The Visual Inference Of Shape: Computation From Local Features

by

Alex Paul Pentland
B.G.S. University Of Michigan
(1976)

SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF PHILOSOPHY IN PSYCHOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHONLOGY
June 1982

Signature of Author____ [redacted]
Department Of Psychology, February 24, 1982

Certified By____ [redacted]
Professor Whitman Richards, Thesis Supervisor

Accepted by____ [redacted]
Professor Richard Held, Chairman, Department of Psychology

THE VISUAL INFERENCE OF SHAPE: COMPUTATION FROM LOCAL FEATURES

by

Alex Paul Pentland

Submitted to the Department of Psychology on February 24, 1982
in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy in Psychology

ABSTRACT

A complete understanding of any visual system requires knowing how sufficient information may be extracted from an image to allow recovery of the structure of the corresponding scene. This thesis examines how shading, illuminant direction and contour in an image may be used to infer the shape of the imaged surfaces without *a priori* knowledge about the scene. Following a theoretical analysis, algorithms are developed which:

(1) estimate surface shape using a local analysis of changes in image intensity,

(2) estimate illuminant direction and use it to constrain surface shape,

(3) estimate three-dimensional shape for imaged contours and use them to help determine surface shape.

The performance of these algorithms is evaluated by using them to recover surface shape from both natural and synthesized images.

Many aspects of human perception of surface shape can be explained by this analysis, including the correct perception of shape in natural images as well as certain failures in shape perception. A new illusion is presented that demonstrates some fundamental limitations on the use of shading information to recover shape.

Thesis Supervisor: Dr. Whitman Richards, Professor of Psychology

Biography

Alex Paul Pentland joined the Environmental Research Institute of Michigan (ERIM) in 1973. He worked as a systems analyst for two years, designing and implimenting a PL-1 type language, and a large-scale data base system, in addition to engaging in research on satillite-base remote sensing of the environment. He then was promoted to Research Associate in 1975, and placed in charge of a research task to develop and evaluate image processing techniques for NASA's Large Area Crop Inventory Experiment (LACIE). His contributions to this effort (in the areas of data clustering, image segmentation and the automatic statistical stratification of data), were critical to the development of the first successful image processing system for the world-wide survey of wheat. For efforts in the development of this system his research group recieved NASA's Group Achievement Award in 1976. That same year, he received a Bachelor in General Studies degree from the University of Michigan.

Having found that current image processing techniques were insufficent to deal successfully with many of the tasks people perform so effortlessly, he concluded that only through study of human visual processing could significant gains be made in image processing technology. He therfore left his position at ERIM and entered the M.I.T. Psychology department in September of 1977, joining the Artificial Intelligence Laboratory at M.I.T. at the same time. He has worked as a consultant in remote sensing and data base design to Arthur D. Little during his time at M.I.T., and will join the Artificial Intelligence Center at SRI International following graduation from M.I.T.

SUMMARY

People are able to recover surface shape from monocular images under very general viewing conditions. Without texture or contour information, however, people make systematic errors. The addition of texture or contour information, which by itself is often insufficient to recover surface shape, can allow the correct recovery of the surface.

The computational basis for this human capability has been investigated by examining the physics of image formation and exploring the use of several possible constraints on the interpretion of shape from local image data. This analysis has revealed the following facts about the local computation of surface shape from shading and contour information.

Determination Of Shape From Local Shading:

(1) The image-plane component of surface orientation (tilt) at a point may be exactly determined from an analysis of shading in the neighborhood of the imaged point, under the assumption that the surface is locally a second-order surface. The tilt of the surface may be inferred from the direction in which d^2I, the second derivative of image intensity, is greatest; it is notable that knowledge of the illuminant direction is unnecessary.

(2) The depth component of surface orientation (slant) may be estimated with reasonable accuracy using a maximum likelihood estimator, which by definition is the minimum variance unbiased estimator. The slant of the surface is estimated to be proportional to $|\nabla^2 I/I|^{-\frac{1}{2}}$, where $\nabla^2 I$ is the Laplacian of the image intensity I; again knowledge of the illuminant direction is unnecessary. Examples of surface shape being recovered are shown using both natural and synthesized images.

(3) The effects of foreshortening on image shading, caused by the surface slanting away from the viewer, are identical to the effects of surface curvature on image shading, so that foreshortening may cancel all or part of the effects of surface curvature, and vise versa. Therefore it is impossible to exactly estimate the depth component (slant) of a surface from shading information alone; it is always

confused with surface curvature. This shows that the previous two results give the best estimate of surface orientation that it is possible to obtain from local shading information, given their assumptions.

(4) The type of the surface up to refections about the image plane may be determined from the number of directions along which d^2I is zero at the point under consideration. If d^2I is zero along one direction, the surface is a cylinder; if d^2I is zero along three (and thus all) directions, a plane; and if d^2I is not zero for any image direction then the surface either a convex, concave or saddle surface.

Constraint On Shape From Illuminant Direction:

(5) The illuminant direction may be closely estimated within a region of the image under the assumption that change in surface orientation is isotropically distributed. Examples are shown using natural and synthetic images.

(6) Information about illuminant direction may be used to provide constraint on surface shape: the convexity of the surface may be estimated along each direction (i.e., whether the surface is curving toward, or away, from the viewer), and from this it can be estimated whether the surface is a convex, concave or saddle surface.

Shape From Contour:

(7) A maximum likelihood estimate of the three-dimensional shape of an imaged contour may be made under the assumption that viewer position and contour generator shape are independent. This estimate is that the contour generator is regionally planar.

Shape From Contour And Shading:

(8) Given the existence of shading information in the neighborhood of an imaged contour, the constraint on surface shape obtained from a local analysis of shading is generally sufficient to determine whether the regions on either side of an imaged contour are part of the same surface, or are from

different surfaces.

(9) Given the existence of shading information in the neighborhood of an imaged contour, the constraint from local analysis of the shading is generally sufficient to determine whether or not the contour is a smooth occluding contour, such as is found at the edge of the image of a sphere. Examples are shown using natural images.

(10) Given the existence of shading information in the neighborhood of an imaged contour, then the surface orientation may be exactly determined if the contour generator is planar (see (7) above), and not parallel to the image plane. Examples of contour being used to constrain the interpretation surface shape are shown, using both natural and synthesized images.

Relevance To Human Perception:

The psychological and biological reality of local estimation of surface shape from shading, illuminant direction and imaged contour is discussed. Many aspects of the human perception of surface shape can be accounted for by this analysis of the problem, including the following:

(11) The inability to utilize shading in photographic negatives, even though the shading in electron microscope images (which appear similar to photographic negatives) can be used to determine surface shape.

(12) The inability of humans to perceive the correct shape in some situations where texture and contour information is lacking. A new illusion is demonstrated, and an explanation for this illusion is provided.

(13) The addition of surface markings is sometimes sufficient to allow the correct perception of surface shape in cases where shading alone is insufficient. An explanation this phenomenon is presented.

(14) The ability of humans to estimate the illuminant direction in novel images is quantified, and a theory advanced which accounts for human performance.

Acknowledgments

I would like to thank my advisor, Professor Whitman Richards, for his helpful suggestions, for prompt and constructive criticism on the innumerable drafts of this thesis, and most especially for providing unstinting and constant support.

I would like to thank the other members of my thesis committe, Professors Azriel Rosenfeld and Shimon Ullman for their participation in forming the balance and direction of this thesis, and particularly for providing me with a standard of excellence in research, a standard which I hope this thesis and all of my subsequent research lives up to.

In the same vein, I would like to thank Professor Berthold Horn for providing both inspiration and example for all students of vision research.

I would also like to remember Professor David Marr, who regretfully passed away during the first phases of the research for this thesis, for bringing together and providing much of the driving force behind the M.I.T. research group which has provided the facilities and intellectual climate which produced this reseach.

And last, but not least, I wish to extend a richly deserved thanks to my fellow students, who profoundly influenced this research through discussions, brain-storming sessions and insightful criticism. Chief among these have been (in cronological order) Andrew Witkin, Donald Hoffman, Joseph Scheuhammer and Eric Grimson.

Contents

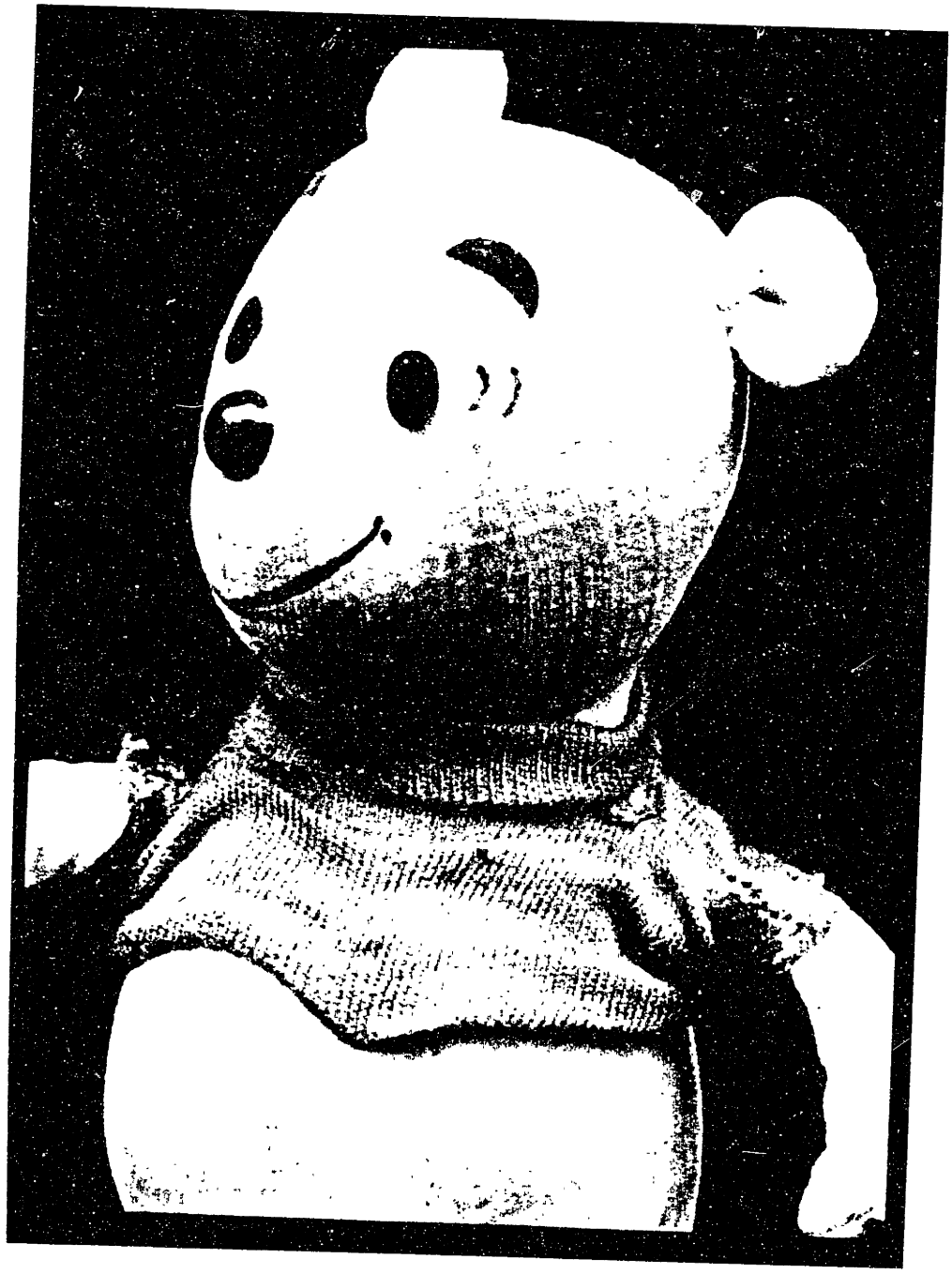


Figure 0-0. Typically, most of an image is shading, and contours are fairly infrequent.

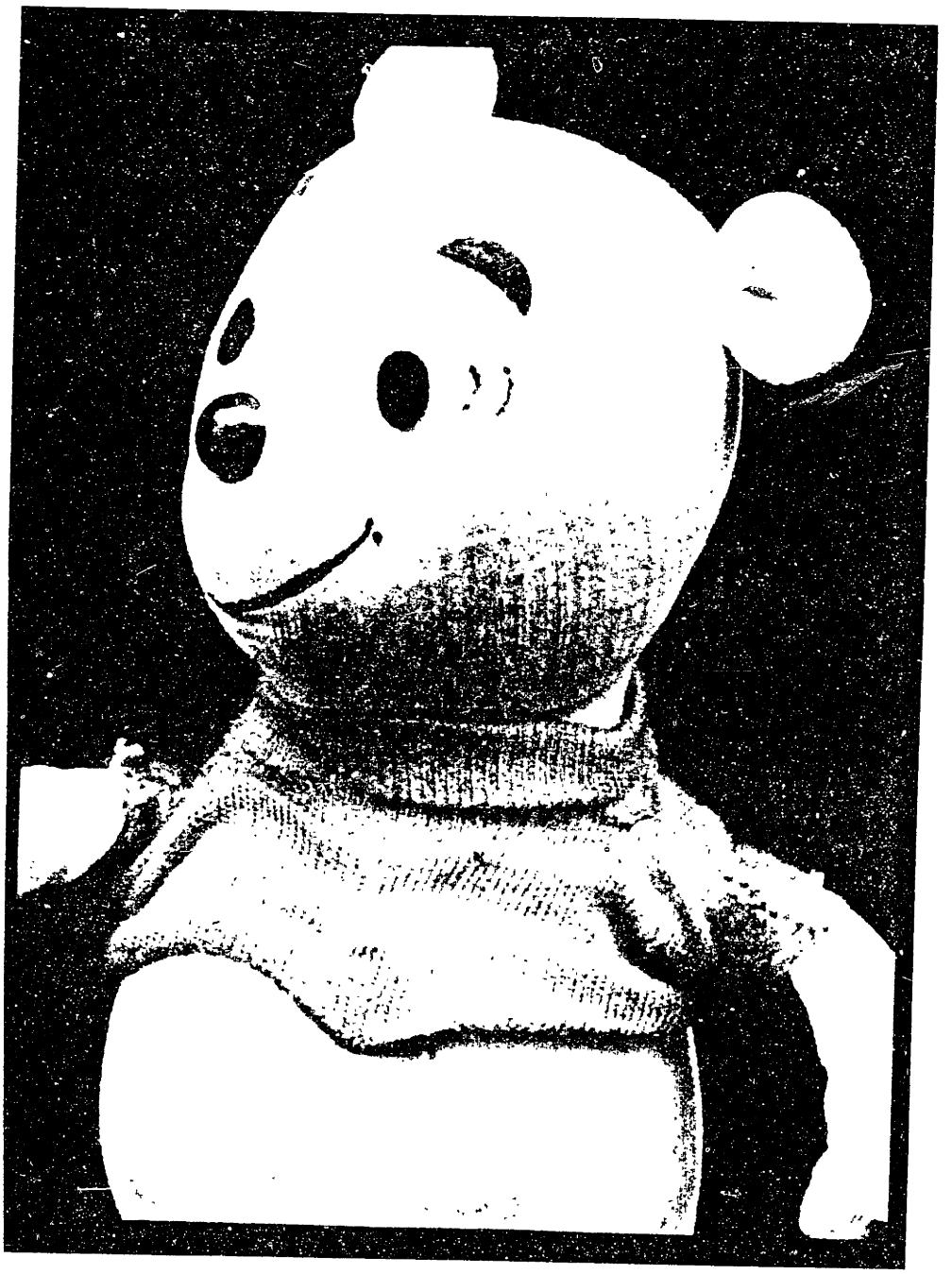


Figure 0-0. Typically, most of an image is shading, and contours are fairly infrequent.

1 Introduction

How do you determine the shape of an unfamiliar object from an image of that object? This is the central question of this thesis. The answer to this question is important because the perception of *shape* is central to the human visual ability.

The full answer to how we perceive shape is very complex. It is not only the image of a scene which determines our perception of shapes in the scene but also higher-level knowledge about the world, for instance, knowledge about gravity. Before examining how such high-level knowledge contributes to the perception of shape, however, it seems that it might be best to first investigate to what extent we can use local image information to directly calculate shape.

A spatially-restricted analysis of a single image is the logical first stage of any visual system. This first stage of analysis is especially important because it determines what information is available to the remainder of the visual system, and therefore determines what is required of the remainder of the system. If the first stage of analysis produces a rich description of the structure of the world, then the remainder of the visual system will be much simpler than if it had to deal directly with all of the ambiguities of the image. It is therefore important to perform as much analysis as is possible in this first stage of processing. If it is possible to calculate shape from local image information in a straightforward, bottom-up manner then the design of any visual system should take full advantage of this possibility. There is overwhelming evidence that biological visual systems are organized in this manner, devoting a large percentage of their neurons to the initial local analysis of the image.

This reasoning sets the scope of inquiry for this thesis. The question of how we perceive shape from an image will be addressed by investigating how shape may be determined locally, that is, how the shape of a surface at a point in the image may be determined using only the image information within a small neighborhood of that point. It is important to understand that limiting the scope of inquiry to a local analysis of the image does not imply that we can obtain only simple or local results; there are many examples of local, parallel processes which exhibit complex, global properties (see Rosenfeld *et*

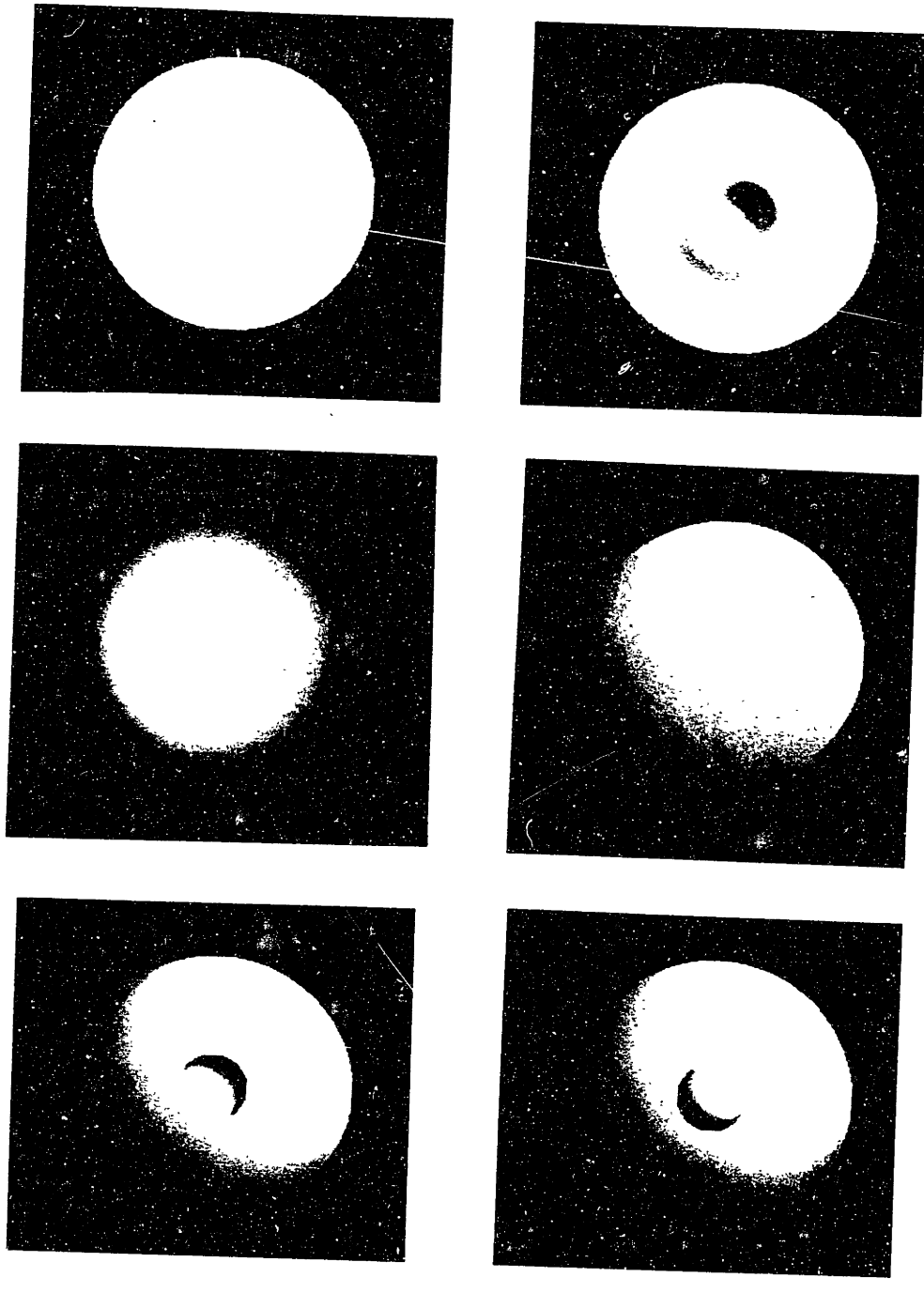


Figure 0-1. People use shading information to help them derive an impression of surface shape. Images of a disk, a sinc function, two spheres, and two other shapes, a sphere with a hollow in it and a sphere with a bump on it. In the first four cases the contours are identical, it is only the shading which differs. In the last two cases the illuminant direction must be used to determine the convexity/concavity of the surface.

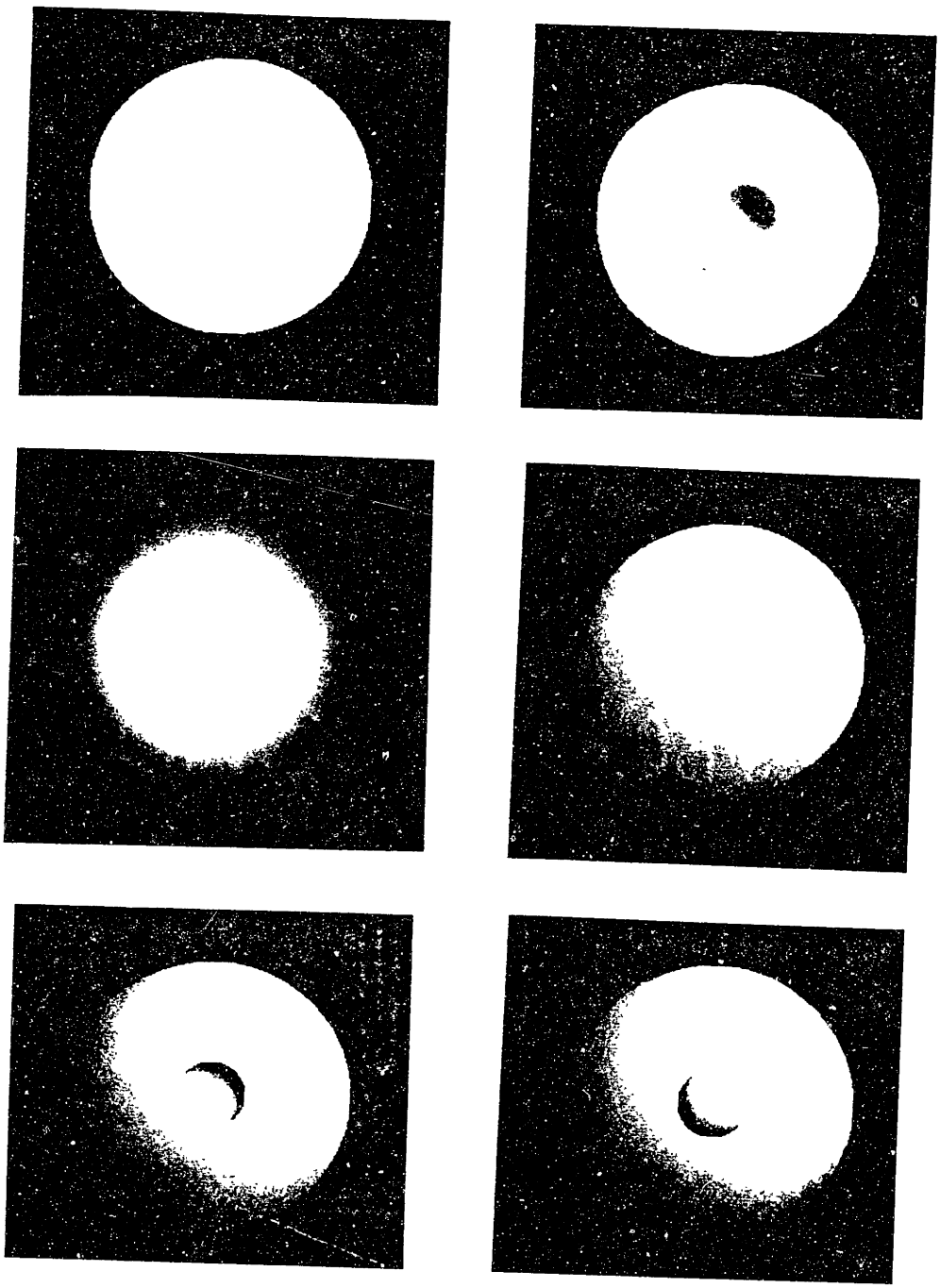


Figure 0-1. People use shading information to help them derive an impression of surface shape. Images of a disk, a sinc function, two spheres, and two other shapes, a sphere with a hollow in it and a sphere with a bump on it. In the first four cases the contours are identical, it is only the shading which differs. In the last two cases the illuminant direction must be used to determine the convexity/concavity of the surface.

al 1976, Ullman 1979)

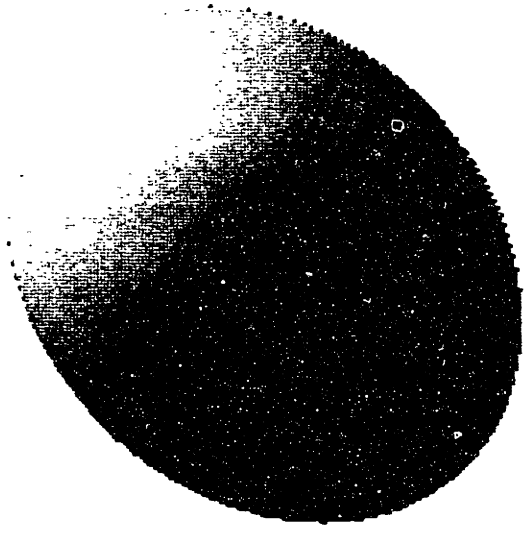
1.1 Shape Perception: Shading, Illuminant Direction and Contour

People use shading information to help them determine the shapes of objects. In figure 0-1 the only difference between the first four shapes is in the shading; the contours are the same. Thus the different perceptions of shape which result from viewing these four images are due to differences in shading. Interpreting shape from shading cannot always be as straightforward as in these first four images. In the bottom two images the shading cannot determine the shape directly; the shading must first be used to determine the illuminant direction, and the illuminant direction can then be used to determine the convexity/concavity of the central bump/hollow.

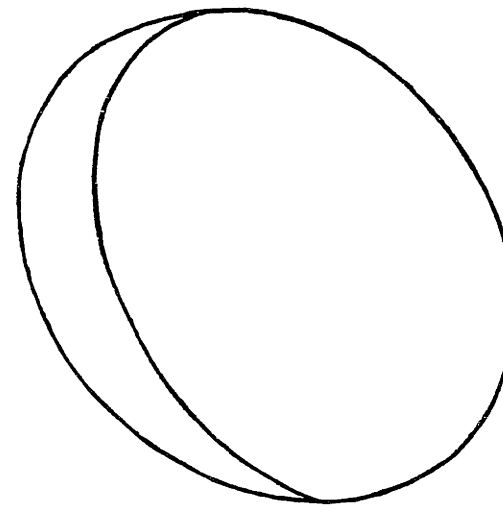
People cannot always use shading to correctly determine shape, however. In figure 0-2, we see an example which shows that shading alone can be insufficient for recovering surface shape. Part (A) of this figure is generally perceived as a flatish, egg-shaped object lying in the image plane. The shape of the object seen in (A) is actually more like the Hindenburg, as shown in figure 0-3 which is a side-view relief map of the object. This Hindenburg illusion shows a basic deficiency in the ability of people to use shading information.

When we examine a small neighborhood around a point in an image, usually all we find in the neighborhood is small changes in shading (changes in image intensity). It is unusual to find a contour within the neighborhood of an image point (see, for instance, figure 0-0). Thus if we are to investigate what we may learn about surface shape from local examination of an image, we must concern ourselves with shading. After determining what information shading gives about surface shape we can examine how imaged contours, and factors such as the scene's illumination, interact with and constrain the interpretation of shape derived from shading.

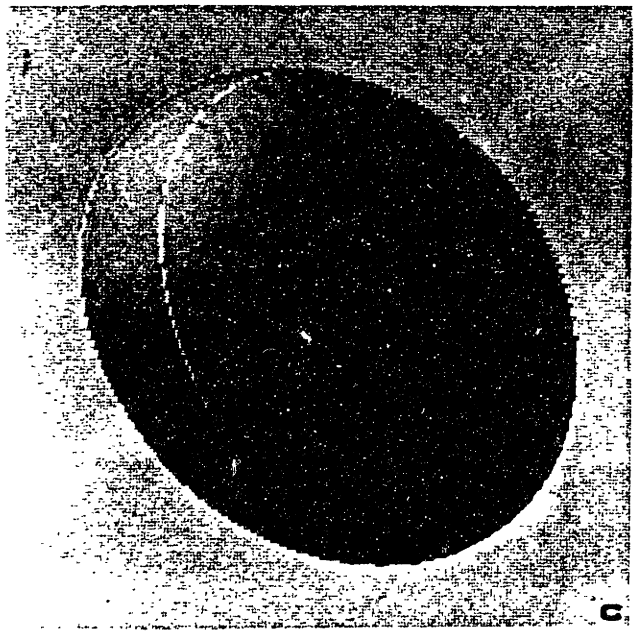
If we add the contour shown in (B) of figure 0-2 to the shaded Hindenburg image of (A), we obtain the image which is part (C) of figure 0-2. This image gives a much greater impression of relief than the shading alone, and certainly more of a perception of depth than the contour by itself. The contour information alone is inadequate to give an impression of three-dimensional shape, yet the addition of this single contour to the shading has changed the perceived shape to a much more correct perception



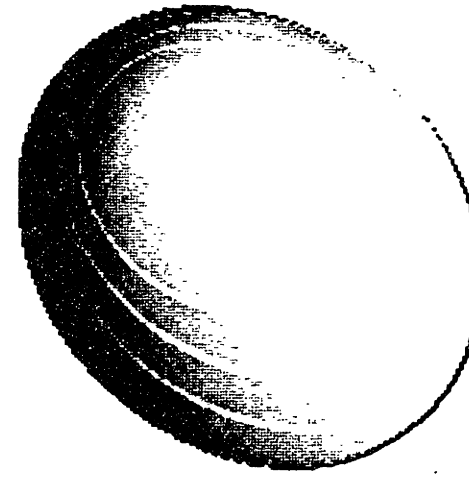
A



B

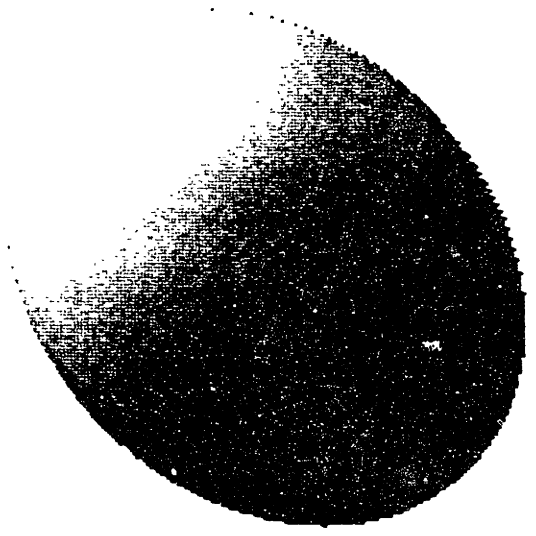


C

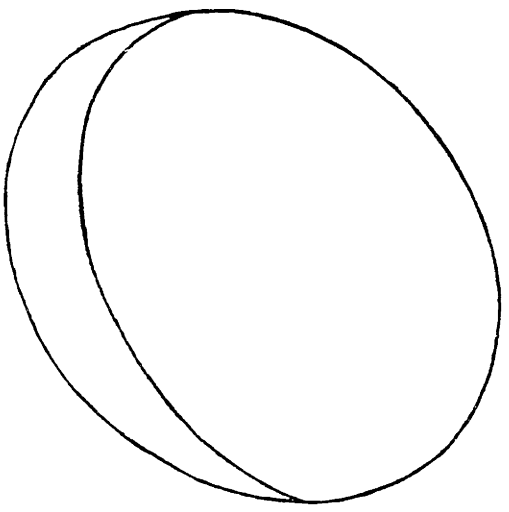


D

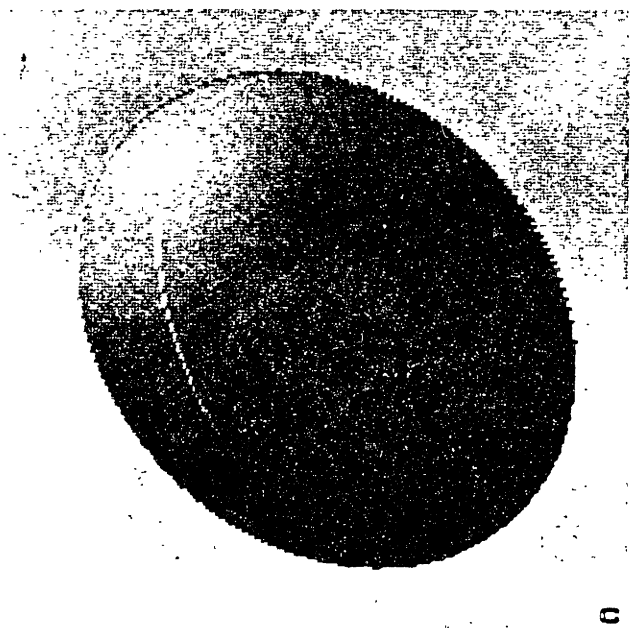
Figure 0-2. (A) a shape with shading, (B) a contour drawn on the surface, together with the outline of (A), (C) a shape with shading plus the contour. (A) is generally perceived as a flattened egg-shape lying in the image plane, and (B) generally gives little impression of three-dimensional shape, yet (C) which is the combination of (A) and (B) produces a fairly strong three-dimensional impression of a elongated egg-shape (a "Hindenburg" shape) pointing out of the image plane at 45°, which is the true shape of this figure. The addition of contour does not always give rise to a correct impression of shape, as shown by (D) which is the same shape as (A) and (C) but with different contour information drawn on the surface of the Hindenburg.



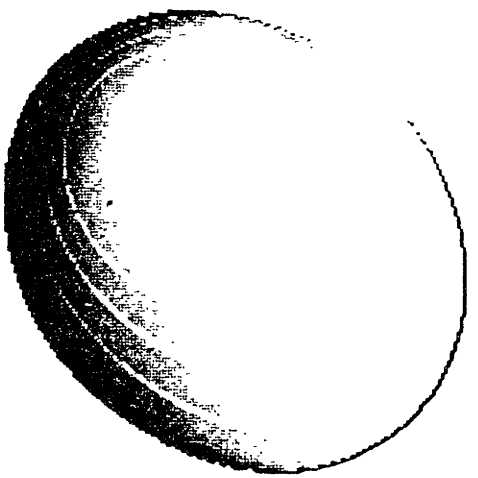
A



B



C



D

Figure 0-2. (A) a shape with shading, (B) a contour drawn on the surface, together with the outline of (A), (C) a shape with shading plus the contour. (A) is generally perceived as a flattened egg-shape lying in the image plane, and (B) generally gives little impression of three-dimensional shape, yet (C) which is the combination of (A) and (B) produces a fairly strong three-dimensional impression of a elongated egg-shape (a "Hindenburg" shape) pointing out of the image plane at 45°, which is the true shape of this figure. The addition of contour does not always give rise to a correct impression of shape, as shown by (D) which is the same shape as (A) and (C) but with different contour information drawn on the surface of the Hindenburg.

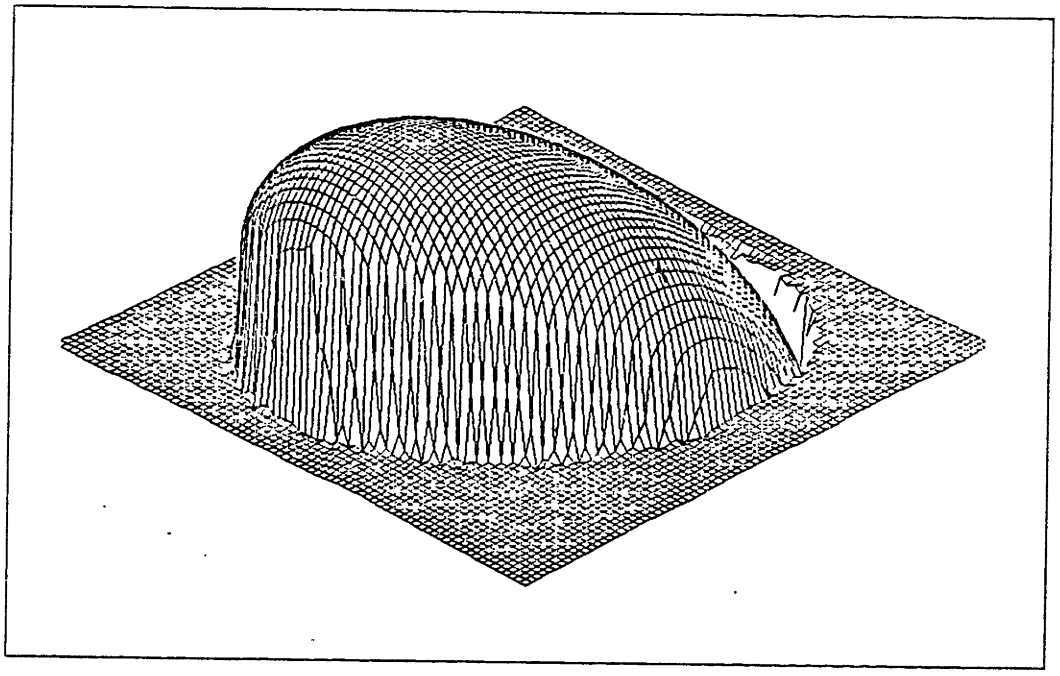


Figure 0-3. Side view of a relief map showing the shape of the objects in figure 0-2. This object is shaped like the Hindenburg, and is slanted out from the image plane at 45°.

of the Hindenburg shape.

One might think of shading information as a glue-like substance which fills in and ties together the image, but which is also quite maleable until it is "set" by the contour information. The shading alone gives a weak impression of the depth in these images, and the contour alone gives almost no impression of depth, but together the shading and contour can give a relatively strong three-dimensional impression of shape. Image (C) shows that it can take relatively little contour information to firmly "set" the shading "glue", however the addition of contour information does not always result in the correct perception of the shape, as shown by (D) of this figure. This image contains more contours than (C), and yet (D) gives only a flat impression of shape. Thus not all contours can be used to help interpret the shape implied by shading.

Sharpening The Question. The examples of figures 0-1, and 0-2 help to sharpen the original question of how do we determine the shape of an object in a picture. We now see that we can obtain a useful answer to the original question by finding the answer to these four more specific questions:

How is it that surface shape may be recovered from an image using a local analysis of shading?

Why is shading insufficient to recover shape in some cases, but not in others?

Why does the addition of contour information give a correct impression of shape in some cases but not in others?

How does knowledge of the illuminant direction provide constraint on the computation of surface shape?

In order to answer these questions, we must first address the *theoretical* problems posed by these questions; that is, we must first understand what information about object shape is contained in the image, and how it is *theoretically* possible that this information could be used to obtain an estimate of the object shape (Horn 1975, Marr 1977). This requires an analysis of the process of image formation which shows explicitly the relationship between image intensity and the physical characteristics of the viewed scene. Once we have derived equations which express the relationship between image and physical parameters of the scene, we may use these equations to deduce the structure of the scene using the evidence of the image. These equations together with the inference scheme which allows us to take image data and deduce the physical structure of the viewed scene form a *computational theory* of shape perception.

When we have developed such a computational theory, describing what information is contained within the image and how this information can be used to recover shape from the image, then we will know *what* it is that the human visual system is doing. Then we can begin to understand *how* it is being done.

1.2 Previous Work

Most directly related to this thesis is the work of Berthold Horn and his associates, who have developed an approach to deriving shape from image intensity measurements.

In 1970 Horn showed that the shape of an object may be recovered from image intensity information, given that several facts are known beforehand about the objects' surface and the illuminant. Horn found that he could compute surface shape from image intensities given *a priori* knowledge of

(1) the distribution of illumination, (2) the bidirectional reflectance function of the objects' surface, and (3) the surface orientation along an initial curve on the objects' surface.

Horn and his colleagues (Horn 1975, 1977, Strat 1979, Brooks 1979) have developed numerical integration schemes for using image intensity to solve for object shape. These numerical algorithms correctly recover the shape of the imaged surface given the required *a priori* knowledge, although these schemes have some numerical stability problems. These algorithms have the particular disadvantages that they require that surface orientation be known along an initial curve on the objects' surface, and that the surface may not contain a smooth occluding contour (such as is found at the edge of an image of a sphere).

In order to extend the general applicability of Horns' shape-from-shading methodology, Ikeuchi and Horn 1981 have developed an iterative numerical scheme which makes use of points at which surface orientation might possibly be determined without *a priori* knowledge (e.g., a smooth occluding contour) rather than simply assuming that an initial curve is known. The illuminant direction and surface reflectance function are still assumed to be known *a priori*. Bruss 1981, however, proves that surface shape is correctly and unambiguously derived using this scheme only for images in which image intensity is determined completely by the slant of the surface[1] . Bruss' result also proves that shape cannot be unambiguously determined from a local analysis of image intensity even given the *a priori* knowledge of illuminant direction and reflectance function.

These shape-from-shading techniques have been shown to be useful in situations where there is sufficient information known *a priori* about the image, such as in a factory setting where one can know about the illumination and the surface reflectance function beforehand. These techniques have also been shown to be useful in analysis of LANDSAT imagery[2] (see Horn and Sjoberg 1979), where shape is known but reflectance is not. In such imagery shape-from-shading techniques may be applied to determine the contribution of shape to image intensity, this contribution may then be subtracted

[1]Specifically, only if the reflectance map R is $R(p,q) = f(p^2 + q^2)$, e.g., a Lambertian surface with the illuminant at $\mathbf{V}$, the viewers' position, or at $-\mathbf{V}$, as in scanning electron microscope images.

[2]imagery collected by NASA's earth-resources satellite, LANDSAT.

from the image intensity resulting in a value which is much more directly related to the reflectance of the viewed material.

The research performed by Horn and his colleges comprises the greater part of our understanding of how surface shape and shading are related. The question of this thesis, however, is how *people* determine shape from shading and other cues. All of the previously mentioned analyses assumed substantial *a priori* knowledge of the scene; such knowledge is simply not available in the typical visual task faced by the human visual system. Thus it remains for us to discover what may be determined about the world without such *a priori* knowledge. In addition, the Horn-type shape-from-shading techniques function by propagating constraint from boundary conditions over the surface whose shape is to be estimated. Thus these techniques use global, rather than local image information to estimate surface shape. We are interested in how purely *local* image information may be used to estimate shape, because local analysis is the logical first stage of any visual system.

In order to address the question of what information about shape may be obtained using only local image information, with no *a priori* information assumed, we must produce a new analysis of image shading. Unfortunately, only limited use can be made of Horn's analysis, because it is based on a mathematical construct called the *reflectance map*, which describes how intensity varies with surface orientation given a particular combination of viewer, surface material and illuminant. In order to answer our question, it will be necessary to examine how changes in image intensity depend on *all* the physical variables which determine the image, not only surface orientation but also surface curvature, illuminant direction and so forth.

Both Witkin (1980, 1981) and Stevens (1979, 1981) have investigated the question of how imaged contour information, such as generated by surface markings, can be used to recover surface shape. For both Witkin and Stevens the primary problem is that the shape of imaged contours results from two sources: the inherent shape of the contour generator, and the foreshortening caused by the angle between the contour generator and the image plane (see also Kender 1980). The amount of foreshortening must be determined before the contour shape, and thus underlying surface shape, may

be determined.

Witkin analyzed the case where nothing is known about the shape of the contour generator, and so its shape might best be treated as a randomly distributed quantity. Because viewer position and contour generator shape are independent of each other, over all scenes the distribution of contour generator tangent direction will be uniform. Thus any anisotropy in the distribution of imaged contour tangent direction is most likely the result of foreshortening. Measurement of this anisotropy allows a maximum likelihood estimate of the foreshortening may be made, and thus the shape of the contour generator determined.

Witkin treated in depth the case of a surface patch covered with random curves. He showed that in scenes where the surface is densely covered with surface markings the foreshortening can be accurately estimated, and thus the surface orientation determined. He showed several examples where the surface orientation was recovered from an image of a naturally occurring scenes, through an analysis of contours found in the scene.

Witkins results show that surface orientation may be reliably recovered from surfaces densely covered with isotropically-distributed surface contours; however surfaces with an anisotropic distribution of surface contours (such as tree bark) produce erroneous estimates, and the variance of the surface orientation estimate increases rapidly as the density of the randomly-distributed surface contours decreases.

Stevens analyzed several cases in which people can recover surface shape from imaged contours and line drawings, such as the relief plots used throughout this thesis. He demonstrated that it is necessary to make assumptions about the relationship of the imaged contour to the contour generator and the surface that the contour generator lies upon. He then investigated what inferences may be made about surface shape given various special configurations of imaged contour and the assumption of general position. These inferences about surface shape from the imaged contours appear to be sufficient to account for perceived surface shape in the examples discussed[3].

[3]Currently Stevens is working toward implementing these inferences in a computer program, so that his inference scheme may be tested.

Some of the inferences about surface type developed by Stevens have analogies in the domain of local analysis of shading. For instance, the inference that if an imaged specularity is linear then the surface must be cylindrical is directly analogous to the method proposed here to determine surface type (the proposition dealing with determination of surface type is proved in chapter 2). More generally, many of the inferences are similar, apparently because the geometry of projection is identical for any image information. Thus, for instance, the theory for estimation of illuminant direction appears analogous to Witkins' estimation of surface orientation from the distribution of imaged contours.

Both the work of Witkin and that of Stevens has important implications for the interpretation of shape from imaged contour. This thesis will investigate the interaction of estimates of surface shape made using imaged contour with estimates made using shading information. This will provide a framework for integrating information about surface shape obtained from inferences about imaged contour with information about surface shape obtained from shading.

Chapter 1

Groundwork For A Theory

1 Deciding Input And Output

What is the appropriate image information to use as input for the computation of surface shape? Because we are limiting this inquiry to local computation of surface shape in novel, unfamiliar images we are limited to image intensity or its derivatives as the input information. Any other information requires either *a priori* information about the scene, or more global processing.

What is the appropriate output representation? The words "surface shape" can mean any of several things, including the surface orientation (or equivalently, the surface normal) at each point, or a map of relative depth from the viewer at each point. Both of these quantities contain the same information; thus for a computational theory it doesn't matter which we deal with. For convenience of explication, I will use the surface orientation (surface normal) as the output of the theory.

One of the consequences of these decisions is that absolute image intensities alone are not sufficient for our purposes. Measuring the absolute image intensity yields only one measurement at each point in the image, whereas surface orientation has two degrees of freedom[1] . It is impossible to determine both of the parameters of surface orientation with only one measurement, so that measurement of image intensity alone is insufficient: we need more measurements at each point. Both the first spatial derivative (which yields two measurements per point), and the second spatial derivative (three measurements per point) both provide sufficient measurements per point to potentially allow determination of both the parameters of surface orientation[2] .

[1]The two degrees of freedom may be expressed as either slant $\sigma = \cos^{-1} dz$ and tilt $\tau = \tan^{-1} dy/dx$ (Stevens 1979) or $p = dz/dx$ and $q = dz/dy$ (Horn 1975), for instance.

[2]Note that usually it requires two points to measure dI and three to measure d^2I. In order to show how we may obtain derivatives at one point requires introducing the notion of *scale*, which is beyond the goals of this section. This problem is treated in chapter 3 section 1.

The use of either the first or second spatial derivatives of image intensity fits well with what is known of the human visual system. The human visual system is sensitive to *changes* in image intensity, rather than to absolute image intensity. In chapter 4 it is shown that the human sensitivity to changes in image intensity may be reasonably characterized by *either* the first derivative of image intensity (Mach 1910), or by the second derivative of image intensity (Kuffler 1952, 1953).

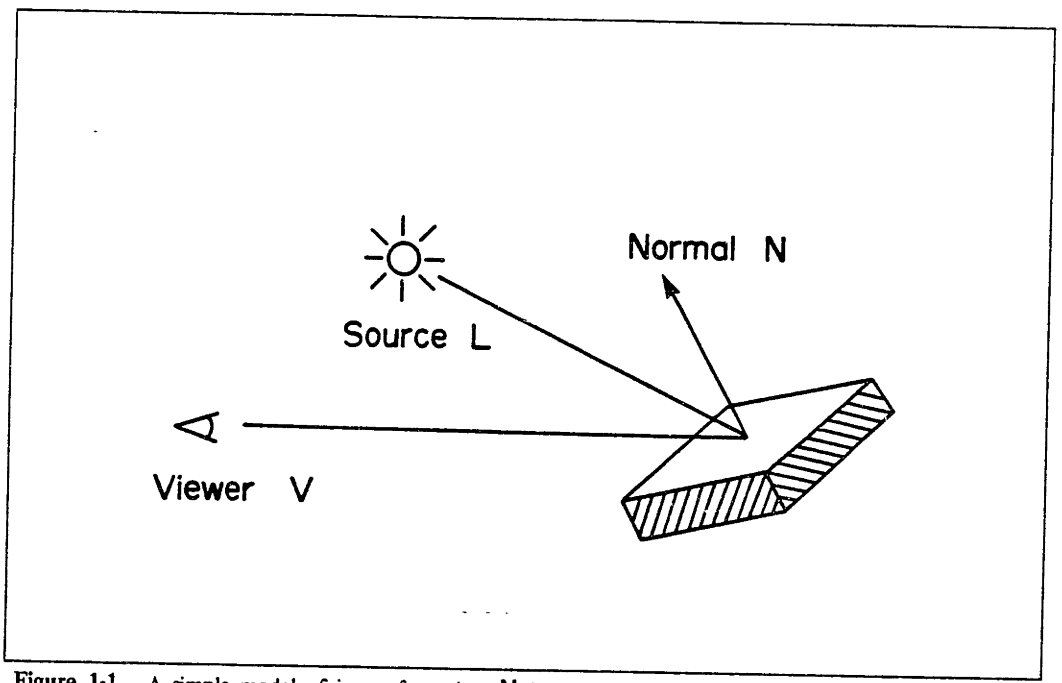


Figure 1-1. A simple model of image formation. **N** is the surface normal, **L** is the illumination direction, **V** is the viewers direction. If λ is the flux emitted toward the surface, ρ the average reflectance of the surface, and we assume distant light source and a Lambertian reflectance function for the surface then the image intensity I is given by $I = \rho\lambda(\mathbf{N} \cdot \mathbf{L})$.

2 The Process Of Image Formation

The original question of this thesis, how we perceive shape, has been sharpened to the question of how changes in image intensity in the neighborhood of a point can be used to recover the surface normal (or equivalently, surface orientation) at that point. In order to answer this question we must discover what information about the three-dimensional surface normal resides in the two-dimensional image, and that means we must investigate how changes in the image reflect the structure of the three-dimensional world.

In order to be able to make quantitative statements about the world from the image, we must develop a mathematical model of the process of image formation. Figure 1-1 shows a simple model of image formation: a distant point source illuminant at direction **L**, a patch of surface with surface normal **N**, and a viewer in direction **V**. Throughout this thesis all bold face variables (e.g., **N**, **L**, **p** etc.) represent three-dimensional vectors, all other variables are scalars.

The surface normal **N**, the viewer direction **V** and the illuminant direction **L** are unit vectors in

Cartesian three-space. As they are unit vectors, two parameters suffice to specify them, the third being determined by the constraint that they have unit magnitude. Two parameters which are often chosen are the slant, σ, and the tilt, τ (see Stevens 1979).

The tilt of a surface is $\tan^{-1}(y_N/x_N)$, where x_N and y_N are the x and y components of the surface normal, and the slant of the surface is $\cos^{-1}(z_N)$, where z_N is the z component of the surface normal. The tilt and slant are a useful way to characterize directions because they divide the components of a direction into the orientation of the image-plane component, which is the tilt, and the slope of the depth component, which is the slant. This image-plane versus depth distinction seems to fit well with some of our intuitions about the world, and with some of the properties of images.

2.1 Three Components Of Image Formation

There are three parts to our model of image formation: the illumination of the surface, the reflection of light falling on the surface back into space, and the projection of this light into the eye, forming the image. Each of these portions of the process has characteristicly different effects on the image, and so it is useful to examine each of them separately.

The Illuminant Component Of Image Formation. The amount of light falling on a surface varies with the angle between the surface and the illuminant: a small patch of the surface which faces away from the light receives less light than it would if it were directly facing the light. This illumination foreshortening is proportional to the cosine of the angle between the surface normal and the illuminant[3] so that the size of an infinitesimal patch of the surface dA_S appears to have the foreshortened area dA_L when seen from the direction of the illuminant:

$$dA_L = \mathbf{N} \cdot \mathbf{L} \quad dA_S$$

Thus if we let λ be the amount of light (flux) per square area arriving in the neighborhood of the surface of an object from a distant point-source illuminant positioned in direction $\mathbf{L}$, then λ_N, the

[3]The cosine of this angle may be expressed as the dot product of $\mathbf{N}$ and $\mathbf{L}$.

amount of light per meter squared incident on a patch of surface with surface normal **N**, will be

$$\lambda_N = \lambda \mathbf{N} \cdot \mathbf{L}$$

The Surface Component Of Image Formation. We now have an expression for the amount of light falling on the surface of the object. The next step is to express how that light is reflected from the surface of the object back into space, and eventually onto the image plane.

To do this we must define ρ, the albedo of the surface, and R the reflectance function[4] of the surface.

The albedo ρ is the portion of incident light which is reflected. Thus if a surface had an albedo of 0.2 (the average for all natural surfaces — see Krimov 1971, Richards 1981), then only twenty percent of the incident light would be reflected, with the rest absorbed into the surface, increasing the temperature of the surface.

The reflectance function $R(\mathbf{N}, \mathbf{L}, \mathbf{V})$ describes how much of the reflected light leaves in each direction. The amount of incident light which is reflected in the viewer direction, **V**, is a function of the illuminant direction, **L**, and the surface normal, **N**. In the case of the reflectance function of a mirror, all incident light is simply reflected at an angle equal to the incident angle. At the opposite end of the spectrum of reflectance functions is the Lambertian reflectance function, where the light is reflected proportional to the cosine of the incident angle, i.e., proportional to $\mathbf{N} \cdot \mathbf{L}$. This reflectance function is an approximation of that typical of rough, matte surfaces. Most reflectance functions are somewhere between these two.

Using the albedo and the reflectance function we may describe how light incident upon the surface is reflected back into space. If we let λ_R be the amount of light reflected from the surface in direction **V**, then

$$\lambda_R = \rho \lambda_N R(\mathbf{N}, \mathbf{L}, \mathbf{V}) = \rho \lambda (\mathbf{N} \cdot \mathbf{L}) R(\mathbf{N}, \mathbf{L}, \mathbf{V})$$

[4]This function is neither the bi-directional reflectance function nor Horn's reflectance map. If we set $\mathbf{N} = (0, 0, 1)$ and vary the values of **L** and **V**, we generate the bi-directional reflectance function. If we fix **L** and **V**, and let **N** vary, we generate Horn's reflectance map.

The Projection Component Of Image Formation. The final portion of the process of image formation is the projection from the surface of the object onto the image plane. As with the projection of the illuminant onto the surface of the object, there is a foreshortening effect which is proportional to the cosine of the angle between the viewer direction **V** and the surface normal **N**, as expressed by $\mathbf{N} \cdot \mathbf{V}$. Thus an infinitesimal patch of surface dA_S is mapped into an infinitesimal patch of the image dA_I according to

$$dA_S = (\mathbf{N} \cdot \mathbf{V})^{-1} dA_I$$

Thus I, the image intensity[5], is given by

$$I = \lambda_R(\mathbf{N} \cdot \mathbf{V})^{-1} = \rho\lambda(\mathbf{N} \cdot \mathbf{L})R(\mathbf{N}, \mathbf{L}, \mathbf{V})(\mathbf{N} \cdot \mathbf{V})^{-1}$$

A Lambertian reflectance function is defined as

$$R(\mathbf{N}, \mathbf{L}, \mathbf{V}) = \mathbf{N} \cdot \mathbf{V}$$

which is the reciprocal of $(\mathbf{N} \cdot \mathbf{V})^{-1}$, the projection foreshortening term. Thus for a Lambertian surface the reflectance function and the effects of projection cancel each other. Because this cancellation considerably simplifies many calculations, a Lambertian reflectance function is often assumed. If we assume a Lambertian reflectance function, then the equation for image intensity becomes

$$I = \rho\lambda\mathbf{N} \cdot \mathbf{L} \tag{1}$$

2.2 The First Derivative Of Image Intensity

The image intensity I and the surface normal **N** are different at each point (x, y) in the image, and so are perhaps better written $I(x, y)$ and $\mathbf{N}(x, y)$. When discussing them at a point P, however, they

[5]In the interests of avoiding confusion, the words "image intensity" will be used throughout this document rather than using the more technically correct "image irradiance" for the flux per unit area falling on the image plane, and "image intensity" for the for the measured image irradiance. The two may be assumed to be numerically equal, and thus the distinction is of little interest to the task at hand.

will be written as simply I and $\mathbf{N}$. Similarly, I will write dI and $d\mathbf{N}$ to designate the first derivative of image intensity and the surface normal, respectively, at a point P in the direction (dx, dy). That is,

$$dI = I_x dx + I_y dy \qquad d\mathbf{N} = \mathbf{N}_x dx + \mathbf{N}_y dy$$

where

$$I_x = \frac{\partial I}{\partial x} \qquad I_y = \frac{\partial I}{\partial y} \qquad \mathbf{N}_x = \frac{\partial \mathbf{N}}{\partial x} \qquad \mathbf{N}_y = \frac{\partial \mathbf{N}}{\partial y}$$

The first derivative of image intensity may also be expressed in terms of the physical parameters of the scene. If we are examining a small, homogeneous region of an image it is reasonable to assume that the illumination and albedo of the surface change very little, and so we may treat $\mathbf{L}$, ρ, and f as constants[6] . If we also assume a Lambertian reflectance function, so that equation (1) applies, then

$$\begin{aligned} dI &= d(\rho\lambda\mathbf{N}\cdot\mathbf{L}) \\ &= \rho\lambda d\mathbf{N}\cdot\mathbf{L} + \rho\lambda\mathbf{N}\cdot d\mathbf{L} \\ &= \rho\lambda d\mathbf{N}\cdot\mathbf{L} \end{aligned} \tag{2}$$

The term $\mathbf{N}\cdot d\mathbf{L}$ is zero because $\mathbf{L}$ was assumed constant. Thus the first derivative of image intensity is a function of the first derivative of the surface normal. We must therefore investigate the nature of $d\mathbf{N}$, and how it relates to the surface normal, $\mathbf{N}$.

$d\mathbf{N}$ In Terms Of Surface Properties. Let us represent the surface in the neighborhood of a point P by a Monge patch, i.e.,

$$\mathbf{p} = u\mathbf{e}_1 + v\mathbf{e}_2 + f(u,v)\mathbf{e}_3$$

where $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ are orthogonal basis vectors, and $\mathbf{p}$ is a vector giving the position of a point on the surface at point (u, v). This is shown in figure 1-2.

[6]This assumption is not reasonable at occluding contours, of course; thus all of the calculations dependent upon these equations break down at occluding contours. This causes no problems if we can identify occluding contours (or potential occluding contours) beforehand and treat them specially. Chapter 2 section 5 addresses the problem of finding and identifying these contours.

If we position the surface with the origin of the (u, v) plane at point P, with the tangent plane to the surface parallel to the (u, v) plane then

$$\mathbf{p}_u = \frac{\partial \mathbf{p}(u,v)}{\partial u} = \mathbf{e}_1 + f_u \mathbf{e}_3 \qquad \mathbf{p}_v = \frac{\partial \mathbf{p}(u,v)}{\partial v} = \mathbf{e}_2 + f_v \mathbf{e}_3$$

where f_u and f_v are the partials of $f(u, v)$ with respect to u and v. Note that at the point P, $f_u = f_v = 0$, so that $\mathbf{p}_u = \mathbf{e}_1$ and $\mathbf{p}_v = \mathbf{e}_2$. Thus the partials $\mathbf{p}_u$, $\mathbf{p}_v$ of the surface vector $\mathbf{p}$ lie in the tangent plane to the surface. This arrangement is also shown in figure 1-2.

There are several quantities which are defined at each point on the surface, and which characterize the surface. At each point there is a unit surface normal $\mathbf{N}$ which defined as

$$\mathbf{N} = \frac{\mathbf{p}_u \times \mathbf{p}_v}{|\mathbf{p}_u \times \mathbf{p}_v|} = \frac{-f_u \mathbf{e}_1 - f_v \mathbf{e}_2 + \mathbf{e}_3}{\sqrt{1 + f_u^2 + f_v^2}}$$

Thus the surface normal is perpendicular to the tangent plane. The normal surface curvature κ_n is also defined in each direction θ at a point P

$$\kappa_n = \mathbf{p}_{uu} \cdot \mathbf{N} \cos^2\theta + \mathbf{p}_{uv} \cdot \mathbf{N} \cos\theta \sin\theta + \mathbf{p}_{vv} \cdot \mathbf{N} \sin^2\theta \qquad (3)$$

where

$$\mathbf{p}_{uu} = \frac{\partial^2 \mathbf{p}}{\partial u \partial u} \qquad \mathbf{p}_{uv} = \frac{\partial^2 \mathbf{p}}{\partial u \partial v} \qquad \mathbf{p}_{vv} = \frac{\partial^2 \mathbf{p}}{\partial v \partial v}$$

and θ is measured from the u axis (see Lipschutz 1969).

The curvature takes on distinct maximum and minimum values in orthogonal directions, unless the surface is a sphere or a plane, in which case the curvature is constant. The maximum and minimum directions are the *principal directions* of the surface at point P, and the maximum and minimum values of the curvature are, respectively, the first and second *principal curvatures*, κ_1 and κ_2. As the principal directions are perpendicular we may align the axes (u, v) with the principal directions and then the principal curvatures simplify to

$$\kappa_1 = \mathbf{p}_{uu} \cdot \mathbf{N} \qquad \kappa_2 = \mathbf{p}_{vv} \cdot \mathbf{N}$$

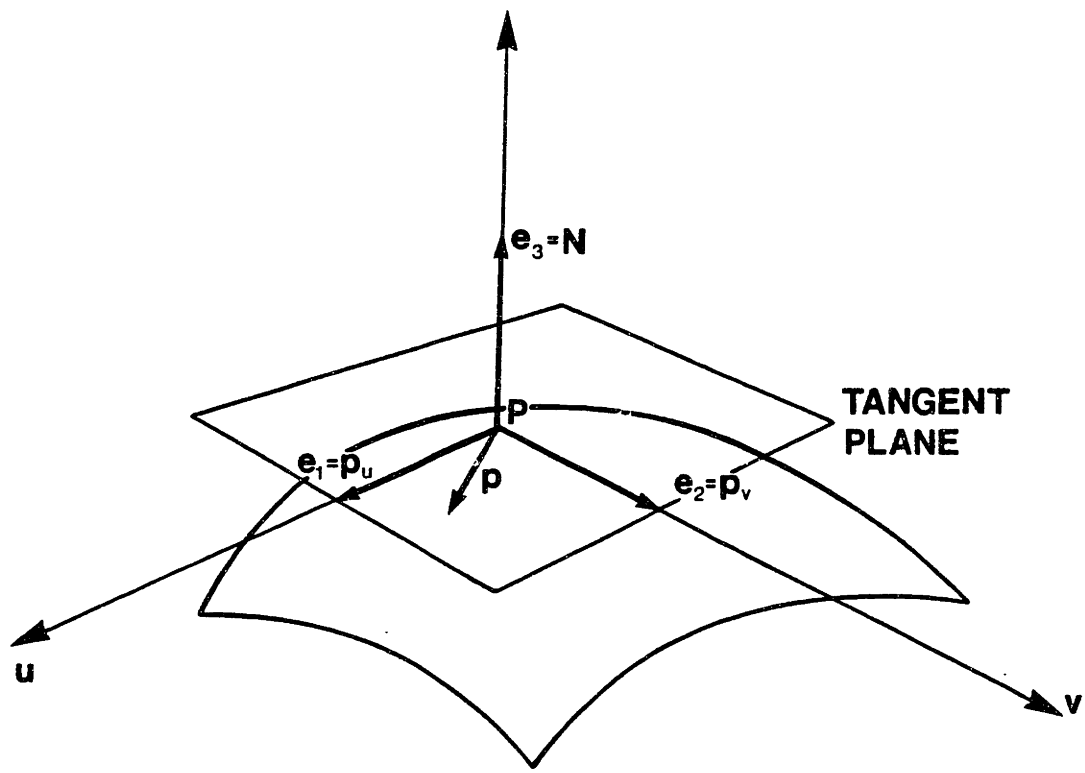


Figure 1-2. The position vector $\mathbf{p} = u\mathbf{e}_1 + v\mathbf{e}_2 + f(u, v)\mathbf{e}_3$ defines the surface, the surface normal $\mathbf{N} = \mathbf{e}_3$, the tangent plane defined by $\mathbf{p}_u = \mathbf{e}_1$, $\mathbf{p}_v = \mathbf{e}_2$, the partials of $\mathbf{p}$ at the point P

Equation (2) showed that the first derivative of image intensity dI in direction (du, dv) depends upon the first derivative of the surface normal $d\mathbf{N}$ in the direction (du, dv), i.e., $dI = I_u du + I_v dv$ depends upon $d\mathbf{N} = \mathbf{N}_u du + \mathbf{N}_v dv$. The vector $d\mathbf{N}$ is a linear approximation to the change in surface normal between (u, v) and $(u + du, v + dv)$, so that

$$d\mathbf{N} = \mathbf{N}(u + du, v + dv) - \mathbf{N}(u, v) + \mathcal{O}((du^2 + dv^2)^{1/2})$$

This is shown in figure 1-3. Note that $d\mathbf{N}$ is perpendicular to $\mathbf{N}$, as

$$2d\mathbf{N} \cdot \mathbf{N} = d(\mathbf{N} \cdot \mathbf{N}) = d(1) = 0$$

Thus $d\mathbf{N}$ lies in the tangent plane to the surface at P, as does $d\mathbf{p}$.

We would like to obtain a concise expression for $d\mathbf{N}$ in terms of surface curvature and orientation. We may obtain such an expression using $\mathbf{p}_u$ and $\mathbf{p}_v$, whose cross-product defines the surface normal. Because $\mathbf{p}_u$ and $\mathbf{p}_v$ are orthogonal vectors in the tangent plane, $d\mathbf{N}$ (which also lies in the tangent plane) may be written as a linear combination of them. Thus

$$d\mathbf{N} = \mathbf{N}_u du + \mathbf{N}_v dv = c_1 \mathbf{p}_u + c_2 \mathbf{p}_v$$

The constants c_1 and c_2, however, are not the same for $d\mathbf{N}$ taken along different directions (du, dv).

Let us assume that u and v are the principal directions at point P. This is shown in figure 1-4. Then Rodreguez formula tells us that

$$\mathbf{N}_u = -\kappa_1 \mathbf{p}_u$$
$$\mathbf{N}_v = -\kappa_2 \mathbf{p}_v$$

(see Lipschutz, 1969, pp. 183-187 for a proof of this theorem). As $d\mathbf{N} = \mathbf{N}_u du + \mathbf{N}_v dv$ then we see that in the direction (du, dv)

$$d\mathbf{N} = -\kappa_1 \mathbf{p}_u du - \kappa_2 \mathbf{p}_v dv \tag{4}$$

This then gives us an expression for $d\mathbf{N}$ in terms of the surface properties of curvature and orientation.

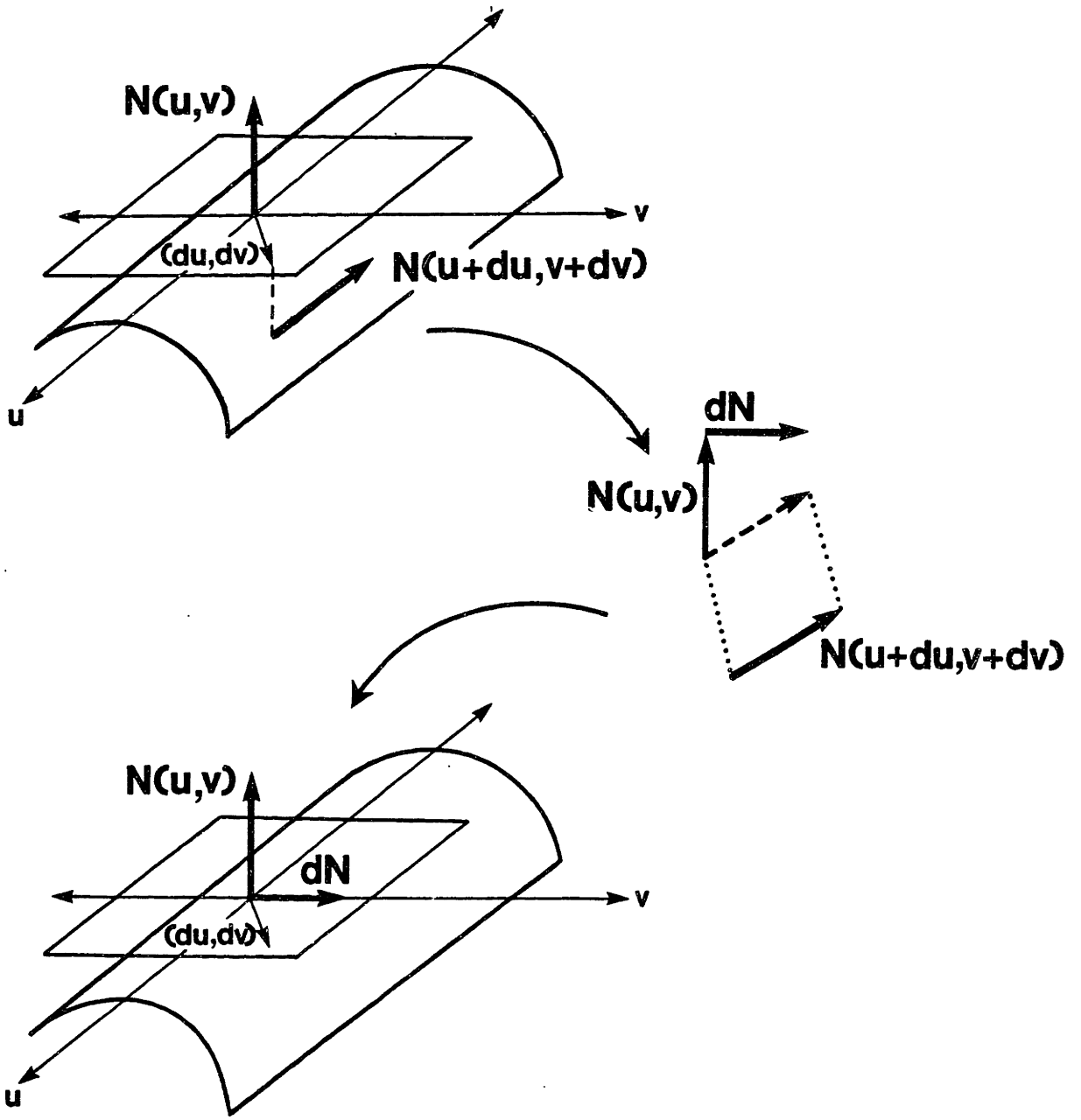


Figure 1-3. $d\mathsf{N} = \mathsf{N}_u du + \mathsf{N}_v dv$, the first derivative of N in the direction (du, dv), is a linear approximation to the change in surface normal along the differential step (du, dv). $d\mathsf{N}$ does not normally point in the direction (du, dv), unless the surface is a sphere.

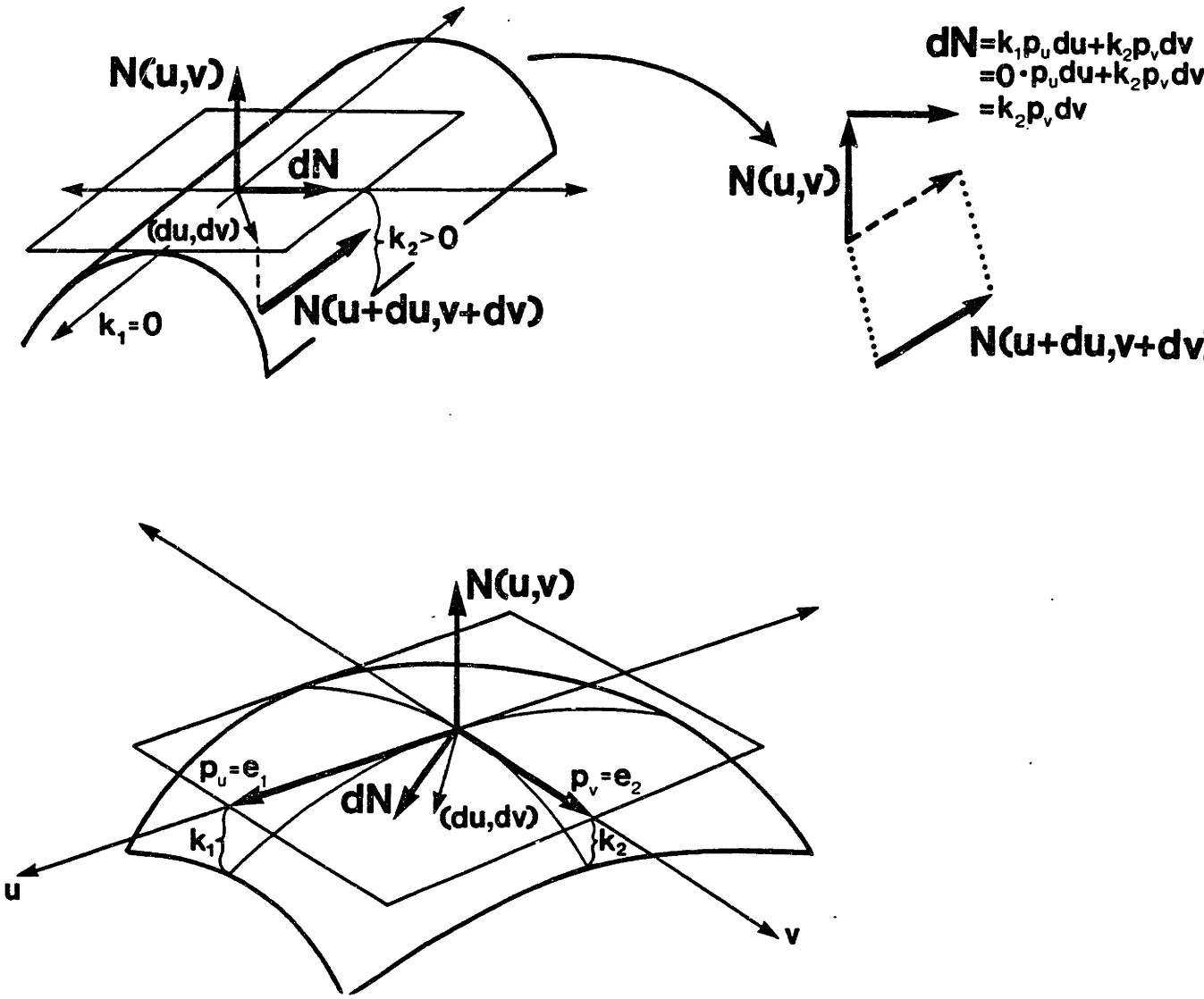


Figure 1-4. $d\mathbf{N}$ along (du, dv) may be expressed as a linear combination of $\mathbf{p}_u$ and $\mathbf{p}_v$, as it lies in the tangent plane. The relationship is $d\mathbf{N} = -\kappa_1 \mathbf{p}_u du - \kappa_2 \mathbf{p}_v dv$, where κ_1 and κ_2 are the principal curvatures, and u, v are the principal directions.

As equation (4) indicates, the vector $d\mathbf{N}$ points parallel to $\mathbf{p}_u$ when measured in the direction $(du, 0)$, and parallel to $\mathbf{p}_v$ when measured in the direction $(0, dv)$. However, as figures 1-3 and 1-4 indicate, the vector $d\mathbf{N}$ is not in general parallel to $d\mathbf{p}$ along any other direction. Thus $d\mathbf{N}$ measured along direction (du, dv) does not in general point along the direction (du, dv).

dI in terms of surface properties. Having derived an expression for $d\mathbf{N}$ in terms of surface properties, we may now derive an expression for dI in terms of surface properties. Using equations (2) and (4) we find dI taken along (du, dv) is

$$\begin{aligned} dI &= \rho\lambda d\mathbf{N} \cdot \mathbf{L} \\ &= \rho\lambda(-\kappa_1 \mathbf{p}_u du - \kappa_2 \mathbf{p}_v dv) \cdot \mathbf{L} \\ &= \rho\lambda(-\kappa_1(\mathbf{p}_u \cdot \mathbf{L})du - \kappa_2(\mathbf{p}_v \cdot \mathbf{L})dv) \end{aligned} \tag{5}$$

We may rewrite equation (5) in a more convenient matrix notation as follows:

$$dI = \rho\lambda\begin{pmatrix} -\kappa_1(\mathbf{p}_u \cdot \mathbf{L}) & -\kappa_2(\mathbf{p}_v \cdot \mathbf{L}) \end{pmatrix}\begin{pmatrix} du \\ dv \end{pmatrix} \tag{6}$$

This equation expresses the derivative of image intensity in a co-ordinate system which is tangent to the surface and which is aligned with the directions of the principal curvatures.

In order to be generally useful, we must obtain an expression for dI in the image plane co-ordinate system (x, y). Thus we require a projection matrix Φ which maps the image plane co-ordinate system (dx, dy) into the tangent plane co-ordinate system (du, dv):

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix}$$

This projection matrix is actually the Jacobian of the transformation from the (x, y) co-ordinate system to the (u, v) co-ordinate system. Given Φ we may write dI in the image co-ordinate system (x, y) as

$$dI = \rho\lambda\begin{pmatrix} -\kappa_1(\mathbf{p}_u \cdot \mathbf{L}) & -\kappa_2(\mathbf{p}_v \cdot \mathbf{L}) \end{pmatrix}\left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix}\right)$$

In order to express dI in the image plane co-ordinate system we must first rotate the co-ordinate system in the (u, v) plane in order to line up with the tilt of the surface, project the co-ordinate

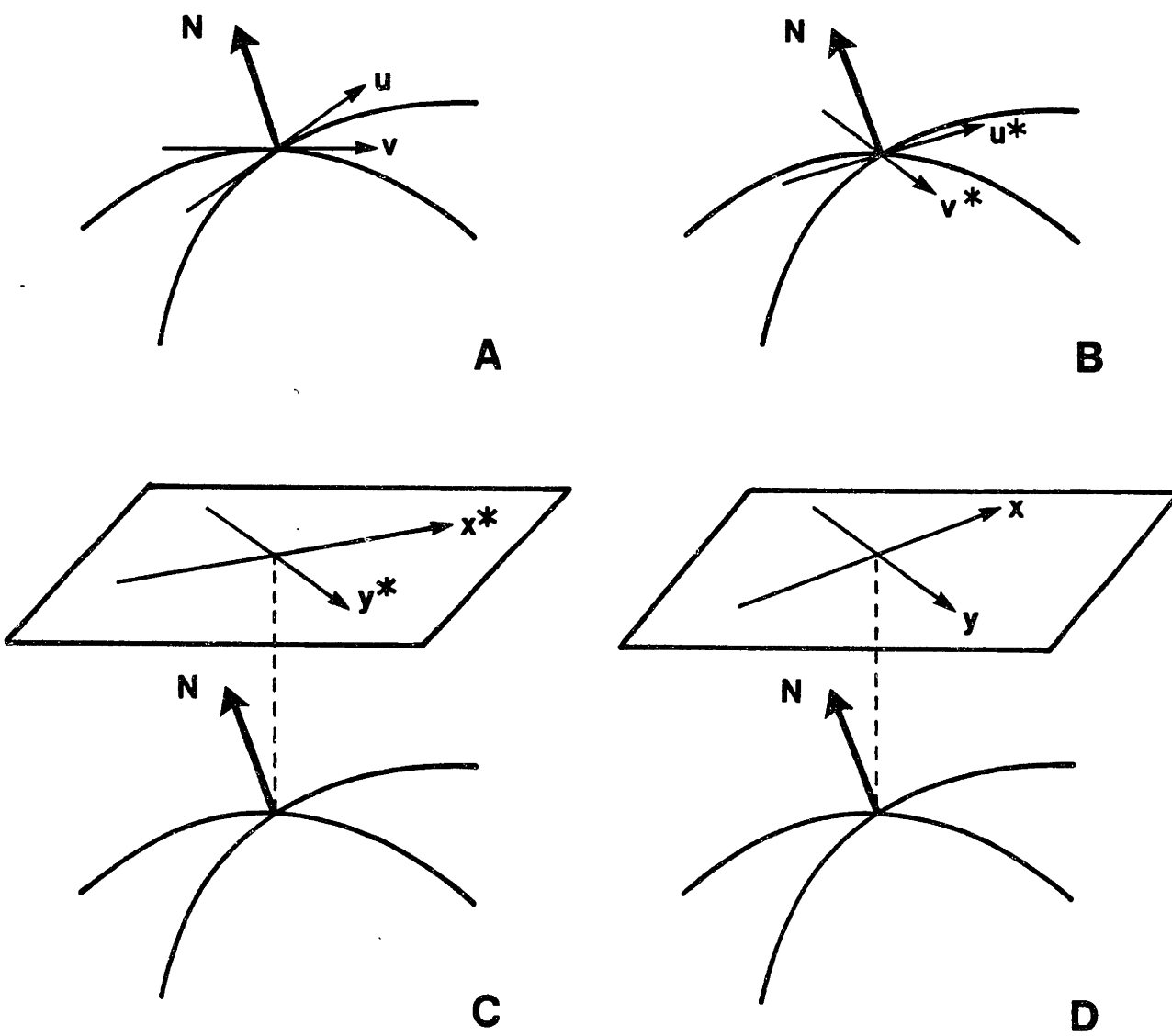


Figure 1-5. Transformation from (A) a co-ordinate patch oriented along the principal curvature directions and the surface normal **N**, to (B) a system (u^*, v^*) still in the tangent plane (u, v) but oriented along the tilt direction of the tangent plane, to (C) a system (x^*, y^*) in the image plane aligned with the tilt direction of the tangent plane, to (D) the (x, y) co-ordinate system.

system onto the image plane, and finally rotate the projected co-ordinate system to match the (x, y) co-ordinates. These successive transformations are shown in figure 1-5.

To obtain an expression for the differential step (du, dv) in terms of the differential step (du^*, dv^*) in the co-ordinate system (u^*, v^*) which is rotated by an angle θ_1 in the tangent plane we simply multiply by a rotation matrix:

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} du^* \\ dv^* \end{pmatrix}$$

In order to account for the projection into the image plane we must first line our co-ordinate system up with the projection of the tilt of the tangent plane (i.e., with the projection onto the tangent plane of an image-plane vector which points in the direction of maximum descent in the tangent plane), which may be done by letting θ_1 above be the angle between the projection of the tangent plane tilt and the direction of the first principal curvature.

Along the surface tilt direction there is a foreshortening effect which is proportional to z_N^{-1} where z_N is the z component of $\mathbf{N}$, the surface normal (i.e., $\mathbf{N} = (x_N, y_N, z_N)$). Along the direction orthogonal to the surface tilt there is no foreshortening effect. We can then account for the forshortening due to projection as follows:

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} z_N^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dx^* \\ dy^* \end{pmatrix}$$

This then leaves us able to express (du, dv) in terms of a differential step (dx^*, dy^*) in a co-ordinate system (x^*, y^*) in the image plane, but with the x^* direction oriented along the tilt direction of the surface. Thus to obtain an expression for (du, dv) which is in the (x, y) co-ordinate system we must rotate the axes by θ_2, the angle between the x^* and the x axes. We can do this by multiplying by a second rotation matrix as follows:

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} z_N^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ -\sin\theta_2 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

This then is the desired projection matrix, Φ.

We may use the projection matrix Φ to give us an expression for dI in the image-plane co-ordinate system by taking equation (6), which expresses dI in the (u, v) co-ordinate system, and multiplying by the projection matrix Φ:

$$dI = \rho\lambda \begin{pmatrix} -\kappa_1(\mathbf{p}_u \cdot \mathbf{L}) & -\kappa_2(\mathbf{p}_v \cdot \mathbf{L}) \end{pmatrix} \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \tag{7}$$

where

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} z_N^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ -\sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

and where

θ_1 = angle between the projection of the surface tilt and the direction of the first principal curvature

z_N = the z component of the surface normal $\mathbf{N} = (x_N, y_N, z_N)$

θ_2 = the tilt of the surface

2.3 The second derivative of image intensity

If, as with the first derivative, we assume that the albedo and illuminant are locally constant then the second derivative of image intensity will be

$$\begin{aligned} d^2I &= d(\rho\lambda d\mathbf{N} \cdot \mathbf{L}) \\ &= \rho\lambda d^2\mathbf{N} \cdot \mathbf{L} + \rho\lambda d\mathbf{N} \cdot d\mathbf{L} \\ &= \rho\lambda d^2\mathbf{N} \cdot \mathbf{L} \end{aligned}$$

Thus the second derivative of image intensity depends upon the second derivative of the surface normal, just as the first derivative depended upon the first derivative of the surface normal.

$d^2\mathbf{N}$ in terms of surface properties. As with the first derivative of the surface normal, we want to express the second derivative of the surface normal, $d^2\mathbf{N}$, in terms of surface curvature and surface orientation. In the co-ordinate system (u, v), the first derivative of the surface normal in direction (du, dv) was

$$d\mathbf{N} = -\kappa_1\mathbf{p}_u du - \kappa_2\mathbf{p}_v dv$$

Therefore the second derivative of the surface normal in direction (du, dv) is

$$\begin{aligned} d^2\mathbf{N} = &-(\frac{d\kappa_1}{du}\mathbf{p}_u + \kappa_1\mathbf{p}_{uu})du^2 - (\frac{d\kappa_2}{du}\mathbf{p}_v + \kappa_2\mathbf{p}_{vu})dudv \\ &-(\frac{d\kappa_2}{dv}\mathbf{p}_v + \kappa_2\mathbf{p}_{vv})dv^2 - (\frac{d\kappa_1}{dv}\mathbf{p}_u + \kappa_1\mathbf{p}_{uv})dvdu \end{aligned} \tag{8}$$

This equation is too complex to be tractable. However considerable simplification is possible without incurring significant error.

Equation (3) gives us that in the (u, v) co-ordinate system the principal curvatures κ_1 and κ_2 are

$$\kappa_1 = \mathbf{p}_{uu} \cdot \mathbf{N} \qquad \kappa_2 = \mathbf{p}_{vv} \cdot \mathbf{N}$$

Thus

$$\left|\frac{d\kappa_1}{du}\right| \leq \|\mathbf{p}_{uuu}\| \quad \left|\frac{d\kappa_2}{du}\right| \leq \|\mathbf{p}_{vvu}\| \quad \left|\frac{d\kappa_2}{dv}\right| \leq \|\mathbf{p}_{vvv}\| \quad \left|\frac{d\kappa_1}{dv}\right| \leq \|\mathbf{p}_{uuv}\|$$

that is, the changes in curvature are less than or equal to the third order derivatives. Thus if we assume that the third derivatives of $\mathbf{p}$ are zero, then the changes in curvature are also zero. If we further assume that the surface has no "twist" in it, i.e., that $\|\mathbf{p}_{uv}\| = 0$, then we obtain the desired simplification of equation (8):

$$d^2\mathbf{N} = -\kappa_1\mathbf{p}_{uu}du^2 - \kappa_2\mathbf{p}_{vv}dv^2 \tag{9}$$

By assuming that the surface we are observing has zero third derivatives and no "twist", we are approximating the observed surface by a second-order surface, that is, a surface which may be described by the Monge patch $\mathbf{p} = u\mathbf{e}_1 + v\mathbf{e}_2 + f(u, v)\mathbf{e}_3$ where $f(u, v)$ is of the form $a_1u^2 + a_2v^2 + a_3u + a_4v + a_5$. The approximation of an arbitrary surface by such a surface typically causes errors on the order of δ^3 where δ is the spacing between observed points on the surface[7] (Isaacson and Keller, 1966). Thus we may assume that we are observing such a second-order surface without incurring undue error.

[7]For surfaces with $\|\mathbf{p}_{uv}\|$ large the errors may become $O(\delta^2)$, they cannot become larger. Thus if ten points are observed across a surface, the maximum error incurred by this approximation is on the order of $1/100^{th}$ of the width of the surface. This occurs only for surfaces like those in figure 1-6 (e) or (f); for all other surfaces having ten observations would limit the maximum error to be on the order of $1/1000^{th}$ of the width of the surface.

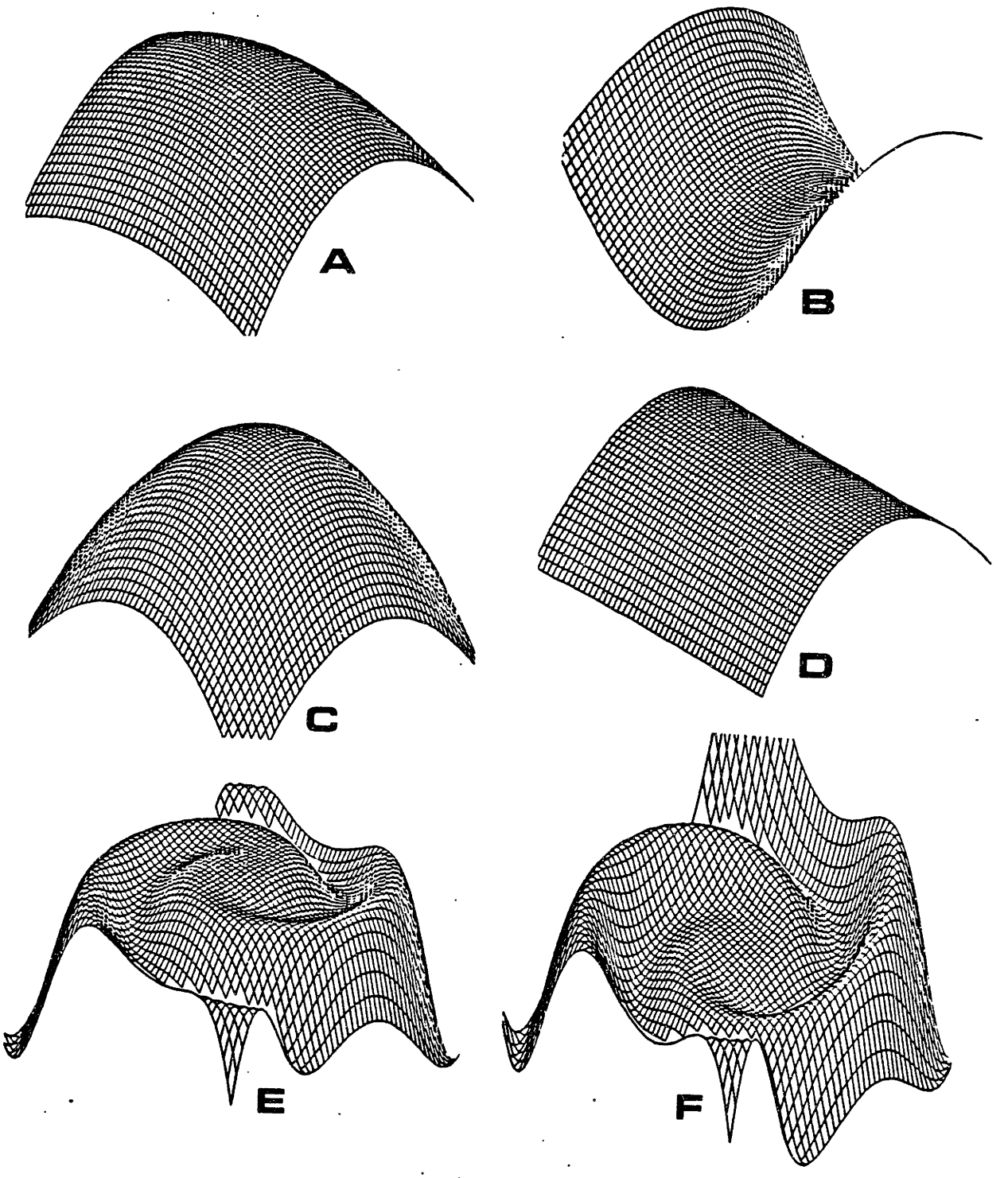


Figure 1-6. Surfaces (a) through (d) have zero third derivatives, and also have no "twist" in them, i.e., $\|\mathbf{p}_{uv}\| = 0$. Surfaces (e) and (f) have "twist" in them, i.e., $\|\mathbf{p}_{uv}\| \neq 0$.

Figure 1-6 illustrates the implications of this assumption. This figure shows four surfaces, (a) through (d), which are second-order surfaces with no cross-terms. These surfaces therefore have zero third derivatives, and do not "twist", i.e., $\|\mathbf{p}_{uv}\| = 0$. In contrast, surfaces (c) and (f) have significant "twist", i.e., the cross-derivative $\mathbf{p}_{uv}$ is non-zero. By assuming that the surface is second order, we in effect approximate surfaces such as (c) or (f) using small patches of surface types (a) — (d). In fact, the surfaces in figure 1-6 (c) and (f) are actually shown as composed of small patches of surface types (a) — (d), demonstrating the effectiveness of such approximation. These figures make it clear that the assumption that a surface is second-order is not an overly restrictive assumption; especially as the assumption typically introduces only $\mathcal{O}(\delta^3)$ error, and gives us considerable simplification of the expression for $d^2\mathbf{N}$.

If we are observing such a second-order surface, then in the co-ordinate system (u, v) the second derivative of the surface normal $d^2\mathbf{N}$ is perpendicular to the first derivative $d\mathbf{N}$, just as the first derivative is perpendicular to the surface normal. Further, under this assumption $d^2\mathbf{N}$ points in the same direction as the surface normal, as shown in figure 1-7. This claim is proven in the appendix. The fact that $d^2\mathbf{N}$ is parallel to $\mathbf{N}$ is important because it shows that the effects of the illuminant are identical upon all three components of d^2I, i.e., I_{uu}, I_{vv} and I_{uv}, and the surface normal.

d^2I in terms of surface properties. Using equation (9) then

$$\begin{aligned} d^2I &= \rho\lambda d^2\mathbf{N} \cdot \mathbf{L} \\ &= \rho\lambda(-\kappa_1\mathbf{p}_{uu}du^2 - \kappa_2\mathbf{p}_{vv}dv^2) \cdot \mathbf{L} \\ &= \rho\lambda(-\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L})du^2 - \kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L})dv^2) \end{aligned}$$

We might alternatiely choose to express equation (9) using the Hessian:

$$\begin{pmatrix} du \\ dv \end{pmatrix}^T \begin{pmatrix} I_{uu} & I_{uv} \\ I_{vu} & I_{vv} \end{pmatrix} \begin{pmatrix} du \\ dv \end{pmatrix} = \rho\lambda \begin{pmatrix} du \\ dv \end{pmatrix}^T \begin{pmatrix} -\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L}) & 0 \\ 0 & -\kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L}) \end{pmatrix} \begin{pmatrix} du \\ dv \end{pmatrix}$$

where the superscript "T" denotes the transpose of the matrix.

We can then use the projection matrix Φ to obtain the Hessian in the image co-ordinate system

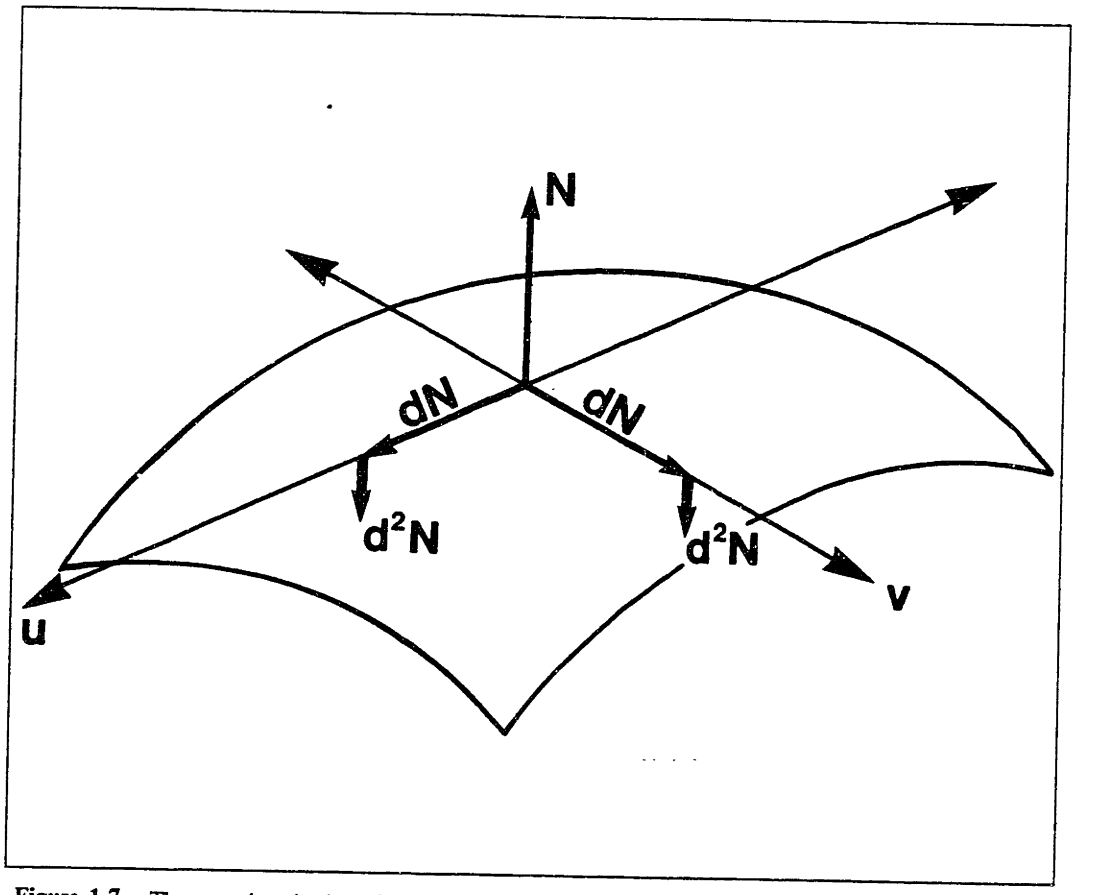


Figure 1-7. The second derivative of the surface normal points in the same direction as the surface normal if the surface has zero third derivatives.

(x, y)

$$\begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} I_{xx} & I_{yx} \\ I_{xy} & I_{yy} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} =$$
$$\rho\lambda \left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \right)^T \begin{pmatrix} -\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L}) & 0 \\ 0 & -\kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L}) \end{pmatrix} \left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \right)$$

For our purposes it will be more useful to have an expression for dI rather than the full Hessian. We may adopt a notation similar to that used for dI to express d^2I in the (u, v) co-ordinate system:

$$d^2I = \rho\lambda \left(\begin{pmatrix} -\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L}) & 0 \\ 0 & -\kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L}) \end{pmatrix} \begin{pmatrix} du \\ dv \end{pmatrix} \right)^T \begin{pmatrix} du \\ dv \end{pmatrix}$$

and then use Φ to express d^2I in the (x, y) co-ordinate system

$$d^2I = \rho\lambda\left(\begin{pmatrix} -\kappa_1(\mathbf{p}_{uu}\cdot\mathbf{L}) & 0 \\ 0 & -\kappa_2(\mathbf{p}_{vv}\cdot\mathbf{L}) \end{pmatrix}\left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix}\right)\right)^T\left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix}\right) \tag{10}$$

This then gives us an expression for d^2I in terms of surface properties, under the assumption that we are observing a second-order surface.

2.4 Generality Of The Assumptions

In developing an equation for the derivative of image intensity we assumed that the surface had a Lambertian reflectance function, and that the surface was illuminated by a distant point-source. These are strong assumptions when applied globally, however they are *much* weaker assumptions when they are only interpreted locally. This section addresses the strength of these local assumptions, that is, how generally true these local assumptions may be expected to be in real scenes. Proofs of the following proposition is given in appendix.

Proposition (Generality Of A Point-Source Illuminant). **Given that a surface has a Lambertian reflectance function, any constant distribution of illumination is equivalent to a distant point source.**

The caveat of a "constant" distribution of illumination means that all of the surface under discussion must be illuminated by the same distribution of light — i.e., no "self-shadowing". When applied globally, the requirement no self-shadowing means that if the surface has a wide range of surface orientations, then it is impossible to find an point-source which is equivalent to any relatively broad distribution of illumination. When applied to a single point, the requirement that there be no self-shadowing only rules out distributions of illumination where a large amount of the incident light is nearly parallel to the points' tangent plane, e.g., points along the terminator. Thus given a Lambertian reflectance function the assumption of a distant point-source introduces little error as long as we are dealing with a single image point.

Generality Of Lambertian Reflectance Function. The following "general" expression for image intensity

$$I = \rho f k(\mathbf{N} \cdot \mathbf{L}) + \rho f(1 - k)\left[\frac{(2\mathbf{N} - \mathbf{V})}{\|2\mathbf{N} - \mathbf{V}\|} \cdot \mathbf{L}\right]^q \tag{11}$$

was suggested by Horn 1977 as a first approximation to most natural reflectance functions.[8] This function contains a matte or Lambertian term (the first term) and a specular or mirror-reflection term (the second term) in the proportion k to $1 - k$. The sharpness of the specularity term (how nearly mirror-like the term is) is given by the power q. Using this function we can obtain reflectance functions which range from purely Lambertian to purely specular or mirror-like.

For most values of $\mathbf{N}$, k and q equation (11) may be reasonably approximated by the Lambertian reflectance function with an appropriate selection of $\mathbf{L}$, because for most values of k, q and $\mathbf{N}$ the specular term is small. Even for the most specular, shiny surfaces (those with low k and high q values), the function is roughly Lambertian over most of the range of $\mathbf{N}$. It is only near those values of $\mathbf{N}$ near where $\mathbf{L}$ is parallel to $2\mathbf{N} - \mathbf{V}$, that is, near where highlights and specular reflection are occurring, that the reflectance function is a quickly changing function of $\mathbf{N}$ and thus cannot be locally approximated by the Lambertian reflectance function.

[8]See Torrance, Sparrow and Birkebak 1966, Torrance and Sparrow 1967.

Chapter 2

A Theory For Determining Shape

1 Introduction

We now have equations for the image intensity, and its first and second derivatives. These were

$$\begin{aligned} I &= \rho\lambda \mathbf{N} \cdot \mathbf{L} \\ dI &= \rho\lambda d\mathbf{N} \cdot \mathbf{L} \\ d^2I &= \rho\lambda d^2\mathbf{N} \cdot \mathbf{L} \end{aligned} \tag{12}$$

The problem of estimating surface shape from local information is, essentially, the problem of determining the unknown surface normal, $\mathbf{N}$, from image measurements of I, dI or d^2I. Solving for the unknowns in any system of equations requires having more measurements than unknowns. As mentioned before, $\mathbf{N}$ has two degrees of freedom (for instance, slant and tilt), and so we require least two measurements at each image point in order to solve for $\mathbf{N}$ locally. A further obstacle to determining surface orientation is that the unknown illuminant direction, $\mathbf{L}$, also appears in these equations, adding another two degrees of freedom to the equations.

This means that we cannot determine the surface normal using image intensity alone, as it provides only one measurement per image point. More measurements per point can be obtained from the first or second derivatives of image intensity, which provide respectively two and three measurements per point[1] However both $d\mathbf{N}$ and $d^2\mathbf{N}$ depend not only upon $\mathbf{N}$ but upon the curvature of the surface and how that curvature is oriented, adding another three degrees of freedom. Thus using dI or d^2I to obtain more measurements does not lead directly to a solution.

We cannot obtain a solution for $\mathbf{N}$ simply by using higher derivatives to obtain yet more measurements per point. Although each additional derivative does give more measurements per point, each

[1] dI is completely characterized by its partials with respect to x and y, I_x and I_y, and d^2I is completely characterized by the partials I_{xx}, I_{yy}, and I_{xy}.

additional derivative also brings an even larger number of unknowns into the equations. Thus in order to solve for **N** locally we must either have additional information about the imaged surface or make some simplifying assumptions, so that the number of unknowns is not larger that the number of measurements available to us.

1.1 Potential For Use Of dI And d^2I

Let us look in more detail at two candidates for use in the local estimation of surface shape: dI and d^2I. Equation (12) shows that these depend on $d\mathbf{N}$ and $d^2\mathbf{N}$, respectively, and their relationship to the illumination.

As shown previously in figure 1-3, the vector $d\mathbf{N}$ points in different directions when measured along different image directions (dx, dy) -- and the direction in which $d\mathbf{N}$ points is not necessarily (dx, dy). The fact that $d\mathbf{N}$ points in a different (unknown) direction for each image direction along which it is measured means that each measurement of dI is a function of two unknown and confounded factors: the magnitude of $d\mathbf{N}$ and the direction of $d\mathbf{N}$ relative to **L**. The dependence of dI upon **L** makes dI useful for finding the illuminant direction and using it to identify surface type, as will be done in the third section of this chapter. However this dependence upon **L** makes the use of dI to estimate surface shape critically dependent upon knowing the exact illuminant direction. Thus it is difficult to obtain accurate estimates of shape using measurements of dI.

By contrast, $d^2\mathbf{N}$ always points in the same direction as the surface normal (see figure 1-7), given the assumption that the surface has zero third derivatives (this is proven in the appendix). This means that the effects of the illuminant are identical upon all three components of d^2I, as well as the surface normal itself. This gives us considerable potential to discount the effects of the illuminant and makes d^2I especially useful for estimating surface shape, as will be shown in the fourth section of this chapter.

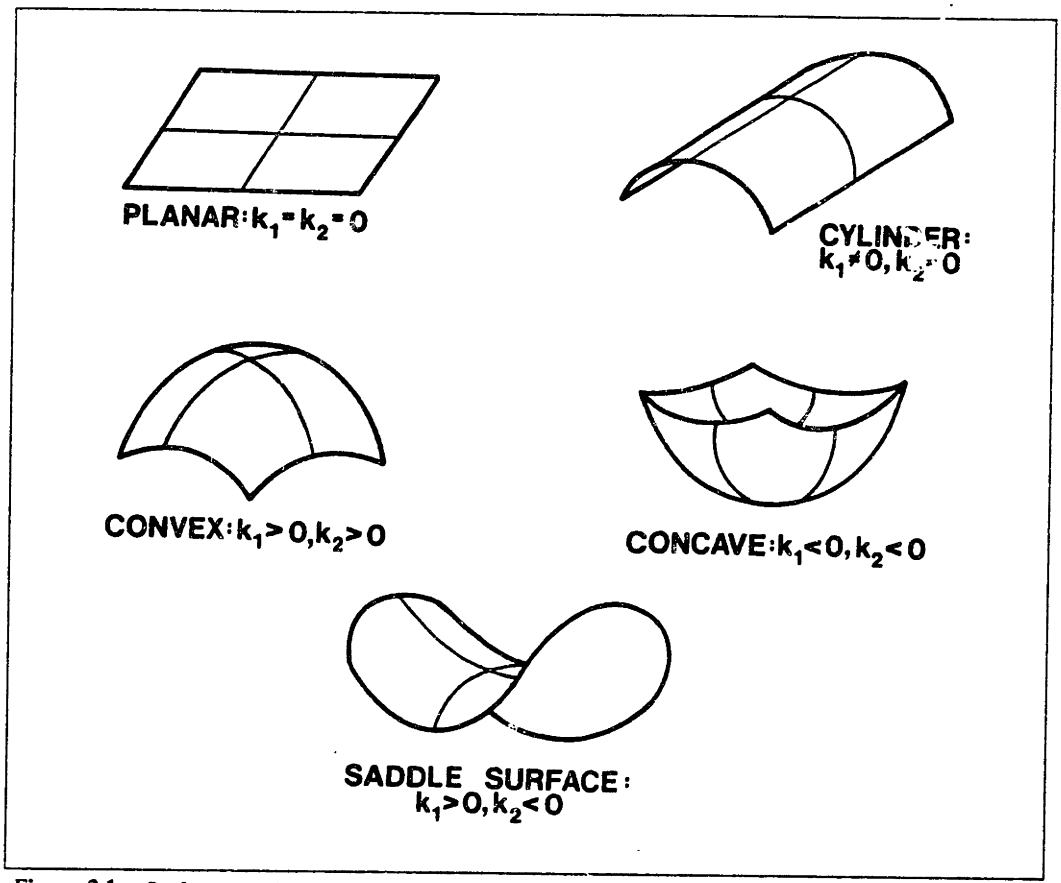


Figure 2-1. Surfaces may be classified into five types: planar, cylindrical, convex, concave, or saddle surface. The classification of a surface depends on whether the two principal curvatures κ_1 and κ_2 are positive, negative, or zero.

1.2 Direct Inference About The Surface

Measurements of d^2I are in a sense direct measurements of the surface — of $d^2\mathbf{N}$. These measurements are relative to the illuminant, in that they are scaled by the cosine of the angle between them and the illuminant, and thus their magnitude is as much a result of the illuminant direction as of the surface. However if d^2I is zero, then it turns out that $d^2\mathbf{N}$ must also be zero[2] , and vise versa. Knowing when $d^2\mathbf{N}$ is zero helps to characterize the surface.

Surfaces may be classified into five types: planar, cylindrical, convex, concave, or saddle surface. These five types are shown in figure 2-1. The classification of a surface depends on whether the two

[2]Because $d^2\mathbf{N}$ is always nearly parallel to $\mathbf{N}$ (see appendix), the $d^2\mathbf{N}$ will not be perpendicular to $\mathbf{L}$ except at or near where $\mathbf{N} \cdot \mathbf{L} = 0$, i.e., at or near terminators in the complete absence of diffuse illumination. In the presence of an atmosphere this practically cannot happen; there will always be diffuse illumination, so that $\mathbf{N} \cdot \mathbf{L} \neq 0$ and thus $d^2\mathbf{N} \cdot \mathbf{L} \neq 0$ unless $\|d^2\mathbf{N}\| = 0$.

principal curvatures κ_1 and κ_2 are positive, negative, or zero:

	plane	$\kappa_1 = 0$	$\kappa_2 = 0$
	cylinder	$\kappa_1 > 0$	$\kappa_2 = 0$
	convex	$\kappa_1 < 0$	$\kappa_2 < 0$
	concave	$\kappa_1 > 0$	$\kappa_2 > 0$
saddle	*surface*	$\kappa_1 > 0$	$\kappa_2 < 0$

We can use equation (9) to show how d^2I relates to the principal curvatures (these conclusions are not dependant upon the assumptions of equation (9), however use of equation (9) allows a clearer exposition). Equation (9) describes d^2I given that we are looking straight down upon the surface:

$$d^2I = \rho\lambda(-\kappa_1 \mathbf{p}_{uu} \cdot \mathbf{L} du^2 - \kappa_2 \mathbf{p}_{vv} \cdot \mathbf{L} dv^2) \tag{9}$$

Because the third derivatives of the surface were assumed to be zero, then $\mathbf{p}_{uu}$, $\mathbf{p}_{vv}$ and $\mathbf{N}$ are parallel, and so $\mathbf{p}_{uu} = \kappa_1 \mathbf{N}$, $\mathbf{p}_{vv} = \kappa_2 \mathbf{N}$. We may then write equation (9) as

$$d^2I = \rho\lambda(-\kappa_1^2 \mathbf{N} \cdot \mathbf{L} du^2 - \kappa_2^2 \mathbf{N} \cdot \mathbf{L} dv^2)$$

From this equation we see that $d^2I = 0$ only if (1) either $\kappa_1 = 0$ or $\kappa_2 = 0$, or (2) both $\kappa_1 = 0$ and $\kappa_2 = 0$[3].

In case one, where either κ_1 or κ_2 are zero (a cylindric surface), then d^2I is zero in only one direction[4] In case two, where both κ_1 and κ_2 are zero (a planar surface) d^2I is zero in all directions. The fact that $d^2I \neq 0$ in all directions for convex, concave and saddle surfaces is a consequence of shading's inherent convexity/concavity ambiguity, discussed in the following section. As the projection matrix Φ cannot alter the number of zero directions, we have

Proposition (Surface Type). The number of directions in which $d^2I = 0$ determines the surface type, up to reflections about the image plane:

$d^2I = 0$ in no directions $\Rightarrow$ *convex/concave/saddle surface*
$d^2I = 0$ in one direction $\Rightarrow$ *cylinder*
$d^2I = 0$ in all directions $\Rightarrow$ *plane*

[3]The case where $\mathbf{N} \cdot \mathbf{L} = 0$ can be ignored, as explained in the last footnote.

[4]Stevens 1981 has noted that linear highlights imply a cylindric surface, this is a special case of this more general result.

Thus if we observe that $d^2I = 0$ along a line in the image, then we know that the surface has a zero principal curvature along that line and thus that the surface is cylindrical along that line. Similarly, if we observe that $d^2I = 0$ along a direction (dx, dy) throughout some region in the image, then the surface is a cylinder with an axis pointing in the (dx, dy) direction.

The detection of lines along which a surface is cylindrical is of considerable importance because it is only at such cylindrical lines that changes in the surface type (e.g., convex, concave, saddle surface) occur. As the surface changes from one type to a different surface type the sign of at least one of the principal curvatures changes from positive to negative, or vise versa. In changing sign the curvature is briefly zero, and so the surface is cylindrical along the locus where the surface changes type (if the change takes place over an extended area, both curvatures will be zero and so the surface will be planar instead of cylindrical). Thus lines along which $d^2I = 0$ are places where the surface is changing type. The set of such lines divides the surface up into regions with the same sign of curvature.

2 Fundamental Limitations Of Shading

Local analysis of shading has several intrinsic limitations for use in the estimation of surface shape (see also Bruss 1981). The first fundamental limitation of shading is apparent as early as equation (5):

$$dI = \rho\lambda(-\kappa_1 \mathbf{p}_u \cdot \mathbf{L}du - \kappa_2 \mathbf{p}_v \cdot \mathbf{L}dv) \tag{5}$$

where as before $\mathbf{L}$ is the illuminant direction, $\mathbf{p}_u$ and $\mathbf{p}_v$ are the partials of the surface vector $\mathbf{p} = u\mathbf{e}_1 + v\mathbf{e}_2 + f(u,v)\mathbf{e}_3$ with respect to u and v, and κ_1, κ_2 are the principal curvatures. Because κ_1 and κ_2 only appear in conjunction with the terms $\mathbf{p}_u \cdot \mathbf{L}$ and $\mathbf{p}_v \cdot \mathbf{L}$ (respectively), we see that for purposes of estimation the signs of κ_1 and κ_2 are hopelessly confounded with the signs of $\mathbf{p}_u \cdot \mathbf{L}$ and $\mathbf{p}_v \cdot \mathbf{L}$. From this observation comes our first proposition about the limitations of a local analysis of shading:

Proposition (Convexity/Concavity Ambiguity). In any analysis of local image shading, there will always remain ambiguity about the sign of both the principal surface curvatures κ_1 and κ_2

This ambiguity is familiar as the "crater illusion" — pictures of lunar craters can look like bumps rather than depressions if we imagine the illuminant to be coming from the bottom of the picture rather than from the top. At this point in the analysis it is not clear whether this ambiguity may be limited to only one, overall ambiguity for a surface, or it may exist at each point. Later it will be shown that the ambiguity is of a more global nature, given certain continuity assumptions. This is consistent with human perception of surfaces — if we "switch" one part of the picture from convex to concave, all of the figure switches.

There another limitation to the use of local shading information for the estimation of surface shape, and this relates to the perception of the relief in an image. The common example of this limitation is *bas relief* sculpture, where the actual relief of the surface is much less than the three-dimensional object being portrayed. Despite the small amount of relief the correct impression of shape may be achieved. This is explained more exactly by the following proposition:

Proposition (Intrinsic Underdetermination Of Surface By Shading). **Foreshortening, caused by the surface slanting away from or toward the viewer, and surface curvature have identical effects on the first derivative of image intensity; either can reinforce or cancel the effects of the other.**

This proposition illustrates the trade-off of forshortening and curvature in determining the first derivative of image intensity. We may illustrate this ambiguity by again examining equation (5), which gives an expression for dI. The partials of I, which are I_u and I_v, are dI along the directions $(du, 0)$ and $(0, dv)$, thus from equation (5):

$$I_u = -\rho\lambda\kappa_1 \mathbf{p}_u \cdot \mathbf{L} \qquad I_v = -\rho\lambda\kappa_2 \mathbf{p}_v \cdot \mathbf{L}$$

The two numbers I_u and I_v completely specify the first derivative of image intensity at a point. It is clear that we may obtain any I_u and I_v by varying either $\mathbf{p}_u$ and $\mathbf{p}_v$, which describe the surface orientation and thus the foreshortening, or by varying the surface curvatures κ_1 and κ_2, or by any combination of $\mathbf{p}_u$, $\mathbf{p}_v$ and κ_1 and κ_2. Thus the proposition.

This demonstrates the equivalence of the effects of foreshortening and curvature, and shows that at an arbitrary point the image effects of curvature (κ_1, κ_2) can modify, and even cancel, the effects foreshortening.

Equation (9), the expression for d^2I which is analogous to equation (5), confounds the factors of foreshortening and surface curvature in a manner similar to that of equation (5) (see the proof of the following proposition). Thus these fundamental limitations also exist for the second derivative of image intensity, and in fact new limitations in the use of a local analysis of shading arise from the greater number of factors contributing to the second derivative. One such limitation is

Proposition (Sufficiency). **The second derivative of any pattern of image intensities can be accounted for by a second-order surface.**

This proposition (proved in the appendix) provides further justification for assuming that the viewed surface is second-order, the assumption used in developing the expression for d^2I, equation (10). This proposition shows that equation (10) allows us to describe all possible local image intensity

configurations; even if we were to view a non-second-order surface, there would be no way to locally distinguish it from a second-order surface on the basis of shading information.

3 Constraint From Illuminant Direction

The first proposition in the last section shows that additional information, such as the illuminant direction, is always required in order to determine the convexity or concavity of a surface. For instance, to determining the convexity/concavity of the central "bump" in the last two images of figure 0-1 *requires* knowledge of the illuminant direction. In this section we will first address the problem of determining the illuminant direction, and then see how the illuminant direction may be used to complete the determination of surface type, i.e., to determine whether the surface is convex or concave.

3.1 Finding The Illuminant Direction

In biological vision **L** is not known *a priori*, and so must be estimated from the image data itself. This is difficult because, as shown in equations (12), image data are largely determined by the angle between the surface normal and the illuminant direction. Because the evidence about illuminant direction is confounded with the unknown direction of the surface normal, estimating the illuminant requires making some assumption about the surface orientation (or its derivatives).

It suffices to estimate the illuminant direction for regions of the image, rather than at each individual point, because the illuminant direction typically varies slowly. Generally the illuminant is distant, and so the light rays are nearly parallel and thus the illuminant direction nearly constant within a small region. The most common nearby "source" of light comes from light reflected off of one surface onto another. However, since the average albedo of natural surfaces is only 0.2, and rarely goes above 0.5 (see Kirnov 1971, Richards 1981) the amount of light contributed by a nearby surface is usually *much* less than the amount of light falling directly onto the surface. Thus the illuminant direction is usually little affected by reflected light, and so may be reasonably treated as constant over small regions of the image.

Having an accurate guess about the mean of $d\mathbf{N}$ within a region of the image allows us to solve for the illuminant direction within that region. Using $d\bar{I}$, the mean value of dI measured along a

particular image direction, together with $d\bar{\mathbf{N}}$, our guess of the mean of $d\mathbf{N}$ along that direction, we can solve for $\mathbf{L}$ using the relation $d\bar{I} = d\bar{\mathbf{N}} \cdot \mathbf{L}$. The accuracy of this solution for $\mathbf{L}$ depends upon how close the value assumed for the mean of $d\mathbf{N}$ is to the actual mean, and upon how constant the illuminant direction is within the region.

One good guess about the mean of $d\mathbf{N}$ within a region comes from asserting that the sum of all $d\mathbf{N}$ over all orientations within a region is zero. This is equivalent to asserting that change in surface orientation is isotropically distributed within the image region. It is true that $d\mathbf{N}$ is isotropically distributed when considered over all scenes, and thus that the sum of $d\mathbf{N}$ over all scenes is zero. In addition to the statistical mean being zero, there is a large class of common objects for which the sum of all $d\mathbf{N}$ is exactly zero. This class of objects includes all images of convex objects bounded entirely by a gradual occluding contour[5], such as the image of any smooth pebble.

Given the assumption that changes in surface orientation are isotropically distributed, so that the sum of all $d\mathbf{N}$ is zero, we can devise a procedure for estimating the illuminant direction $\mathbf{L}$ by looking for the regular biasing effect of the illuminant direction on $d\bar{I}$, the mean value of dI, along various image directions (dx, dy). The effect of the illuminant direction is to make $d\bar{I}$, vary according to

$$\begin{aligned} d\bar{I} &= \rho\lambda d\bar{\mathbf{N}} \cdot \mathbf{L} \\ &= \rho\lambda(d\bar{x}_N x_L + d\bar{y}_N y_L + d\bar{z}_N z_L) \end{aligned}$$

where $d\bar{\mathbf{N}} = (d\bar{x}_N, d\bar{y}_N, d\bar{z}_N)$ and $\mathbf{L} = (x_L, y_L, z_L)$ is the illuminant direction. Under the assumption that within a region

$$\sum_{x,y,\theta} d\bar{\mathbf{N}} = \mathbf{0}$$

i.e., that the sum of all $d\mathbf{N}$ along all directions is the zero vector, then along any *one* image direction $d\bar{x}_N$ is proportional to the x- component of the image direction, dx, and $d\bar{y}_N$ is proportional to y-component of the image direction, dy, while $\bar{dz}_N$ is zero. Therefore

$$d\bar{I} = k(d\bar{x}_N x_L + d\bar{y}_N y_L)$$

[5]This may be proved by noting that the surface normals on such an object are perpendicular to $\mathbf{V}$ at the boundary of such an object, and thus (given that the object is strictly convex) we may form a 1-1 onto map between the surface normals of the object and the Gaussian sphere, which has mean zero.

where k is a constant determined by the albedo, illuminant strength and the variance of the distribution of $d\mathbf{N}$ within the region.

By using this relation, we can set up a linear regression using the mean of dI as measured along various image directions to solve for the ratio of the unknowns x_L and y_L. The constant k, and from this the exact values of x_L,y_L and z_L, can be estimated from the mean and variance of the distribution of dI along any one image direction, as specified in the following proposition, proved in the appendix:

Proposition (Illuminant Direction) Under the assumption that change in surface normal is isotropically distributed, the following regression gives a maximum-likelihood estimate of τ_L, the tilt of the illuminant direction within a region:

$$\begin{pmatrix} \hat{x}_L \\ \hat{y}_L \end{pmatrix} = \left[\beta^T \beta\right]^{-1} \beta^T \begin{pmatrix} d\bar{I}_1 \\ d\bar{I}_2 \\ \vdots \\ d\bar{I}_n \end{pmatrix} \tag{13}$$

where $d\bar{I}_i$ are the means of dI over the region along each of n image directions (dx_i, dy_i), β is a $2 \times n$ matrix of directions (dx_i, dy_i) and β^T indicates the transpose of β. The illuminant tilt, τ_L is given by

$$\tau_L = \tan^{-1}\left(\frac{\hat{y}_L}{\hat{x}_L}\right)$$

and the full illuminant direction $\mathbf{L} = (x_L, y_L, z_L)$ is estimated by

$$x_L = \frac{\hat{x}_L}{k} \qquad y_L = \frac{\hat{y}_L}{k} \qquad z_L = \sqrt{1 - \frac{(\hat{x}_L^2 + \hat{y}_L^2)}{k^2}}$$

where

$$k = \sqrt{E(dI^2) - E(dI)^2}$$

Note that this estimation procedure determines the tilt of the illuminant direction to within $\pm\pi/2$, leaving an ambiguity about illuminant position which is identical to the human perceptual ambiguity.

This estimation procedure has been applied to several synthetic images, and to images of real-world scenes, consisting of rocks and logs (see Pentland 1980, 1982a, b). In the synthetic images this maximum likelihood estimation procedure correctly determined the illuminant direction, as was to be expected for an image which fulfilled the assumptions on which the estimation procedure was based. The performance of this algorithm on images of real-world scenes is discussed in chapter 3, section 3.

3.2 Using The Illuminant Direction

It is reasonable to expect that if the illuminant direction were known then we would have additional constraint available to help in the estimation of surface shape, simply because there would be two fewer degrees of freedom left in equations (12). In the first section of this chapter we saw that the surface type could be identified directly from measurement of d^2I, up to reflections about the image plane. Thus what remains undetermined is the convexity of the surface, i.e., whether along a particular direction the surface is curving toward the viewer (concave), or away from the viewer (convex). One of the important ways knowledge of $\mathbf{L}$ may be used is to complete this typing of the surface, by giving us sufficient constraint to identify the surface as convex or concave.

Using constraint from $\mathbf{L}$ to determine surface convexity. The convexity of the surface, as well as the surface type, is determined by the signs of κ_1 and κ_2. In the first section of this chapter we were able to determine whether one or both of the surface curvatures κ_1, κ_2 were zero. We could not, however, determine whether the curvatures were both positive (a concave surface), both negative (a convex surface) or if they were of opposite sign (a saddle surface). To complete the typing of the surface we must determine the sign of the curvatures.

If we examine equation (7), we find that dI is dependent upon the angle between $\mathbf{p}_u$ or $\mathbf{p}_v$ and $\mathbf{L}$, the curvatures κ_1 and κ_2, and the projection matrix Φ:

$$dI = \rho\lambda\begin{pmatrix} -\kappa_1(\mathbf{p}_u \cdot \mathbf{L}) & -\kappa_2(\mathbf{p}_v \cdot \mathbf{L}) \end{pmatrix}\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix} \tag{7}$$

Even if we are given $\mathbf{L}$ there remain too many unknown factors to exactly determine the sign of the

curvatures, because some configurations of these unknown factors produce quite atypical patterns of image intensity.

Given **L**, each arrangement of curvature has its own typical or generic appearance. Whether dI is positive or negative along a particular direction depends upon whether $d\mathbf{N}$ points toward, or away from, the illuminant **L**. If we assume that surface orientation and curvature are randomly distributed, then for a convex surface dI measured in the direction (dx, dy) will typically be positive if (dx, dy) is toward **L**. This is illustrated in figure 2-2 (A). The sign of dI is positive because for a convex surface $d\mathbf{N}$ measured along (dx, dy) typically points in the direction (dx, dy), so that $dI = \rho\lambda d\mathbf{N} \cdot \mathbf{L}$ is positive. On the other hand, if the surface is concave (curving toward the viewer) then dI will typically be negative, because for a concave surface $d\mathbf{N}$ measured along (dx, dy) typically points in the direction $(-dx, -dy)$, so that $dI = \rho\lambda d\mathbf{N} \cdot \mathbf{L}$ is negative. Thus the sign of dI as we measure toward and away from the illuminant gives us an estimate of the surface convexity in that direction:

Proposition (Surface Convexity). **The maximum likelihood estimate of surface convexity in a direction (dx, dy), i.e., whether the surface is becoming closer (concave) or farther (convex) from the viewer as one moves along that direction, assuming that κ_1, κ_2 are identically distributed and that surface orientation is uniformly distributed, is**

$$\begin{aligned} \textit{convex:} \quad & sgn(x_L dx + y_L dy) = sgn(dI) \\ \textit{concave:} \quad & sgn(x_L dx + y_L dy) \neq sgn(dI) \end{aligned}$$

where $\mathbf{L} = (x_L, y_L, z_L)$ and dI is measured in the direction (dx, dy)[6] .

Having an estimate of the sign of the surface curvature in each direction allows to complete the surface typing begun in the first section of this chapter. The angle between τ_0, the direction along which $dI = 0$, and τ_L, the tilt of the illuminant ($\tau_L = \tan^{-1}(y_L/x_L)$), provides sufficient information to estimate surface type, as the sign of dI is positive on one side of τ_0, and negative on the other side. Thus knowing τ_0 and τ_L allows us to estimate the convexity of the surface along each image direction, and thus estimate the surface type. Figure 2-2 (B) shows the probability distribution of τ_0 for each

[6]The function $sgn(x)$ is $+1$ when $x > 0$, -1 when $x < 0$ and zero when $x = 0$.

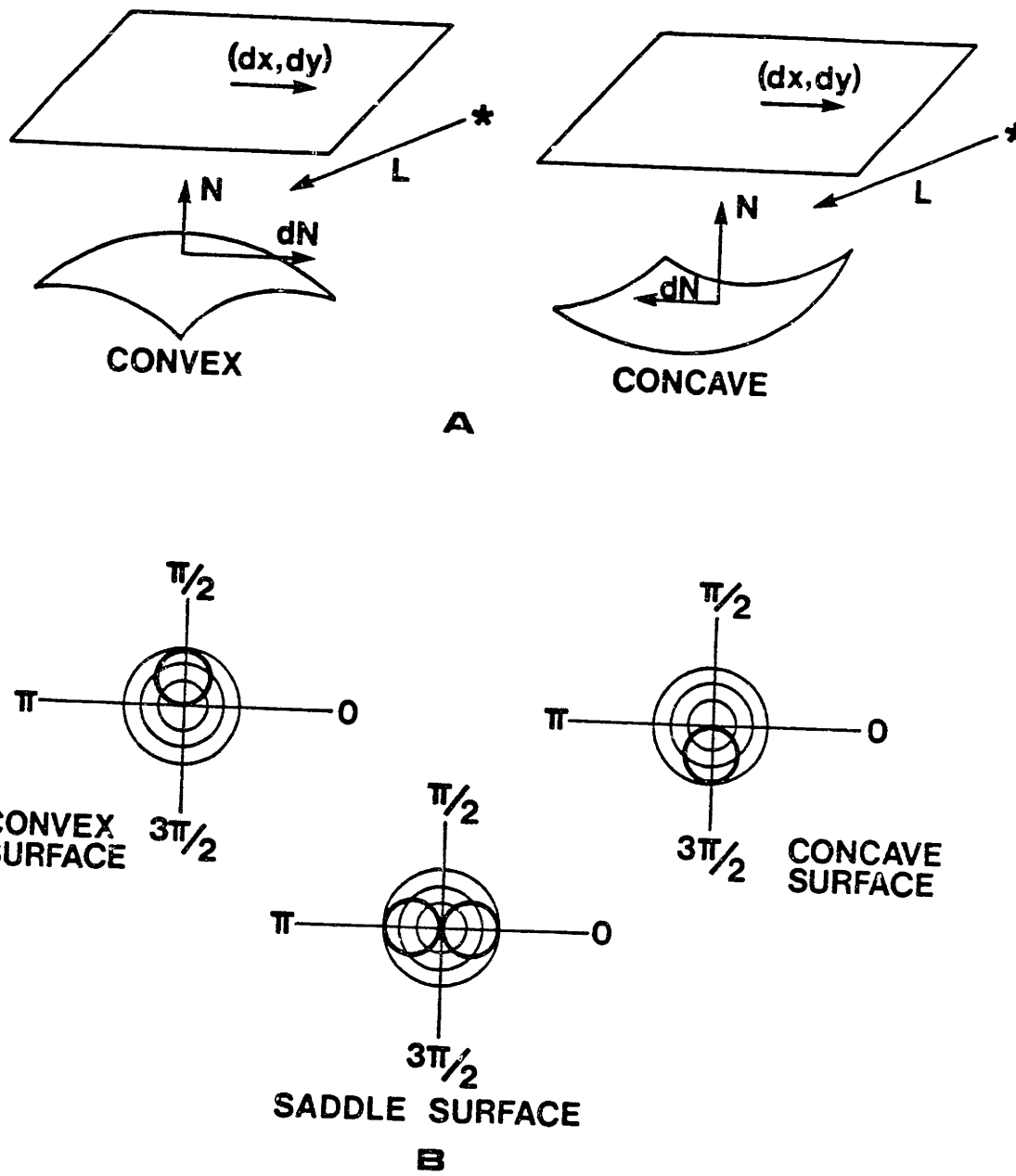


Figure 2-2. (A) For a convex surface, $d\mathbf{N}$ measured along (dx, dy) typically points in the direction (dx, dy), so that if the direction (dx, dy) is toward the illuminant then $dI = \rho\lambda d\mathbf{N} \cdot \mathbf{L}$ is positive. For a concave surface, $d\mathbf{N}$ measured along (dx, dy) typically points in the direction $(-dx, -dy)$, so that $dI = \rho\lambda d\mathbf{N} \cdot \mathbf{L}$ is negative. Thus the sign of dI along (dx, dy) relative to the illuminant direction gives an estimate of the surface convexity along that direction. (B) Use of constraint from the illuminant direction to determine the qualitative type of surface. Each type of surface has a generic appearance, which may be characterized by the angle between τ_0, the direction in which $dI = 0$, and τ_L, the illuminant direction. The distribution of $\tau_0 - \tau_L$ is shown for each surface type, assuming that $dI > 0$ to the right of τ_0. It can be seen that the appearance of the different types does not overlap much, so that a good identification of the surface type may be made from this angle.

surface type given τ_L, the tilt of the illuminant[7] , and assuming that all other variables are uniformly distributed. As can be seen by comparing the overlap between these probability distributions, the probability of a correct identification is quite good.

Note that the ambiguity of $\pm\pi/2$ in the estimation of illuminant direction leads to a global convexity/concavity ambiguity. Thus, just as with the human percept, when scene is sufficiently simple that it is not certain that convex objects dominate the scene, the direction of the illuminant may be "switched" by $\pi/2$, and this causes all of the convexity/concavity determinations to switch sign.

[7]These distributions were determined by a Monte Carlo simulation.

4 Surface Orientation From Changes In Shading

We now have developed a theory to determine the type of the surface, estimate the illuminant direction, and use the illuminant direction to determine the convexity of the surface. We know the sign of the principal surface curvatures, κ_1 and κ_2, which constitutes a good description of the *qualitative* shape of the surface. Knowing the sign of the principal curvatures, however, does not provide constraint on what we set out to determine — the surface orientation. What is needed now is a method of estimating surface orientation, as described by **N**, the surface normal.

Section 2 of this chapter, on the intrinsic limitations of shading, showed that locally foreshortening and surface curvature have the same effects, and thus it is impossible to distinguish between them for the purposes of estimating surface shape. Therefore we cannot hope to solve completely for either the curvatures κ_1 and κ_2 or for the surface normal **N** on the basis of shading information alone; there will always be at least one degree of freedom remaining. We may however make an *estimate* of **N**, and that is the goal of this section of the thesis.

We cannot use the image intensity alone to estimate **N** because it does not provide us with enough measurements at each point in the image. Use of the first derivative of image intensity to estimate shape also has an intrinsic problem; as we measure dI in various image directions both the magnitude of $d\mathbf{N}$ and the angle between $d\mathbf{N}$ and **L** vary. Thus there are two unknown factors which determine each measurement of dI. This makes it difficult to use dI for estimation of surface shape.

The same problems do not apply to the second derivative. We have seen that the second derivative of image intensity, d^2I, depends upon the second derivative of the surface normal, $d^2\mathbf{N}$:

$$d^2I = \rho\lambda d^2\mathbf{N} \cdot \mathbf{L}$$

If a surface has zero third derivatives (as illustrated by figure 1-6), then $d^2\mathbf{N}$ points in the same direction as the surface normal (see figure 1-7) *regardless of the direction in which it is measured.* This means that the effects of the illuminant are identical upon all three components of d^2I, as well as upon the surface normal itself. This fact means that we may "factor out" the effects of **L** on d^2I by

dividing by the image intensity, I. Removing the effects of illumination is an important step toward solving the problem of estimating surface orientation. As discussed in chapter 1, the assumption of zero third derivatives does not introduce significant error in our estimation of the surface even if the assumption is incorrect.

We can show how the effects of the illuminant direction may be "factored out" of d^2I by dividing by I by expanding d^2I/I using equation (12):

$$\frac{d^2I}{I} = \frac{\rho\lambda d^2\mathbf{N}\cdot\mathbf{L}}{\rho\lambda\mathbf{N}\cdot\mathbf{L}} \tag{14}$$

This expansion shows that the surface albedo term, ρ, and the illuminant strength term, λ, divide out so that d^2I/I is independent of surface albedo and illuminant strength. If we then use equation (9) to expand $d^2\mathbf{N}$ in terms of surface curvature and the derivatives of $\mathbf{p}$ we obtain from equation (14):

$$\frac{d^2I}{I} = \frac{-\kappa_1(\mathbf{p}_{uu}\cdot\mathbf{L})du^2 - \kappa_2(\mathbf{p}_{vv}\cdot\mathbf{L})dv^2}{\mathbf{N}\cdot\mathbf{L}} \tag{15}$$

Equation (9) assumes that the third derivatives of $\mathbf{p}$ are zero. Under this assumption $d^2\mathbf{N}$ is parallel to $\mathbf{N}$ (as shown in the appendix), and so

$$\mathbf{p}_{uu}\cdot\mathbf{N} = \pm\|\mathbf{p}_{uu}\| \qquad \mathbf{p}_{vv}\cdot\mathbf{N} = \pm\|\mathbf{p}_{vv}\|$$

where the variable sign has the same sign as the surface curvature in that direction. Noting that by definition in the (u, v) co-ordinate system

$$\kappa_1 = \mathbf{p}_{uu}\cdot\mathbf{N} \qquad \kappa_2 = \mathbf{p}_{vv}\cdot\mathbf{N}$$

we see that equation (15) may be written

$$\frac{d^2I}{I} = \frac{-\kappa_1^2(\mathbf{N}\cdot\mathbf{L})du^2 - \kappa_2^2(\mathbf{N}\cdot\mathbf{L})dv^2}{\mathbf{N}\cdot\mathbf{L}}$$

where the variable sign of each term is taken to be negative if the curvature is greater than zero, and positive otherwise. The terms $(\mathbf{N}\cdot\mathbf{L})$ then cancel in the numerator and denominator, leaving

$$\frac{d^2I}{I} = -\kappa_1^2du^2 - \kappa_2^2dv^2$$

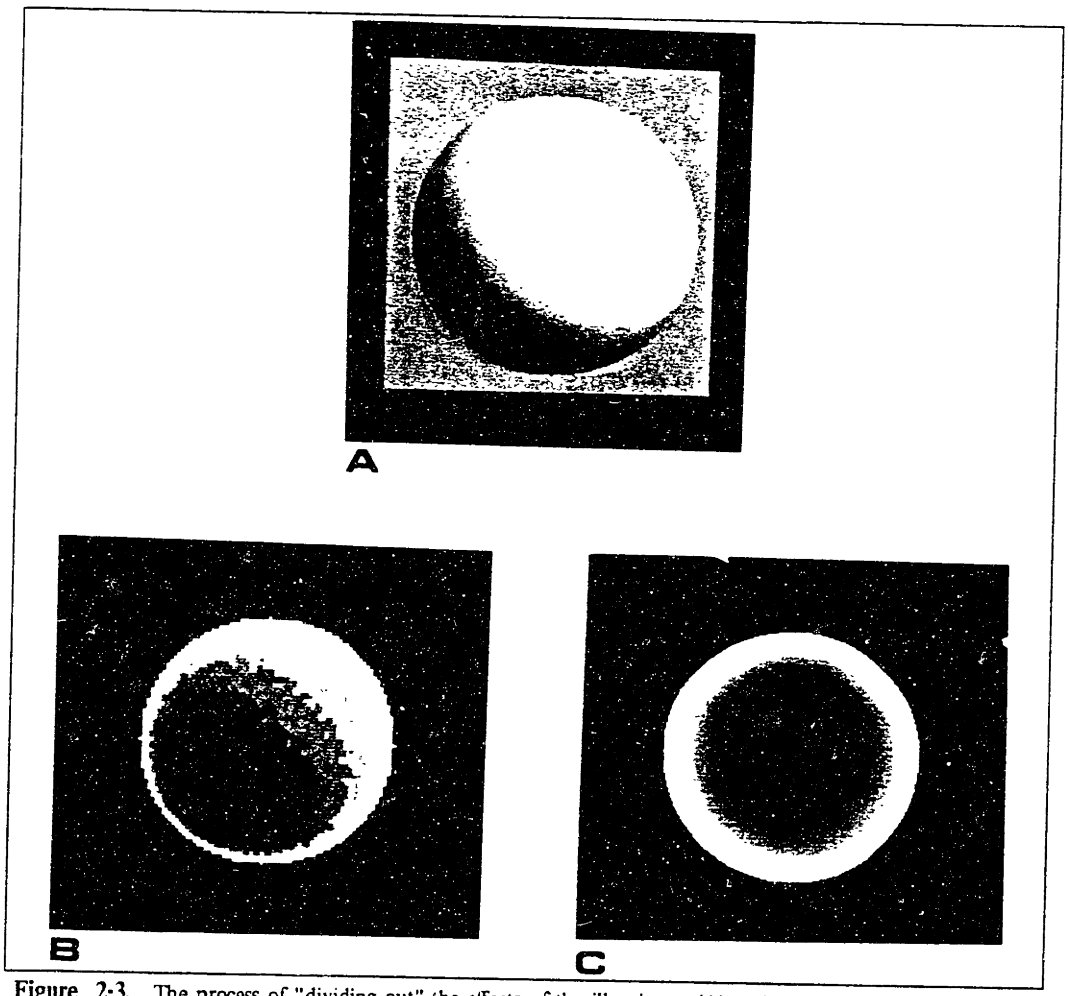


Figure 2-3. The process of "dividing out" the effects of the illuminant. (A) an image of a sphere illuminated from the side. (B) shows $\nabla^2 I = I_{xx} + I_{yy}$ as calculated from the image in (A). Larger values of $\nabla^2 I$ are lighter, smaller values are darker. Note the asymmetry in this figure caused by the sphere being illuminated from the side. (C) shows $\nabla^2 I / I$ as calculated from the image in (A). After division by I, there is almost no asymmetry in the figure; the illuminant effects have been "divided out". The remaining asymmetry occurs only around the edge of the sphere, where the assumption of a homogeneous surface is violated.

which does not contain terms for the illuminant direction, illuminant strength, or surface albedo. This example shows how the effects of the illuminant may removed in the surface tangent-plane coordinate system. Exactly the same operations apply to the general image-plane expression, equation (10), and render it free of illuminant and albedo effects (see the appendix).

This process of "dividing out" the effects of the illuminant is illustrated by figure 2-3. Part (A) of this figure shows a sphere illuminated from the side. Part (B) of this figure shows $\nabla^2 I = I_{xx} + I_{yy}$ as calculated from the image in (A). Larger values of $\nabla^2 I$ are lighter, smaller values are darker. Note

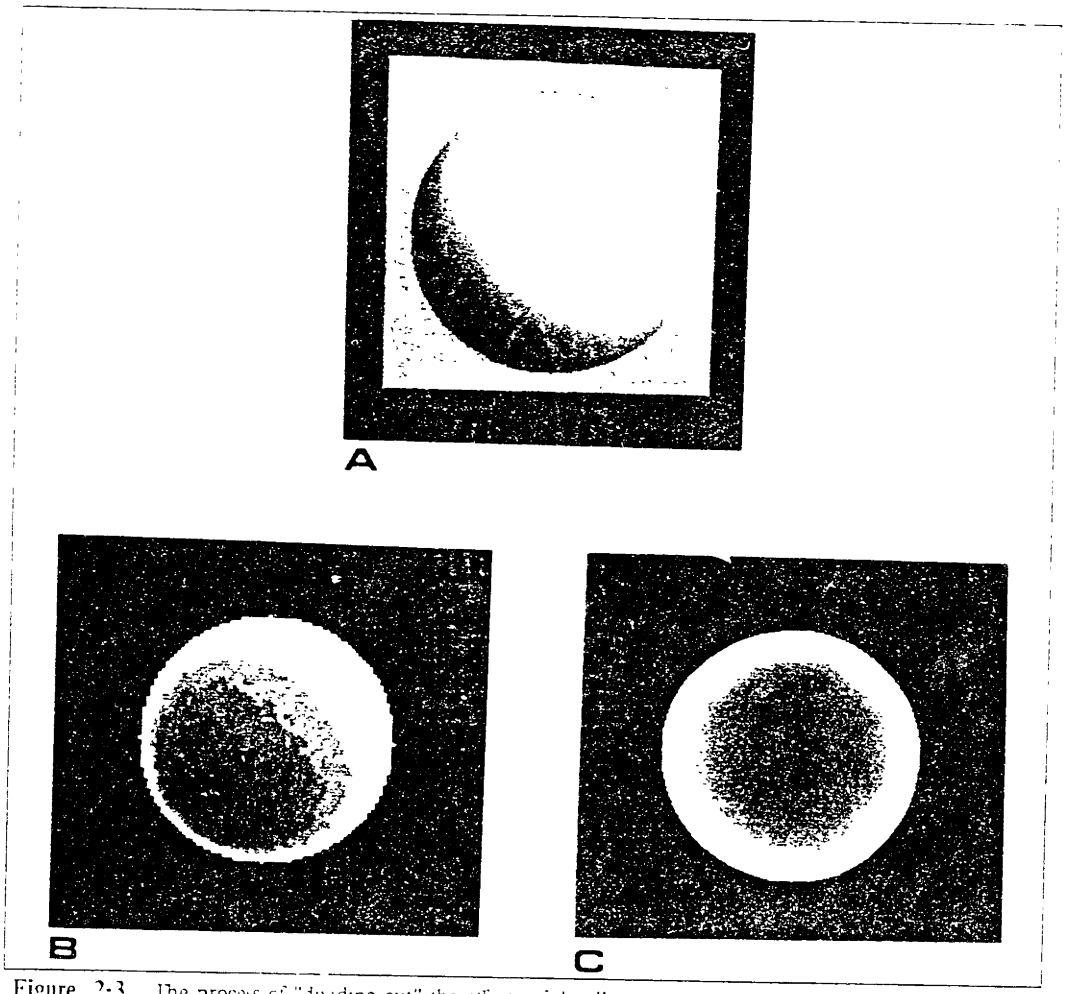


Figure 2-3. The process of "dividing out" the effects of the illuminant. (A) an image of a sphere illuminated from the side. (B) shows $\nabla^2 I = I_{xx} + I_{yy}$ as calculated from the image in (A). Larger values of $\nabla^2 I$ are lighter, smaller values are darker. Note the asymmetry in this figure caused by the sphere being illuminated from the side. (C) shows $\nabla^2 I / I$ as calculated from the image in (A). After division by I, there is almost no asymmetry in the figure, the illuminant effects have been "divided out". The remaining asymmetry occurs only around the edge of the sphere, where the assumption of a homogeneous surface is violated.

which does not contain terms for the illuminant direction, illuminant strength, or surface albedo. This example shows how the effects of the illuminant may removed in the surface tangent-plane coordinate system. Exactly the same operations apply to the general image-plane expression, equation (10), and render it free of illuminant and albedo effects (see the appendix).

This process of "dividing out" the effects of the illuminant is illustrated by figure 2-3. Part (A) of this figure shows a sphere illuminated from the side. Part (B) of this figure shows $\nabla^2 I = I_{xx} + I_{yy}$ as calculated from the image in (A). Larger values of $\nabla^2 I$ are lighter, smaller values are darker. Note

that the asymmetry in this figure is caused by the sphere being illuminated from the side. Part (C) of figure 2-3 shows $\nabla^2 I/I$ as calculated from the image in (A). After division by I, there is almost no asymmetry in the figure; the illuminant effects have been "divided out". The remaining asymmetry occurs only around the edge of the sphere, where the assumption of a homogeneous surface is violated (see chapter 3 section 1).

4.1 The Tilt Of The Surface

When we observe a smooth surface, we obtain a strong impression of the *tilt* of the surface — that is, which direction the surface is slanting away from us. Because we have such a strong impression of the tilt, it seems that there might be some way of directly computing the tilt of the surface. If we could exactly determine τ, the tilt of the surface, then there would be only one degree of freedom (the slant) left undetermined in **N**. The propositions in section 2 of this chapter prove that this is the most information about surface orientation which can be unambiguously determined from local shading information.

Let us develop an intuition of how we might go about estimating the tilt. Imagine that we could directly observe the lines of curvature on a surface. These lines of curvature would look like the lines drawn in figure 2-4. If we were looking straight down on a surface with no twist, the lines of curvature would appear perpendicular, as in figure 2-4 (A). As we tilted the surface off to one side, the lines of curvature would appear progressively more spread apart, as in figure 2-4 (B) and (C). Different directions of tilt cause spreading in different directions[8], as shown by (D).

We can't directly observe lines of curvature on the surface, of course, but we can observe the interaction of surface curvature with the illuminant in the second derivatives of image intensity. The second derivative of image intensity has three components: I_{xx} and I_{yy}, the "curvature" of image intensity along the x and y axes, and I_{xy}, the "spread" of those curvatures. Just as with the spreading of the lines of curvature, the direction in which this spread term is the greatest is the direction of the

[8]Stevens 1980 observes that this "spreading" effect may be used to constrain surface orientation given imaged contours known *a priori* to be perpendicular on the surface.

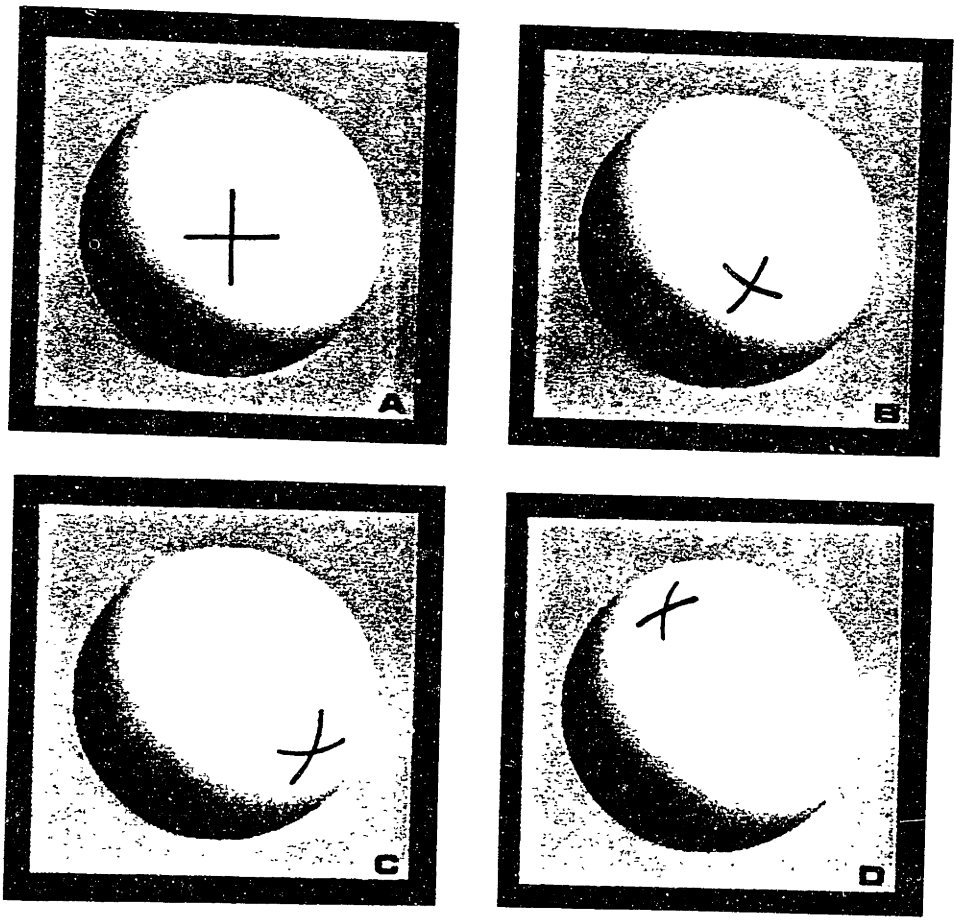


Figure 2-4. The manner in which image curvature "spreads" indicates the tilt of the surface. This may be understood by imagining that we could directly observe the lines of curvature on a surface. These lines of curvature would look just like the lines drawn in this figure. If we were looking straight down on a surface with no twist, the lines of curvature would appear perpendicular, as in (A). As we tilted the surface off to one side, the lines of curvature would appear progressively more spread apart, as in (B) and (C). Different directions of tilt would cause spreading in different directions, as shown by (D).

surface tilt. The direction in which the spread is the greatest is also the direction along which d^2I is the greatest, and so the following proposition (whose proof is given in the appendix):

Proposition (Tilt of the Surface). Given the smooth, homogeneous surface assumed by equation (10), and $I_{xx} \neq I_{yy}$, $I_{xy} \neq 0$), then the tilt of the surface is the image direction in which the second derivative of image intensity, d^2I, is greatest.

Thus one of the two components of surface orientation, the tilt, may be directly determined from the second derivative of image intensity *without knowledge of the illuminant direction.* This leaves only the slant of the surface to be determined.

4.2 The Slant Of The Surface

In the previous sections we have proven that while the tilt of the surface may be exactly determined, the slant of the surface cannot be completely disentangled from the curvature of the surface. Still we may hope to make an unbiased estimate of the surface slant. We would then have the best estimate of **N** which it is possible to obtain from local shading information. This section develops a method for making such an estimate of surface slant. First we observe that:

Proposition (Normalized Laplacian). Given the smooth, homogeneous surface assumed by equation (10), then

$$\frac{\nabla^2 I}{I} = -\kappa_n^2 z_N^{-2} - \kappa_m^2$$

where κ_n is the surface curvature along the surface tilt direction and κ_m is the surface curvature in the orthogonal direction, z_N is the z component of the surface normal, equal to the cosine of the slant of the surface.

What this proposition shows is that $\nabla^2 I/I$ is a function of the squared curvatures of the surface, κ_n^2 and κ_m^2, and the foreshortening.

There is a limited range of relationships between the surface curvatures and the foreshortening

for any particular observed value of d^2I/I. Therefore if we were given that the magnitude of the surface curvature had a particular *a priori* distribution, say a uniform distribution, then for any observed value of $\nabla^2 I/I$ we could make a maximum likelihood estimate of the foreshortening. As the foreshortening is proportional to z_N^{-2}, this then gives us an estimate of the slant of the surface (which is $\cos^{-1}(z_N)$). This leads to the following proposition:

Proposition (Estimation Of Slant). **Assuming a uniform distribution of surface curvature, and the smooth, homogeneous surface assumed by equation (10), then the maximum-likelihood estimate of z_N, where z_N is equal to the arccosine of the slant of the surface, is**

$$z_N = \sigma_\kappa \left(\left| \frac{\nabla^2 I}{I} \right| - \sigma_\kappa^2 \right)^{-\frac{1}{2}} \tag{16}$$

where σ_κ^2 is the variance of the distribution of surface curvatures.

Section 2 of this chapter showed the intrinsic underdetermination of shape by local shading information, demonstrating that at least one degree of freedom will remain undetermined by the shading information. The tilt proposition showed that one of the two parameters of surface orientation can be determined exactly, leaving only the slant undetermined. This proposition gives a maximum likelihood estimate of the slant, which by definition is the minimum-variance unbiased estimate. *Therefore the slant and the tilt propositions together constitute the best estimate of surface that it is theoretically possible to make from local shading information.* Note that neither the slant estimate nor the tilt estimate require knowledge of the illuminant direction.

5 Shape From Contours

We have seen that local analysis of shading can produce useful information about surface shape, however we have also seen that shading information alone is intrinsically insufficient to exactly determine surface shape. Imaged contours can also be used to obtain information about shape, information which is independent of the sources of error inherent in shading information. Thus use of information derived from imaged contours in conjunction with shape information derived from shading can provide the constraint necessary to determine surface shape more completely. This section therefore addresses the problems of finding contours in the image, and then obtaining shape information from imaged contours.

5.1 Locating Imaged Contours

If we are to use imaged contours, first we must find them. Further, all of the results developed so far have depended upon the assumption that the region under examination is homogeneous. The only way to insure homogeneity is to localize all of the discontinuities surrounding the homogeneous regions — that is, find the "edges".

The problem of locating contours in the image has been investigated by many researchers (Roberts 1965, Horn 1968, Herskovitz and Binford 1970, Rosenfeld *et al* 1971a, 1971b, Heuckel 1971, Macleod 1972, Binford and Horn 1971, Shirai 1973, Duda and Hart 1973, Davis 1975, Marr 1976b, Rosenfeld and Kak 1976, Pratt 1978, Marr and Hildreth 1980), and many different solutions have been proposed (see Marr and Hildreth for a good survey of the field). Because we must find the imaged contours in order to accomplish our goals, it is useful examine the problem of edge detection in light of the results developed so far.

In all of the earlier portions of this chapter, it was assumed that we were observing a homogeneous, continuous patch of Lambertian surface, with constant albedo and illumination. Thus in order to apply the results obtained so far, we must try to locate all discontinuities in albedo (e.g., changes in material), illumination strength (such as shadows), surface normal, and reflectance function (e.g., a

highly specular point). Thus it is the occurance of these events in the course of image formation which we will label "imaged contours".

How are we to find these imaged contours? Marr and Hildreth 1980 have proposed using the zero-crossings of $\nabla^2 I$ to locate image contours. Their edge-finding technique has met with particular success, especially as employed in the Marr-Poggio-Grimson stereo algorithm (Marr and Poggio 1979, Grimson 1980, 1981), and so we will investigate its functioning for use in this application.

Let us begin by examining the image formation equation developed in the beginning of chapter 1:

$$I = \rho\lambda(\mathbf{N}\cdot\mathbf{L})R(\mathbf{L},\mathbf{V},\mathbf{N})(\mathbf{N}\cdot\mathbf{V})^{-1}$$

If we combine the functions $R(\mathbf{L},\mathbf{V},\mathbf{N})$ and $(\mathbf{N}\cdot\mathbf{V})^{-1}$ into a single function $\Re(\mathbf{L},\mathbf{N})$ describing how light is reflected from the surface onto the image ($\mathbf{V}$ need not appear as it is constant), we see that

$$\begin{aligned}\nabla^2 I =& \nabla^2\rho\lambda\mathbf{N}\cdot\mathbf{L}\Re(\mathbf{L},\mathbf{N}) + \rho\nabla^2\lambda\mathbf{N}\cdot\mathbf{L}\Re(\mathbf{L},\mathbf{N}) \\ &+ \rho\lambda\nabla^2(\mathbf{N}\cdot\mathbf{L})\Re(\mathbf{L},\mathbf{N}) + \rho\lambda\mathbf{N}\cdot\mathbf{L}\nabla^2(\Re(\mathbf{L},\mathbf{N})) \\ &+ \textit{cross terms}\end{aligned}$$

The "cross terms" in this equation are terms with differentials of *two* factors, rather than just one. Let us assume that it is unlikely that two of the factors will vary together, so that we make take these cross-terms to be zero.

Under the assumption that it is unlikely that two factors vary together, there will be a zero-crossing of $\nabla^2 I$ whenever any of the factors $\nabla^2\rho$, $\nabla^2\lambda$, $\nabla^2(\mathbf{N}\cdot\mathbf{L})$ or $\nabla^2\Re(\mathbf{L},\mathbf{N})$ undergoes a zero-crossing, i.e., whenever any of ρ, λ, $\mathbf{N}\cdot\mathbf{L}$ or $\Re(\mathbf{L},\mathbf{N})$ undergoes a change in convexity[9] Such changes in convexity are found at discontinuities in these variables[10] , so that the Marr and Hildreth operator should localize the image formation events that we have defined as imaged contour. It is clear, however, that the Marr and Hildreth operator will also localize inflections in these quantities, and we do not wish to label these inflections as imaged contours.

[9]That is, one or both of the partials I_{xx}, I_{yy} changes sign.

[10]While not true of ideal images, it is true in real images, because they always have some amount of blurring. See section 1 of the next chapter.

A second potential problem with using zero-crossings of $\nabla^2 I$ stems from the effects of image noise. It is easily confirmed that a small noise spike in an otherwise perfectly constant region of of the image will generate a small zero-crossing surrounding it. Thus noise is a problem which must be considered in using zero-crossings of $\nabla^2 I$.

We can circumvent the problem of noise by discarding weak zero-crossings. Given a bound on the noise amplitude, δ, we can effectively exclude the zero-crossings generated by noise by discarding zero-crossings which cross the zero-value point with a slope less than the slope of the zero-crossing generated by a noise pulse of size δ. Discarding such weak zero-crossings means that only robust contours, those of magnitude greater than δ, will be localized. Thus this procedure will discard some "real" (although weak) zero-crossings along with the noise-generated zero-crossings. It appears that for our purposes this is an advantage, rather than a drawback, as most of these weak "real" zero-crossings correspond to the unwanted inflections in ρ, λ, etc. Thus discarding weak zero-crossings appears to be a method of ridding ourselves of both noise-produced zero-crossings and inflection-produced zero-crossings (see chapter three for an evaluation of this discarding procedure).

5.2 Shape From Imaged Contour

There are two categories of imaged contour that are relevant to the estimation of surface shape. The first category is smooth occluding contours, such as occur at the edge of an image of a sphere. The second category contains all varieties of sharp contours, including sharp occluding contours, such as found at the edge of an image of a disk, and sharp surface contours, such as are generated by surface markings. These two categories, smooth contours and sharp contours, are distinguished by the different constraint which they provide on the interpretation of shape.

Constraint from smooth occluding contour. A smooth occluding contour provides the strongest possible constraint on surface shape: the surface normal along that contour is known exactly; it lies perpendicular to the contour and parallel to the image plane. Thus the identification of smooth contours is of great importance to the interpretation of surface shape.

The smooth occluding contour, by itself, provides only local constraint on shape; there are an infinite number of shapes bounded by the same smooth occluding contour. By using shading information the powerful constraint from smooth occluding contours may be propagated across the surface of an object. The fact that shading may be used to propagate constraint has been successfully used by Horn, Grimson and others.

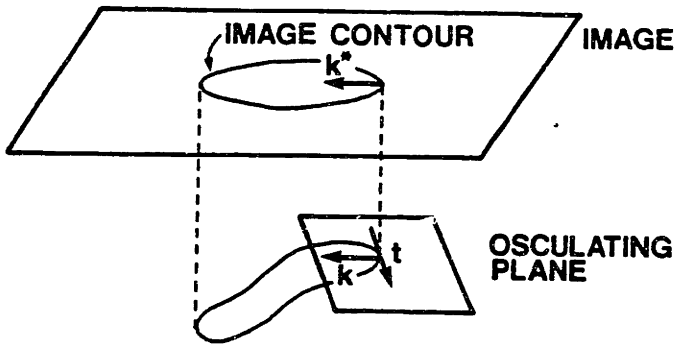
Constraint from sharp contours. Sharp contours do not provide as strong constraint on shape as do smooth occluding contours. There is no absolute information about orientation or shape which may be gained from a sharp contour, as is provided by a smooth occluding contour. Nonetheless, important constraints may be derived from a sharp contour, as is discussed in the following subsection.

5.3 Three-dimensional interpretation of sharp contours

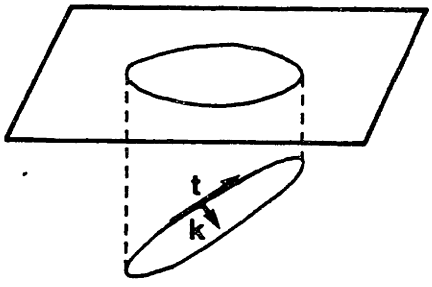
Not all three-dimensional interpretations of a sharp contour are equally likely. Given an imaged sharp contour one may make a reasonable guess about the three-dimensional shape of the contour using the fact that viewer position is independent of the shape of the contour generator.

Figure 2-5 (A) shows the relationships which exist between the contour generator and the imaged contour. Curvature in the image of a three-dimensional curve is a function of κ, the magnitude of the contour generators curvature vector, **k**, and the foreshortening along both the contour generators curvature vector and its tangent vector, **t**. Foreshortening along the curvature vector causes κ^*, the magnitude of the imaged curvature, to be less than κ, the magnitude of the three-dimensional curvature vector as shown by figure 2-5 (B). Foreshortening along the tangent vector **t** causes κ^* to be greater than κ, as shown by figure 2-5 (C).

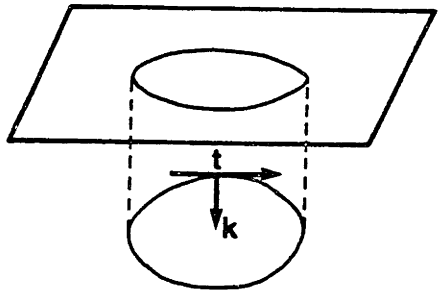
The torsion, τ, of the contour generator is the amount which the curvature vector deviates from a single plane (called the osculating plane). The torsion therefore describes the "twist" in the contour generator. The presence of torsion gives no direct evidence in the image, it simply results in more or less foreshortening as the plane of the contour varies. The effects of torsion can therefore be exactly mimicked by changes in the curvature κ, and vise versa. We may however make an estimate of the



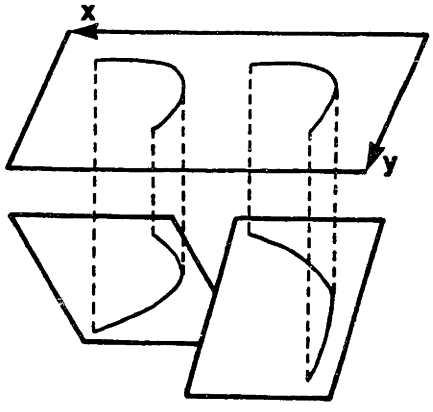
A. IMAGING OF 3-D CONTOUR GENERATORS



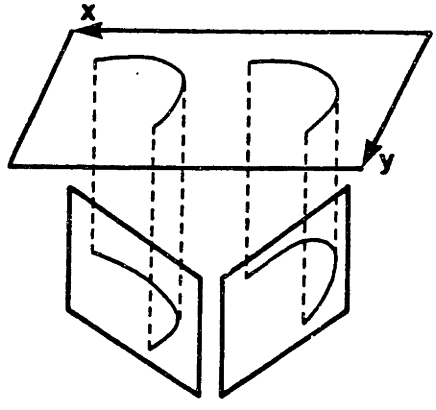
B. FORESHORTENING ALONG t MAKES IMAGED CURVATURE GREATER THAN 3-D CURVATURE



C. FORESHORTENING ALONG k MAKES IMAGED CURVATURE LESS THAN 3-D CURVATURE



MOTION AROUND X-AXIS



MOTION AROUND Y-AXIS

D. TWO DIMENSIONAL FAMILY OF POSSIBLE PLANAR CONTOUR GENERATORS FOR ANY IMAGED CONTOUR

Figure 2-5. The geometry of imaged contours and contour generators.

torsion in an imaged contour:

Proposition (Zero Torsion). The maximum likelihood estimate of the torsion of the contour generator is zero (i.e., no "twisting" of the contour generator).

This proposition follows from the independence of viewer position and contour generator shape — if the shape is independent of the viewer position, then the torsion of the contour generator is also independent. Therefore it is equally likely that the torsion of the contour generator is toward the viewer as away, and the most likely estimate is the mean torsion: zero. This proposition also follows from the assumption of general position (the assumption that the viewed scene is stable under small changes in viewpoint). A curve with substantial torsion will alter its appearance considerably with small changes in viewer position; thus if we assume general position then the torsion in a imaged contour must be small.

Note that at places where the curvature vector is zero — straight segments and inflection points — the torsion is not defined and may be taken to be zero. Because torsion is not defined along straight segments and at inflection points, even with zero torsion the osculating plane may change freely at these points. As a consequence, we will normally estimate that most of the "twisting" or changes in osculating plane of the contour generator will occur in places where the imaged curvature is zero, i.e., at straight lines and inflection points.

If the torsion is zero, then for each uninflected arc of an imaged contour there is a two-parameter set of planar contour generators which could have generated that imaged arc. Those two parameters correspond to the two degrees of freedom of the osculating plane. This is shown in figure 2-5 (D).

Three-dimensional estimation. Assume that we are given the image of an uninflected arc, and asked to estimate its three-dimensional shape. Let us start to determine the three-dimensional shape of the arc by examining the first five points of the arc, which always define an ellipse in the image plane. On the basis of the previous proposition, we would guess that the generator of the arc had zero torsion, i.e., it was planar.

The case of imaged arcs which are assumed to be planar was examined by Witkin 1980. He observed that as all contour generator orientations are equally likely (because viewer position is independent of generator orientation), it is unlikely that one would get the elliptical, anisotropic distribution of orientations which is characteristic of foreshortened images of uniformly-distributed contour generators. It is most likely that the anisotropy which looks like foreshortening is in fact due to foreshortening. Using this result, Witkin constructed a maximum-likelihood estimator for the contour generator shape, under the assumption that the contour generator is planar. The maximum-likelihood estimate for the ellipse defined by the first five points of our imaged arc is that the arc is generated by a segment of a circle. This estimate corresponds well to human perception — when presented with an ellipse and asked for a three-dimensional interpretation, people interpret the ellipse as a tilted circle. Nonetheless, it is clear that this estimate is weak because *any* five points define an ellipse.

Let us now examine the next point on the imaged arc, the sixth. If this point lies on the same ellipse defined by the first five points, we have much stronger evidence of the contour generators shape, because it is very unlikely that by chance the sixth point would fall on the same ellipse as the first five if either the curvature or the osculating plane of the contour generator was changing. If the osculating plane was changing (i.e., there was torsion in the contour generator), then the curvature would have to be changing also, and in such a manner as to exactly cancel the image effects of the changing osculating plane. Similarly, if the curvature of the contour generator was changing, then the osculating plane would have to change exactly enough to cancel out the image effects of the changing curvature. As both of these cancellations are unlikely, we must conclude that the sixth point lies on the same circular contour generator defined by the first five points of the arc.

But what if the sixth point lies off the ellipse defined by the first five? Then we can be certain that either the curvature or the osculating plane (or both) of the contour generator has changed. We must measure another five points before we can estimate a new orientation for the contour generator.

These arguments leads to the following proposition:

Proposition (Planar Interpretation). Given an elliptical segment of imaged contour (five or

more points) the maximum likelihood estimate for the three-dimensional shape of the contour generator is a planar, constant-curvature contour — i.e., a segment of a circle.

As any curve may be closely approximated by a contour made up of portions of ellipses and straight lines, these two results form an interpretation strategy for imaged contour which encompasses all configurations of imaged contour. This estimation strategy results in an estimate which is quite similar to that obtained using the estimation algorithm proposed by Barrow and Tennenbaum 1981. What is novel about this treatment of the problem is the probalistic motivation for the estimation strategy.

6 Shape From Contour And Shading

There are three main ways in which shading and contour interact: (1) shading may be used to identify whether or not a particular contour is a smooth occluding contour or a sharp contour, (2) the shading on either side of a sharp contour may be be used to determine whether or not the two sides join together smoothly, and (3) the estimate of three-dimensional shape for an imaged contour, developed in the last section, may be used to help constrain the estimate surface shape derived from shading to a unique, correct interpretation of surface shape. Each of these three cases of shading and contour interacting will be dealt with separately.

6.1 Identification Of Smooth Occluding Contours

If we can identify smooth occluding contours in the image, then we can make use of the strong constraint on surface shape provided by them. The remaining contours can then be treated as sharp contours. But how are we to identify them?

In a previous section it was proven that local shading is insufficient to disambiguate between surface curvature and foreshortening. Thus we have proven that it is impossible to unambiguously identify smooth occluding contours on the basis of local shading information alone. We have also, however, presented an estimator of surface orientation which will be shown to work quite well, especially for the highly slanted areas which are an intrinsic characteristic of smooth occluding contours.

The idea for identification of smooth occluding contours is simple: the defining characteristic of smooth occluding contours is that the slant of the surface approaches $\pi/2$, and the tilt of the surface is perpendicular to the contour. Using the results of section 4 of this chapter, we may determine the surface tilt and estimate the surface slant using local shading information near a contour. If the estimated slant and tilt are appropriate for a smooth occluding contour, then the contour is identified as such; otherwise the contour is not a smooth occluding contour. Because the estimate of surface orientation developed in section 4 of this chapter is the best estimate possible from local shading information, this identification of smooth occluding contour is also the best that can be made using

local shading information.

Constraint from illuminant direction may also useful for helping to identify smooth occluding contours. The magnitude of dI along a smooth occluding contour depends on the orientation of the contour relative to the illuminant, with dI along contours perpendicular to the illuminant tilt being the smallest, and dI along contours parallel to the illuminant tilt being the largest. This is similar to an idea advanced by Barrow and Tennenbaum 1978, and like that proposal this constraint is critically dependent on constant albedo and illuminant direction. Thus this constraint from illuminant direction is useful only if applied in a qualitative manner.

6.2 Joining Regions Together

One of the most important interactions of contour and shading occurs when determining whether regions on either side of a sharp contour join together smoothly, as with the regions on either side of a surface marking, or whether the regions are separate, as on either side of a sharp occluding contour. It is clear that if the regions on either side of a contour are "the same", that is, if the regions to either side of a contour have exactly the same properties (brightness, texture, etc.) then these regions will appear identical in the image. Thus adjoining regions which appear identical may be presumed to join together smoothly, i.e., along the boundary the two regions will have the same surface normal. If, however, the regions on either side of an imaged contour are not identical, then no such constraint need apply — but often the surfaces do in fact join together smoothly, such as in the case of the dissimilar regions on either side of a shadow boundary.

The problem in determining whether two regions join together smoothly is that we are concerned only with whether or not the surface normals are the same, and not with other factors such as the surface albedo or the illumination. These other factors strongly influence the appearance of a surface in the image and so confound us. Our estimate of surface shape from local shading information, however, is not affected by surface albedo or illumination. Thus it may be used to determine whether or not two regions join together smoothly, even if they have different albedos or different illumina-

tion.

We may produce an estimate of the surface orientation along both sides of a contour using the local shading information. While such estimates may be in error because of the curvature of the surface, they are not affected by the albedo or the illumination: the estimates are purely a function of the shape of the surface. Further, it is a measure-zero occurance that two different surfaces will produce the same estimate of surface shape. Thus if along the two sides of a contour we obtain identical estimates of surface orientation, then we know with certainty that the two sides of the contour have the same shape — and thus the two sides of the contour must have the same surface orientation along that contour. Note that because we cannot produce an estimate of surface orientation for either planar or cylindrical surfaces, we cannot apply this reasoning to the boundaries between such surfaces.

If the estimates are not the same along the two sides of the contour, the the two sides of the contour have different shapes, and therefore cannot join together smoothly. Note that this inference is stronger than the inference available using texture; if the texture is identical on both sides then the two sides must join together smoothly, but one cannot assert that if the textures differ then the two sides do not join smoothly.

Thus, given that we may use shading information to estimate surface orientation, we may determine whether or not the two regions join together smoothly regardless of differences in surface albedo or illumination by comparing the estimates of surface orientation along both sides of the contour. A determination made in this manner will virtually always be correct.

6.3 Using Contour And Shading To Constrain Surface Shape

The estimates of the three-dimensional shape of imaged contour produced by the estimator developed above is subject to error, because the probabilistic constraints which allowed us to assume torsion to be zero are weak. Similarly, we have proven that it is impossible to determine surface orientation exactly using only local shading information. Fortunately, the causes of error when interpreting shape from shading (e.g., unequal surface curvatures) are different and *independent* of the causes of

error in determining the three-dimensional shape of a contour (e.g., torsion). Thus if we could use shading to check for the existence of torsion in an contour generator, or use imaged contour to check for cancellation of foreshortening effects by unequal surface curvatures, then we could obtain a better estimate of surface shape by using both sources of image information.

Let us consider what is implied by interpreting an imaged arc as planar. Planarity means that all of the points (x_i, y_i) along the contour must have z co-ordinates such that all of the points fall in a single plane. This constraint may be expressed as

$$(x_i \cos \tau_o + y_i \sin \tau_o) \tan^{-1} \sigma_o = z_i \tag{17}$$

for all (x_i, y_i), where τ_o and σ_o are the tilt and slant of the osculating plane of the contour generator.

Taking the derivative of (17) along the contour C, we obtain

$$\left(\frac{dx}{dC} \cos \tau_o + \frac{dy}{dC} \sin \tau_o\right) \tan^{-1} \sigma_o = \frac{dz}{dC} \tag{18}$$

The quantity dz may also be expressed in terms of the surface that the contour generator is embedded in:

$$dz = z_N^{-1}(\cos \tau \frac{dx}{dC} + \sin \tau \frac{dy}{dC}) \tag{19}$$

were $\mathbf{N} = (x_N, y_N, z_N)$, τ is the tilt of the surface, and the imaged direction of the contour is (dx, dy). Putting equations (18) and (19) together, we obtain the following proposition:

Proposition (Contour And Shading). **Given an imaged contour C and shading information about a point, then under the assumption that the maximum-likelihood estimate of contour generator shape is correct (i.e., the contour generator is planar) then z_N, the cosine of the slant of the surface, is**

$$z_N = \frac{(\cos \tau \frac{dx}{dC} + \sin \tau \frac{dy}{dC})}{\left(\frac{dx}{dC} \cos \tau_o + \frac{dy}{dC} \sin \tau_o\right) \tan^{-1} \sigma_o} \tag{20}$$

where dx/dC and dy/dC are the change in x and y along the contour C, τ_o and σ_o are the tilt and slant of the contour generators osculating plane (from the maximum likelihood estimate of the imaged contours shape), and τ is the tilt of the surface (from the local shading information).

Under the assumption of planarity, the quantities τ_o and σ_o, the tilt and slant of the contour generators osculating plane, may be determined exactly from the imaged contour. The contour direction, (dx, dy), and the surface tilt, τ, may be determined exactly, leaving z_N as the only quantity which is estimated and could therefore be in error. Thus equation (20) allows us to use these known quantities to determine z_N under the assumption of planarity, as long as σ_o is not zero, in which case the equation is degenerate.

This solution would be of considerably more use if we could detect violations of the assumption of planarity. We may find evidence for such violations by examining the curvature of the surface: by taking the derivative of equation (18), we obtain

$$(\frac{d^2x}{dC^2}\cos\tau_o + \frac{d^2y}{dC^2}\sin\tau_o)\tan^{-1}\sigma_o = \frac{d^2z}{dC^2}$$

The quantity d^2z/dC^2 is approximately equal to d^2I/I along the imaged contour, as for a second-order surface d^2I/I is equal to the foreshortened surface curvature (see the proof of the normalized Laplacian proposition in the appendix) which is exactly d^2z/dC^2 when C is linear. Thus if we observe that either d^2I/I varies when the curvature of the contour does not, or if the curvature of the contour varies while d^2I/I remains constant, then we have evidence of violation of the planarity assumption.

Chapter 3
Computation Of Shape

The first two chapters of this thesis developed a theory for the inference of shape from local image data. In this chapter we will discuss a practical algorithm which accomplishes the computations required by the theory. This algorithm has been implemented as a computer program on a Lisp Machine[1] , and tested on both synthetic and natural images. Comparison of the surface shape estimates produced by the computer program to the actual shapes appearing in these test images has allowed considerable evaluation of the algorithms performance, and thus indirectly an evaluation of the validity and robustness of the theory.

The first section of this chapter will discuss several issues concerning the design of the algorithm. In particular, this section will discuss the measurement of the image information required by the theory, and will address issues concerning the scale of examination (the resolution of estimation algorithm) which are raised when discussing how to measure the required image information.

The second section of this chapter will describe the algorithm by presenting block-diagram schematics showing the flow and successive transformation of information from raw image data to a description of the viewed surface. Each of the elements of this block-diagram will then be briefly discussed, in each case showing how a sequence of relatively simple, parallel operations results in the calculations required by the theory. The portions of the algorithm which utilize imaged contour are difficult to describe in such a block-diagram, as they require analysis at more than one point of the imaged contour. Therefore description these portions of the algorithm will be delayed until the third

[1]The Lisp Machines are single-user digital computers designed and built by the Artificial Intelligence Laboratory of the Massachusetts Institute Of Technology.

section of this chapter, where we may describe the algorithms operation on specific examples.

The third section of this chapter will show examples of application of the algorithm to both synthetic and natural image data, and discuss the performance of the algorithm on these images. The examples presented in this section do not constitute an exhaustive evaluation of the performance of the algorithm on image data, nonetheless I believe that one may obtain a clear impression of both the performance capabilities of the algorithm and the general validity of the theory.

1 Measuring Image Information

The problems of measuring dI, d^2I or $\nabla^2 I$ and of identifying imaged contours have both been addressed by many different researchers. Measurement of dI, d^2I or $\nabla^2 I$ by convolving the image with an appropriate filter is a standard image processing technique (see Horn 1974). Marr and Hildreth 1980 propose locating contours in a similar fashion, by convolving the image with a $\nabla^2 G(x, y)$ filter. Although the measurement of the required image information has become standard practice in the image processing community, it is unclear how the approximations made in performing measurements might influence the results of our theory. Thus we must investigate the implications of various methods of measuring the input information required in the theory.

1.1 The Notion Of Scale

In all imaging devices there are imperfections which result in a blurring of the image. This blurring is described by the blur function, or the point-spread function, and may be characterized as a filter which is applied to the image. This blurring limits the scale at which we can examine imaged surfaces by averaging out all finer details of the surface. Thus an imaging systems blur function defines a scale of examination (Mandelbrot 1977, Witkin 1980).

Let us assume that the blur function of a particular imaging system were characterized by the filter $G(x, y)$, a Gaussian[2] as shown in figure 3-1 (A), and let us also assume that we can characterize the scene and the image with infinite resolution, so that we can let $\mathbf{N}^*(x, y)$ be the surface normal at the image point (x, y), and we can let $I^*(x, y)$ be the infinite-resolution image intensity at the point (x, y)

Let us continue to assume that the image intensity values I^* are determined by the cosine of the angle between the surface normal $\mathbf{N}^*$ and the illuminant direction $\mathbf{L}$ in the following manner:

$$I^*(x, y) = \mathbf{N}^*(x, y) \cdot \mathbf{L}$$

[2]The use of a Gaussian is unimportant; any unimodal filter would suffice. However, if there is high-frequency noise in the imaging system, it is desirable to have a blur function which attenuates higher frequencies quickly. The Gaussian is commonly employed as a filter for attenuating such noise so that accurate measurements of the derivative can be made.

Then the actual image intensities $I(x, y)$ available to us after passing through our imaging system will be

$$\begin{aligned} I(x,y) &= G(x,y) \otimes I^*(x,y) \\ &= G(x,y) \otimes (\mathbf{N}^*(x,y) \cdot \mathbf{L}) \end{aligned} \tag{21}$$

where '$\otimes$' signifies the operation of convolution. If continue to assume that $\mathbf{L}$ is a constant we can re-write equation (21) by defining a new surface normal

$$\mathbf{N}(x,y) = G(x,y) \otimes \mathbf{N}^*(x,y) \tag{22}$$

where in this case the convolution operation is applied to each of the components of $\mathbf{N}^*$ individually. The new surface defined by $\mathbf{N}$ is quite different from the original surface, as the smoothing which occurs is dependent upon the value of $\mathbf{N}^*$. Regions facing away from the image plane will be much more affected by this smoothing than regions facing the image plane. Equation (22) allows us to write

$$\begin{aligned} I(x,y) &= G(x,y) \otimes (\mathbf{N}^*(x,y) \cdot \mathbf{L}) \\ &= \mathbf{N}(x,y) \cdot \mathbf{L} \end{aligned}$$

This shows that the image intensities available to us in the imaging system can be thought of as a function of the surface defined by $\mathbf{N}(x, y)$ in equation (22). If we refer only to the surface $\mathbf{N}(x, y)$ defined using the scale set by the imaging systems' blur function, we do not need to be concerned with the imaging systems' blur and may treat the imaged intensities $I(x, y)$ as exact.

The surface $\mathbf{N}(x, y)$ is the *only* surface which is available for observation in the image because all of the additional detail of $\mathbf{N}^*$ has been attenuated to below the background noise level by the imaging systems blur function[3] The fact that $\mathbf{N}$ has only limited detail available has important implications for the validity of the assumptions made in the first chapter. This is discussed at the end of this chapter.

The use of scale to define surface orientation splits the world into two parts: *Macro*structure, those features of the surface which are large enough to not be "smoothed over" by our averaging, and

[3]If we let $S(f)$ be the frequency distribution of the signal, $B(f)$ that of the blur function, and $N(f)$ that of the noise, then the signal irretrievably lost will be those frequencies for which $S(f)B(f)N^{-1}(f) < 1$ Thus the resulting signal-to-noise ration of the imaged system is $B^{-1}(f)N(f)$. It is perhaps better to speak of the inverse of this function as the blur function, rather than as we do now.

*Micro*structure, those features which are too small to appear in our average[4] . When we speak of the "shape" of a surface, it is the macrostructure to which we refer. When we speak of the (three-dimensional) "surface texture", it is microstructure to which we refer. That is to say, the details of the surface shape which are the three-dimensional texture are too small to be distinguished within our notion of the surface shape at the particular scale of examination we are currently using. The small variations which constitute the three-dimensional surface texture have somehow been smoothed over in arriving at our conception of the surface shape. The size of the region used in this smoothing operation defines the *scale* at which we are representing or examining the surface shape. This division on the basis of scale defines both what constitutes the surface orientation, and what constitutes surface texture.

1.2 Measuring The Derivatives Of Image Intensity

Using equation (22), we may calculate that the first directional derivative of image intensity, $dI(x, y)$, taken in the image direction (dx, dy) as a function of the change in surface normal (see also chapter 1):

$$dI(x, y) = d\mathbf{N}(x, y) \cdot \mathbf{L}$$

If we now were to build a imaging system whose blur function was the first directional derivative of the Gaussian, $dG(x, y)$ (shown in figure 3 - 1 (B)), then $\mathcal{I}(x, y)$, the image resulting from this blur function, would be

$$\mathcal{I}(x, y) = dG(x, y) \otimes I^*(x, y)$$

But from the Convolution theorem we have

$$dG(x, y) \otimes I^*(x, y) = d\Big[G(x, y) \otimes I^*(x, y)\Big] = dI(x, y)$$

[4]In order to make this notion of scale fully coherent, we must also assume noise in the system so that features which have been sufficiently smoothed over by the averaging process can no longer be distinguished from the background noise.

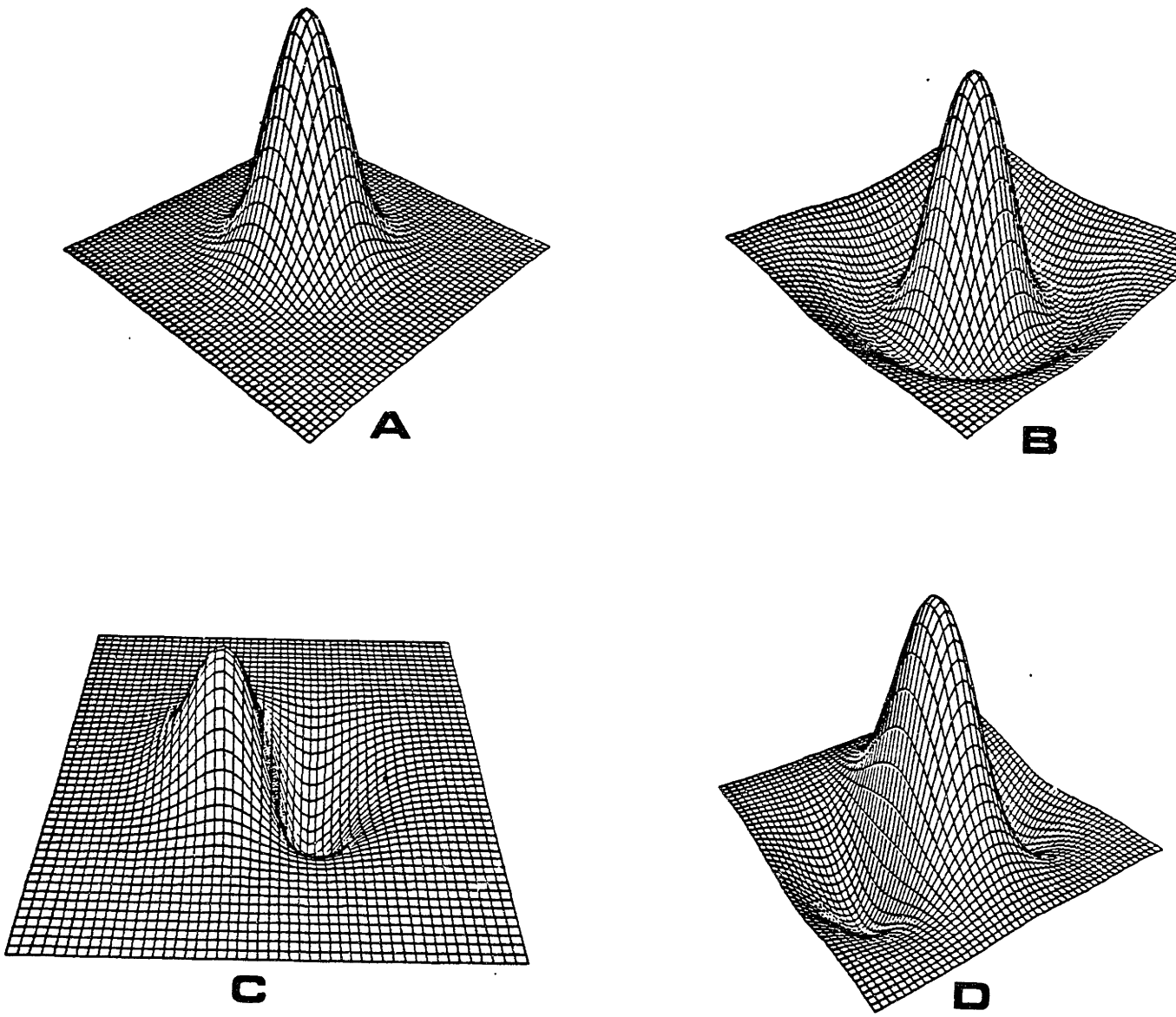


Figure 3 - 1. (A) the blur function $G(x, y)$, (B) its first directional derivative $dG(x, y)$, (C) its circularly-symmetric second derivative (the Laplacian) $\nabla^2 G(x, y)$, and (D) the second directional derivative, $d^2G(x, y)$. The function $G(x, y)$ is the most common approximation for the average blur function, because when several stages are involved in an imaging system the resulting blur function always tends toward $G(x, y)$ (by the Central Limit Theorem).

thus we see that the output of this new imaging system is the first directional derivative of the output of an imaging system whose blur function is $G(x, y)$.

If the blur function of a system were the Laplacian of a Gaussian, $\nabla^2 G(x, y)$ as shown in figure 3-1 (C), then the output $\mathfrak{I}(x, y)$ of the imaging system would be

$$\begin{aligned}\mathfrak{I}(x, y) &= \nabla^2 G(x, y) \otimes I^*(x, y) \\ &= \nabla^2 \Big[G(x, y) \otimes I^*(x, y) \Big] = \nabla^2 I(x, y)\end{aligned}$$

so that the output of this imaging system is the Laplacian of the output of an imaging system whose blur function is $G(x, y)$. Similarly, if the blur function were $d^2 G(x, y)$, the second directional derivative of a Gaussian, as shown in figure 3-1 (D), then the output would be the second derivative of an imaging system with the blur function $G(x, y)$.

Thus in order to calculate dI, $d^2 I$ or $\nabla^2 I$, the input information required for the computations of our theory, we need only convolve the original image I^* with dG, $d^2 G$ or $\nabla^2 G$. The result of these convolutions is the exact derivatives of $I(x, y)$, a smoothed version of this original image, specifically $G(x, y) \otimes I^*(x, y)$. This smoothed image, in turn, corresponds to the image produced by $\mathbf{N}(x, y)$, a smoothed version of the original surface $\mathbf{N}^*$, specifically $G(x, y) \otimes \mathbf{N}^*(x, y)$.

1.3 Finding Imaged Contours

The previous chapter discussed the edge-detection theory of Marr and Hildreth 1980. Marr and Hildreth propose defining contours in the image by calculating the zero-crossings of $\nabla^2 I$, that is, where the value of $\nabla^2 I$ changes from positive to negative or vise versa.[5] As shown in chapter 2, the zero-crossings of $\nabla^2 I$ occur whenever image contours occur. For the purposes of locating imaged contours, however, zero-crossings have the problem that they also occur whenever there is a rapid inflection in surface orientation, surface reflectance or illumination characteristics. In addition, there are often many "weak" zero-crossings, caused by small amounts of noise in the image, which appear randomly strewn across the image.

[5]Actually, they used $\nabla^2 G \otimes I$, a linear operator which they characterized as an approximately-bandpass filter. As shown above, this operator may also be thought of as calculating $\nabla^2 I$.

As mentioned in chapter 2, it appears that we can solve the problems of weak and inflection-caused zero-crossings by discarding those zero-crossings for which $\nabla^2 I$ has negligible slope across the zero-crossing contour. The threshold value for discarding weak zero-crossings is determined by the amount of noise in the system. In checking for such weak zero-crossings, it is important to consider the average slope of the zero-crossing along *segments* of the zero-crossing, rather than the slope individual points only, so that local fluctuations in image intensity will not cause perfectly valid contour points to be discarded. Zero-crossing contours may be usefully segmented by cutting them wherever sharp changes in contour direction occur.

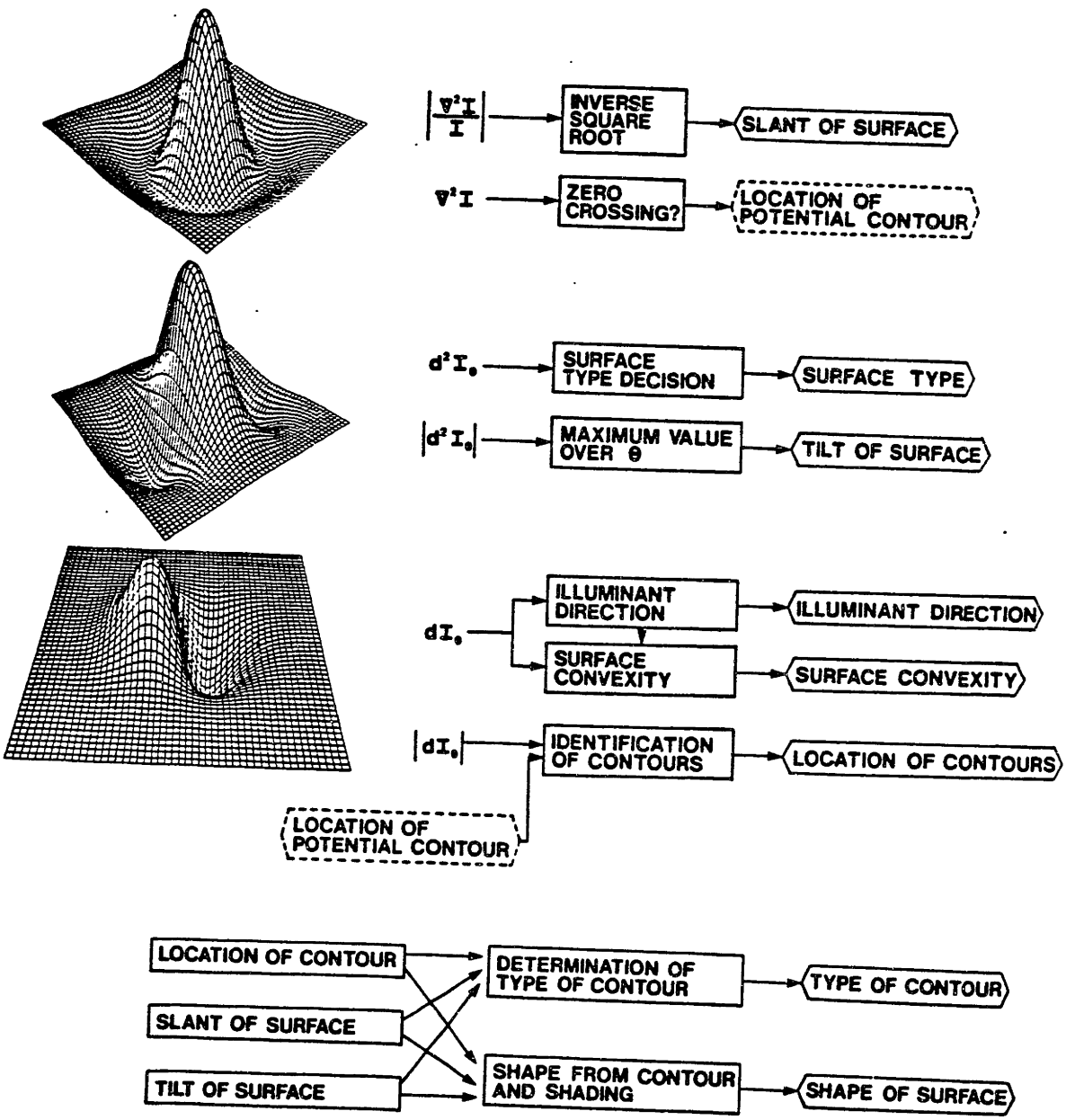


Figure 3-2. Block diagram incorporating all of the image measurement and estimation algorithms developed here into a single integrated system. See description of individual algorithms for explanation.

2 Description Of The Estimation Algorithm

Having described how to obtain the required input information, we can describe how to use this information to carry out the computations required by the theory. The estimation algorithms may be grouped into three "modules" on the basis of their input information; these three modules are (1) the shape-from-shading and surface type estimators, which use d^2I and $|\nabla^2 I|$, (2) the illuminant direction and surface convexity estimators, which use dI, and (3) the contour-and-shading estimators, which use the results of the previous two estimation modules together with imaged contour shape.

This section will describe each of these modules in turn. The final result is an integrated algorithm which combines of all of these image measurement and estimation algorithms into a single information processing system. The block structure diagram of this integrated system is shown in figure 3-2.

Shape-From-Shading And Surface Type Algorithms

Surface type. The theory for determining surface type is simple; it was proven that it only requires counting the number of image directions along which d^2I is zero to determine the surface type. If d^2I is zero along three or more directions, then the surface is a plane; if d^2I is zero along one direction then the surface is a cylinder, and if d^2I is non-zero for all image directions then the surface is convex, concave or a saddle surface. This portion of the algorithm is diagrammed in figure 3-3. These last three types (convex, concave and saddle surface) can be distinguished only by using additional information, such as knowledge of the illuminant direction. The separation of these three types is accomplished by the surface convexity portion of the algorithm.

The surface slant. The theory for estimation of surface orientation from local shading information proved that z_N, the surface slant (how much the surface is slanted away from the viewer or

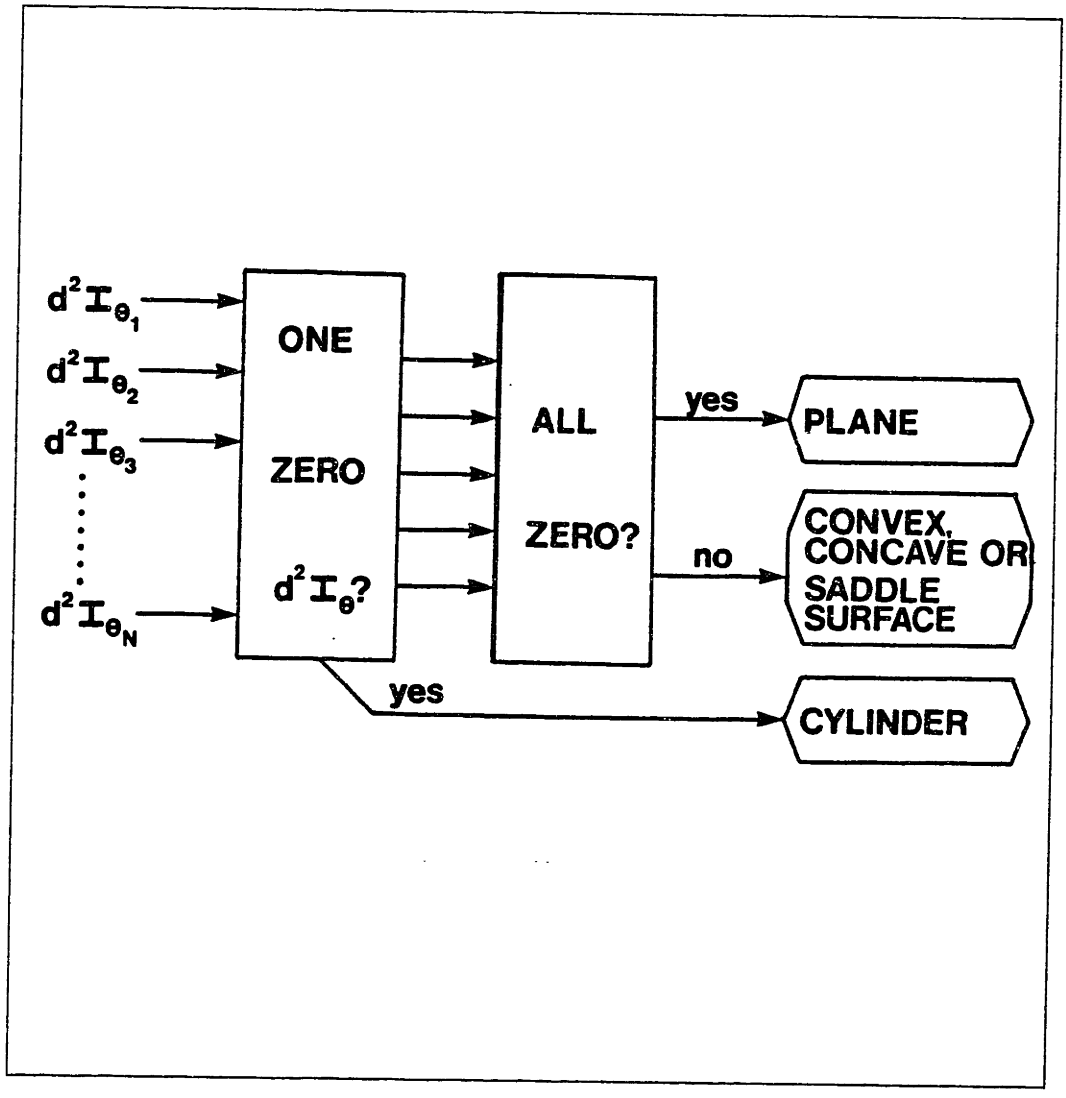


Figure 3-3. Surface type identification portion of the algorithm.

toward the viewer) may be estimated by

$$z_N = \sigma_\kappa \left(\left| \frac{\nabla^2 I}{I} \right| - \sigma_\kappa^2 \right)^{-\frac{1}{2}} \tag{16}$$

where σ_κ^2 is the variance of the distribution of surface curvatures. We have already shown how to calculate $|\nabla^2 I|$, thus the only difficulty left is to determine the constant σ_κ^2. This constant is a function of the size of the imaging systems blur function. Thus if one knew the blur function, one could determine σ_κ^2.

Alternatively, this constant may be calculated by making use of the constraint that $0 \leq z_N \leq 1$

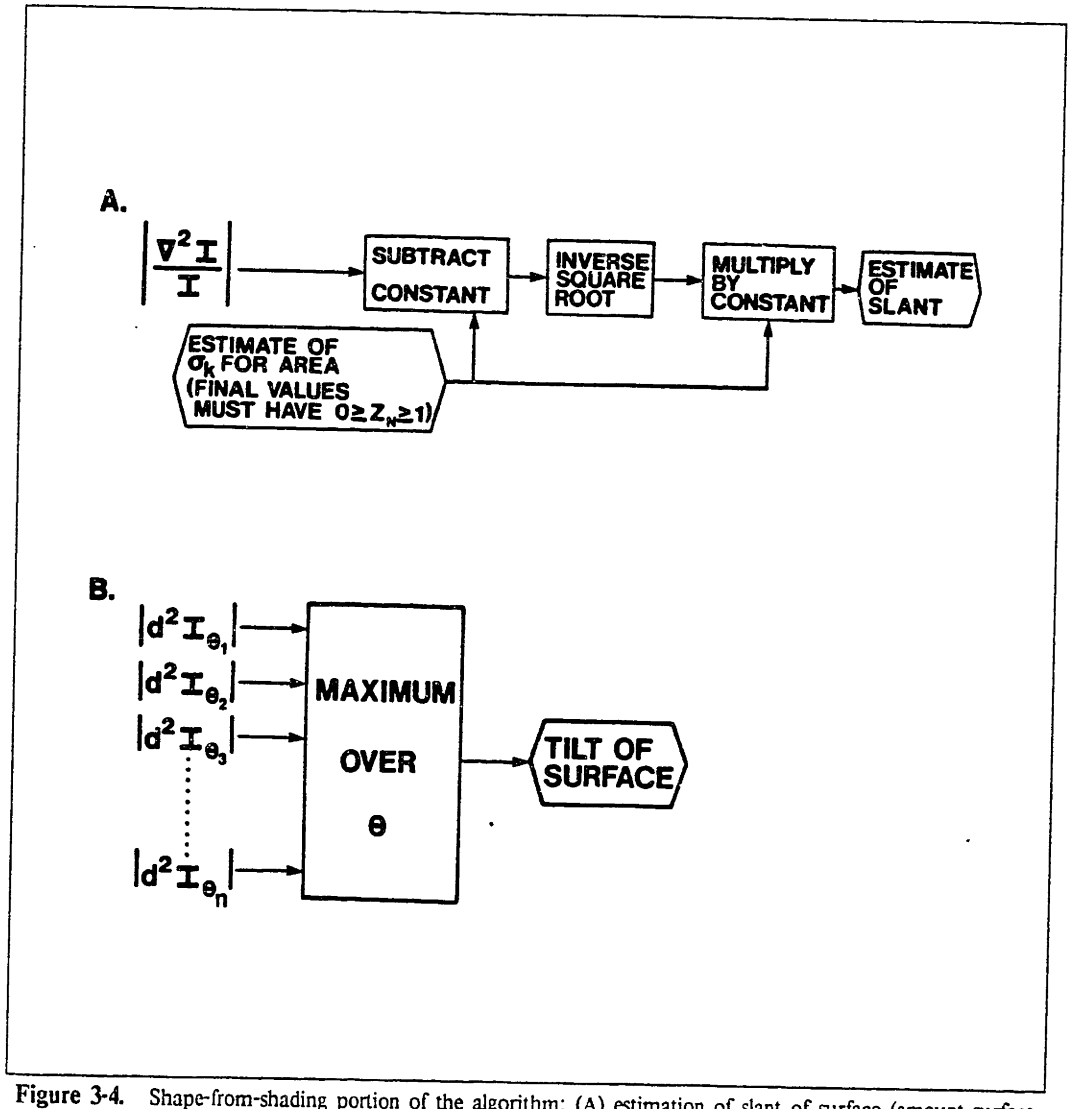


Figure 3-4. Shape-from-shading portion of the algorithm: (A) estimation of slant of surface (amount surface is oriented away from or toward the viewer) and (B) estimation of tilt of the surface (which image direction the surface is slanting away or toward the viewer).

and the identified smooth occluding contours (see the section on identifying smooth occluding contours). The maximum value of z_N within a region is one, and the minimum value ($z_N = 0$) must occur at the identified smooth occluding contours. The two resulting equations allow us to determine the value of σ_κ^2. Therefore an estimate of the surface slant is obtained by subtracting this constant (or adding it, if the surface is identified as a saddle surface) from $|\nabla^2 I/I|$, computing the inverse square root and multiplying by the constant $1/\sigma_\kappa$. This portion of the algorithm devoted to estimating the slant of the surface is diagrammed in figure 3-4 (A).

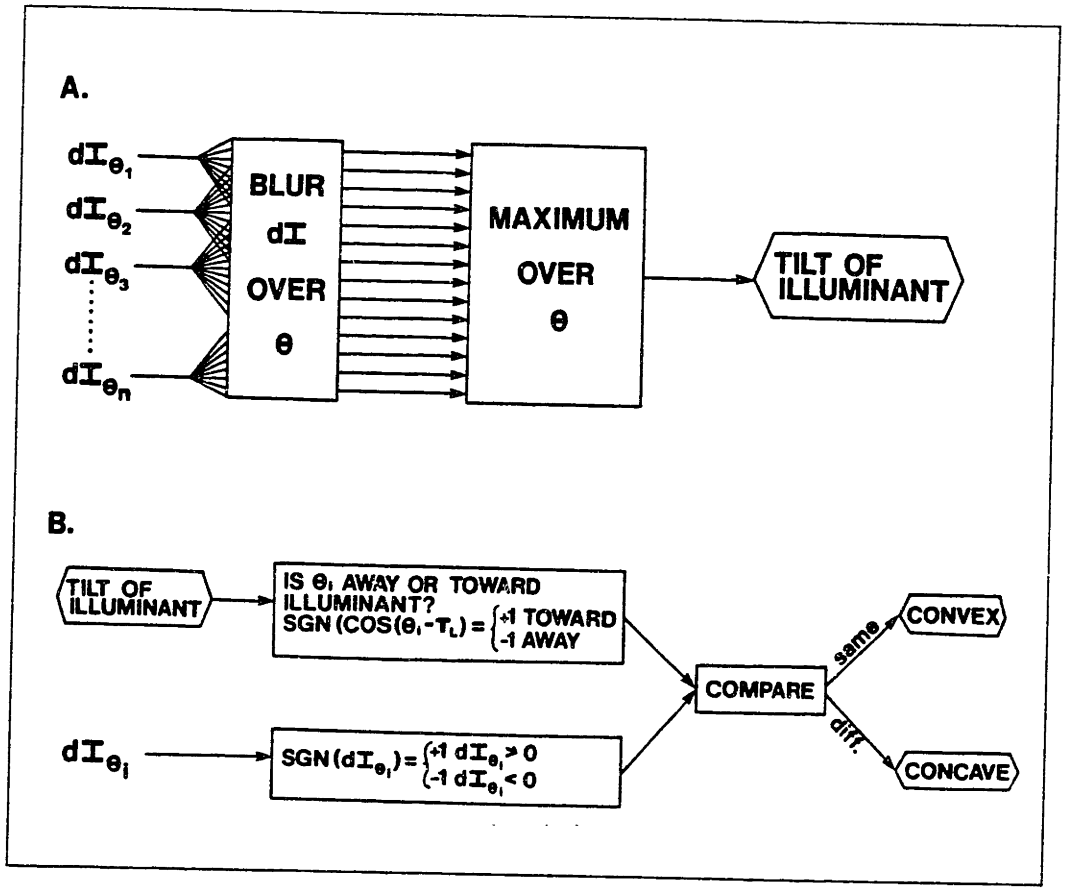


Figure 3-5. Illuminant direction and surface convexity portions of algorithm.

The surface tilt. In the theory chapter it was proven that the tilt of the surface (the direction along which the surface slants away from or toward the viewer) may be found by determining the direction, at each point, in which $|d^2I|$ is largest. Thus we need only compare $|d^2I|$ across image orientations to exactly determine the surface tilt at each point, as shown in figure 3-4 (B). The surface tilt and slant completely specify surface orientation at a point, therefore these two outputs at each point in the image constitute a complete description of surface shape.

Illuminant Direction And Surface Convexity Algorithms

Illuminant direction. The previous chapter developed a theory for the estimation of the illuminant direction. This theory consisted of two steps: first, calculate the mean value of dI along each image orientation, and second, determine the direction along which the mean of dI is greatest.

This direction is the tilt (the image-plane component of the direction) of the illuminant. Only determination of the tilt will be discussed here, as the tilt is all that is required by the other portions of the algorithm. The theory used a linear regression to determine the direction in which the mean of dI is maximum, because variations in surface shape introduce changes in the mean of dI which make it necessary to treat the mean of dI a noisy signal.

The mean value of dI along a particular image direction may be found by summing all of the appropriately-oriented dI-operators throughout the entire region for which the illuminant direction is to be estimated. We don't need to divide this sum by the number of image points, as the number of points is constant for all orientations.

The second step in estimating the illuminant direction is to find the direction for which the sum of dI is largest. If we blur the response profile of dI across orientation and then pick the maximum of the blurred response profile, we accomplish the same result as the regression used in the theory. The blurring smoothes out any random variations in the sums of dI, and allows us to pick the orientation with the maximum sum, and thus mean, of dI. The portion of the algorithm dealing with finding the illuminant direction is diagrammed in figure 3-5.

Surface convexity. The theory for estimating surface convexity along a particular direction (whether the surface is curving toward the viewer, or away from the viewer) stated that if dI was positive when measured along a direction which is toward the illuminant then the surface is convex[6], if dI is negative then the surface is concave. Similarly, if dI is negative when measured away from the illuminant, then the surface is convex, otherwise it is concave.

This comparison may be made by creating a set of reference signals which specify whether each direction is toward or away from the illuminant. Such a signal would be $+1$ if the direction is toward the illuminant, and -1 otherwise. Such a signal could then be compared to the sign of dI along the same direction to determine the convexity of the surface; if both dI and the reference signal have the same sign then the surface is convex, otherwise it is concave. This comparison operation is

[6]The 180° ambiguity in illuminant direction shown by both people and this algorithm results here in the "crater illusion", i.e., if we decide that the illuminant is actually in the mirror-image position then we get a reversal of convexity also.

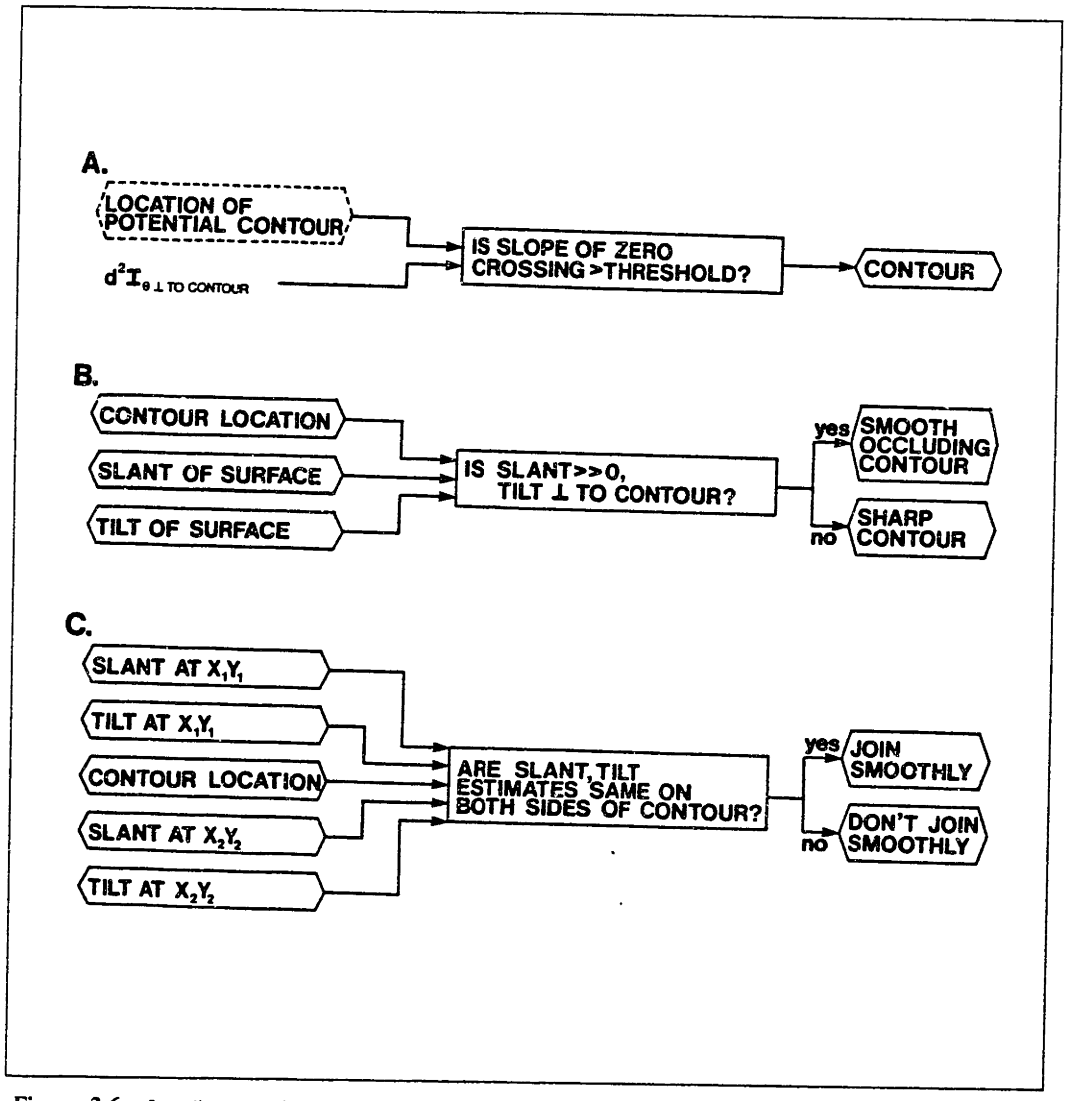


Figure 3-6. Localization of contours, detection of smooth occluding contours, determining whether regions should be joined together.

diagrammed in figure 3-5. To determine whether the surface is convex, concave or a saddle surface one only need examine the estimates of convexity along various directions at a point to see whether they are all convex, all concave, or mixed (implying a saddle surface).

Detecting Smooth Occluding Contours, Region-Joining Algorithms

The theories of detecting smooth occluding contours and of determining whether or not the regions on either side of a contour join together smoothly may be diagrammed and described in a manner similar to the above algorithms. Figure 3-6 shows diagrams of these portions of the algorithm.

However, the use of the constraint contour gives on shading to determine surface shape, and the estimation of three-dimensional shape from an imaged contour are not easily diagrammed, as they involve integration of information from several points along an imaged contour. In the next section of this chapter the workings of these algorithms will be illustrated by going through specific examples of their application.

Detecting smooth occluding contours. The theory developed in the previous chapter showed that we may determine whether or not a particular contour is a smooth occluding contour (such as found at the edge of the image of a sphere) by examining the surface orientation estimates made from local shading information along the contour. If the surface orientation estimates show that the surface is "rolling off" in the manner that defines a smooth occluding contour, then the contour may be taken to be a smooth occluding contour; otherwise it is not.

Thus determining whether a contour is a smooth occluding contour is straightforward if we have available reliable estimates of surface orientation. If the estimates (especially the slant) were known to have a significant variance, it would be desirable to be cautious about such identifications, because the mis-identification of a sharp contour as a smooth occluding contour can cause serious errors in later estimates of surface shape. However, if we adopt the strategy of identifying only those areas with the largest slant estimates and most appropriate tilt estimates as smooth occluding contours then we can obtain arbitrarily reliable identifications of smooth occluding contours, at the cost of failing to identify a significant portion of them. As long as a reasonable fraction of the smooth occluding contours are identified, there seems to be no great penalty incurred by failing to identify some of these contours, and no false information is introduced into the system[7] .

Joining surfaces together. In theory chapter it is shown that the determination of whether or not the two regions on either side of a contour join together smoothly can be made in the same

[7]In this algorithm the estimate of surface slant requires calculation of σ_κ^2, the variance of the distribution of surface curvatures. The calculation of this constant used here requires the identification of smooth occluding contours — which seems to require the estimation of slant. There is no problem however, as the largest values of $|\nabla^2 I/I|$ will correspond to the most slanted regions of the image regardless of the value of σ_κ^2, so that the most reliably-identified smooth occluding contours can still be found.

manner as the determination of whether or not a contour is a smooth occluding contour. The two regions on either side of an imaged contour *must* join together smoothly if the shape-from-shading estimates of surface orientation are the same along both sides of the contour. If the estimates are not the same, then the regions on either side of the imaged contour *cannot* join together smoothly. Thus if local estimates of surface orientation can be made along both sides of a contour[8] then we can determine with virtual certainty whether or not the two regions join together smoothly.

[8]I.e., there is shading information, and the surfaces are not planar or cylindrical.

3 Evaluating The Algorithm

In order to evaluate the performance of the algorithm it is important that it be tested on both natural and synthetic images. Synthetic images allow a carefully controlled, noise-free test of the algorithms, while natural images provide a critical test of the algorithms performance under real-world conditions.

The algorithm described in the previous section has been applied to a number of natural and synthetic images, allowing evaluation of the performance of the algorithm. This section presents and discusses several examples where the shading and contour information in an image were used to estimate the shape of the surfaces in the image. From these examples one may obtain a clear impression of the performance and capabilities of the algorithm.

In the following examples, the images for which shape was estimated were typically 50×50 pixels in size, with the largest being the Tuckermans Ravine image, which was 100×80 pixels. The filters which were convolved with these images were 11×11 pixels, giving a support of 121 pixels. The images for which the illuminant direction was estimated were approximately 200×200 pixels, and the images for which contours were localized were typically 400×400 pixels. The filters used for these images were typically 21×21 pixels, giving a support of 441 pixels.

3.1 Shape From Local Shading

The estimation of surface orientation from local shading information is an important part of the theory, and so deserves considerable examination. The slant (depth component) and tilt (image-plane component) of surface orientation have been estimated for synthetic images of a sphere and an ellipsoidal object (a "Hindenburg" shape), images of several natural objects, and an Electron Microscope (EM) image. The success at recovering surface orientation from these images indicates that local shading may be used to give significant information about surface shape.

Artificial Images. Synthetic images of a sphere and an ellipsoidal shape were used in order to

allow a careful examination of the local shape-from-shading estimators. These images are shown in figure 3-7 (A).

Because it is difficult to graphically show the surface orientation estimates for an object in such a manner that the correctness or incorrectness of the estimate is apparent, the surface orientation estimates produced by the algorithm (slant and tilt) were converted to another form ($p = dz/dx$ and $q = dz/dy$) and then were integrated, yielding map of the relative depth for the surface. This depth map was then converted into a line drawing which shows the relief of the surface, a graphic representation of the surface which has been found to give people a good impression of the estimated shape. Figure 3-7 (B) shows such relief maps, corresponding to the sphere and ellipsoidal shape images of figure 3-7 (A), and figure 3-7 (C) shows relief maps of the shape estimated using a local analysis of the shading information in the images of figure 3-7 (A). The steps of conversion to p, q and then integrating to form a depth map are *not* part of the theory under discussion here, but merely a step taken to provide the reader with an easily-interpreted graphic representation of the estimated surface shape.

The left portion of figure 3-7 (C) shows the surface shape estimate for the sphere image. It can be seen that for the sphere image this estimation strategy works well; in the error analysis section an example will be shown in which the deviation from spherical is less than 0.01%. The right side of figure 3-7 (C) shows the surface shape estimate for the image of the ellipsoidal shape, a figure with unequal surface curvatures. It can be seen that for this image the estimate of surface shape is qualitatively correct, although the overall relief of the surface is somewhat underestimated. In the ellipsoidal image of figure 3-7 (A) the arrangement of unequal surface curvatures somewhat cancels the effects of the foreshortening, thus any local analysis of the shading "sees" somewhat less foreshortening in the shading than it would if the curvatures were equal. As predicted by the theory, this partial cancellation leads to underestimation of the slant of the surface. Because of this intrinsic underdetermination of shape by shading (see section 2 of chapter 2), any local analysis of shading information will produce such an underestimation of the relief of the surface (chapter 4 discusses the fact that people also

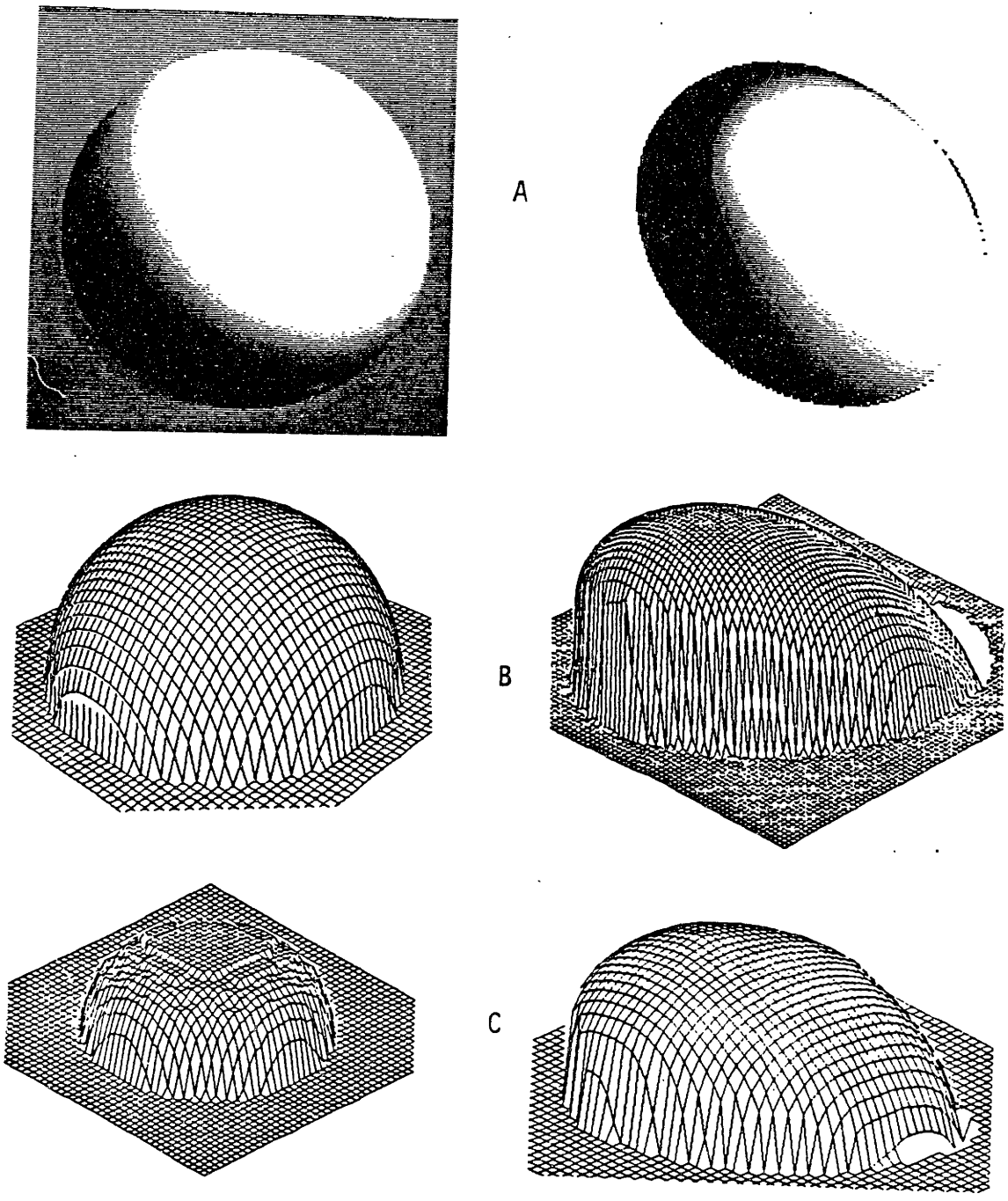


Figure 3-7. (A) Images of a sphere and an ellipsoidal object. (B) Relief maps showing a side view of the shapes of the objects whose images appear in (A). (C) Result of applying shape-from-local-shading algorithms to estimate shape for the images in (A). *Left side.* Shape estimate for the sphere image. The shape is successfully recovered. Note that illuminant direction and surface albedo are not used, and thus need not be known. *Right side.* Shape estimate for the image of the ellipsoidal object. The estimated shape is qualitatively correct, and corresponds closely to the shape perceived by human subjects. The slant of the estimated surface is somewhat underestimated because the unequal surface curvatures partially cancel the effects of foreshortening, as predicted by the theory.

underestimate the relief in this situation).

These two synthetic images conform to the assumptions of the theory. The fact that the slant and tilt estimates are as predicted by the theory demonstrates that the mathematics of the theory are fundamentally sound. Making conclusions about the general applicability of the theory on the basis of testing on synthetic images has a significant potential problem, namely that the synthetic images may not be typical of natural images in some critical manner. It does not appear that the assumptions of the theory are overly restrictive (see section 3 of chapter 1), however the only real test is to apply the estimation methods to images of real-world scenes.

Natural Images. In order to test the performance of the algorithm on images of real-world scenes, four natural images were selected for processing: a picture of a log, a picture of a rock, a close-up picture of a young woman (Lisa), and a mountain scene, the headwall at Tuckermans Ravine[9] . Digitized versions of the rock and log pictures are shown in figure 3-8 (A) and 3-8 (B). Figure 3-8 (C) shows a picture of Lisa, and the digitized close-up view of her face for which a shape estimate was made. Figure 3-8 (D) shows a picture of Tuckermans Ravine, and and the digitized image of the headwall for which an estimate of shape was made.

As described previously, the estimates of slant and tilt cannot be displayed in such a manner that the correctness or incorrectness of the estimates is apparent. Therefore the slant and tilt estimates were converted to dz/dx and dz/dy and integrated to obtain the relative depth of the surface, which can be satisfactorily displayed. This process in illustrated in figures 3-9 and 3-10.

Figure 3-9 shows (A) the digitized log image, (B) a relief map of the estimated slant of the surface, and (C) a relief map showing a side view of the surface shape estimated for the image, produced by integrating the estimated slant and tilt of the surface to obtain relative depth. The estimated slant of the surface is shown by the relief map in 3-9 (B); the higher areas of the surface shown in this figure correspond to portions of the image in which the surface is facing the viewer, while the lower portions

[9]Tuckermans Ravine is one of two major ravines on Mt. Washington, a member of the Presidential range in New Hampshire and the tallest mountain in the eastern United States. Tuckermans ravine is famous for downhill skiing, it is generally considered the steepest slope (averaging 60° for long distances) commonly skied on this continent.

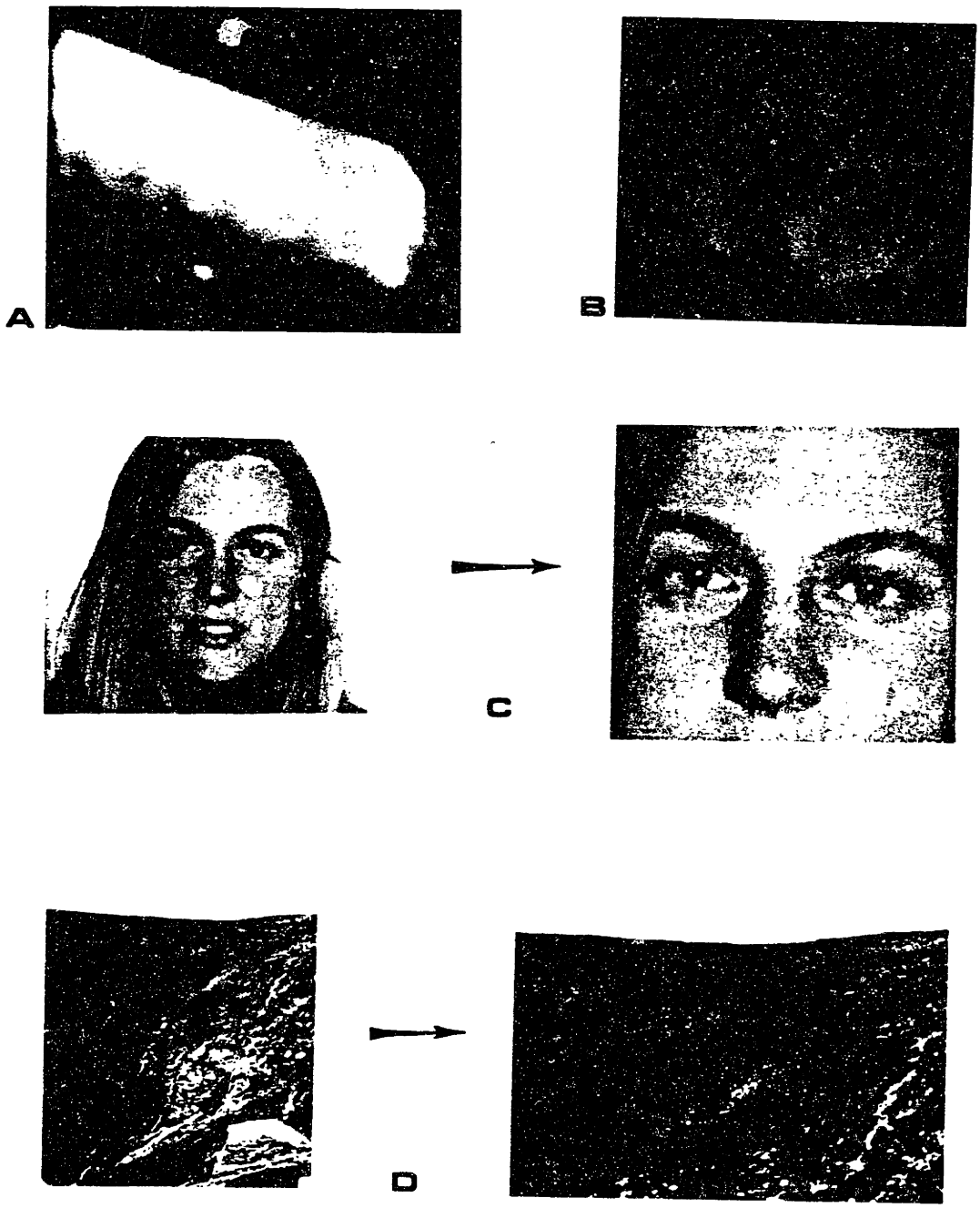


Figure 3-8. (A) Digitized image of a log. (B) Digitized image of a rock. (C) A picture of Lisa, and the digitized portion of this image for which shape was estimated. (D) A picture of Tuckermans ravine, and the digitized portion of this image (the headwall) for which an estimate of shape was made.

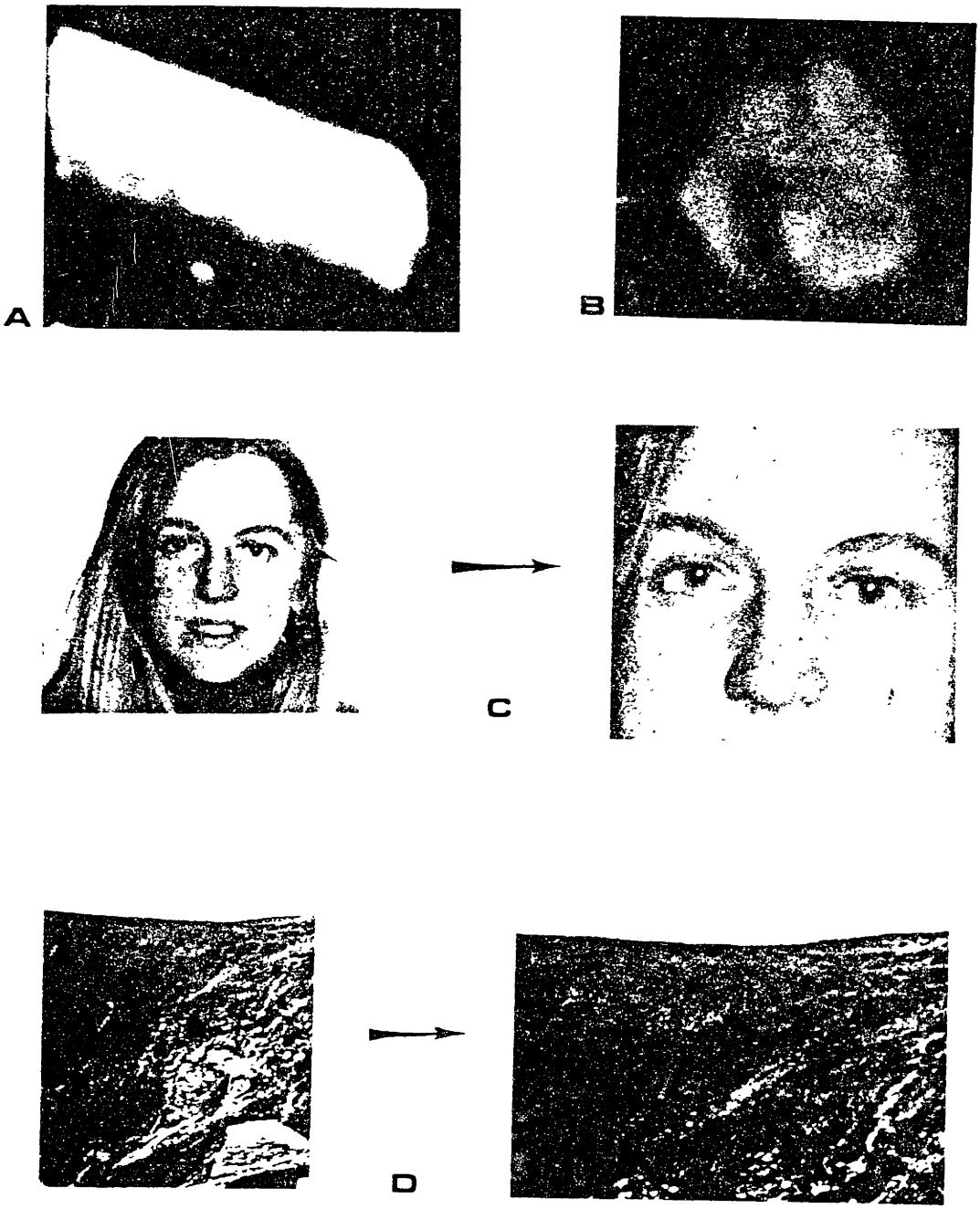


Figure 3-8. (A) Digitized image of a log. (B) Digitized image of a rock. (C) A picture of Lisa, and the digitized portion of this image for which shape was estimated. (D) A picture of Tuckermans ravine, and the digitized portion of this image (the headwall) for which an estimate of shape was made.

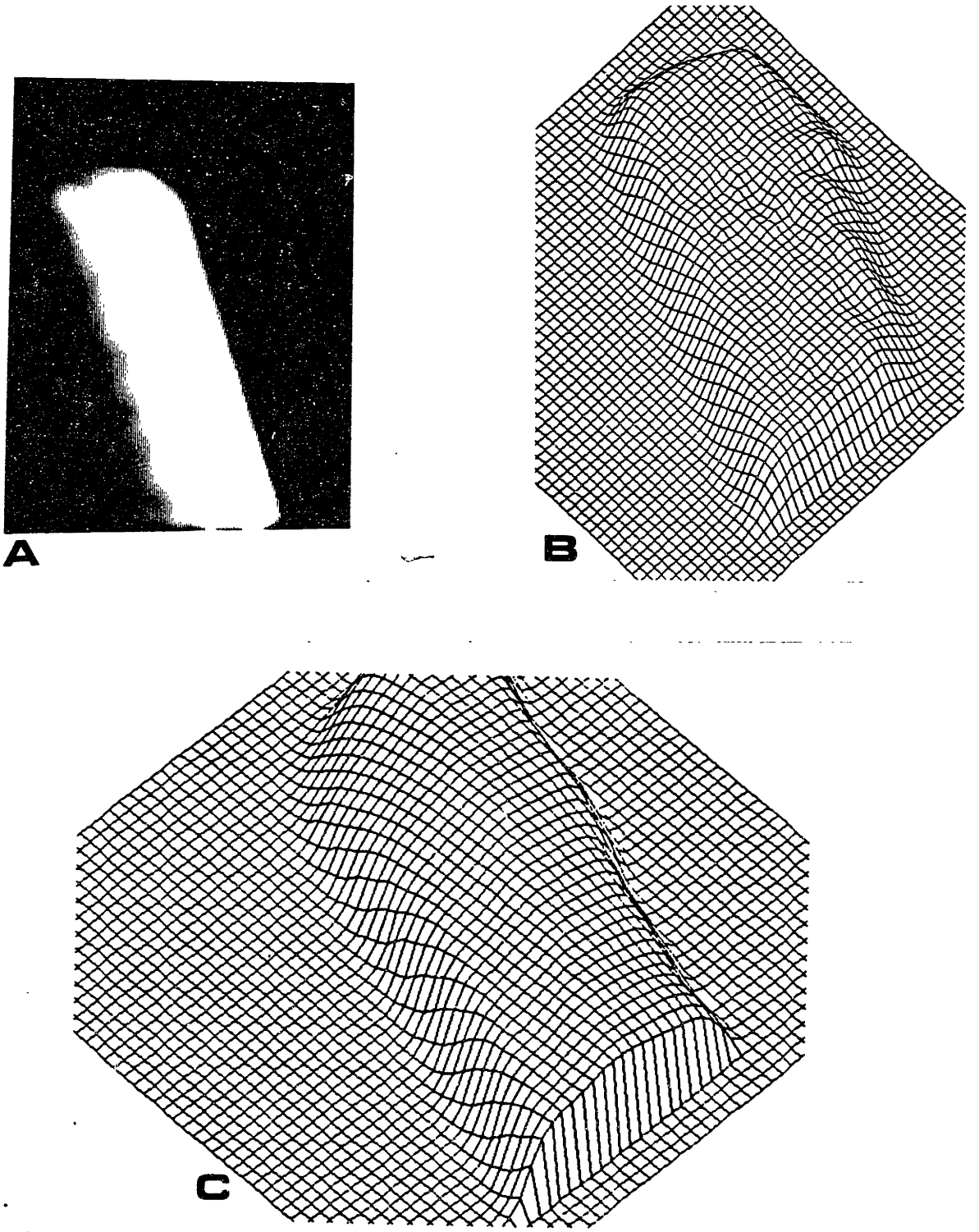
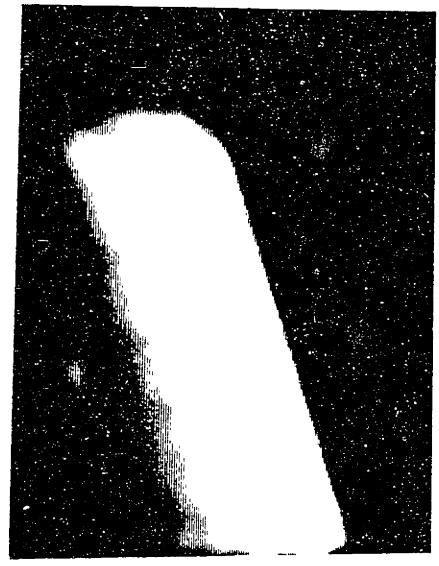
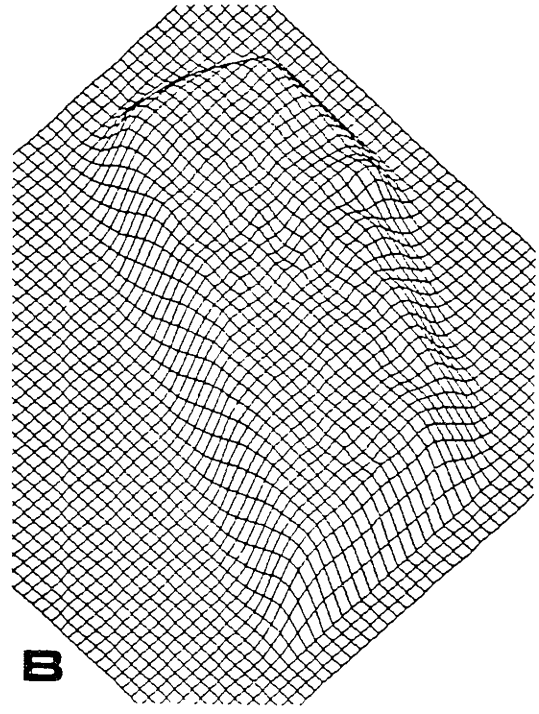


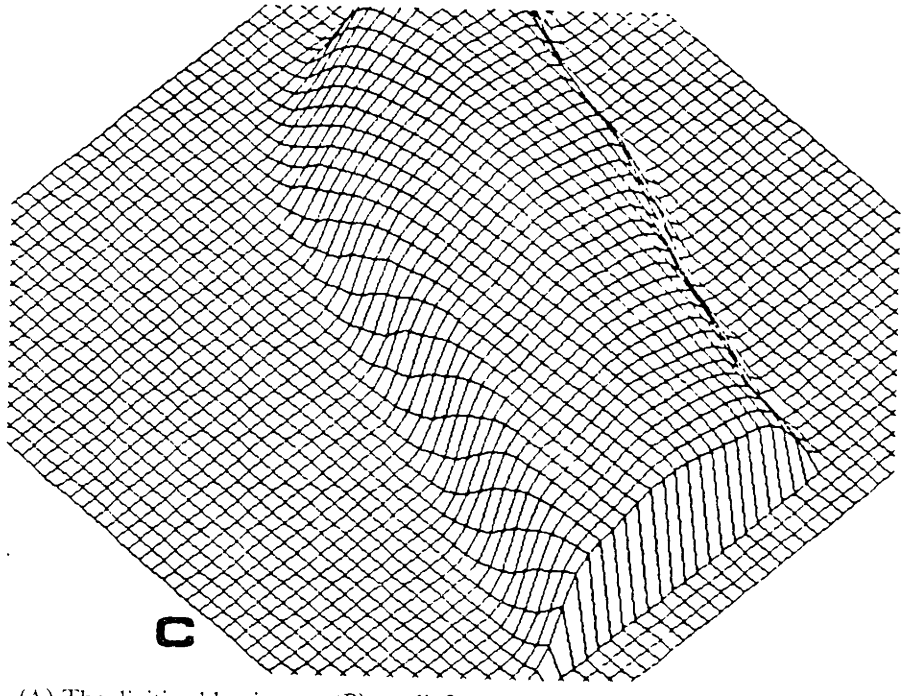
Figure 3-9. (A) The digitized log image, (B) a relief map showing the estimated slant of the surface (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer), and (C) a relief map showing a side view of the surface estimated for this image.



A



B



C

Figure 3-9. (A) The digitized log image, (B) a relief map showing the estimated slant of the surface (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer), and (C) a relief map showing a side view of the surface estimated for this image.

correspond to regions in which the surface slants away from the viewer. Figure 3-9 (C) shows a side view of the estimated surface shape; it is clear from this figure that this estimated shape is the correct one.

The slant map of figure 3-9 (B) gives the impression of containing all the information needed to recover the shape of the log; it seems that we only need to "puff up" this slant map to obtain the surface shape shown in figure 3-9 (C). If the surface has no "twist" in it, i.e., that $\|\mathbf{p}_{uv}\| = 0$ everywhere on the viewed surface, then the slant map shown in figure 3-9 (B) is by itself a sufficient representation of the surface shape. This fact may be important in understanding the representations of shape built by the human visual system even though this condition is is not true of all surfaces. Examples of surfaces without such twist are shown in figure 1-6 (A) - (D), while the surfaces in figure 1-6 (E) and (F) demonstrate surfaces with considerable twist in them. Note that we have already made the assumption that $\|\mathbf{p}_{uv}\|$ is negligible in developing equation (10), the expression for d^2I.

Figure 3-10 shows (A) the digitized rock image, (B) a relief map of the slant estimated for the surface, and (C) two views relief map showing the surface shape estimated for the rock image, as obtained by integrating the slant and tilt estimates. Again, it is apparent that the estimated surface shape corresponds to the human percept of shape for these images. The estimated surface shape is also quite accurately the objectively correct surface shape. The fact that shape is successfully estimated for both the log and rock images, neither of which have uniform albedo, Lambertian reflectance function or distant point-source illuminant, demonstrates the robustness of the theory.

Figure 3-11 shows (A) the digitized picture of Lisa, and (B) a relief map of the surface slant estimated for the Lisa picture. No relief map of the estimated surface shape is shown because the step of integrating slant and tilt estimates (a step which is not part of this theory, but done only to allow the reader to evaluate the estimated shape) proved difficult to do because of the complexity of the surface shape. If a relief map was desired for such complex surfaces, it would be better to use a relaxation-type integration algorithm rather than the one-pass, serial algorithm employed here. See Rosenfeld and Kak 1976. As before, in the slant-map representation regions with higher relief face toward the

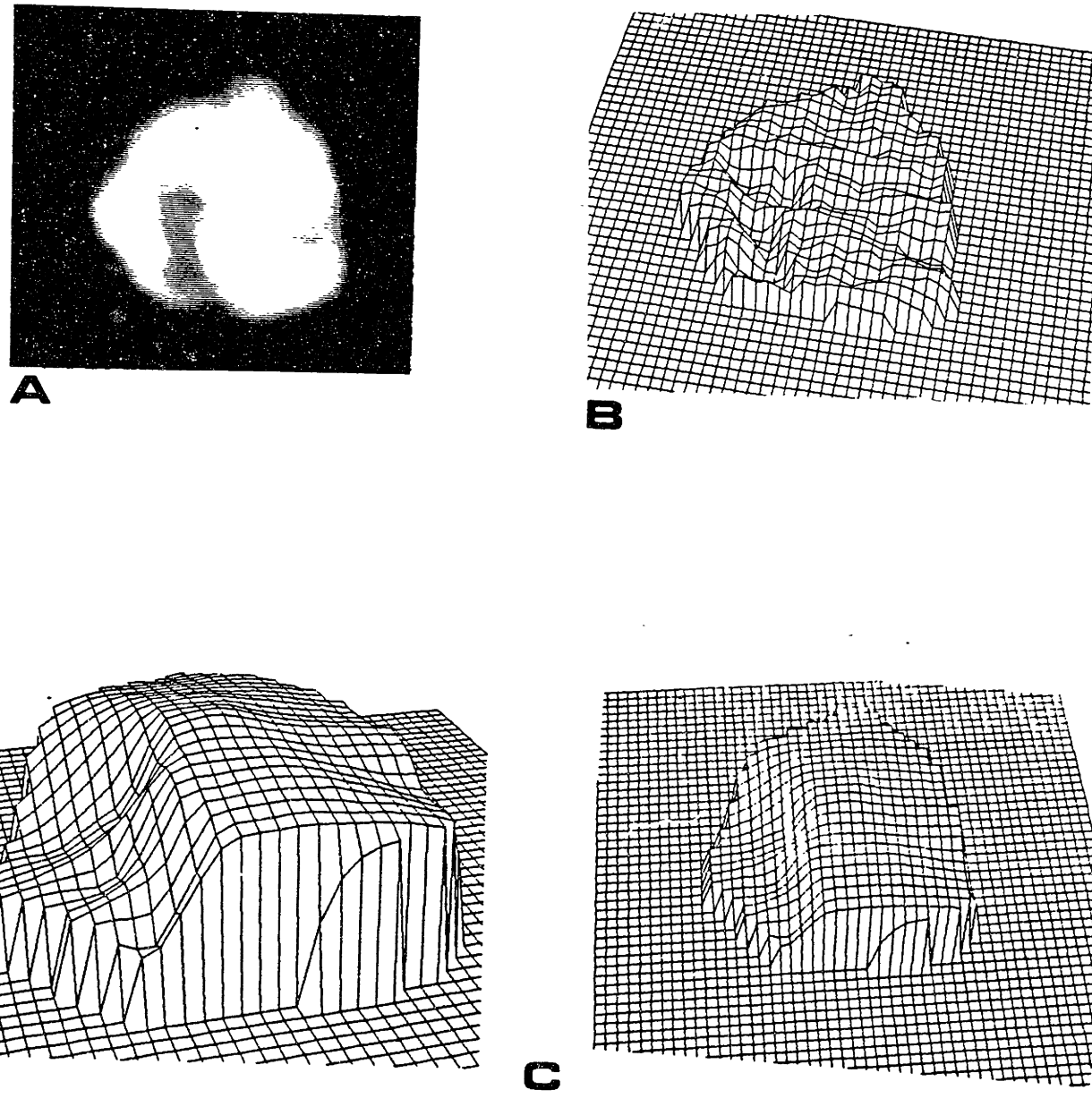


Figure 3-10. (A) The digitized rock image, (B) a relief map showing the estimated slant of the surface (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer), and (C) a relief map showing a side view of the surface estimated for this image.

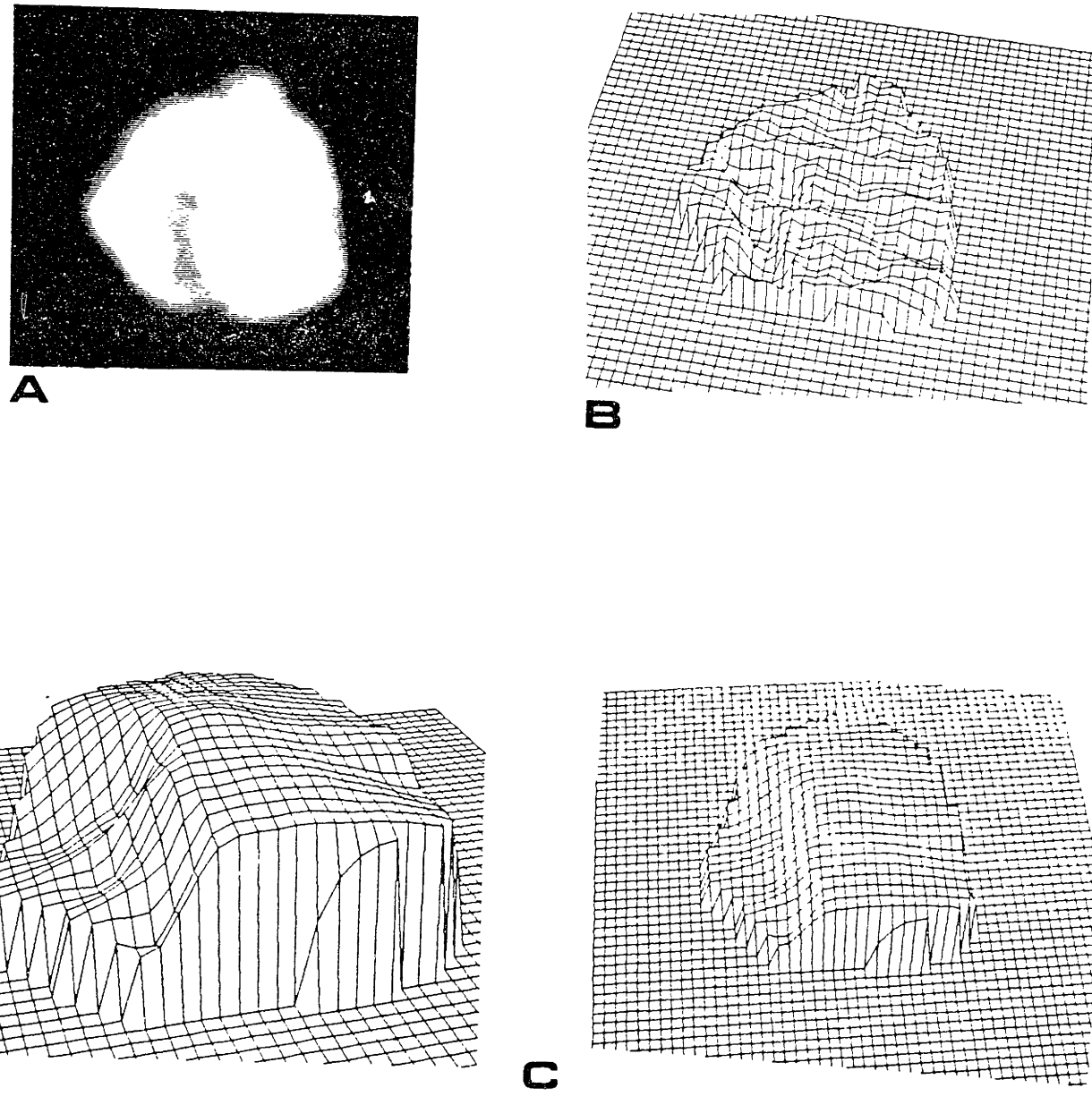


Figure 3-10. (A) The digitized rock image, (B) a relief map showing the estimated slant of the surface (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer), and (C) a relief map showing a side view of the surface estimated for this image.

viewer, lower relief regions face away from the viewer. Note that again, as with figures 3-9 and 3-10 all important details of the surface shape seem to be available in this representation; for instance, the structure of the nose and eyebrow ridges is plainly visible. Thus it seems that for human faces the slant estimate alone might serve as a reasonable representation of surface shape.

Figure 3-12 shows (A) the digitized image of Tuckermans ravine, (B) a relief map showing a side view of the estimated surface shape, obtained by integrating the slant and tilt estimates, and (C) a topographic map of the Tuckermans ravine area in which the imaged area is outlined and the camera position marked, together with a relief map showing the estimated surface shape as it would be seen from above. This relief map may be directly compared to the outlined area of the topographic map. Comparing the topographic map with the estimated surface shape, we see that the roll-off at the top of figure 3-12 (B) and the steepness of the estimated surface are correct for this surface — this area of the ravine has a slope which averages 60°. And people ski down this!

Looking at the original image, it seems that the lower right portion of the image has considerably more relief than is estimated. In comparing the estimated shape with the outlined area in the topographic map, we see that there is indeed greater relief in the lower right-hand of the image, confirming our visual impression. The underestimation of relief in this portion of the image when using only shading information will allow a demonstration of the important role of imaged contours in determining shape. In the lower right portion of the image is a region of markedly different albedo, which in fact generates a contour as defined in this algorithm. In order to produce a correct estimate of shape for this image, it is necessary to locate the contours which bound this region, and then use them to constrain the estimate of shape. This will be accomplished in the following sections of this chapter.

An Electron Microscope image. In addition to natural images, an electron microscope (EM) image was selected. Ikeuchi and Horn 1981 have measured the reflectance function for this image, and found that the image intensities may be reasonably well described by

$$I = k(\mathbf{N} \cdot \mathbf{V})^{-1}$$

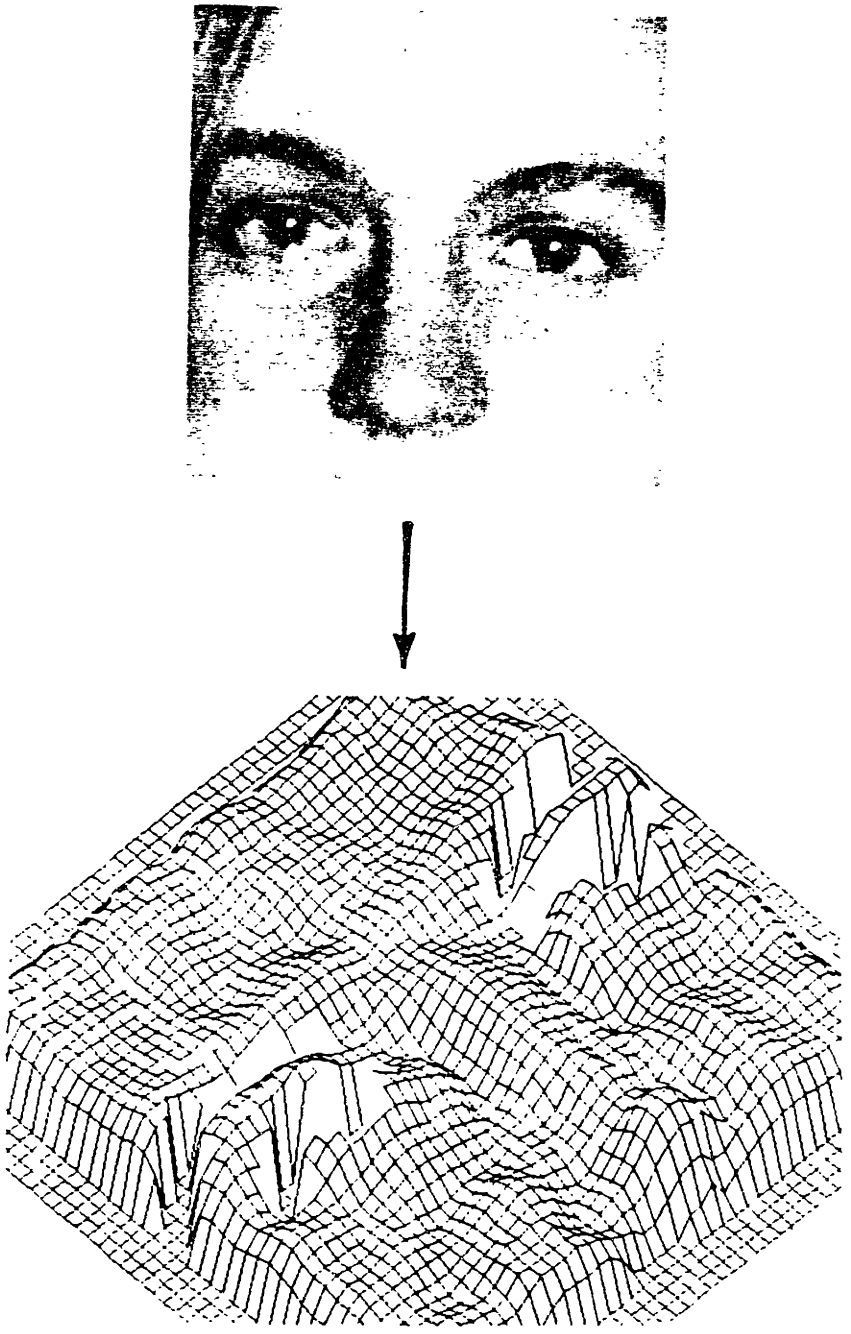


Figure 3-11. (A) The digitized image of Lisa and (B) a relief map showing the estimated surface slant in this image (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer). The integration of estimated slant and estimated tilt to show surface shape proved impossible because of the complexity of this surface.

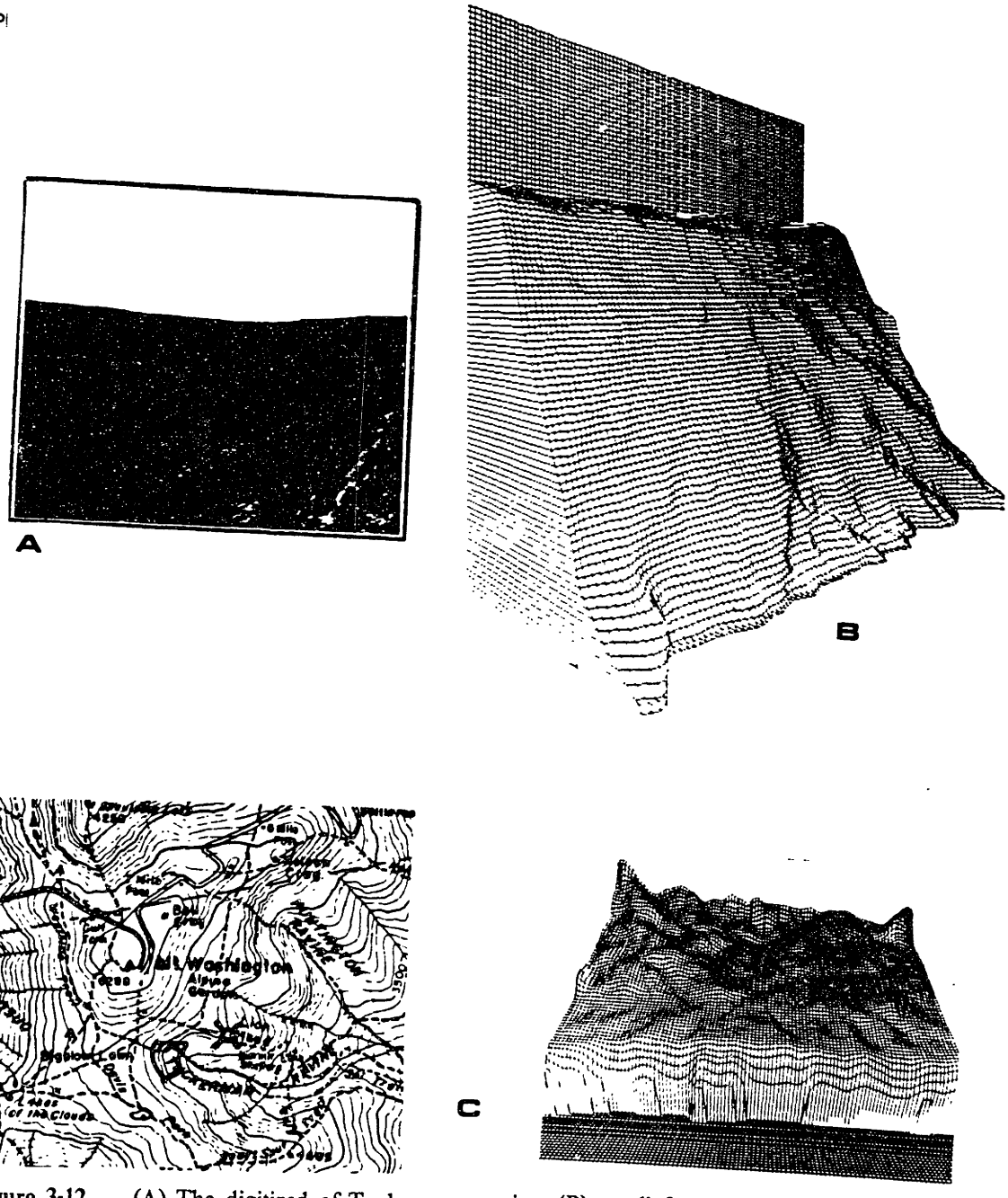


Figure 3-12. (A) The digitized of Tuckermans ravine, (B) a relief map showing a side view of the surface estimated for this image, and (C) a topographic map of the Tuckermans ravine area in which the imaged area is outlined and the camera position marked with a small dot, together with a relief map showing the estimated surface shape as seen from above. This relief map may be directly compared to the outlined area of the topographic map. Comparing the topographic map with the estimated surface shape, we see that the roll-off at the top of (B) and the steepness of the estimated surface are correct for this surface. Comparing the estimated surface to the topographic map of the area, or to our own impression of the shape of the surface, we see that the relief of the lower right-hand portion of the image is underestimated. We must use both the shading and the contour information in this image before the relief of this portion of the image can be correctly recovered. The use of the contour information to constrain the estimate of shape is accomplished in the following sections.

P

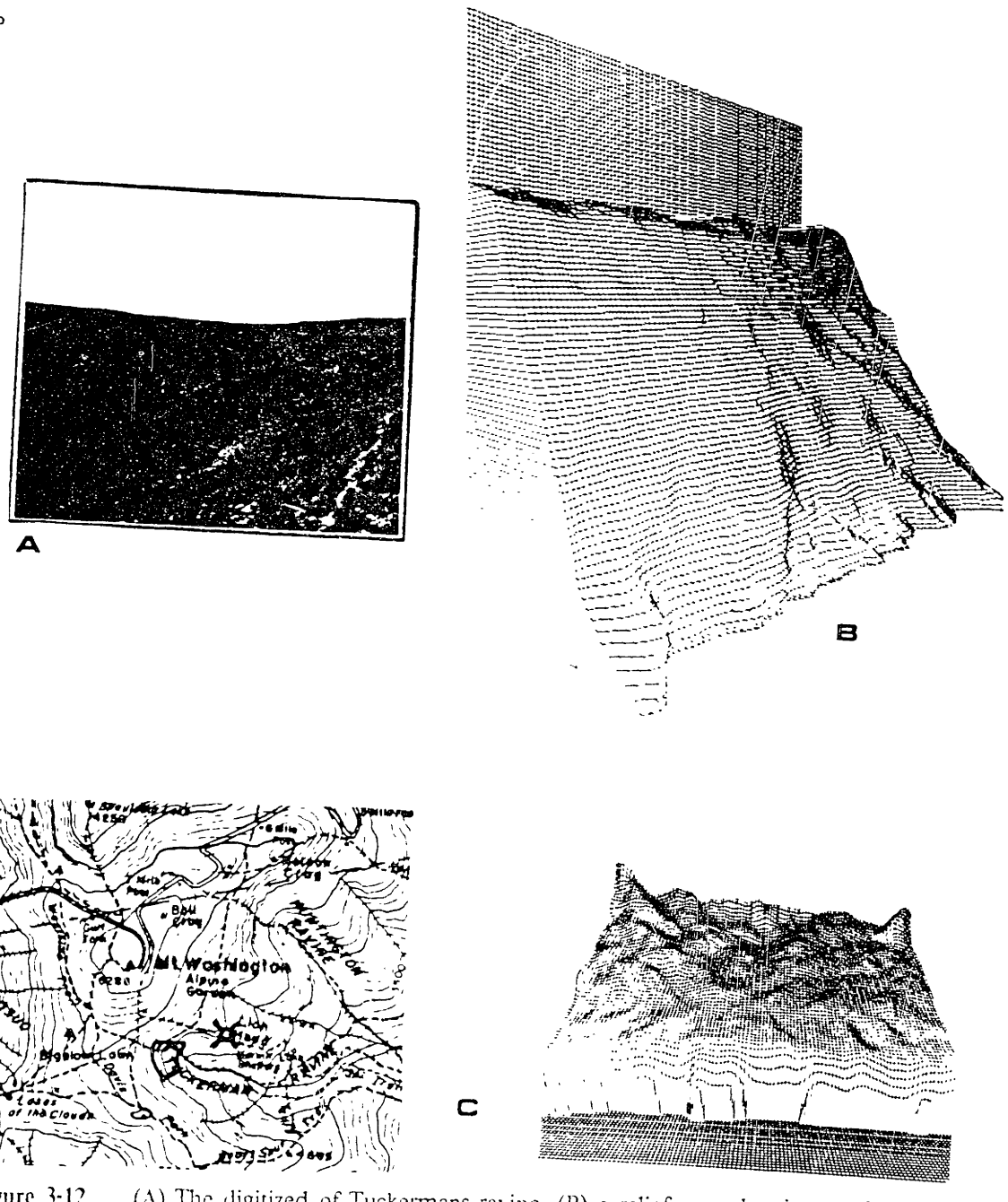


Figure 3-12. (A) The digitized of Tuckermans ravine, (B) a relief map showing a side view of the surface estimated for this image, and (C) a topographic map of the Tuckermans ravine area in which the imaged area is outlined and the camera position marked with a small dot, together with a relief map showing the estimated surface shape as seen from above. This relief map may be directly compared to the outlined area of the topographic map. Comparing the topographic map with the estimated surface shape, we see that the roll-off at the top of (B) and the steepness of the estimated surface are correct for this surface. Comparing the estimated surface to the topographic map of the area, or to our own impression of the shape of the surface, we see that the relief of the lower right-hand portion of the image is underestimated. We must use both the shading and the contour information in this image before the relief of this portion of the image can be correctly recovered. The use of the contour information to constrain the estimate of shape is accomplished in the following sections.

where k is approximately 0.8. If we carry out the required computations, we see that the tilt is still the direction along which d^2I is greatest, and the z component of the surface normal is approximately proportional to $\nabla^2 I/I$, as in normal images[10] . Therefore we can expect to obtain a reasonable shape estimate for EM images using the estimation techniques we developed for normal images. Figure 3-13 (A) shows this image and the portion of the image for which shape was estimated. People can use the shading information in EM images to perceive shape, as figure 3-13 (A) shows. This fact is surprising, because these images look very much like photographic negatives, and people are very poor at using the shading in negatives to determine shape. This image therefore is a critical test of the similarity between the human use of shading and this theory.

Figure 3-13 (B) shows a relief map of the slant estimated for the surface, and figure 3-13 (C) a relief map showing a side view of the estimated surface shape, again obtained by integrating the slant and tilt estimates for this image. It can be seen that the estimated surface shape is quite accurate. The success of the shape estimation theory on this image is of special significance because of the bizzare nature of EM images. The fact that both people and the theory presented here function successfully on such an unnatural image and not on photographic negatives (which are superficially similar to EM images) has important implications for understanding the human perception of shape, as discussed in chapter 4.

[10]See the discussion of EM images in chapter 4, especially figure 4-3.

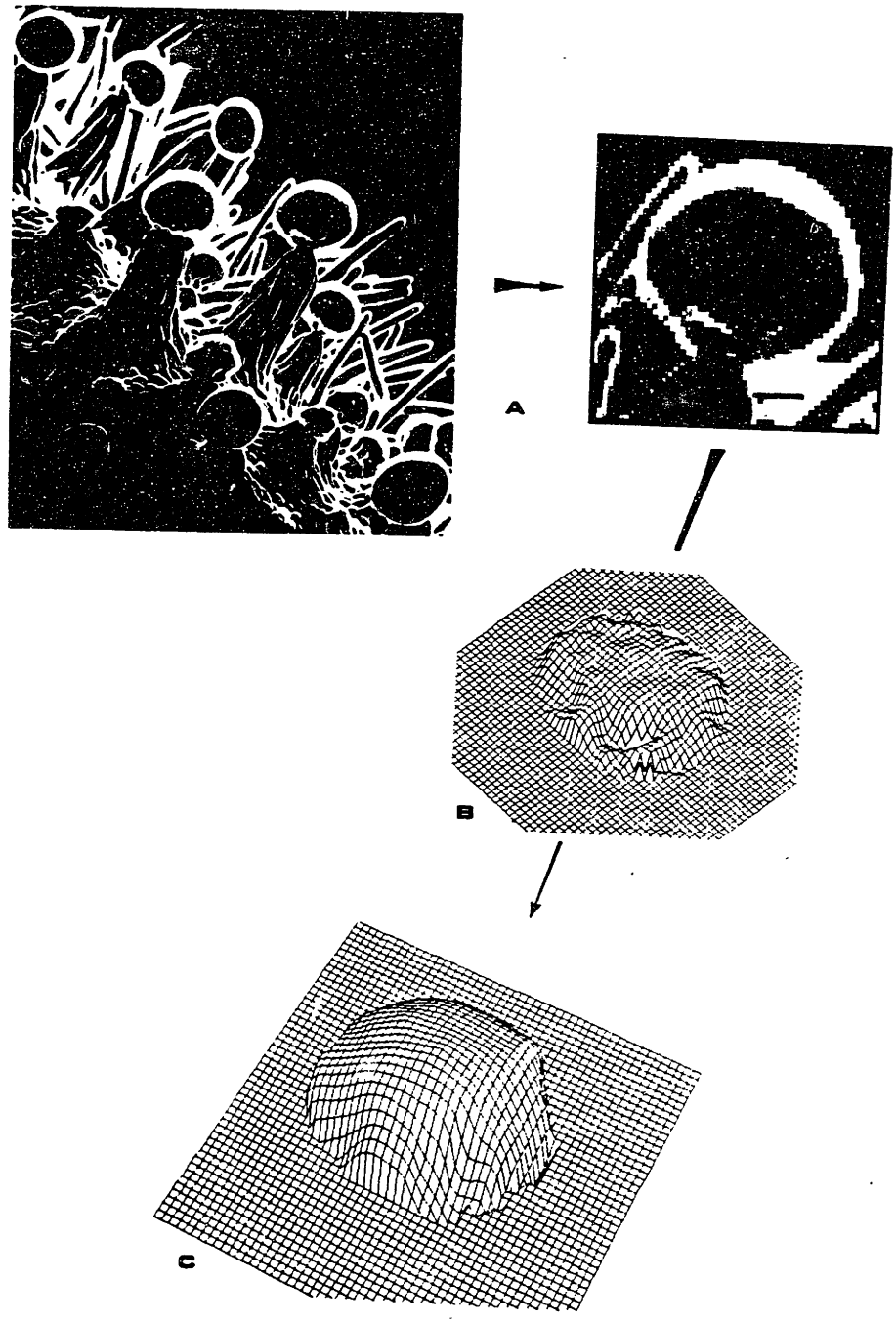


Figure 3-13. (A) An Electron Microscope image of resin nodules on a flower of *cannabis sativa*, and the portion of the image for which shape was estimated, (B) a relief map showing the estimated slant of the surface (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer), and (C) a relief map showing a side view of the surface estimated for this image. The fact that both people and this algorithm can correctly use the shading information in Electron Microscope images, but not in photographic negatives (which are superficially similar to EM images), has important implications for understanding human vision.

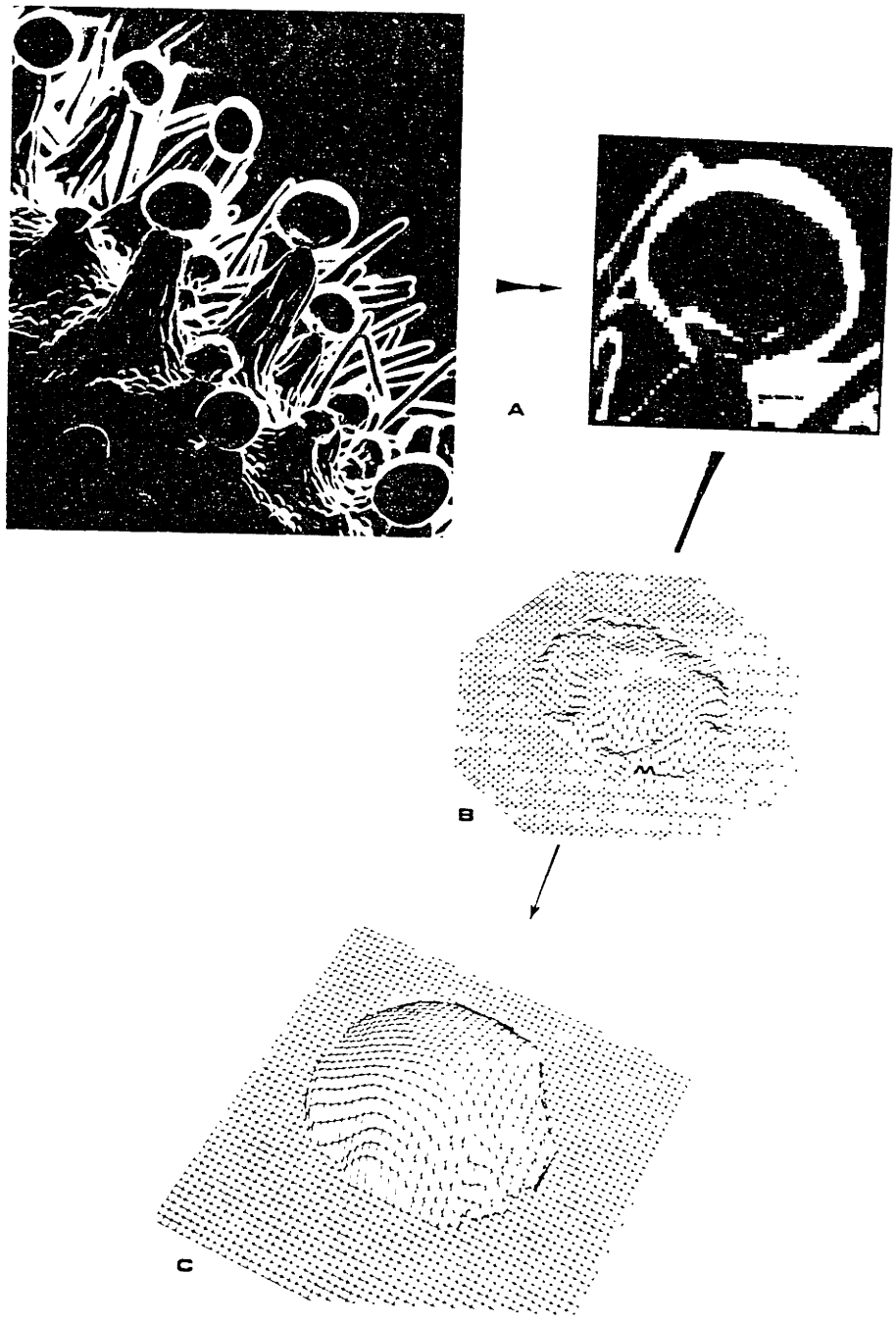


Figure 3-13. (A) An Electron Microscope image of resin nodules on a flower of *cannabis sativa*, and the portion of the image for which shape was estimated, (B) a relief map showing the estimated slant of the surface (high-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer), and (C) a relief map showing a side view of the surface estimated for this image. The fact that both people and this algorithm can correctly use the shading information in Electron Microscope images, but not in photographic negatives (which are superficially similar to EM images), has important implications for understanding human vision.

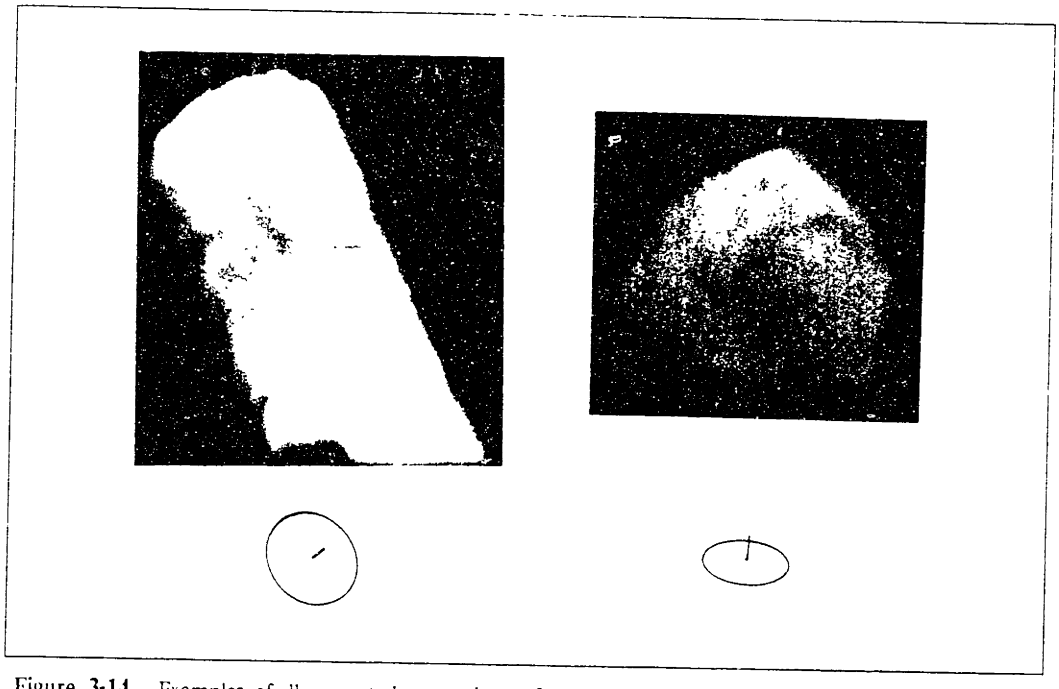


Figure 3-14. Examples of illuminant direction being found for digitized images of natural objects (a rock and a log). The ellipses show the estimated illuminant direction.

3.2 Finding The Illuminant Direction

The illuminant direction portion of the algorithm has been applied to several synthetic and natural images (Pentland 1980, 1982a, 1982b). On synthetic images of a sphere illuminated from various directions, the algorithm performs correctly — as it should in instances where the image fits the assumptions of the theory perfectly. On natural images the estimated illuminant direction was compared to the direction of the sun. Figure 3-14 shows two examples of digitized images of natural objects (a rock and a log), together with two ellipses which show the estimated light direction.

In every natural image but one, the estimated and actual illuminant directions are the same to within experimental accuracy (about $\pm 15°$). In the one case in which the illuminant direction was erroneously estimated, people also err in their judgment of illuminant direction — in fact both the algorithm and people give the same estimate of the illuminant direction for this image. The next chapter discusses the similarity between human perception of the illuminant and the estimate produced by this algorithm.

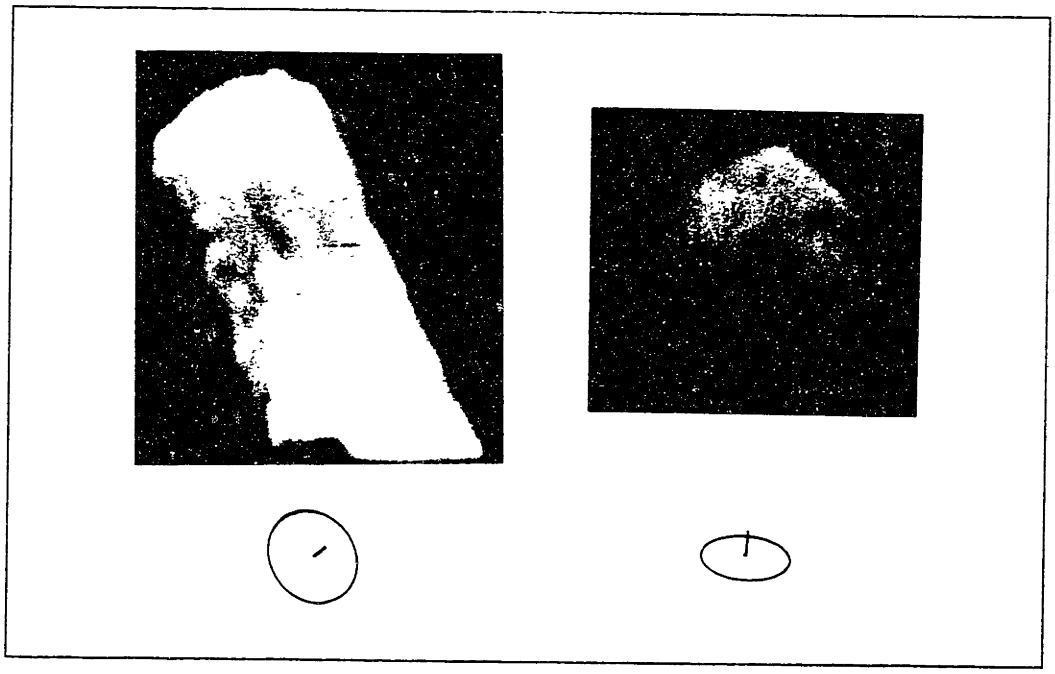


Figure 3-14. Examples of illuminant direction being found for digitized images of natural objects (a rock and a log). The ellipses show the estimated illuminant direction.

3.2 Finding The Illuminant Direction

The illuminant direction portion of the algorithm has been applied to several synthetic and natural images (Pentland 1980, 1982a, 1982b). On synthetic images of a sphere illuminated from various directions, the algorithm performs correctly — as it should in instances where the image fits the assumptions of the theory perfectly. On natural images the estimated illuminant direction was compared to the direction of the sun. Figure 3-14 shows two examples of digitized images of natural objects (a rock and a log), together with two ellipses which show the estimated light direction.

In every natural image but one, the estimated and actual illuminant directions are the same to within experimental accuracy (about $\pm 15°$). In the one case in which the illuminant direction was erroneously estimated, people also err in their judgment of illuminant direction — in fact both the algorithm and people give the same estimate of the illuminant direction for this image. The next chapter discusses the similarity between human perception of the illuminant and the estimate produced by this algorithm.

3.3 Location And Identification Of Contours

In order to apply the shape-from-shading techniques used in previous sections of this chapter, we must first identify regions of differing albedo. After the boundary between such regions has been found and its type (smooth occluding, surface marking, etc) determined, we may then estimate shape separately for each region and then "hook them together" properly. In addition, the imaged contour within such a region is important in the estimation of shape, as shown by the theory and demonstrated by the Hindenburg illusion, presented in the introduction to this thesis. Thus it is of great importance that we be able to (1) locate imaged contours, and (2) determine whether or not they are smooth occluding contours (the type of contour found at the edge of an image of a sphere).

The second section of this chapter describes algorithms for locating contours, and then determining whether or not the contour is a smooth occluding contour. These algorithms have been applied to several synthetic images, and have performed admirably. However the location and identification of contours is well known to be much easier in synthetic images than in in natural images. Therefore several natural images have been processed, as a critical test of the algorithms ability to locate and identify contours.

Location of contours. The process of locating contours may best be shown by stepping through some examples. Figure 3-15 (A) shows the Henry Moore sculpture which sits in the M.I.T. Great Courtyard. Figure 3-15 (B) shows the zero-crossings of $\nabla^2 I$ calculated for this image, as per Marr and Hildreth 1980. These zero-crossings can be seen to localize many more "contours" than those we would like to consider. In figure 3-15 (C) shows those zero-crossings segments with a significant average slope across the zero-crossing. It can be seen that these contours correspond quite closely to people's intuitive definition of "contour", and that these contours are the one we need to find in order to be able to apply our shape-from-shading algorithm. Similarly, figure 3-15 (D) shows the Lisa picture, figure 3-15 (E) shows the zero-crossings of $\nabla^2 I$, and figure 3-15 (F) shows the zero-crossing segments with a significant average slope across the zero-crossing. Figure 3-15 (E) shows

many "weak" zero-crossings due to image noise. Again, the contours of figure 3-15 (F) seem to be the correct ones, as well as corresponding closely with peoples intuitive notion of an image contour.

The number of zero-crossing segments with a particular slope across them is a relatively smooth, monotonic function of the slope across the zero-crossing segment. Typically, the number of zero-crossing segments with a particular slope declines rapidly as slope increases up to a certain point above which the number of zero-crossing segments is relatively constant regardless of slope. The decline of the number of segments with increasing slope fits well with the hypothesis that these segments are due to noise, whereas a constant number of zero-crossing segments for any particular slope is to be expected for zero-crossings which correspond to imaged contours as the difference in light reflected from adjacent imaged surfaces has a uniform distribution. Thus the threshold for discarding weak zero-crossing segments is set at the slope above which the number of zero-crossing segments is constant.

Further examples of contour localization are shown in figure 3-16. Figures 3-16 (A) and (C) show LANDSAT[11] imagery of fields in North Dakota and Kansas. Figures 3-16 (B) and (D) show the contours localized this imagery (see Pentland 1980). In (B) and (D) the contours are overlayed as black lines on the image itself, in order to show the correspondence between the image and the localized contours. It is notoriously difficult to define reasonable contours in LANDSAT imagery, because the signal (inter-field differences) is the same size as the noise (the inter-pixel variation) (see Richardson and Pentland 1976, Pentland 1977, Kauth *et al* 1978); typically only people are able to discern the field structure. This extremely poor signal-to-noise ratio makes the accuracy of contour localization[12] in these examples all the more impressive.

Identification of smooth occluding contours. The theory chapter showed that smooth occluding contours could be differentiated from other contours by the surface orientation estimates on either side of the contour. It was proven that this differentiation cannot identify all smooth

[11]LANDSAT is NASA's earth resources survey satellite.

[12]Contour location was compared to ground-truth maps gathered as part of NASA's LACIE (Large Area Crop Inventory Experiment).

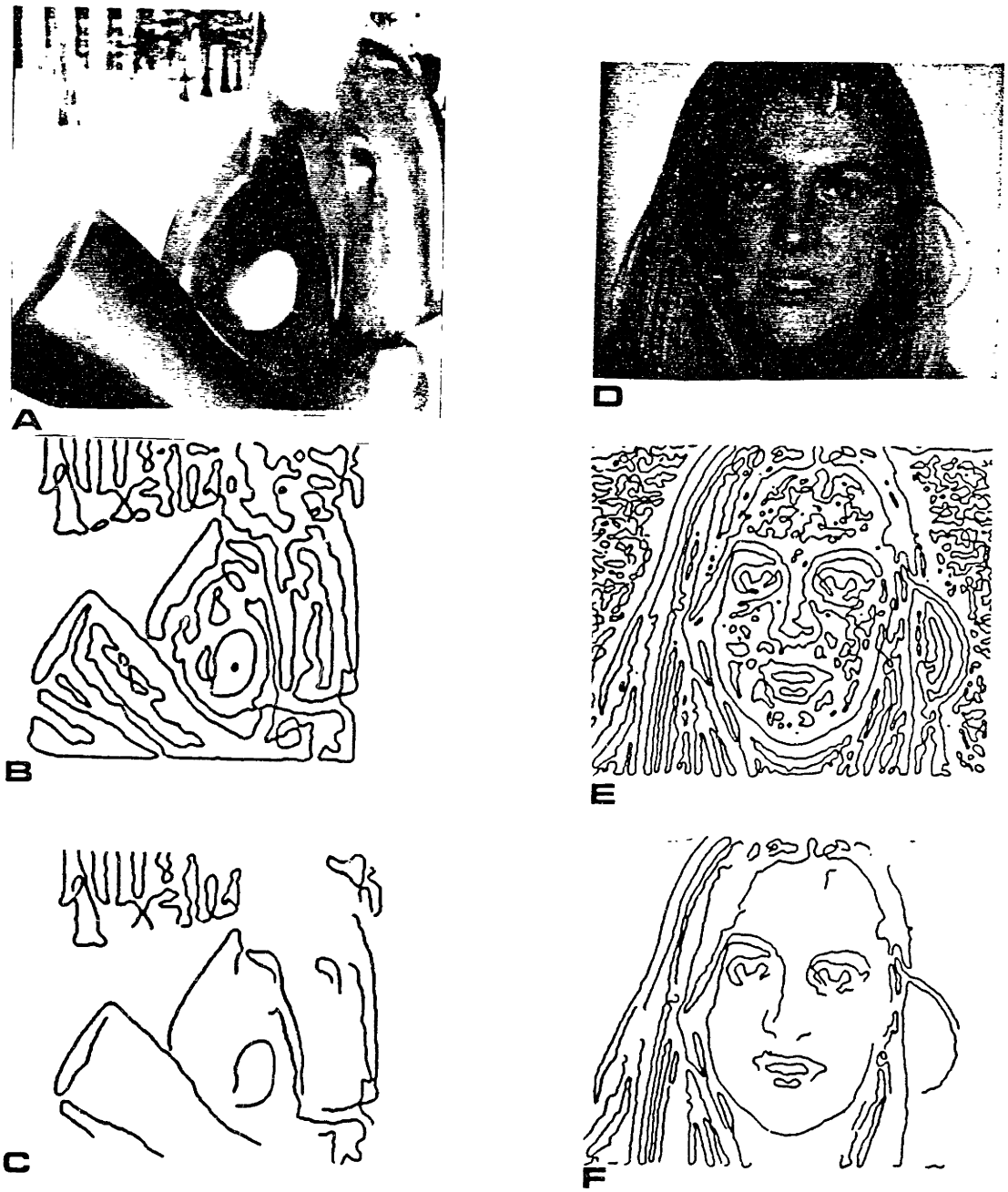


Figure 3-15. (A) Moore sculpture, (B) zero-crossings, (C) the localized contours. (D) The Lisa image, (E) zero-crossings, (F) the localized contours.

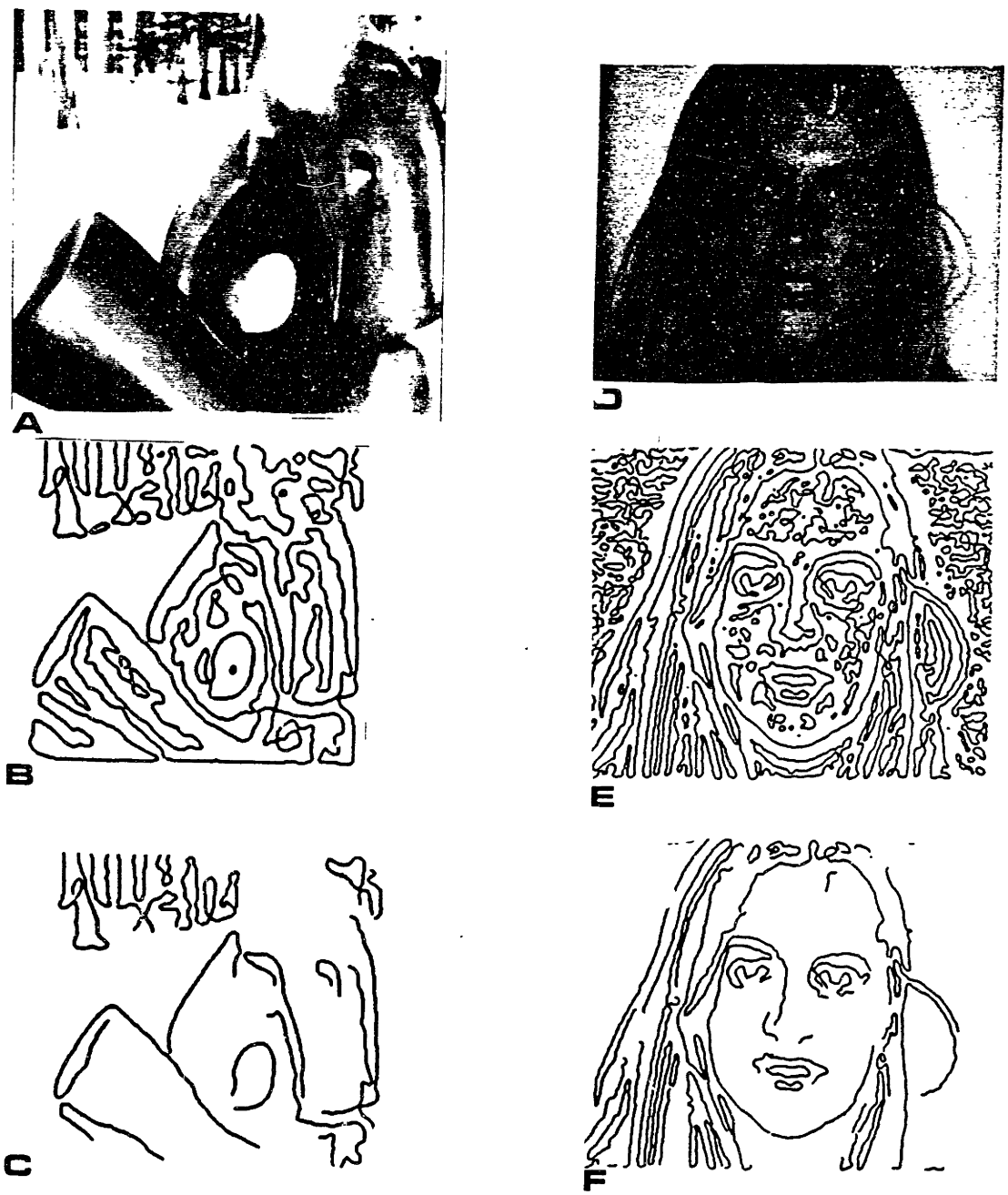


Figure 3-15. (A) Moore sculpture, (B) zero-crossings, (C) the localized contours. (D) The Lisa image, (E) zero-crossings, (F) the localized contours.

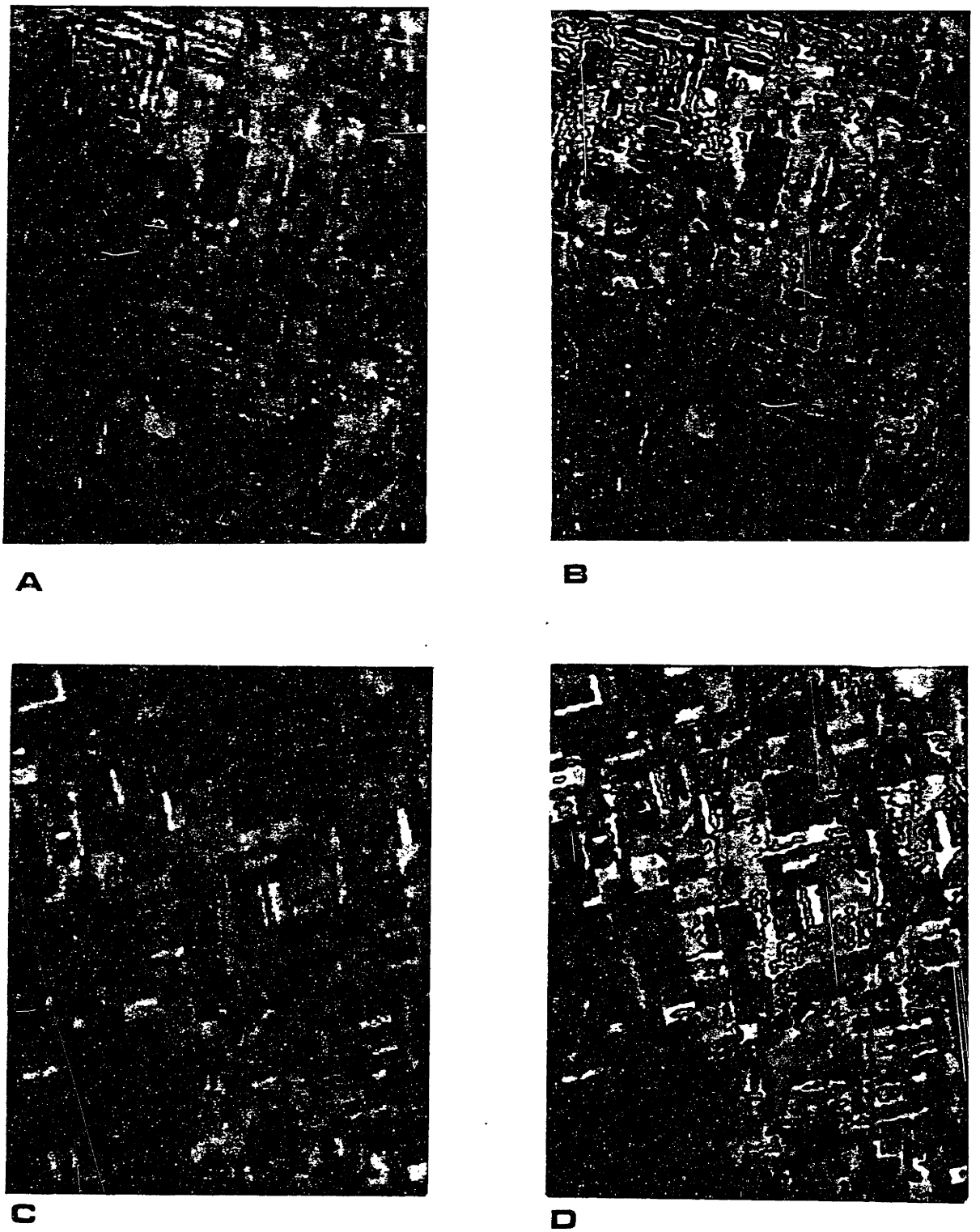


Figure 3-16. (A) and (C) show LANDSAT imagery of fields in North Dakota and Kansas. (B) and (D) show the contours localized in this imagery (see Pentland 1980).

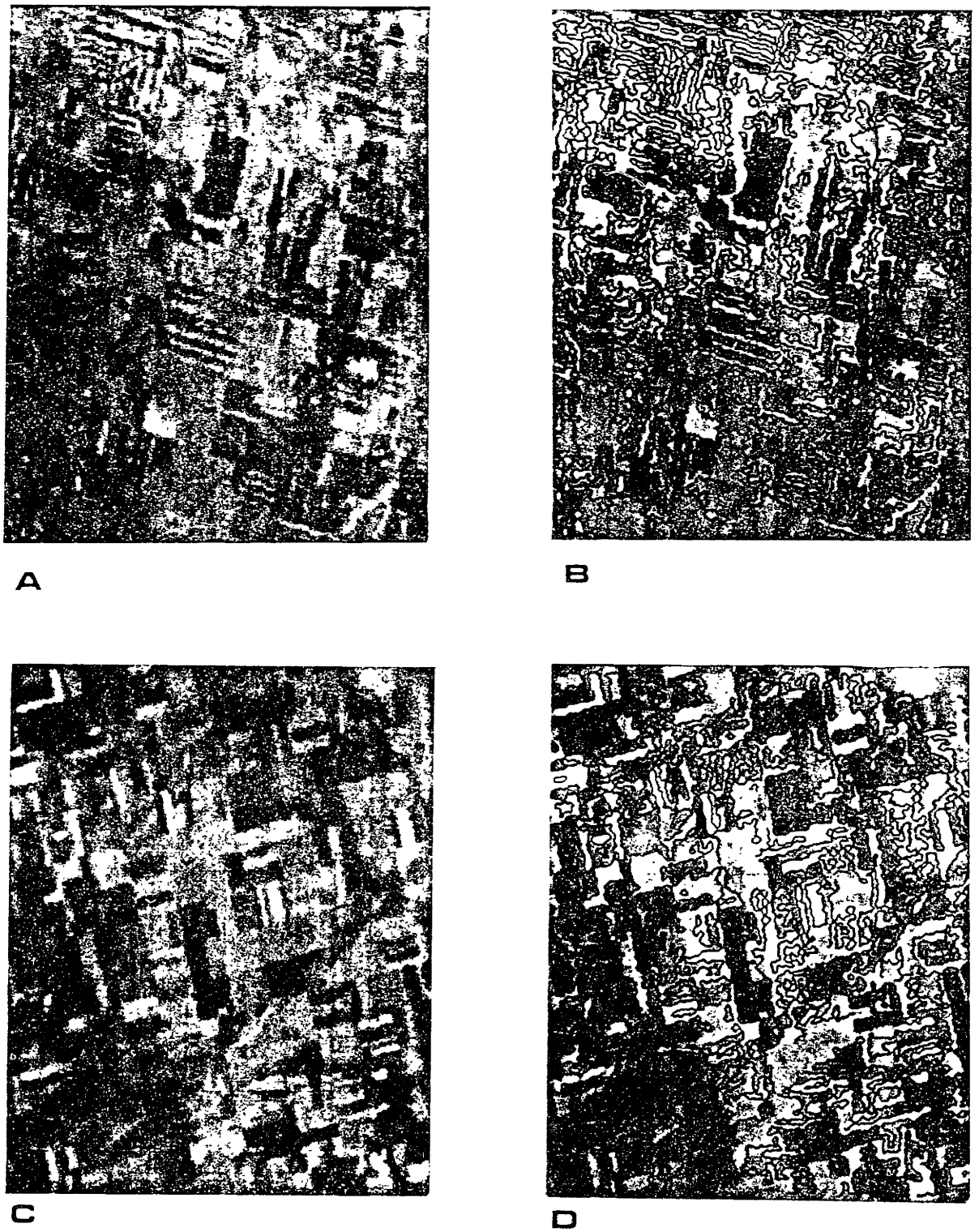


Figure 3-16. (A) and (C) show LANDSAT imagery of fields in North Dakota and Kansas. (B) and (D) show the contours localized in this imagery (see Pentland 1980).

occluding contours, because the slant may not be absolutely determined but only estimated. Thus this identification strategy will *mis*identify some contours with unusual surface shape configurations or unusual surface textures unless we are cautious, that is, identify only those contours which we are most certain of at the cost of missing the others. Because misidentification of sharp contours as smooth occluding contours can result in serious mistakes, it it important to only identify those contours bordered by the largest slant values and most appropriate tilt values as smooth occluding contours. Such an identification criterion can result in reliable identification of a significant portion of the smooth occluding contours in both synthetic and natural images, and thus provides significant information about shape.

Figure 3-17 presents two natural images for which the contour type has been estimated. Figure 3-17 (A) shows the Henry Moore sculpture, and figure 3-17 (B) shows the image of Tuckermans ravine. The contours localized in the Moore and Tuckerman images are shown in figures 3-17 (B), and 3-17 (E), respectively. The identified smooth occluding contours are shown in figures 3-17 (C) and 3-17 (F). The smooth occluding contours appear to be correctly identified, as can be seen from this figure[13], although in the more complicated Moore image many of the smooth occluding contours have been missed. The performance of this algorithm on these test images supports the claim that many of the smooth occluding contours in an image may be reliably identified in this manner.

[13]The correctness of the identification of the left-most portion of the Moore sculptures' bounding contour may be debated; however its surface orientation is certainly near $\pi/2$, so even if incorrect this error is not serious.

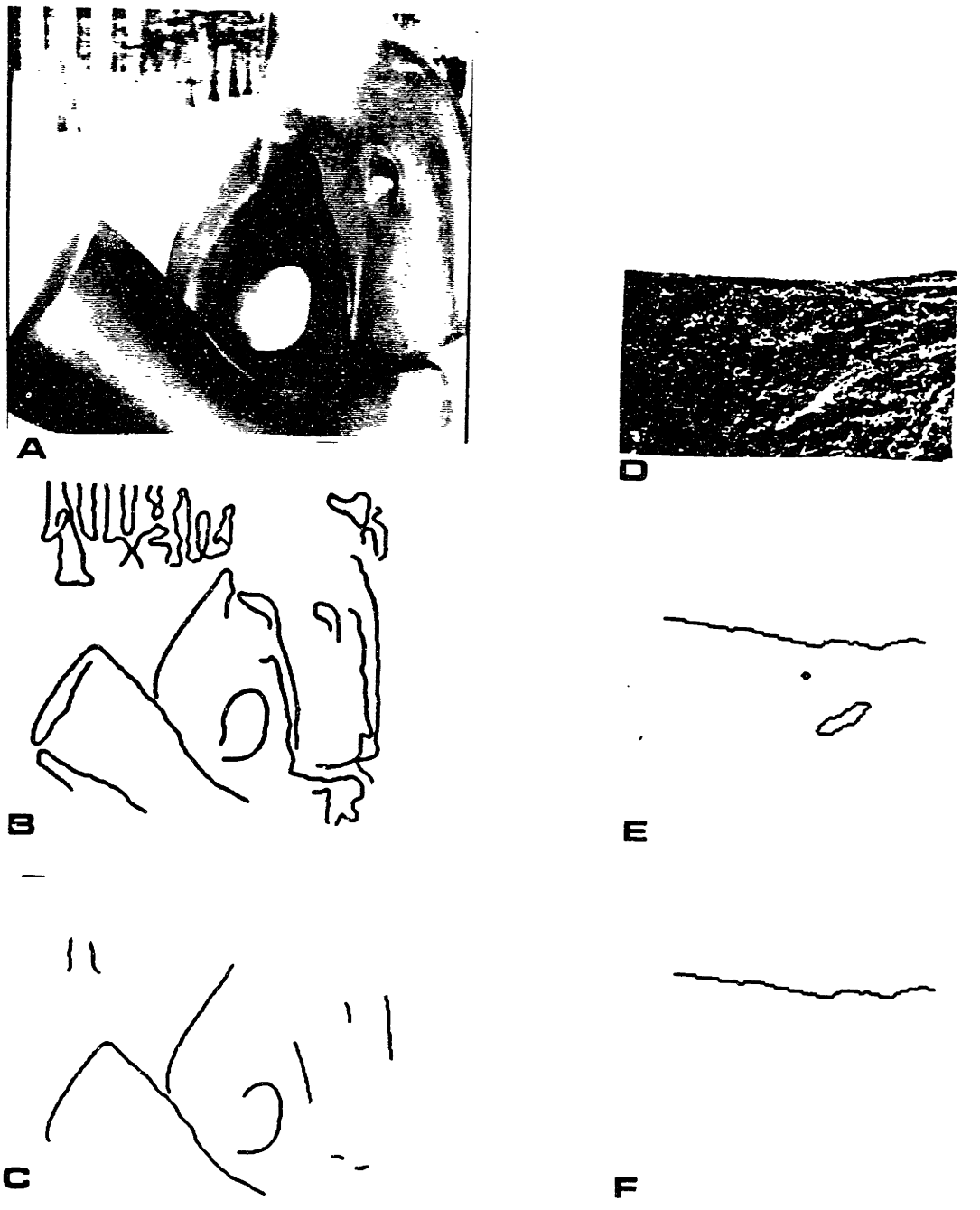


Figure 3-17. Determination of contour type. Figure 3-17 (A) shows the Henry Moore sculpture, figure 3-17 (D) shows the image of Tuckermans ravine. The contours which were identified for the Moore and Tuckerman images are shown in figure 3-17 (B), and figure 3-17 (E) respectively. The identified smooth occluding contours are shown in figures 3-17 (C) and 3-17 (F). The smooth occluding contours appear to be correctly identified, as can be seen from this figure.

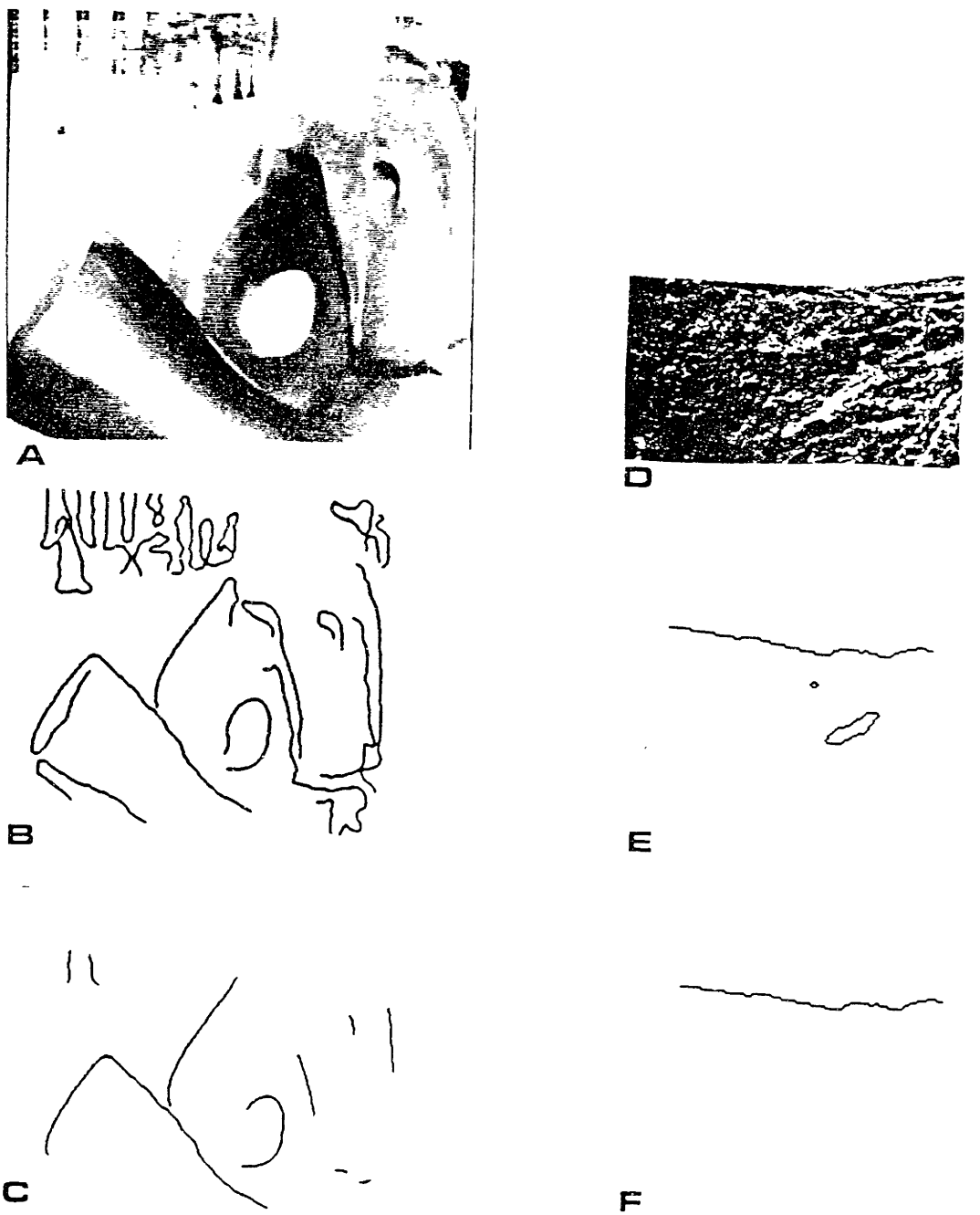


Figure 3-17. Determination of contour type. Figure 3-17 (A) shows the Henry Moore sculpture, figure 3-17 (D) shows the image of Tuckermans ravine. The contours which were identified for the Moore and Tuckerman images are shown in figure 3-17 (B), and figure 3-17 (E) respectively. The identified smooth occluding contours are shown in figures 3-17 (C) and 3-17 (F). The smooth occluding contours appear to be correctly identified, as can be seen from this figure.

3.4 Shape From Contour And Shading

We have developed a theory which shows how three-dimensionally interpreted contour may be used to constrain the interpretation of surface shape from shading, yielding a correct estimate of shape. Unfortunately, it is difficult to produce a simple diagram of how such constraint may be implemented, as was done with the other portions of the theory. In lieu of such a diagram, the algorithm can perhaps best be explained by demonstrating its action in several examples. Thus this section will first demonstrate the performance of the shape-from-contour-*and*-shading portion of the algorithm on several synthetic images, and at the same time explain the operation of the algorithm.

Figure 3-18 shows three examples of contour and shape together in images of the "Hindenburg" shape, shown without contour in part (A) of this figure. In (B) and (C) of this figure, the addition of contour to the shaded image gives a correct impression of the surface shape, however in (D) the addition of contour leaves the impression of shape unaffected[14]. These examples are especially interesting because they show that the human visual system uses some inter-constraint from contour and shading, and that this inter-constraint is sufficient to recover shape in some cases but not in others. Because these examples are so interesting, they were selected to demonstrate the functioning of the estimation algorithm.

We may use local shading information to determine the tilt of the surface and the convexity with great confidence, however there is a wide range of surface slants which are possible for any given pattern of local shading information; thus we may only estimate the slant of the surface. Because only an *estimate* is possible, there are a wide range of reasonable interpretations of the surface shape. We can illustrate the range of possible surface shape interpretations using the Hindenburg image of figure 3-18 (A) as an example. The extremes of interpretation for this image are illustrated by figure 3-19 (A) and 3-19 (B). Either of these shapes, and all the shapes intermediate between them, are possible interpretations of 3-18 (A).

[14]Note that in (C) the contours are simply iso-intensity lines from the shading, and thus *objectively* the shading plus contour image contains no more information about shape than the shading image alone; yet the shading alone does not give the correct impression of shape. Thus the interaction of contours with shading must be more than the simple addition of information about shape.

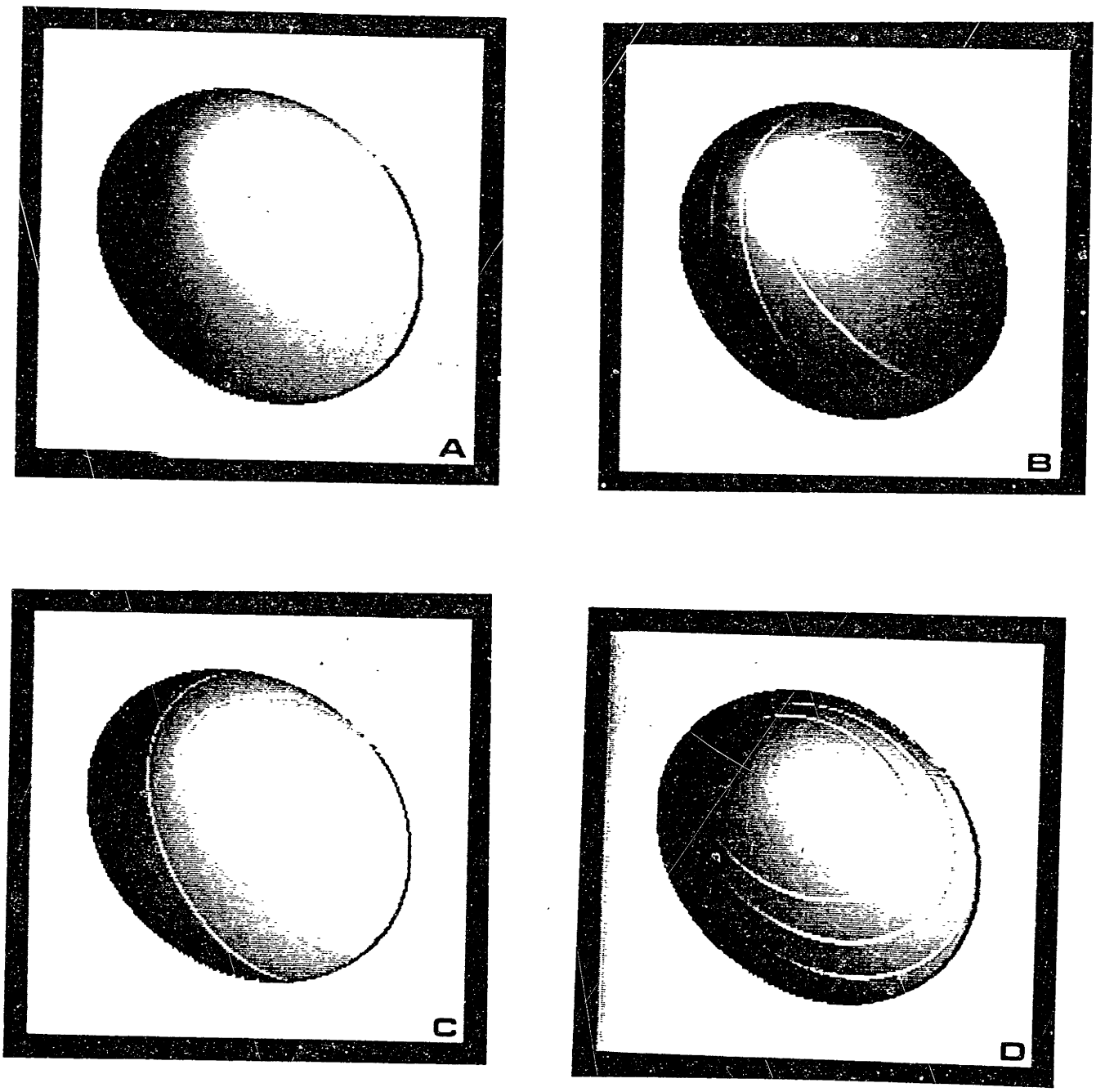


Figure 3-18. Part (A) of this figure, the Hindenburg with shading alone, gives a very flat impression of shape. Part (B) of this figure gives a much greater (and more correct) three-dimensional percept of the shape, as does (C). Part (C), however, contains no more information about shape than (A), as the only difference between (A) and (C) is that one of the iso-intensity contours in (A) have been made brighter in (C). (D) demonstrates that the presence of imaged contours does not always help the percept of shape. This image contains more contours than (C), and in fact the contours give additional information about shape in a way that the contour in (C) doesn't, and yet (D) gives only a very flat impression of shape.

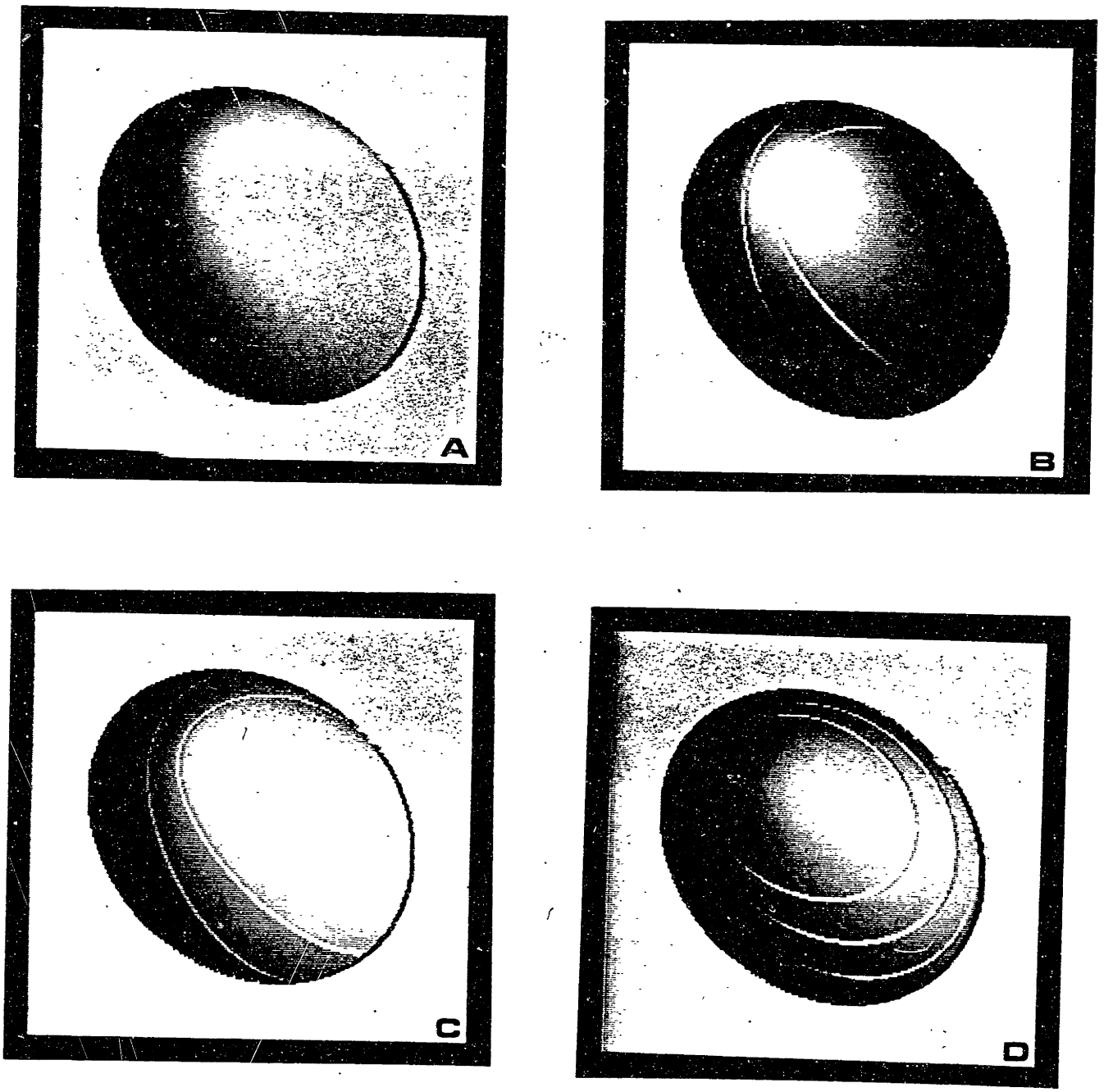


Figure 3-18. Part (A) of this figure, the Hindenburg with shading alone, gives a very flat impression of shape. Part (B) of this figure gives a much greater (and more correct) three-dimensional percept of the shape, as does (C). Part (C), however, contains no more information about shape than (A), as the only difference between (A) and (C) is that one of the iso-intensity contours in (A) have been made brighter in (C). (D) demonstrates that the presence of imaged contours does not always help the percept of shape. This image contains more contours than (C), and in fact the contours give additional information about shape in a way that the contour in (C) doesn't, and yet (D) gives only a very flat impression of shape.

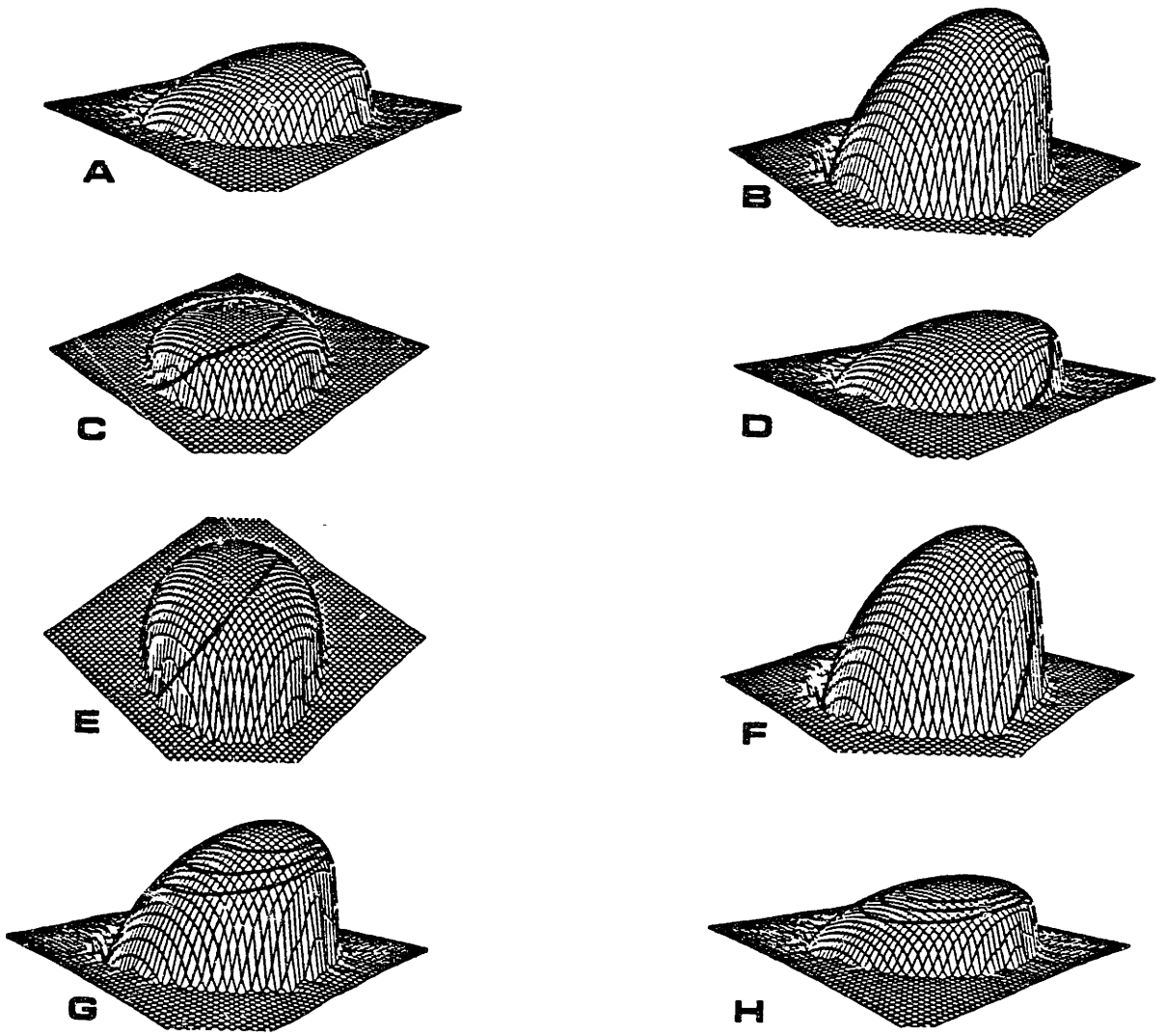


Figure 3-19. (A) shows a low-relief interpretation of surface shape which is locally consistent with the shading evidence. (B) shows a high-relief interpretation of surface shape which is also consistent with the shading. These two interpretations show extremes of the range of surface shapes which are locally consistent with the shading information. (C) and (D) show the imaged contours from figure 3-18 (B) and (C) projected onto the low-relief interpretation of surface shape provided by the shading. It can be seen that the requirement that these curves be planar is not consistent with this interpretation of surface shape, in fact this interpretation of surface shape requires a violation of general position. (E) and (F) show the imaged contours from figure 3-18 (B) and (C) projected onto the high-relief interpretation of surface shape provided by the shading. It can be seen that the requirement that these curves be planar is consistent with this interpretation of surface shape. (G) and (H) show the imaged contours from figure 3-21 (D) projected onto both the high- and low-relief interpretations of surface shape provided by the shading. It can be seen that the requirement that these curves be planar provides no constraint on the interpretation of surface shape in this image.

The three-dimensional interpretation of imaged contour developed in chapter 2 section 5, constrained the three-dimensional shape of the contour generator to be regionally planar. Let us see how such a planar curve fits with shape interpretations (A) and (B) of figure 3-19.

In figures 3-19 (C) and (D), we see the imaged contour of figures 3-18 (B) and (C) projected down onto the surface shape interpretation of 3-19 (A). As can be seen from this figure, if we accept this low-relief interpretation of surface shape then the curve generating the imaged contour cannot be planar; it must contain significant curvature oriented exactly away from the viewer, so that it cannot be seen in the image. Thus the low-relief interpretation of surface shape is inconsistent with the planar, maximum-likelihood interpretation of the contour generators three-dimensional shape. In fact, the low-relief interpretation requires a violation of general position in that significant features of the contour generator are hidden by being exactly lined up with the line of sight.

Figures 3-19 (E) and (F) show the imaged contour projected down onto the high-relief interpretation of surface shape, figure 3-19 (B). In this figure the three-dimensional shape of the contour generator is planar, as required by the maximum-likelihood interpretation of the imaged contour. Thus the high-relief interpretation of surface shape is consistent with the maximum-likelihood estimate of the contour generators three-dimensional shape, and requires no violation of general position. Therefore in this example the interaction of the planarity of the contour generator with the tilt and convexity determinations obtained from shading results in a unique, and correct, interpretation of surface shape in figures 3-18 (B) and (C). This is consistent with the human percept for these images.

How does this algorithm function in figure 3-18 (D), where the human visual system was unable to use the imaged contour to correctly estimate surface shape? Figure 3-19 (G) and (H) show the contours from figure 3-18 (D) projected onto both the high- and low-relief interpretion of surface shape. It can be seen that *both* interpretations of surface shape are consistent with the planar, maximum-likelihood estimate of the contour generator. Thus these imaged contours provide no constraint on the shading-obtained qualitative surface shape in this image. This failure is predicted in the theory; equation (20) which expresses the constraint between contour and shading information becomes

degenerate for contour generators which are parallel to the image plane; i.e., which run perpendicular to the tilt of the surface, as is the case with the contours in figure 3-18 (D). Thus the functioning of the algorithm in this example is also consistent with the human percept. The similarity of the algorithms performance to the human perception of shape is discussed in chapter 4 section 2.

Contour and shading in a natural image. The image of Tuckermans ravine was used in the portion of this chapter which evaluated the local shape-from-shading estimators. The result of estimating surface shape for this image is shown in figure 3-12 (B). Looking at the original image (figure 3-12 (A), or better, figure 3-8 (D) which shows a larger portion of the ravine), it seems that the lower right portion of the image has considerably more relief than is estimated. This underestimation was confirmed by comparing an aerial view of the estimated surface shape to a topographic map of the area, shown in figure 3-12 (C). The underestimation of the relief in this image when using only shading information allows a demonstration of the important role imaged contours play in determining shape.

In the lower right portion of the image (shown in figure 3-20 (A)) is a region of markedly different albedo, which generates a sharp contour as defined by this algorithm (the previous section of this chapter showed the location and identification of the contours in this image; the sharp contours found in this figure are shown as solid lines in figure 3-20 (B), the smooth occluding contour is shown as dotted line). We can use the contours which bound this region to obtain a more correct estimate of shape.

The sharp contours around the region of markedly differing albedo form an ellipse-like shape. The estimate of orientation for the plane in which this contour is estimated to lie is shown by figure 3-20 (C) (which may be thought of as a tin-can lid with a nail through it). Using equation (20), we then calculate that the region of the image for which this contour gives constraint must have the same orientation as the contour plane itself. We may therefore replace the shading-derived surface orientation estimates with these new estimates of surface orientation, taking care to join them smoothly with the shading-derived estimates of surface orientation, as no discontinuities in surface orientation were

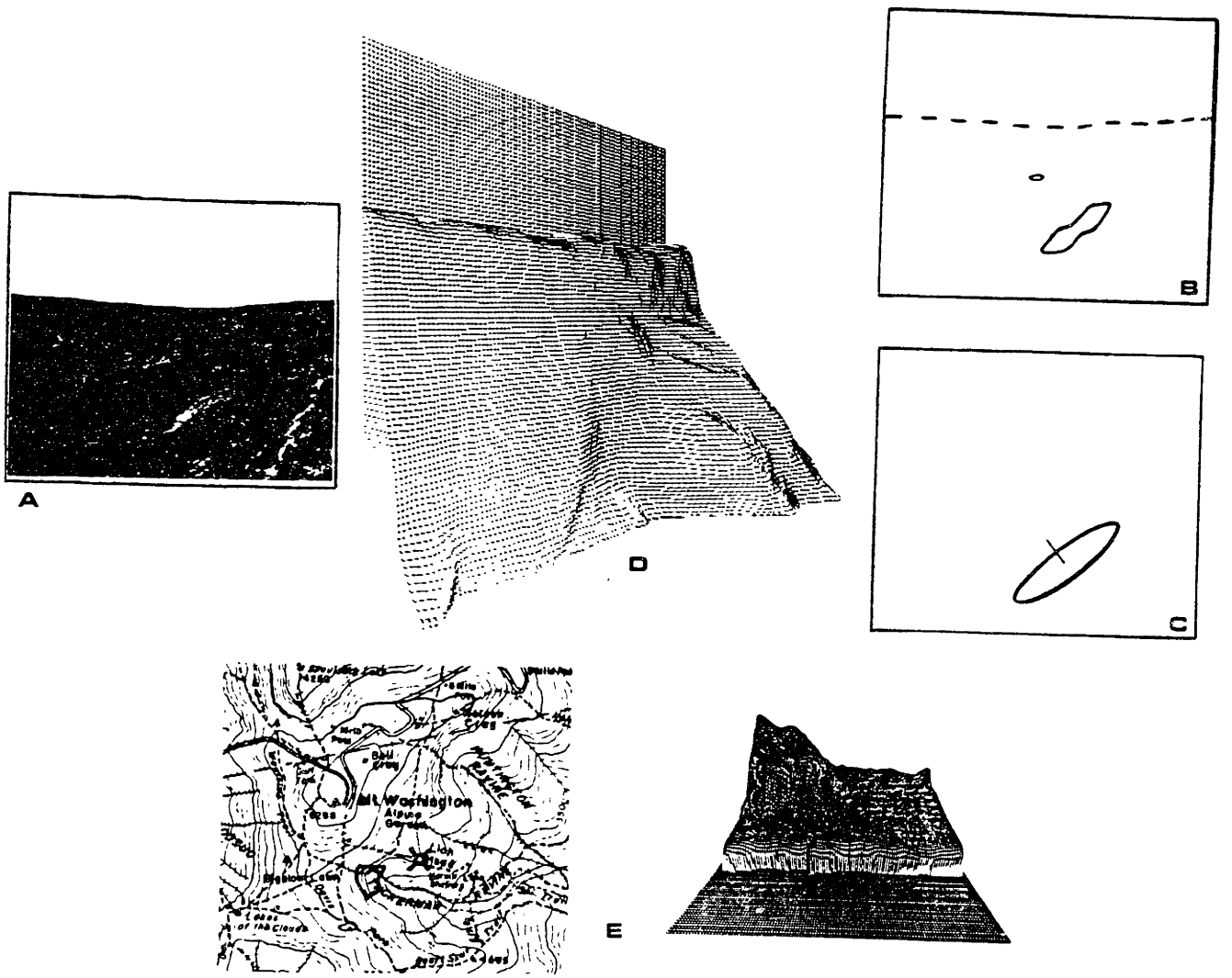


Figure 3-20. (A) The image of Tuckermans ravine (figure 3-8 shows the context of this image), (B) the contours located and typed in this image. Dotted contours have been identified as smooth occluding contours, solid contours are sharp occluding contours. (C) The estimated orientation of the osculating plane containing the sharp occluding contour, (D) a relief map showing a side view of the estimated surface shape. (E) a topographic map of the Tuckermans ravine area in which the imaged area is outlined and the camera position marked with a small dot, together with a relief map showing an aerial view of the estimated surface shape which may be directly compared to the outlined area of the topographic map. Comparing the topographic map with the estimated surface shape, we see that the estimated surface shape is correct. Thus the interaction of the shading and contour information in this image has allowed the shape of the imaged surface to be correctly recovered.

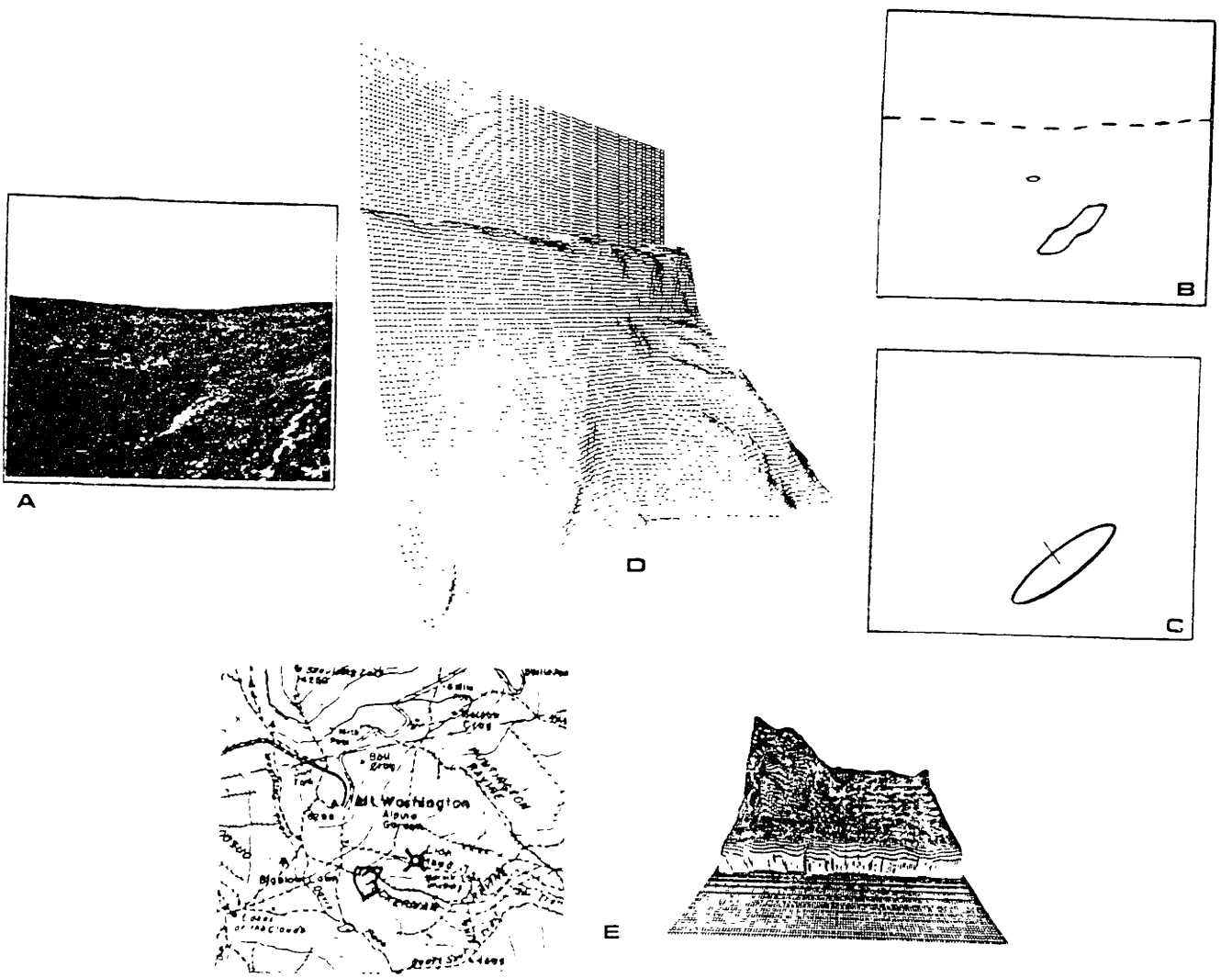


Figure 3-20. (A) The image of Tuckermans ravine (figure 3-8 shows the context of this image), (B) the contours located and typed in this image. Dotted contours have been identified as smooth occluding contours, solid contours are sharp occluding contours. (C) The estimated orientation of the osculating plane containing the sharp occluding contour, (D) a relief map showing a side view of the estimated surface shape. (E) a topographic map of the Tuckermans ravine area in which the imaged area is outlined and the camera position marked with a small dot, together with a relief map showing an aerial view of the estimated surface shape which may be directly compared to the outlined area of the topographic map. Comparing the topographic map with the estimated surface shape, we see that the estimated surface shape is correct. Thus the interaction of the shading and contour information in this image has allowed the shape of the imaged surface to be correctly recovered.

indicated by the shading. We may now integrate using this new estimate of slant and tilt to produce a new relief map showing a side view of the estimated surface shape, figure 3-20 (D).

We may compare this estimated shape with the actual shape of the imaged region. Figure 3-20 (E) shows the estimated shape as viewed from above, together with a topographic map of the area. The region corresponding to the image is outlined, and the camera position is marked. Comparing the estimated shape with the topographic map, we see that there is an impressive agreement between the two. Thus the addition of contour information has greatly improved the accuracy of our algorithm. It is clear that the addition of contour information has also greatly improved the agreement between the estimated surface shape and the impression of surface shape we obtain when viewing the image.

3.5 Robustness And Error Analysis

Scale and the computational theory: Contours. The selection of a scale, that is, the specification of the size (variance) of $G(x, y)$, has an influence on the validity of some of the assumptions of the theory, and on the algorithms sensitivity to image noise. For instance, Marr and Hildreth 1980 have analyzed how zero-crossings of $\nabla^2 G(x, y) \otimes I^*(x, y) = \nabla^2 I(x, y)$ are influenced by image noise as a function of scale. They found that the larger the size of $G(x, y)$, the less the effect of a specific amount of image noise.

A similar analysis was carried out for the contour localization algorithm developed here. The results of this analysis are illustrated in figure 3-21. Part (A) of this figure again shows the Moore sculpture image, and part (B) shows the same image after the addition of a large amount of uniformily distributed noise. The signal-to-noise ratio in (B) is one, so that performance on this image is a severe test of the stability of the contour localization algorithm. Part (C) shows the zero-crossings for the original image, (A), using a filter with a support of 121 pixels and a center region four pixels across. Shown below the zero-crossings is the distribution of zero-crossing slope, with number of zero-crossing segments along the vertical axis, and slope (equivalent to d^2I across the contour) along the horizontal axis. Part (D) of this figure shows the zero-crossings of the noisy image, (B), again using a filter with a support of 121 pixels. Below this is shown the distribution of zero-crossing slope for the noisy image.

Comparing the distribution of zero-crossing slopes in (C) with that in (D) we see that as noise is added to the image the proportion of zero-crossings classified as noise by the threshold rule also grows. This fact adds creedence to the claim that the zero-crossings discarded by the threshold rule are primarily due to noise.

Part (E) of figure 3-21 shows the contours localized in the original Moore image, i.e., those zero-crossings segments with slope greater than the threshold, and part (F) of figure 3-21 shows the contours localized in the noisy Moore image. This example shows that contours localized in this manner

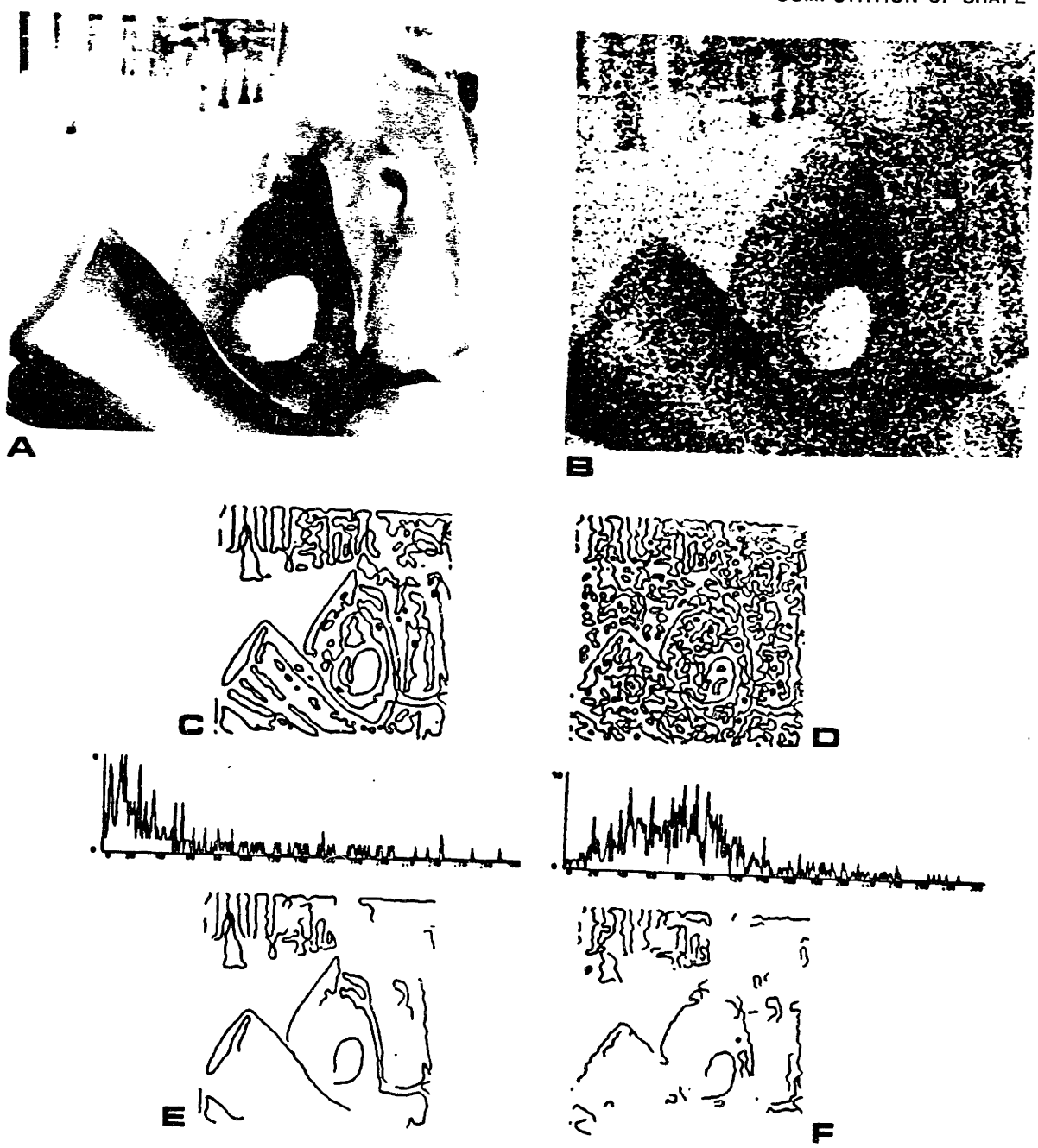


Figure 3-21. (A) The Moore sculpture image, (B) The same image after addition of uniformily distributed noise, so that the signal-to-noise ratio is one. (C) The zero-crossings for the original image, (A), and the distribution of zero-crossing slope, with number of zero-crossing segments along the vertical axis, and slope (equivalent to d^2I across the contour) along the horizontal axis. (D) The zero-crossings of the noisy image, (B), and the distribution of zero-crossing slope for the noisy image. Comparing the distribution of zero-crossing slopes in (C) with that in (D) we see that as noise is added to the image the proportion of zero-crossings classified as noise by the threshold rule also grows. This fact adds creedence to the claim that the zero-crossings discarded by the threshold rule are primarily due to noise.

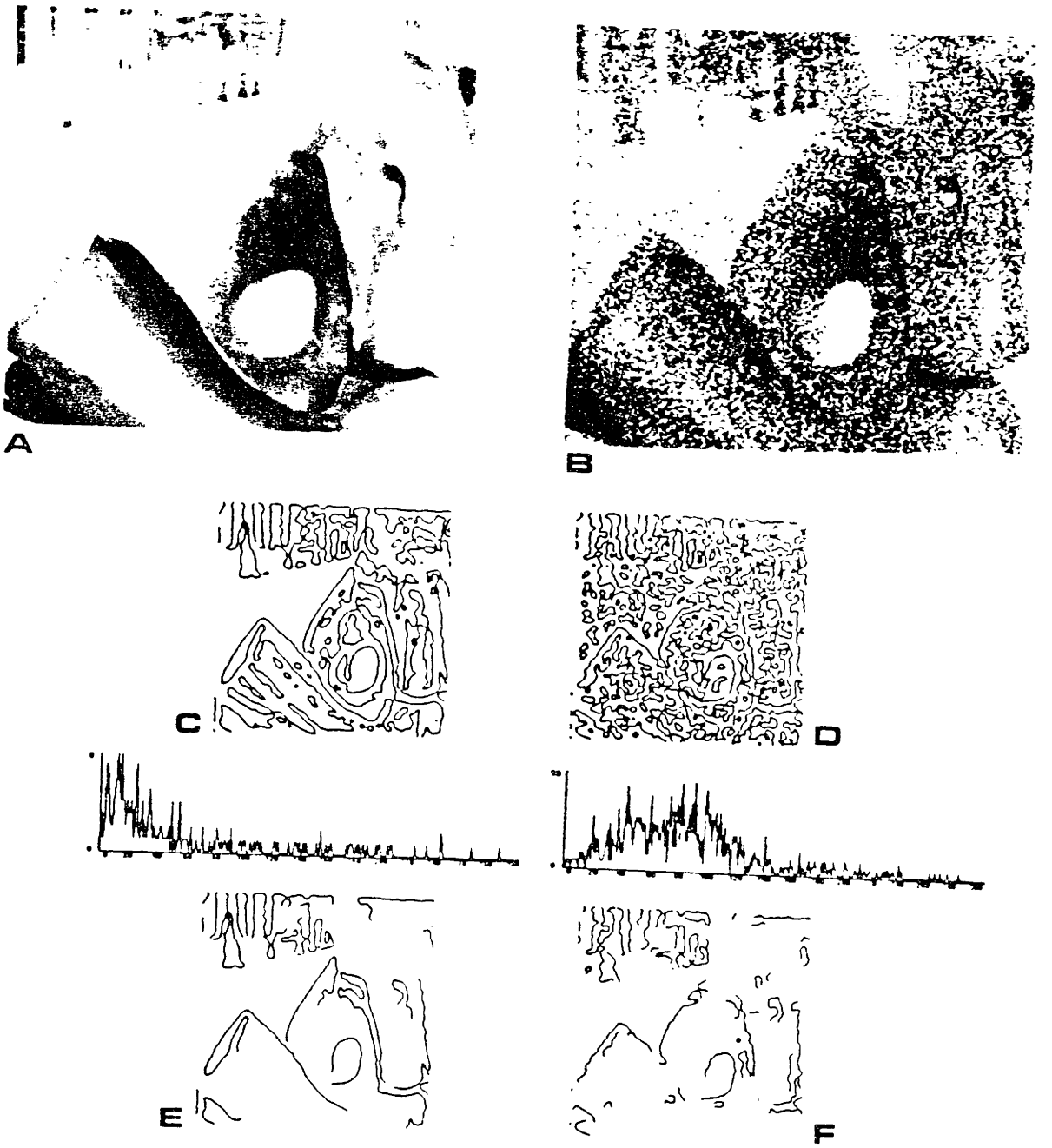


Figure 3-21. (A) The Moore sculpture image, (B) The same image after addition of uniformly distributed noise, so that the signal-to-noise ratio is one. (C) The zero-crossings for the original image, (A), and the distribution of zero-crossing slope, with number of zero-crossing segments along the vertical axis, and slope (equivalent to d^2I across the contour) along the horizontal axis. (D) The zero-crossings of the noisy image, (B), and the distribution of zero-crossing slope for the noisy image. Comparing the distribution of zero-crossing slopes in (C) with that in (D) we see that as noise is added to the image the proportion of zero-crossings classified as noise by the threshold rule also grows. This fact adds credence to the claim that the zero-crossings discarded by the threshold rule are primarily due to noise.

are stable even for the extreme signal-to-noise ratio of one. Figure 3-16 showed the performance of this algorithm in two naturally occuring cases where the signal-to-noise ratio is one. If we compare the unthresholded zero-crossings in (C) and (D), we see that there is a substantial, even qualitative difference caused by the addition of this much noise. Thus if we are to obtain a consistent description of the contours in a viewed scene despite varying noise conditions we cannot use unthresholded zero-crossings; it seems that the thresholding operation must be applied in order to obtain a stable description of the scene.

Another important aspect of robustness is *scale invariance*; we must obtain similar descriptions of the contours in a viewed scene as we move closer or farther from the objects in the scene. Without such invariance, we would require a separate description of each object at every possible viewing distance, and could not perform a recognition task without first knowing the distance to the viewed object. Figure 3-22 shows how the localized contours vary with scale. Parts (A), (B) and (C) of figure 3-22 show the zero-crossings of the Moore sculpture image at three different scales, each one octave apart. The filter used to generate (A) had a center region four pixels across, for (B) the filter had a center eight pixels across, and for (C) the filter had a center sixteen pixels across. The zero-crossings produced by these three filters is identical to the zero-crossings which would be produced by the smallest filter convolved with an image taken at one, two and four times the actual distance to the Moore sculpture. Parts (D), (E) and (F) show how the localized contours (thresholded zero-crossings) vary for these three filters or distances. It can be seen from this figure that although the unthresholded zero-crossings vary considerably with distance, the localized contours (thresholded zero-crossings) remain quite stable.

Scale and the computational theory: Shading. Marr and Hildreth 1980 have shown how the relatively large support of the $\nabla^2 G$ filters such as used here act to reduce the effects of noise. Because the shape-from-shading estimators developed here use the same filter, one would expect that they would also benifit from the noise-reducing averaging effects of filters with such large supports. In addition, however, the size of filter selected affects the truth of the assumptions that the cross-term

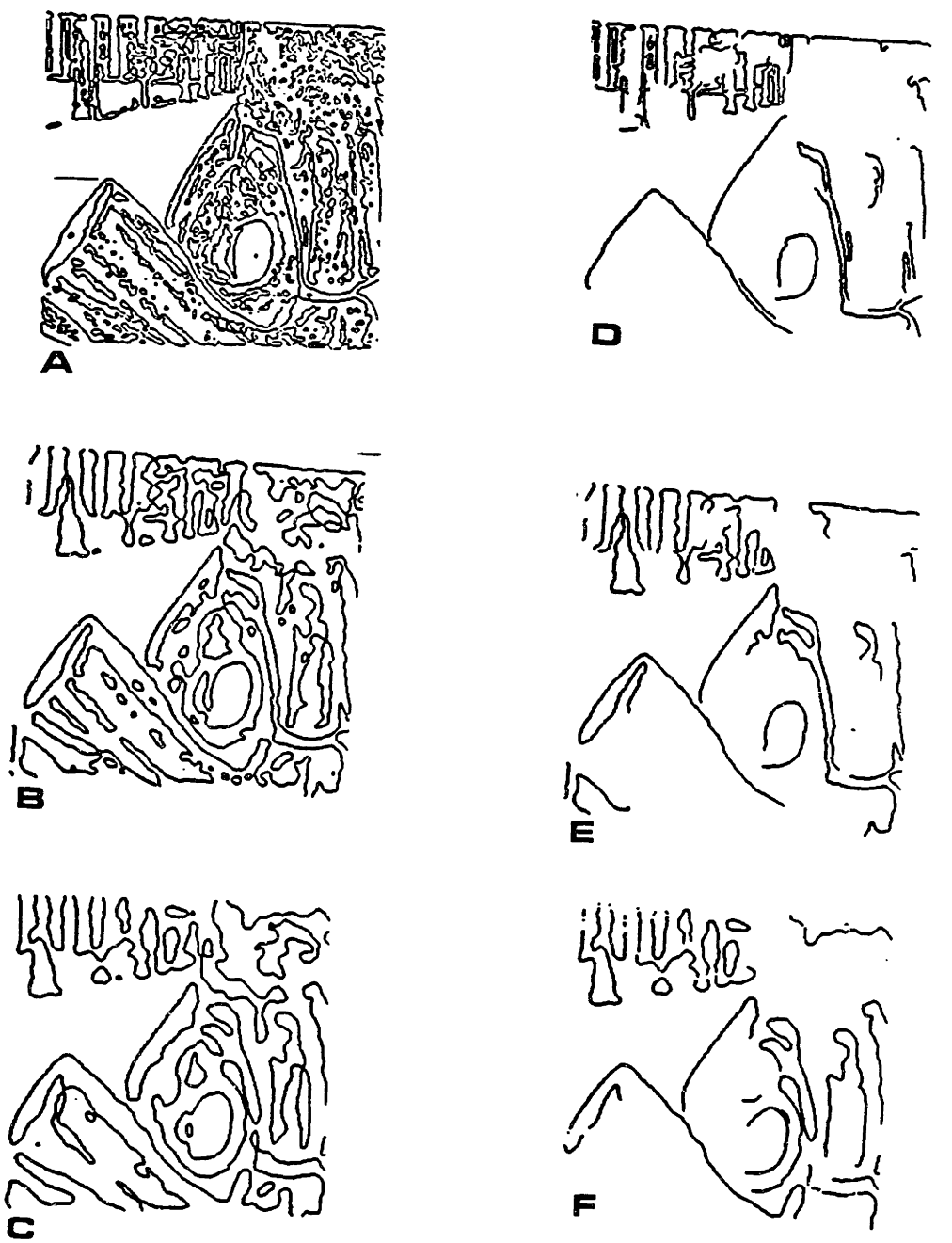


Figure 3-22. Invariance of contours with scale or viewing distance. (A), (B) and (C) show the zero-crossings of the Moore sculpture image at three different scales. The filter used to generate (A) had a center region four pixels across, for (B) the filter had a center eight pixels across, and for (C) the filter had a center sixteen pixels across. The zero-crossings produced by these three filters is identical to the zero-crossings which would be produced by the smallest filter convolved with an image taken at one, two and four times the actual distance to the Moore sculpture. Parts (D), (E) and (F) show how the localized contours (thresholded zero-crossings) vary for these three filters or distances. It can be seen from this figure that although the unthresholded zero-crossings vary considerably with distance, the localized contours (thresholded zero-crossings) remain quite stable.

$\mathbf{p}_{uv}$ and that the third derivatives of $\mathbf{p}(u, v)$ are zero.

In developing an expression for the second derivative of image intensity, it was assumed that the third derivatives of the surface $\mathbf{p}(u, v)$ with respect to the co-ordinate system (u, v) were zero, and that $\|\mathbf{p}_{uv}\| = 0$. This assumption is, of course, not always true of real surfaces. When we pick a scale of examination (i.e., a particular $G(x, y)$), however, we also limit all of the derivatives of image intensity. This can be shown by convolving an image consisting of a single delta function of unit amplitude with derivatives of $G(x, y)$:

$$\begin{aligned}\delta(x, y) \otimes G(x, y) &= G(x, y) \\ \delta(x, y) \otimes dG(x, y) &= dG(x, y) \\ \delta(x, y) \otimes d^2G(x, y) &= d^2G(x, y)\end{aligned}$$

The convolution of the delta function with the various filters G, dG, or d^2G gives the maximum values of the 0^{th}, 1^{st}, 2^d derivatives for a fixed amplitude of δ. Thus if we pick a filter with small third derivatives, we will be observing a surface $\mathbf{N}(x, y)$ with at most those third derivatives, as

$$\begin{aligned}I(x, y) &= G(x, y) \otimes I^*(x, y) \\ &= (G(x, y) \otimes \mathbf{N}^*(x, y)) \cdot \mathbf{L} \\ &= \mathbf{N}(x, y) \cdot \mathbf{L}\end{aligned}$$

Thus the third derivatives of $\mathbf{p}(u, v)$ can be made arbitrarily small by choosing a large enough G. The same arguments apply to minimizing $\|\mathbf{p}_{uv}\|$.

We can carry out an investigation similar to the one shown in figures 3-21 and 3-22 using the shape-from-shading algorithm developed here. The results of such a study are shown in figure 3-23. Two different levels of uniformly distributed image noise were added to an image of a sphere. The size of these images was 200×200 pixels, and the the ∇^2G filter had a center 21 pixels across. The signal-to-noise ratios used were infinite (no noise), ten-to-one, and one-to-one. The shape was then estimated for each of these images, and the resulting estimates integrated to produce an estimated surface shape. The estimated surfaces are shown in the relief maps of figure 3-23 (A), (B), and (D).

The no-noise case produced a shape estimate whose relief map is shown in figure 3-23 (A), The estimated shape is perfect to within the numerical error expected for the type of integration program

used. The deviation from spherical averaged 0.0094%within the central 150° of the surface, increasing somewhat near the edges.

The ten-to-one signal-to-noise ratio case produced a shape estimate whose relief map is shown in figure 3-23 (B). As can be seen, "gullies" have appeared in the sphere, although the overall shape is still fairly spherical. The deviation from spherical is averages 10.15%over the central 150° degrees. The size of this error, however, appears to be largely a function of the integration procedure, and not of the estimator itself. Figure 3-23 (C) shows the difference between a perfect sphere and the shape of figure 3-23 (B). It can be seen that in the central region the error is very small; it is only as one gets further from the center that the errors build up, causing the gullies. Because the integration program starts in the center and "builds" the surface radially by adding successive patches of surface, it appears that what is happening is that once a small error is made, the integration program carries it along and adds to it, thus causing the gullies and an overall trend toward higher error as one gets further away from the center.

The one-to-one signal-to-noise ratio case produced a shape whose relief map is shown in figure 3-23 (D). As can be seen, the gullies have deepened, and the general spherical shape further degraded. In this case the deviation from spherical averages 26.37%over the central 150° degrees. While this level of performance is poor, it should be remembered that both this case and the ten-to-one case are very difficult tests of the algorithm. Additionally, use of a different integration program seems likely to significantly improve these results.

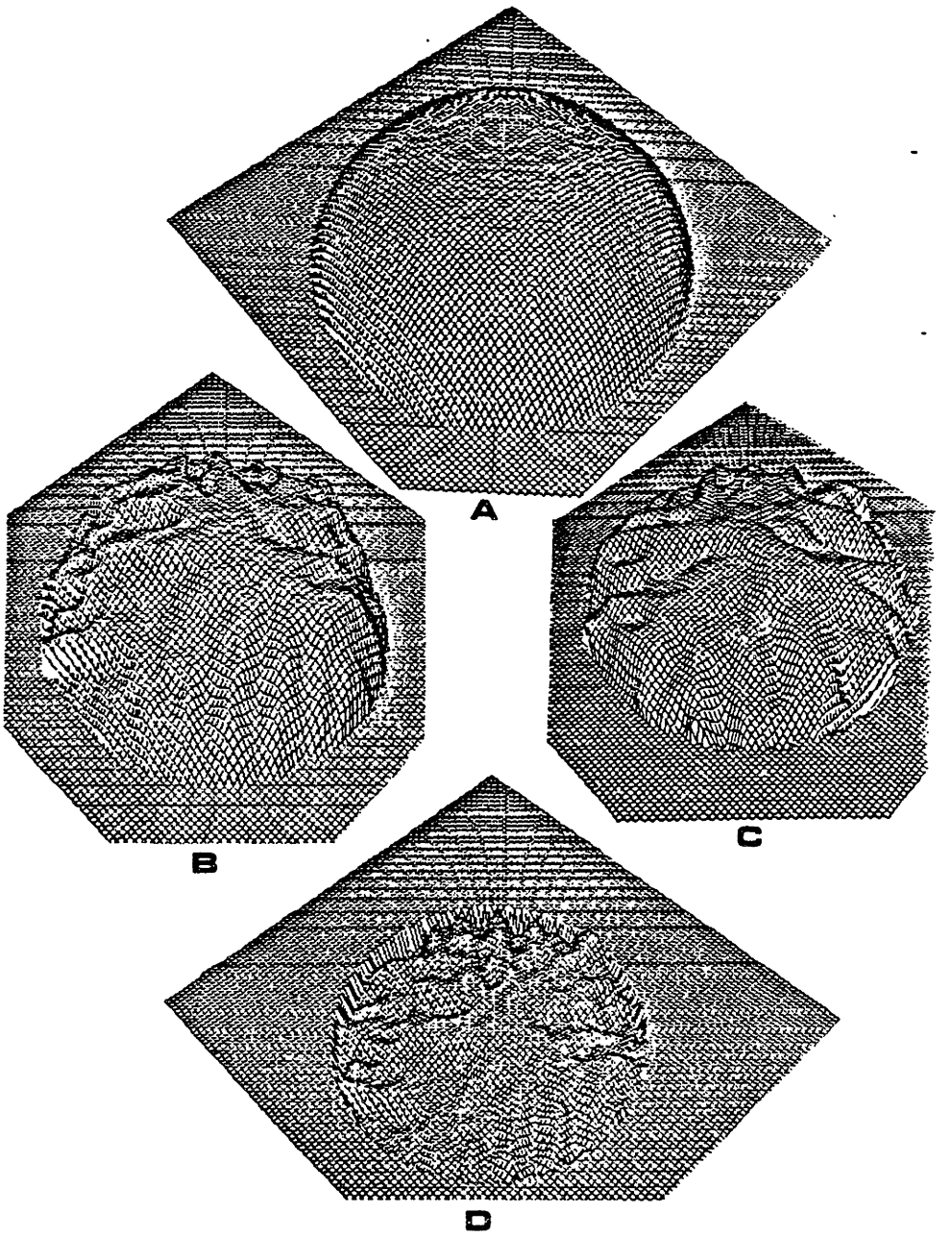


Figure 3-23. Two different levels of uniformily distributed image noise were added to an image of a sphere. (A) shows the no-noise case; the estimated surface shape is nearly perfect. (B) shows the surface shape estimated in a ten-to-one signal-to-noise ratio image. The deviation from sherical is 10.15%(C) shows the difference between a perfect sphere and the shape shown in (B). As can be seen, the errors are very small in the center and increase as one moves toward the edge. This pattern of error seems to implicate the integration procedure, rather than the shape estimator, as the major source of error. (D) shows the surface shape estimated in a one-to-one signal-to-noise ratio image. The deviation from spherical averages 26.37%. While this level of performance is poor, it should be remembered that both this case and the ten-to-one case are very difficult tests of the algorithm. Additionally, use of a different integration program seems likely to significantly improve these results.

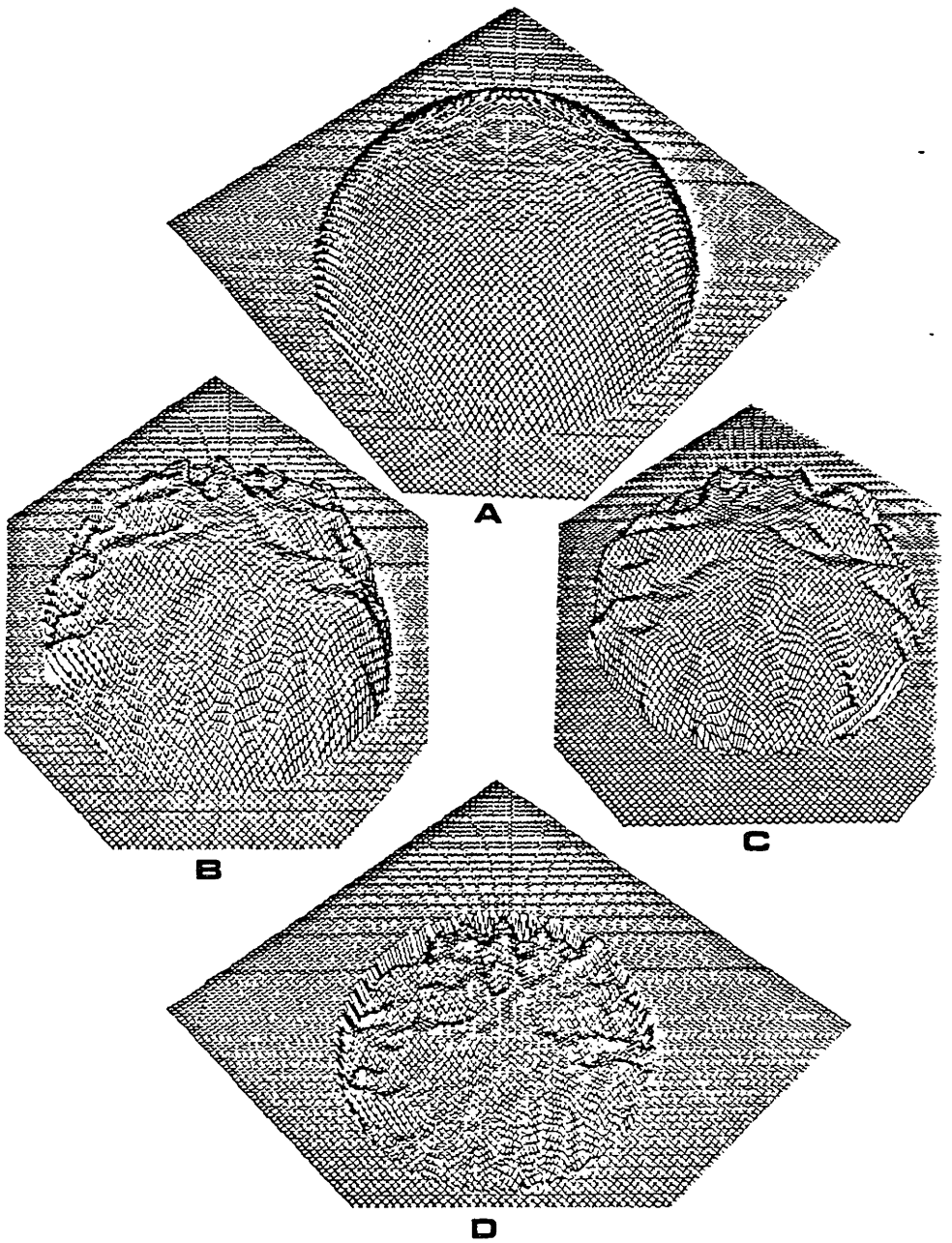


Figure 3-23. Two different levels of uniformily distributed image noise were added to an image of a sphere. (A) shows the no-noise case; the estimated surface shape is nearly perfect. (B) shows the surface shape estimated in a ten-to-one signal-to-noise ratio image. The deviation from sherical is 10.15%(C) shows the difference between a perfect sphere and the shape shown in (B). As can be seen, the errors are very small in the center and increase as one moves toward the edge. This pattern of error seems to implicate the integration procedure, rather than the shape estimator, as the major source of error. (D) shows the surface shape estimated in a one-to-one signal-to-noise ratio image. The deviation from spherical averages 26.37%. While this level of performance is poor, it should be remembered that both this case and the ten-to-one case are very difficult tests of the algorithm. Additionally, use of a different integration program seems likely to significantly improve these results.

Chapter 4

Human Perception

1 Introduction

The task of this thesis has been investigate how it is that the human visual system is able to estimate surface orientation from a single, unfamiliar image, seemingly in a local, bottom-up manner. In order to investigate this question, we first addressed the *theoretical* problem. Chapter 1 examined the physics and geometry of the process of image generation to determine how image features are affected by surface orientation and curvature. Chapter 2 then determined how these features could be used to estimate the surface shape.

After the theoretical problem was dealt with, an algorithm was designed to implement the computations required by the theory. This algorithm was described and evaluated in chapter 3. On the basis of the examples presented in chapter 3, it appears that useful estimates of surface shape may be generally obtainable from a local analysis of the image.

We now face the question of how this theory and algorithm relate to the human perception of shape. In order to answer this question we must first be clear about what claims are reasonable to make, about what evidence might bear upon these claims, and about the extent to which we can potentially prove or disprove such claims. Therefore we must first investigate the logical relations between the theory, the algorithm, the human mechanism and experimental evidence (see Marr and Poggio 1977).

1.1 Theory, Algorithm, Mechanism And Evidence

There are three levels at which we can make the claim that our theory corresponds to the human

visual system:

The Computational Theory. We can claim that both our theory and the function of the human system are predicated upon the same laws of physics and statistical regularities of the world in order to deduce, from the same image information, similar physical interpretations of each image. This is to claim that the *computational theories* are the same.

We could test this claim by determining whether the two estimation "machines" (the computer implementation of our algorithm and the human visual system) obtain the same estimates given the same images under "normal" operating conditions. Both machines will normally produce the same estimates if and only if they are both predicated upon the same assertions about the world and operate on the same body of image information. Further, we might expect that the variance of the estimates would be correlated — if they both use the same image information then when one has enough information to make a confident estimate, so should the other.

We would expect that both estimation machines would make major errors on the same images — those images which violate the assumptions on which the theory is predicated. The exact nature of the major mistakes might not be the same, however, because differences in algorithm might cause the two estimates to break down in different ways. Similarly, we would *not* necessarily expect the two estimation machines to make absolutely identical estimates; small differences may result from limitations inherent in the particular algorithm or mechanism chosen to carry out the computations of the theory. An example of this is two computers carrying out the same computation with different precisions (number of bits per number). The two computers would give typically give slightly different answers because of round-off errors; and they might give quite different answers for computations involving very large or very small numbers, yet they both are still performing the same computation.

The Algorithm. We could expand our theoretical claims beyond those of the computational theory by claiming that our algorithm for carrying out the computational theory was identical to the algorithm used by the human brain. We would then be claiming that (1) both our theory and the

human system are predicated upon the same information about the world and use the same image information, and (2) both our algorithm and the human system extract the same output information through isomorphic transformations of the same image features.

Note that in order to make the notions of "computational theory" and "algorithm" coherent it is necessary that the term "computational theory" refer *only* to the generic solution of scene structure from image data. If we were to interpret the computational theory as being the complete inference scheme, from image data to interpretation, then the notion of "theory" swallows up the notion of "algorithm".

If we expand our claims to include a proposal about the algorithm, then we would reasonably expect that both our theory and the human visual system would make the *same* mistakes under the same conditions, at least under conditions where the hardware or mechanism is in its normal operating range and thus presumably not contributing significantly to the errors. Further, we would expect that our algorithm could account for (or at least be consistent with) many of the secondary characteristics of the human visual system. For example, our proposed algorithm and the human visual system should have similar sensitivities to changes in image intensity.

The Mechanism. Finally, we might wish to expand our claims still further to include a description of the mechanism which implements the algorithm. We could then propose that this mechanism is that used by the human visual system. We would then expect to account for anatomical data, and data from neurophysiology about neural characteristics.

These then are the three levels at which we can entertain a claim that we have an explanation for some portion of the human visual system. The first two levels, theory and algorithm, are classically addressed through psychophysical experiments. The level of mechanism may be addressed through the disciplines of anatomy and neurophysiology. Therefore the next section will address itself to the psychophysical evidence concerning similarity between the human visual system and the theory developed in this thesis, and between the human visual system and the algorithm implementing this

theory. The last section will deal with similarities between this algorithm and the neuronal organization observed in biological visual systems.

2 Psychophysics

There are three major portions of the theory which we will investigate as possible theories of human perception. These three are

(1) Estimation of surface orientation.

(2) The constraint of contour on shading.

(3) Estimation of illuminant direction.

For each of these three parts of the theory we may make a claim that the theory on which the human visual system is predicated similar to the theory developed here; and we may also make the separate (but dependent) claim that the algorithm employed by the human visual system is similar to the algorithm described here.

Similarity of theory. The preceding discussion showed that the only methods by which we may comment on the similarity between the theory of this thesis and the theory on which human vision is based is to demonstrate (1) that both yield similar estimates given the same image information, and (2) that both make mistakes on the same examples (although the mistakes don't necessarily have to be the same). Such a demonstration does not *prove* that the theories are identical; such definite proof appears to be impossible to obtain without exhaustive evaluation of all cases. Even if such exhaustive comparison were possible, the complexity of the human visual system would insure that our results would still be open to doubt. Nonetheless, we may still seek strong correlational evidence of similarity.

Similarity of algorithm. The preceding discussion also showed that the methods available to show similarity between the algorithm developed in this thesis and the algorithm used by the human visual system are (1) demonstrate that they both make the same mistakes, and (2) demonstrate they both have the same "secondary characteristics", e.g., similar contrast sensitivity. Again, we cannot conclusively demonstrate equivalence of algorithm, but again we can hope to obtain strong correlational evidence of similarity.

2.1 Estimation Of Surface Orientation: Similarity Of Theory

Chapter 3 presented several examples of surface shape being estimated using a local analysis of shading information. For the shading-only examples (figures 3-7, 3-9, 3-10), the estimates of surface shape are very much like the shape people perceive. This is true even for images such as the Hindenburg image (see figure 3-7) where people incorrectly perceive the surface shape. From these examples it appears that the theory of this thesis may provide a useful explanation for the human perception of shape from shading information.

Unfortunately, even in these simple images people have considerable additional information available to them, including the global configuration of shading, the bounding contour, general knowledge about object shape and so forth. Thus the similarity of the perceived shape to the shape estimated by this theory might well be the result of a different analysis of the image which also yields a generally correct estimate of shape. Brady (1981) has suggested that an algorithm such as developed by Horn and Ikeuchi, using a few standard reflectance functions, might provide such an alternative explaination; and Ullman (1981) raised the question of whether the data might reasonably be accounted for by a relatively crude, qualitative shape-from-shading algorithm. Thus before we can claim that this theory provides an account of one aspect of human perception we must (1) demonstrate that people are able to use a local analysis of shading information to determine surface shape, and (2) demonstrate that there are aspects of human perception of shape which cannot be accounted for by other theories.

Can people use local shading information to estimate surface shape?. In order to answer this question the following experiment was designed: Four small, square patches were cut from an image of a sphere. The size of the patches was chosen to be just large enough that changes in the shading were visible, thus allowing the second derivative of image intensity to be observed. No contour or texture information was present in the patches. Subjects were told that they were going to be shown several small patches cut from some pictures, and that they were to indicate the orientation of the imaged surface by positioning the image patch at the appropriate place (the place with the same surface orientation) on a drawing showing the bounding contour of a sphere. They were not told the

illuminant direction, however all subjects who reported considering the illuminant direction made the correct assumption about where the illumination was coming from. Subjects were then presented with the image patches one at a time, they indicated the surface orientation for each patch, and then the patch was removed from sight. Presentation order was randomized.

Figure 4-1 shows the results of this experiment. The surface tilt estimated by the subjects ($N = 12$) is plotted against the true surface tilt. The error bars show the standard deviation of the estimates. As can be seen, the subjects showed impressive accuracy in estimating the surface tilt; in fact the correlation between true and estimated tilt is 0.9926, so that the variable of surface tilt accounts for 98.5%of the variance in this data. The dotted line in figure 4-1 shows the regression line between the true and estimated tilt; the coefficients are different from zero at the $p = 0.01$ confidence level.

The slant estimates cannot be so easily determined from this data, because the experimental paradigm makes some slant estimates unlikely — for instance, people resist placing the patches on top of the occluding contour, so that the largest surface slants will never be chosen. Subjects show a similar resistance to placing the patches in the center of the occluding contour, so that the smallest surface slants are never chosen. If we take the range of slants subjects will actually chose to correspond to the range $0 - \pi/2$, then we can obtain a value for the subjects estimates of surface slant. In this case the subjects showed a bias of $+6.69°$, and a standard deviation of $16.35°$.

This experiment demonstrates that people are able to utilize purely local image information to determine the orientation of the surface. The fact that people are able to estimate surface shape from a local analysis of shading, and the fact that the estimates of shape produced by such an analysis are similar to the human percept make it reasonable to believe that people obtain information about surface shape through use of the local analysis of shading developed in this thesis[1] .

Errors in shape perception. Do the errors people make in estimating shape from shading occur as predicted by the theory? Two examples have been found of images for which people are unable to correctly utilize shading: one is when curvature and foreshortening "cancel" as in the Hindenburg

[1]It is not reasonable to claim that people use *only* local analysis to determine surface shape, of course, as will be discussed in the following sections.

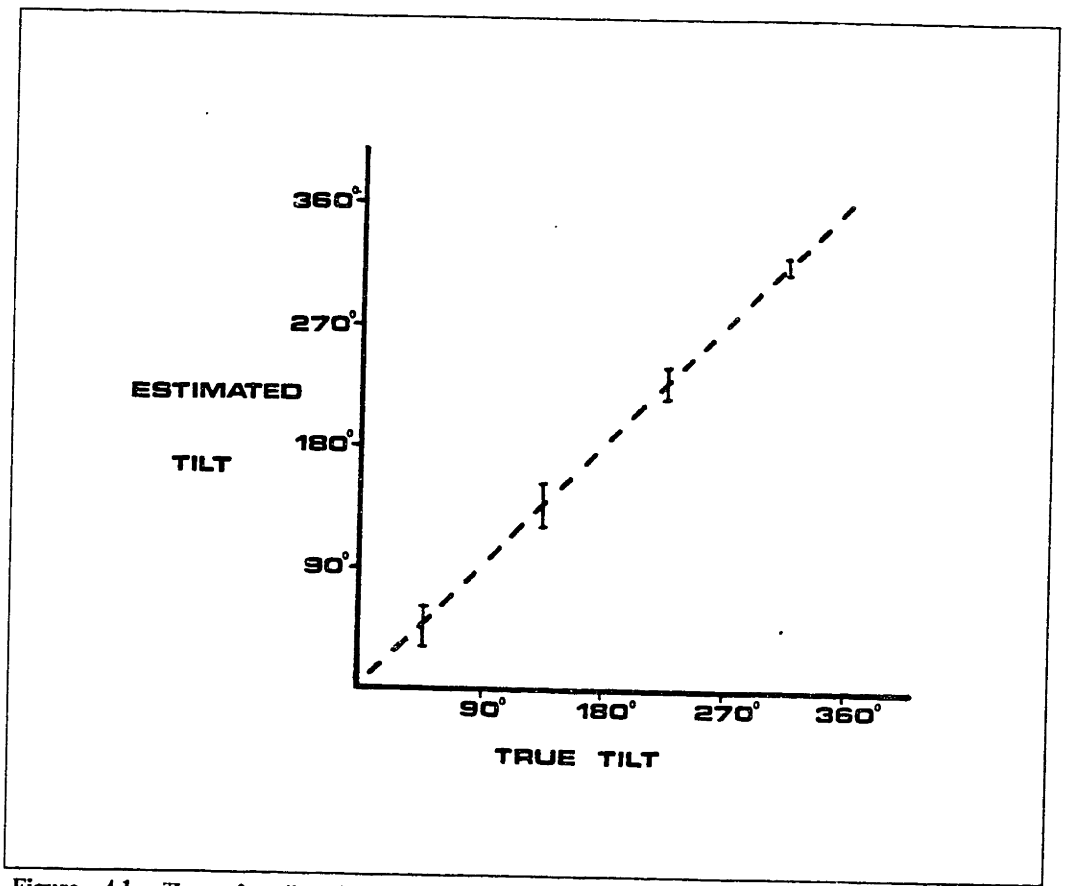


Figure 4-1. The surface ilt estimated by the subjects ($N = 12$) is plotted against the true surface tilt. The error bars show the standard deviation of the estimates. As can be seen, the subjects showed impressive accuracy in estimating the surface tilt; in fact the correlation between true and estimated tilt is 0.9926, so that the variable of surface tilt accounts for 98.5%of the variance in this data. The dotted line shows the regression line between the true and estimated tilt; the coefficients are different from zero at the $p = 0.01$ confidence level.

illusion (see figure 0-1 (A)), and the other is the inability of people to correctly utilize the shading in photographic negatives. In both of these cases the theory predicts the human percept.

Underestimation of surface relief. In the Hindenburg illusion the relief of the surface is underestimated, so that the figure looks flatter than it actually is. In this image shading is the only source of information about surface shape interior to the bounding contour. This underestimation of relief is as predicted by the theory, as demonstrated by figure 3-8 which shows the result of local estimation of shape for the Hindenburg image. Thus both the human visual system and this theory show the same underestimation of surface relief when surface curvature is arranged so as to partially cancel the effects of foreshortening. This underestimation would not be expected if people were performing

a global analysis of the shading information; for such an analysis the relationship of curvature and foreshortening is a matter of complete indifference.

Inability to use shading in photographic negatives. People are very poor at using the shading in photographic negatives, as demonstrated by figure 4-2 (A). On the left side of figure 4-2 (A) is a normal image, while on the right side is a photographic negative. The photographic negative is disturbing to most people, and the shape of the surface is unclear even though the contour and texture information is the same as in the positive image.

In photographic negatives, the image intensity of a surface patch i with a Lambertian reflectance function, albedo ρ_i, and illuminated with intensity λ is

$$I = \max_j(\rho_j \lambda(\mathbf{N}_j \cdot \mathbf{L})) - \rho_i \lambda(\mathbf{N}_i \cdot \mathbf{L})$$

Thus the point k where $k = \max_j(\rho_j \lambda(\mathbf{N}_j \cdot \mathbf{L}))$, i.e., the brightest point in the original image, now assumes an intensity value of zero, and the darkest point in the original image now becomes the brightest point. Let us investigate what happens to our estimator of slant with photographic negatives. The slant estimator was

$$z_N = \sigma_\kappa \left(\| \frac{\nabla^2 I}{I} \| - \sigma_\kappa^2 \right)^{-\frac{1}{2}} \tag{16}$$

where z_N is the z component of the surface normal, equal to the arccosine of the surface slant, and σ_κ^2 is the variance of the surface curvature distribution. This estimator of $\mathbf{z}_N$ was predicated upon the following identities, as explained in chapter 2 section 4:

$$\begin{aligned} \frac{d^2 I}{I} &= \frac{\rho \lambda d^2 \mathbf{N} \cdot \mathbf{L}}{\rho \lambda \mathbf{N} \cdot \mathbf{L}} \\ &= \frac{-\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L})du^2 - \kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L})dv^2}{\mathbf{N} \cdot \mathbf{L}} \\ &= \frac{-\kappa_1^2(\mathbf{N} \cdot \mathbf{L})du^2 - \kappa_2^2(\mathbf{N} \cdot \mathbf{L})dv^2}{\mathbf{N} \cdot \mathbf{L}} \\ &= -\kappa_1^2 du^2 - \kappa_2^2 dv^2 \end{aligned}$$

The key steps in deriving this relationship is that the terms ρ, λ and $\mathbf{N} \cdot \mathbf{L}$ all divide out. But if we

substitute the equation for image intensity we derived for photographic negatives we find:

$$I = \max_j(\rho_j \lambda(\mathbf{N}_j \cdot \mathbf{L})) - \rho_i \lambda(\mathbf{N}_i \cdot \mathbf{L})$$

$$dI = -\rho_i \lambda(d\mathbf{N}_i \cdot \mathbf{L})$$
$$d^2I = -\rho_i \lambda(d^2\mathbf{N}_i \cdot \mathbf{L})$$

and thus

$$\frac{d^2I}{I} = \frac{-\rho\lambda d^2\mathbf{N} \cdot \mathbf{L}}{\max_j(\rho_j\lambda(\mathbf{N}_j \cdot \mathbf{L})) - \rho\lambda\mathbf{N} \cdot \mathbf{L}}$$
$$= \frac{\rho(\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L})du^2 + \kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L})dv^2)}{\max_j(\rho_j(\mathbf{N}_j \cdot \mathbf{L}) - \rho\mathbf{N} \cdot \mathbf{L}}$$
$$=??????$$

We can go no farther in the chain of deductions we forged before; the terms stubbornly refuse to disappear, and thus the quantity $\nabla^2 I/I$, which before was a measure of the slant of the surface, is now a meaningless quantity. Thus when applied to photographic negatives, this algorithm for estimating surface shape produces erroneous results (see also figure 4-3).

Thus both the theory and human perception are unable to utilize the shading in photographic negatives. Note that peoples inability to utilize shading information in negatives cannot be explained by the fact that brightness relations are reversed in negatives (i.e., they are generally brighter at edges and darker where the surface faces the viewer), because people are fully able to utilize shading information in electron microscope images, which have the same brightness relations as photographic negatives, as shown by figure 4-2 (B) (*left side*).

The failure of human perception on negatives, but not on electron microscope images, rules out any theory in which shading information is used in only a rough, qualitative manner. Any such theory would perforce perform as well on both photographic negatives and electron microscope images. In addition, the ability of people to utilize the shading in EM images shows that people cannot be using shading in a Ikeuchi and Horn style algorithm, because such an algorithm functions by making use of a few standard reflectance maps. EM images have a reflectance map radically unlike the reflectance map of any naturally occuring scene, thus it is implausable that the human visual system has somehow provided such a reflectance map. Further, if by some chance people had use of such a reflectance map they would be able to adequately function on photographic negatives — because, as shown by Ikeuchi

and Horn, their algorithm is quite insensitive to small changes in the reflectance map so that the difference between the map for an EM image and that for a photographic negative should produce only a small difference in the estimated shape, instead of the relatively massive collapse experienced by people.

Thus there are three points of evidence which indicate similarity between the theory of this thesis and the theory which accounts for the human perception of shape from shading information:

(1) People are able to use local shading information to estimate surface shape.

(2) Estimates of surface shape made from a local analysis of shading as per this theory are similar to human estimates of shape — even for EM images.

(3) Both make errors on the same images: when curvature cancels foreshortening, and on photographic negatives.

These three points appear to indicate that people do in fact make substantial use of a local analysis of shading information, as described in this thesis, to determine shape.

2.2 Estimation Of Surface Orientation: Similarity Of Algorithm

One of the more interesting aspects of this theory of local shape estimation is that is does not work for photographic negatives — just as people cannot use the shading information in photographic negatives. This observation is evidence of similarity between the theory of this thesis, and that of the human visual system. An example demonstrating the similarity of algorithm is offered by electron microscope (EM) images.

Electron microscopes do not produce pictures by a normal imaging process, rather they operate by bombarding a sample with electrons and recording voltage potentials nearby, as diagrammed in figure 4-2 (C). This unusual process produces images where the reflectance function is approximately the reciprocal of the Lambertian reflectance function, i.e.

$$I = k(\mathbf{N} \cdot \mathbf{V})^{-1}$$

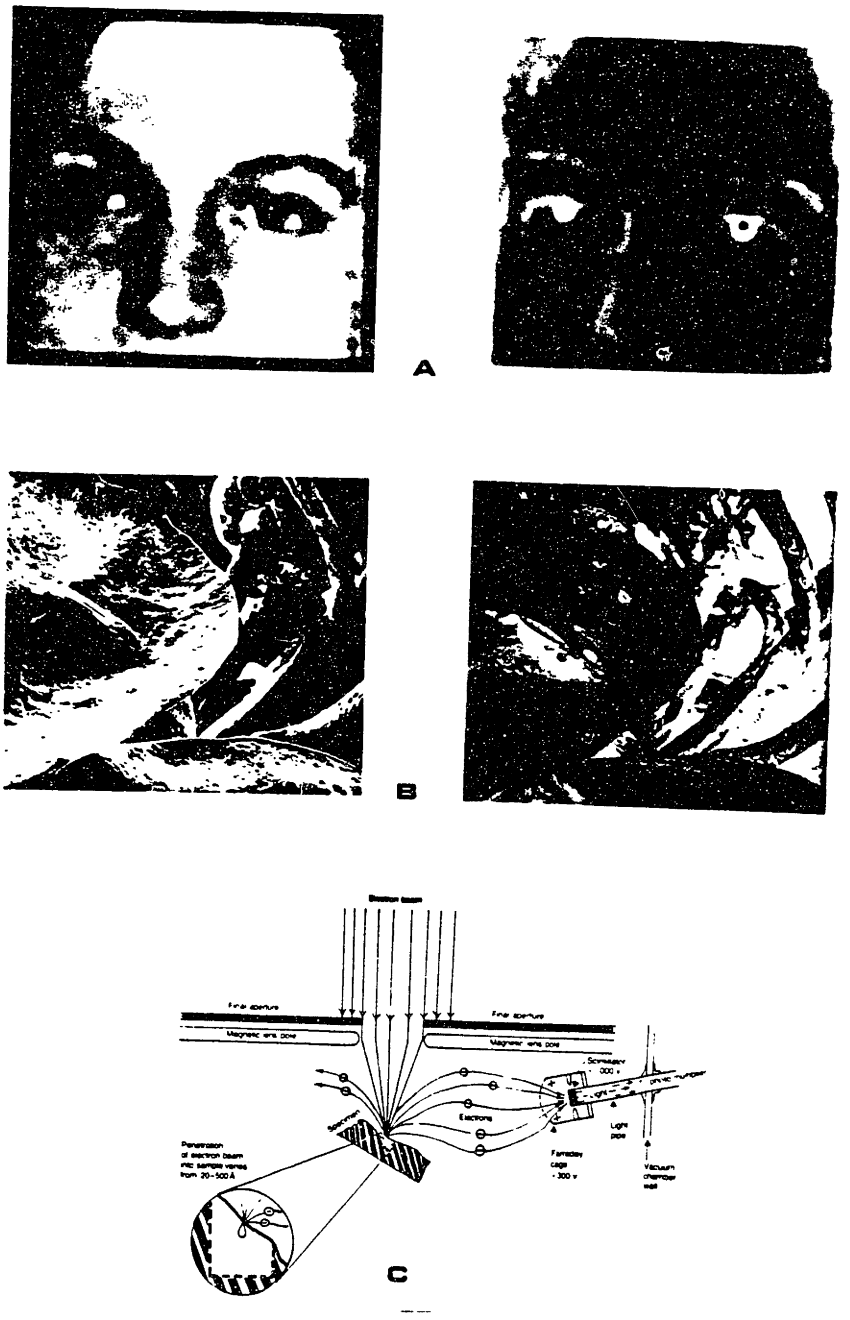


Figure 4-2. (A) A normal photograph (*left*) and its photographic negative (*right*). Note that shape is difficult to perceive in the the negative image, even though both the positive and the negative have the same contour information. (B) An Electron Microscope (EM) image (*left*) and its photographic negative (*right*). Note that even though the EM image superficially looks like the photographic negative, we can easily use the shading information in this image. Further, we seemingly cannot use the shading information in the photographic negative of the EM image, even though this negative superficially looks like a normal image. (C) Electron microscopes do not produce pictures by a normal imaging process, rather they operate by bombarding a sample with electrons and recording voltage potentials nearby.

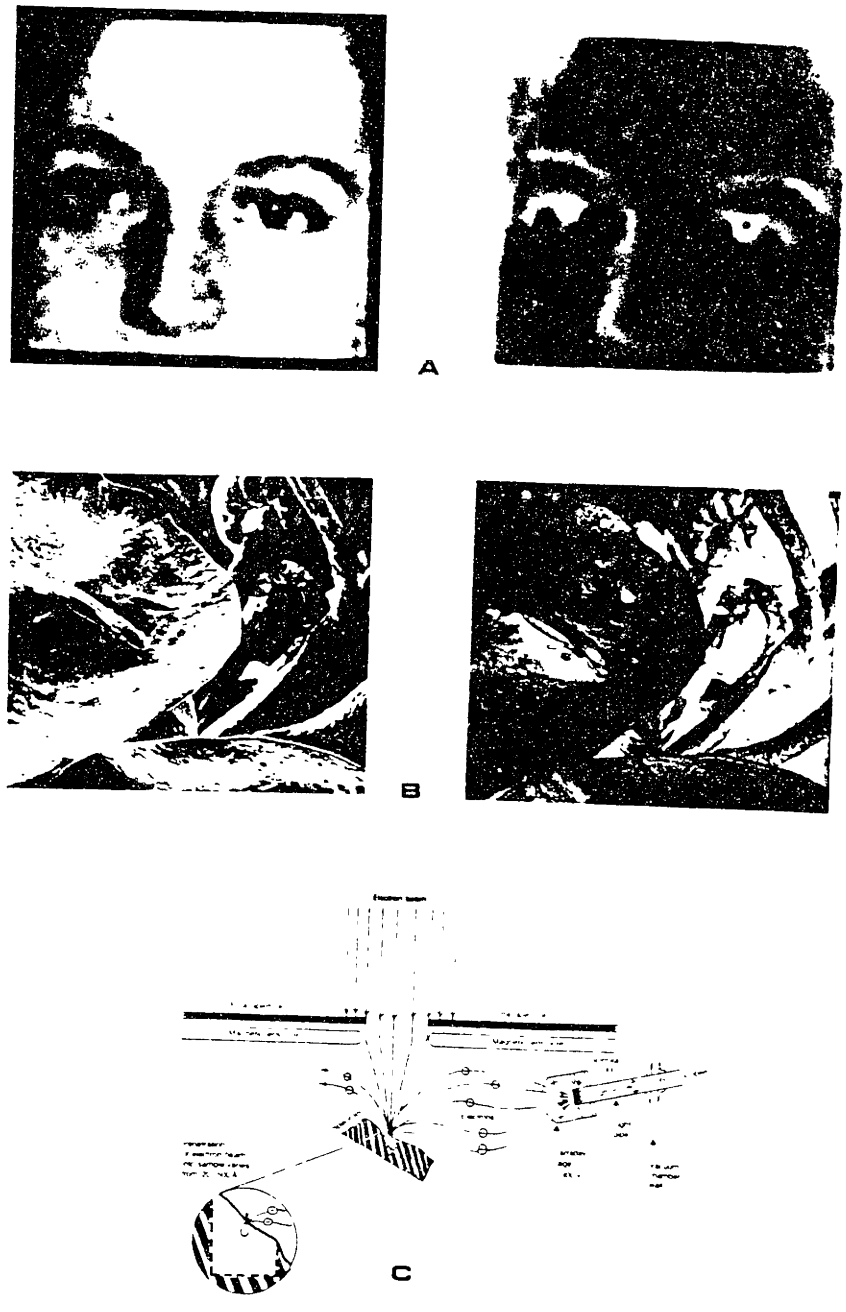


Figure 4-2. (A) A normal photograph (*left*) and its photographic negative (*right*). Note that shape is difficult to perceive in the the negative image, even though both the positive and the negative have the same contour information. (B) An Electron Microscope (EM) image (*left*) and its photographic negative (*right*). Note that even though the EM image superficially looks like the photographic negative, we can easily use the shading information in this image. Further, we seemingly cannot use the shading information in the photographic negative of the EM image, even though this negative superficially looks like a normal image. (C) Electron microscopes do not produce pictures by a normal imaging process, rather they operate by bombarding a sample with electrons and recording voltage potentials nearby.

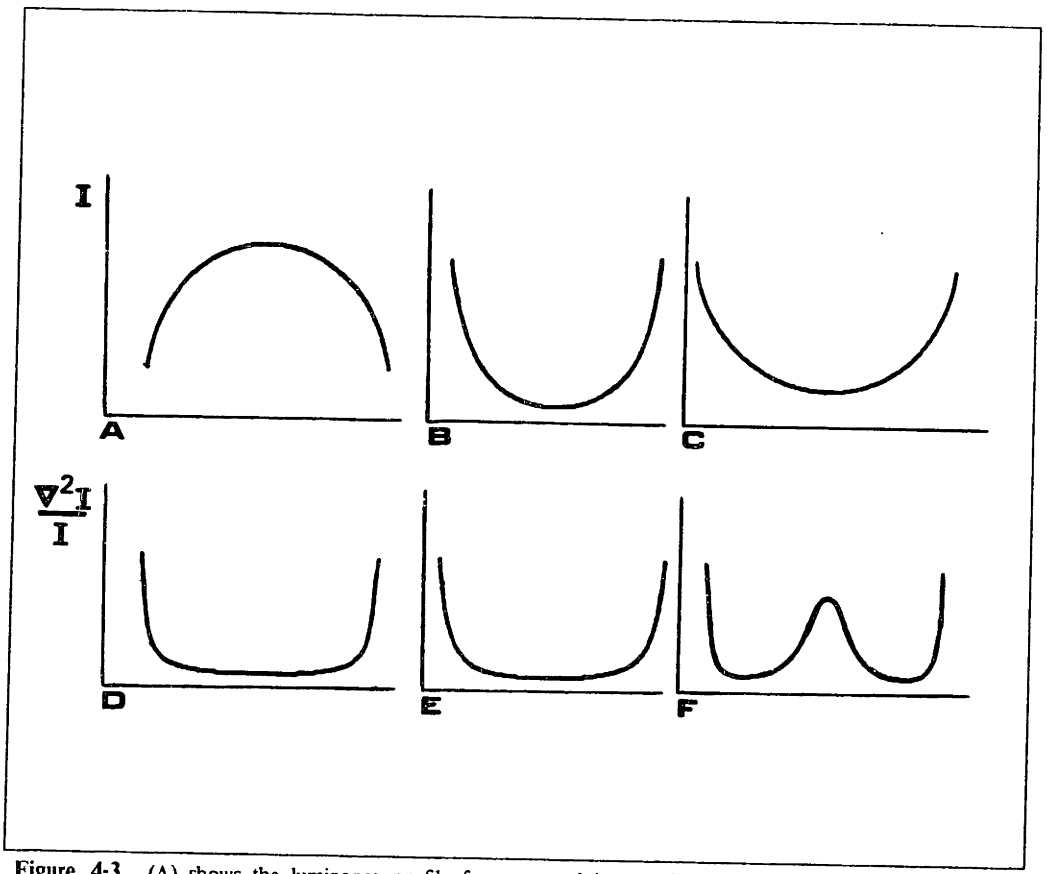


Figure 4-3. (A) shows the luminance profile for a normal image of a sphere, (B) shows the luminance profile for an EM image of a sphere, and (C) shows the luminance profile for a photographic negative of a normal image. Note that while both (B) and (C) have reversed contrast relative to (A), the profile of (B) is both narrower than (C) and flatter on the top. (D) shows the values of $\nabla^2 I/I$ across the normal sphere image, (E) shows the values of $\nabla^2 I/I$ across the simulated EM image of a sphere, and (F) shows the values of $\nabla^2 I/I$ across the photographic negative of the sphere image. This difference between (F) and (D) or (E) proves to be critical when we go to estimate the object's shape, as the slant of the surface is estimated using $\nabla^2 I/I$. Thus for (D) we correctly obtain a sphere, for the EM image we use (E) and obtain a shape much like (B), whereas for (F) we obtain only bizzare and inconsistent surfaces. Photographic negatives of EM images suffer the same fate as photographic negatives of normal images, producing values of $\nabla^2 I/I$ much like those shown (F). Thus the shape estimators developed in this thesis produce a correct estimate of shape for a normal sphere image, an almost correct estimate of shape for an EM image of a sphere, and fail for the photographic negatives of both normal and EM images — which agrees with human perception (see section 4 of the appendix for more detail).

where k is a constant (Ikeuchi and Horn 1981). These images look superficially very much like photographic negatives, as shown by the normal EM image on the left side of figure 4-2 (B).

Due to some quirk of the human visual system, people are able to correctly perceive shape these odd images. Further, people are *unable* to utilize the shading information in the photographic negatives of EM images, even though negatives of EM images have the same brightness relations as in

normal images (see figure 4-2 (A) – on the left is a normal image, on the right a photographic negative, and (B) — on the left is a normal EM image, on the right the photographic negative of the image).

The ability of people to use the shading in an EM image is a *mistake* of the visual system, as such images are impossible in normal human experience. It is however a fortuitous mistake which results in the correct impression of shape. The inability of people to utilize the shading in the negatives of EM images (which otherwise look much more like normal images than do EM images) underlines the accidental nature of peoples ability to use the shading in EM images.

It is therefore of considerable interest to note that the estimators of surface shape developed in this thesis still function with the reflectance function of an EM image — and not with the negative of an EM image. The correct functioning of the shape estimators on an EM image was demonstrated in figure 3-13. The reason the shape estimators function with EM images is shown in figure 4-3. Figure 4-3 (A) shows the luminance profile for a normal image of a sphere, figure 4-3 (B) shows the luminance profile for an EM image of a sphere, and figure 4-3 (C) shows the luminance profile for a photographic negative of a normal image. Note that while both (B) and (C) have reversed contrast relative to (A), the profile of (B) is both narrower than (C) and flatter on the top. This difference in luminance profile proves to be critical when we go to estimate the object's shape. Recall that the slant of the surface was found to be inversely proportional to the square root of $\nabla^2 I/I$, and so to estimate the slant we first computed $\nabla^2 I/I$. Figure 4-3 (D) shows the magnitude of $\nabla^2 I/I$ across the normal sphere image, figure 4-3 (E) shows the magnitude of $\nabla^2 I/I$ across the EM image of a sphere, and figure 4-3 (F) shows the magnitude of $\nabla^2 I/I$ across the photographic negative of the sphere image. It is clear that (D) and (E) are quite similar, there is only a small difference between them, which leads to a very slight flattening of the estimated surface in the EM image (i.e., a shape much like the EM image's luminance profile). On the other hand, (F) is quite different from (D), so that a shape estimate based on (F) would result in bizzare and inconsistent surfaces. Photographic negatives of EM images suffer the same fate as photographic negatives of normal images, producing values of $\nabla^2 I/I$ much

like those shown in figure 4-3 (F). Thus the shape estimators developed in this thesis produce a correct estimate of shape for a normal sphere image, an almost-correct estimate of shape for an EM image of a sphere, and fail miserably for the photographic negatives of both normal and EM images — just like people (for more detail see section 4 of the appendix).

The fact that both the estimators of this theory and the human visual system make the same fortuitous "mistakes" which result in the correct interpretation of the shading in EM images, and incorrect interpretations for photographic negatives, is strong evidence of the similarity between the algorithm developed here and the algorithm employed by the human visual system.

2.3 Constraint Of Contour On Shading: Similarity Of Theory

We have seen that people are able to use local shading to estimate surface shape — and that people make systematic errors in the interpretation of shape when *only* shading information is present. In chapter 2 we showed that the three-dimensional interpretation of imaged contour may be used to constrain the shading-derived estimates of surface shape, and examples of this constraint were shown in chapter 3. This constraint provides a sufficient explanation of the human perception of shape, both correct and incorrect, in all of the examples examined.

The Hindenburg image provides a controlled situation in which we may investigate the interaction of contour with shading in a situation where the shading alone is insufficient to provide the correct impression of shape. Figure 3-18 (A) shows a shading-only Hindenburg image, in which the true relief of the surface is not apparent. In figures 3-18 (B) and (C) a contour has been added to the shading-only image; this contour brings out the correct relief of the Hindenburg figure.

Figure 3-18 (C) is interesting because although it gives a correct impression of shape, it actually contains no more information about shape than figure 3-18 (A) which gives a flat impression of shape. All that has been changed between figures 3-18 (A) and (C) is to distinguish some of the iso-brightness contours of (A). Because distinguishing some of the iso-brightness contours does not add any information about shape, this example demonstrates that the improved impression of shape

to the image.

Figure 3-18 (D) shows an example where the addition of contours to the image does not result in an improved impression of shape. Thus the addition of contour information is not always sufficient to show the relief of the imaged surface; any theory of the interaction of contour and shading must be able to explain failures such as this.

Chapter 3 demonstrated how the theory of this thesis uses imaged contour to provide the necessary constraint to obtain a correct estimate of shape in figures 3-18 (B) and (C), and why the theory cannot employ the imaged contours in figure 3-18 (D) to obtain an accurate estimate of shape. A graphic explanation of how the theory uses contour to constrain the interpretation of shape in these images is contained in figure 3-19. In these three examples the constraint of contour on shading, as developed in chapter 2, is sufficient to account for the human percept[2] .

Several other synthetic images have been examined, and in each case this mechanism correctly accounts for the human percept. Figure 3-20 shows a natural image in which the the contour and shading information have been used together to produce a correct estimate of shape, an estimate which agrees with the human percept.

Thus the theory of chapter 2 provides a sufficient account in the examples for which people are able to employ constraint from contour to determine shape correctly. Further, in those examples where the theory is unable to use the contour information to determine shape, people are also unable to make use of contour information. The agreement of this theory with the human percept, on both examples where people are successfully able to estimate shape and on examples where people incorrectly estimate shape, provides strong evidence that the theory by which people employ contour to estimate shape is similar to the theory developed here. There are of course likely to be several ways in which people use imaged contour to constrain the interpretation of shape; in light of these examples,

[2]People appear to use certain assumptions in order to interpret line drawings without shading (Stevens 1980), for instance, that the contours are geodesics or that the surface is "locally cylindrical" in the neighborhood of the contour. In these examples the contours are not seen to be geodesics, nor is the surface "locally cylindrical", thus people seem to make stronger assumptions about the relationship between contour and surface in the absence of shading than when shading is present.

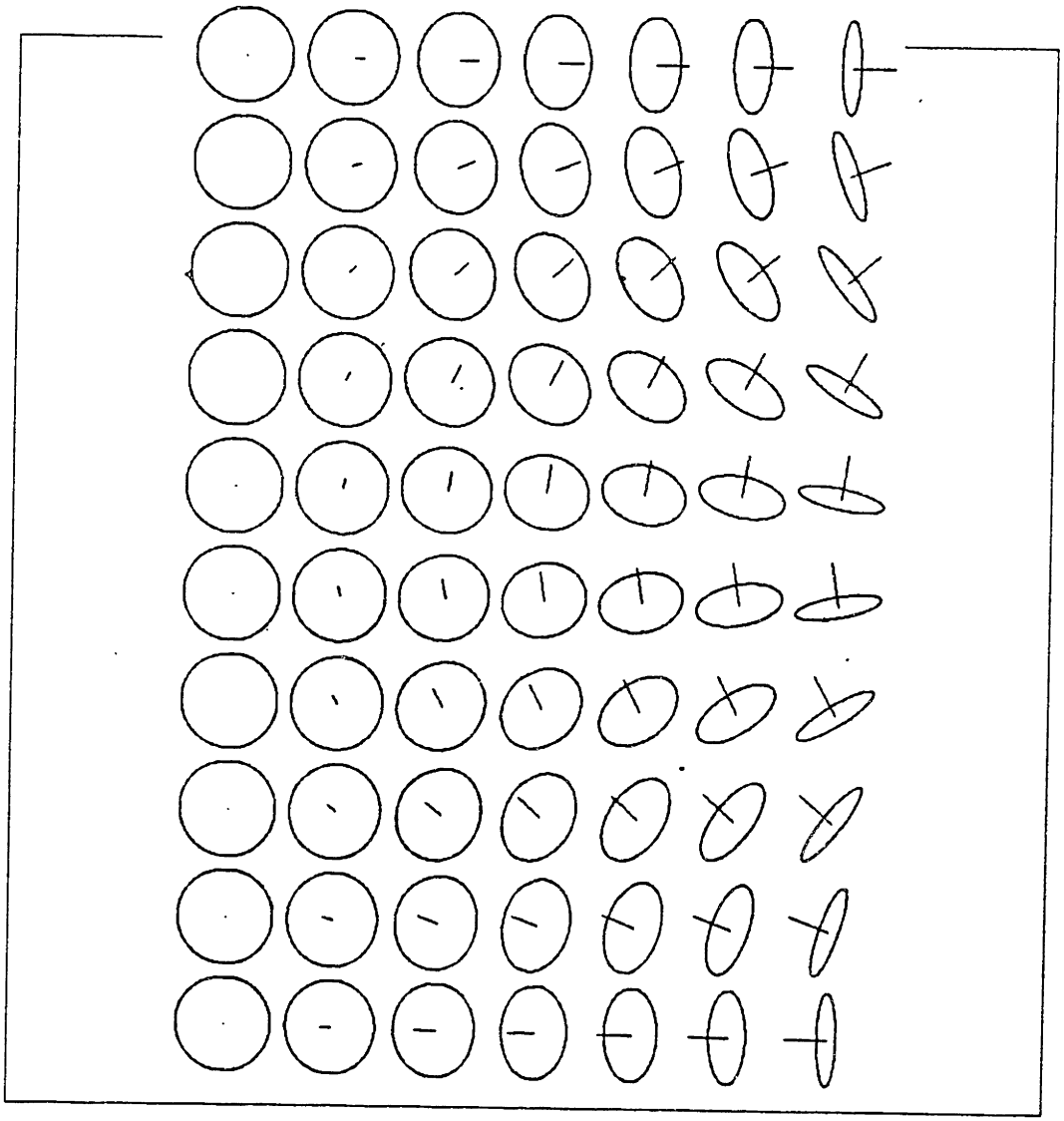


Figure 4-4. The probe sheet given to human subjects in order to estimate illuminant direction. Subjects were asked to pick the disk whose surface normal (the "nail" stuck into the disk) pointed toward the illuminant.

however, it seems likely that the mechanism discussed here plays a significant part in the process.

2.4 Estimation Of Illuminant Direction: Similarity Of Theory

Chapter 2 developed a maximum-likelihood estimator for the illuminant direction; in chapter 3 this estimator was tested on synthesized images and natural images and found to accurately estimate the illuminant direction. In order to compare the theory on which this estimator was developed to the human visual system we must first determine how well people can estimate the illuminant direc-

tion, after we have accomplished this we can compare human performance to that of the maximum-likelihood estimator (see also Pentland 1979, 1982 a, b).

Human Accuracy at estimating illuminant direction. Subjects ($N = 15$) were given a probe sheet which contained pictures of disks at various orientations, with the surface normal drawn in. This is shown in figure 4-4. Subjects were then shown a series of pictures of spheres illuminated from various directions, and asked to pick the disk-probe whose surface normal most closely pointed toward the illuminant. The results of this experiment show that the human subjects can estimate both the tilt τ_L of the illuminant direction and the slant σ_L of the illuminant direction, with a standard deviation of less than twelve degrees.[3] This experiment provides us with an lower limit on the accuracy of the estimates of illumination direction given by subjects when using the probe-sheet method, and thus an estimate of the experimental error inherent in this testing situation.

Estimates of Illuminant Direction for Natural Images. The next experiment was to have the human subjects (again, $N = 15$) use the probe sheet to estimate the illuminant direction for a series of digitized pictures of natural objects such as rocks and logs. The photographs of these objects were made under natural illumination conditions, so that these images of natural objects do not have a point-source illuminant Digitized pictures were used so that both the human subjects and our maximum-likelihood estimator, which was implemented on a computer, would receive exactly the same image information.

Figure 4-5 shows a comparison of the estimates of illuminant direction given by human subjects and the estimate of illuminant direction given by the maximum-likelihood estimator. The slant component of the illuminant direction (the inclination in depth, away from the image plane) and the tilt component (the image-plane component of the direction) are shown separately. The previous experiment found that the mean of the estimates given by human subjects had an error of up to ten degrees when using the probe-sheet method to estimate the illuminant direction in images of spheres.

[3]This figure of twelve degrees is actually somewhat *better* than the subjects ability to estimate local surface orientation under the same conditions (Stevens, 1979). This shows that the estimates given by subjects couldn't have been made by, for instance, determining the surface orientation of the brightest point on the surface.

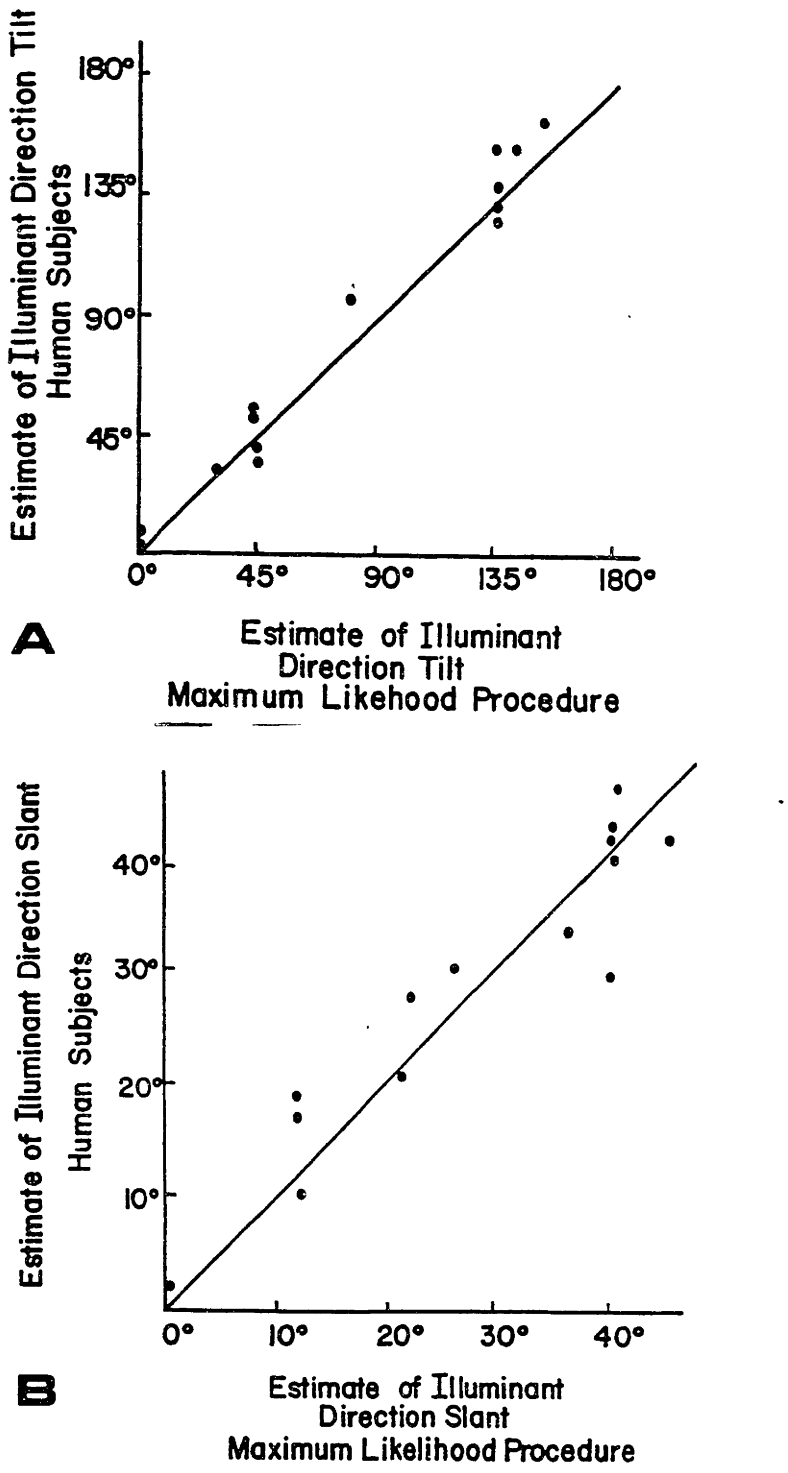


Figure 4-5. A comparison of the estimates of illuminant direction given by human subjects and the maximum-likelihood estimator, for fourteen images of natural objects. (A) shows the comparison for the tilt component of the illuminant direction, and (B) shows the comparison for the slant component of the illuminant direction.

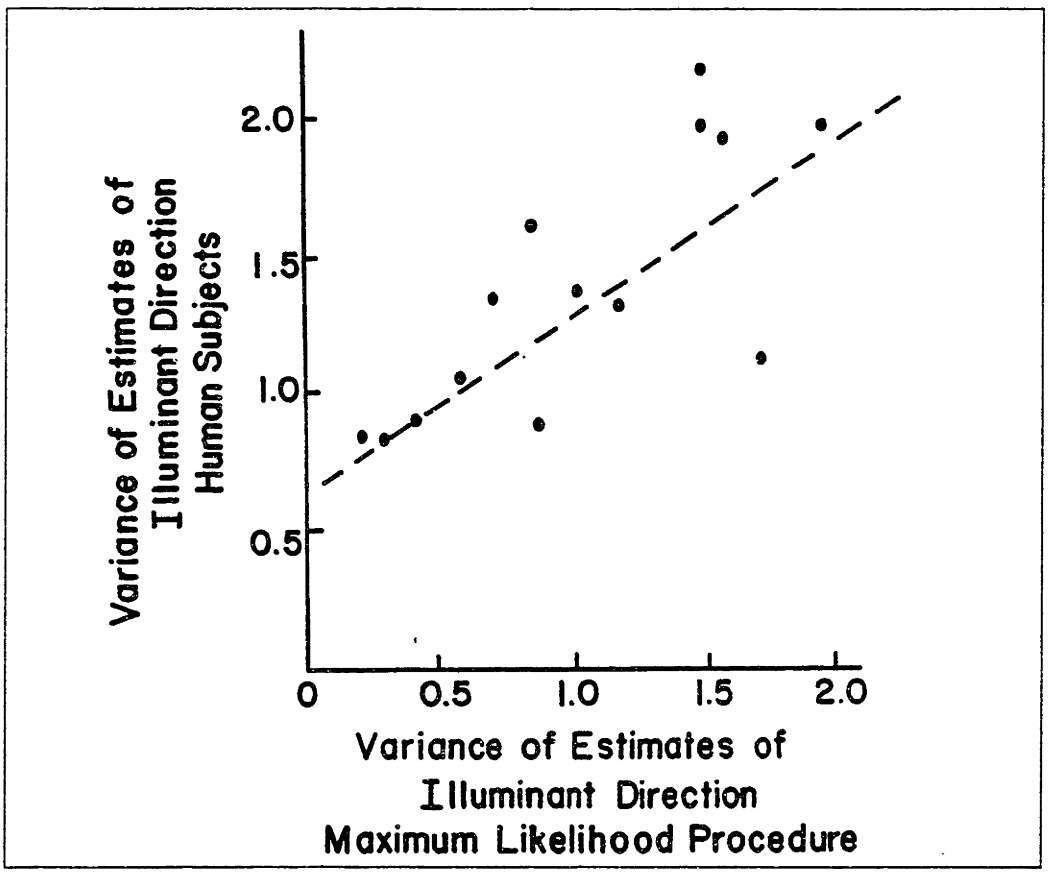


Figure 4-6. The relationship between the variance of the estimates of illuminant direction given by human subjects, and the variance of the estimate of illuminant direction produced by the maximum-likelihood estimator. There is a correlation of 0.63 between the variance of the two sets of estimates, significant at the $p \leq 0.05$ level. The dashed line is the linear regression line relating the variance of the two estimation procedures.

Thus figure 4-5 shows that for images of natural objects, the estimates given by human subjects and the computer-generated estimates agree to within the experimental error which would be expected on the basis of the results of the previous experiment. We conclude then that there is no significant difference between the two estimates of illuminant direction in this experiment.

One further piece of evidence of the equivalence of the theory underlying the human and maximum-likelihood estimates comes from the variance of the two estimates. The maximum-likelihood estimator generates, for each image, a confidence statistic along with its estimate. This confidence statistic is proportional to the variance of the estimate for that image (given the assumptions of the procedure). We can compare the variance of peoples estimates for a particular picture to the variance of the maximum-likelihood estimate (as predicted by the confidence statistic).

Figure 4-6 shows the results of comparing the variance of the estimates of illuminant direction given by human subjects to the variance of the estimate of illuminant direction produced by the maximum-likelihood estimator. There is a correlation of 0.63 between the variance of the two sets of estimates, significant at the $p = 0.05$ level. The linear regression line relating the human and maximum-likelihood variance is shown as a dashed line in figure 4-6; the coefficients of the regression are significantly different from zero at the $p = 0.01$ level. The significant relationship between the variance of the two estimation procedures (human and maximum-likelihood) shows that when one of the two finds enough information in the image to make a low-variance estimate, so does the other estimation procedure. This then is evidence showing that the human and maximum-likelihood estimation procedures utilize equivalent information for making their estimates.

Thus there are two points of evidence which show that the theory developed in this thesis is and the method used by human subjects for estimating illuminant direction are based on equivalent assumptions and utilize equivalent information:

(1) Both give the same estimates (to within experimental error) for both artificial scenes of illuminated spheres, and for images of natural scenes consisting of pictures of rocks and logs.

(2) There is a significant correlation between the variance of the human and maximum-likelihood estimates of illuminant direction. This shows that the two methods of estimation utilize equivalent image information.

2.5 Estimation Of Illuminant Direction: Similarity Of Algorithm

One of the example images is of particular importance, because it is an example where people incorrectly estimate the illuminant direction. Because people mis-estimate the illuminant direction in this image, it may be used as a test of the similarity of algorithm. When the image of the rock shown in figure 4-7 was given to human subjects, they erroneously estimated the illuminant direction by about 120 degrees (it is actually illuminated from the left top, not the right top as all but two of the fifteen subjects reported). What is significant is that the estimate given by the maximum

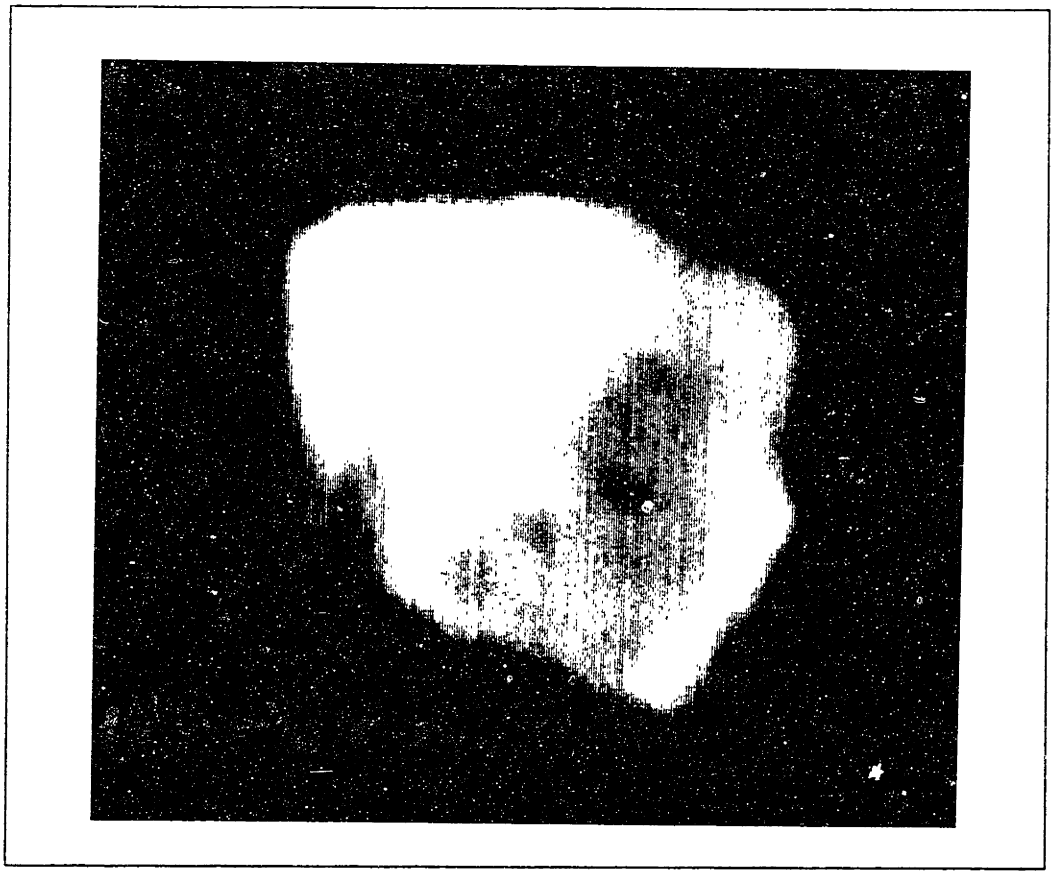


Figure 4-7. Picture of a rock for which both human estimates of illuminant direction and the maximum likelihood estimate agreed, but were objectively wrong. Actual illumination direction is top right, not top left as reported by all but two of the fifteen human subjects

likelihood estimation procedure agreed with the estimate given by the human subjects, even though both estimates of illuminant direction were objectively wrong. This image is a case which violates the assumptions on which the human estimates of illuminant direction are based; the human estimate was objectively wrong. The special significance of this case is that it also violates the assumptions of our maximum likelihood estimation procedure in such a manner that it produces exactly the same estimate as the human subjects. This is strong evidence that the algorithm by which people estimate illuminant direction is similar to the algorithm described here.

2.6 Secondary Characteristics

One of the prime tests of the similarity of algorithm is that the secondary characteristics of the

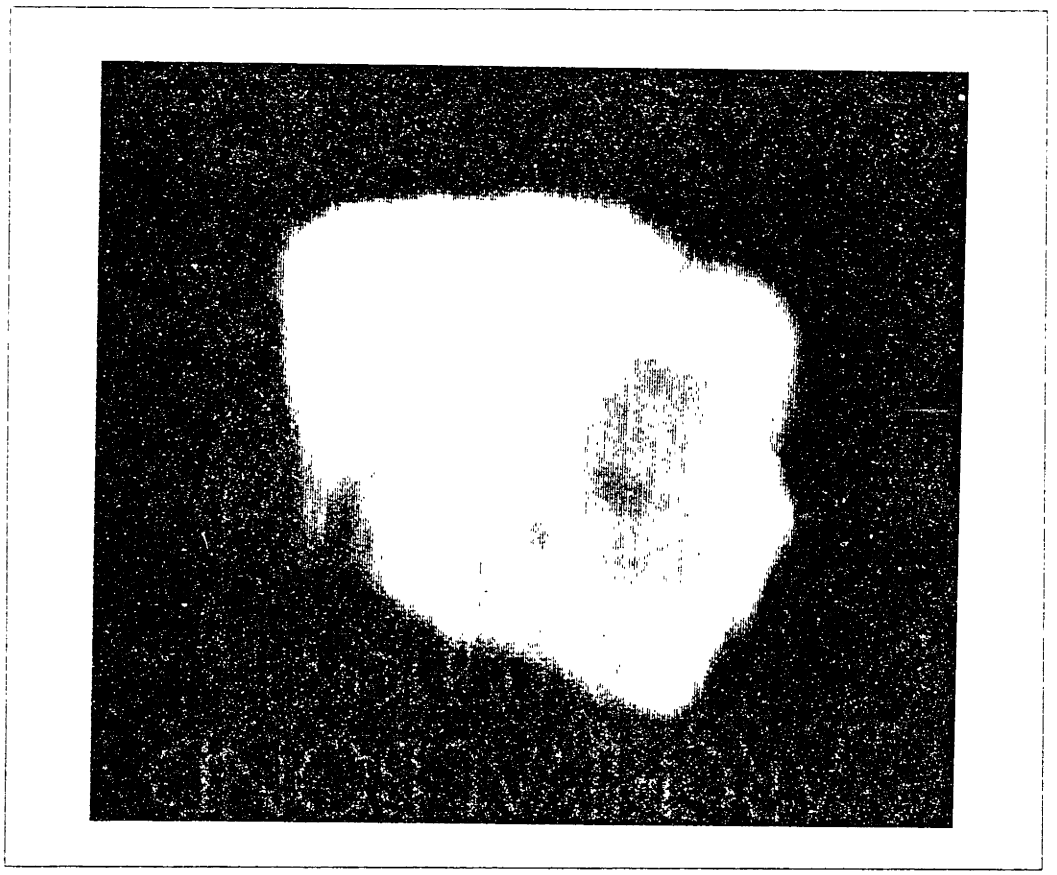


Figure 4-7. Picture of a rock for which both human estimates of illuminant direction and the maximum likelihood estimate agreed, but were objectively wrong. Actual illumination direction is top right, not top left as reported by all but two of the fifteen human subjects.

likelihood estimation procedure agreed with the estimate given by the human subjects, even though both estimates of illuminant direction were objectively wrong. This image is a case which violates the assumptions on which the human esumates of illuminant direction are based: the human estimate was objectively wrong. The special significance of this case is that it also violates the assumptions of our maximum likelihood esumation procedure in such a manner that it produces exactly the same estimate as the human subjects. This is strong evidence that the algorithm by which people estimate illuminant direction is similar to the algorithm described here.

2.6 Secondary Characteristics

One of the prime tests of the similarity of algorithm is that the secondary characteristics of the

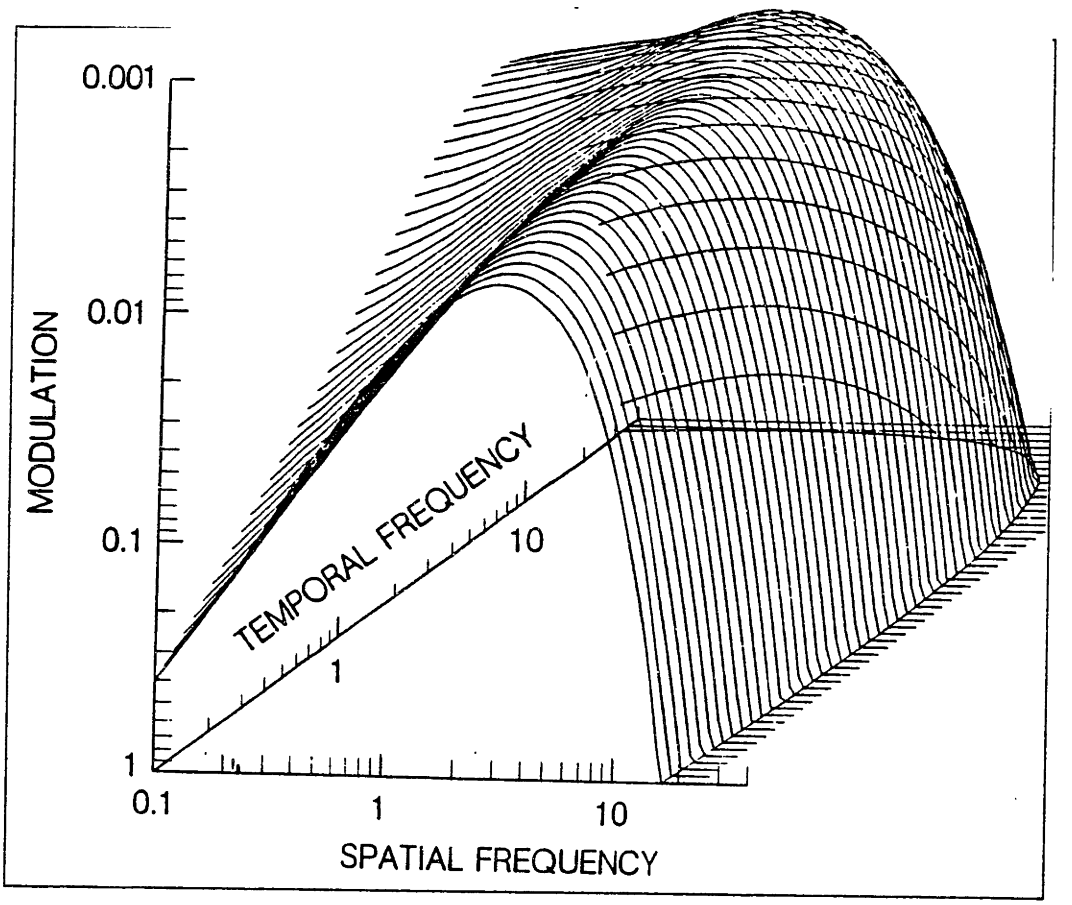


Figure 4-8. The human spatio-temporal threshold surface, measured using stabilized drifting gratings (Kelly 1979).

two systems must be the same. In the case of vision, the most thoroughly investigated "secondary characteristic" is spatial frequency sensitivity. Thus we should ask whether the algorithm proposed in chapter 3 has the same spatial frequency sensitivity as the human visual system.

The algorithm proposed in chapter 3 has multiple possible spatial frequency characteristics. If one considers the $\nabla^2 G(x, y)$ filter used to measure $\nabla^2 I$, or the $d^2 G(x, y)$ filter used to measure $d^2 I$, then the sensitivity to changes in image intensity is best characterized as a sensitivity to the second derivative of image intensity. If one considers the $dG(x, y)$ filter used to measure dI however, then the sensitivity is best characterized as a sensitivity to the first derivative of image intensity. Thus for the algorithm to match human data, the human data will have to show sensitivity to *both* first and second derivatives.

Good measures of human spatial frequency sensitivity have been available for over fifteen years,

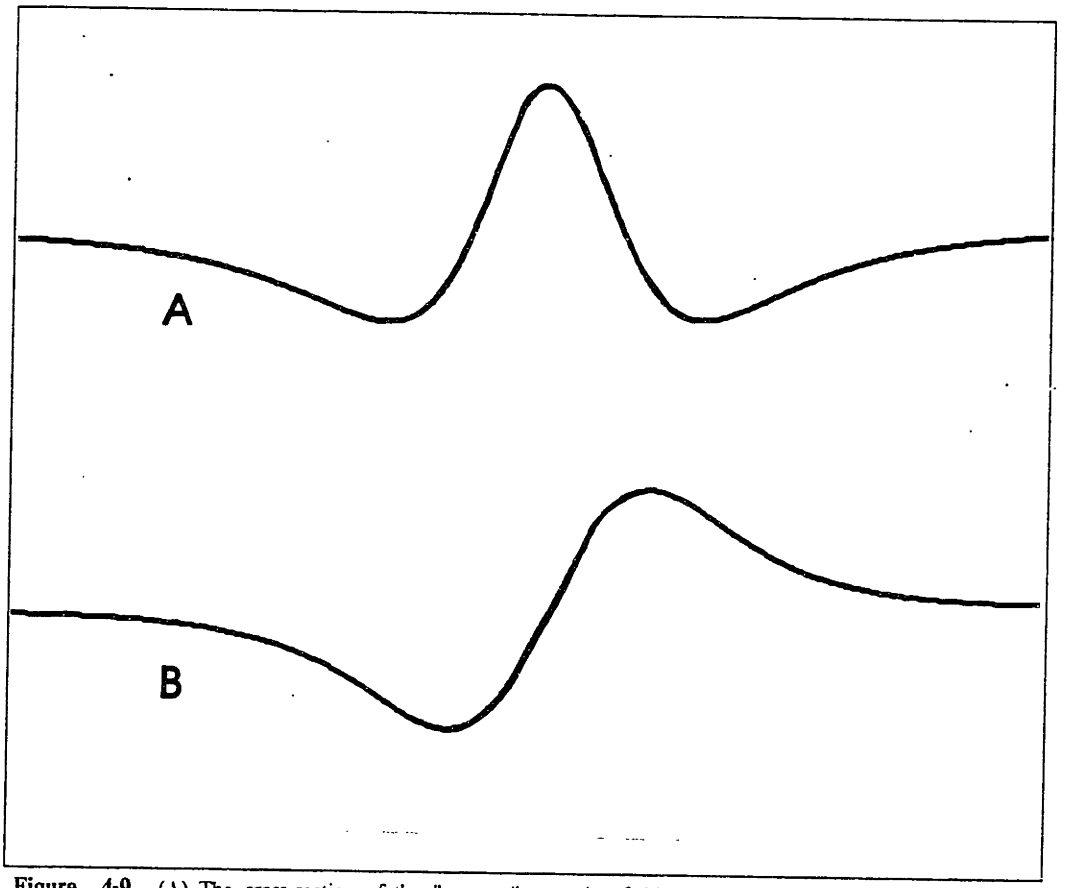


Figure 4-9. (A) The cross-section of the "average" receptive field, assuming detection is mediated by an isotropic detector. This receptive field structure implies that the human visual system may be considered to be sensitive to the second derivative of image intensity. (B) The cross-section of the "average" receptive field, assuming detection is mediated by a directional detector, as it must be at some stage of the visual system if the direction of motion is to be discerned. This receptive field structure implies that the human visual system may also be considered to be sensitive to the first derivative of image intensity. This is not contradictory; different stages of the visual system may use the input information differently, resulting in both first and second derivative sensitivity.

and these show that humans have an approximately bandpass sensitivity to the rate of change in image intensity for a fixed stimulus velocity (see Robson 1966 and Kelly 1966). Kelly 1979 recently reported completing the measurement of the entire spatio-temporal threshold surface, using counter-phase sine wave gratings which were stabilized on the retina so that both spatial and temporal variation could be carefully controlled. This threshold surface is shown in in figure 4-8.

Kelly points out that one may take the inverse Fourier transform of this sensitivity function to obtain the cross-section of the "average" receptive field, shown in figure 4-9 (A). The equation of this

inverse transform is

$$y = \frac{1}{2\pi}(1 + 6x^2)(1 + x^2)^{-3}$$

In performing this inverse transform one must make an assumption about how sensitivity varies with direction; a standard assumption is that detection sensitivity is the same for all orientations, which results in the symmetric receptive field shown. Kelly 1975 and later Marr and Hildreth 1980 point out that this receptive field structure is well modeled by $\nabla^2 G(x, y)$, the Laplacian of a two-dimensional Gaussian. This implies that people are sensitive to the second derivative of image intensity.

People, however, are also able to determine the direction of motion in such detection tasks, implying that the human sensitivity to changes in image intensity must be directional. As shown by Levinson and Sekuler 1975, the variation in sensitivity for moving sine wave gratings is exactly half that of counter-phase gratings. They present this result together with other evidence to show that counter-phase gratings such as used by Kelly are analyzed in the human visual system by independent motion-sensitive detectors.

Thus it appears that not only the moving sine wave gratings employed by Levinson and Sekuler but also the counter-phase gratings employed by Kelly are analyzed by motion-sensitive detectors. If we accept this, then in performing the inverse Fourier transform we must assume that the sensitivity is directional rather than isotropic, and as a result we obtain the asymmetric "average" receptive field cross-section shown in figure 4-9 (B) (Bracewell 1978). The equation of this inverse transform is

$$y = \frac{1}{2\pi}(12x(1 + x^2) - 6x - 36x^3)(1 + x^2)^{-4}$$

This receptive field structure is best modeled by a first derivative operator such as $dG(x, y)$, implying that peoples sensitivity is also proportional to the first derivative of image intensity.

It therefore seems that the human sensitivity to changes in image intensity may be reasonably characterized by *either* the first or second derivatives, as typified by the linear operators $\nabla^2 G(xy)$ and $dG(x, y)$, depending on how we conceptualize the detector. If we assume that detection is mediated by an isotropic detector, then the sensitivity is best characterized as second derivative; if however

the detector is directional, then the sensitivity is best characterized by the first derivative. Perhaps this should not be too surprising; different stages of the visual system may make different uses of the same input, resulting in functionally different sensitivities at the various stages. How the visual system might reasonably obtain both first and second derivative outputs from the same retinal organization is discussed in the next section.

3 Neurophysiology

Chapter 3 described an algorithm for performing the computations required by the theory on a digital computer. These computations may equally well be carried out by neurons in a biological visual system (see Ullman 1979). To address the question of whether there is evidence that they actually are carried out by a biological visual system, we will first have to propose a biological implementation of the algorithm, and then compare that implementation to the neurophysiological evidence.

3.1 Neurophysiology And Measurement Of Image Information

In chapter 3 the measurement of dI, d^2I and $\nabla^2 I$ was discussed. It was not specified exactly how these quantities were to be calculated, except that they were the results of convolutions with the image. Such convolutions are exactly analogous to receptive fields with sensitivity characteristics such as are shown in figure 3-1. It is inefficient and costly to have a separate "retinas" with different receptive fields, we would rather have a single retina with a limited number of types of receptive field which we could then combine to produce the desired *effective* receptive field. One such organization of receptive fields which implements the algorithm of chapter 3 is described here; it consists of three separate steps:

Step 1. The first step is to construct the circularly-symmetric "on-center" and "off-center" cells which are known to populate the retina, First discovered by Kuffler 1952, 1953, the response characteristics of these cells were described by Roderick and Stone 1965 a, b, and by Enroth-Cugell and Robson 1966 as the interaction of a smaller central excitatory receptive field with a larger surrounding inhibitory receptive field, resulting a receptive field structure typified by figure 4-10 (A). Working from anatomical and neurophysiological data (Cajal 1911, Dowling and Boycott 1966, Boycott and Dowling 1969) Richter and Ullman 1981 have shown that the characteristics of the X- and Y-type retinal ganglion cells may be closely modeled by subtracting the output of central photoreceptors

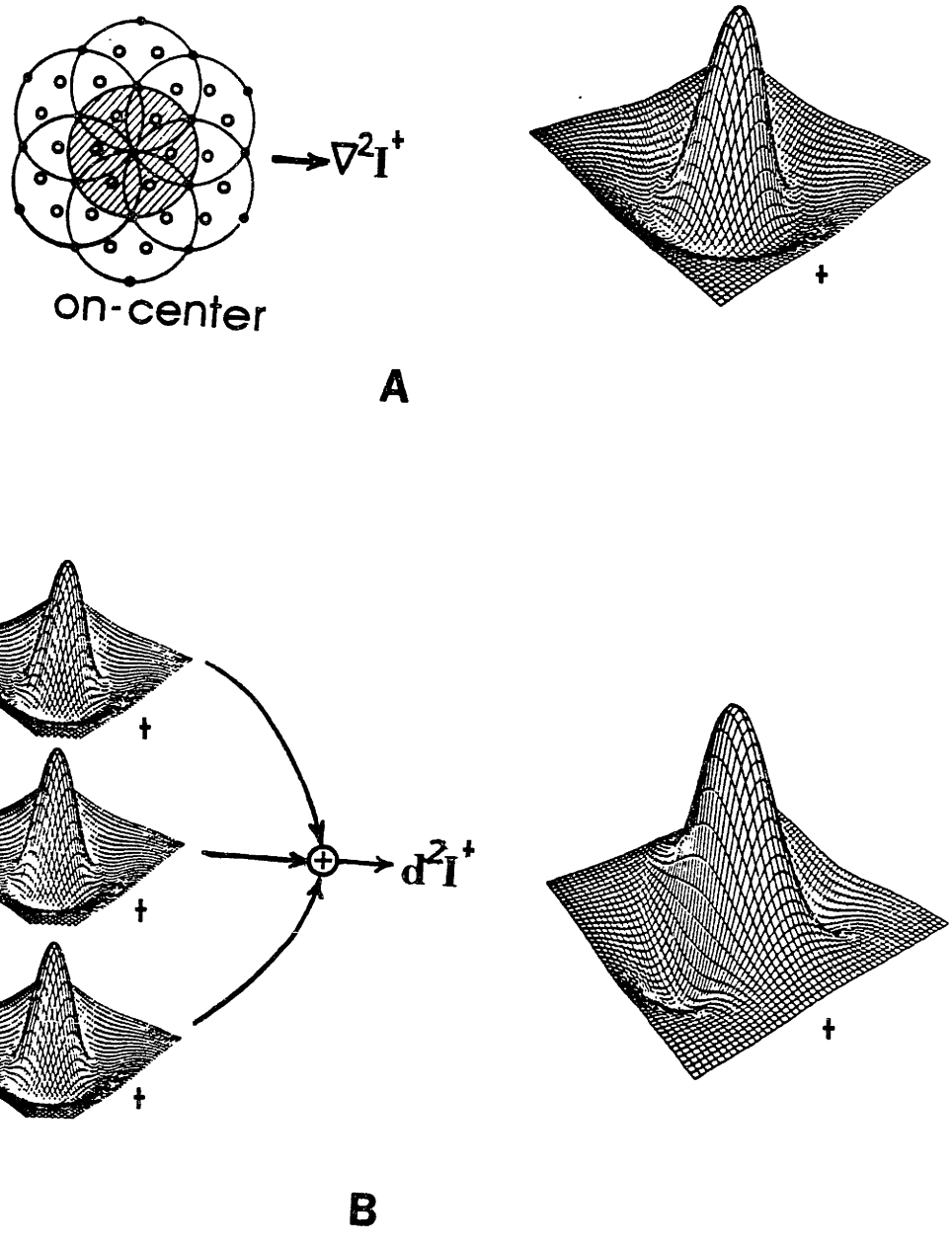


Figure 4-10. (A) Making circularly-symmetric on-center and off-center retinal ganglion neurons, whose response is proportional to $\nabla^2 I^+ = \max(\nabla^2 I, 0)$ and $\nabla^2 I^- = \max(-\nabla^2 I, 0)$ respectively, as a consequence of their receptive field shape. (B) Making oriented on-center and off-center cortical neurons by summing circularly-symmetric on-center and off-center neurons, their response is proportional (for a fixed stimulus velocity) to $d^2 I_\theta^+ = \max(d^2 I_\theta, 0)$ or $d^2 I_\theta^- = \max(-d^2 I_\theta, 0)$ depending upon whether on-center or off-center neurons were summed.

from the output of the surrounding photoreceptors.

Accepting these descriptions of the on-center and off-center cells, we find that these cells response is proportional to the positive and negative portions of $\nabla^2 I$, that is, the on-center respond proportional to $\nabla^2 I^+ = \max(\nabla^2 I, 0)$ and the off-center proportional to $\nabla^2 I^- = \max(-\nabla^2 I, 0)$. As the on- and off-center cells occur across the entire retina, this results in the outputs $\nabla^2 I^+(x, y)$ and $\nabla^2 I^-(x, y)$ at each point (x, y) in the image.

Step 2. The second step is to combine both on-center and off-center outputs into linear arrays, by summing the outputs of several of units which lie along a line oriented in a particular direction, resulting in the receptive field characteristics shown in figure 4-10 (B). Cortical neurons with such a receptive field were first described by Hubel and Wiesel 1962, who suggested that these receptive fields were the result of the summation of the outputs of circularly-symmetric on- and off-center cells along a particular orientation. The response of cortical neurons is, of course, conditional upon stimulus motion as well as stimulus shape, and so this receptive field shape may not be taken as a complete description of cortical neurons but only as a description of their response to a constant-velocity stimulus.

The particular receptive field organization shown here was suggested by Bishop *et al* 1973, Glezer *et al* 1973 and Shiller *et al* 1976a. Shiller *et al* 1976b suggest that all cortical receptive fields may be constructed of subfields of this type. Rose 1979, Daugman 1980 show that the observed cortical neuron receptive field shape can indeed be constructed by the linear summation of circularly-symmetric on- and off-center neurons.

The spatial frequency versus orientation sensitivity expected of such a receptive field is consistent with the actual sensitivity data as reported by De Valois *et al* 1977, 1979. This is important because it shows that the image information carried by the cell is completely described by its receptive field response characteristics, as shown in figure 4-10 (B). This is the expected result if the input to these neurons is the linear summation of circularly-symmetric on- and off-center neurons.

Accepting these descriptions of cortical receptive fields, we find that the response of the cortical

neurons to a constant-velocity stimulus is proportional to $d^2I_\theta^+ = \max(d^2I_\theta, 0)$ if the neuron is constructed from entirely on-center neurons, or $d^2I_\theta^- = \max(-d^2I_\theta, 0)$ if constructed from entirely off-center neurons, where θ is the image direction perpendicular to the direction along which the circularly-symmetric neurons were summed.

Step 3. The final step in obtaining the image information required is to combine the responses of the neurons of steps 1 and 2. If a cells response is proportional to the sum (or difference) of the responses of the neurons from steps 1 and 2, then the activity of that cell is a measure of the following types of image information:

The summed response of on- and off-center retinal ganglion cells is proportional to $|\nabla^2 I|$, while the difference between such cells response is proportional to $\nabla^2 I$. This is shown in figure 4-11 (A).

Similarly, the summed response of on- and off-center cortical neurons with orientation θ is $|d^2I_\theta|$, while the difference between the responses of such neurons is d^2I_θ. This is shown in figure 4-11 (B).

If on- and off-center cortical neurons of orientation θ are summed with an offset of two-thirds their width, then the result is $|dI_\theta|$, and the difference of such neurons is dI_θ. This is shown in figure 4-11 (C). Obtaining dI_θ in this manner is not rigorously exact, as with $\nabla^2 I$ and d^2I_θ, however it is a very close approximation as demonstrated by figure 4-11 (D) which compares the receptive field necessary to exactly obtain dI_θ to this approximation to dI_θ. As this figure shows, the approximation proposed here is very good; the difference averages less than 2 percent of the maximum value.

The receptive fields of the neurons constructed by adding and subtracting the basic oriented cortical neurons of step two correspond quite well to the cortical neuron receptive fields observed by Shiller *et al* 1976b. Their S_1 cell type corresponds to the basic cortical neuron receptive field constructed in step 2 (figure 4-10 (B)); their S_2 receptive field is the same as that shown in figure 4-11 (C), and their complex cell receptive field is the same as shown in figure 4-11 (B). All but one of their remaining receptive field types (S_7) are constructed of the S_1 and S_2 receptive field types with various combinations of directionality.

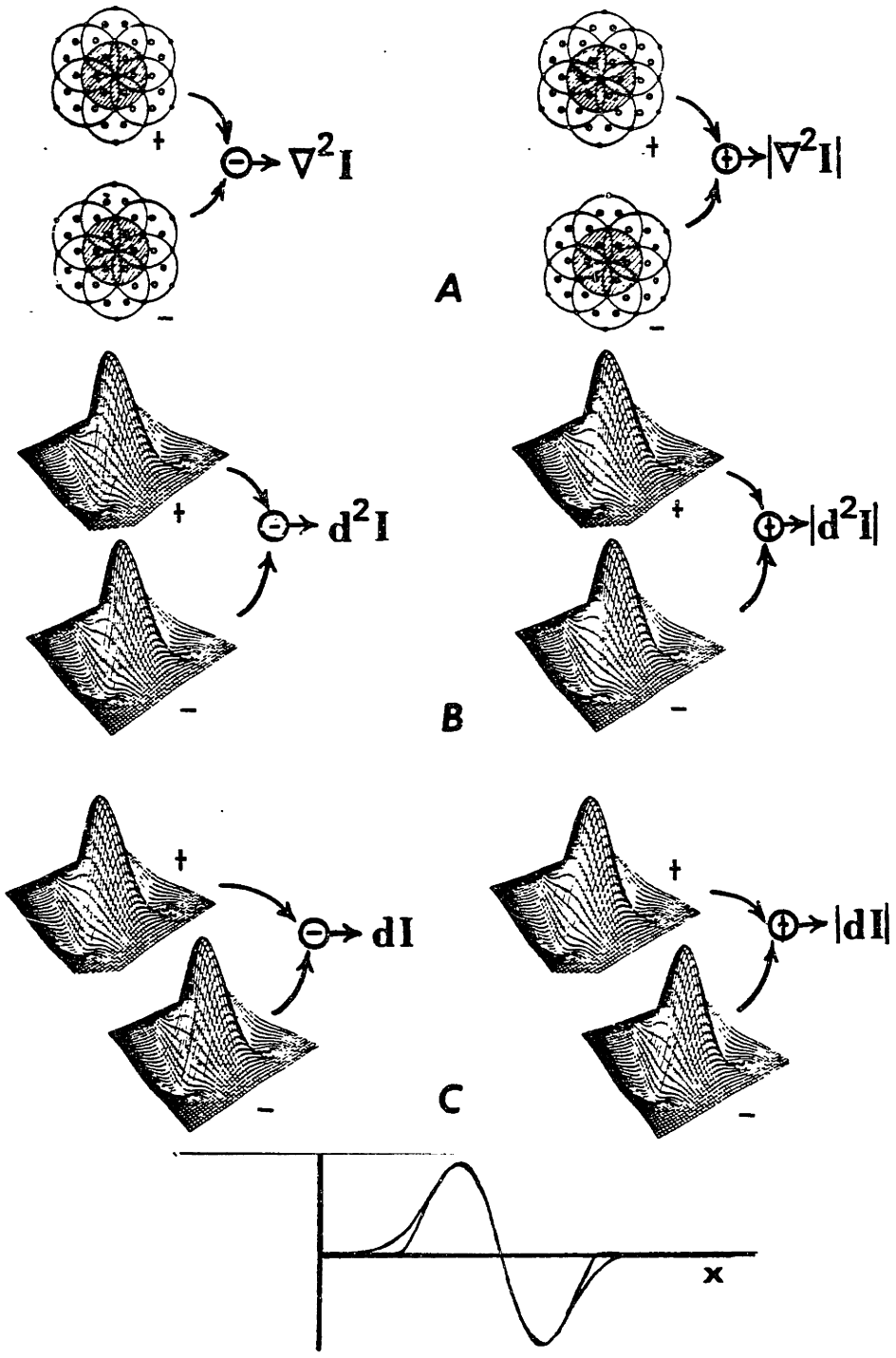


Figure 4-11. The response of a cell is completely determined by its receptive field; thus any cell with a receptive field such as pictured here will carries the image information shown in this figure. Thus we may use the circularly-symmetric on-center and off-center retinal neurons and the oriented on-center and off-center cortical neurons to (A) Compute $\nabla^2 I$ and $|\nabla^2 I|$, (B) Compute $d^2 I_\theta$ and $|d^2 I_\theta|$, (C) Compute dI_θ and $|dI_\theta|$ The graph in (D) is a comparison between the exact receptive field required to compute dI and the approximation used here. As can be seen, the approximation is a good one.

For both the X-type retinal ganglion cell described in step 1, and the simple cortical neuron described in step 2, the receptive field response characteristics of a cell seem to completely determine the responses of the cell[4] . The image information carried by such a cell is therefore purely a consequence of its receptive field response characteristics. Thus the responses of neurons driven by adding or subtracting the output of these neurons will be proportional to one of the types of image information assumed by the theory of this thesis, and thus a measure of the image information.

The assertion that the response of these cells is proportional to certain image information is a direct consequence of the organization of the constituent receptive fields, and the linear summing of these fields. We need assume nothing else. Thus any cell with the receptive fields described here will have a response which is a measure of the particular image information.

What remains uncertain is whether the remainder of the visual system is constructed so as to utilize this information, e.g., are complex cells responding to d^2I_θ or spatial frequency? They certainly have tuning curves for both properties; however the correct interpretation of the response of such a cell must depend upon how such responses are used by cells later in the system. Thus the answer to whether or not the information carried by a cell is used must wait until a significantly larger portion of the visual system is well understood.

3.2 Neurophysiology And The Estimation Algorithm

We have seen that the measurement of the required image information (specifically, dI_θ, d^2I_θ, and $\nabla^2 I$) is a plausible hypothesis for the early stages of visual processing. Is the estimation of surface shape as described in this thesis also plausible?

The computations required to estimate surface orientation from shading information are simple once given the input information. The tilt estimator simply requires comparing the output of several differently-oriented complex-type cortical neurons with aligned receptive fields; the orientation with the largest response is the direction of surface tilt. Similarly, determining the surface type only re-

[4]The response of a Y-type retinal ganglion cell is strongly dependent upon temporal aspects of the stimulus; see Richter and Ullman 1981 for a discussion of the implications this has for the information carried by such cells.

quires counting the number of orientations which have zero response within a group of complex cells with aligned receptive fields.

The estimator of slant proposed in this theory is slightly more complicated, with two operational components: the receptive field structure shown in figure 4-11 (A), constructed from adding on- and off-center retinal ganglion cell outputs, and division by the image intensity, I. We have already shown that the receptive field structure is plausible. Division by I is equivalent to Webers Law, and so it appears quite reasonable to posit such division by I occurring early in the system.

The simplicity of the required operations, the fact that the image information is available, and the fact that these operations *actually recover surface shape from the image* makes it seem likely that these estimation operations are being performed. Unfortunately, there is little direct neurophysiological evidence for or against the proposition that the visual system is performing these estimation operations. We must therefore wait for further neurophysiological evidence.

Summary. I believe that the relevance of this thesis to the study of neurophysiology is as follows:

(1) Any neuron in early in the visual system is performing a local analysis of the image, because its receptive field is limited.

(2) The theory presented here describes what kind of information about the viewed scene is *theoretically* available from such a limited receptive field.

(3) This theory is a *working* theory which actually recovers information about the shape of the world by analyzing the data available with a limited receptive field.

(4) This *working* theory has strong resemblances to well-known structure in the visual system; the computations are simple and therefore it is plausible that the desired image information concerning shape concerning shape can be extracted early in the visual system.

From these four points it seems that it is worth examining the hypothesis that neurons early in the visual system are carrying out the local analyses of the image suggested by this theory.

Appendix

1 Introductory Calculations

There are some special relationships which hold between d^2I and the properties of the surface, relationships which we will use in many of the following proofs. We may discover these relations by expanding the terms of the Hessian of image intensity in terms of θ_1, θ_2 and z_N (see the development of equation (10)):

$$\begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} I_{xx} & I_{yx} \\ I_{xy} & I_{yy} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} =$$
$$\rho\lambda \left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \right)^T \begin{pmatrix} -\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L}) & 0 \\ 0 & -\kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L}) \end{pmatrix} \left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \right)$$

We start by expanding the terms of the projection matrix Φ, letting $c_1 = \cos\theta_1$, $s_1 = \sin\theta_1$, $c_2 = \cos\theta_2$ and $s_2 = \sin\theta_2$:

$$\begin{aligned} \phi_{11} &= z_N^{-1} c_1 c_2 - s_1 s_2 \\ \phi_{12} &= z_N^{-1} c_1 s_2 + s_1 c_2 \\ \phi_{21} &= -z_N^{-1} s_1 c_2 - c_1 s_2 \\ \phi_{22} &= -z_N^{-1} s_1 s_2 + c_1 c_2 \end{aligned}$$

If we let now let $\zeta_{uu} = -\rho\lambda \mathbf{p}_{uu} \cdot \mathbf{L}$ and $\zeta_{vv} = -\rho\lambda \mathbf{p}_{vv} \cdot \mathbf{L}$ then we may obtain from the Hessian the following three expressions

$$\begin{aligned} I_{xx} &= \kappa_1 \zeta_{uu} \phi_{11}^2 + \kappa_2 \zeta_{vv} \phi_{21}^2 \\ I_{yy} &= \kappa_1 \zeta_{uu} \phi_{12}^2 + \kappa_2 \zeta_{vv} \phi_{22}^2 \\ I_{xy} &= \kappa_1 \zeta_{uu} \phi_{11} \phi_{12} + \kappa_2 \zeta_{vv} \phi_{21} \phi_{22} \end{aligned}$$

If we expand the various ϕ terms we obtain

$$\begin{aligned} \phi_{11}^2 &= z_N^{-2} c_1^2 c_2^2 - 2 z_N^{-1} c_1 c_2 s_1 s_2 + s_1^2 s_2^2 \\ \phi_{12}^2 &= z_N^{-2} c_1^2 s_2^2 + 2 z_N^{-1} c_1 c_2 s_1 s_2 + s_1^2 c_2^2 \\ \phi_{21}^2 &= z_N^{-2} s_1^2 c_2^2 + 2 z_N^{-1} c_1 c_2 s_1 s_2 + c_1^2 s_2^2 \\ \phi_{22}^2 &= z_N^{-2} s_1^2 s_2^2 - 2 z_N^{-1} c_1 c_2 s_1 s_2 + c_1^2 c_2^2 \\ \phi_{11}\phi_{12} &= z_N^{-2} c_1^2 s_2 c_2 - z_N^{-1} s_1 c_1 s_2^2 + z_N^{-1} s_1 c_1 c_2^2 - s_1^2 s_2 c_2 \\ \phi_{21}\phi_{22} &= z_N^{-2} s_1^2 s_2 c_2 + z_N^{-1} s_1 c_1 s_2^2 - z_N^{-1} s_1 c_1 c_2^2 - c_1^2 s_2 c_2 \end{aligned}$$

From these expressions we can obtain the following special relationships:

$$\nabla^2 I = I_{xx} + I_{yy} = \kappa_1 \zeta_{uu}(z_N^{-2} c_1^2 + s_1^2) + \kappa_2 \zeta_{vv}(z_N^{-2} s_1^2 + c_1^2)$$

At $c_2 = \pm 1$,

$$I_{xy} = (\kappa_1 \zeta_{uu} - \kappa_2 \zeta_{vv}) z_N^{-1} s_1 c_1$$

while at $c_2 = 0$

$$\begin{aligned} I_{xx} &= \kappa_1 \zeta_{uu} s_1^2 + \kappa_2 \zeta_{vv} c_1^2 \\ I_{yy} &= \kappa_1 \zeta_{uu} z_N^{-2} c_1^2 + \kappa_2 \zeta_{vv} z_N^{-2} s_1^2 \\ I_{xy} &= (\kappa_2 \zeta_{vv} z_N^{-1} - \kappa_1 \zeta_{uu} z_N^{-1}) s_1 c_1 \end{aligned}$$

and at $c_2 = s_2 = \pm\sqrt{2}/2$

$$I_{xy} = \frac{\kappa_1}{2} \zeta_{uu}(z_N^{-2} c_1^2 - s_1^2) + \frac{\kappa_2}{2} \zeta_{vv}(z_N^{-2} s_1^2 - c_1^2)$$

2 Proof Of Claims

Claim. Given that $\mathbf{p}_{uuu} = \mathbf{p}_{vvv} = \mathbf{p}_{uuv} = \mathbf{p}_{vuu} = 0$ then $d^2\mathbf{N}$ is parallel to $\mathbf{N}$, and perpendicular to $d\mathbf{N}$.

Proof

We can show that $d^2\mathbf{N}$ is perpendicular to $d\mathbf{N}$ and parallel to $\mathbf{N}$ as follows:

Continuing in our use of the Monge patch defined by $\mathbf{p} = u\mathbf{e}_1 + v\mathbf{e}_2 + f(u,v)\mathbf{e}_3$ to represent the surface, then

$$\mathbf{N} = \frac{-f_u\mathbf{e}_1 - f_v\mathbf{e}_2 + \mathbf{e}_3}{\sqrt{1+f_u^2+f_v^2}}$$

Note that this definition of the surface does not require that $\mathbf{e}_1$ or $\mathbf{e}_2$ be along a principal direction, but only that they lie in the tangent plane to the surface. Therefore any property of the derivatives of $\mathbf{p}$ with respect to u are also properties of the dirivatives along any direction of interest; we need only change the definitions of $\mathbf{e}_1$ and $\mathbf{e}_2$ so that $\mathbf{e}_1$ lies along the direction of interest to demonstrate this.

Letting $M = 1 + f_u^2 + f_v^2$, and M_u be the partial of M with respect to u then in the direction $\mathbf{e}_1$

$$\begin{aligned}
d\mathbf{N} =& (-f_{uu}M^{-1/2} + \frac{1}{2}f_uM^{-3/2}M_u)\mathbf{e}_1 \\
&+ (-f_{vu}M^{-1/2} + \frac{1}{2}f_vM^{-3/2}M_u)\mathbf{e}_2 \\
&- \frac{1}{2}M^{-3/2}M_u\mathbf{e}_3 \\
d^2\mathbf{N} =& (-f_{uuu}M^{-1/2} + f_{uu}M^{-3/2}M_u - \frac{3}{4}f_uM^{-5/2}M_u^2 + \frac{1}{2}f_uM^{-3/2}M_{uu})\mathbf{e}_1 \\
&+ (-f_{vuu}M^{-1/2} + f_{vu}M^{-3/2}M_u - \frac{3}{4}f_uM^{-5/2}M_u^2 + \frac{1}{2}f_uM^{-3/2}M_{uu})\mathbf{e}_2 \\
&- (-\frac{3}{4}M^{-5/2}M_u^2 + \frac{1}{2}M^{-3/2}M_{uu})\mathbf{e}_3
\end{aligned}$$

Because the surface is tangent to the plane of the co-ordinate system (u, v), the partials f_u and f_v are zero, $M = 1, M_u = 0$ and $M_{uu} = 2f_{uu} + 2f_{vu}$. Thus

$$d\mathbf{N} = -f_{uu}\mathbf{e}_1 + -f_{vu}\mathbf{e}_2 \qquad d^2\mathbf{N} = -f_{uuu}\mathbf{e}_1 - f_{vuu}\mathbf{e}_2 + (f_{uu} + f_{vu})\mathbf{e}_3$$

If the third derivatives of $\mathbf{p}$ are zero then $\mathbf{N}_{uu} \cdot \mathbf{N}_u = 0$, showing that $d^2\mathbf{N}$ is perpendicular to $d\mathbf{N}$, and $d^2\mathbf{N}$ points in the same direction as the surface normal •.

3 Proofs Of The Propositions

Proposition (Sufficiency). The second derivative of any pattern of image intensities can be accounted for by a second-order surface.

Proof

The second derivative of image intensity is specified completely by three numbers, I_{xx}, I_{yy} and I_{xy}. The images corresponding to second-order surfaces are described by the Hessian derived in developing equation (10):

$$\begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} I_{xx} & I_{yx} \\ I_{xy} & I_{yy} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} =$$
$$\rho\lambda \left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \right)^T \begin{pmatrix} -\kappa_1(\mathbf{p}_{uu} \cdot \mathbf{L}) & 0 \\ 0 & -\kappa_2(\mathbf{p}_{vv} \cdot \mathbf{L}) \end{pmatrix} \left(\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \right)$$

If any three numbers I_{xx}, I_{yy} and $I_{xy} = I_{yx}$ can be accounted for by this Hessian, then the proposition is proved. In fact, it suffices to show that any three numbers I_{xx}, I_{yy} and I_{xy} can be accounted for by this equation with the tilt aligned with the y axis, i.e., with $\theta_2 = \pi/2$.

From previous calculations, at $\theta_2 = \pi/2$ so that $c_2 = 0$

$$\begin{aligned} I_{xx} &= \kappa_1 \zeta_{uu} s_1^2 + \kappa_2 \zeta_{vv} c_1^2 \\ I_{yy} &= \kappa_1 \zeta_{uu} z_N^{-2} c_1^2 + \kappa_2 \zeta_{vv} z_N^{-2} s_1^2 \\ I_{xy} &= (\kappa_1 \zeta_{uu} z_N^{-1} - \kappa_2 \zeta_{vv} z_N^{-1}) s_1 c_1 \end{aligned}$$

where $\zeta_{uu} = -\rho\lambda \mathbf{p}_{uu} \cdot \mathbf{L}$ and $\zeta_{vv} = -\rho\lambda \mathbf{p}_{vv} \cdot \mathbf{L}$. We may assume that κ_1, κ_2, ζ_{uu} and ζ_{vv} may take on any values, and $1 \geq z_N^{-2}$, $-\pi \geq \theta_1 \geq \pi$.

Given any two values for I_{xx}, I_{yy}, we can use the above equations expressing these quantities in physical variables to solve for first for c_1^2 and s_1^2, and then for $\kappa_1 \zeta_{uu}$ I_{xy}. Doing this we obtain

$$c_1^2 = \frac{I_{xx} - \kappa_1 \zeta_{uu}}{\kappa_2 \zeta_{vv} - \kappa_1 \zeta_{uu}} \qquad s_1^2 = \frac{\kappa_2 \zeta_{vv} - I_{xx}}{\kappa_2 \zeta_{vv} - \kappa_1 \zeta_{uu}}$$

$$\kappa_1 \zeta_{uu} = \kappa_2 \zeta_{vv} - I_{xx} - I_{yy} z_N^2$$

We then are left with one equation in two unknown quantities, z_N and $\kappa_2\zeta_{vv}$:

$$I_{xy}^2 = (\kappa_2\zeta_{vv} - I_{xx})(z_N^2\kappa_2\zeta_{vv} - I_{yy})$$

Thus whatever the values of I_{xx} and I_{yy}, we may still pick values of z_N and $\kappa_2\zeta_{vv}$ such that the correct value of I_{xy} is obtained. •

Proposition (Generality Of A Point-Source Illuminant). Given that a surface has a Lambertian reflectance function, any constant distribution of illumination is equivalent to a distant point source.

Proof

Let us represent the distribution of illumination by the function $\mathbf{L}(\sigma, \tau)$, where $\mathbf{L}(\sigma, \tau)$ is a vector pointing in the direction (σ, τ) of magnitude equal to the amount of flux coming from that direction. Given a Lambertian reflectance function such that for a distant point-source

$$I = \rho\lambda \mathbf{N} \cdot \mathbf{L}$$

where $\rho\lambda$ is a constant, then the image intensity for an arbitrary distribution of illumination will be

$$\begin{aligned} I &= \int_{\sigma,\tau} \rho\lambda \mathbf{N} \cdot \mathbf{L}(\sigma, \tau) d\sigma d\tau \\ &= \rho\lambda x_N \int_{\sigma,\tau} x_L(\sigma, \tau) d\sigma d\tau + \rho\lambda y_N \int_{\sigma,\tau} y_L(\sigma, \tau) d\sigma d\tau + \rho\lambda z_N \int_{\sigma,\tau} z_L(\sigma, \tau) d\sigma d\tau \end{aligned}$$

where $\mathbf{N} = (x_N, y_N, z_N)$ and $\mathbf{L}(\sigma, \tau) = (x_L(\sigma, \tau), y_L(\sigma, \tau), z_L(\sigma, \tau))$. By the mean-value theorem, this is equivalent to the image intensity which would be caused by a Lambertian surface illuminated by a point-source somewhere in the hemisphere which is the range of (σ, τ), i.e.,

$$I = \rho\lambda^* \mathbf{N} \cdot \mathbf{L}^*$$

where λ^* is a scalar, and $\mathbf{L}^* = (x_{L^*}, y_{L^*}, z_{L^*})$ is a unit vector whose components are

$$\begin{aligned} x_{L^*} &= \int_{\sigma,\tau} x_L(\sigma, \tau) d\sigma d\tau \\ y_{L^*} &= \int_{\sigma,\tau} y_L(\sigma, \tau) d\sigma d\tau \\ z_{L^*} &= \int_{\sigma,\tau} z_L(\sigma, \tau) d\sigma d\tau \end{aligned}$$

This result applies throughout all areas of the surface which can "see" all of the light sources (i.e., as long as the surface is not self-occluding). For a Lambertian surface the equivalent point-source is at the weighted mean of the distribution of illumination. •

Proposition (Illuminant Direction). Under the assumption that change in surface normal is isotropically distributed, the following regression gives a maximum-likelihood estimate of τ_L, the tilt of the illuminant direction within a region:

$$\begin{pmatrix} \hat{x}_L \\ \hat{y}_L \end{pmatrix} = \left[\beta^T \beta\right]^{-1} \beta^T \begin{pmatrix} d\bar{I}_1 \\ d\bar{I}_2 \\ \vdots \\ d\bar{I}_n \end{pmatrix}$$

where $d\bar{I}_i$ are the means of dI over the region along each of n image directions (dx_i, dy_i), β is a $2 \times n$ matrix of directions (dx_i, dy_i) and β^T indicates the transpose of β. The illuminant tilt, τ_L is given by

$$\tau_L = \tan^{-1}\left(\frac{\hat{y}_L}{\hat{x}_L}\right)$$

and the full illuminant direction $\mathbf{L} = (x_L, y_L, z_L)$ is given by

$$x_L = \frac{\hat{x}_L}{k} \qquad y_L = \frac{\hat{y}_L}{k} \qquad z_L = \sqrt{1 - \frac{(\hat{x}_L^2 + \hat{y}_L^2)}{k^2}}$$

where

$$k = \sqrt{E(dI^2) - E(dI)^2}$$

Proof

If we measure the derivative of image intensity along a particular image direction (dx, dy), we are are observing

$$dI = \rho \lambda d\mathbf{N} \cdot \mathbf{L}$$

We will assume that the mean of $d\mathbf{N}$ is zero, because the expected value of $d\mathbf{N}$ is zero on average over all directions in all scenes, and the mean is zero for images of a large class of common objects. Thus the value of the z component of $d\mathbf{N}$, dz, measured along any one image direction, has a zero-mean distribution, and thus $E(dz)$, the expected value of dz, is also zero.

A maximum likelihood estimate of the illuminant tilt can be found by performing a least-squares regression of the mean of the distribution of dI along each image direction onto a model of the expected variation in the mean as a function of image direction. If we write $\mathbf{L} = (x_L, y_L, z_L)$ and at a point $d\mathbf{N} = (dx, dy, dz)$ then

$$dI = \rho\lambda d\mathbf{N} \cdot \mathbf{L} = \rho\lambda(dx x_L + dy y_L + dz z_L)$$

$E(dI)$, the expected value of dI along a particular image direction given n measurements of dI along that image direction, is

$$\begin{aligned} E(dI) &= \rho\lambda \frac{1}{n-1} \sum^{n} (dx dx_L + dy y_L + dz z_L) \\ &= \rho\lambda \frac{1}{n-1} \left(x_L \sum^{n} dx + y_L \sum^{n} dy + z_L \sum^{n} dz \right) \\ &= \rho\lambda (x_L E(dx) + y_L E(dy) + z_L E(dz)) \\ &= \rho\lambda (x_L d\bar{x} + y_L d\bar{y} + 0) \end{aligned} \tag{23}$$

where $d\bar{x}$ and $d\bar{y}$ are the mean values of dx and dy as observed along that image direction, and $E(dz) = 0$.

Under the assumption that $E(d\mathbf{N}) = 0$, then $\bar{\kappa} dx_I = d\bar{x}$ and $\bar{\kappa} dy_I = d\bar{y}$, where $\bar{\kappa}$ is the mean projected surface curvature and (dx_I, dy_I) the differential step in the image along which dI was measured. If we introduce a differential dr which may be thought of as the expected magnitude of $d\mathbf{N}$, i.e., as $E(|d\mathbf{N}|)$, then we may also write this as $x dr = d\bar{x}$ and $y dr = d\bar{y}$, where $x^2 + y^2 = 1$ and $x/y = dx_I/dy_I$. Using dr, equation (23) provides a model which we may use in a regression to recover the values $\hat{x}_L = \rho\lambda x_L dr$ and $\hat{y}_L = \rho\lambda y_L dr$, and thus the tilt of the illuminant, as it is equal to $\tan^{-1}(\hat{y}_L/\hat{x}_L)$. This regression is as follows.

Letting $d\bar{I}_i$ be the average of dI over the region along image direction (dx_i, dy_i), then our regres-

sion model is

$$\begin{pmatrix} d\bar{I}_1 \\ d\bar{I}_2 \\ \vdots \\ d\bar{I}_n \end{pmatrix} = \begin{pmatrix} dx_1 & dy_1 \\ dx_2 & dy_2 \\ \vdots & \vdots \\ dx_n & dy_n \end{pmatrix} \begin{pmatrix} \hat{x}_L \\ \hat{y}_L \end{pmatrix}$$

If we denote the matrix of directions (dx_i, dy_i) by β, and let β^T indicate the transpose of β, then the maximum likelihood estimate of $\hat{x}_L$ and $\hat{y}_L$ is the following least-squares regression

$$\begin{pmatrix} \hat{x}_L \\ \hat{y}_L \end{pmatrix} = \left[\beta^T \beta\right]^{-1} \beta^T \begin{pmatrix} d\bar{I}_1 \\ d\bar{I}_2 \\ \vdots \\ d\bar{I}_n \end{pmatrix}$$

The variance of our estimate of the illuminant direction can be calculated from this least-squares regression, giving us confidence intervals for the illuminant direction. It remains to obtain an estimate of $\rho\lambda dr$ in order to obtain x_L, y_L and $z_L = \sqrt{1 - x_L^2 - y_L^2}$ from $\hat{x}_L$ and $\hat{y}_L$.

If we again let $d\mathbf{N} = (dx, dy, dz)$ and $\mathbf{L} = (x_L, y_L, z_L)$ then $E(dI^2)$, the variance of dI along a particular image direction is

$$\begin{aligned} E(dI^2) &= E\Big(\rho^2\lambda^2(x_L dx + y_L dy + z_L dz)^2\Big) \\ &= E\Big(\rho^2\lambda^2(x_L^2 dx^2 + y_L^2 dy^2 + z_L^2 dz^2 + 2x_L y_L dx dy + 2x_L z_L dx dz + 2y_L z_L dy dz)\Big) \\ &= \rho^2\lambda^2 \frac{1}{n-1}\Big(x_L^2 \sum^n dx^2 + y_L^2 \sum^n dy^2 + z_L^2 \sum^n dz^2 \\ &\qquad + 2x_L y_L \sum^n dx dy + 2x_L z_L \sum^n dx dz + 2y_L z_L \sum^n dy dz\Big) \end{aligned} \tag{24}$$

and as dx and dy are both independent of dz, and $E(dz) = 0$, the final two cross-terms vanish.

We now require estimates of the variance of the dx, dy and dz components of $d\mathbf{N}$. Using a sphere as a model, we find that we may reasonably set variances $Var(dx) = Var(dy) = Var(dz) = dr^2$, so that by using the relations $x dr = d\bar{x}$ and $y dr = d\bar{y}$ we obtain $E(dx^2) = x^2 dr^2 + dr^2$,

$E(dy^2) = y^2 dr^2 + dr^2$, and $E(dz^2) = dr^2$. Thus from equation (24)

$$E(dI^2) = \rho^2\lambda^2\Big(x_L^2(x^2dr^2 + dr^2) + y_L^2(y^2dr^2 + dr^2) + z_L^2 dr^2 + 2x_L y_L xy dr^2\Big)$$
$$= \rho^2\lambda^2\Big((xx_L + yy_L)^2 dr^2 + (x_L^2 + y_L^2 + z_L^2)dr^2\Big)$$

As **L** is a unit vector, we then have

$$E(dI^2) = \rho^2\lambda^2\Big((xx_L + yy_L)^2 dr^2 + dr^2\Big)$$

We then note that

$$E(dI)^2 = (\rho\lambda(d\bar{x}x_L + d\bar{y}y_L))^2 = (\rho\lambda(xx_L + yy_L)dr)^2 = \rho^2\lambda^2(xx_L + yy_L)^2 dr^2$$

and thus for any particular image direction we have

$$E(dI^2) - E(dI)^2 = \rho^2 f^2 dr^2$$

We may now estimate x_L, y_L and z_L

$$x_L = \frac{\hat{x}_L}{k} \qquad y_L = \frac{\hat{y}_L}{k} \qquad z_L = \sqrt{1 - \frac{(\hat{x}_L^2 + \hat{y}_L^2)}{k^2}}$$

where

$$k = \rho\lambda dr = \sqrt{E(dI^2) - E(dI)^2}$$

This completes the proof of this proposition. •

Proposition (Surface Convexity). The maximum likelihood estimate of surface convexity in a direction (dx, dy), i.e., whether the surface is becoming closer (concave) or farther (convex) from the viewer as one moves along that direction, assuming that κ_1, κ_2 are identically distributed and that surface orientation is uniformly distributed, is

$$\begin{aligned} &convex && sgn(x_L dx + y_L dy) = sgn(dI) \\ &concave && sgn(x_L dx + y_L dy) \neq sgn(dI) \end{aligned}$$

where $\mathbf{L} = (x_L, y_L, z_L)$ and dI is measured in the direction (dx, dy)

Proof

Consider the expression for dI, equation (7):

$$dI = \rho\lambda\begin{pmatrix} -\kappa_1(\mathbf{p}_u \cdot \mathbf{L}) & -\kappa_2(\mathbf{p}_v \cdot \mathbf{L}) \end{pmatrix}\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix} \tag{7}$$

where

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix}\begin{pmatrix} z_N^{-1} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ -\sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

and where

θ_1 = angle between the projection of the surface tilt and the direction of the first principal curvature

z_N = the z component of the surface normal $\mathbf{N} = (x_N, y_N, z_N)$

θ_2 = the tilt of the surface

If we assume that θ_1, z_N and θ_2 are uniformly distributed, then the distribution of surface orientation and the distribution of orientation of the principal directions is isotropic from the standpoint of the viewer and thus the mean surface orientation will be directly toward the veiwer. Thus Φ, the projection matrix, will not affect the mean value of dI.

To show this we note that if θ_1 is taken to be uniformly distributed, then the first rotation matrix in Φ will on average be the identity matrix as is as likely that there be an occurance of a surface with θ_1 as an occurance of a surface with $-\theta_1$. Similarly, the probability of an occurance of a surface with a particular z_N and θ_2 is balanced by the probability of a surface with z_N and $\theta_2 + \pi/2$. Thus the mean surface orientation is directly toward the viewer, and Φ has no effect on the mean of dI.

This means that it suffices to analyze the simpler equation (5), which gives dI in the surface co-ordinate system:

$$dI = \rho\lambda(-\kappa_1(\mathbf{p}_u \cdot \mathbf{L})du - \kappa_2(\mathbf{p}_v \cdot \mathbf{L})dv) \tag{5}$$

From (5) we may obtain

$$I_u = -\rho\lambda\kappa_1\mathbf{p}_u \cdot \mathbf{L} \qquad I_v = -\rho\lambda\kappa_2\mathbf{p}_v \cdot \mathbf{L} \tag{25}$$

But at the point P in a co-ordinate system tangent to the surface, $\mathbf{p}_u = \mathbf{e}_1$ and $\mathbf{p}_v = \mathbf{e}_2$, thus $\mathbf{p}_u \cdot \mathbf{L} = x_L$ and $\mathbf{p}_v \cdot \mathbf{L} = y_L$, where $\mathbf{L} = (x_L, y_L, z_L)$. As κ_1 and κ_2 are assumed to be uniformily distributed,

$$E(|\kappa_1|) = E(|\kappa_2|) = \bar{\kappa} > 0$$

where E is the expected value operator, and $\bar{\kappa}$ is the expected magnitude of κ_1 and κ_2. Thus we may reduce equation (25) to

$$E(I_u) = sgn(-\kappa_1)\rho\lambda\bar{\kappa}x_L \qquad E(I_v) = sgn(-\kappa_2)\rho\lambda\bar{\kappa}y_L$$

Because $\rho\lambda\bar{\kappa} > 0$, then on average

$$sgn(-\kappa_1) = sgn(\frac{I_u}{x_L}) \qquad sgn(-\kappa_2) = sgn(\frac{I_v}{y_L})$$

Thus on average if dI along some direction is of the same sign as the projection of the illuminant vector onto that direction, then the surface is convex; if they are of opposite sign then the surface is concave. Thus on average

$$\begin{array}{ll} convex & sgn(x_L dx + y_L dy) = sgn(dI) \\ concave & sgn(x_L dx + y_L dy) \neq sgn(dI) \end{array}$$

where dI is measured along direction (dx, dy). •

Proposition (Tilt of the Surface). Given the smooth, homogeneous surface assumed by equation (10), and $I_{xx} \neq I_{yy}$, $I_{xy} \neq 0$, then the tilt of the surface is the image direction in which the second derivative of image intensity, d^2I, is greatest.

Proof

As in the introductory calculations, we will let $\zeta_{uu} = -\rho\lambda \mathbf{p}_{uu} \cdot \mathbf{L}$ and $\zeta_{vv} = -\rho\lambda p_{vv} \cdot \mathbf{L}$, and will adopt the notation $c_1 = \cos\theta_1$, $s_1 = \sin\theta_1$, $c_2 = \cos\theta_2$, $s_2 = \sin\theta_2$. We then recall these results proven in the introductory calculations:

At $c_2 = \pm 1$,

$$\begin{aligned} 2I_{xy} &= 2\kappa_1\zeta_{uu}\phi_{11}\phi_{12} + 2\kappa_2\zeta_{vv}\phi_{21}\phi_{22} \\ &= 2(\kappa_1\zeta_{uu} - \kappa_2\zeta_{vv})z_N^{-1}s_1c_1 \end{aligned}$$

while at $c_2 = s_2 = \pm\frac{\sqrt{2}}{2}$

$$\begin{aligned} I_{yy} - I_{xx} &= \kappa_1\zeta_{uu}\phi_{11}^2 + \kappa_2\zeta_{vv}\phi_{21}^2 \\ &\quad - \kappa_1\zeta_{uu}\phi_{12}^2 - \kappa_2\zeta_{vv}\phi_{22}^2 \\ &= 2(\kappa_1\zeta_{uu} - \kappa_2\zeta_{vv})z_N^{-1}s_1c_1 \end{aligned}$$

Thus $2I_{xy}$ at $c_2 = \pm 1$ is equal to $I_{yy} - I_{xx}$ at $c_2 = s_2 = \pm\frac{\sqrt{2}}{2}$.

Given I_{xx}, I_{yy} and I_{xy} we may solve for the angle θ_2, which is the tilt of the surface, as follows.

We know that given I_{xx}, I_{yy} and I_{xy} we may obtain these quantitites in any other image plane coordinate system (x^*, y^*) which is a rotation of (x, y) by the angle τ. First we note that

$$\begin{aligned} I_{x^*} &= I_x\frac{dx}{dx^*} + I_y\frac{dy}{dx^*} \\ I_{y^*} &= I_x\frac{dx}{dy^*} + I_y\frac{dy}{dy^*} \end{aligned}$$

The standard rotation transformation is

$$\begin{aligned} x^* &= xc_\tau + ys_\tau \\ y^* &= -xs_\tau + yc_\tau \end{aligned}$$

where s_τ and c_τ are the sine and cosine of the angle τ, and the inverse of this rotation transformation is

$$\begin{aligned} x &= x^*c_\tau - y^*s_\tau \\ y &= x^*s_\tau + y^*c_\tau \end{aligned}$$

Thus

$$\frac{dx}{dx^*} = c_\tau \qquad \frac{dy}{dx^*} = s_\tau \qquad \frac{dx}{dy^*} = -s_\tau \qquad \frac{dy}{dy^*} = c_\tau$$

and so

$$I_x^* = I_x c_\tau + I_y s_\tau$$
$$I_y^* = -I_x s_\tau + I_y s_\tau$$

Similarly,

$$I_{x^*x^*} = \frac{d(I_{x^*})}{dx}\frac{dx}{dx^*} + \frac{d(I_{x^*})}{dy}\frac{dy}{dx^*}$$
$$I_{y^*y^*} = \frac{d(I_{y^*})}{dx}\frac{dx}{dy^*} + \frac{d(I_{y^*})}{dy}\frac{dy}{dy^*}$$
$$I_{x^*y^*} = \frac{d(I_{x^*})}{dx}\frac{dx}{dy^*} + \frac{d(I_{x^*})}{dy}\frac{dy}{dy^*}$$

resulting in:

$$I_{x^*x^*} = I_{xx}c_\tau^2 + I_{yy}s_\tau^2 + 2I_{xy}s_\tau c_\tau$$
$$I_{y^*y^*} = I_{xx}s_\tau^2 + I_{yy}c_\tau^2 - 2I_{xy}s_\tau c_\tau$$
$$I_{x^*y^*} = -I_{xx}s_\tau c_\tau + I_{yy}s_\tau c_\tau + I_{xy}(c_\tau^2 - s_\tau^2)$$

Thus to obtain θ_2, the tilt of the surface, we need to find τ such that the quantity $I_{y^*y^*} - I_{x^*x^*}$, at a rotation of the co-ordinates by $\tau + \pi/4$, is equal to the quantity $2I_{x^{**}y^{**}}$, at a rotation of the co-ordinates of τ. That is, we find τ such that

$$\begin{aligned} I_{y^*y^*} - I_{x^*x^*} &= (I_{xx}s^2_{(\tau+\pi/4)} + I_{yy}c^2_{(\tau+\pi/4)} - 2I_{xy}s_{(\tau+\pi/4)}c_{(\tau+\pi/4)}) \\ &\quad - (I_{xx}c^2_{(\tau+\pi/4)} + I_{yy}s^2_{(\tau+\pi/4)} + 2I_{xy}s_{(\tau+\pi/4)}c_{(\tau+\pi/4)}) \\ &= -2I_{xx}s_\tau c_\tau + 2I_{yy}s_\tau c_\tau + 2I_{xy}(c_\tau^2 - s_\tau^2) = I_{x^{**}y^{**}} \end{aligned}$$

Using the identities $c_\tau^2 - s_\tau^2 = c_{2\tau}$ and $2s_\tau c_\tau = s_{2\tau}$, we may reduce this equation to obtain

$$(I_{yy} - I_{xx})c_{(2\tau+\pi/2)} - 2I_{xy}s_{(2\tau+\pi/2)} = (I_{yy} - I_{xx})s_{2\tau} + 2I_{xy}c_{2\tau}$$

then noting that $c_{2\tau+\pi/2} = -s_{2\tau}$ and $s_{2\tau+\pi/2} = c_{2\tau}$ we obtain

$$-(I_{yy} - I_{xx})s_{2\tau} - 2I_{xy}c_{2\tau} = (I_{yy} - I_{xx})s_{2\tau} + 2I_{xy}c_{2\tau}$$

and thus

$$-2(I_{yy} - I_{xx})s_{2\tau} - 4I_{xy}c_{2\tau} = 0$$

Solving for τ we find

$$\tau = \frac{1}{2}\tan^{-1}\left(\frac{2I_{xy}}{I_{xx} - I_{yy}}\right)$$

The angle τ is then the direction of the tilt of the surface. Note that there is an ambiguity of $\pm\pi$ in this solution; this reflects the inherent convexity/concavity ambiguity of the imaged surface. Note also that when $I_{xy} = 0$ and $I_{xx} = I_{yy}$ that both this solution and the tilt are undefined.

This angle is also the direction of maximum d^2I. In order to show this, we find the angle τ for which $I_{x'x'}$ attains a maximum over all rotations of the image plane co-ordinate system. $I_{x'x'}$ is equal to

$$I_{x'x'} = I_{xx}c_\tau^2 + I_{yy}s_\tau^2 + 2I_{xy}s_\tau c_\tau$$

The maximum of $I_{x'y'}$ occurs at

$$\begin{aligned} 0 &= (I_{yy} - I_{xx})2s_\tau c_\tau + 2I_{xy}(c_\tau^2 - s_\tau^2) \\ &= (I_{yy} - I_{xx})s_{2\tau} + 2I_{xy}c_{2\tau} \end{aligned}$$

so that the angle τ for which d^2I attains its maximum is also

$$\tau = \frac{1}{2}\tan^{-1}\left(\frac{2I_{xy}}{I_{xx} - I_{yy}}\right)$$

Thus the direction τ in which d^2I attains its maximum, is also the direction of the tilt of the surface. •

Proposition (Normalized Laplacian). Given the smooth, homogeneous surface assumed by equation (10) then

$$\frac{\nabla^2 I}{I} = -\kappa_n^2 z_N^{-2} - \kappa_m^2$$

where κ_n is the surface curvature along the surface tilt direction and κ_m is the surface curvature in the orthogonal direction, z_N is the z component of the surface normal, equal to the cosine of the slant of the surface.

Proof

In chapter 2 section 4 it was shown that in the surface co-ordinate system

$$\begin{aligned}\frac{d^2 I}{I} &= \frac{\rho\lambda d^2\mathbf{N}\cdot\mathbf{L}}{\rho\lambda\mathbf{N}\cdot\mathbf{L}} \\ &= \frac{-\kappa_1(\mathbf{p}_{uu}\cdot\mathbf{L})du^2 - \kappa_2(\mathbf{p}_{vv}\cdot\mathbf{L})dv^2}{\mathbf{N}\cdot\mathbf{L}} \\ &= \frac{-\kappa_1^2(\mathbf{N}\cdot\mathbf{L})du^2 - \kappa_2^2(\mathbf{N}\cdot\mathbf{L})dv^2}{\mathbf{N}\cdot\mathbf{L}} \\ &= -\kappa_1^2 du^2 - \kappa_2^2 dv^2\end{aligned}$$

We may repeat these operations using the general expression for $d^2 I$, equation (10):

$$d^2 I = \rho\lambda\left(\begin{pmatrix}-\kappa_1(\mathbf{p}_{uu}\cdot\mathbf{L}) & 0\\ 0 & -\kappa_2(\mathbf{p}_{vv}\cdot\mathbf{L})\end{pmatrix}\left(\begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix}\begin{pmatrix}dx\\ dy\end{pmatrix}\right)\right)^T\left(\begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix}\begin{pmatrix}dx\\ dy\end{pmatrix}\right) \tag{10}$$

to obtain

$$\frac{d^2 I}{I} = -\left(\begin{pmatrix}-\kappa_1^2 & 0\\ 0 & -\kappa_2^2\end{pmatrix}\left(\begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix}\begin{pmatrix}dx\\ dy\end{pmatrix}\right)\right)^T\left(\begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix}\begin{pmatrix}dx\\ dy\end{pmatrix}\right) \tag{26}$$

In the first section of this appendix we calculated that

$$\nabla^2 I = I_{xx} + I_{yy} = \kappa_1\varsigma_{uu}(z_N^{-2}c_1^2 + s_1^2) + \kappa_2\varsigma_{vv}(z_N^{-2}s_1^2 + c_1^2)$$

where $\varsigma_{uu} = -\rho\lambda\mathbf{p}_{uu}\cdot\mathbf{L}$ and $\varsigma_{vv} = -\rho\lambda\mathbf{p}_{vv}\cdot\mathbf{L}$. Thus using equation (26) we find that

$$\frac{\nabla^2 I}{I} = -\kappa_1^2(z_N^{-2}c_1^2 + s_1^2) - \kappa_2^2(z_N^{-2}s_1^2 + c_1^2)$$

Using Eulers theorm,

$$\kappa_n = \kappa_1 \cos^2 \theta_n + \kappa_2 \sin^2 \theta_n$$

we find that for κ_n along the tilt direction, $\theta_n = \theta_1$ and thus for κ_n along the tilt direction and κ_m in the perpendicular direction, that

$$\frac{\nabla^2 I}{I} = -\kappa_n^2 z_N^{-2} - \kappa_m^2$$

which completes the proof. •

Proposition (Estimation Of Slant). Assuming a uniform distribution of surface curvature, and the smooth, homogeneous surface assumed by equation (10), then the maximum-likelihood estimate of z_N, where z_N is equal to the arccosine of the slant of the surface, is

$$z_N = \sigma_\kappa^2 \left(\left| \frac{\nabla^2 I}{I} \right| - \sigma_\kappa^2 \right)^{-2}$$

where σ_κ^2 is the variance of the distribution of surface curvatures.

Proof

From the previous proposition,

$$\left| \frac{\nabla^2 I}{I} \right| = \| -\kappa_n^2 \bar{z_N}^2 - \kappa_m^2 \|$$

Assuming that κ_1 and κ_2 are uniformly distributed, then

$$E(\kappa_1^2) = E(\kappa_2^2) = \sigma_\kappa^2$$

where σ_κ^2 is the variance of the distribution of κ_1 and κ_2. Thus for a fixed z_N,

$$\begin{aligned} E\left(\left| \frac{\nabla^2 I}{I} \right| \right) &= E(\| -\kappa_n^2 \bar{z_N}^2 - \kappa_m^2 \|) \\ &= \sigma_\kappa^2 (\bar{z_N}^2 + 1) \end{aligned}$$

Thus

$$z_N = \sigma_\kappa^2 \left(\left| \frac{\nabla^2 I}{I} \right| - \sigma_\kappa^2 \right)^{-2}$$

This completes the proof of the proposition. •

4 Positive Images, EM Images And Photographic Negatives

Figure 4-3 of chapter 4 shows how the slant estimator $|\nabla^2 I/I|$ functions correctly on positive images, approximately correctly on Electron Microscope (EM) images, and fails on photographic negatives. This section of the appendix goes through the examples of figure 4-3 more explicitly, so that the reader may fully understand why the estimator works on positive images and EM images, but not on photographic negatives.

Figures 4-3 (A), (B) and (C) show the image intensity profiles across the middle of a sphere in, respectively, a positive image, EM image and photographic image. In (A) a Lambertian surface with unit albedo and incident illumination is assumed, with the illuminant at the viewers position, so that $I = \mathbf{N} \cdot \mathbf{V}$. In (B) the image intensity is given by $I = (\mathbf{N} \cdot \mathbf{V})^{-1}$, which is the form of the reflectance function for EM images. In (C), the same imaging geometry was assumed as in (A), except that the brightest point in the original, positive image was taken to have an intensity of 1.05, so that the image intensity is given by $I = 1.05 - \mathbf{N} \cdot \mathbf{L}$.

Figures 4-3 (D), (E) and (F) show the result of applying the slant estimator, $|\nabla^2 I/I|$, to the profiles in (A), (B) and (C). Note that as the theory requires that the range of $|\nabla^2 I/I|$ be scaled so that z_N lies between zero and one, the values of $|\nabla^2 I/I|$ shown in (D), (E) and (F) have been scaled to the same range, 0.0 — 1.0. In chapter 2 section 4 it was shown that for normal images $|\nabla^2 I/I|$ is proportional to the foreshortened curvature, i.e.,

$$\frac{d^2 I}{I} = \frac{d^2 \mathbf{N} \cdot \mathbf{L}}{\mathbf{N} \cdot \mathbf{L}} = \|d^2 \mathbf{N}\|$$

so that we expect to see a smooth cup-like shape as the result of calculating $|\nabla^2 I/I|$ for the profile of (A); this is shown in figure 4-3 (D).

For the EM image $I = (\mathbf{N} \cdot \mathbf{V})^{-1}$ (figure 4-3 (B)), we have that

$$\frac{d^2 I}{I} = \frac{-(\mathbf{N} \cdot \mathbf{V})^{-2}(d^2 \mathbf{N} \cdot \mathbf{V}) + 2(\mathbf{N} \cdot \mathbf{V})^{-3}(d\mathbf{N} \cdot \mathbf{V})^2}{(\mathbf{N} \cdot \mathbf{V})^{-1}} = (\mathbf{N} \cdot \mathbf{V})(d^2 \mathbf{N} \cdot \mathbf{V}) + 2(\mathbf{N} \cdot \mathbf{V})^{-2}(d\mathbf{N} \cdot \mathbf{V})^2$$

It just so happens that for regions of small slant, the values of this function are small, and as the slant becomes larger the values increase approximately proportional to z_N^{-2} (just as for normal images), as shown by figure 4-3 (E). It is worth emphisizing that there is no particular *reason* why this should be so; it is simply a result of the rather peculiar process of image generation used by the scanning electron microscope.

For the photographic negative $I = 1.05 - \mathbf{N} \cdot \mathbf{L}$ (figure 4-3 (C)), we have

$$\frac{d^2I}{I} = \frac{-d^2\mathbf{N} \cdot \mathbf{L}}{1.05 - \mathbf{N} \cdot \mathbf{L}}$$

Because of the constant 1.05, no further reduction of this function is possible. For **N** near **L**, the small values of I result in a blow-up of the quantity $|d^2I/I|$, as shown by figure 4-3 (F). In this example the blow-up occurs for $\mathbf{N} = \mathbf{V}$, producing a symmetric result, however with any other illuminant direction the result would not be symmetric. Further, the amount of blow-up is dependant upon what other objects are in the scene, and if the brightest object in the original scene is very much brighter than the object under consideration the result will be uniformily small estimates of slant regardless of the actual surface slant. Thus in a photographic negative of a complex scene the slant estimator will give a different pattern of error for each object.

If we propose a slightly more complex scheme we can obtain results which are, perhaps, more in line with informally reported subjective experience. Let us first normalize image intensities by setting the darkest point in the image to the value one and then apply the shape estimators. In normal images this procedure achieves correction for the deleterious effects of large amounts of diffuse illumination, which can sometimes cause the effective point-source illuminant to shift rapidly. In EM images this procedure has no effect, as for all objects in an EM image the darkest point already has a value of one. In photographic negatives, however, this procedure results in the correct estimation of shape only in those scenes with one albedo and one illuminant. In scenes with a wide range of albedos or considerable diffuse illumination the surface relief will be underestimated for all but the brightest objects. This normalization scheme produces results which correspond more closely with informally

reported subjective experience; however additional experimentation is needed before this slightly more complex scheme can be seriously advanced.

REFERENCES

Binford, T. and Horn, B. K. P., "The Binford-Horn LINEFINDER," *A.I. Vision Flash 16* (1971).

Barrow, H.G. and Tenenbaum, J.M., "Interpreting Line Drawings As Three-Dimensional Surfaces," *Artificial Intelligence* 15, Special Issue on Computer Vision (1981). 75-116.

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

Bishop, P.O., Coombs, J.S., and Henery, G.H., "Receptive Fields Of Simple Cells In Cat Striate Cortex," *J. Physiol.* 231 (1973), 31-60.

Boycott, B.B. and Dowling, J.E., "Organization of the primate retina: light microscopy," *Phil. Trans. Roy. Soc. B* 255 (1969), 109-184.

Bracewell, R.N., *The Fourier Transform And Its Applications*, McGraw-Hill Book Company, New York, New York, 1978.

Brady, M., "personal communication," (1981).

Bruss, A., "Shape From Shading And Bounding Contour," *Ph.D. Thesis, Dept. Elec. Engr.& Comp. Sci., MIT* (1981).

Cajal, S. R., "Histologie du systeme nerveux de l'homme et des vertebrates," 2, Paris, Maloine (1911).

Daugman, J.G., "Two-Dimensional Spectral Analysis Of Cortical Receptive Field Profiles," *Vis. Res.* 20 (1980), 847-856.

Davis, L., "A Survey of Edge Detection Techniques," *Computer Graphics and Image Processing* 4

(1975), 248-270.

DeValois, R.L., Albrect D.G., and Thoull, L.G., "Spatial Tuning of LGN and Cortical Cells In The Monkey Visual System," *Spatial Contrast* , Ed. Spekreijse, H. and van der Tweel, L.H., North-Holland, Amsterdam (1977).

DeValois, K., DeValois R.L., and Yund, E.W., "Responses Of Striate Cortex Cells To Grating And Checkerboard Patterns," *J. Physiol.* **291** (1979), 483-505.

Dowling, J.E. and Boycott, B.B., "Organization Of Primate Retina: Electron Microscopy," *Proc. R. Soc. Lond. B* **166** (1966), 80-111.

Duda, R.O., and Hart, P.E., *Pattern Recognition and Scene Anaylsis*, John Wiley and Sons, New York, New York, 1973.

Enroth-Cugell and Robson, J.G., "The Contrast Sensitivity Of Retinal Ganglion Cells Of The Cat," *J. Physiol.* **187** (1966), 517-552.

Glezer, V.D., Ivanov, V.A., and Tscherback, T.A., "Investigation Of Complex And Hypercomplex Receptive Fields Of Visual Cortex Of The Cat As Spatial Frequency Filters," *Vis. Res.* **13** (1973), 1875-1904.

Herskovitz, A. and Binford, T.O., "On Boundary Detection," , M.I.T. AI Lab Memo 183 (1970).

Horn, B.K.P., "Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from one View," *A.I. Technical Report 79* , Project MAC, M.I.T. (1970).

Horn, B.K.P.H, "Linear Analysis And Image Processing," *M.I.T. A.I. Working Paper 100* (1974).

Horn, B.K.P., "Obtaining Shape from Shading Information", *The Psychology of Computer Vision, P. H. Winston (ed.)*, McGraw Hill Book Company Inc., 1975, 115–155.

(1975), 248-270.

DeValois, R.L., Albrect D.G., and Thoull, L.G., "Spatial Tuning of LGN and Cortical Cells In The Monkey Visual System," *Spatial Contrast* , Ed. Spekreijse, H. and van der Tweel, L.H., North-Holland, Amsterdam (1977).

DeValois, K., DeValois R.L., and Yund, E.W., "Responses Of Striate Cortex Cells To Grating And Checkerboard Patterns," *J. Physiol.* **291** (1979), 483-505.

Dowling, J.E. and Boycott, B.B., "Organization Of Primate Retina: Electron Microscopy," *Proc. R. Soc. Lond. B* **166** (1966), 80-111.

Duda, R.O., and Hart, P.E., *Pattern Recognition and Scene Anaylsis*, John Wiley and Sons, New York, New York, 1973.

Enroth-Cugell and Robson, J.G., "The Contrast Sensitivity Of Retinal Ganglion Cells Of The Cat," *J. Physiol.* **187** (1966), 517-552.

Glezer, V.D., Ivanov, V.A., and Tscherback, T.A., "Investigation Of Complex And Hypercomplex Receptive Fields Of Visual Cortex Of The Cat As Spatial Frequency Filters," *Vis. Res.* **13** (1973), 1875-1904.

Herskovitz, A. and Binford, T.O., "On Boundary Detection," , M.I.T. AI Lab Memo 183 (1970).

Horn, B.K.P., "Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from one View," *A.I. Technical Report 79* , Project MAC, M.I.T. (1970).

Horn, B.K.P.H, "Linear Analysis And Image Processing," *M.I.T. A.I. Working Paper 100* (1974).

Horn, B.K.P., "Obtaining Shape from Shading Information", *The Psychology of Computer Vision, P. H. Winston (ed.),* McGraw Hill Book Company Inc., 1975, 115–155.

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

Horn, B.K.P. and Sjoberg, R.W. , "Calculating the Reflectance Map," *Applied Optics* **18** (1979), 1770-1779.

Hubel, D.H., and Wiesel, T.N., "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex," *J. Physiol.* **160** (1962), 106-154.

Hubel, D.H., and Wiesel, T.N., "The Ferrier Lecture: Functional architecture of macque monkey visual cortex," *R. Soc. Lond. B.* **198** (1977), 1-59.

Hueckel, M.H., "A Operator Which Locates Edges In Digital Pictures," *J. ACM* **18** (1971), 113-125.

Ikeuchi, K. and Horn, B.K.P. , "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence* **15**, Special Issue on Computer Vision (1981), 141-184.

Issacson, E. and Keller, H.B., *Analysis Of Numerical Methods*, John Wiley and Sons, New York, New York, 1966.

Kauth,R.J., Lambeck, P.F., Richardson, W., Thomas, G.S., and Pentland, A.P., "Feature Extraction Applied To Agricultural Crops As Seen By Landsat," *Proceedings of The LACIE Symposium* , Lyndon B. Johnson Space Center, Houston, Texas. (1978).

Kelly, D. H., "Frequency Doubling in Visual Responses," *J. Opt. Soc. Am.* **56** (1966), 1628-1633.

Kelly, D. H., "Spatial Frequency Selectivity In The Retina," *Vis. Res.* **15** (1975), 665-672.

Kelly, D. H., "Motion and vision. II. Stabilized spatio-tenporal threshold surface," *J. Opt. Soc. Am.* **69**, 10 (1979), 1340-1349.

Kender, S., *Ph.D Thesis, Carnigie Mellon Univerisity*, 1980.

Kirnov, E. L., "Spectral Reflectance Properties Of Natural Formations," , G. Belkov, translator. NRC of Canada Technical Translation 439 (1971).

Kuffler, S.W., "Neurons In The Retina: Organization, Inhibition And Excitation Problems," *Cold Spring Harbor Symp. on Quant. Biology* **17** (1952), 281-292.

Kuffler, S.W., "Discharge Patterns And Functional Organization Of Mammalian Retina," *J. Neurophys.* **16** (1953), 37-68.

Levinson, E. and Sekuler, R., "The Independence Of Channels In Human Vision Selective For Direction Of Movement," *J. Physiol. (Lond)* **250** (1975), 347-366.

Lipshitz, M.M., *Differential Geometry,* McGraw-Hill Book Company, New York, New York, 1969.

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

Macleod, I.D.G., "Comments on 'Techniques For Edge Detection'," *Proceedings of the IEEE* **60** (1972), 344.

Marr,D., "Early Processing Of Visual Information," *Phil. Trans. R. Soc. Lond. B.* **275** (1976), 483-524.

Marr,D., "Analysis Of Occluding Contour," *Proc. R. Soc. Lond. B.* **197** (1977), 441-475.

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

Marr and Poggio, "A Computational Theory of Human Stereo Vision," *Proc. R. Soc. Lond.* **204**, B (1979), 301-328.

Pentland, A., "Human Visual Processing And Remote Sensing: Finding The Edges," *Proceedings of the Fourteenth International Symposium On Remote Sensing Of The Environment, San Jose, Costa*

Rica. (1980).

Pentland, A., "Estimating The Illuminant Direction To Obtain Shape From Shading," *Opt. Soc. Am.*, Recent Advances In Vision (1980).

Pentland, A., "Evaluation Of Signature Extension Algorithms," , National Aeronautics and Space Administration report no. ERIM 109600-39-F, Environmental Research Institute of Michigan, Ann Arbor Michigan (1977).

Pentland, A., "Finding The Illuminant Direction," *M.I.T. AI Laboratory Memo 584* (1982).

Pentland, A., "Finding The Illuminant Direction," *J. Opt. Soc. Am* (forthcoming).

Pratt, W., *Digital Image Processing.* John Wiley and Sons, New York, New York, 1978.

Richards, W. A., Hoffman, D. D., and Rubin, J. M., "Equation Counting And The Interpretation Of Sensory Data," *M.I.T. A.I. Memo 614* (1981).

Richardson, W. and Pentland, A., "Evaluation Of Algorithms For Estimating Wheat Acreage From Multispectral Scanner Data," , National Aeronautics and Space Administration report no. ERIM 109600-26-F, Environmental Research Institute of Michigan, Ann Arbor Michigan. (1976).

Richter, J. and Ullman, S., "A Model For The Spatio-Temporal Organization Of X And Y Ganglion Cells In The Primate Retina," *A.I. Memo 573(A)* (1981).

Roberts, L.G., "Machine Perception of Three-Dimensional Solids," *Optical and Electro-Optical Information Processing*, (J. Tippet, D. Berkowitz, L. Clapp, C Koester, A. Vanderbergh Eds.) MIT Press (1965), 159-197.

Robson, J.G., "Spatial and Temporal Contrast Sensitivity Functions of the Visual System," *J. Opt. Soc. Am.* 56 (1966), 1141-1142.

Roderick, R.W., and Stone, J., "Responses Of Cat Retinal Ganglion Cells To Moving Visual Patterns," *J. Neurophys.* **28** (1965a), 819-832.

Roderick, R.W., and Stone, J., "Analysis Of Receptive Fields Of Cat Retinal Ganglion Cells," *J. Neurophys.* **28** (1965b), 833-849.

Rose, D., "Mechanisms Underlying The Receptive Field Properties Of Neurons In Cat Visual Cortex," *Vis. Res.* **19** (1979), 533-544.

Rosenfeld, A. and Thurston, M., "Edge and Curve Detection for Visual Scene Analysis," *IEEE Transactions on Computers* **C-20** (1971), 562-569.

Rosenfeld, A., Thurston, M., and Lee, Y., "Edge and Curve Detection: Further Experiments," *IEEE Transactions on Computers* **C-21** (1971), 677-715.

Rosenfeld, A. and Kak, A. C., "Digital Picture Processing," *Academic Press, New York* (1976).

Rosenfeld, A., Hummel, R. A., and Zucker, S. W., "Scene Labeling By Relaxation Operations," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6 (1976), 420-433.

Scharf, D., *Magnifications -- Photography With The Scanning Electron Microscope,* Schocken Brooks, New York, New York, year of publication unknown.

Shiller, P.H., Findlay, B.L., and Volman, S.F., "Quantitative Studies Of Single Cell Properties In Monkey Striate Cortex. I. Spatio-Temporal Organization Of Receptive Fields," *J. Neurophys.* **39** (1976a), 1288-1319.

Shiller, P.H., Findlay, B.L., and Volman, S.F., "Quantitative Studies Of Single Cell Properties In Monkey Striate Cortex. II. Orientation Specificity And Ocular Dominance," *J. Neurophys.* **39** (1976b), 1334-1351.

Shirai, Y.A., "A Contex-Sensitive Line-finder for Recognition of Polyhedra," *Artificial Intelligence* **4** (1975), 95-113.

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

Stevens, K., "The Visual Interpretation of Surface Contours," *Artificial Intelligence* **17** (1981), 47-73.

Strat, M.T., "A Numerical Method for Shape-From-Shading From a Single Image," *Artificial Intelligence Laboratory, M.I.T.*, Working Paper 191 (1979).

Torrance, K.E. and 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. and 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., "On Visual Detection Of Light Sources," *Biological Cybernetics* **21** (1976), 205-212.

Ullman, S., "Relaxation And Constrained Optimization By Local Processes," *Computer Graphics and Image Processing* **10** (1979), 115-125.

Ullman, S., "personal communication," (1981).

Witkin, A., "Shape From Contour," *Ph.D. Thesis, Dept. Psychology, MIT* (1980).

Witkin, A., "Shape From Contour," *Artificial Intelligence* **17** (1981), 1-47.

Woodham, R.J., "Reflectance Map Techniques for Analyzing Surface Defects in Metal Casting," *Artificial Intelligence Laboratory, M.I.T.*, TR-457 (1978).

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.