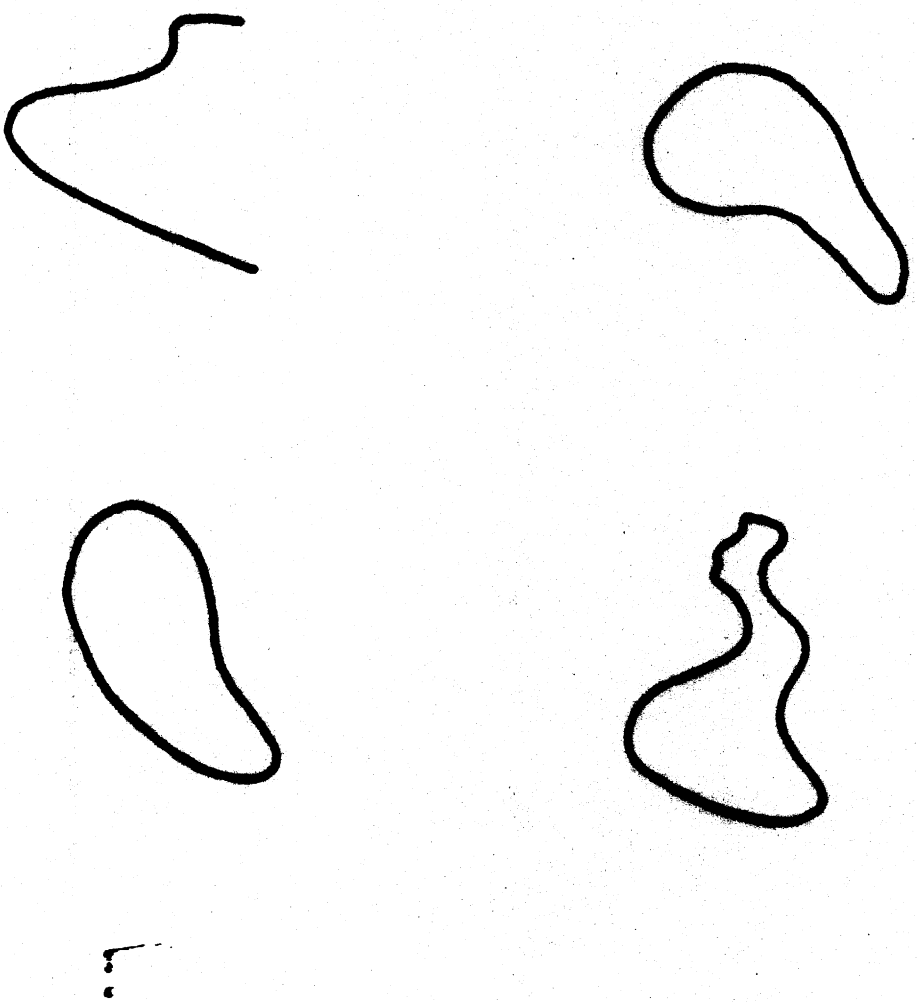

Figure 1. Unfamiliar curves often appear slanted in space

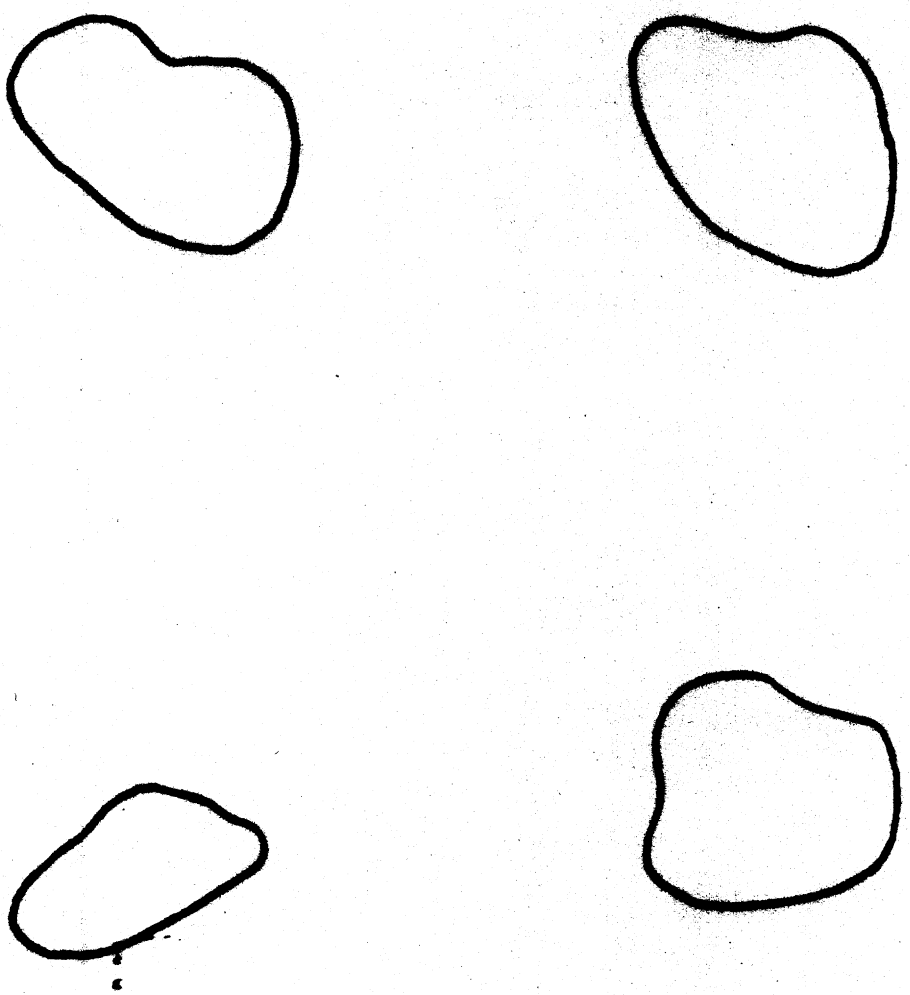

Figure 2. Examples of the "random" curves used as stimuli.

subject. The probe gave the compelling appearance of a figure rotating rigidly in space. (see Fig. 3). Subjects were instructed to adjust the probe until the two crossed lines appeared to lie flat on the surface defined by the experimental shape, with the third line normal to that surface, and orientation was recorded when a match was achieved.

Twenty shapes were shown to ten subjects, in each of eight picture-plane orientations (i.e. rightside up, sideways, upside down, etc.—not slanted in depth), for a total of 1600 trials. The order of presentation was randomized with respect to shape and picture plane orientation.

A.2.3 Results

Maximum likelihood estimates of orientation were first computed for each experimental shape, by the method described in Chapter 2.

The judgments of tilt (τ) were highly consistent across subjects, for each shape, but the judgments of slant were much more variable. In view of the high variance of slant judgments, only τ was compared to the predictions.[1] Figure 4 shows a representative sample of the experimental shapes, each with polar histograms of the tilt data and the tilt vectors obtained by the estimation strategy. Also shown are standard deviations of the data from the predicted values. The data and predictions are clearly in close accord. Figure 5 shows a histogram of the tilt data combined across stimuli, and centered on the predicted values.

A.3 Discussion

These data show that the statistical strategy of Chapter 2 accurately predicts human observers' tilt judgments for a class of unfamiliar shapes whose "real" orientations all lie in the frontal plane. Thus the geometric/statistical model underlying that strategy may be taken as a succinct description of the assignment of orientations to these shapes by human observers. Perhaps more important, the model explains *why* the strategy reflected in that pattern of judgments is an effective one for interpreting natural images, because the model is cast in terms of assumptions about the visual world.

[1]Since orientation was measured by a matching procedure, the variance of slant judgments might be due to variability in the perceived orientation of the probe, or to variability in the orientation matching itself, as well as variability in the perceived slants of the experimental stimuli. Subsequent work (Pentland, 1980) suggests that far lower variance in slant judgments can be achieved using an ellipse-shaped probe.

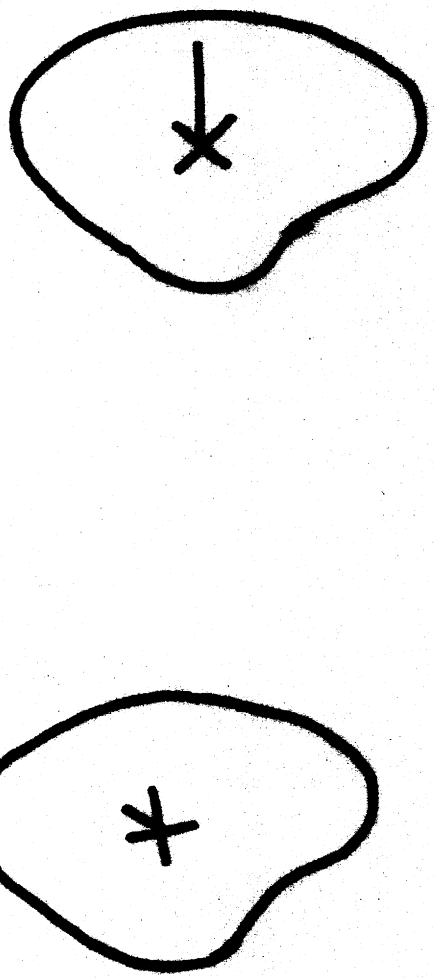

Figure 3. Orientation was measured by concurrent display of the experimental shape and an adjustable probe shape. Above is shown an experimental shape, with the probe in two positions.

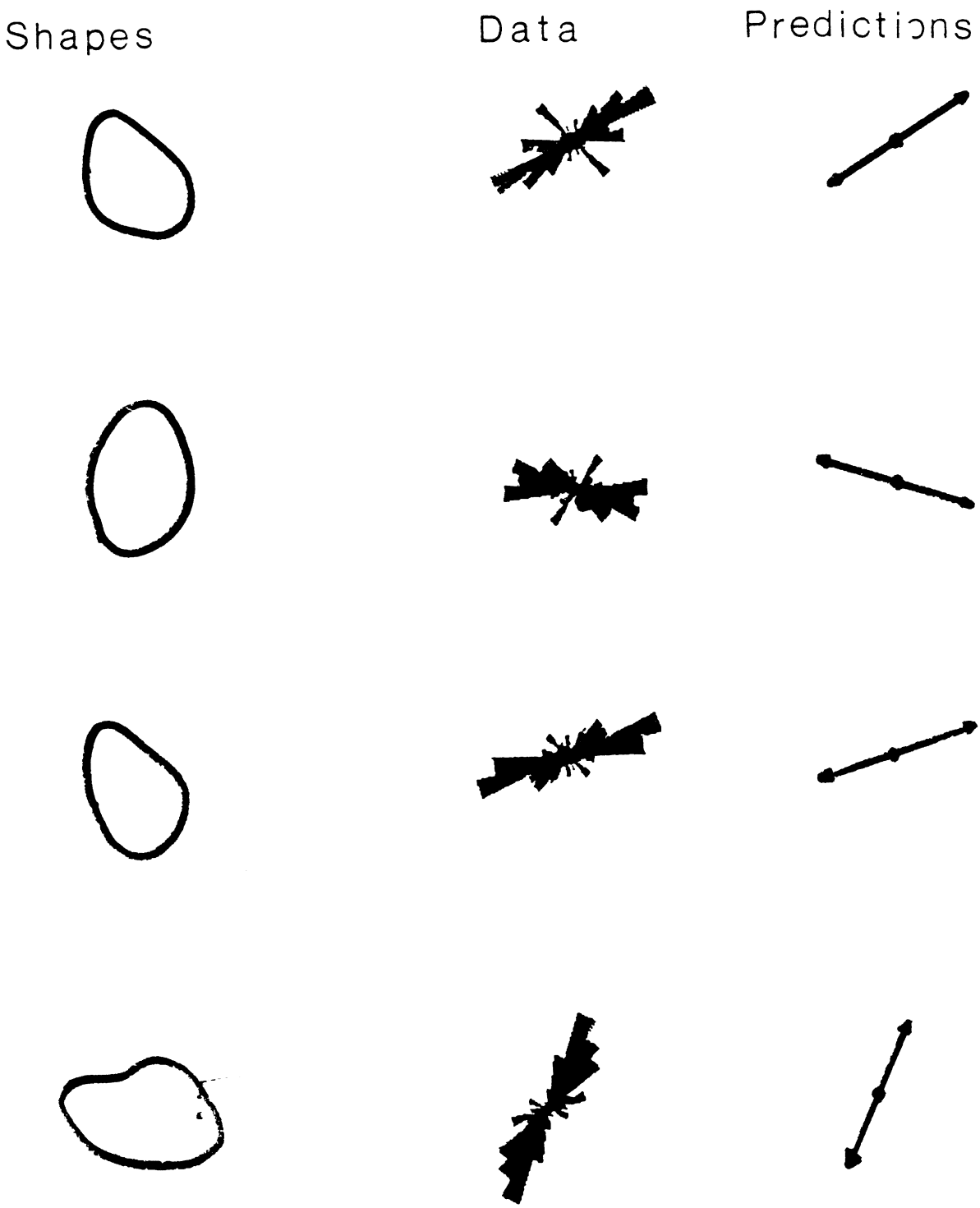

Figure 4. Several of the experimental shapes, with polar histograms of the tilt data, and tilt vectors predicted by the estimation strategy.

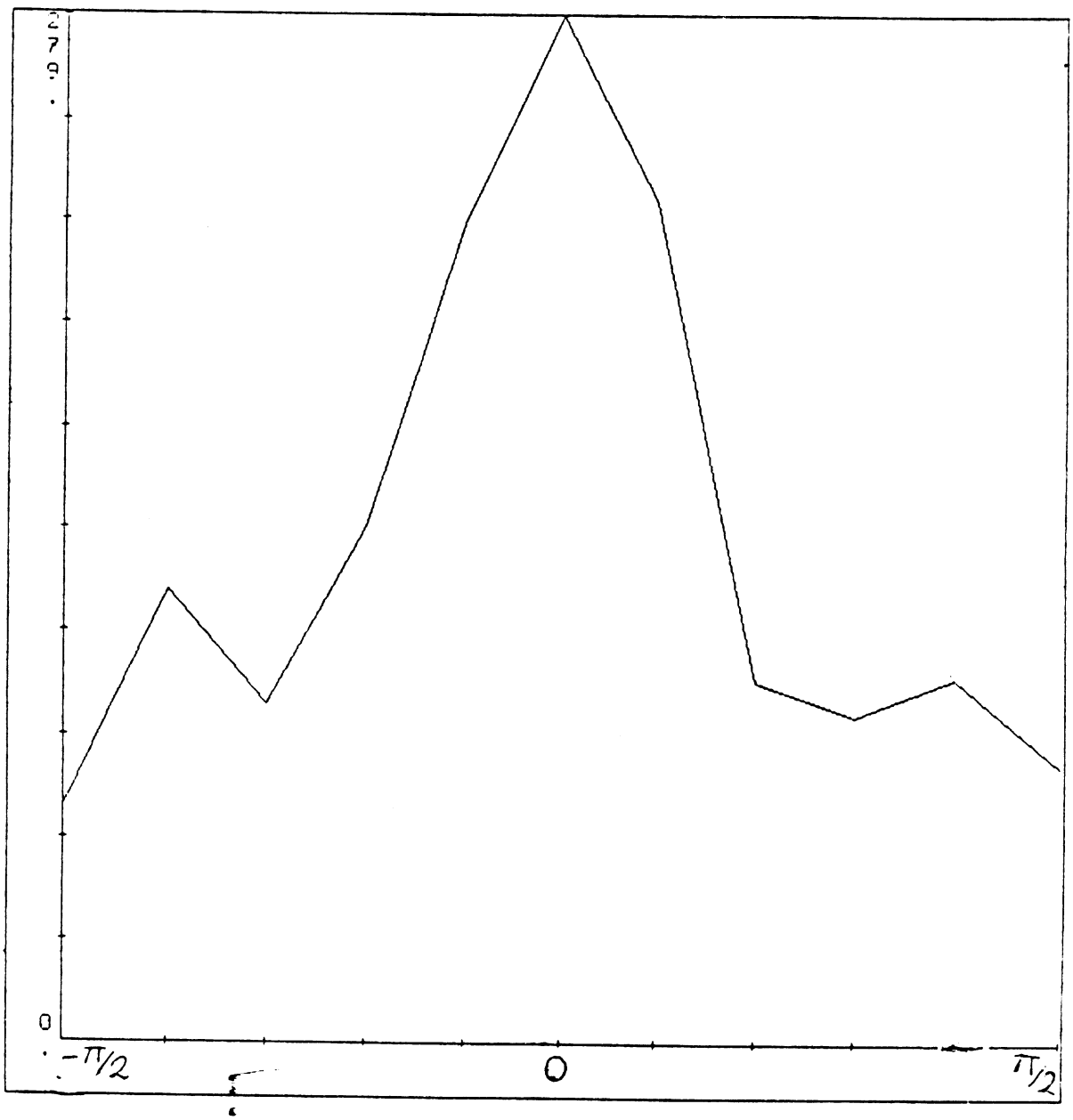

Figure 5. Combined histogram of tilt data. The data from each shape were centered on the predicted values before summing across shapes, so distance from the histogram's center corresponds to deviation from the predicted value.

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 2/29/96

Report # AI-TR-589

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

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☒ Technical Report (TR) ☐ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 99 (107-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☐ Laser Print
☐ InkJet Printer ☐ Unknown ☒ Other:_____

Check each if included with document:

☒ DOD Form (2) ☐ Funding Agent Form ☒ Cover Page
☒ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages (by page number): _____

Photographs/Tonal Material (by page number): _____

Other (note description/page number):

Description : Page Number:

(A) IMAGE MAP: (1-99) UNNUMBERED TITLE PAGE, 2-99 (100-107) SCANCONTROL, COVER, SPINE, DOD (2), TRGT'S (3)

(B) NOTE: PAGE SEVENTEEN IS A XEROX'ED REPLACEMENT PAGE.

Scanning Agent Signoff:

Date Received: 2/29/96 Date Scanned: 3/18/96 Date Returned: 3/21/96

Scanning Agent Signature: Michael W. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AI-TR-589	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Shape from Contour		5. TYPE OF REPORT & PERIOD COVERED Technical Reprot 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Andrew P. Witkin		8. CONTRACT OR GRANT NUMBER*(s)* N00014-75-C-0643 MCS79-23110 5T32GM07484
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 November 1980 13. NUMBER OF PAGES 100
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)*

Shape
Surface orientation
Curvature
Image Registration

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

The problem of using image contours to infer the shapes and orientations of surfaces is treated as a problem of statistical estimation. The basis for solving this problem lies in an understanding of the geometry of contour formation, coupled with simple statistical models of the contour generating process. This approach is first applied to the special case of surfaces known to be planar. The distortion of contour shape imposed by projection is treated as a signal to be estimated, and variations of non-projective origin are treated as noise. The resulting method is then extended to the estimation

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)*

of curved surfaces, and applied successfully to natural images. Next, the geometric treatment is further extended by relating contour curvature to surface curvature, using cast shadows as a model for contour generation. This geometric relation, combined with a statistical model, provides a measure of goodness-to-fit between a surface and an image contour. The goodness-of-fit measure is applied to the problem of establishing registration between an image and a surface model. Finally, the statistical estimation strategy is experimentally compared to human perception of orientation: human observers' judgements of tilt correspond closely to the estimates produced by the planar strategy.

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 3/18/1996

M.I.T. Libraries
Document Services