Determining Intrinsic Scene Characteristics from Images

by

Ting-Chuen Pong

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science and Applications

APPROVED:

Robert M. Haralick

Linda G. Shapiro

Roger W. Ehrich

James R. Craig

Monte B. Boisen

August, 1984
Blacksburg, Virginia

DETERMINING INTRINSIC SCENE CHARACTERISTICS FROM IMAGES

by

Ting-Chuen Pong

Committee Chairman: Robert M. Haralick

Computer Science and Applications

(ABSTRACT)

Three fundamental problems in computer vision are addressed in this dissertation. The first deals with the problem of how to extract and assemble a rich symbolic representation of the gray level intensity changes in an image. Results show that the facet model based feature extraction scheme proposed here is superior to the other existing techniques. The second problem addressed deals with the interpretation of the resulting structures as three-dimensional object surfaces. The three different shape modules described in this dissertation are found to be useful in the recovery of intrinsic scene characteristics. Finally, mechanisms for interaction among different sources of information obtained from different shape modules are studied. It is demonstrated that interactions among shape modules can enhanced the data acquired by different means.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Robert Haralick, for his guidance from entrance to final completion of my doctoral program. Without his advice and support, this dissertation would not have been possible.

Special appreciation is expressed to Professor Linda Shapiro. She gave unselfishly of her time, provided advice and assistance for beyond her obligation as a member of my graduate committee. I would also like to thank Professor Roger Ehrich, Professor James Craig, and Professor Monte Boisen for serving on my committee. Thanks also to Professor for helping me solve many mathematical problems.

This research was made possible because of the unique facilities and atmosphere of the SDA Laboratory. For this, I would like to thank my friends, and all other people who have made my work there enjoyable.

To , my wife, and , my eight month old son, I owe them a lot of love, affection, and shared time. I hope I'll be able to compensate them enough for everything they missed during this period. I would like to tell my wife how

much her love and understanding have helped me through the graduate program.

Most of all, I need to express my sincere gratitude to my parents, to whom this dissertation is dedicated. Their love and support have made this dissertation possible.

TABLE OF CONTENTS

<u>Appendix</u>

page

LIST OF FIGURES

Chapter I

INTRODUCTION

During the past decade, computer vision systems have been studied with increasing interest by a large number of researchers in the field of artificial intelligence. In general, a scene of our three-dimensional world is input to a computer vision system in terms of digitized sensory information which is contained in a large two-dimensional array of image elements or pixels. The system is expected to output a description of the contents of the scene that produced the array. Despite considerable progress in recent years, solving this seemingly easy task has proven to be extremely difficult.

A trend of the recent mid-level computer vision research is to study the recovery of intrinsic scene characteristics such as surface orientations, reflectance properties and distance information from images [Ba78]. This recovery process is important because knowing these intrinsic scene characteristics will be a definite aid to higher-level scene analysis problems. Apparently, humans have the ability to determine these characteristics even for unfamiliar scenes. Since information about the scene being viewed is captured as observed gray level intensities in the image, it is reasonable to believe that some of these characteristics are recoverable from image intensity variations.

Many "shape-from-..." approaches have been proposed for determining intrinsic scene characteristics. Among these approaches, stereopsis ([Ma76] and [Gr81]) and motion ([Ul79] and [Ho81]) are currently the most studied subjects. Four other important sources of shape information are shading, texture, shadow and contour.

Smooth intensity variation (or shading) is an important clue for determining surface orientation. The shape-from-shading idea was first formulated by Horn [Ho75]. Since then, a great deal of work has been done in this area ([Ik81], [Wo80] and [Pe82]). If we assume a uniformly texture surface, surface orientations may be inferred from the way the coarseness of the image texture changes across the image. Shape-from-texture is another area of recent research ([Wi81] and [Ke82]). When shadows are located in an image, the shapes of the shadows can be used to determine three-dimensional information about the objects in the scene. Shadow analysis can be referred to as the process of locating shadow regions, finding correspondences between shadow casting objects and shadow regions, and deducing three-dimensional information about the objects involved in the shadow formation process. Theoretical work on shadow analysis can be found in [Sh83]. Shape-from-shadow methods have been found to be useful for estimating heights of objects in aerial images [Hu83]. Three-dimensional surface shape can also be inferred from the two-dimensional shapes of edges or curves in an image. Shape-from-contour methods ([St81] and [Br84]) have been found to be effective in determining the shape of a visible surface.

The shape modules use various sources of information in the images to infer intrinsic scene characteristics. Since intrinsic scene characteristics of the visible surface in an image are captured as observed intensities in the image, an important goal of early vision is to extract a rich symbolic representation of the gray tone intensity changes in the image. Among the many representations proposed, the following are the most important ones: (1) the primal sketch [Ma76] or zero-crossing edges [Ma80] of Marr, (2) the intrinsic images of Barrow and Tenenbaum [Ba78], (3) the relational trees of Ehrich and Foith [Eh78], and (4) the topographic primal sketch of Haralick, Watson and Laffey [Ha83].

The idea of the facet model ([Ha80] and [Ha81]) is used in the construction of the topographic primal sketch. The facet model assumes that image intensity values are noisy sampled observations of an underlying intensity surface. Thus, any interpretation made on the basis of a neighborhood of pixel values should be understood through the analysis of its underlying intensity surface. In the past, the facet model has been proven to be useful in edge and region analysis ([Ha80], [Ha84] and [Po84]), noise removal [Ha83], corner detection [Zu83] and optical flow [Ha83]. These successes suggest that there is indeed something fundamental about the approach. We believe that a unified approach based on the facet model and the topographic primal sketch can be developed and used to solve the computer vision problem.

1.1 OVERVIEW

Many processes can contribute to the recovery of intrinsic scene characteristics. In Chapters 3-5, three approaches for determining surface orientation are investigated. The most pertinent related literature is briefly discussed in Chapter 2.

In Chapter 3, the facet model ([Ha80],[Ha81]) is used in the recovery of surface orientation from single and multiple images. Two methods which determine surface orientation from image shading are presented. The first method works in the single image domain and is formulated as a non-linear optimization problem. Since three-dimensional scene information available in a single image is usually ambiguous, the optimization procedure can result in multiple solutions. The possibility of adding boundary constraints to the optimization process is also investigated. In the second method, additional images are obtained from the same viewing position, but with changing illumination direction. With these additional images, local surface orientation is determined uniquely by a linear optimization procedure. Experimental results of the shape-from-shading methods are then described.

A stereo matching procedure is described in Chapter 4. Three-dimensional positions of the visible surfaces in stereo images can be determined by simple trigonometric transformation once the stereo matching problem is solved. Most stereo matching methods involve matching low-level image features ([Ma76], [Ba79] and [Gr81]). The method to be described in Chapter 4 matches high-level image features. The extraction of high-level image features is described in Section 4.1. The high-level image features extracted are topographic structures, arc segments and region segments. The overall matching scheme is described in Section 4.2. Matching results on simple and complex images are discussed in section 4.3.

In Chapter 5, a technique for estimating object shape from topographic primal sketch is investigated. Given a gray tone image of a three-dimensional object, a topographic labeling of the image indicates the peaks and pits, ridges and valleys, and flats and hillsides of the underlying, continuous, gray tone surface. The patterns of these topographic labels capture information about the original three-dimensional object in the scene and about the illumination. In order to determine if estimation of three-dimensional shape from a topographic labeling is feasible, we have both analytically and experimentally determined the topographic labelings for images of some mathematically generated surfaces with varied directions of illumination. Our results indicate that such patterns do exist and will be useful in determining three-dimensional shape from two-dimensional images. A scheme for partial classification of three-dimensional object surface is proposed. Preliminary results are illustrated.

Chapter II

LITERATURE REVIEW

2.1 SHAPE FROM SHADING

A great deal of work has been done in determining surface orientation from image shading. The shape from shading idea is first formulated in Horn's paper [Ho75]. Horn approaches the shape from shading problem by solving the image irradiance equation which is expressed in the form of a first-order partial differential equation. He suggests this equation be solved using a modified characteristic strip-expansion method. While this method works well in the absence of noise, its behavior is uncertain when applied to a noisy image.

While Horn estimates surface orientation by solving the image irradiance equation by direct integration, Ikeuchi and Horn [Ik81] estimate surface normal by an iterative method. The iterative method attempts to minimize the difference between the observed image intensities and the predicted values which are obtained iteratively from two constraints. The first constraint is based on the image irradiance equation. The second constraint is provided by the assumption that the surface is smooth. As an additional constraint for both Horn's and Ikeuchi's methods, it is required that both the reflectance property of the surface and the illumination condition of the light source be known. Ikeuchi's method requires that occluding boundary information be specified as boundary conditions.

Stereographic projection is used in Ikeuchi's method because the gradients of an occluding boundary map to infinity in gradient space. In stereographic projection, orientations of surface normals are represented uniquely by points on the Gaussian sphere. A point on the Gaussian sphere is then projected onto a plane tangent at the north pole with the center of projection at the south pole. Therefore, an occluding boundary, which is represented by points on the equator of the Gaussian sphere, is projected into a circle of radius two. A modification of Ikeuchi's method can be found in [Sm83]. Other results in shape from shading from a single image include the work of Woodham [Wo81] and Pentland [Pe82]. These studies determine surface orientations of visible surfaces through local analysis of image shading.

Woodham [Wo80] introduces a technique to determine local surface orientation from images which are taken by varying the illumination direction, while keeping the viewing direction fixed. This technique is referred to as photometric stereo. In this method, since images are taken without altering the camera position, there is no problem in finding corresponding points in the images. This technique determines surface orientation at each pixel by finding the intersection of the reflectance maps of the images in the gradient space. In general, at least three images are needed to determine a unique surface orientation.

While Woodham's method is designed for matte surfaces, Ikeuchi [Ik81] applies the photometric stereo technique to determine surface orientations of specular surfaces. The point light source used in Woodham's method is unsuitable for specularly reflecting surfaces, since the specular component in the image due to point light source only occupies a very small area. Instead, Ikeuchi employs a distributed planar light source with spatially varying irradiation.

Although both Woodham's and Ikeuchi's methods work well for noise free images, these methods obtain surface orientation in an image on a pixel-by-pixel basis. Therefore, we would expect these methods to be very sensitive to noise. The problem is yet to be solved if the input images are known to be noisy.

2.2 STEREO MATCHING

Stereo matching has been approached mainly in two directions. The first approach matches areal features, while the second approach matches lineal features such as edges. The area based approach attempts to find a match for each neighborhood in the first image by searching for neighborhoods of maximum similarity in the second image. A typical similarity measure is the mean-square difference of the gray levels within the two neighborhoods. Since the search process is computationally expensive and not all pixels can be unambiguously matched, matches should be restricted to neighborhoods with high information content. One such technique for finding candidate points for matching is Moravec's interest operator [Mo77]. Moravec's operator computes an initial interest measure for each pixel by finding the minimum of the four directional variances (horizontal, vertical and two diagonals) over the neighborhood centered around the pixel. A pixel is of interest if its interest measure attains a local maximum. Moravec's interest operator tends to select corners.

Barnard and Thompson [Ba80] use a relaxation labeling technique to find matches for areal features. They use Marvavec's interest operator to select matchable points. The searching space is restricted by assuming a maximum disparity. An initial probability is given to each match based on the similarity of the neighborhoods around the pixel. A relaxation labeling technique is then used to improve the initial probabilities by using the local continuity property of disparity. A common problem with area-based approaches is that they are sensitive to changes in contrast and brightness between the pair of images.

The edge-based approach attempts to search for edge correspondences between edges detected from the stereo pair. Since an edge often corresponds to a discontinuity in the depth of an object, edge-based methods are less sensitive to photometric characteristics of the object in the scene. Furthermore, edge-based matchings are in general more accurate and less expensive. This is because there are fewer edges than pixels and edges can usually be detected at reasonably precise locations.

A computational theory of human stereo vision was formulated by Marr and Poggio [Ma76]. They suggest matching features which are located as zero-crossings of the Laplacian of a Gaussian filtered image. This scheme works well on random-dot stereograms, but fails for real images. Grimson [Gr81] implemented a modified version of Marr's algorithm. Grimson's approach seems to generate reasonable results for some real images. A coarse to fine approach is used in these methods. This is done by convolving the image with filters of various resolutions. This approach tends to fail for complex scenes because edges detected from low resolution filters do not usually correspond to any true edges and incorrect matches are often introduced. Thus, correct matches as well as incorrect matches propagate from the low levels to the higher levels.

More recently, Baker and Binford [Ba81] and Arnold [Ar83] developed two similar methods for stereo matching based on dynamic programming technique. The Viterbi dynamic programming algorithm is used in the matching process. They attempt to find an optimum match for each pair of epipolar lines in the image by assuming that there is no order reversal of edges. The optimum match is obtained by maximizing a measure derived from local properties of edges and by requiring that an edge angle and interval constraint be satisfy. One limitation of these methods is that a dynamic programming implementation is possible only if the edge ordering assumption is not violated.

Until recently, most feature based stereo matching systems match low level edge primitives. Medioni and Nevatia

[Me83] proposed a technique for matching higher level primitives, namely, line segments. Matching line segments from the two images are searched within a parallel sided window which has two of its sides parallel to the orientation of a line segment and the other two sides horizontal. By using such a window, all epipolar lines that intersect the line segment to be matched are considered and a maximum disparity is allowed for each match. This is an iterative technique. At each iteration, each possible match is evaluated by assuming a minimal differential disparity. A preferred match for each segment is selected to be the one which is most consistent with the disparities suggested by it neighbors.

While Mediono and Nevatia solve the correspondence problem by matching line segments, Goshtasby [Go83] suggests using region segments. Goshtasby first uses an image segmentation procedure to determine optimally similar regions, by an appropriate measure, from multiple images. Region correspondences are then achieved by a probabilistic relaxation labeling process.

Cheng and Huang [Ch82] present a method for registering images. A relational description is used in their method to represent structures extracted from images. They extract straight line segments as primitives and use relations such as parallel, antiparallel, collinear and adjacent to describe the interrelationships between line segments. The image correspondance problem is then solved by matching relational structures. As another step toward using higher level primitives, the matching scheme to be described in Chapter 4 matches arc segments, region segments, and topographic structures, and uses spatial relations among these structures to resolve ambiguous matches.

Chapter III

A FACET APPROACH TO SHAPE FROM SHADING

3.1 INTRODUCTION

Of the clues available in an image, image shading is one of the important sources of information for determining these scene characteristics [Ho75]. In the present work, the intrinsic property which we are primarily interested in is surface orientation. In this chapter, the facet model ([Ha80],[Ha81]) is used in the recovery of surface orientations from single or multiple images. The facet model assumes that image intensity values are noisy sampled observations of an underlying intensity surface. Thus, any interpretation made on the basis of a neighborhood of pixel values should be understood through the analysis of its underlying intensity surface.

In past facet work [Ha83], the underlying intensity surface was estimated by a least squares fit with a functional form consisting of a linear combination of the tensor products of discrete orthogonal polynomials. In the new facet shape-from-shading formulation, the three-dimensional object surface is assumed to be locally fit by a quadratic surface. The underlying gray tone intensity surface is then modeled using a quadratic object surface having Lambertian reflectance. Having estimated the free parameters of the Lambertian intensity surface, the surface orientation around each pixel is readily determined.

3.1.1 Overview

Two methods for determining surface orientation from image shading information are presented in this chapter. The first method works in the single image domain and the second method works for multiple images. Before discussing these techniques in detail, the image geometry and the illumination model to be used in our presentation are defined in Section 3.1.2 and 3.1.3, respectively. The basic theory of the facet based shape-from-shading method is given in section 3.2.1. In section 3.2.1.1, the selection of starting value, which is required in our method, is discussed. A modification of the original technique is also suggested. The second method which works for multiple images is described in section 3.2.2. A possible extension of these methods to handle specular reflection is given in section 3.3. In section 3.5, results of some preliminary experiments are discussed.

3.1.2 Imaging Geometry

The relationship between scene coordinates and image coordinates is illustrated in Figure 1. We assume that the camera lens is at the origin and that the z-axis is directed towards the image plane which is in front of the lens. The image plane is placed at a distance f, the focal length of the lens, in front of the origin so that the image is oriented in the same way as the scene. As seen from Figure 1, the following relations hold for perspective projection:

$$u = \frac{fx}{z} \quad \text{and} \quad v = \frac{fy}{z} .$$

In our discussion, the perspective projection is approximated by an orthographic projection. This approximation is good when the size of the objects being imaged is small compared to the viewing distance. In this case, appropriate coordinate systems can be chosen such that the following relations hold:

$$u = x \quad \text{and} \quad v = y .$$

3.1.3 Illumination Model

In the following discussion, we will use a simple illumination model that assumes a distant point light source and a Lambertian reflectance model. A Lambertian surface

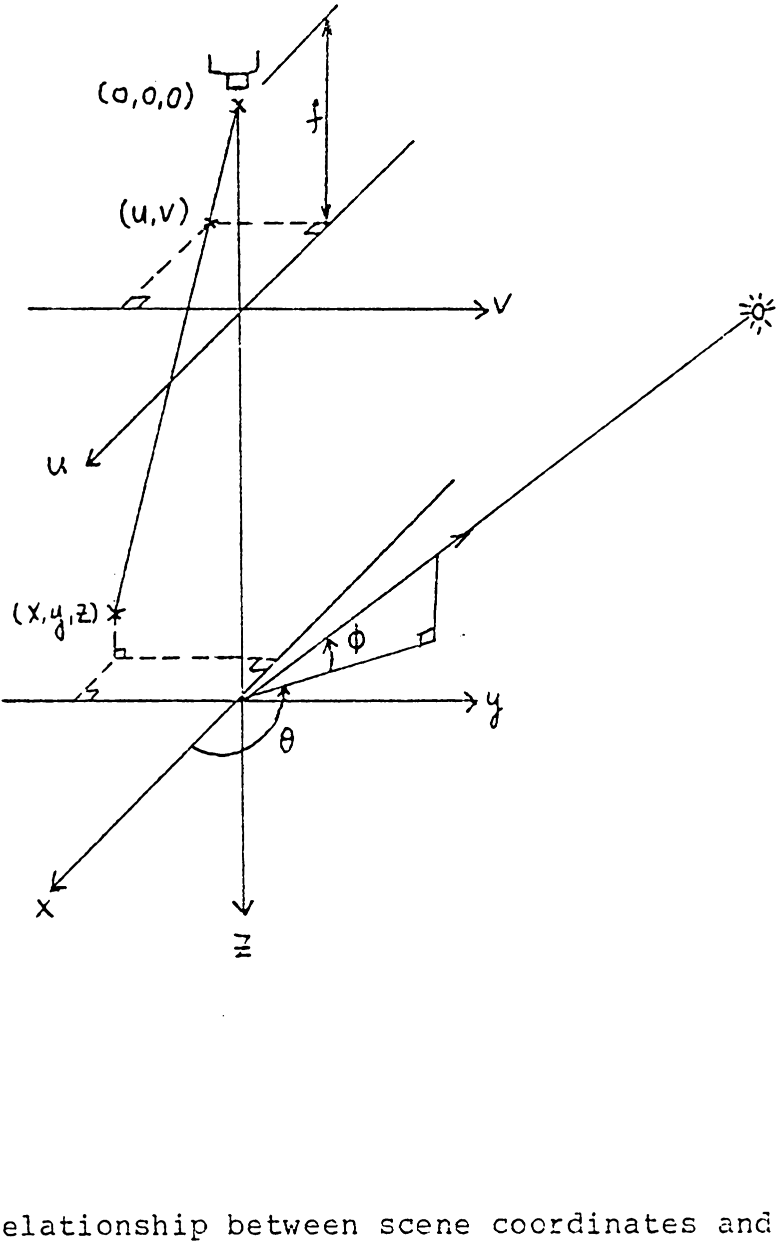


Figure 1: Relationship between scene coordinates and image coordinates.

scatters light equally in all direction. The brightness of a Lambertian surface illuminated by a distant point light source is given by:

$$I = I_0 \ N \cdot L \qquad (1)$$

where I_0 is a constant depending on the surface albedo and the intensity of the light source, N is the unit surface normal vector, and L is the unit vector of the illumination direction.

The unit vector which points in the direction of the light source can be specified by the two angles shown in Figure 1. The first is the azimuth (θ) which is the angle between the x-axis and the projection of the vector onto the x-y plane, while the second is the angle of elevation (ϕ) of the light source. If we represent this unit vector by [a,b,c], then

$$a = \cos\theta \ \cos\phi ,$$

$$b = \sin\theta \ \cos\phi , \text{ and}$$

$$c = -\sin\phi . \qquad (2)$$

In our discussion, we will consider only positive values of ϕ. Therefore, c is always less than zero.

If the height of the object surface above the x-y plane is expressed as a function of x and y,

$$z = S(x,y) ,$$

then the surface normal is given by the vector:

$$N = [\, S_x,\ S_y,\ -1 \,]/(1 + S_x^2 + S_y^2)^{1/2}$$

where S_x and S_y denote first partials of S with respect to x and y, respectively. By carrying out the dot product in Eq.(1), it follows that

$$I = I_0 \frac{aS_x + bS_y - c}{(1 + S_x^2 + S_y^2)^{1/2}} \quad .$$

3.2 THE FACET MODEL APPROACH

In the following sections, two methods which determine surface orientation from image shading information are presented. The first method works in the single image domain and is formulated as a non-linear optimization problem. Since three-dimensional scene information available in a single image is usually ambiguous, the optimization procedure can result in multiple solutions. In the second method, additional images are obtained from the same viewing position, but with changing illumination direction. With these additional images, the non-linear system of image irradiance equations are transformed into a set of linear equations. The least squares procedure described in section 3.2.2 enables a unique solution to be obtained from this linear system of equations.

3.2.1 Shape from Shading from a Single Image

The first method determines surface orientation from the shading information of a single image. It is formulated as a non-linear optimization problem. According to the facet model, image intensity values are the results of noisy sampling of an underlying intensity surface. Therefore, the image intensity of a Lambertian surface illuminated by a distant point light source under orthographic projection can be expressed as:

$$I(x,y) = I_0 \frac{aS_x(x,y) + bS_y(x,y) - c}{(1 + S_x^2(x,y) + S_y^2(x,y))^{1/2}} + n(x,y) \quad (3)$$

where n denotes the independent and identically distributed noise. Furthermore, a smoothness constraint on the object surface is imposed in the following way. Let R be an arbitrary neighborhood over which equation (3) holds. We assume that

$$S(x,y) = \sum_{i=1}^{n} k_i \, h_i(x,y) \qquad \text{for } (x,y) \in R,$$

where the h_i's are some basis functions and the k_i's are constants. Different neighborhoods have different coefficients. The form for S that we currently use is a bivariate quadratic polynomial:

$$S(x,y) = k_0 + k_1x + k_2y + k_3x^2 + k_4xy + k_5y^2$$

By using the intensity values $\{I(x,y) \mid (x,y) \in R\}$, we can determine a least squares estimator $\hat{\underline{k}}$ of $\underline{k}$ (note that $\underline{k}$ denotes the vector $[k_1,k_2,\ldots,k_n]$).

The least squares approximation problem is to determine coefficients k_i such that

$$e^2 = \sum_{(x,y)\in R} \left| I(x,y) - I_0 \frac{aS_x(x,y) + bS_y(x,y) - c}{(1 + S_x^2(x,y) + S_y^2(x,y))^{1/2}} \right|^2 \qquad (4)$$

is minimized. In the case of a quadratic object surface,

$$S_x(x,y) = k_1 + 2k_3x + k_4y$$

and $$S_y(x,y) = k_2 + k_4x + 2k_5y \text{ .}$$

where S_x and S_y denote the partial derivatives of S with respect to x and y. Equation (4) can be rewritten as:

$$e^2 = \sum_{(x,y)\in R} \left| I(x,y) - F(x,y,\underline{k}) \right|^2 \qquad (5)$$

where

$$F(x,y,\underline{k}) = I_0 \frac{a(k_1+2k_3x+k_4y) + b(k_2+k_4x+2k_5y) - c}{[1 + (k_1+2k_3x+k_4y)^2 + (k_2+k_4x+2k_5y)^2]^{1/2}}$$

Equation (5) is non-linear in the k_i's. One of the methods for solving this non-linear system is the Marquardt Method [Dr81]. This method is basically an improvement of the Taylor series linearization method and the steepest descent method. The ideas of the linearization method and the steepest descent method are explained briefly here.

The linearization method begins by expanding $F(x,y,\underline{k})$ into a Taylor Series, that is,

$$F(x,y,\underline{k}) \cong F(x,y,\underline{k}_0) + \sum_{i=1}^{n} \left[\left. \frac{d\, F(x,y,\underline{k})}{d\, k_i} \right|_{\underline{k}=\underline{k}_0} (k_i - k_0(i)) \right] \quad (6)$$

where $\underline{k}_0$ is a starting value of $\underline{k}$ and $k_0(i)$ denotes the i^{th} component of $\underline{k}_0$. If we let $g_i = k_i - k_0(i)$, then equation (6) becomes linear in the g_i's. Therefore, the original problem becomes a linear regression problem and can be solved by the least squares procedure for linear models. The least squares estimator thus obtained is used as a new starting value for the Taylor series expansion and the process is repeated until convergence (i.e. until e^2 is less than some predetermined tolerance).

The idea of the steepest descent method is to move the estimate, from a starting point k_0, along a direction of steepest descent as long as e^2 decreases. The direction of steepest descent is estimated as the direction obtained by differentiating (5) with respect to $\underline{k}$. In general, the steepest descent method works well when the starting value is far from the desired solution and the linearization method does a better job when a good initial value is provided. However, Marquardt's method combines the good features of the two methods and works well for most practical problems.

The problem with using the above method for the shape from shading task is that it is very starting-value sensitive. A bad choice of starting value may lead to only a local minimum instead of a global minimum. Furthermore, multiple solutions may exist if the information available from the input image is ambiguous.

3.2.1.1 Initial Estimate and Boundary Constraints

The nonlinear optimization procedure described above is an iterative method. An important starting point is to obtain a good initial estimate. Better starting values not only increase the rate of convergence, they also reduce the chances of converging to local minima. We also believe that supplying additional constraints to the shape recovery process would result in an improved accuracy in the reconstructed surface. We want to explore in this section the possibility of using two kinds of boundary constraints, depth and surface orientation constraints, to bootstrap the shape-from-shading process. Sparse depth values can be obtained from stereo computations. A detailed discussion of stereo computation can be found in Chapter 4 of this thesis. The second constraint can be obtained from local analysis of image shading. Local analysis of image shading is given in [Pe82] and Chapter 5 of this thesis.

3.2.1.2 Depth Constraints

In binocular stereopsis, surface depth can be determined by matching image features which are usually extracted as zero-crossing edges [Ma80]. Given a pair of stereo images, depth of object surface can be determined by simple trigonometric transformation once the correspondence problem is solved. In stereo computations, corresponding points in the two images are determined only at locations with high information content. Since stereo matching can produce only a sparse depth map, some kind of interpolation technique is required to compute a complete surface description.

A computational theory of visual surface interpolation is proposed by Grimson [Gr83] and refined by Terzopoulos [Te83]. While this interpolation method constructs surface based on information available along zero-crossing contours, its reconstructed surface fails to account fully for the intensity changes which occur away from the zero-crossing contours. Therefore, the reconstructed surface may not be consistent with the observed intensity variations. Our intention here is to compute a complete surface reconstruction by augmenting shading information with depth constraints.

Let R be an arbitrary connected region over which the smoothness constraint (Equation 4) is satisfied. Let B be a set of pixels for which the depths are known. We have for (xi,yi) ∈ B,

$$k_0+k_1x_i+k_2y_i+k_3x_i^2+k_4x_iy_i+k_5y_i^2 \cong s_i \quad \text{for } i=1,..,n$$

where n is the number of pixels in B, and s_i gives the depth of the surface at (x_i,y_i). In matrix notation, we have

$$A \underline{x} \cong \underline{y}$$

where

$$\underline{x} = (k_0,k_1,k_2,k_3,k_4,k_5)^t,$$

$\underline{y}i$ = si, and

$a_{i.}$ = (1, xi, yi, xi2, xiyi, yi2), i=1,...,n

Given enough points in B, an initial estimate for the surface in R can be determined by a least-squares method which minimizes

$$e^2 = \sum_{i=1}^{n} ||(k_0+k_1x_i+k_2y_i+k_3x_i^2+k_4x_iy_i+k_5y_i^2) - s_i||^2$$

or equivalently,

$$e2 = || Ax - y||^2.$$

The solution to this least-squares problem is discussed in section 3.2.1.4.

3.2.1.3 Surface Orientation Constraints

In some locations of an image, surface orientations can be estimated locally. Pentland [Pe82] presents a method for determining surface orientation through local analysis of image shading. Ikeuchi and Horn [Ik81] suggest that constrained values of surface orientation can be obtained at occluding boundaries and singular points ([Ba81] and [Ma77]). In Chapter 5 of this thesis, an estimate of surface orientation is obtained through the analysis of the topographic primal sketch. The set of surface orientations estimated through local analysis provides another means for determining a good initial estimate.

Let R be an arbitrary connected region over which Equation 4 is satisfied. Let B be a set of pixels for which the surface orientations are known. We have for (xi,yi) $\in$ B,

$$k_1 + 2k_3x_i + k_4y_i \cong p_i, \qquad \text{for } i=1,\ldots,n ,$$

and $$k_2 + k_4x_i + 2k_5y_i \cong q_i, \qquad \text{for } i=1,\ldots,n ,$$

where n is the number of pixels in B, and (p_i,q_i) gives the surface orientation at (x_i,y_i). In matrix notation, we have

$$A\,\underline{x} \cong \underline{y}$$

where

$$\underline{x} = (k_1,k_2,k_3,k_4,k_5)^t,$$

$$\underline{y}i = \begin{cases} p_i & \text{if } i \leq n \\ q_{i-n} & \text{if } i > n, \text{ and} \end{cases}$$

$$a_{i.} = \begin{cases} (1,\ 0,\ 2x_i,\ y_i,\ 0) & \text{if } i \leq n \\ (0,\ 1,\ 0,\ x_{i-n},\ 2y_{i-n}) & \text{if } i > n\ . \end{cases}$$

As for the depth constraints, an initial estimate is determined by minimizing

$$e^2 = \sum_{i=1}^{n} ||(k_1+2k_3x_i+k_4y_i)-p_i||^2 + \sum_{i=1}^{n} ||(k_2+k_4x_i+2k_5y_i)-q_i||^2$$

or equivalently,

$$e2 = ||\ Ax - y||^2.$$

3.2.1.4 Singular Value Decomposition

There are many algorithms for solving the least-squares problems. The most reliable method is the singular value decomposition method. A singular value decomposition of a rectangular mxn matrix A with $m \geq n$ is a factorization of the form

$$A = U\ S\ V^t$$

where U is an mxm orthogonal matrix, V is an nxn orthogonal matrix, and S is an mxn diagonal matrix with nonnegative diagonal entries. The diagonal elements of S are called singular values of A.

Using the singular decomposition of A, the linear system $Ax = y$ becomes

$$U\,S\,V^t\,x = y.$$

The error,

$$e^2 = ||\,A\hat{x} - y\,||^2,$$

to be minimized in the least-squares problem becomes

$$e^2 = ||\,U\,S\,V^t\,\hat{x} - y\,||^2 .$$

Since orthogonal matrices preserve norm, we have

$$\begin{aligned} e &= ||\,U^t\,(U\,S\,V^t\,x - y)\,|| \\ &= ||\,S\,V^t\,x - U^t\,y\,||. \end{aligned}$$

If we let $z = V^t x$ and $v = U^t y$, we then have

$$e = ||\,S\,z - v\,||.$$

Since S is a diagonal matrix, it is easily seen that the vector z that produces the minimum e is given by

$$\begin{aligned} z_i &= v_i/s_i \quad &&\text{if } s_i \neq 0, \text{ and} \\ z_i &= 0 &&\text{if } s_i = 0. \end{aligned}$$

The transformation $x = Vz$ then determines the least-squares solution.

3.2.1.5 Possible Modification

We have shown how boundary constraints can be used to obtain initial estimates. As an extension of the algorithm described in Section 3.2.1, we may also consider a minimization of the form

$$w \sum_{(x,y) \in R} ||I(x,y) - F(x,y,k)||^2 + (1-w) \sum_{(x',y') \in B} ||Ax - y||^2$$

where w is a constant between zero and one. The first part of the above expression is acquired from Equation (5) which gives the error between the observed and the estimated gray level intensities. The second part designates the deviation between the estimates and the boundary conditions. The constant w is set close to one if the boundary conditions are not certain to be reliable. The same Marquardt method can be used to solve this minimization problem. Experiments show that this formulation is superior to the original formulation.

3.2.1.6 Extracting Regions

So far, we have described a shape-from-shading method. One problem remains to be solved is to find regions over which the smoothness constraints are satisfied. This region extraction process can be accomplished by analyzing the topographic primal sketch of the image intensity surface. While a detail description of the region extraction process is given in Chapters 3 and 4 of this thesis, the overall process is summarized as follows:

1. extract edge structures based on the facet model second directional derivative zero-crossing edge operator;
2. extract maximally connected components of non-edge structures as an initial set of region segments; and
3. Extract regions by assembling topographic structure within the initial set of region segments.

3.2.2 Shape from shading from multiple images

As mentioned in Section 3.2.1, three-dimensional scene information available in a single image is sometimes ambiguous. Woodham [Wo80] suggests a method to resolve this ambiguity by using multiple images. In this method, multiple images are obtained by varying the illumination direction, while keeping the viewing direction fixed. In this section, we will describe a variation of Woodham's method. This method is also based on the facet model. We believe that our method is more robust than Woodham's method.

Suppose two images $I_1(x,y)$ and $I_2(x,y)$ are obtained by varying the position of a point light source. The irradiance equations corresponding to $I_1(x,y)$ and $I_2(x,y)$ are

$$I_1(x,y) = I_0 \frac{S_x(x,y)a_1 + S_y(x,y)b_1 - c_1}{(1 + S_x^2(x,y) + S_y^2(x,y))^{1/2}}$$

and

$$I_2(x,y) = I_0 \frac{S_x(x,y)a_2 + S_y(x,y)b_2 - c_2}{(1 + S_x^2(x,y) + S_y^2(x,y))^{1/2}} ,$$

where $[a_1,b_1,c_1]^t$ and $[a_2,b_2,c_2]^t$ are the unit vectors which point in the directions of the light sources.

In the case of a quadratic object surface,

$$I_1(x,y) = I_0 \frac{a_1(k_1+2k_3x+k_4y) + b_1(k_2+k_4x+2k_5y) - c_1}{[1 + (k_1+2k_3x+k_4y)^2 + (k_2+k_4x+2k_5y)^2]^{1/2}} \quad (7)$$

$$I_2(x,y) = I_0 \frac{a_2(k_1+2k_3x+k_4y) + b_2(k_2+k_4x+2k_5y) - c_2}{[1 + (k_1+2k_3x+k_4y)^2 + (k_2+k_4x+2k_5y)^2]^{1/2}} \quad (8)$$

Note that $I_1(x,y)$ and $I_2(x,y)$ do not depend on k_0. This is expected because a distant point light source is assumed; a translation of the object along the viewing direction does not affect the observed image intensity.

The above irradiance equations are nonlinear in the k_i's. By taking the ratio of the two irradiance equations, we will show that local surface orientation can be determined by a linear optimization procedure.

3.2.2.1 Estimation of Surface Orientation

Although (7) and (8) are nonlinear in the k_i's, we can linearize the problem by taking their ratio. That is

$$\frac{I_1(x,y)}{I_2(x,y)} = \frac{(k_1+2k_3x+k_4y)a_1 + (k_2+k_4x+2k_5y)b_1 - c_1}{(k_1+2k_3x+k_4y)a_2 + (k_2+k_4x+2k_5y)b_2 - c_2} . \quad (9)$$

If we let $I(x,y) = I_1(x,y)/I_2(x,y)$, equation (9) can be rewritten as:

$$\begin{aligned}&(I(x,y)a_2 - a_1)k_1 + (I(x,y)b_2 - b_1)k_2\\&+ 2x(I(x,y)a_2 - a_1)k_3\\&+ [x(I(x,y)b_2 - b_1) + y(I(x,y)a_2 - a_1)]k_4\\&+ 2y(I(x,y)b_2 - b_1)k_5 - (I(x,y)c_2 - c_1) = 0 \qquad (10)\end{aligned}$$

Equation (10) is linear in the k_i's. We can estimate the k_i's by fitting (10) to the observed $I(x,y)$s in the neighborhoods of I1 and I2. For simplicity, we will consider only odd sized rectangular neighborhoods and the coordinates of the center of the neighborhood RxC will be set to (0,0). It is important to note that the surface normal at the center of the neighborhood is given by $[k_1,k_2,-1]^t$. For $(x,y) \in RxC$, the system of equations can be represented in matrix notation by

$$B\ \underline{a} = \underline{y}\ , \qquad (11)$$

where, if $\underline{b}(x,y)$ is the row vector

$$[\ I(x,y)a_2-a_1,\ I(x,y)b_2-b_1,\ 2x(I(x,y)a_2-a_1),\\ x(I(x,y)b_2-b_1)+y(I(x,y)a_2-a_1),\ 2y(I(x,y)b_2-b_1)\ ],$$

then

$$B = \begin{vmatrix} \underline{b}(-R,-C) \\ \underline{b}(-R,-C+1) \\ \vdots \\ \vdots \\ \underline{b}(0,0) \\ \vdots \\ \vdots \\ \underline{b}(R-1,C-1) \\ \underline{b}(R,C) \end{vmatrix}, \quad \underline{y} = \begin{vmatrix} I(-R,-C)c_2-c_1 \\ I(-R,-C+1)c_2-c_1 \\ \vdots \\ \vdots \\ I(0,0)c_2-c_1 \\ \vdots \\ \vdots \\ I(R-1,C-1)c_2-c_1 \\ I(R,C)c_2-c_1 \end{vmatrix}$$

and $\underline{a} = [\, k_1, k_2, k_3, k_4, k_5 \,]^t$.

A least squares procedure which minimizes

$$e^2 = |\, \underline{y} - B\, \hat{\underline{a}} \,|^2$$

is used to determine the estimator $\hat{\underline{a}}$. The normal equation estimator for a is given by

$$\hat{\underline{a}} = (\, B^t B \,)^{-1}\, B^t\, \underline{y} \, . \qquad (12)$$

If there is exact linear dependency between the columns of B, the rank of B is less than 5 and B^tB is singular. In general, the columns of B are near linearly dependent. In this case, B^tB is ill-conditioned and the estimate of $\underline{a}$ suffers from large error. A more reliable method for solving this least-squares problem is based on a matrix factorization known as the singular value decomposition which is described in section 3.2.1.4. However, if the problem is rank deficient (i.e. at least one of the singular values of A is zero), the solution which minimizes e^2 is not unique. Two approaches are suggested to avoid this problem.

The first approach is to use a third image. An additional set of equations can be obtained by taking the ratio of the irradiance equation of the third image and that of the first (or second). A new matrix B is obtained by combining these equations. It is clear that the effect of multicollinearity in the new B can be reduced, provided the vectors of the incident illuminations do not all lie in a plane. The second approach is to constrain the locations of the light sources.

3.2.2.2 Locations of the light sources

Recall from section I.1 that the direction of a distant point light source is specified by two angles: the azimuth (θ) and the elevation (ϕ). It follow from Eq.(2) that if θ is 90^o, then a becomes zero, if θ is 0^o then b becomes zero and if ϕ is 90^o both a and b are zero. The angle of elevation of the light source is always made greater than zero, therefore, c is always non-zero.

We can simplify equation (10) by restricting the positions of the light sources. If both a_1 and a_2 are set to zero, equation (10) can be rewritten as

$$(I(x,y)b_2 - b_1)k_2 + x(I(x,y)b_2 - b_1)k_4$$
$$+ 2y(I(x,y)b_2 - b_1)k_5 - (I(x,y)c_2 - c_1) = 0$$

or

$$k_2 + xk_4 + 2yk_5 = J(x,y) \quad ,$$

where $J(x,y) = (I(x,y)c_2 - c_1)/(I(x,y)b_2 - b_1)$. Therefore, the matrix B in (11) can be replaced by

$$B = \begin{vmatrix} 1 & -R & 2(-C) \\ 1 & -R & 2(-C+1) \\ & \vdots & \\ 1 & 0 & 0 \\ & \vdots & \\ 1 & R & 2(C-1) \\ 1 & R & 2C \end{vmatrix},$$

$\underline{y}$ becomes a column vector of the J(x,y)'s and $\underline{a}$ becomes $[k_2, k_4, k_5]^t$. Note that B is independent of the I's and because the center of each neighborhood has (x,y) equal to (0,0), B^tB is the same for all the neighborhoods in the image. Furthermore, B^tB in (12) is always invertible and can be computed accurately. Therefore, the computation of $\hat{\underline{a}}$ from equation (12) is stable and efficient.

We can determine k_2, k_4 and k_5 accurately if a_1 and a_2 are both zero. Similarly, k_1, k_3 and k_4 can be obtained by making both b_1 and b_2 equal to zero. Hence, if we obtain a first image with $\theta=90$ and $\phi=90$, a second image with $\theta=90$ and arbitrary ϕ, and a third image with $\theta=0$ and arbitrary ϕ, then $\hat{\underline{a}}$ can be determined accurately and efficiently. If (k_1,k_3,k_4) and (k_2,k_3,k_5) are determined independently as described, it is possible that the k_4's computed from the two processes disagree. In the case where the k_4's differ by a significant amount, we can either take their average or use the three images together to solve for a unique $\underline{a}$.

If only two images are to be used, we can first compute k_2, k_4 and k_5 by setting a_1 and a_2 to zero (or k_1, k_3 and k_4 by setting b_1 and b_2 to zero) and then obtain k_1 (or k_2) from the following derivations.

If $a_1=0$ and $a_2=0$, one obtains from (7)

$$I(0,0) = \frac{I_0(k_2b_1 - c_1)}{(1 + k_1^2 + k_2^2)^{1/2}} \tag{13}$$

or

$$1+k_1^2+k_2^2 = [I_0(k_2b_1-c_1)]^2 \,/\, I(0,0)^2 \tag{14}$$

Similarly,

$$1+(k_1^2-2k_1k_4+k_4^2)+(k_2^2-4k_2k_5+4k_5^2) = [I_0((k_2-2k_5)b_1-c_1)]^2 \,/\, I(0,-1)^2 \tag{15}$$

and

$$1+(k_1^2+2k_1k_4+k_4^2)+(k_2^2+4k_2k_5+4k_5^2) = [I_0((k_2+2k_5)b_1-c_1)]^2 \,/\, I(0,1)^2 \tag{16}$$

If I_0 is known, one can obtain k_1 directly from (13). Otherwise,

let $h_0 = (k_2b_1-c_1)^2 \,/\, I(0,0)^2$,

$h_{-1} = ((k_2-2k_5)b_1-c_1)^2 \,/\, I(0,-1)^2$ and

$h_1 = ((k_2+2k_5)b_1-c_1)^2 \,/\, I(0,1)^2$,

then

$$(16)-(15) \Rightarrow 4k_1k_4+8k_2k_5 = I_0^2(h_1-h_{-1}) \tag{17}$$

and

$$(14)-(15) \Rightarrow 2k_1k_4-k_4^2+4k_2k5-4k_5^2 = I0^2(h_0-h_{-1}). \tag{18}$$

If $k_4 \neq 0$, by taking the ratio of (17) and (18), one finds

$$k_1 = \frac{h(4k_2k_5-k_4^2-4k_5^2) - 8k_2k_5}{2k_4(2-h)},$$

where $h = (h_1-h_{-1})/(h_0-h_{-1})$.

If $k_4=0$, by taking the ratio of (15) and (16), one finds

$$k1^2 = \frac{h[1+(k_2^2-4k_2k_5+k_5^2] - [1+(k_2^2+4k_2k_5+4k_5^2]}{1-h}.$$

where $h = h_{-1}/h_1$. In this case, the sign of k_1 is undetermined.

In this section, a photometric stereo method has been presented. By taking the ratio of the irradiance equations of a pair of images, we are able to determine local surface orientation uniquely by a linear optimization procedure. It was also shown that adding a third image and restricting the position of the light sources allows us to determine surface orientations accurately and efficiently. The photometric stereo technique is particularly useful in controlled environments where the intensity and position of light source can be adjusted and measured accurately. Some possible applications of this technique include robotic hand-eye systems and industrial inspection tasks.

3.3 SPECULAR REFLECTANCE MODEL

A more general shading model consists of two components. The first component is Lambertian reflection. Lambertian surfaces (or matte surfaces) scatter light equally in all directions. Therefore, they are seen as equally bright from all directions. Specular or mirror-like reflection accounts for the second component. A surface having specular reflection reflects light unequally in different directions. Specular reflection is usually observed as highlights in the image.

Let θ and ϕ be the angles defined in Figure 2. The angle θ is the angle between the incident ray (L) and the surface normal (N). Note that the angle between the reflected ray (R) and the surface normal is also θ. The angle ϕ is the angle between N and the vector half way between the viewing direction (V) and L. According to the shading model described in [Ha83], the intensity of a specular surface illuminated by a point light source can be expressed as:

$$I = I_1 \cos\theta + I_s \cos^n \phi \qquad (1)$$

where the values of Il and Is depend on the properties of the surface material, the distance of the light source to the surface element, and the intensity of the light source, and n is a constant. The value of n ranges roughly from 1 to 200, depending on the surface. A shiny metallic surface will have a very large n. A dull surface will have a small value. If we let $\hat{L}$, $\hat{N}$ and $\hat{V}$ be, respectively, the unit vectors of L, N and V, then equation (1) can be rewritten as:

$$I = I_1(\bar{L}.\hat{N}) + I_s(\bar{N}.\hat{H})^n.$$

where

$$\hat{H} = (L+V)/|L+V| \ .$$

It is clear that $I_1(\bar{L}.\hat{N})$ is the component due to Lambertian reflection, and $I_1(\bar{N}.\hat{H})^n$ is the component due to specular reflection. Note that the intensity of the specularly reflected light falls off rapidly as ϕ increases.

The shape-from-shading method described above can easily be extended to handle specular reflection. The observed intensity of a specular quadratic object surface can be approximated by

$$\begin{aligned} I(x,y) = I_1 & \frac{a(k_1+2k_3x+k_4y) + b(k_2+k_4x+2k_5y) - c}{[1 + (k_1+2k_3x+k_4y)^2 + (k_2+k_4x+2k_5y)^2]^{1/2}} \\ I_s & \left\{\frac{d(k_1+2k_3x+k_4y) + e(k_2+k_4x+2k_5y) - f}{[1 + (k_1+2k_3x+k_4y)^2 + (k_2+k_4x+2k_5y)^2]^{1/2}}\right\}^n \end{aligned}$$

where $L=(a,b,c)$ and $H=(d,e,f)$ are unit vectors. The same optimization procedure as described in section x can be used to determine surface orientations for specular surfaces by minimizing the sum of squares of the differences between the

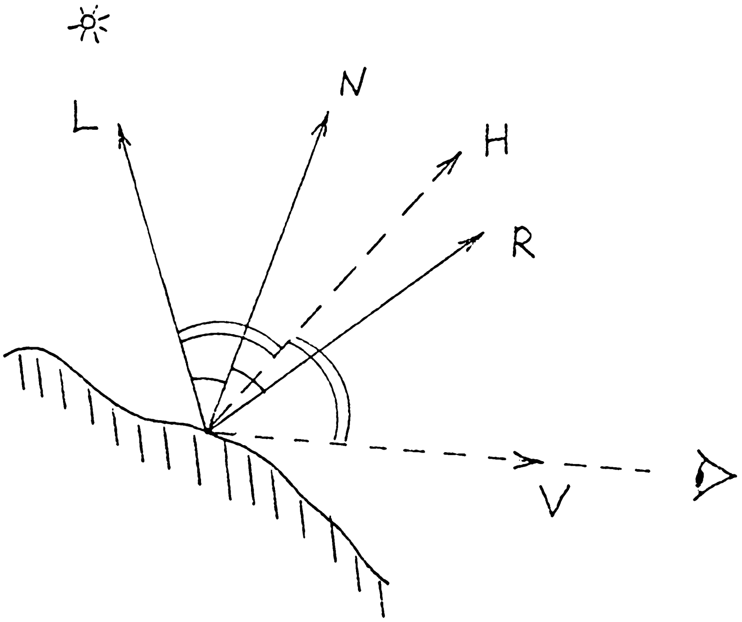


Figure 2: Relationship between θ and ϕ.

observed intensity data and the estimated intensity values computed from the equation above.

3.4 JUSTIFICATION

The methods presented here have some advantages over the previous shape-from-shading approaches. Since a neighborhood of pixels are considered when the surface orientation at a particular location is estimated, results show that this method is less sensitive to noise.

Furthermore, the above methods can produce more than just the surface normal at each pixel. For each neighborhood, it also produces the coefficients of the fitting polynomial. Using these coefficients permits us to determine whether or not neighboring pixels belong to the same surface patch.

3.5 RESULTS

The non-linear optimization procedure described in section III.1 has been tested on a 50x50 image of a simple surface. Figure 3 shows the surface generated by the polynomial: $40x+40y-2x^2-xy+2y^2$. The shaded image resulting from illuminating the surface of Figure 3 from light direction (0,0,-1), which means directly above the center of the surface, is shown in Figure 4. By inputting the starting value $\underline{k}_0=(1,1,1,1,1)$ or $\underline{k}_0=(40,40,-2,-1,2)$ to the optimization procedure, $\hat{\underline{k}}$ was found to be $(38.5,43.8,-2.0,-1.3,2.0)$, which is very close to the original surface within the range of (x,y) used. By using $\underline{k}_0=(-40,-40,2,1,-2)$, the estimated surface is shown in Figure 5. By using $\underline{k}_0=(0.86,-1.61,-0.03,0.04,-0.01)$, which was obtained by fitting a second degree polynomial to the intensity surface of the shaded image, the estimated surface is shown in Figure 6. By illuminating the three estimated surfaces from light direction $(0,0,-1)$, we obtained essentially the same shaded image. This shows the ambiguity in image shading information.

Figure 7 shows a pair of synthetic stereo images. A sparse depth map was determined by the stereo matching procedure described in Chapter 4. The depth map is illustrated in Figure 8. The depth constraints obtained from the stereo pair of images were used to determine an initial estimate of the object surface in the images. The singular value decomposition method was used to obtain such an initial estimate. By incorporating the depth constraints into the shape-from-shading algorithm, an accurate reconstructed surface was obtained (Figure 9). It is worth mentioning that multiple surface reconstructions are possible if boundary constraints are unavailable.

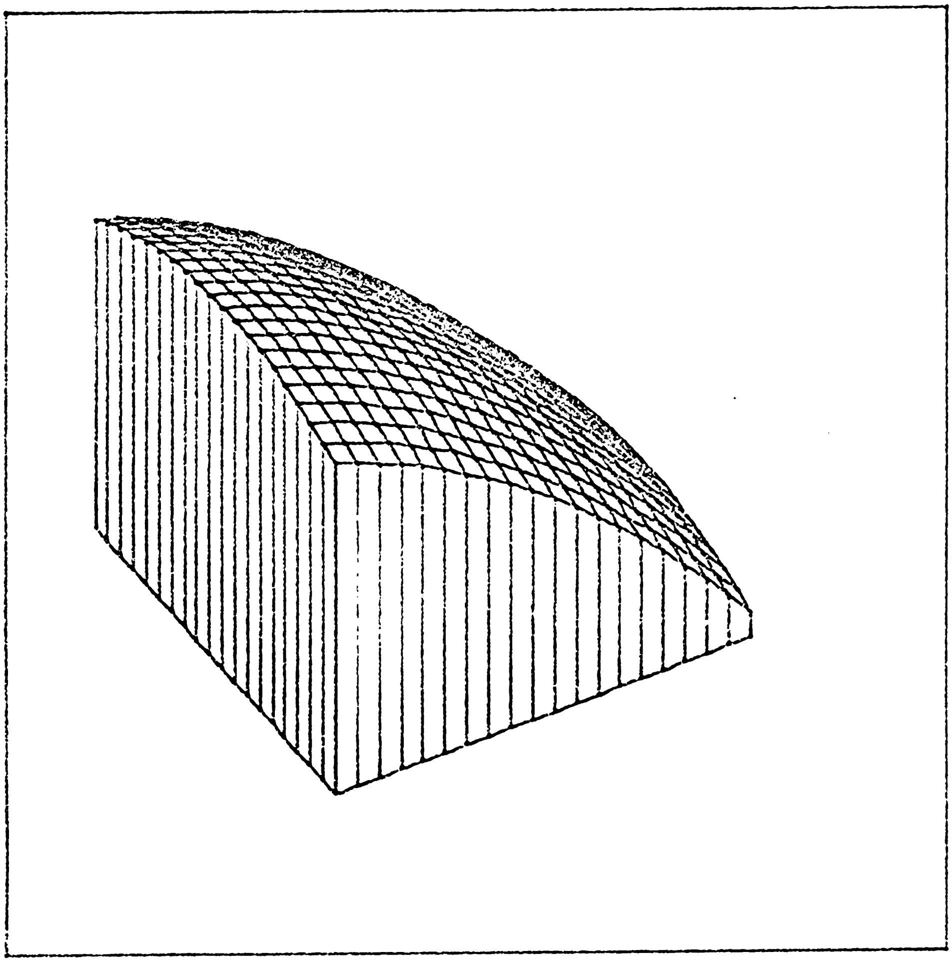


Figure 3: The polynomial surface: $40x+40y-2x^2-xy+2y^2$.

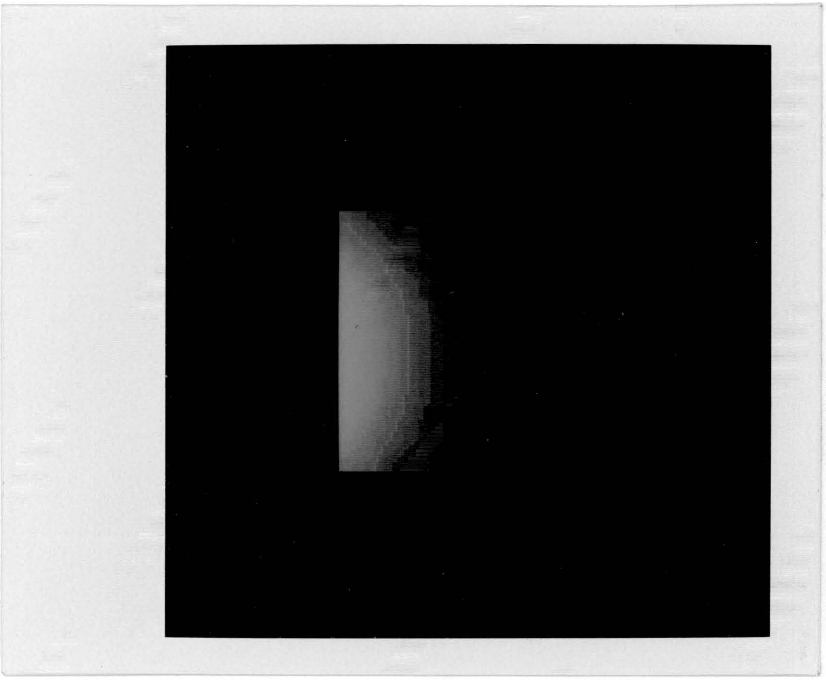


Figure 4: The shaded image of the surface of Figure 3.

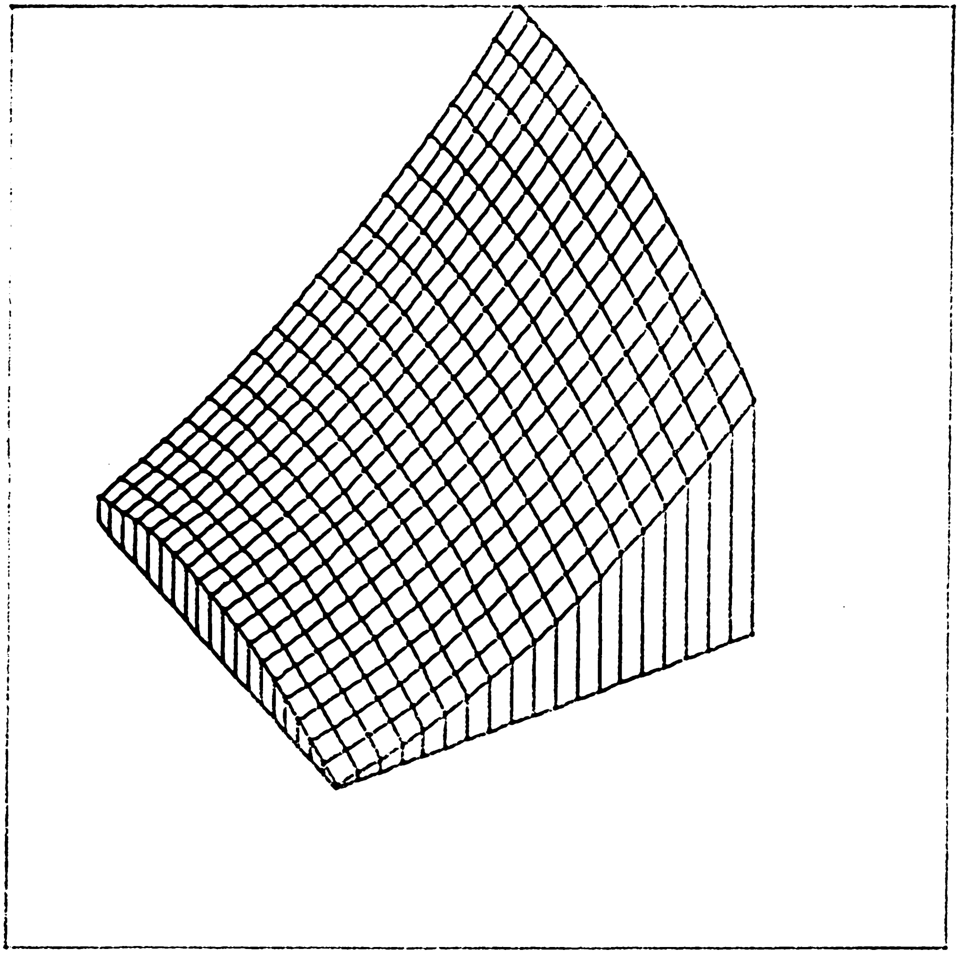


Figure 5: The reconstructed surface for the shaded image of Figure 4.

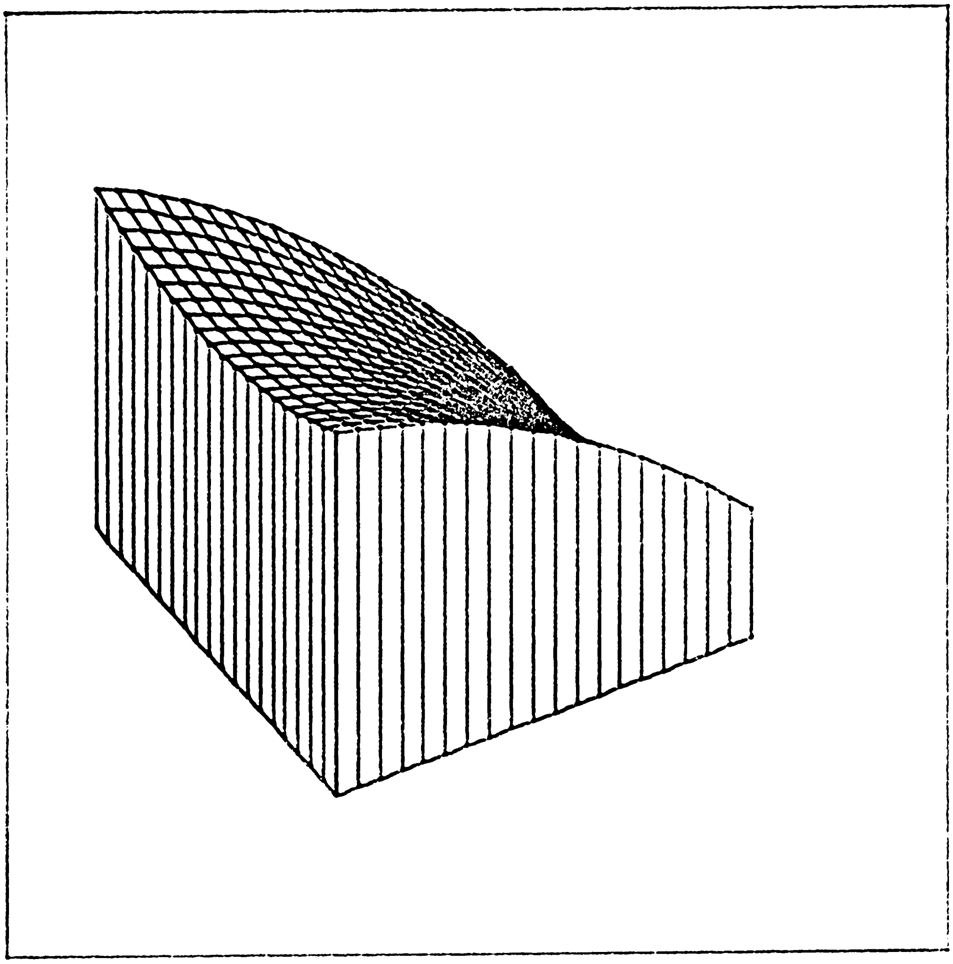


Figure 6: Another reconstructed surface for the shaded image of Figure 4.

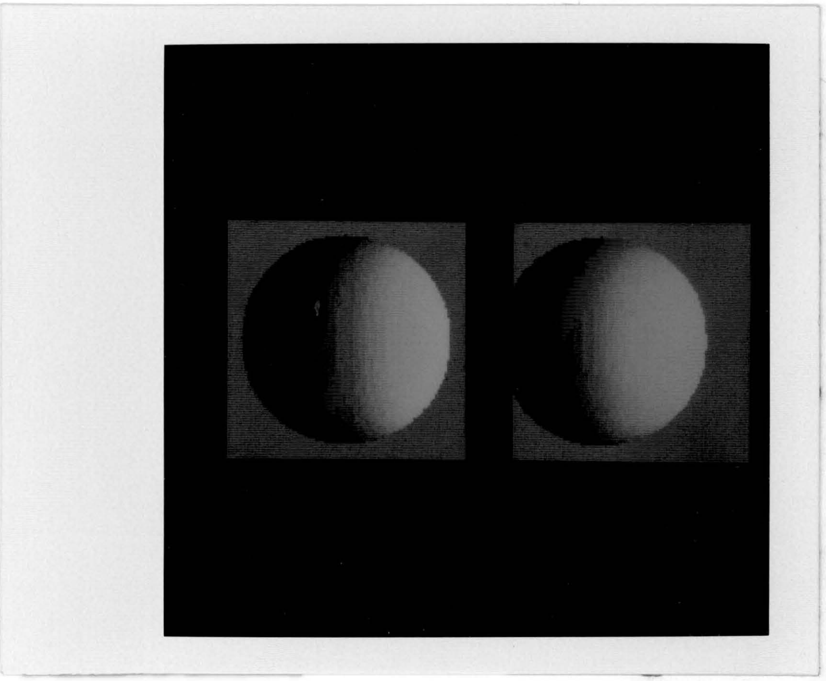


Figure 7: Synthetic stereo images of a spherical surface.

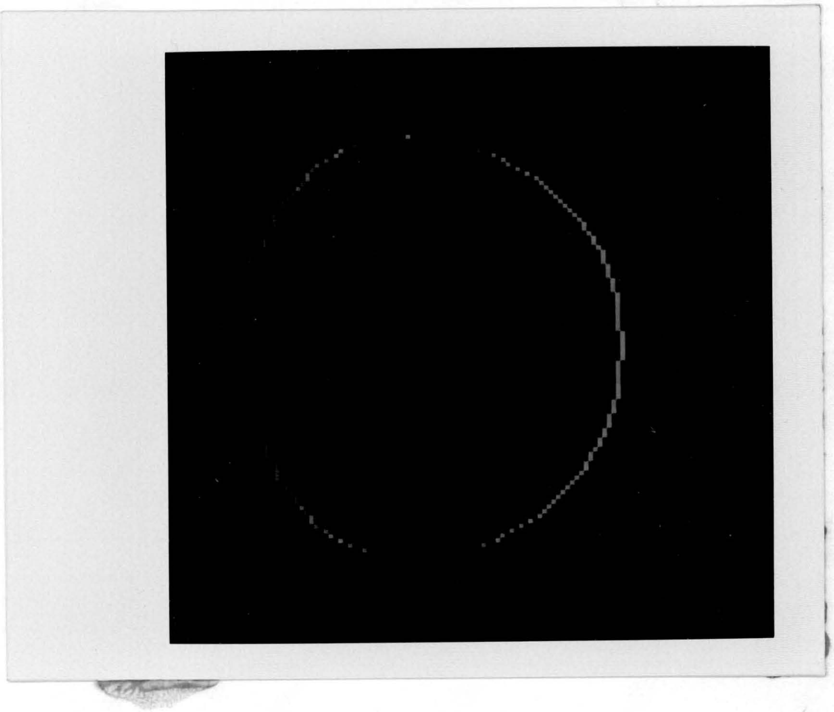


Figure 8: The depth map resulting from matching the stereo images of Figure 7.

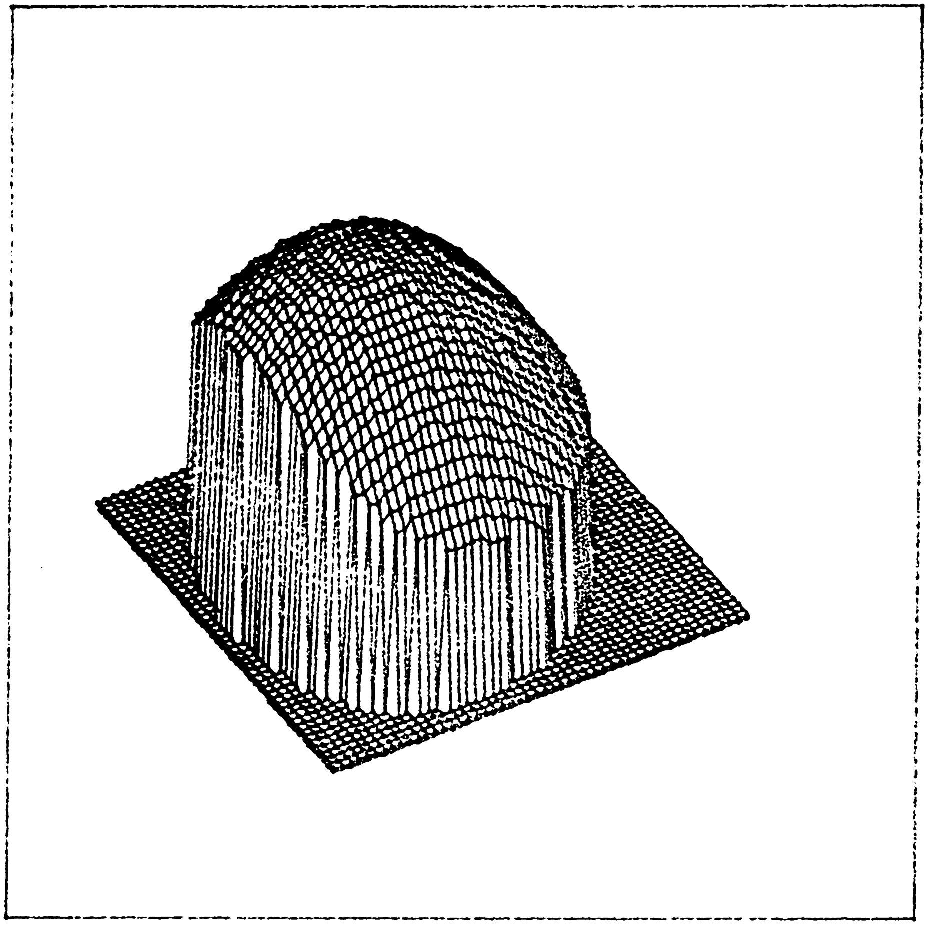


Figure 9: The reconstructed surface for object in the stereo images.

To investigate the effect of boundary constraints on the optimization process, the shape-from-shading algorithms with and without boundary constraints have been tested on the image of Figure 10a. One-hundred sets of randomly generated initial estimates were input to the methods described in Section 3.2.1 and Section 3.2.1.5. A set of boundary constraints was obtained along the object boundary. Results show that the method converged to three different but correct solutions for eighty-four out of the one hundred initial estimates if no boundary constraint was used. The method converged to a unique solution ninty-four times if the set of boundary constraints was included in the optimization process. We also found that the average rate of convergence was improved by a factor of about two if the boundary constraints were used. Similar results were obtained when these methods were run on the image of Figure 10b.

The photometric stereo method described in section 3.2.1 has been tested on both artificially generated images and real images. An object surface corresponding to a portion of a spherical surface is shown in Figure 11. Three shaded images were obtained by illuminating the surface of Figure 11 from three different positions: (1) the light direction is (0,0,-1) which means directly above the center of the

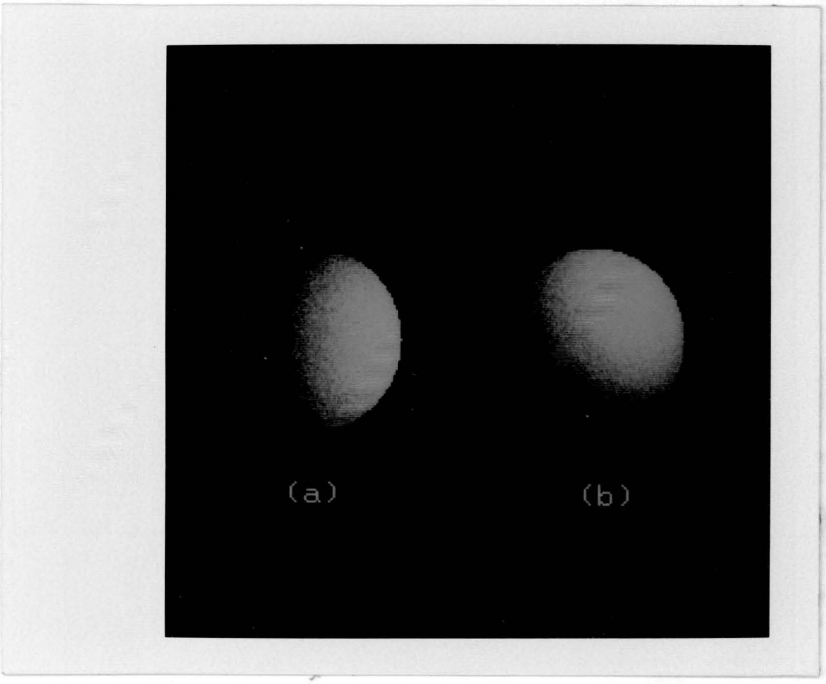


Figure 10: Shaded images of a curved surface.

surface; (2) the light direction is (0,0.259,-0.966) which translates to azimuth 90 degrees and elevation 75 degrees; (3) the light direction is (0.259,0,-0.966) which translates to azimuth 0 degrees and elevation 75 degrees. The shaded images are shown in the first row of Figure 12.

In spite of the fact that a spherical surface cannot be fit exactly by a second degree polynomial, the reconstructed surface by the photometric stereo method is extremely close to the original one. The root mean square difference between the elevation data of the original surface and the reconstructed surface is 1.20, which is .77 percent of the surface elevation range 100-255.

To examine the performance of our method on noisy images, random noise of mean=0 and standard deviation=10 was added to the shaded images of Figure 12. The noisy images are shown in the second row of Figure 12. The reconstructed surface by our method is shown in Figure 13. The RMS difference between the original surface and the reconstructed surface is 2.67. As a comparison, Woodham's method was implemented and was run on the same noisy images. The reconstructed surface by Woodham's method is shown in Figure 14. The RMS difference between the original surface and the reconstructed surface is 5.82. These results show that our method is less sensitive to noise.

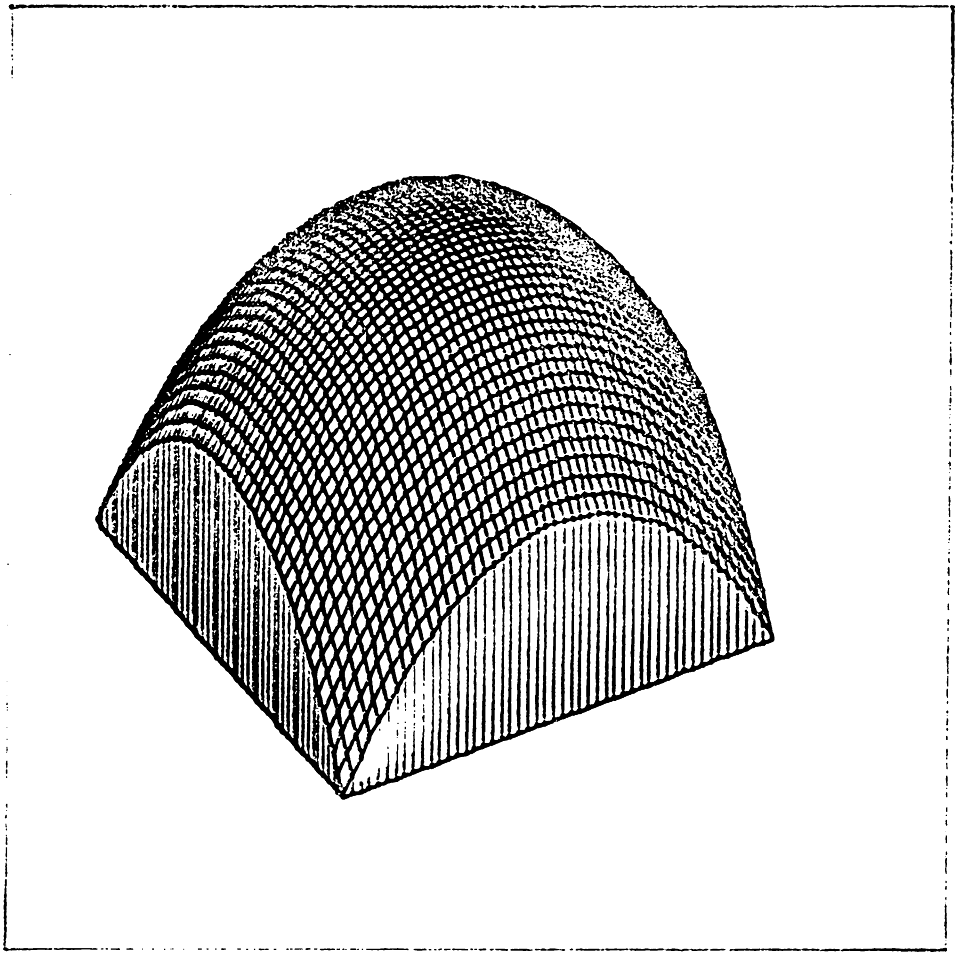


Figure 11: Portion of a spherical surface.

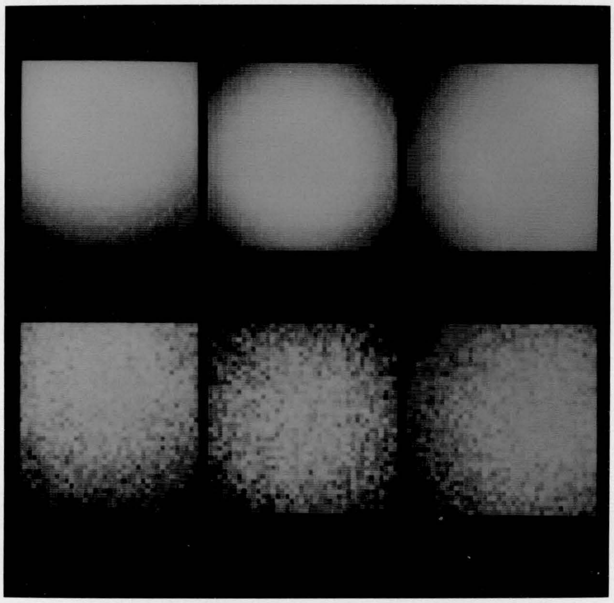


Figure 12: Shaded images and noisy images of the surface of Figure 11.

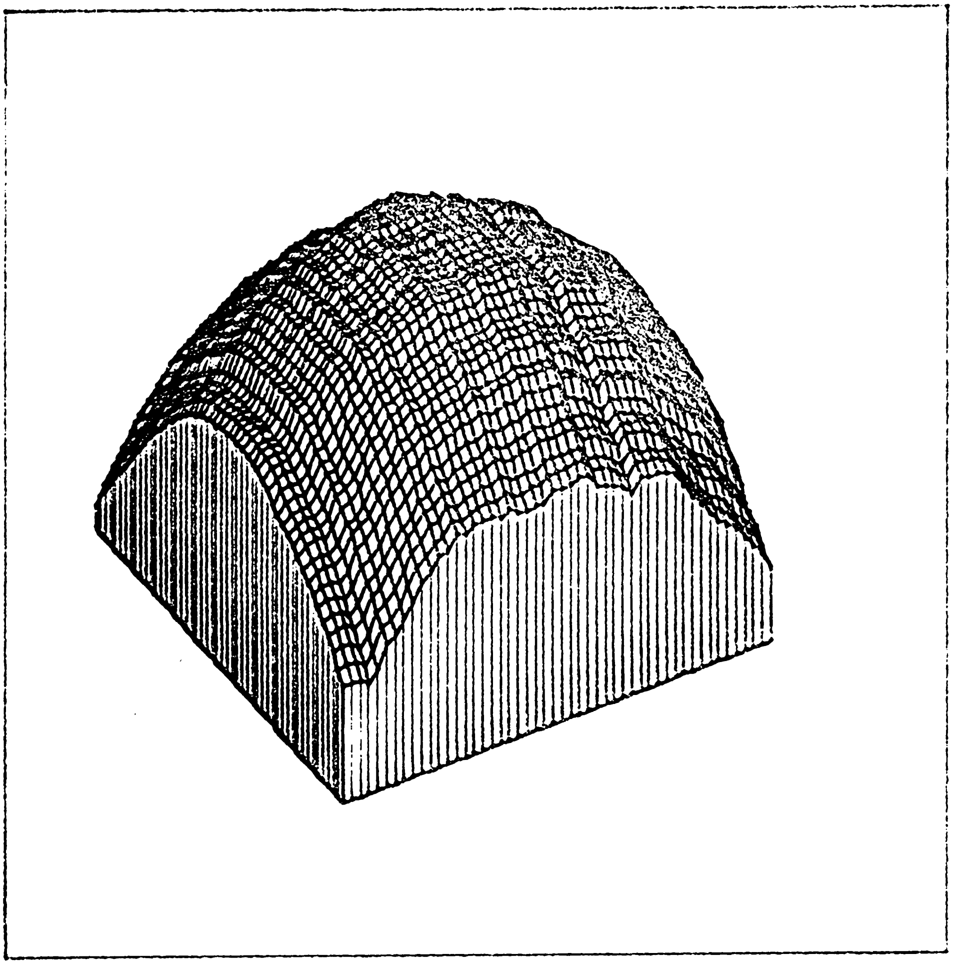


Figure 13: The reconstructed surface by the photometric stereo method.

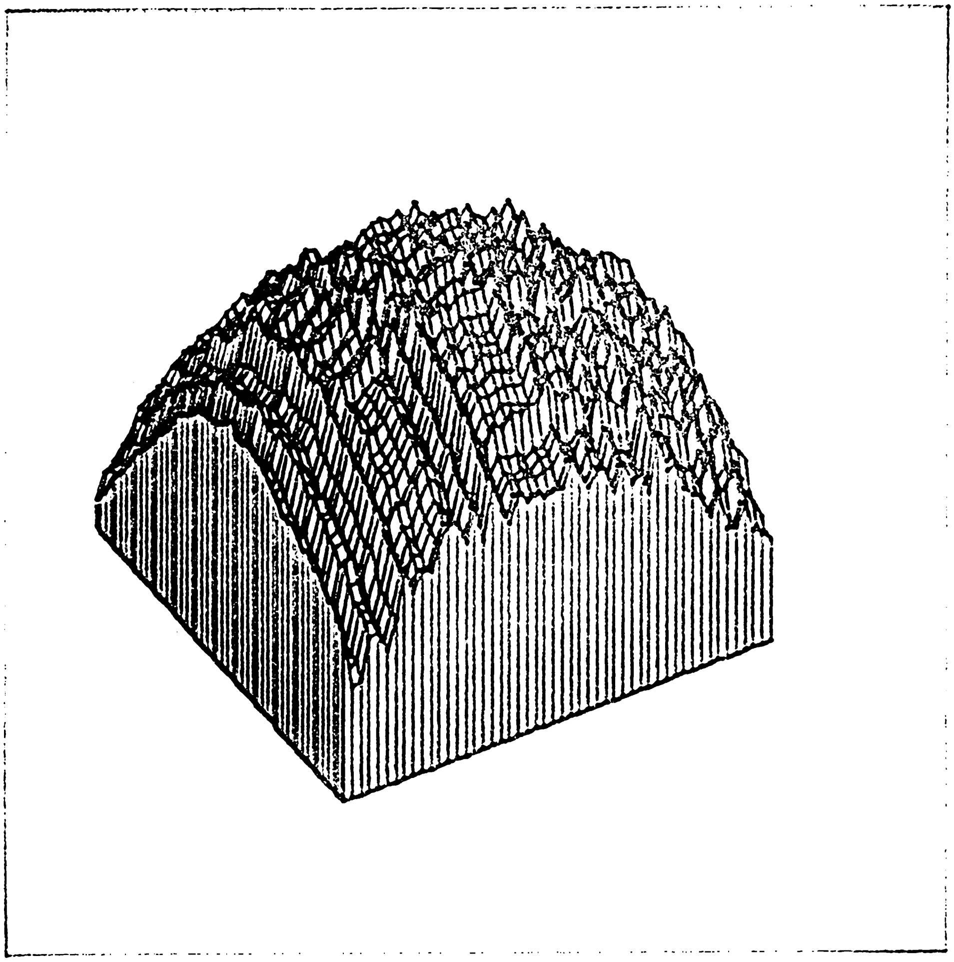


Figure 14: The reconstructed surface by Woodham's method.

Another experiment has been performed on the shaded images of an egg. The images were taken with the camera positioned directly above the egg and the light sources positioned at $(\theta,\phi)=(90^\circ,90^\circ)$, $(\theta,\phi)=(90^\circ,45^\circ)$, and $(\theta,\phi)=(0^\circ,45^\circ)$. The corresponding images are shown in Figure 15. The recovered surface is shown in Figure 16. The result is considered satisfactory because the light source used was not a point light and the illumination angles measured may differ from the true angles by as much as $\pm 10^\circ$.

In additional to the above experiments, images of a more complex object were also used in testing the photometric stereo method. Three shaded images were obtained by illuminating the object from $(\theta,\phi)=(90^\circ,90^\circ)$, $(\theta,\phi)=(90^\circ,60^\circ)$, and $(\theta,\phi)=(0^\circ,60^\circ)$. The shaded images are shown in Figure 17. Since the object is composed of different types of surfaces, the smoothness constraint was not expected to hold at locations where these surfaces met. Thus, an edge operator was first used to detect these discontinuities, and the shape-from-shading method was only applied to locations where no discontinuity was detected. The reconstructed surface obtained by using the photometric stereo method with a 5x5 window is shown in Figure 18.

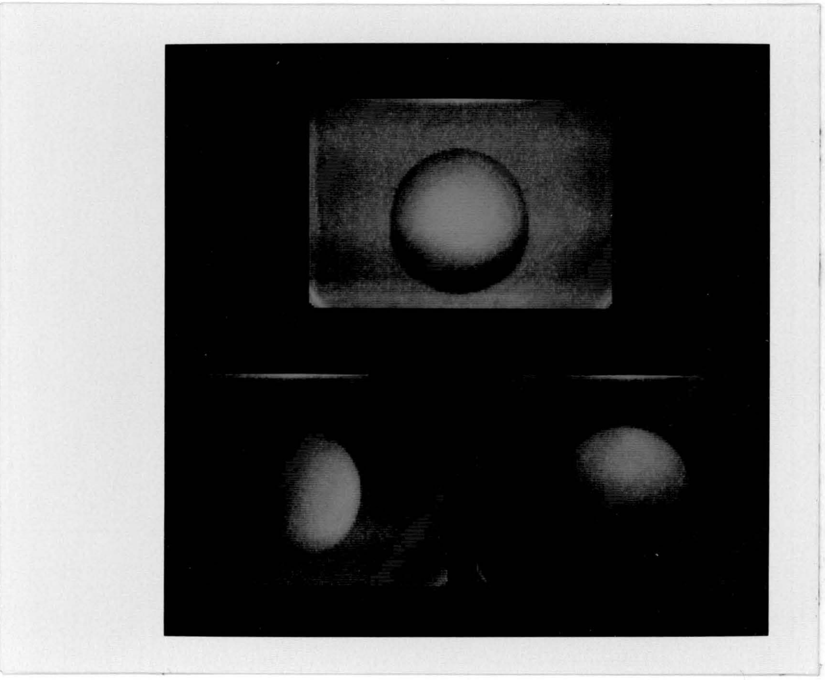


Figure 15: Shaded images of an egg.

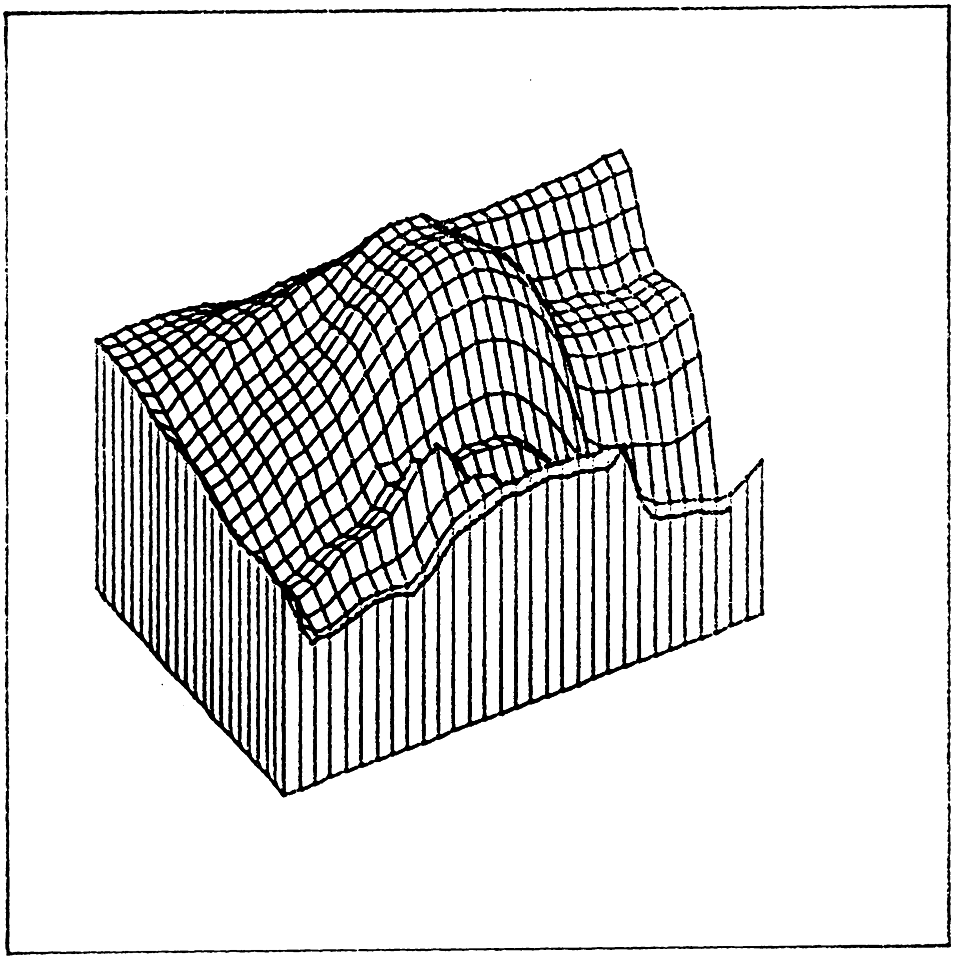


Figure 16: The reconstructed surface from the images of Figure 15.

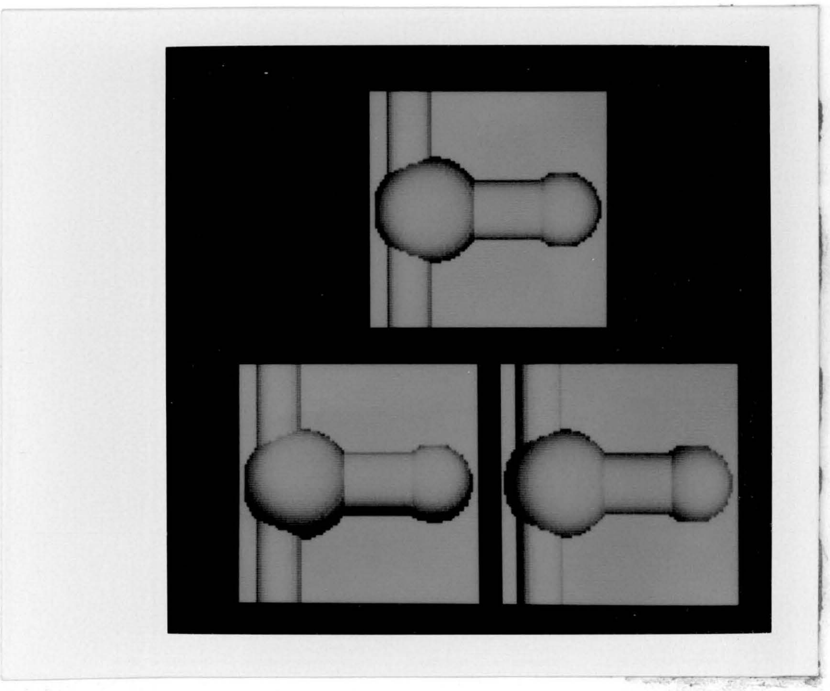


Figure 17: Shaded images of a complex object.

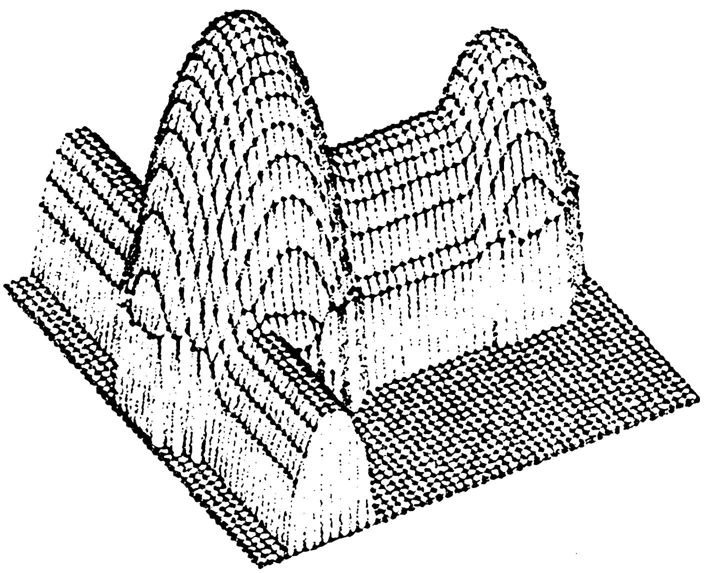


Figure 18: The reconstructed surface from the images of Figure 18.

3.6 SUMMARY

Two methods to recover 3-D surface orientation from image shading information have been presented. The first method works in the single image domain and is posted as a non-linear optimization procedure. Experiments have been performed on synthetic and real images and the results are found to be satisfactory. Since inherent ambiguity exists in the image, the optimization procedure may lead to undesirable or multiple solutions. By using additional boundary constraints, good initial estimates are obtained for the optimization process and the performance of the method is improved.

By using additional images as constraints, the second method linearizes the optimization process. Unique and accurate solutions have been obtained by experimenting with the second method on simple images. So far, the proposed methods have been tested only on images which are obtained from orthographic projection. Using perspective projection, a possible extension of the shape-from-shading method is outlined in appendix A. In order for these methods to handle perspective projection and more complex images, more work needs to be done in the future.

Chapter IV

STEREO MATCHING

A major industrial application of computer vision systems is in automatic inspection. One approach to such an inspection task is a coordinated system that employs robot arms with measurement devices for inspection along with a stereo vision system that provides information to a planner that guides the robot arms [Sh84]. The functions of the vision system are to determine the position of the objects in the scene in relation to the camera(s) and to determine the portion of object which the camera is viewing. The first problem can be solved by finding correspondences between features extracted from the stereo pair of images. The second problem can be approached by matching structures extracted from the images to parts in a three-dimension description of the object to be inspected. One such object model can be found in [Sh83] which employs a hierarchical relational model to describe a complex F-15 bulkhead.

A stereo approach to automatic inspection must solve the following problems. First, how are structures in the images to be extracted? In particular, we are interested in extracting structures which are useful for finding correspondences between the images of the stereo pair and for matching structures in the three-dimensional object model. To extract structures from an image usually involves some kind of segmentation procedure. Unfortunately, none of the existing segmentation technique seems to produce satisfactory results. In the first part of this chapter, we will describe a scheme for extracting line-like and surface-like structures which are used frequently in modeling three-dimensional objects [Sh83]. This scheme is essentially a combination of many of the existing segmentation techniques. Experiments show that this scheme produces improved results over the existing techniques.

The second problem is to find correspondences between structures in the stereo pair. Most existing stereo matching techniques use local features as primitives for matching. Although continuity constraints are imposed in some systems to resolve ambiguity, it seems that the use of local feature is far from adequate. In the second part of this chapter, we will propose a stereo matching technique based on matching higher-level primitives such as arc segments, region segments and topographic structures. The problem of matching extracted structures to a stored model is beyond the scope of this work.

4.1 FEATURE EXTRACTION

The goal of early visual processing is to extract a rich symbolic representation of the gray tone intensity changes in an image. Marr [Ma76] names this representation primal sketch. It is important for the primal sketch to capture all gray tone intensity variations in an image because any intensity changes can be related to changes in scene characteristics such as surface orientation, surface discontinuity, surface reflectance and illumination. Marr's primal sketch relies solely on a special form of intensity changes which are detected as zero-crossings of the Laplacian of a Gaussian filtered image [Ma80]. Unfortunately, the use of edges alone can not provide a rich enough description for intensity changes, since it fails to account for smooth intensity changes or shadings which frequently appear in images of curved surfaces.

Haralick [Ha83] proposed a rich and robust representation for all types of two-dimensional intensity variations. This representation is called the topographic primal sketch. We will describe here a feature extraction scheme which extracts topographic structures such as edges, ridges, valleys and hillsides. Notice that edge is included in the set of topographic structures. Edges, which usually correspond to sharp changes in gray tone intensities, are used to describe the basic structure of an image. The rest of the topographic structures are used to describe the intensity variations within regions which are extracted as connected sets of non-edge structures. In the first part of this section, a segmentation scheme is used to construct edge structures and regions (or non-edge structures). The second part of this section will give a detailed discussion of the topographic labeling process.

4.1.1 Edge and Non-edge structures

Many segmentation schemes have been studied through two main approaches: edge analysis and region analysis. Unfortunately, edge operators are usually too sensitive to noise while region growers usually grow too far and do not always produce good region-object boundary correspondance. In this section, an edge-based segmentation procedure is described. We choose to use an edge-based segmentation scheme because structures extracted here will have important impacts on the outcome of the stereo matching which is to be done in a later stage. It is generally agreed ([Ba81],[Me83]) that edge-based stereo matching gives more accurate results, therefore, it is important to position edges at more precise locations.

The segmentation process can basically be outlined as follows:

1. detect raw edge elements by a local edge operator;
2. extract arc segments by linking fragmented edges and cleaning noisy edges; and
3. assemble non-edge pixels into regions by a connected components algorithm.

4.1.1.1 Extracting Edge Elements

Edge detection was pioneered by Roberts [Ro65]. A comprehensive survey of edge detection can be found in [Da75]. A set of edge detection criteria that capture the desirable properties of an edge operator is formulated in [Ca83]. A comparison of the various edge operators is given by Haralick [Ha84]. Based on the results of [Ha84], Haralick's second directional derivative edge operator was selected for our initial edge detection process. A precise mathematical definition of the operator can be found in [Ha84].

The idea of the facet model ([Ha80] and [Ha81]) is used in Haralick's edge operator. The facet model is a model for image data. It assumes that image intensity values are noisy sampled observations of an underlying intensity surface. Thus, any interpretation made on the basis of a neighborhood of pixel values should be understood through the analysis of its underlying intensity surface. To determine whether or not a pixel should be marked as an edge pixel, its underlying intensity surface must be estimated on the basis of the pixel values in its neighborhood. For this, a least squares fit with a functional form consisting of a linear combination of the tensor products of discrete orthogonal polynomials is used. The highest order tensor product we currently use is cubic. The required directional derivatives are easily computed for this class of function.

Pixels which are part of regions have simple gray tone intensity surfaces over their areas. Pixels which have an edge in them have complex gray tone intensity surfaces over their areas. Specifically, an edge occurs in a pixel if there is some point in the pixel's area having a zero crossing of the second directional derivative taken in the direction of a non-zero gradient at the pixel's center. The results of the second directional derivative edge operator applied to the images of Figure 19 are illustrated in Figure 20.

Although the second directional derivative zero-crossing edge operator is less sensitive to noise than the classical

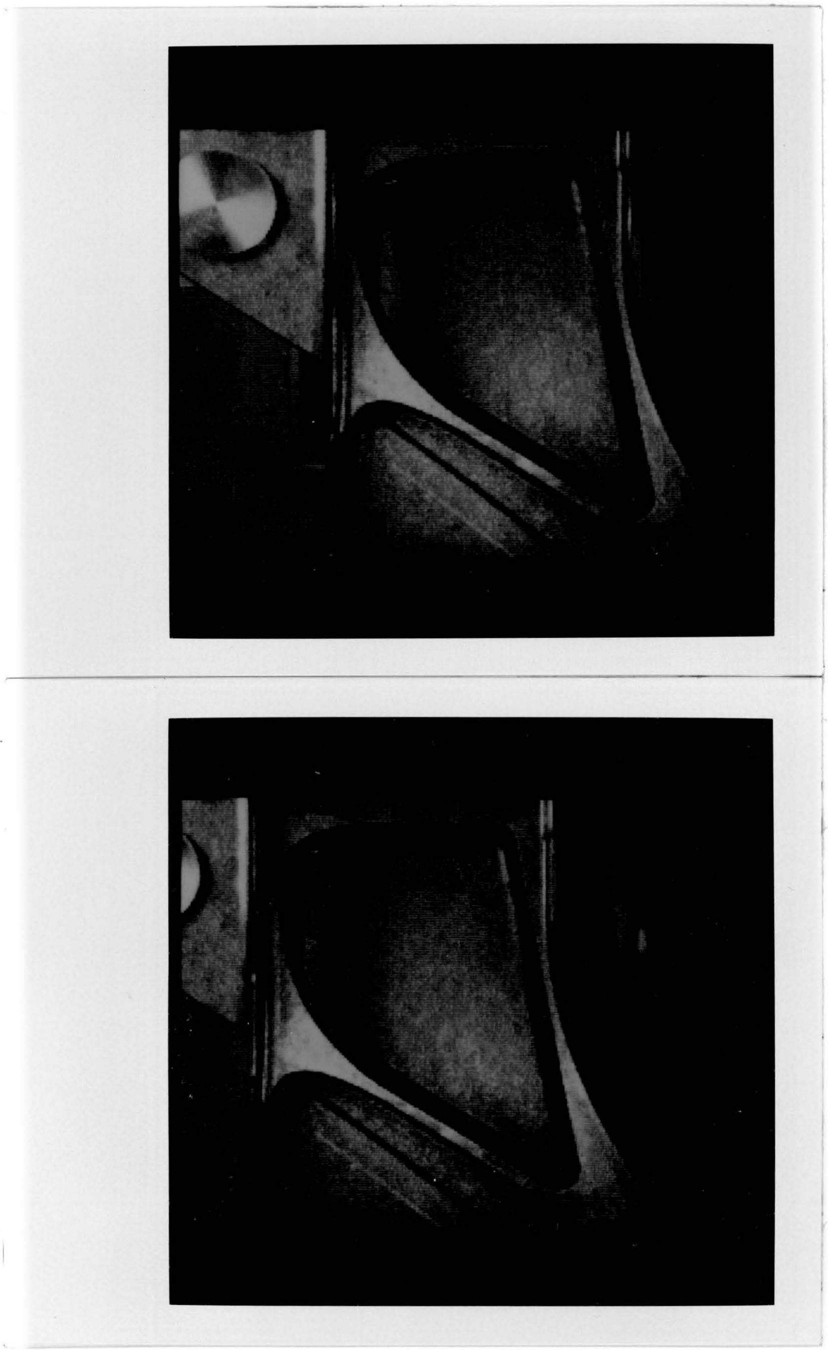


Figure 19: Stereo images of a F-15 bulkhead.

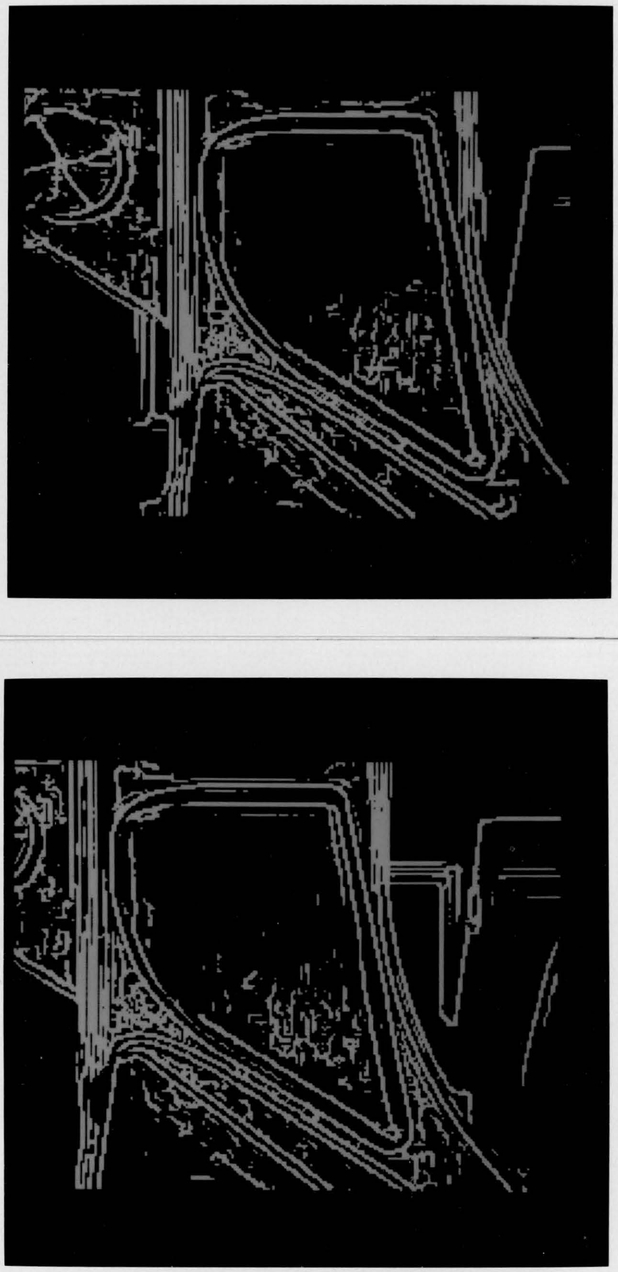


Figure 20: Edges detected from the images of Figure 19.

gradient edge operators, it still produces undesired and fragmented edges for a complex and noisy image such as the one shown in Figure 19. Moreover, industrial objects often consist of shiny metallic parts. Specular reflection creates further problem in edge detection. In order to obtain more meaningful edge structures, a procedure is proposed for extracting arc segments by linking fragmented edge segments and cleaning noisy edges.

4.1.1.2 Extracting Arc Segments

Edge cleaning and edge linking are two difficult problems which may lead to conflicting consequences. Edge cleaning tends to create wider gaps between already broken edges. Linking noisy edges often results in more undesired edges. An attempt is made here to give an even-handed treatment of both. The proposed technique is based on the observation that true edges are usually detected as sequences of orderly oriented pixels, while noisy edges are usually oriented in some random manner. The idea behind this scheme is to first link edge pixels into arc segments based on their detected orientations; isolated or short segments will then be deleted and a final gap filling scheme is used to create closed boundaries.

The edge linking process attempts to link together edge pixels which are detected as edges with similar orientations. This edge linking process is in some sense similar to Burns et al's [Bu84] method for extracting straight lines. Our approach is more powerful than theirs in the context that our method extracts arc segments other than straight lines. Instead of estimating edge orientation by a 2x2 mask as in [Bu84], we compute edge orientation directly from the facet fitting coefficients. Specifically, edge orientation is defined to be the direction perpendicular to the direction which extremizes the first directional derivative. A region growing scheme is employed for the edge linking process. This region growing scheme differs from the traditional region growing scheme in that the similarity measure is derived directly from edge orientation instead of from gray level intensity and that the resulting regions are long-thin edge regions.

An image is scanned left-to-right and top-to-bottom. Each edge pixel's orientation is compared to the mean orientation of all its neighboring edge segments. An edge segment is selected such that its mean orientation is closest to the edge pixel's orientation. The edge pixel is merged into the edge segment provided that their orientations are not too different. If no merging is possible, the edge pixel starts a new segment.

One way to detect closeness between edges is to compute the difference[1] between the two orientations. A more sophisticated approach is to employ statistical testing. Let $\bar{X}$ and S^2 be the mean and variance of the orientation of an edge pixel with n pixels and x be the orientation of the pixel to be tested. By assuming that the orientations of the n pixels in the edge segment and the pixel to be tested are independent and identically distributed normal, the statistic

$$T = \frac{(\bar{X}-x)}{s} \left[\frac{(n-1)n}{(n+1)}\right]^{1/2}$$

has a T_{n-1} distribution. The pixel is added to the segment if T is smaller than some predetermined value.

After a single scan through the image, a set of initial segments is created. In order to produce more meaningful and longer arc segments, a second merging processing is performed. In this step, mean orientations of adjacent edge segments are compared. Adjacent edge segments are merged if their mean orientations are not too different. A similar T-statistic can be constructed and tested for the mean orientations of two edge segments.

---

[1] It is worth noting that sums and differences of edge orientations should be done in modulo 360 arithmetic.

Figure 21 shows the resulting arc segments of the bulkhead image of Figure 19. Figure 22 shows the cleaned edge image by eliminating arc segments with size less than 10 pixels. The results shown in Figure 22 are pleasing in the sense that our technique not only removes most of the undesired edges, but also groups edges into meaningful arc segments which are useful for higher level matching.

One undesirable result of removing all short segments is that it disconnects edges at locations where there are sharp changes in orientation, for example, at corners. This problem can be fixed easily by joining nearby end points of neighboring arc segments by straight lines. A more effective way to link fragmented arc segments is to first extend their end points along pixels which were originally detected as edges by the edge operator, before nearby end points are joined by straight lines.

To determine whether an edge pixel is an end point of an arc segment, the mean orientation of the arc segment to which the pixel belongs is quantized into either vertical, horizontal or one of the two diagonal directions. Consider the following configuration for the neighborhood of an edge pixel e,

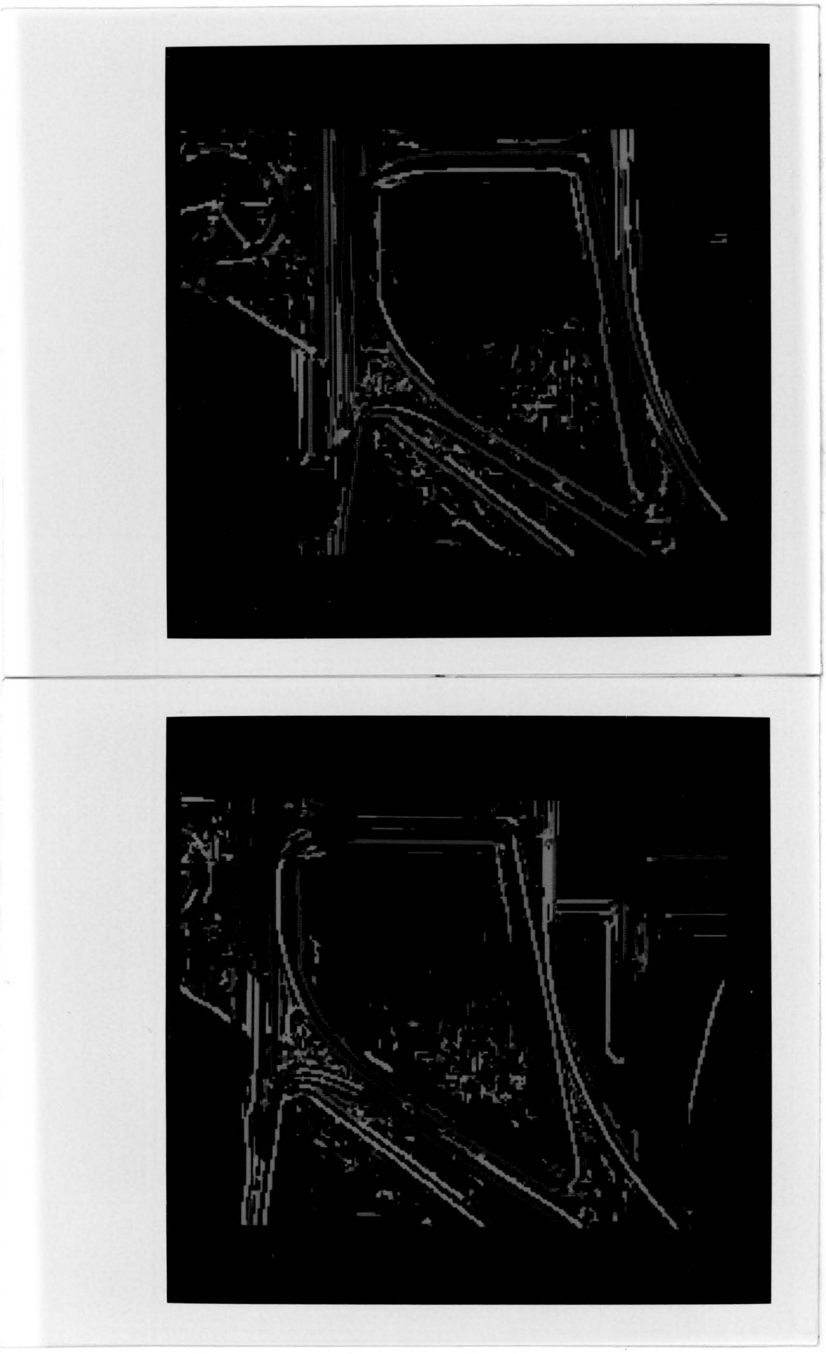


Figure 21: Edge segments of the images of Figure 19.

Figure 22: The cleaned edge images resulted from the images of Figure 21.

a	b	c
d	e	f
g	h	i

we define the two <u>directed neighborhoods</u> of the edge pixel e with orientation θ to be the sets

[{a,b,c} , {g,h,i}] if θ is vertical,
[{a,d,g} , {c,f,i}] if θ is horizontal,
[{a,b,d} , {f,h,i}] if θ is NW diagonal,
and [{b,c,f] , {d,g,h}] if θ is NE diagonal.

An edge pixel is defined to be an end point if none of the members in either one of its directed neighborhoods is an edge pixel.

The end points detected from the arc segments of the images of Figure 22 are shown in Figure 23. The results of the final gap filling process are illustrated in Figure 24. Our method has the advantage over Perkin's expansion and contraction [Pe79] method in that our method does not displace or thicken edges since the end points of arc segments are the only locations where activities are taking place.

Figure 23: End points of the edge segments of Figure 22.

Figure 24: Closed edge segments of the edge images of Figure 23.

4.1.1.3 Extracting Region Segments

Once a final set of arc segments is determined, the regions extracted are the largest connected areas of pixels which are entirely surrounded by arc segments. This process is accomplished by a connected components algorithm which assigns a unique label to each maximally connected group of non-edge pixels. An efficient memory-limited connected components algorithm can be found in [Lu83]. This algorithm requires only one top-down scan and one bottom-up scan of the entire image. The non-edge pixels are labelled as regions, and the arc segments which exist between regions remain unaffected. This region extracting scheme can sometimes create small regions which may not contain enough information for meaningful matching, therefore, regions of sizes less than a certain threshold are usually eliminated. The region segments of the images of Figure 19 are shown in Figure 25.

Figure 25: Region segments extracted from the images of Figure 19.

4.1.2 Extracting Topographic Structures

Multiple images of the same scene can be obtained under different illumination conditions and with a variety of camera gain settings. Such photometric variations can create problems if raw intensity values are used for stereo matching. We are interested in describing intensity variations within region segments in a way which is insensitive to photometric variations. The major categories {peak, pit, ridge, valley, saddle, hillside and flat} of the topographic primal sketch [Ha83] have been proven to be invariance under monotonic gray tone transformations. Although its subcategories {edge, slope, convex hill, concave hill and saddle hill} may change under gray tone transformation, to the extent that we deal only with slight alterations in imaging condition (which is usually the case for stereo vision), it is reasonable to assume that such changes are tolerable.

The topographic primal sketch can be used to represent the underlying intensity surface of a digital image. A complete mathematical treatment of the topographic primal sketch is given in [Ha83]. We will summarize here the concepts necessary to understand the remainder of the paper. A digital image may be interpreted as a sampling and quantizing of a real-valued function f. While the image is a discrete matrix of values, the underlying surface is continuous. Since the underlying surface is continuous, we can work with such well-defined concepts as its gradient magnitude and its first and second directional derivatives.

The topographic labeling scheme is based on the estimation of the values of the gradients and the directional derivatives of the surface. In order to obtain these values, we need first to assume some kind of parametric form for the underlying function f. If we assume that the neighborhood around each pixel is suitably fit by the bivariate cubic:

$$f(x,y) = k_1 + k_2x + k_3y + k_4x^2 + k_5xy + k_6y^2$$
$$k_7x^3 + k_8x^2y + k_9xy^2 + k_{10}y^3 ,$$

then the parameters k_1 through k_{10} can be estimated by a least squares fit to each pixel in the neighborhood. Once these parameters are estimated, the gradient vector (∇f) is given by (df/dx, df/dy) and its magnitude ($||\nabla f||$) is

$$[\ (df/dx)^2 + (df/dy)^2\ ]^{1/2} .$$

The first and second directional derivatives may be calculate by forming the Hessian matrix

$$H = \begin{vmatrix} d^2f/dx^2 & d^2f/dxdy \\ & \\ d^2f/dxdy & d^2f/dy^2 \end{vmatrix} .$$

The gradient magnitude and the directional derivatives obtained from the Hessian are used in determining the topographic labeling of the surface. The mathematical properties of the topographic structures defined in [Ha83] are summarized in Table 1. Each entry in the table is either 0, *, + or -. The 0 means an entry is not significantly different from zero, * means an entry does not matter, and +(-) means an entry is significantly different from zero on the positive (negative) side. If the values of the gradient magnitude, eigenvalues and eigenvectors of the Hessian at a given point of the surface satisfy the constraints defined by one row of Table 1, then that point is classified by the label at the end of that row. It is important to note that any combination of $||\nabla f||$, X1, X2, w1 and w2 corresponds to one and only one row of Table 1. Figures 26 - 31 illustrate some of the topographic features extracted from the images of Figure 19.

In summary, the feature extraction procedure is divided into six steps:

(1) Detect edge elements by using the second directional derivative zero-crossing edge operator.

(2) Form arc segments by linking edges of similar orientations.

(3) Clear edges by deleting isolated short arc segments.

(4) Fill gaps between arc segments.

(5) Extract regions by forming maximum connected sets of non-edge pixels.

(6) Compute topographic structures within regions.

Table 1: Mathematical Properties of Topographic Structures

$\|\nabla f\|$	X_1	X_2	$\nabla f \cdot w^{(1)}$	$\nabla f \cdot w^{(2)}$	Label
0	-	-	0	0	Peak
0	-	0	0	0	Ridge
0	-	+	0	0	Saddle
0	0	0	0	0	Flat
0	+	-	0	0	Saddle
0	+	0	0	0	Ravine
0	+	+	0	0	Pit
+	-	-	-,+	-,+	Hillside
+	-	*	0	*	Ridge
+	*	-	*	0	Ridge
+	-	0	-,+	*	Hillside
+	-	+	-,+	-,+	Hillside
+	0	0	*	*	Hillside
+	+	-	-,+	-,+	Hillside
+	+	0	-.+	*	Hillside
+	+	*	0	*	Ravine
+	*	+	*	0	Ravine
+	+	+	-,+	-,+	Hillside
+	*	*	0	0	Impossible

where

∇f = gradient vector of the graytone intensity function f

$\|\nabla f\|$ = gradient magnitude

H = Hessian matrix of f.

$w^{(1)}$ = direction in which the second directional derivative has greatest magnitude (first eigenvector of H).

$w^{(2)}$ = direction orthogonal to $w^{(1)}$ (second eigenvector of H).

X_1 = value of the second directional derivative in the direction $w^{(1)}$ (first eigenvalue of H).

X_2 = value of the second directional derivative in the direction $w^{(2)}$ (second eigenvalue of H).

$\nabla f.w^{(1)}$ = value of the first directional derivative in the direction of $w^{(1)}$.

$\nabla f.w^{(2)}$ = value of the first directional derivative in the direction of $w^{(2)}$.

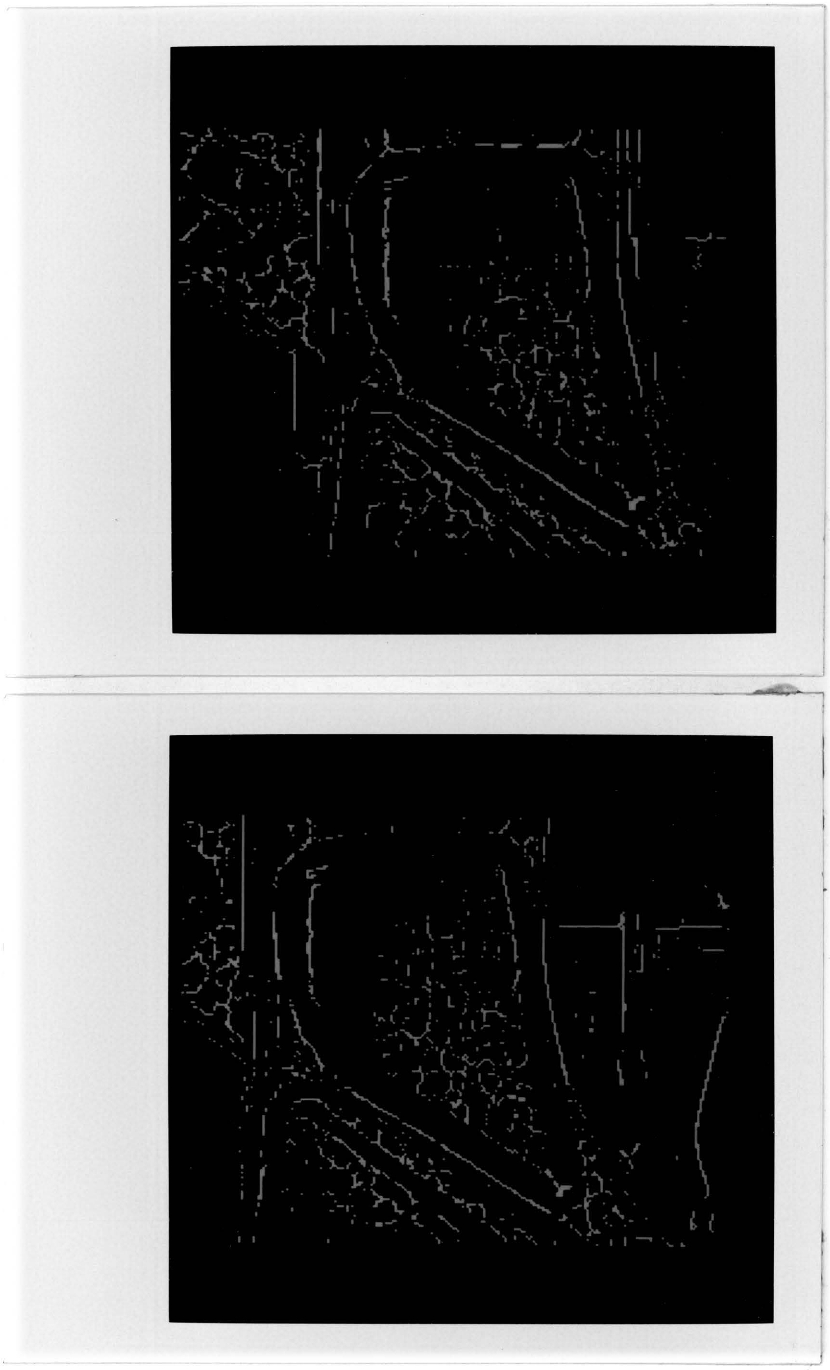


Figure 26: Ridge labels in the images of Figure 19.

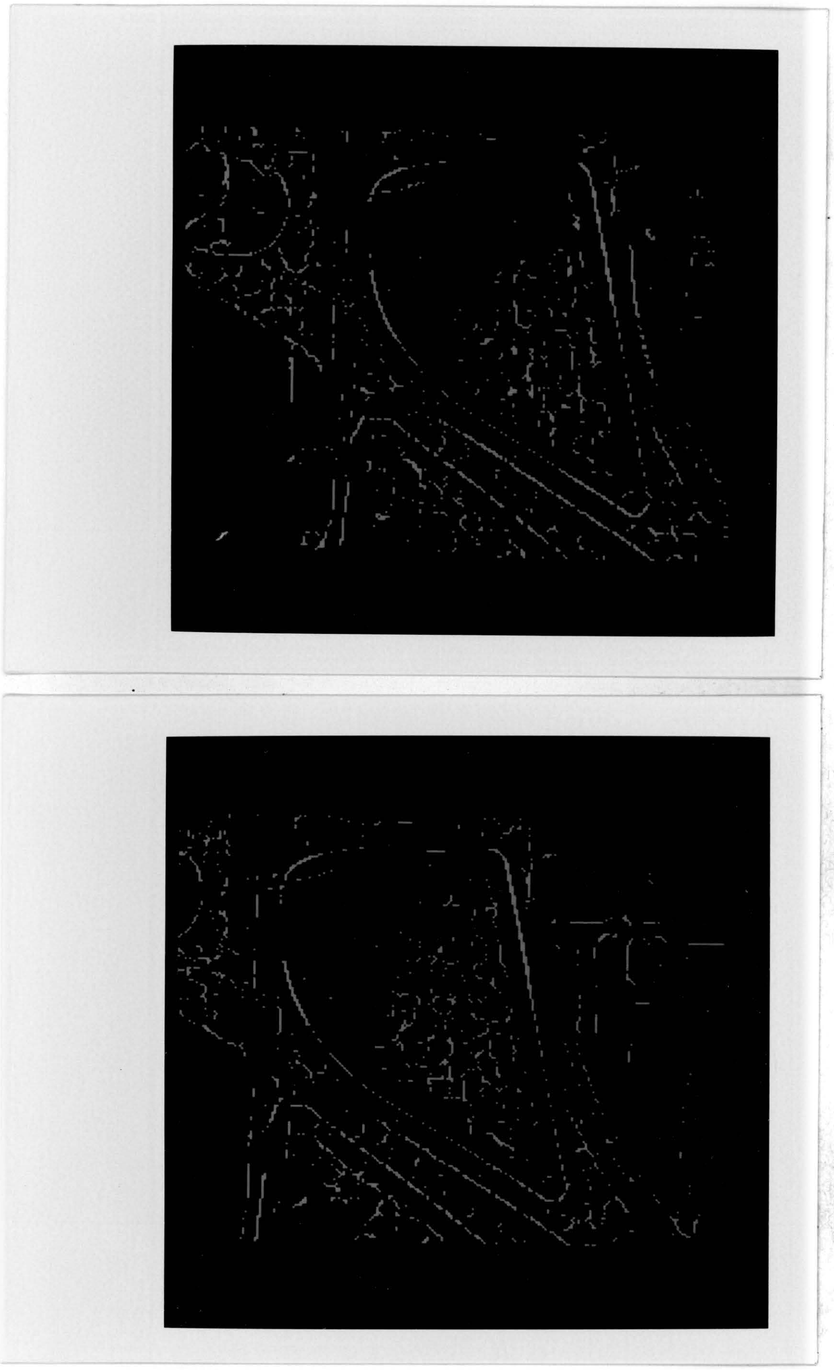


Figure 27: Valley labels in the images of Figure 19.

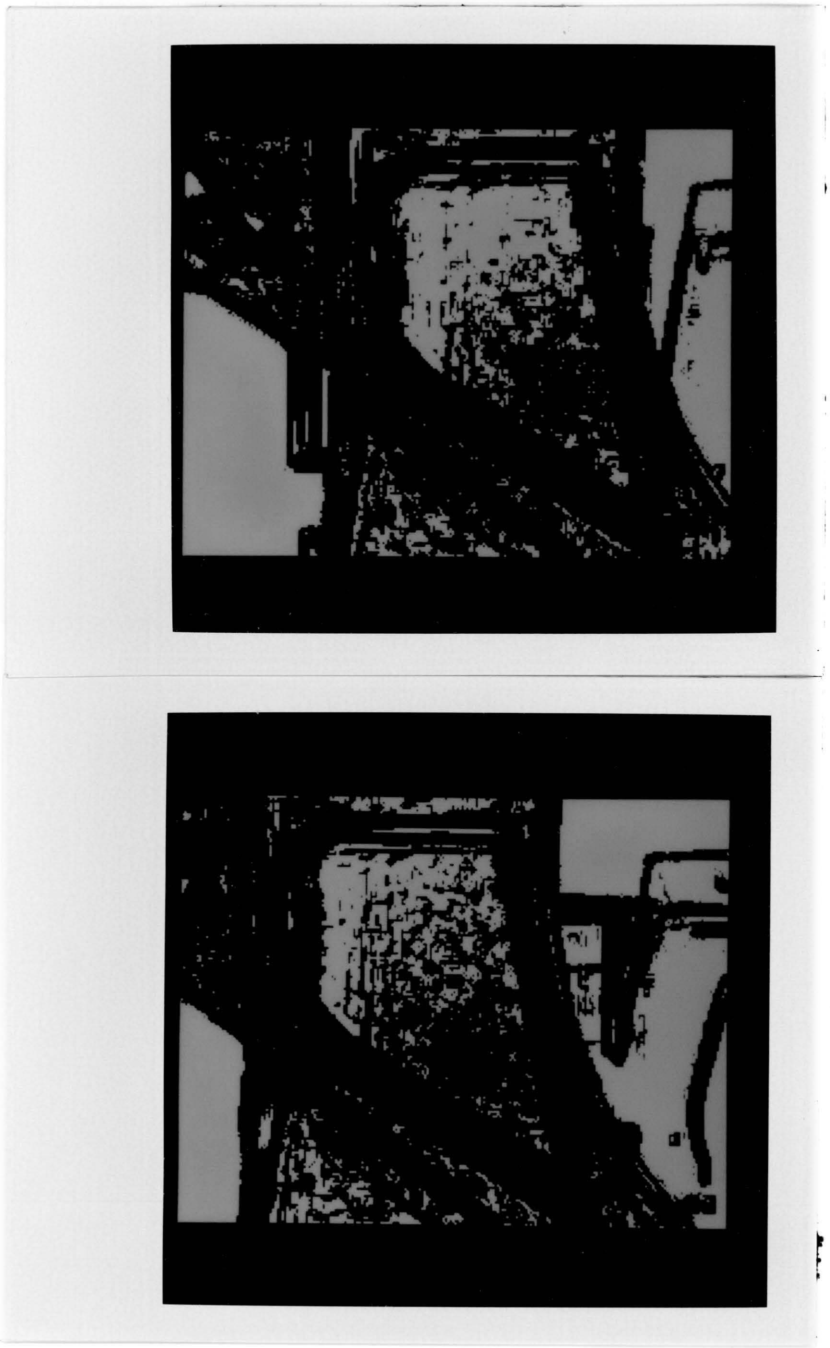


Figure 28: Flat labels in the images of Figure 19.

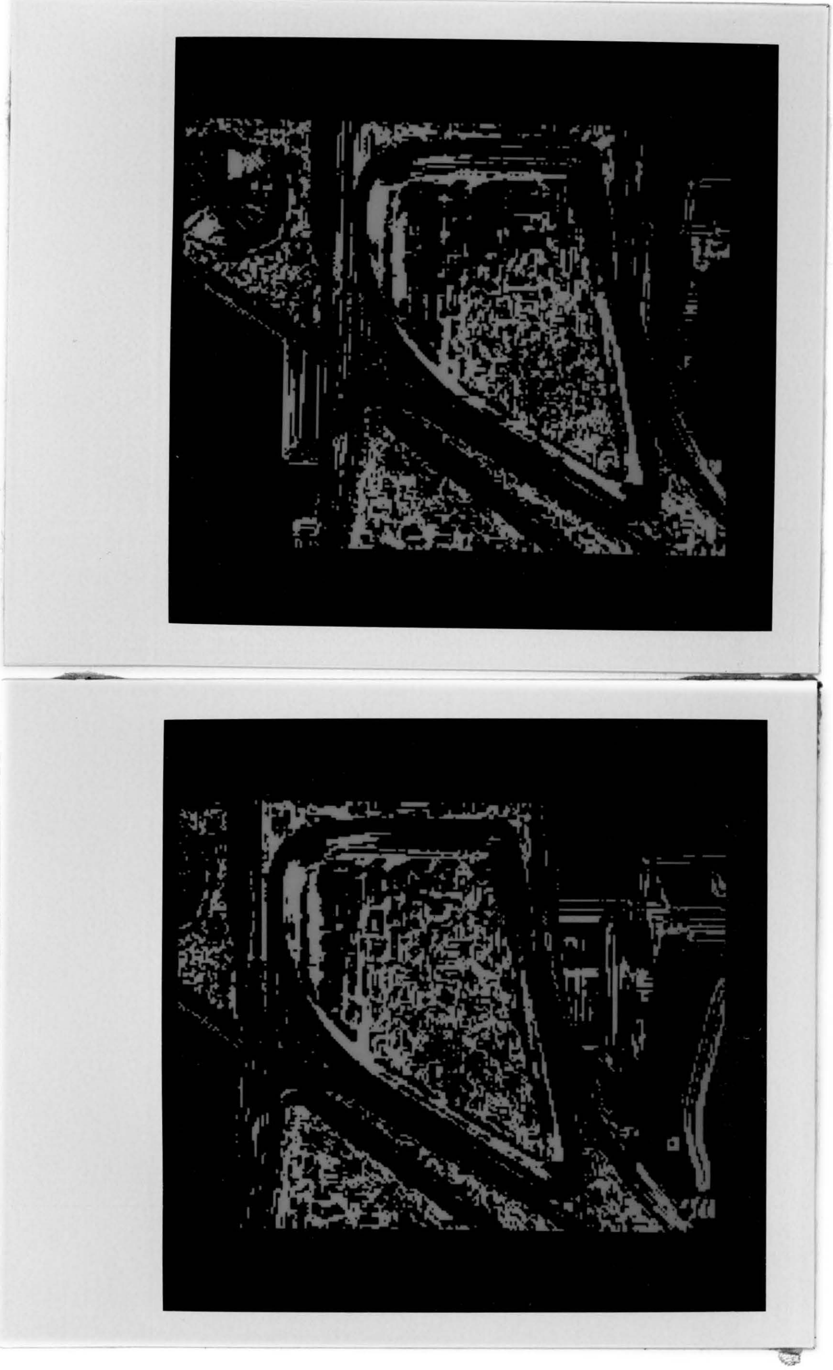


Figure 29: Slope labels in the images of Figure 19.

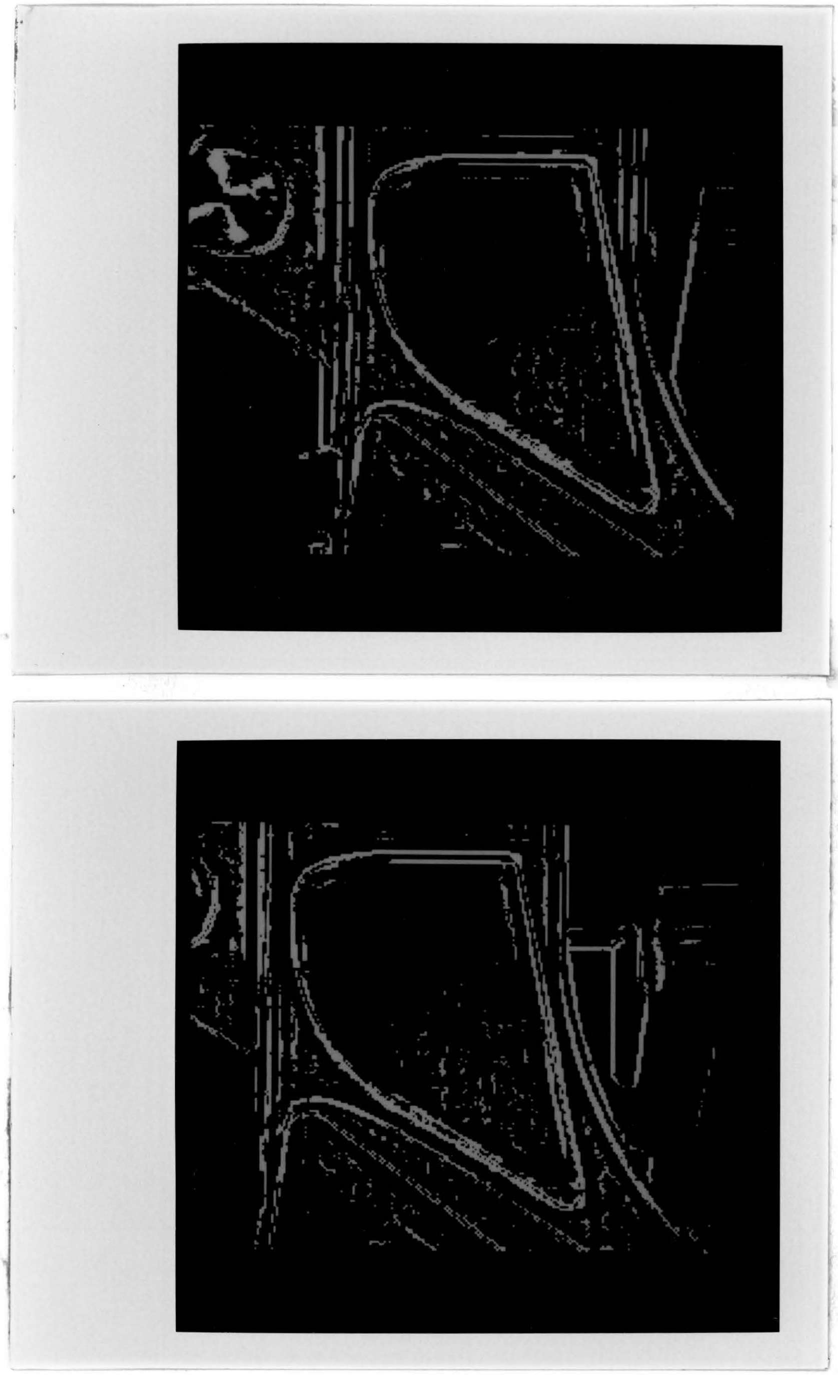


Figure 30: Concex hillside labels in the images of Figure 19.

Figure 31: Concave hillside labels in the images of Figure 19.

4.2 STEREO MATCHING

4.2.1 Stereo Imaging Geometry

The stereo imaging geometry is illustrated in Figure 32. We assume that the image planes of the stereo cameras are coplanar and are placed at a distance f, the focal length of the lens, in front of the two camera foci. The cameras are arranged so that the camera baseline, the line joining the two camera foci, is parallel to the row scan direction of the image (the x-axis in Figure 32). Given any point in the scene, the epipolar plane is defined as the plane determined by the point and the camera foci. It is clear that the projections of a point in 3-space onto the two image planes must lie somewhere along the intersecting lines between the epipolar plane and the image planes. The intersecting lines on the left and right image planes are called epipolar lines. Therefore, the corresponding problem is reduced to finding matches along corresponding epipolar lines. If the input images are taken under a different arrangement, a preprocessing step is required to account for the differences. One such technique can be found in [Ge77].

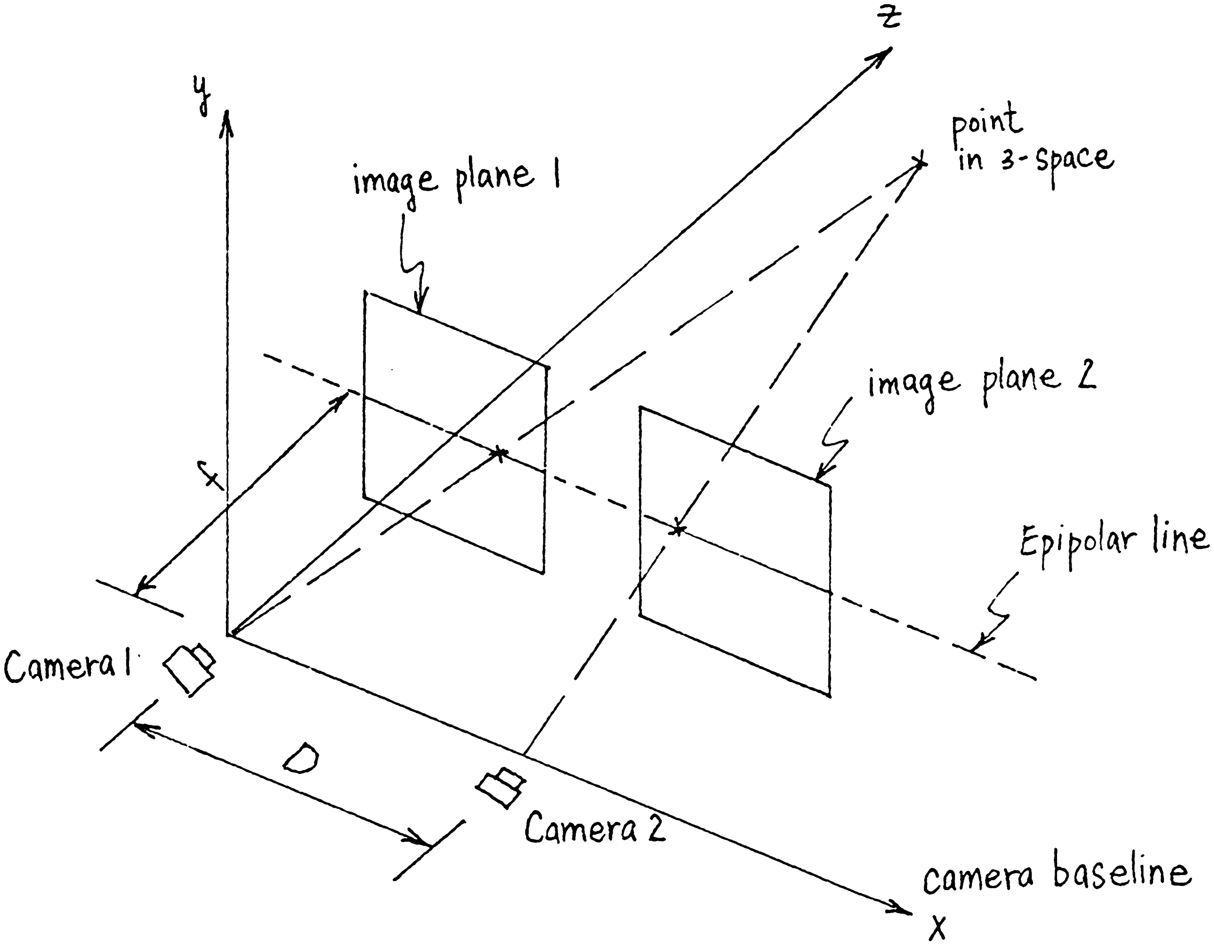


Figure 32: Stereo imaging geometry.

4.2.2 Notation

The segmentation process partitions the images I^l and I^r into arc and region segments. As a notation, we will use superscripts l and r to denote, respectively, descriptors for the left and right image (descriptors without superscript can be applied to either image). The matching process to be described is performed in both direction, that is, from the left to the right image, and vice versa. In the following discussion, we will concentrate on finding matches from the left image to the right.

Let N_A denote the number of arc segments,

$A = \{a_i \mid i = 1,\ldots,N_A\}$ denote the set of arc segments,

N_R denote the number of region segments,

and $R = \{r_i \mid i = 1,\ldots,N_R\}$ denote the set of region segments.

Associated with each arc segment, the mapping $o: A \rightarrow [0,360]$ gives the mean orientation for each arc segment in A. Similarly, the mapping $g: R \rightarrow [I_{min}, I_{max}]$ gives the mean gray level for each region in R.

Define a boolean function p(x,y) for $x \in A^l$ and $y \in A^r$ by

$$p(x,y) = \begin{cases} \text{true} & \text{if } d(x,y) < t_o \\ \text{false} & \text{otherwise} \end{cases}$$

where t_o is a threshold for edge orientation and

$$d(x,y) = \begin{cases} |o(x)-o(y)| & \text{if } |o(x)-o(y)| \geq 180 \\ 360 - |o(x)-o(y)| & \text{otherwise} \end{cases} .$$

Thus p(x,y) is true if and only if arcs x and y are oriented similar enough. Define a boolean function q(x,y) for $x \in R^l$ and $y \in R^r$ by

$$q(x,y) = \begin{cases} \text{true} & \text{if } |g(x)-g(y)| < t_g \\ \text{false} & \text{otherwise} , \end{cases}$$

where t_g is a threshold for mean gray level. Thus q(x,y) is true if the mean gray tones of regions x and y are similar. The selection of the above thresholds is quite loose. Although tight thresholds can reduce the search space, it increases the risk of ruling out potential matches. In the examples presented in this paper, t_o is set to 30 and t_g is set to 50.

4.2.3 Matching Epipolar Line

Since arc and region segments can be very large in size, it will be expensive in terms of both memory and computation to correlate segments. Instead of treating segments as whole units, we will divide the matching problem into subproblems by first matching epipolar lines. Region or arc segment correspondences, which is discussed in the next section, will then be determined globally by combining matching results on the epipolar lines.

Since the images are partitioned into arc and region segments, an image line is divided into connected arc-sections and region-sections. Let L^l and L^r be a pair of corresponding epipolar lines from the topographic images. We have

$$L^l(i) = t^l \quad \text{and} \quad L^r(i) = t^r, \qquad \text{for } i=1,\ldots,\text{nppl}$$

where nppl is the number of pixels per line, and t^l and $t^r \in$ {ridge, valley, convex hillside, concave hillside, saddle hillside, slope, flat}. For an epipolar line,

let n_a be the number of arc-sections,

$S_{arc} = \{b_i \mid i=1,..,n_a\}$ be the set of arc-sections,

n_r be the number of region-sections,

and $S_{reg} = \{c_i \mid i=1,..,n_r\}$ be the set of region-sections. Notice that each region section corresponds to part of a region segment and each region section is bounded by a pair of arc segments except at the boundaries of the image. We thus have for $y \in S_{reg}$, the mapping REG: $S_{reg} \to R$, which gives the region segment to which y belongs. Correspondingly, for $x \in S_{arc}$, the mapping ARC: $S_{arc} \to A$ gives the arc segment to which x belongs.

Let BEGIN: $S_{reg} \to [1,\text{nppl}]$ and END: $S_{reg} \to [1,\text{nppl}]$ be functions which return respectively the beginning and ending locations of a region-section. We can then define a function, LARC: $S_{reg} \to \{A^r, 0\}$, which returns the arc segment to the immediate left of x. If BEGIN(x)=1, LARC(x)=0. Similarly, the function RARC: $S_{reg} \rightarrow \{A^r, 0\}$ returns the arc segment to the immediate right of x. If END(x)=nppl, RARC(x)=0.

We can now define the set of all possible matches for $x \in S^l_{reg}$ as

$$M^l(x) = \{y \mid y \in S^r_{reg}, \quad q(REG^l(x), REG^r(y)) = \text{true}, \quad p(LARC(x), LARC(y)) = \text{true or} \quad p(RARC(x), RARC(y)) = \text{true}, \quad |BEGIN(x)-BEGIN(y)| < d_{max}, \quad \text{and } |END(x)-END(y)| < d_{max}\}$$

where d_{max} is a maximum allowable disparity. That is, $\{x \in S^l_{reg}, y \in S^r_{reg}\}$ is a feasible match if

1. the mean gray levels of the region segment segments to which the region sections belong are not too different,
2. the mean orientations of the arc segments to the immediate left (or right) of the matching region sections are not too different, and
3. the disparity of the end points of the matching region sections is less than a certain maximum allowable disparity.

For each possible match, a similarity measure is computed by correlating topographic structures within the corresponding region-sections. If we let x $\in S^{l}_{reg}$ and y $\in M^{l}(x)$, the lengths of x and y are given by

$$d^{l} = \mathrm{END}(x) - \mathrm{BEGIN}(x) \quad \text{and} \quad d^{r} = \mathrm{END}(y) - \mathrm{BEGIN}(y),$$

and the arc-sections adjacent to x and y are related by

$$p_1 = p(\mathrm{LARC}(x),\mathrm{LARC}(y))$$

$$\text{and} \quad p_2 = p(\mathrm{RARC}(x),\mathrm{RARC}(y)) \ .$$

It is worth noting that p_1 (or p_2) is true if the arc segments to the immediate left (or right) of x and y are matchable. There are three cases to be considered:

1. if p_1 is true and p_2 is false, we compute a similarity measure which is proportional to the number of times that the topographic labels to the immediate right of LARC(x) and LARC(y) are compatible;
2. if p_1 is false and p_2 is true, we computer a similarity measure which is proportional to the number of times that the topographic labels to the immediate left of RARC(x) and RARC(y) are compatible; and
3. if both p_1 and p_2 are true, we computer a similarity measure which is proportional to the number of compatible topographic labels at both ends of the region-sections.

Formally, we compute a similarity measure $m(x,y)$ in the following way:

$$m(x,y)=\begin{cases} \frac{1}{\max(d^l,d^r)}\left[\sum_{k=0}^{\min(d^l,d^r)-1} h(\mathrm{BEGIN}(x)+k,\mathrm{BEGIN}(y)+k)\right], & \text{if } p_1=\text{true and } p_2=\text{false},\\ \frac{1}{\max(d^l,d^r)}\left[\sum_{k=0}^{\min(d^l,d^r)-1} h(\mathrm{END}(x)-k,\mathrm{END}(y)-k)\right], & \text{if } p_1=\text{false and } p_2=\text{true},\\ \frac{1}{\max(d^l,d^r)}\left[\sum_{k=0}^{\min(d^l,d^r)/2-1} h(\mathrm{BEGIN}(x)+k,\mathrm{BEGIN}(y)+k) + \sum_{k=0}^{\min(d^l,d^r)/2-1} h(\mathrm{END}(x)-k,\mathrm{END}(y)-k)\right], & \text{otherwise.}\end{cases}$$

where

$$h(i,j) = \begin{cases} 1 & \text{if } L^l(i) = L^r(j) \\ 0 & \text{otherwise} \end{cases}$$

Notice that $h(i,j)$ determines whether the topographic labels at the i^{th} location of the first region-section and the j_{th} location of the second region-section are the same. The motivation for using m as a similarity measure is that if the topographic structures around two potentially matching segments are similar, the likelihood that they really match is high.

We call y_* a preferred match of x if $m(x,y_*)$ is the largest among all possible matches. We also call LARC(y_*), the arc-section to the immediate left of y_*, a preferred match of LARC(x) if p_1 is true. Similarly, RARC(y_*) is a preferred match of RARC(x) if p_2 is true. In order to avoid ambiguous matches, no preferred match is assigned to x if $m(x,y_*)$ is smaller than some predetermined threshold.

4.2.4 Matching Segments

At this point, each region (or arc) section in the epipolar line of one image is associated with a preferred matching section, if one exists, in the other image. The next attempt is to find global matches for the segments as a whole. Since each segment (region or arc) is decomposed into a set of subsections during the epipolar line matching process, we define SECTIONS(x) to be the set of all subsections of x, that is

$$\text{SECTIONS}(x) = \{ y \mid \text{REG}(y)=x \} \text{ for } x \in R, \text{ or}$$

$$\text{SECTIONS}(x) = \{ y \mid \text{ARC}(y)=x \} \text{ for } x \in A.,$$

Since a preferred match is determined for each $y \in$ SECTIONS(x), we can define a mapping M_*: SECTIONS -> R^r which gives the segment to which the preferred match of y belongs. The set of all possible matches for x is then

$$\text{ALLMATCH}(x) = \{ M_*(y) \mid y \in \text{SECTIONS}(x) \} .$$

We now have for each segment in one image a set of potentially matching segments in the other image. In order to obtain the best of the potential matches, each possible match, $x \in R^1$ and $y \in$ ALLMATCH(x), is evaluated. Two criteria are used in determining the best matches for region segments:

1. the number of times that y matches each of the region sections in SECTIONS(x); and
2. the widths of region sections of which y is the prefered match.

Specifically, the similarity measure sm(x,y) is given by

$$\mathrm{sm}(x,y) = \frac{1}{|x|} \sum_{z \in \mathrm{SECTIONS}(x)} \mathrm{len}(z,y)$$

where $|x|$ denotes the size of x and

$$\mathrm{len}(z,y) = \begin{cases} \mathrm{END}(z)-\mathrm{BEGIN}(z) & \text{if } M_*(z)=y \\ 0 & \text{otherwise .} \end{cases}$$

Notice that len(z,y) returns the length of z if y is a preferred match of z and returns zero otherwise. The best match is considered to be the one which gives the highest similarity measure. We also require that the similarity measure corresponding to the best match be greater than some predetermined value.

A similar evaluation can be computed for matching arc segments, except that for an arc segment, its width or thickness does not play as important a role as for region segments. Therefore, len for an arc is defined to be

$$\mathrm{len}(z,y) = \begin{cases} 1 & \text{if } M_{*}(z)=y \\ 0 & \text{otherwise} . \end{cases}$$

A final global matching process is performed as an attempt to resolve ambiguity and to preserve consistent matches. To measure the degree of consistency between two sets of matches, we examine the region adjacency graphs of the matching segments. Let RAG(x) denote the region adjacency list of x. That is,

$$\mathrm{RAG}(x) = \{x' \mid x' \text{ is adjacent to } x\}.$$

Consider a best match (x,y). We also have for each element in RAG(x) and RAG(y) a best match or "no" match. We can define a consistency measure cm(x,y) by

$$\frac{\# \{u \in \mathrm{RAG}(x) \mid u \text{ is a best match of } v \text{ for some } v \in \mathrm{RAG}(y)\}}{\# \mathrm{RAG}(x)} .$$

Ideally, if x matches y, then each neighbor of x should match some neighbor of y. Therefore, the match (x,y) is highly consistent if cm(x,y) is near one.

In summary, the stereo matching process is divided into three stages:

1. an epipolar line matching process which uses epipolar geometry to reduce the problem to an one-dimensional matching process;
2. a segment matching process which establishes globally optimal matches for region and arc segments; and
3. a global matching process which uses spatial relations among high-level structures to resolve ambiguous matches.

4.3 RESULTS

The stereo matching procedure has been performed on the stereo images of the F-15 bulkhead (Figure 19). Figures 33 and 34 illustrate respectively the matching arc segments and matching region segments. Matching segments in the pictures are displayed with the same color. All matching segments have been identified correctly. It is interesting to note that horizontal arc segments were also matched correctly. Matching between horizontal arc segments was possible because segment adjacency information was incorporated in the global matching process.

Two views of a different portion of the F-15 bulkhead are shown in the images of Figure 35. The arc and region segments extracted from the stereo images of Figure 35 are

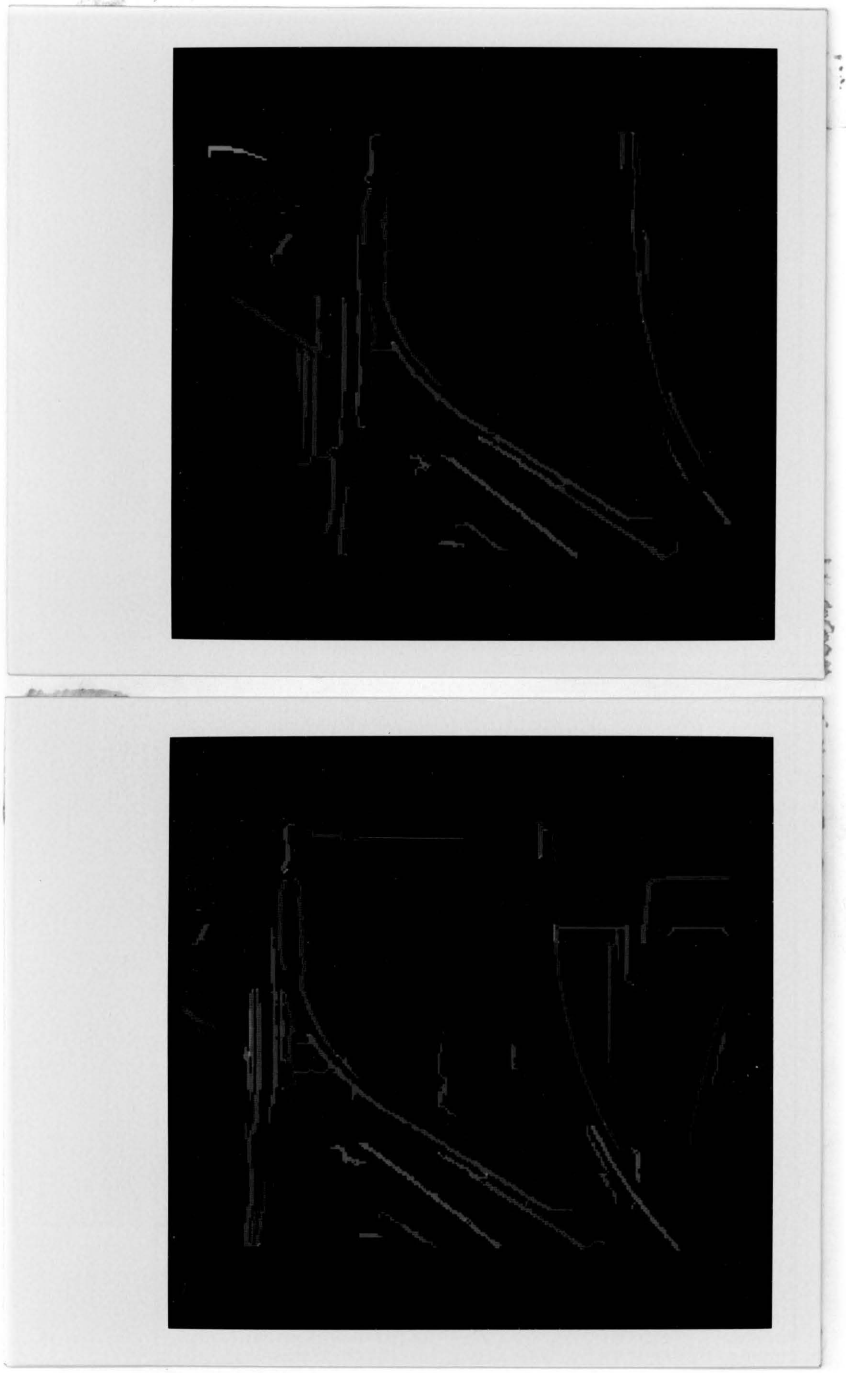


Figure 33: Matching results of the arc segments of Figure 24.

Figure 34: Matching results of the region segments of Figure 25.

shown respectively in Figure 36 and Figure 37. Topographic structures within the region segments are illustrated in Figure 38 where different topographic structures are displayed with different colors. Matching results are shown in Figures 39 and 40.

A pair of stereo images of two curved objects are shown in Figure 41. The arc, region, and topographic structures extracted from the stereo images are illustrated respectively in Figures 42, 43 and 44. Figure 45 illustrates the disparity map resulting from the stereo matching process. By incorporating the depth constraints into the shape-from-shading algorithm as described in Chapter 3, a surface reconstruction was obtained (Figure 46).

4.4 SUMMARY

Instead of using low level image features, we have proposed here a stereo matching procedure based on high level structures such as arcs, regions and topographic structures. Results show that this is a promising step toward using high level features for stereo matching.

Since the feature extraction process is performed independently on the stereo pair of images, one segment

Figure 35: Stereo images of part of a F-15 bulkhead.

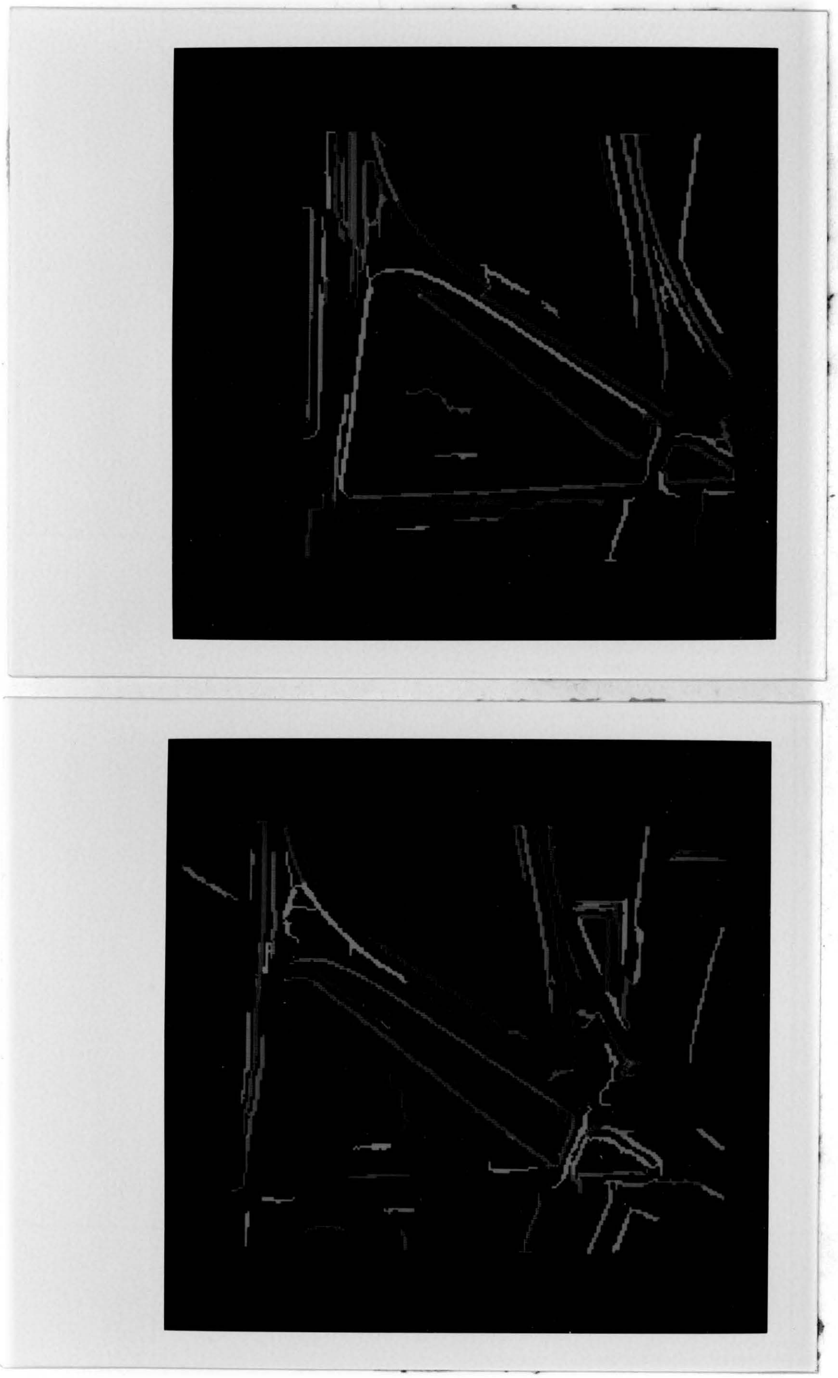


Figure 36: Arc segments extracted from the images of Figure 35.

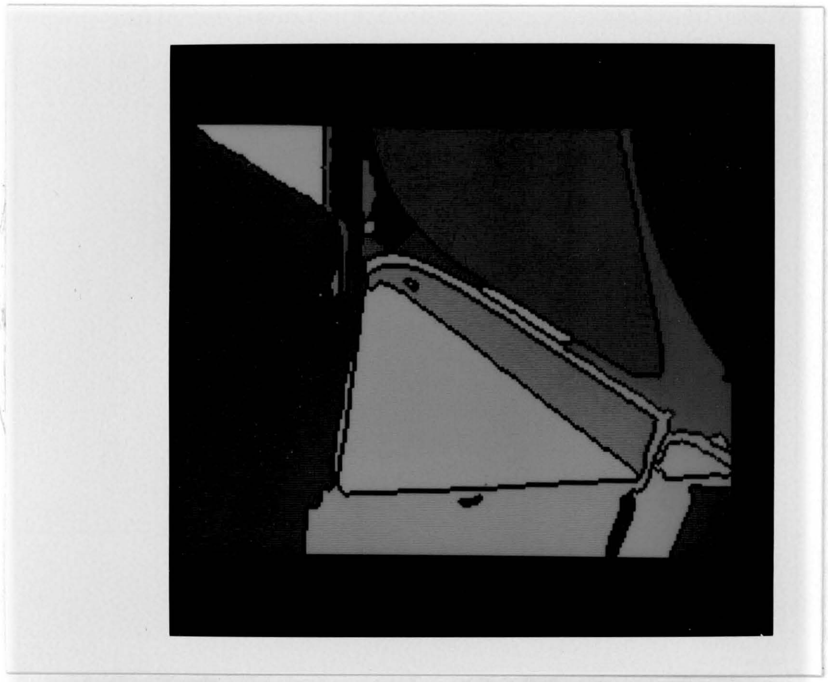


Figure 37: Region segments extracted from the images of Figure 35.

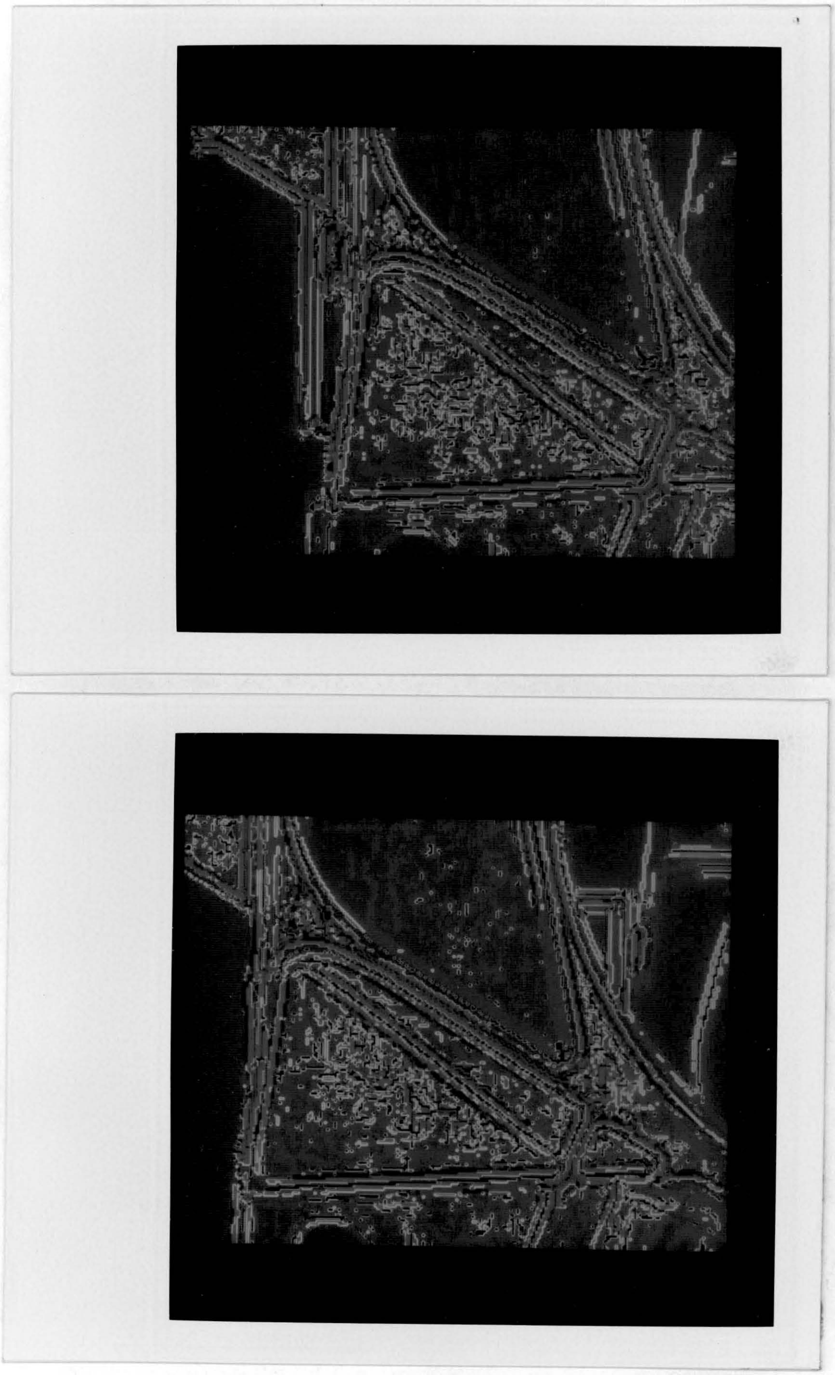


Figure 38: Topographic structures extracted from the images of Figure 35.

Figure 39: Matching results of the arc segments of Figure 36.

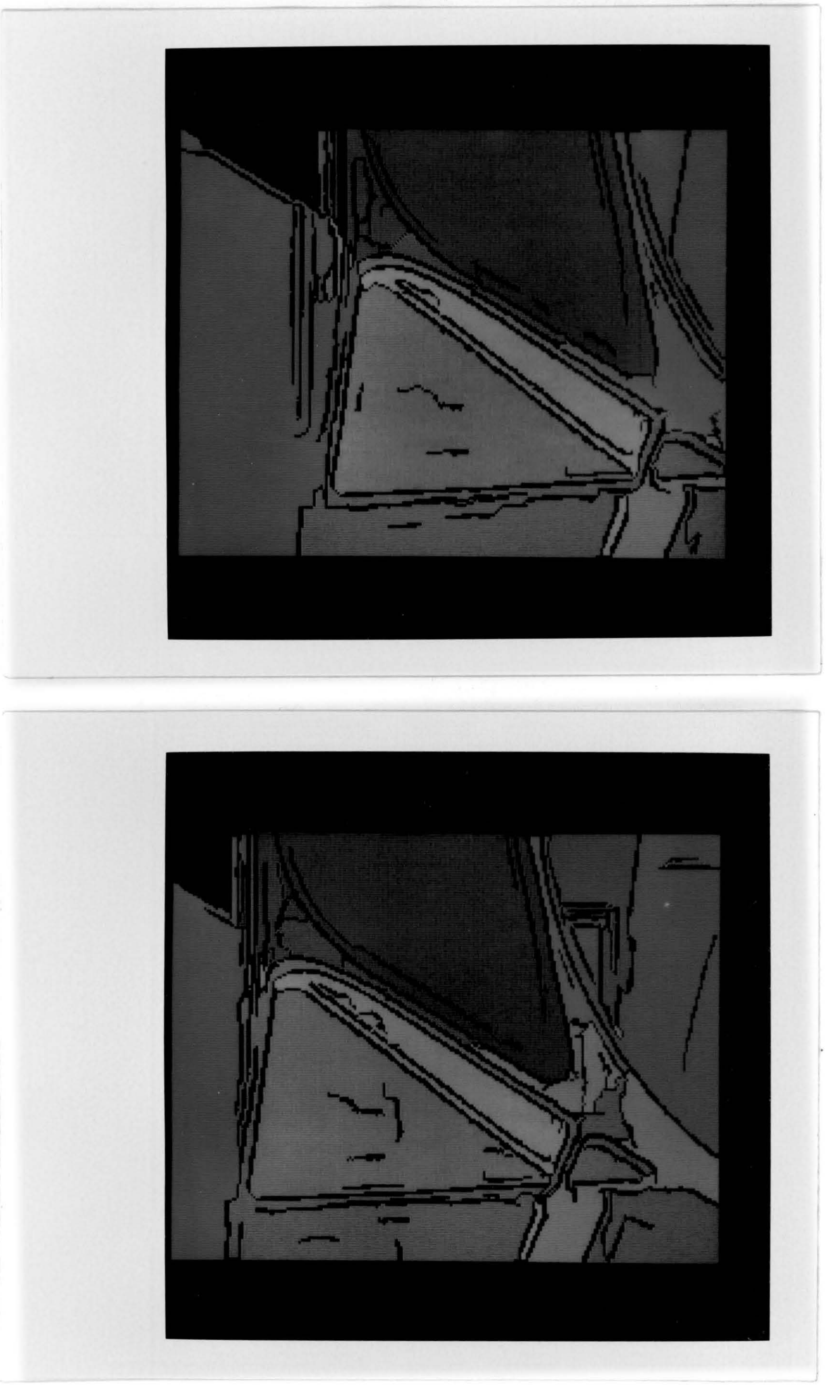


Figure 40: Matching results of the region segments of Figure 37.

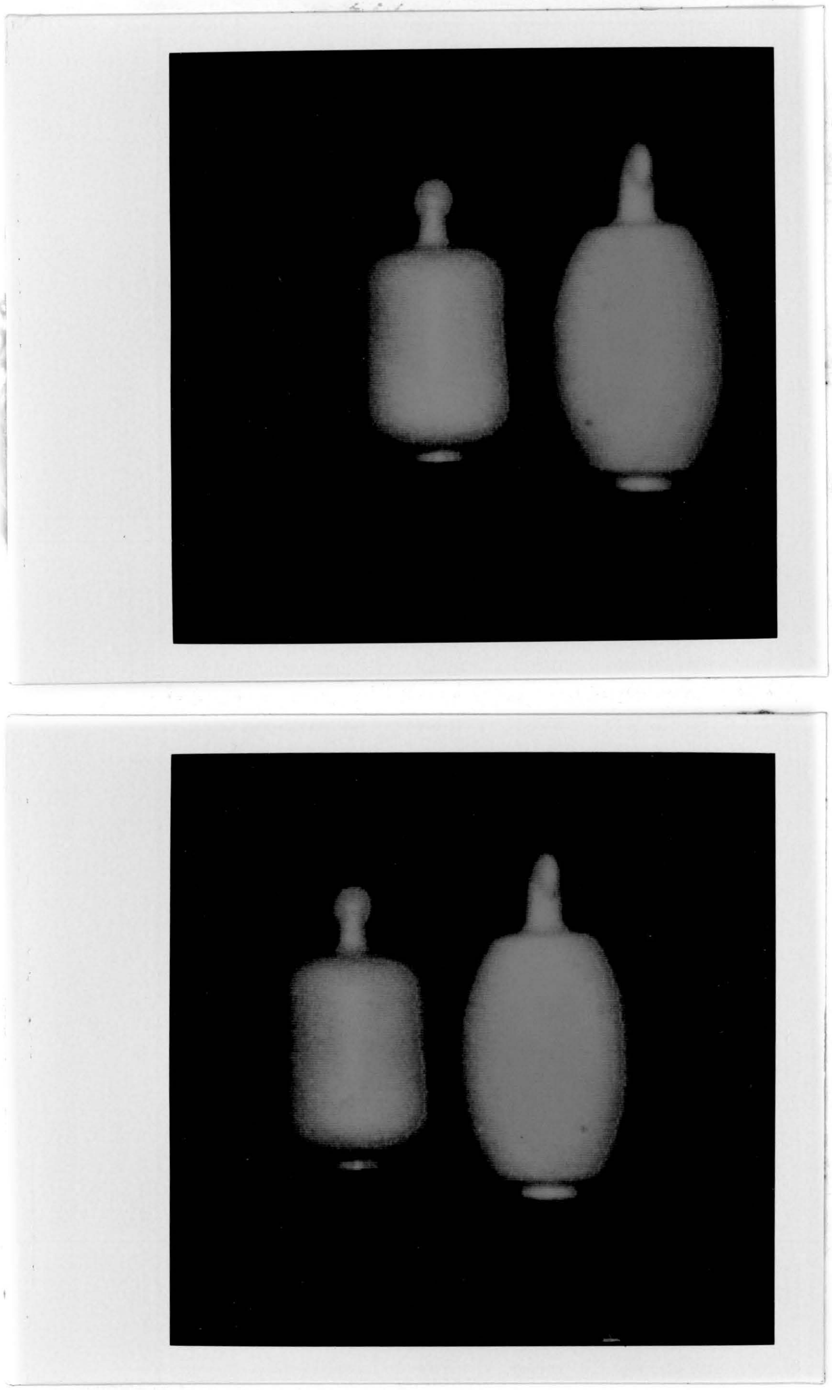
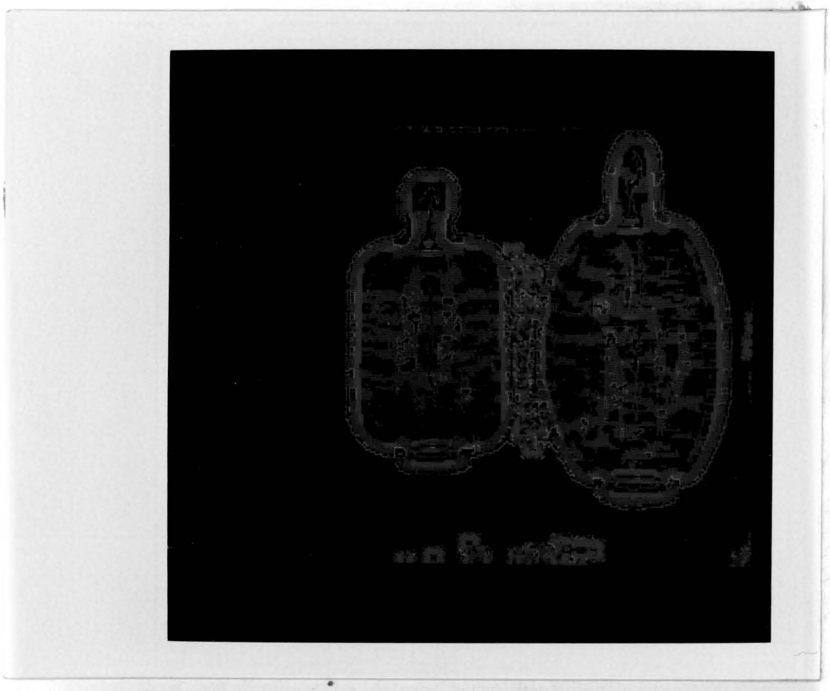


Figure 41: Stereo images of two curved surfaces..

Figure 42: Arc segments extracted from the images of Figure 41.

Figure 43: Region segments extracted from the images of Figure 41.



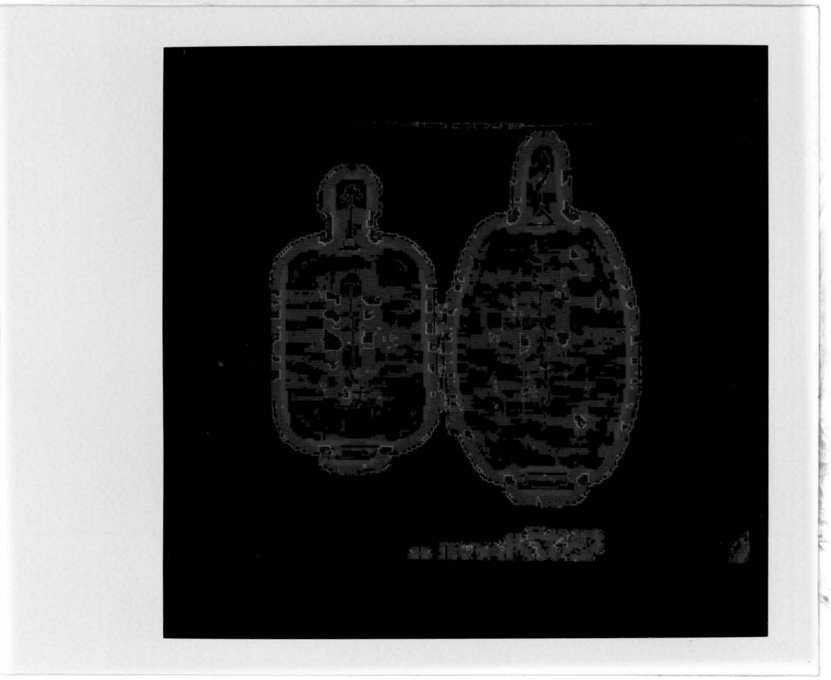


Figure 44: Topographic structures extracted from the images of Figure 41.

Figure 45: Disparity map of the stereo images of Figure 41.

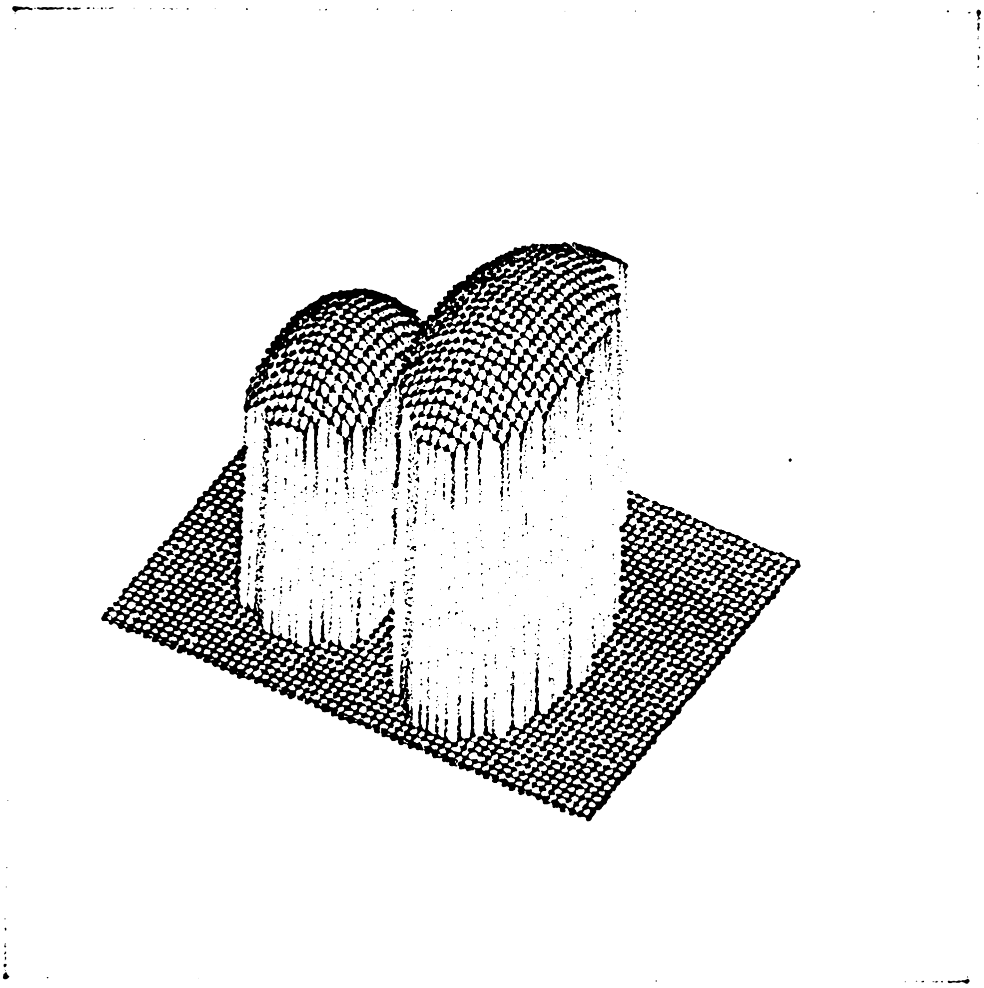


Figure 46: Reconstructed surfaces of the curved objects of Figure 41.

extracted in one image may correspond to multiple segments in the other image. More accurate segmentation can be achieved if matching results are incorporated into a post matching segmentation process.

So far, we have employed only the segment adjacency relation for the global matching process. A possible extension of the current matching scheme is to include relations such as parallel, antiparallel, and collinear.

Although we are matching both region and arc segments, the computations of reliable disparity measures are possible only at the locations of the arc segments. However, if point correspondences are desired for the whole image, the search space is reduced since we need only to search for point correspondences within matching segments. Methods for generating a full disparity map from the sparse disparity values should be explored. A cooperative shape-from-shading algorithm which uses the sparse disparity maps as boundary constraints is proposed in Chapter 3.

Chapter V

SHAPE ESTIMATION FROM TOPOGRAPHIC PRIMAL SKETCH

5.1 INTRODUCTION

Consider an image of a three-dimensional object illuminated by an arbitrary light source and viewed from an arbitrary position. Although ambiguities are possible, frequently the human viewer can estimate a) the three-dimensional shape of the object, b) the camera position, and c) the location of the light source. The original "shape-from-shading" techniques [Ho77] solve systems of differential equations to derive three-dimensional shape from gray tone intensity variations and operate under a limiting set of restrictions. In additional to low level shading cues, we believe that the human viewer also recognizes patterns in the image that give cues leading to estimation of the shape of the object.

Extracting patterns from the original gray tone image is, in most nontrivial cases, an impossible task. In fact, it is for this reason that syntactic pattern recognition systems have had to first extract descriptions consisting of primitives, their properties, and their interrelationships from the image and then to parse these descriptions according to the rules of a grammar. Instead of trying to recognize patterns at the gray-tone intensity level, we propose to work at the topographic labeling level.

To obtain a topographic labeling, a gray tone image may be viewed as a three-dimensional surface whose height at each row-column position corresponds to the intensity value of the image at that position. While the image is a discrete matrix of values, the underlying surface is continuous. Each point of the surface may be labeled as part of a peak, pit, ridge, valley, saddle, hillside, or flat area. Hillsides can be further broken down into the subcategories inflection point, convex hill, concave hill, saddle hill, and slope. In [Ha83], these categories are defined mathematically and the topographic classification of image pixels is summarized in Section 4.1.2.

Our goal is to use patterns expressed in terms of ridges and valleys, peaks and pits, flats and hillsides to estimate three-dimensional shape. In section 5.5 of this chapter, results employing two methods for determining such topographic patterns from gray tone intensity images of simple surfaces are described. Section 5.3 discusses how surface orientations can be estimated from topographic structures. Section 5.4 describes an object surface classification scheme based on the topographic structures extracted from the image.

5.2 SHAPE FROM TOPOGRAPHIC PATTERNS

There are two possible methods for determining the pattern of topographic labels that will appear, given a particular three-dimensional shape category, a particular reflectance model, a particular light source, and a particular viewpoint. The first method is to work the problem analytically, obtaining exact equations for the illuminated surface. At each point the gradient, eigenvectors, and eigenvalues can be computed in order to determine precisely which sets of points have the various topographic labels. The second method is to work the problem experimentally, using software to generate digital images of illuminated three-dimensional surfaces, to fit these image with either polynomials, splines, or discrete cosines, and to assign topographic labels to each pixel. The first method has the advantage of exactness and the disadvantage of becoming extremely difficult for all but the simplest surfaces. The second method has the advantage of being applicable to a wide variety of surfaces and illuminating conditions and the disadvantage of yielding some inaccurate results due to possible errors in fitting the gray tone image. We have begun to experiment with both methods, starting with very simple surfaces, the Lambertian reflectance model, and point light sources. We have worked with four simple surfaces: (1) the top half of a cylinder, (2) the upper hemisphere of a sphere, (3) the top half of an ellipsoid, and (4) the upper half of a hyperboloid. Figures 47-50 illustrate the four three-dimensional surfaces.

5.2.1 Method 1: The Experimental Approach

The process for topographic classification can be done in one pass through the image. At each pixel of the image, the following four steps, which are discussed in more detail in [Ha83], need to be performed.

1. Calculate the least-squares fitting coefficients of a two-dimension cubic polynomial in an nxn neighborhood around the pixel.
2. Use the coefficients calculated in step 1 to find the gradient, the gradient magnitude, and the eigenvalues and eigenvectors of the Hessian at the center of the pixel's neighborhood.
3. Search in the direction of the eigenvectors calculated in step 2 for a zero-crossing of the first directional derivative within the pixel's area.

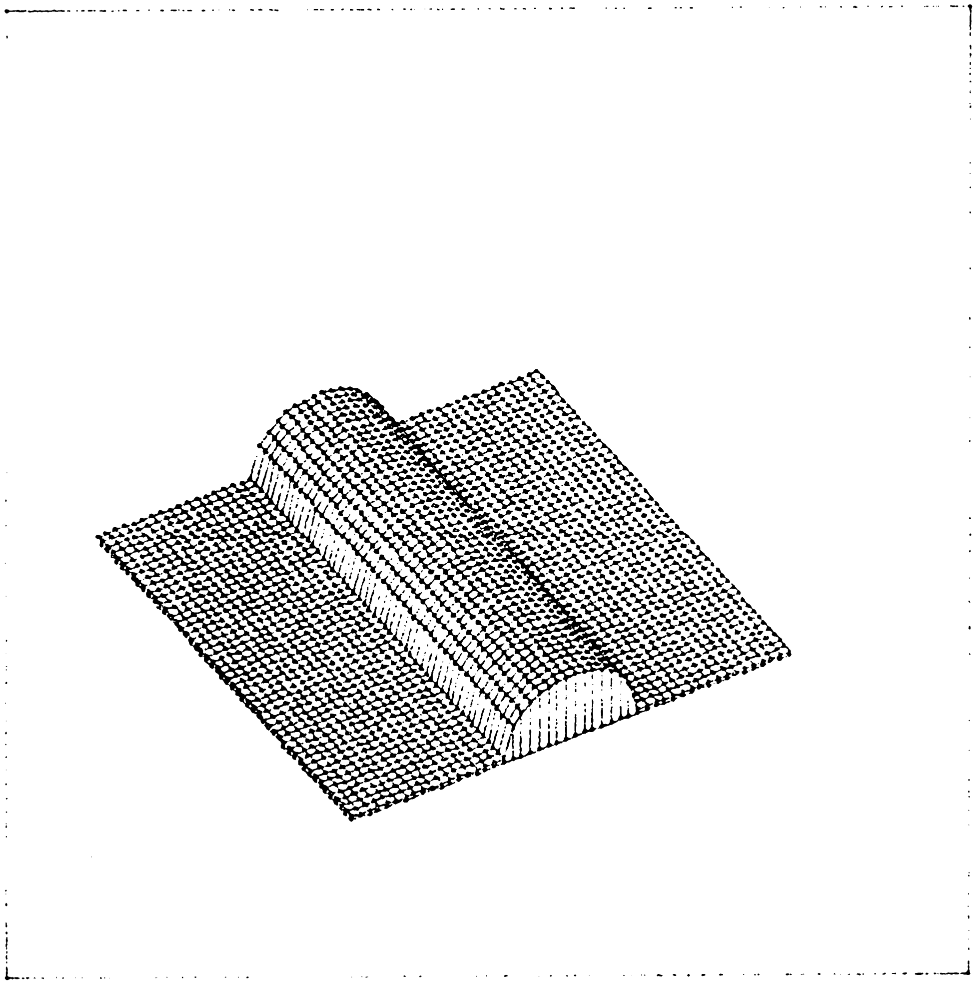


Figure 47: The cylindrical object used in our experiments.

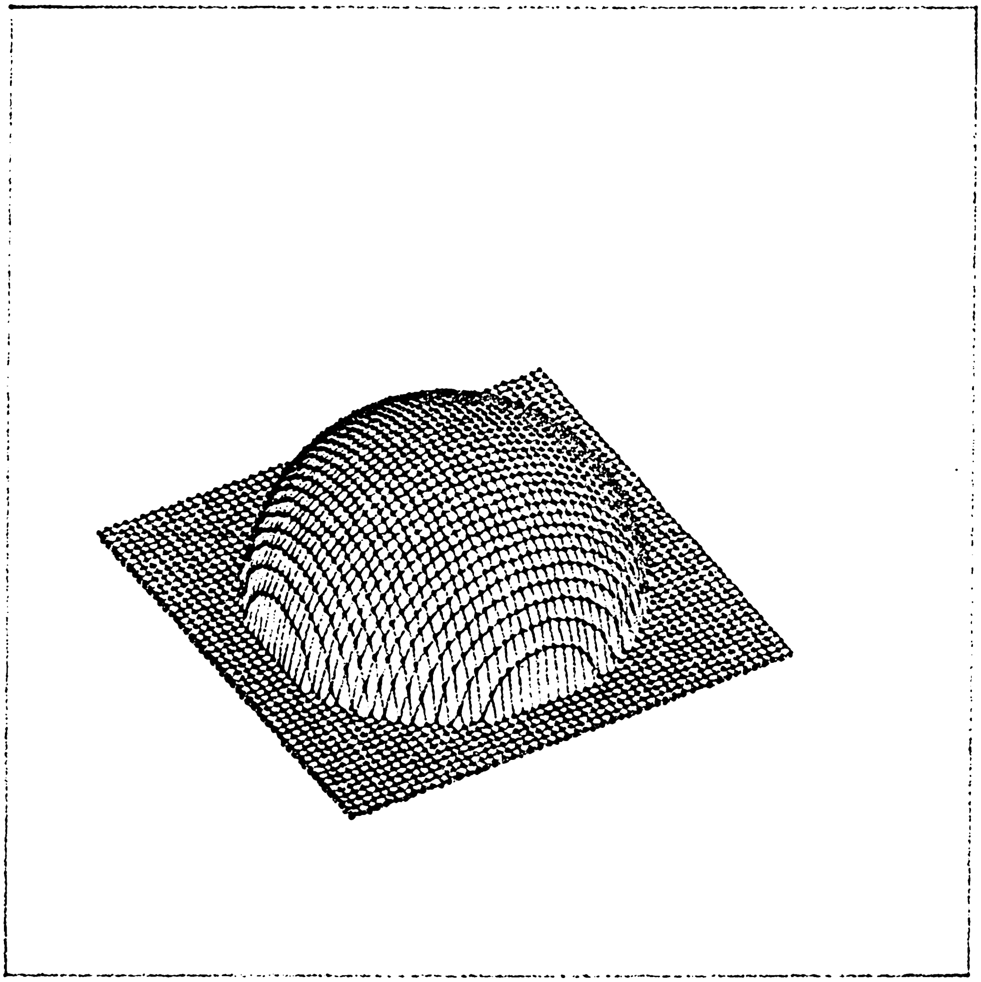


Figure 48: The spherical object used in our experiments.

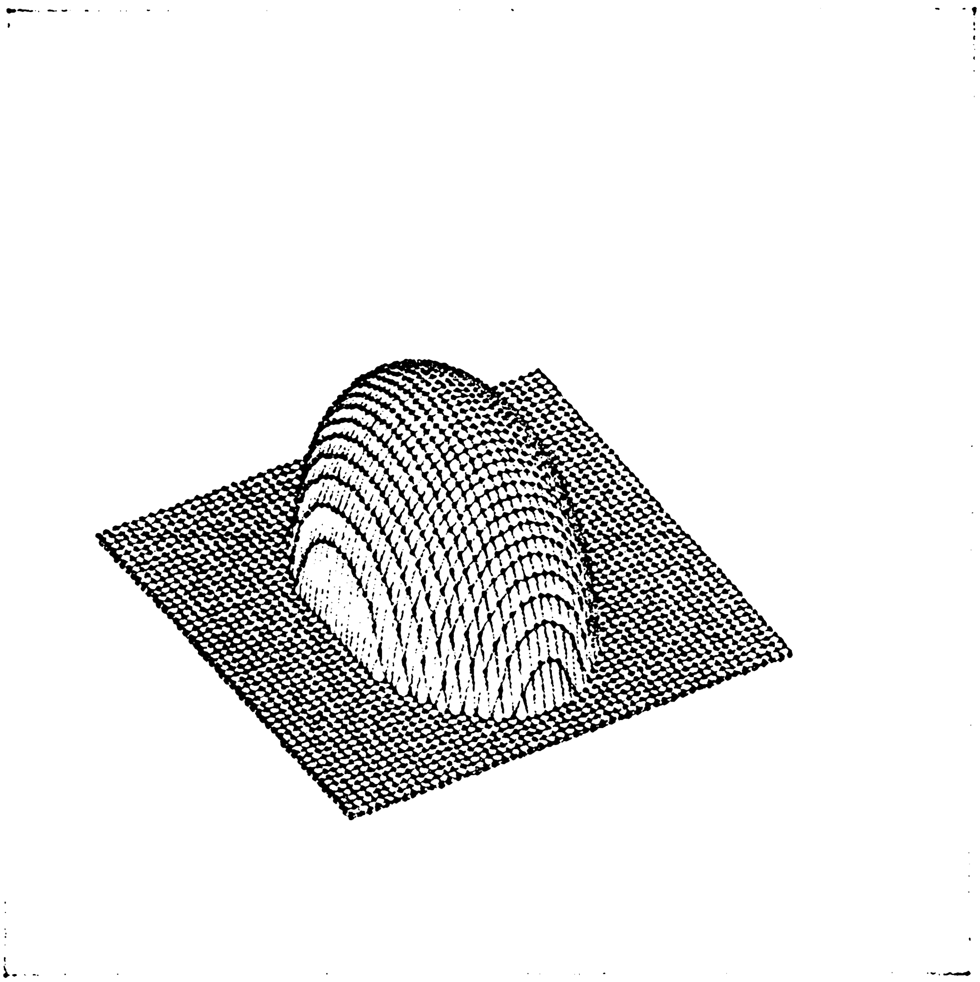


Figure 49: The ellipsoid used in our experiments.

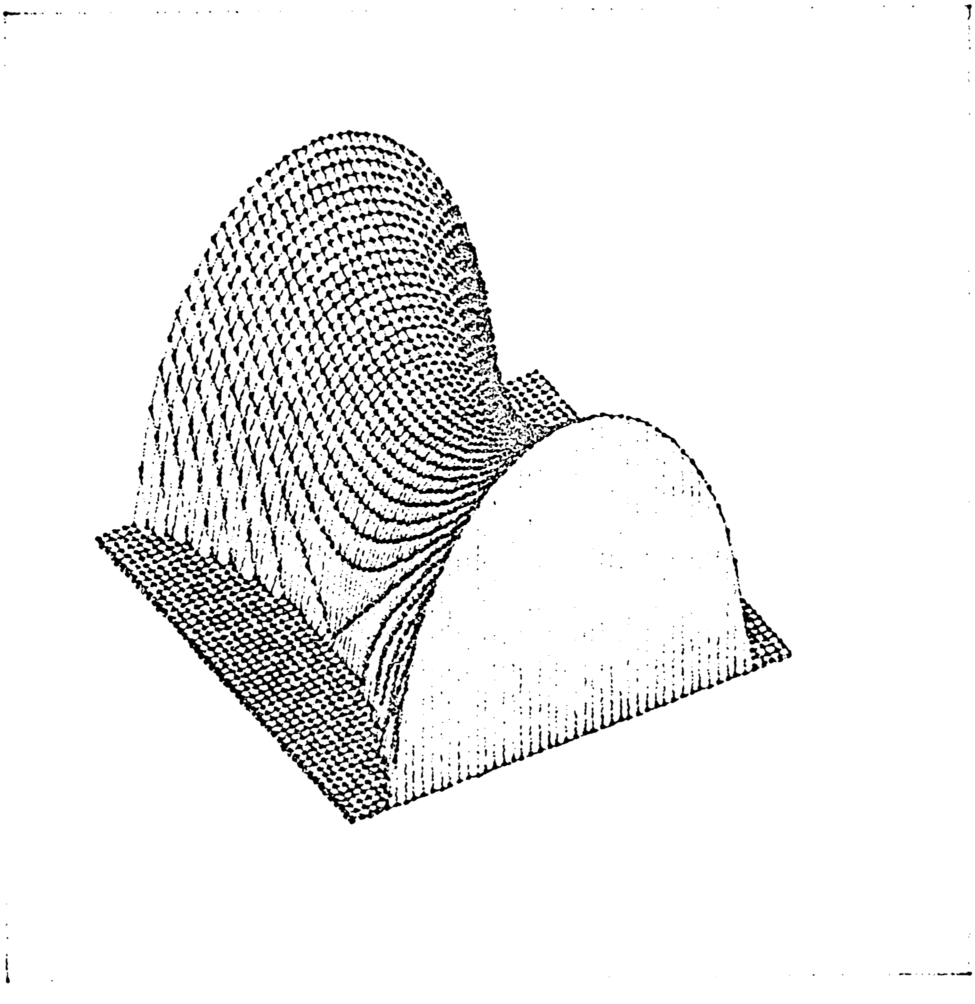


Figure 50: The hyperboloid used in our experiments.

4. Recompute the gradient, gradient magnitude, and values of second directional derivative extrema at each zero crossing. Then classify the pixel based on Table 1.

5.2.2 Method 2: The Analytical Approach

5.2.2.1 Topographic Labels on the Cylinder

Consider a cylindrical surface given by:

$$S(x,y) = d - (r^2-y^2)^{1/2} \qquad \text{for} \quad -r \leq y \leq r \qquad (3)$$

where d is the distance of the x-y plane from the camera down the z-axis and r is the radius of the cylinder. This surface, in which the axis of the cylinder lies along the x-axis, was chosen to simplify calculations. Notice that since only the top half of the cylinder is considered, the sign of the square root in Eq.(3) is taken as positive. By differentiating S with respect to x and y, we obtain

$$S_x = 0$$

and $$S_y = y(r^2-y^2)^{-1/2} .$$

It follows from Eq.(2) that the intensity of the cylinder illuminated from direction (a,b,c) is given by:

$$I(x,y) = I_0(by-c(r^2-y^2)^{1/2})/r \qquad (4)$$

After some simplifications, the first and second partials of I are found to be:

$$Ix = Ixx = Ixy = Iyx = 0 ,$$

$$Iy = I_0(b+cy(r^2-y^2)^{-1/2})/r ,$$

and $Iyy = I_0cr(r^2-y^2)^{-3/2}$.

where the subscripted I's denote partial differentiation with respect to the subscript(s).

Since Ix is equal to zero, the gradient magnitude ($||\nabla f||$) is equal to the absolute value of Iy. Therefore, $||\nabla f|| = 0$ when

$$Iy = I_0(b+cy(r^2-y^2)^{-1/2})/r = 0$$

which implies

$$b(r^2-y^2)^{1/2} = -cy \ . \qquad (5)$$

Upon solving Eq.(5), we obtain

$$y^2 = r^2b^2/(b^2+c^2) \ .$$

Because c is always negative, the sign of y is taken to be the same as that of b in order for Eq.(5) to be satisfied.

To determine the second directional derivative extrema values and the first directional derivatives taken in the directions which extremize second directional derivatives, we form the Hessian:

$$H = \begin{vmatrix} 0 & 0 \\ 0 & I_0cr(r^2-y^2)^{-3/2} \end{vmatrix} .$$

The eigenvalues of the Hessian are obtained as:

$$X_1 = I_0cr(r^2-y^2)^{-3/2} \qquad (6)$$

and $X_2 = 0;$ (7)

their associated eigenvectors are:

$$w^1 = (0,1)$$

and $\qquad w^2 = (1,0)$.

Recall that c is always negative, therefore, X_1 is always negative for $-r<y<r$. By taking the dot product of the gradient with the eigenvectors, we obtain:

$$\nabla I.w^1 = Iy = I_0(b+cy(r^2-y^2)^{-1/2})/r$$

and $\qquad \nabla I.w^2 = 0.$

To determine the topographic labels, we need to consider two cases: (1) zero gradient magnitude and (2) positive gradient magnitude.

CASE 1: Zero Gradient Magnitude

If we let $y_0 = rb(b^2+c^2)^{-1/2}$, it follows from Eq.(5) that $||\nabla I|| = 0$ when $y=y_0$. By Eq.(6-7), the second directional derivative extrema values at y=y0 are

$$X_1 = I_0cr(r^2-y_0{}^2)^{-3/2} \qquad (8)$$

and $\qquad X_2 = 0. \qquad\qquad (9)$

Since X_1 is always less than zero, it follows directly from Table 1 that a ridge is located at $y=y_0$.

CASE 2: Positive Gradient Magnitude

If the gradient magnitude ($||\nabla I||$) is taken to be positive, then the value of the first directional derivative in the direction of w^1 ($\nabla I.w^1$) is always non-zero because $\nabla I.w^1=Iy$ and $||\nabla f||=|Iy|$. In this case, since X_1 is always negative and X_2 is always zero, it follows from row 11 of Table 1 that hillsides are located at those places where the gradient magnitudes are positive.

5.2.2.2 Topographic Labels on the Sphere

In the case of the sphere, the equation of a spherical surface with radius r is given by:

$$S(x,y) = d - (r^2-x^2-y^2)^{1/2} \quad \text{for } -r \leq x \leq r \text{ and } -r \leq y \leq r \tag{10}$$

Its intensity illuminated from direction [a,b,c] is given by:

$$I(x,y) = I_0 * [ax+by-c(r^2-x^2-y^2)^{1/2}]/r \tag{11}$$

After some simplifications, the first and second partials of I are found to be:

$$Ix = I_0[a+cx(r^2-x^2-y^2)^{-1/2}]/r \ ,$$

$$Iy = I_0[b+cy(r^2-x^2-y^2)^{-1/2}]/r \ ,$$

$$Ixx = I_0c(r^2-y^2)(r^2-x^2-y^2)^{-3/2}/r \ ,$$

$$Ixy = Iyx = I_0cxy(r^2-x^2-y^2)^{-3/2}/r \ ,$$

and

$$Iyy = I_0c(r^2-x^2)(r^2-x^2-y^2)^{-3/2}/r \ .$$

The gradient magnitude ($||\nabla I||$) is given by:

$$||\nabla I|| = (Ix^2 + Iy^2)^{1/2} \ ,$$

which is zero when

$$a(r^2-x^2-y^2)^{1/2} + cx = 0$$

and

$$b(r^2-x^2-y^2)^{1/2} + cy = 0$$

are satisfied simultaneously. By squaring and invoking the constraint $a^2+b^2+c^2=0$ on the unit vector $[a,b,c]$, the solution to the simultaneous equations is found to be:

$$x = ra$$

and $$y = rb \ .$$

The Hessian for the intensity surface of the illuminated sphere is given by:

$$H = \frac{I_0 c}{r(r^2-x^2-y^2)^{3/2}} * \left| \begin{matrix} r^2-y^2 & xy \\ xy & r^2-x^2 \end{matrix} \right| .$$

Its eigenvalues are found to be:

$$X_1 = I_0 cr(r^2-x^2-y^2)^{-3/2}$$

and $$X_2 = I_0 c(r^2-x^2-y^2)^{-1/2}/r \ .$$

Notice that both eigenvalues are always less than zero since c is always less than zero. The eigenvector corresponding to X_1 is given by:

$$w^1 = [x(x^2+y^2)^{-1/2}, \ y(x^2+y^2)^{-1/2}]$$

and the eigenvector corresponding to X_2 is given by:

$$w^2 = [-y(x^2+y^2)^{-1/2}, \ x(x^2+y^2)^{-1/2}] \ .$$

The dot product of the gradient with w^1 is

$$\nabla I.w^1 = \frac{I_0}{r(x^2+y^2)^{1/2}}[x(a+\frac{cx}{(r^2-x^2-y^2)^{1/2}})+y(b+\frac{cy}{(r^2-x^2-y^2)^{1/2}})] \qquad (12)$$

and the dot product of the gradient with w^2 is

$$\nabla I.w^2 = I_0(-ay+bx)(x^2+y^2)^{-1/2}/r \ . \qquad (13)$$

We determine the topographic labels by considering two cases.

CASE 1: Zero Gradient Magnitude

The gradient magnitude is equal to zero when $(x,y)=(ra,rb)$. Since both eigenvalues are less than zero on the illuminated sphere, it follows directly from Table 1 that a peak is located at $(x,y)=(ra,rb)$.

CASE 2: Positive Gradient Magnitude

In the case when the gradient magnitude is given to be positive, since both eigenvalues are known to be negative, it follows from Table 1 that there is a ridge at those locations where either $\nabla f.w^1=0$ or $\nabla f.w^2=0$ is satisfied. We obtain from Eq.(12) and Eq.(13) that

$$\nabla I.w^1=0 \text{ when } (ax+by)(r^2-x^2-y^2)^{1/2}+c(x^2+y^2)=0$$

$$\text{and } \nabla I.w^2=0 \text{ when } -ay+bx=0.$$

Table 1 also says that hillsides appear at places where both $\nabla f.w^1$ and $\nabla f.w^2$ are non-zero.

5.2.2.3 Other Conic Surfaces

We have also attempted to derive analytical results for the following conics surfaces:

(1) Ellipsoid: $S(x,y) = d - (l^2-m^2x^2-n^2y^2)^{1/2}$,

(2) Hyperboloid of one sheet:

$$S(x,y) = d - (l^2-m^2x^2+n^2y^2)^{1/2},$$

(3) Hyperboloid of two sheets:

$$S(x,y) = d - (m^2x^2-n^2y^2-l^2)^{1/2}, \text{ and}$$

(4) Paraboloid: $S(x,y) = d - (nx^2 + my^2)$.

Unfortunately, no simple closed form solution has been obtained for the first and second directional derivatives of these surfaces. Some of the analytical equations of these surfaces are summarized in the appendix. Results of the topographic patterns on these surfaces are described in Section 5.5. Results show that patterns also emerged for these surfaces.

5.3 ESTIMATION OF SURFACE ORIENTATION

In this section, we will show that the topographic labels along with their quantitative measurements bear a strong relationship to the surface orientation of the three-dimensional object in the scene. Consider a spherical surface as described in Section 5.2.2.2. The unnormalized surface orientation of such a surface can be represented in the gradient space by the vector [p,q,-1], where

$$p = \frac{x}{(r^2 - x^2 - y^2)^{1/2}} \quad \text{and } q = \frac{y}{(r^2 - x^2 - y^2)^{1/2}}$$

An alternative way of specifying surface orientation is the tilt and slant representation. Tilt specifies the orientation of the projection of the surface normal onto the image plane. Slant is the angle between the surface normal and the viewing direction. The tilt and slant representation and the gradient space representation can be related by the following formulas:

$$\text{Tan}\,\tau = q/p$$

and $$\text{Tan}\,\sigma = (p^2+q^2)^{1/2} \text{ or}$$

$$\text{Cos}\,\sigma = (1+p^2+q^2)^{-1/2} .$$

In the case of a sphere,

$$\text{Tan}\,\tau = y/x$$

and $$\text{Cos}\,\sigma = (r^2-x^2-y^2)/r^2$$

To see how the surface orientation of a spherical Lambertian surface can be derived from the topographic analysis of the image intensity surface, we need first to complete the analytical results of Section 5.2.2.2 by considering the lower half of the sphere. The equation of the lower hemisphere of a sphere whose center is at (0,0,d) is given by:

$$S(x,y) = d + (r^2 - x^2 - y^2)^{1/2} \quad \text{for } -r \le x \le r$$
$$\text{and } -r \le y \le r .$$

Differentiating the above equation with respect to x and y, we obtain

$$p = \frac{-x}{(r^2 - x^2 - y^2)^{1/2}} \qquad q = \frac{-y}{(r^2 - x^2 - y^2)^{1/2}}$$

Notice the sign difference between the surface orientations of the upper and lower hemispheres.

As for the upper hemisphere, after some simplification, we obtain for the lower hemisphere a similar set of expressions for I and its partials,

$$I = I_0[-ax-by-c(r^2-x^2-y^2)^{1/2}]/r,$$
$$I_x = I_0[-a+cx(r^2-x^2-y^2)^{-1/2}]/r,$$
$$I_y = I_0[-b+cy(r^2-x^2-y^2)^{-1/2}]/r,$$
$$I_{xx} = I_0c(r^2-y^2)(r^2-x^2-y^2)^{-3/2}/r,$$
$$I_{xy} = I_{yx} = I_0cxy(r^2-x^2-y^2)^{-3/2}/r,$$
and $$I_{yy} = I_0c(r^2-x^2)(r^2-x^2-y^2)^{-3/2}/r.$$

Notice that the second partials of I are the same for both halves of the sphere. Since the second partials make up the Hessian, it follows that the eigenvalues and eigenvectors for the two hemispheres are also identical. Recall that the eigenvalues and eigenvectors are given by

$$X_1 = I_0cr(r^2-x^2-y^2)^{-3/2},$$
$$X_2 = I_0c(r^2-x^2-y^2)^{-1/2}/r,$$
$$w_1 = [x(x^2+y^2)^{-1/2},\ y(x^2+y^2)^{-1/2}],$$
and $$w_2 = [-y(x^2+y^2)^{-1/2},\ x(x^2+y^2)^{-1/2}].$$

If we take the ratio of the smaller over the larger eigenvalue, we obtain

$$\frac{X_2}{X_1} = \frac{r^2-x^2-y^2}{r^2} .$$

This ratio is the square root of the cosine of the surface slant. Note that the signs of both X_1 and X_2 depend only on the sign of c which is the negative of the sine of the angle of elevation of the light source. Therefore, the ratio is always positive and its square root is always justifiable. Furthermore, the ratio is always less than or equal to one since X_2 is the smaller eigenvalue. Thus, we can obtain surface slant by taking the arc-cosine of the square root of X_2/X_1. The resulting angle is determined uniquely because the surface slant for a visible surface always lies between 0 and $\pi/2$.

The remaining component to be determined for the unit surface normal is the surface tilt. By considering w^1, the eigenvector corresponding to the larger eigenvalue, we can obtain the direction θ in which the second directional derivative of I is extremized. That is,

$$\tan \theta = \frac{y}{(x^2+y^2)^{1/2}} \cdot \frac{(x^2+y^2)^{1/2}}{x} = \frac{y}{x}$$

which is identical to the tangent of the surface tilt. Thus we have $t = \theta$ or $t = \theta+\pi$. Unfortunately, there are two possible solutions. This is expected because each solution corresponds to one half of the sphere. This shows the ambiguity in local analysis of image shading.

Recently, Pentland [Pe83] proposed a method that estimates surface orientation locally at each image point. This method is based on the claim that the image of a point on a spherical Lambertian surface can produce any combination of image intensity I and derivatives I_x, I_y, I_{xx}, I_{yy} and I_{xy}. The implication of this claim is that it is impossible to determine, based on local analysis of shading, whether or not an image point is resulted from a point on a spherical Lamberian surface. In what follows, we will show that Pentland's claim is incorrect and we will then suggest a scheme for partial classification of three-dimensional object surfaces.

Without any assumption about the location of the light source, we found from the above analytical results that the topographic labels on the underlying intensity surface resulting from a spherical Lambertian surface can only be peak, pit, ridge, valley, convex or concave hillside. This is because any combination of I_{xx}, I_{yy} and I_{xy} resulting from a spherical Lambertian surface can produce only either a semi-positive or semi-negative definite Hessian.

Therefore, not all combinations of I_{xx}, I_{yy} and I_{xy} are possible.

Furthermore, if we approximate three-dimensional surfaces locally by spherical surfaces, it is expected that the radii of the approximating spheres for points on a spherical surface are constant. Recall that the radius, r, and the eigenvalues, X_1 and X_2, of a spherical Lambertian surface are related by the following expressions:

$$X_1 = \frac{I_0 cr}{(r^2-x^2-y^2)^{3/2}} \quad \text{and} \quad X_2 = \frac{I_0 c}{r(r^2-x^2-y^2)^{1/2}} \ .$$

We obtain from the expression of X_1

$$(r^2-x^2-y^2)^{1/2} = \frac{I_0 c}{X_2 r} \ .$$

We then have from the expression of X_2

$$X_1 = I_0 cr \left(\frac{I_0 c}{X_2 r}\right)^{-3} ,$$

or

$$r^4 = \frac{X_1 I_0^{\,2} c^2}{X_2^{\,3}} \ .$$

Since I_0 and c are fixed, we conclude that $X_1/X_2^{\,3}$ is constant for a spherical Lambertian surface. Therefore, an image point can be determined as resulting from a point on a spherical Lambertian surface only if it is labeled as a peak, pit, ridge, valley, convex hillside or concave hillside, and the radii of the approximating spheres at pixels within the neighborhood around that point are similar enough. We thus showed that Pentland's claim is incorrect. What this suggests is that we should estimate surface tilt and surface slant locally from the eigenvalues and eigenvectors of the Hessian of the underlying intensity surface only if the underlying intensity surface is compatible with that of a spherical Lambertian surface.

It can be observed from the expressions of the eigenvalues that a pit, valley and convex hillside classification of the intensity surface of a spherical Lambertian surface corresponds to a positive c. This implies a light source below the object surface. Although this is physically possible, such illuminating condition can usually be ignored when solving practical problems. We thus further assume that a spherical Lambertian surface can only result in peak, ridge and concave hillside classifications.

5.4 CLASSIFICATION OF OBJECT SURFACES

We will propose here a scheme for partial classification of three-dimensional object surfaces. The basic goal of this classification scheme is to group together pixels which are likely to come from the same surface patch. We will limit our consideration to five types of object surfaces. They are namely, planar, developable[1], spherical, elliptical, and hyperbolic surfaces.

Based on the above discussions, it is evident that topographic labels together with the signs and magnitudes of their second directional derivatives bear a strong relationship to the nature of the three-dimensional object surface in the scene. This evidence leads to the assumption that maximally connected sets of pixels having the same topographic label belong to the same surface patches. To extract these connected sets of topographic structures, the feature extraction process described in Section 4.2 of Chapter 4 is employed. The resulting structures of this feature extraction process are arcs, regions and topographic labels. The desired topographic structures are then determined by applying a connected components algorithm [Lu83] to the topographic labels within each region segment.

The assembled topographic structures may be divided into three categories: (1) areal structures which consist of convex hillsides, concave hillsides, saddle hillsides, flat surfaces, and sloped surfaces; (2) arc structures which include edges, ridges, and valleys; and (3) point structures

---

[1] A developable surface is one which may be unrolled onto a plane without distortion.

which include peaks, pits, and saddle points. In what follows, we suggest hypotheses that can be made about the three-dimensional objects based on the analytical results that we have derived and the results of the experiments that we have performed. We believe that three-dimensional object shape can be inferred by feeding this knowledge into a hypothesis based reasoning system such as [Mu84].

5.4.1 Areal Structures

Flat

A flat is a simple surface with zero gradient and no curvature. That is, the gray level intensity is constant in a connected flat structure. Since the surface normal vectors within a planar surface are constant, we can be almost certain that pixels belonging to a connected flat structure come from the same planar surface. Although this may not hold for shadow areas, we can usually separate shadow areas by identifying flat structures with relatively low intensity averages.

Hillsides

We first hypothesize that a concave/convex hillside assembly is part of a spherical, elliptical or developable surface, and a saddle hillside assembly is part of a hyperbolic surface. Our first hypothesis is driven by the analytical and experimental results of the cylindrical, spherical, elliptical and hyperbolic surfaces that we have considered in Section 5.2.2.

We further postulate that a concave/convex hillside assembly is part of a developable surface if it is adjacent to a straight and horizontal ridge. In particular, it is part of a cylindrical surface if it is concave and the second directional derivative of the hillside in the direction of the ridge is zero.

As a result of Section 5.2.3, a hillside assembly can be considered as part of a spherical surface if it is concave and the radii of the approximating spheres within the hillside assembly are similar enough.

We have not been able to derive a complete classification scheme for all the areal structures. Nonetheless, since the assembled regions are likely to come from the same surface patches, they are good starting regions for the shape-from-shading method described in Chapter 2.

5.4.2 Line and Point Structures

While edges are considered to be good indications of surface discontinuities, peaks and ridges are found to be significant structures in the images of the conic surfaces that we have considered. The following observations are gathered from the topographic structures of the conic surface:

1. The ridge arcs obtained from the images of the sphere are found to be symmetrical around the peaks.
2. Ridges for the images of the ellipsoid and the hyperboloid are found to be symmetrical around the peaks only if the projection of the light vectors are parallel to one of the axes of these conic surfaces.
3. Straight ridge lines are found in the images of the cylinder. The gray tone intensities along the ridge lines are found to be constant.
4. While the ridges around the peaks found in the image of the ellipsoid curve away from the light sources, those of the hyperboloid curve towards the light sources.
5. The peaks located in the images of the conic surfaces correspond to locations where the surface normals are pointing towards the light sources.

A pictorial illustration of the above observations are given in Figure 51.

In this section, we have listed a set of criteria that can be used to infer shapes of three-dimensional objects. These criteria are supported by the experimental results to be described in Section 5.5. We believe that an integrated knowledge based vision system can be benefited by effectively using the knowledge about these topographic structures.

5.5 RESULTS

In sections 5.5.1-5.5.3, we show the analytical and experimental results of the topographic patterns on the cylinder and sphere of Figures 47 and 48. Three illumination conditions are considered for each surface: (1) the light direction is $(0,0,-1)$ which means directly above the center of the surface; (2) the light direction is $(0,3^{1/2}/2,-1/2)$ which translates to azimuth 0 degrees and elevation 30 degrees; (3) the light direction is $(1/2,1/2,-1/2^{1/2})$ which translates to azimuth 45 degrees and elevation 45 degrees. The illuminated surfaces of the cylinder and the sphere are shown in Figure 52 and Figure 53 respectively.

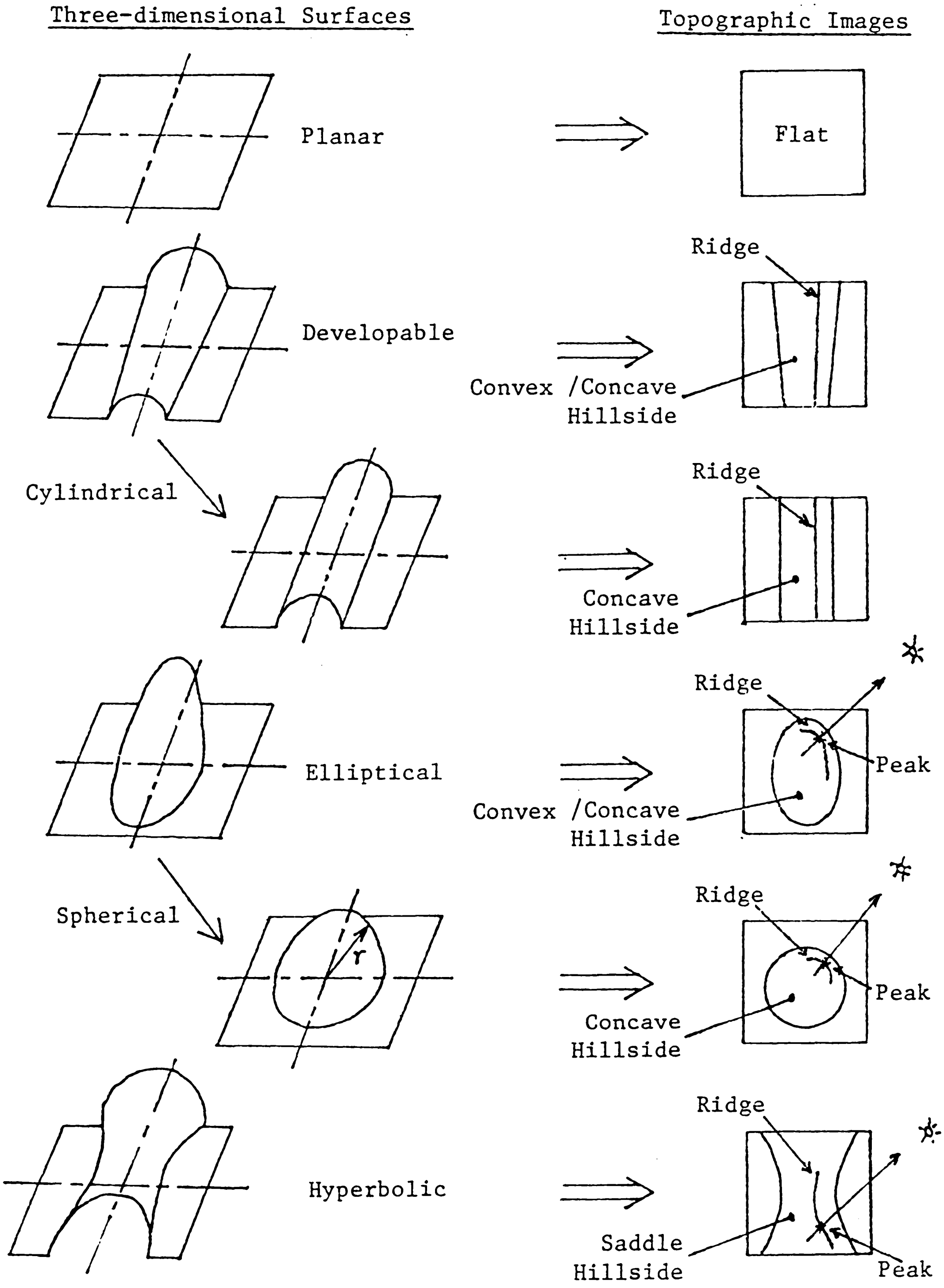


Figure 51: An illustration of the topographic patterns.

5.5.1 Analytical Results for the Cylinder

When the light direction is from azimuth 0 degrees, elevation 90 degrees, analytical results in Section 5.2.2.1 indicate a ridge parallel to the axis of the cylinder and running along the center of the top half as shown in Figure 54a. When the light direction is from azimuth 0 degrees and elevation 30 degrees, the ridge appears as in Figure 54b. When the light direction is from azimuth 45 degrees and elevation 45 degrees, the ridge appears as in Figure 54c. In all three cases, the remaining points of the cylinder are hillsides.

5.5.2 Analytical Results for the Sphere

When the light source is directly above the center of the sphere, the gradient magnitude is zero at (0,0), therefore, a peak is located at the center of the sphere. The gradient magnitude is positive and the first directional derivative in the direction w^2 is zero at the remaining points of the sphere. It follows from our analytical results that ridges locate at these points[1].

---

[1] This result is due to the analysis used in [Ha83] and may not be intuitively obvious to all readers.

When the light direction is $(0,3^{1/2}/2,-1/2)$, a peak is found at $(0,3^{1/2}r/2)$. At the remaining points,

$$\nabla I.w^1 = 0 \text{ when } (x^2+y^2) = 3^{1/2}/2y(r^2-x^2-y^2)^{1/2}$$

and $\nabla I.w^2 = 0$ when $x = 0$.

Therefore, there are ridges when either one of the above two equations is satisfied and hillsides otherwise.

Similarly, a peak is found at $(r/2,r/2)$ when the light direction is $(1/2,1/2,-1/2^{1/2})$. Ridges can be located at places where either

$$2^{1/2}(x^2+y^2) = (x+y)(r^2-x^2-y^2)^{1/2}$$

or $x = y$ is satisfied.

At the remaining points, hillsides are the correct categories. Figure 55 shows the topographic labels for the illuminated spheres.

5.5.3 Experimental Results

Experimentally, we are working in the GIPSY (General Image Processing System) environment [Kr83]. There currently exist GIPSY commands to construct three-dimensional surfaces, to produce images of these surfaces from various viewpoints and light directions, to fit these images with either cubic polynomials, splines, or discrete cosines, and to calculate the topographic labelings. Figures 56 and 57 show experimental results for the cylinder and sphere using cubic polynomial surface fitting. Experimental results show very good correspondence with the analytical results, except for the sphere when the light direction is (0,0,-1). In this case, when points are labeled ridge and have neighboring points in a direction orthogonal to the gradient that are also labeled ridge, our software reclassifies these ridge continuums as hillsides. A detailed discussion of ridge and valley continua can be found in [Ha83].

In additional to the images of the two simple surfaces, a synthetic image of a more complex surface was also used in testing. The surface of Figure 58 is composed of cylindrical and spherical surface patches. Figure 59 shows the image of the surface when illuminated from azimuth 45 degrees and elevation 45 degrees. Figure 60 illustrates the topographic labels that resulted from the experimental method. Most of the resulting topographic labels are located at places where they are predicted by the analytical method.

Our results show that the most informative features found in the images of the cylinder and sphere are ridges and peaks. While the ridges found in the cylinder images are intuitive, the ellipse-like ridges found in the sphere images are unexpected. Although most ridge points found in the sphere images are weak ridges, experimental results show that these ridges are detectable. These ellipse-like ridges will be a definite clue to 3-dimensional surface identifications. Once the shape of the 3-dimensional surface is hypothesized as cylindrical or spherical, information such as the direction of the light source and the cylinder/sphere radius may also be estimated by examining the topographic labels.

5.5.4 Results for the other Conic Surfaces

Experiments have been performed on the ellipsoid and hyperboloid of Figures 49 and 50. Shaded images of these conic surfaces are shown in Figures 61 and 62. Figures 63 and 64 illustrate the topographic labels that resulted from the analytical equations given in the appendix. Figures 65 and 66 show the topographic labels that resulted from the facet surface fitting model. Results show that the dominant structures in the images of the ellipsoid and hyperboloid are respectively concave and saddle hillsides. It was also found that only portions of the ridges that appear in Figures 63 and 64 were detected by the experimental method. Careful examinations showed that the ridges detected by the experimental method were dominant ridges which gave high curvatures (or eigenvalues), and the undetected ridges were weak ridges which gave very small curvatures. It was also observed that the intersection of the two ridge lines produced a peak.

To examine the performance of the facet fitting topographic labeling scheme on noise images, random noise of mean=0 and standard deviation=100 was added to the shaded images of an ellipsoid. The gray level intensities of the shaded images range from 0 to 1000. The noisy images are shown in Figure 67. Figure 68 shows the topographic labels that resulted from the facet surface fitting model. It was found that the ridge labels detected were fragmented into small pieces. In order to group the fragmented ridges into more meaningful structures, a modification of the arc segment extraction procedure as described in Chapter 4 was used to link the fragmented ridges. The modified procedure differs from the original procedure in that linking is done based on ridge orientations instead of edge orientations. In particular, ridge orientation is defined to be the direction perpendicular to the eigen direction which gives the first directional derivative zero-crossing. Results of the ridge linking process are shown in Figure 69. It was also found that the detected ridges became unrecognizable when more noise was added to these images.

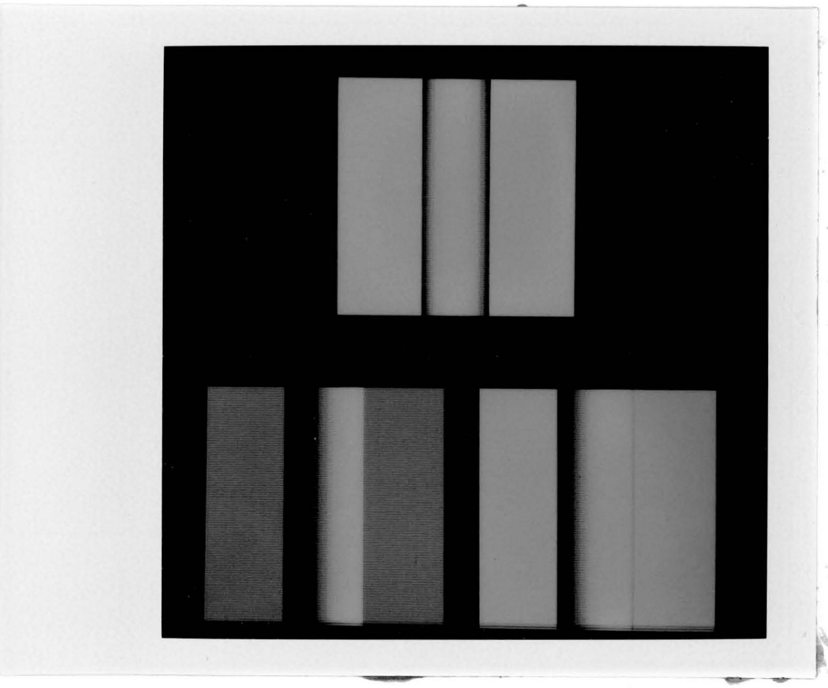


Figure 52: Shaded images of the cylinder of Figure 47.

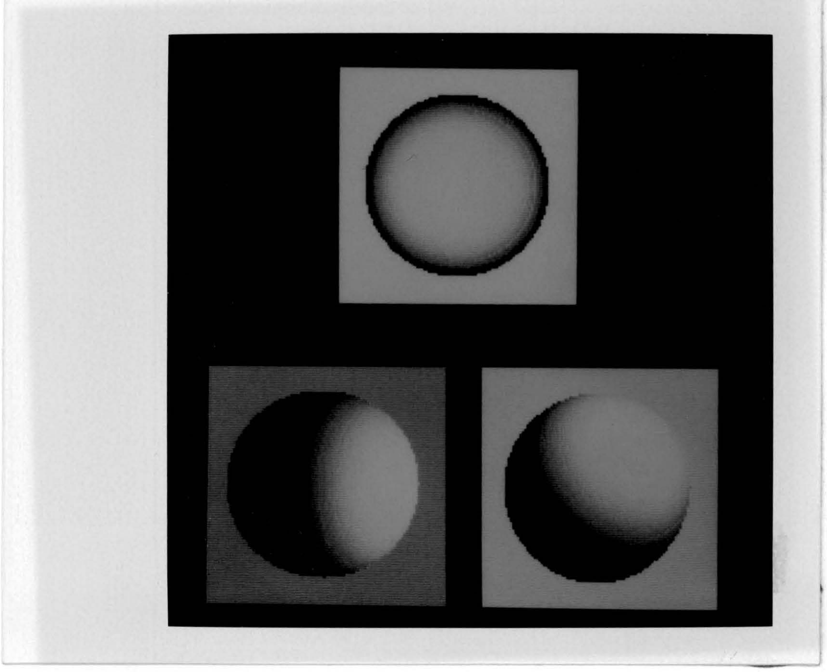


Figure 53: Shaded images of the sphere of Figure 48.

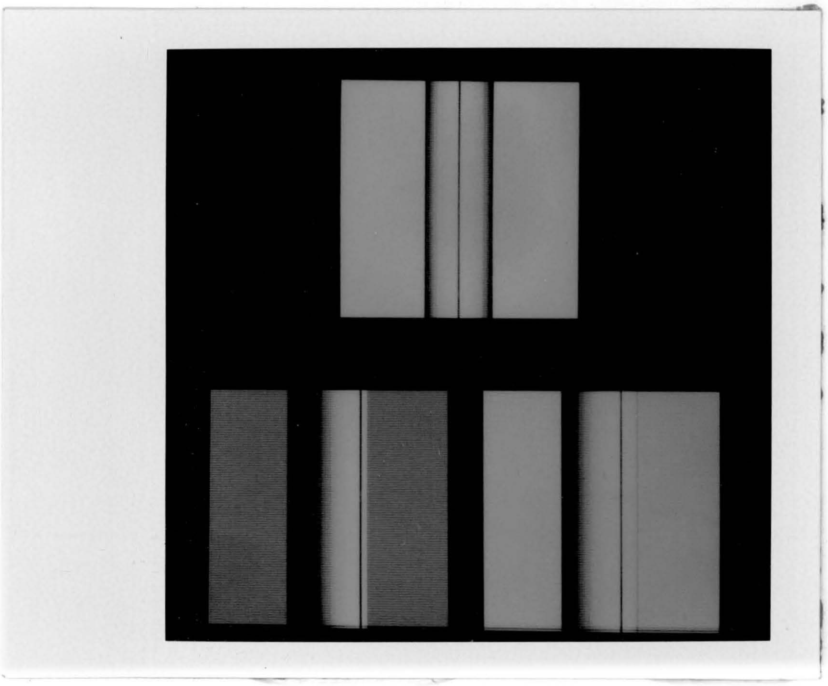


Figure 54: The analytically derived topographic labeling of the cylinder.

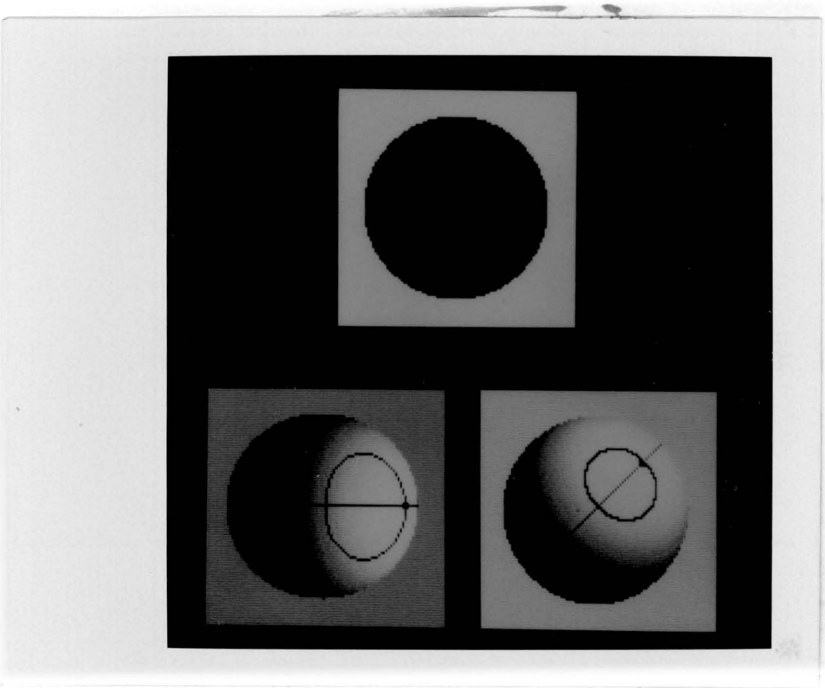


Figure 55: The analytically derived topographic labeling of the sphere.

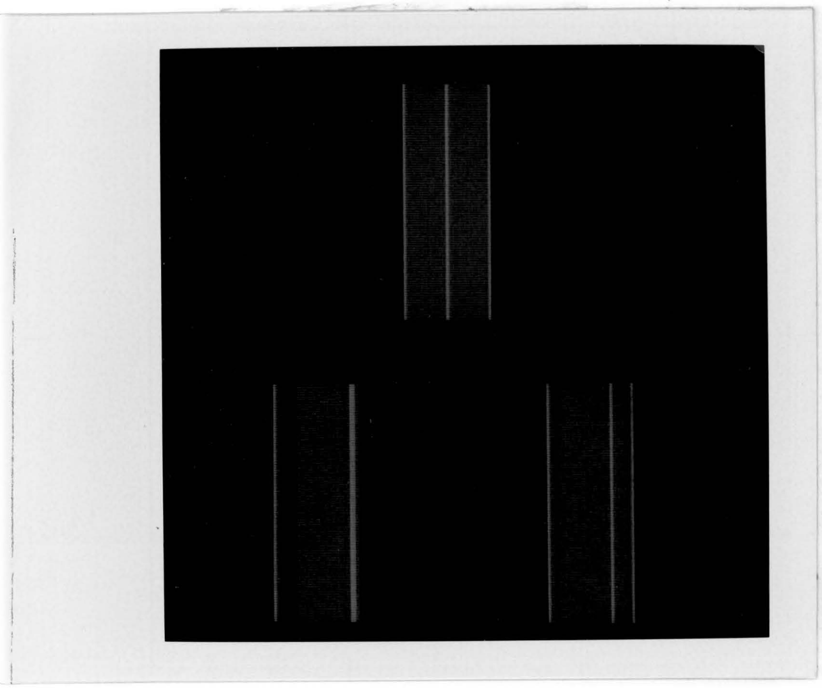


Figure 56: The experimental results for the cylinder.

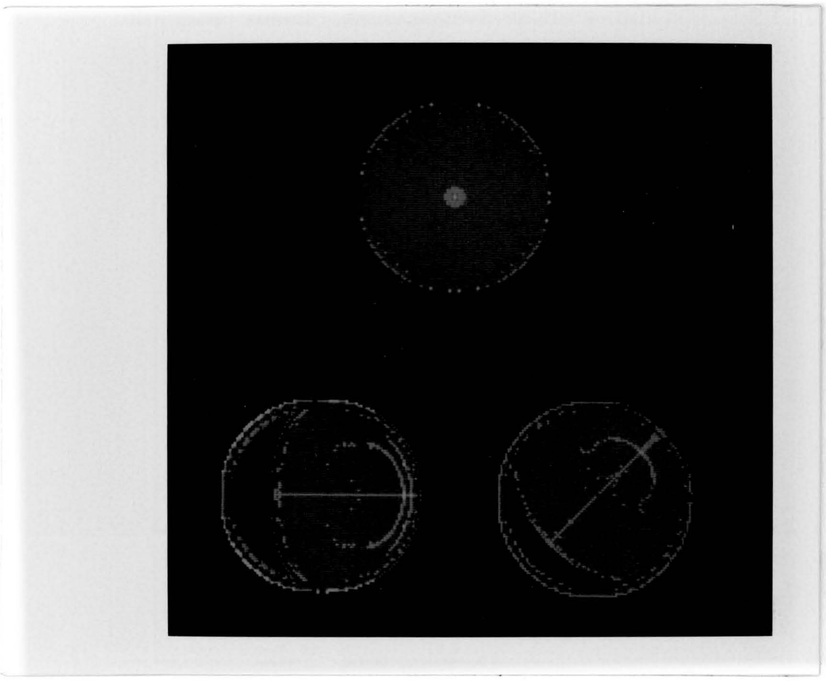


Figure 57: The experimental results for the sphere.

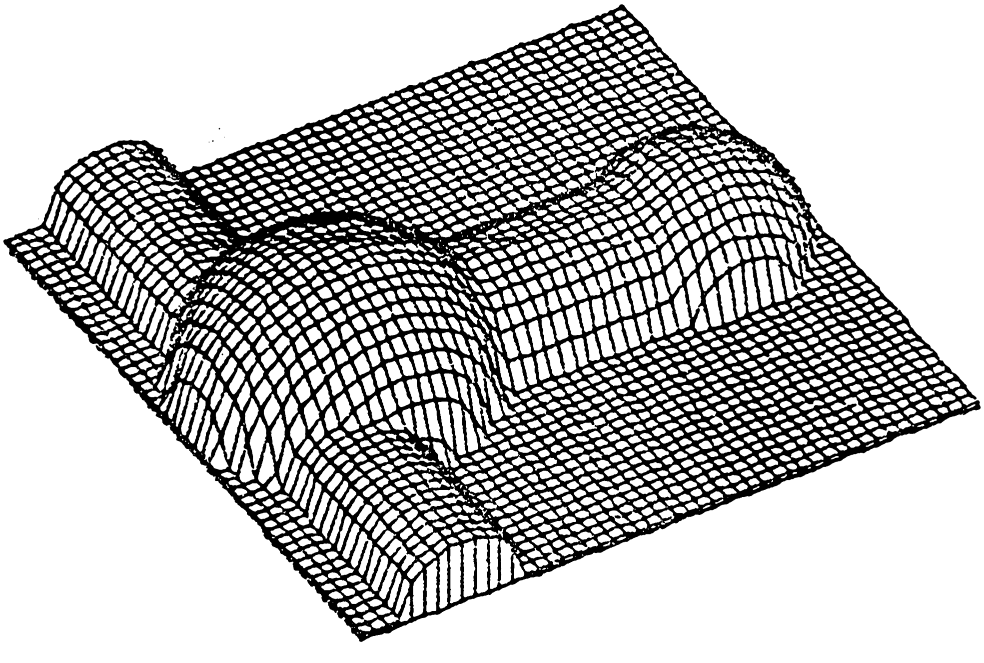


Figure 58: A synthetic three-dimensional object.

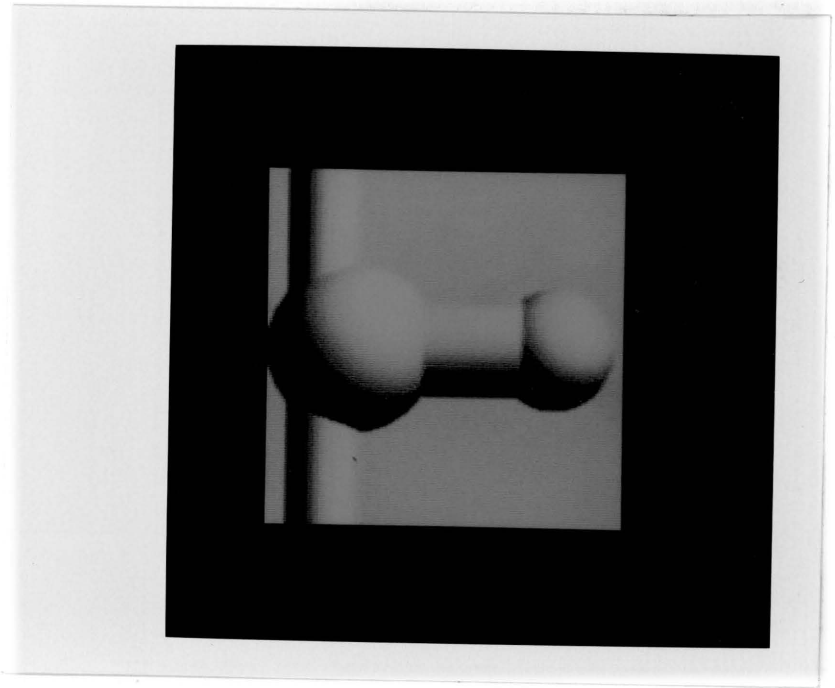


Figure 59: Gray tone image of the synthetic object illuminated from azimuth 45 degrees and elevation 45 degrees.

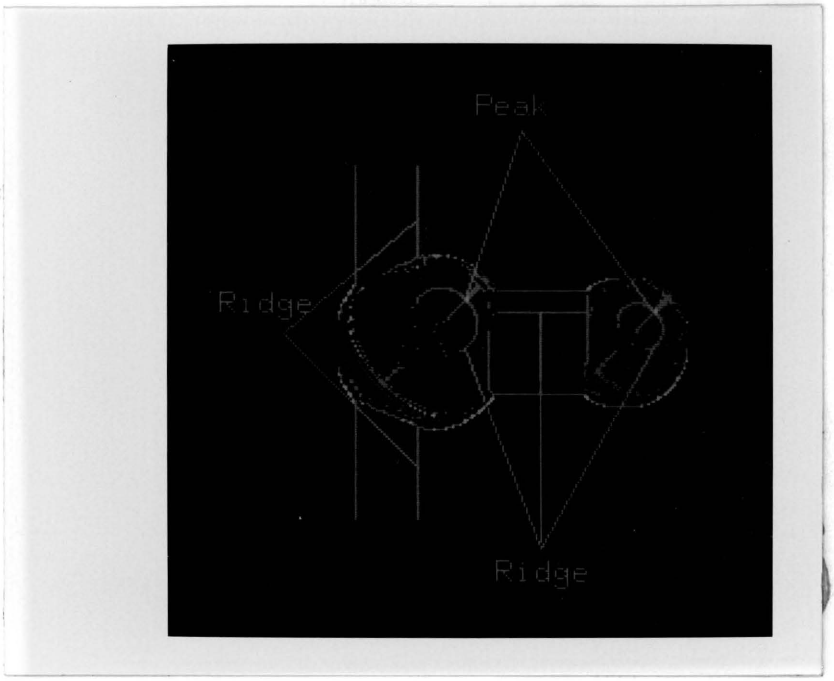


Figure 60: The topographic labeling of the synthetic object image.

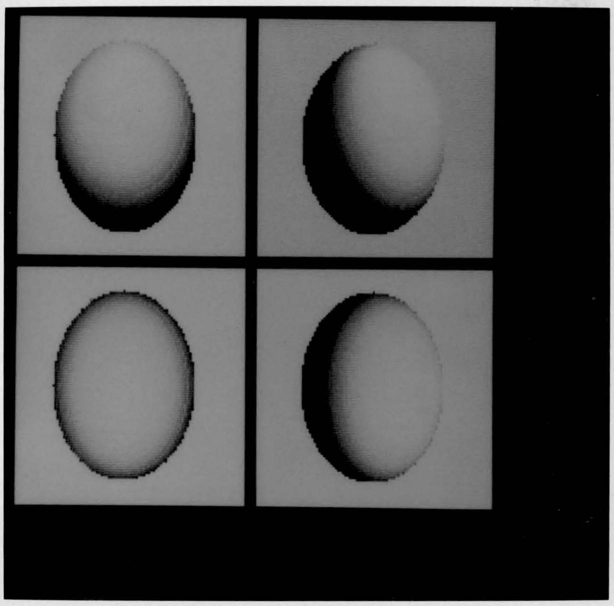


Figure 61: Shaded images of the ellipsoid of Figure 49.

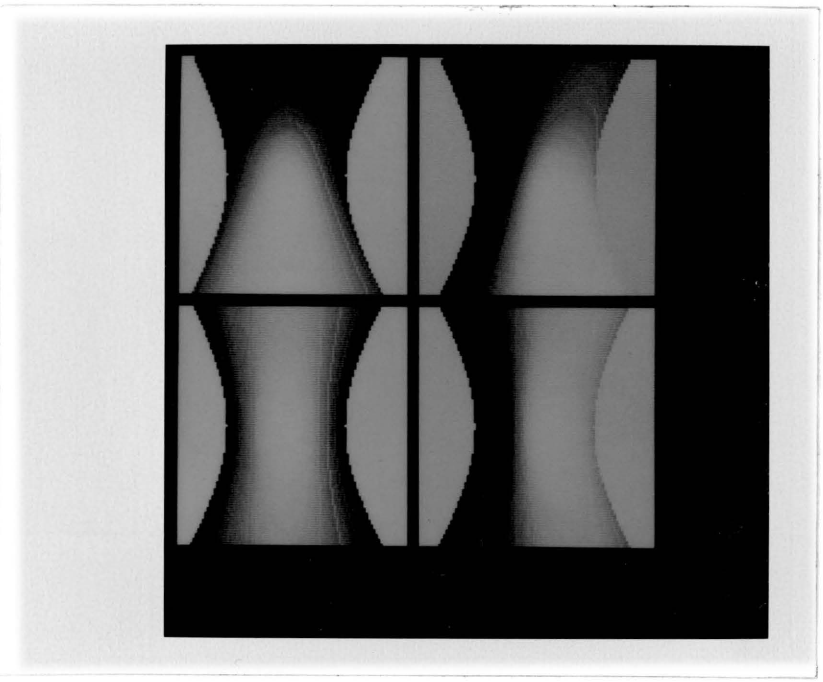


Figure 62: Shaded images of the hyperboloid of Figure 50.

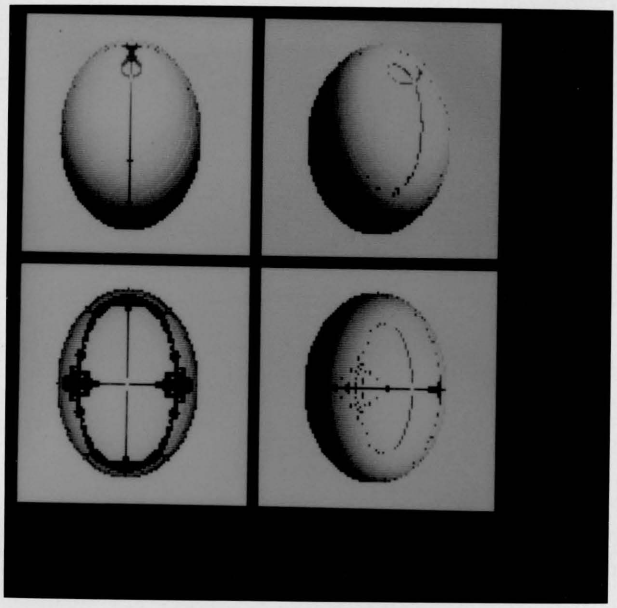


Figure 63: The analytically derived topographic labelings of the images of the ellipsoid.

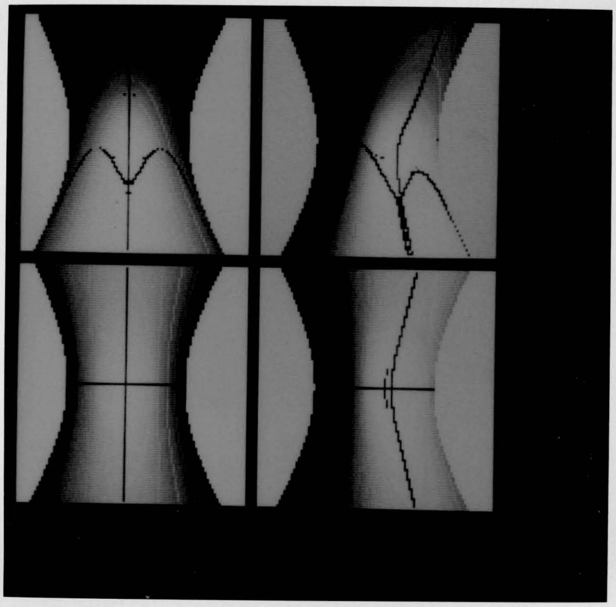


Figure 64: The analytically derived topographic labelings of the images of the hyperboloid.

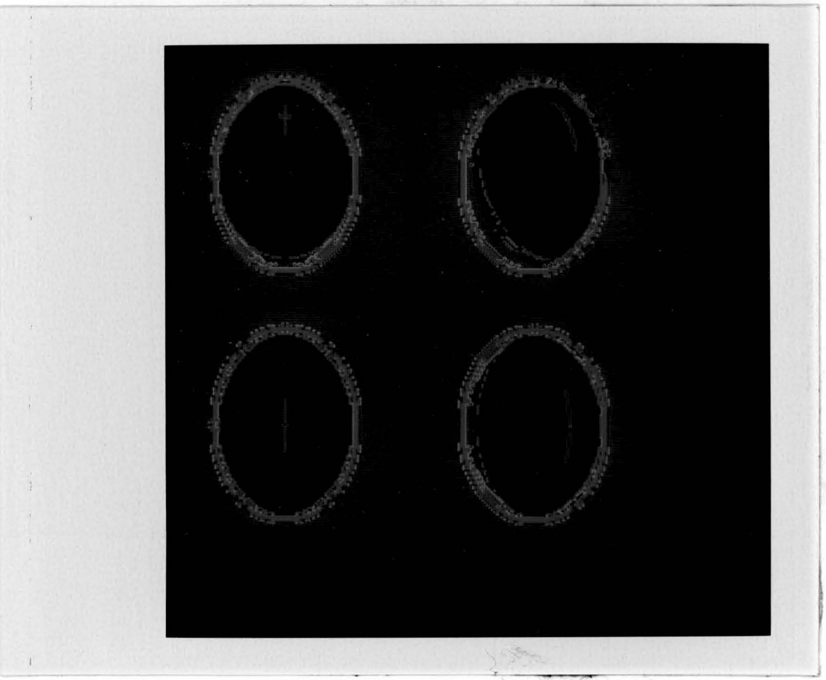


Figure 65: The experimental results of the images of the ellipsoid.

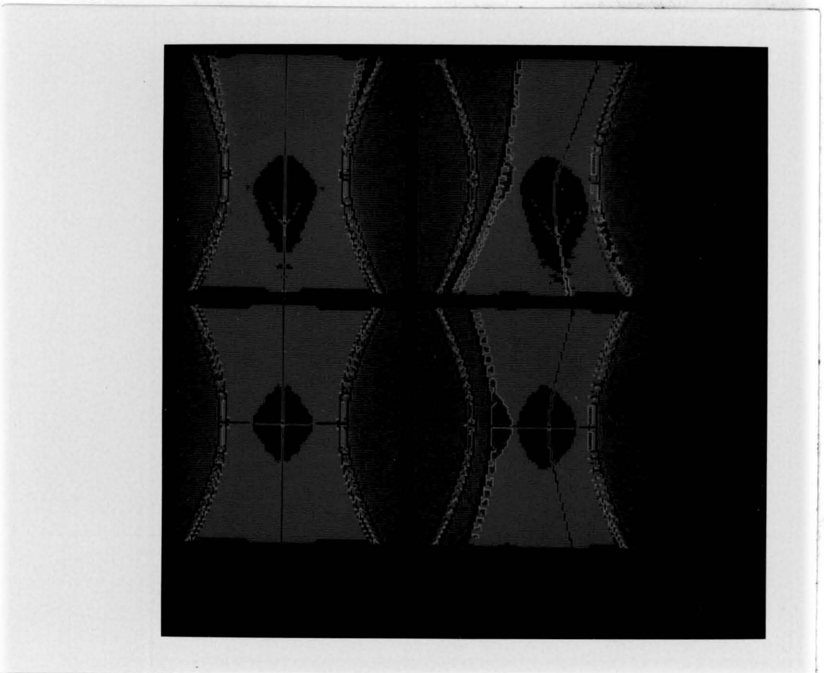


Figure 66: The experimental results of the images of the hyperboloid.

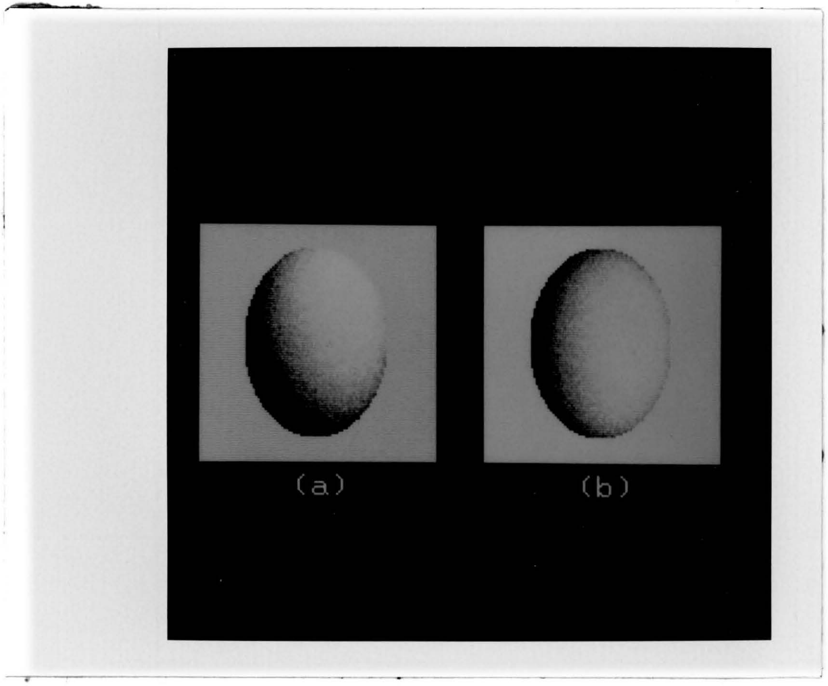


Figure 67: Noisy images of an ellipsoid.

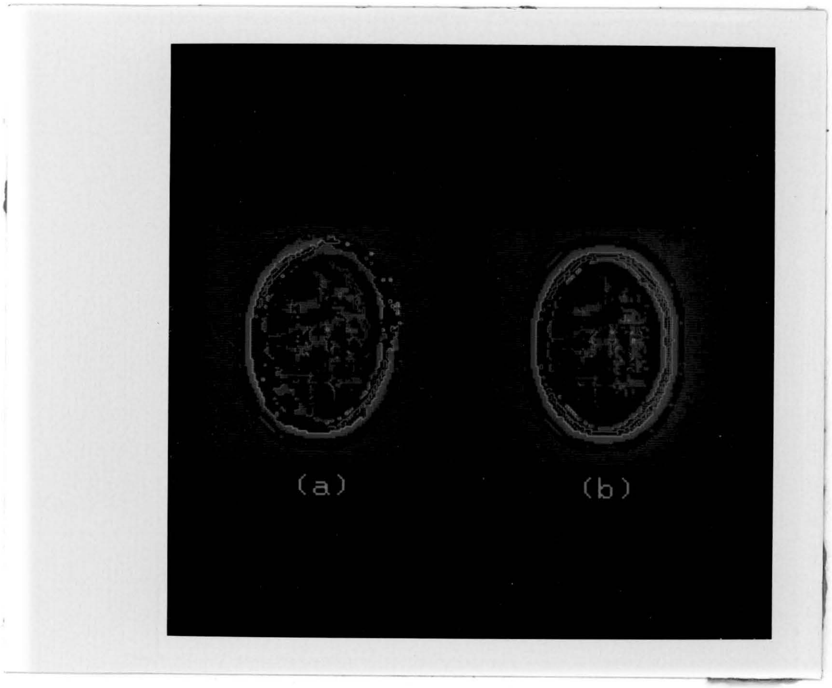
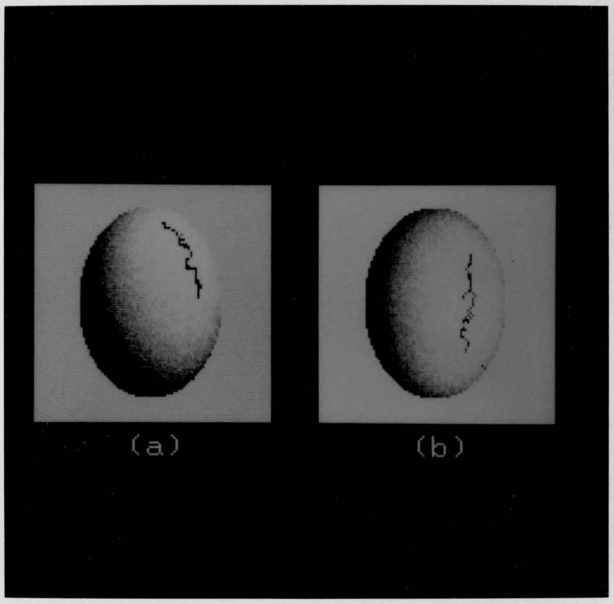


Figure 68: The experimentally derived topographic labelings of the images of Figure 67.



(a) (b)

Figure 69: Results of the ridge linking process.

5.6 SUMMARY

Both the analytical and experimental results so far indicate that there are definite patterns emerging that can help indicate the shape of the original three-dimensional surface and the direction of the light source. For both methods, we anticipate a great deal more work. We would like to carry out the analytic approach for several more simple surfaces, perhaps using a symbol manipulator to help solve the equations. For the experimental methods, we need to work on getting results as close as possible to the analytic results. We will then perform a large series of experiments with various surfaces, combinations of surfaces forming objects, viewpoints, reflectance functions, and lighting conditions. Only then will we be able to complete the work of analyzing the patterns of topographic labels produced and predicting three-dimensional shape from these patterns.

Chapter VI

CONCLUSIONS

Three fundamental problems in computer vision are addressed in this dissertation. The first deals with the problem of how to extract and assemble a rich symbolic representation of the gray level intensity changes in an image. Although there are many approaches available for feature extraction, none of them can result in a rich enough description of the gray level intensity changes in an image. A facet model based feature extraction scheme is proposed in Chapter 4. The scheme extracts edge and region (non-edge) segments as prominent structures in an image. While an edge segment is described by its orientation and contrast, intensity variations within a region segment are represented by the topographic primal sketch. Furthermore, the object surface classification scheme proposed in Chapter 5 groups the topographic structures within each region segment into topographic assemblies. The resulting topographic assemblies are found to be likely to come from the same surface patches.

The second problem addressed deals with the interpretation of the resulting structures as three-dimensional object surfaces. Three different shape modules are described in Chapters 3-5. Surface orientations within topographic assemblies can be estimated by the shape-from-shading methods described in Chapter 3. Depth information can be obtained by matching topographic structures by the stereo matching procedure described in Chapter 4. A rough estimate of surface orientation can also be computed by topographic analysis of conic surfaces.

Finally, mechanisms for interaction among different sources of information obtained from different shape modules are studied. The study of interactions among different shape modules is an area in which very little research results are available. Shape modules used by previous investigators for determining intrinsic scene characteristics have generally been considered to be independent of one another. Although the shape modules can act independently, it is important to maintain consistency and contiguity among the data computed from the different modules. It is demonstrated in Chapter 3 that the shape-from-shading techniques can be enhanced by utilizing the boundary constraints provided by the stereo matching process and/or the analysis of the topographic primal sketch. A more complete analysis of the interactions among shape modules is an area that needs future research.

In this dissertation, some of the ways in which the facet model and the topographic primal sketch can be used for the recovery of intrinsic scene characteristics have been investigated. These methods have been successfully tested on a number of images. However, this success is still to be regarded only as a promising step towards a unified approach to computer vision.

BIBLIOGRAPHY

[Ar83] Arnold, R.D., "Automated Stereo Perception", Stanford University, Ph.D. Dissertation, March 1983.

[Ba81] Baker, H.H. and T.O. Binford, "Depth from Edge and Intensity Based Stereo", Proc. Seventh International Joint Conference on Artificial Intelligence, Vancouver, Canada, August 1981, pp.631-636.

[Ba80] Barnard, S. and W. Thompson, "Disparity Analysis of Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 1980, pp.333-340

[Ba82] Ballard, D.H. and C.M. Brown, Computer Vision, Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

[Ba78] Barrow, H. and J.M. Tenenbaum, "Recovery of Intrinsic Scene Characteristics from Images", Computer Vision Systems, A. Hanson and E. Risemen (ed.), Academic Press, New York, 1978, pp.3-26.

[Ba81] Barrow, H. and J.M. Tenenbaum, "Interpreting Line Drawings as Three-dimensional Surfaces", Artificially intelligence, Vol. 17, 1981, pp.75-116.

[Br84] Brady, M. and A. Yuille, "An extremum Principle for Shape from Contour", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, May 1984, pp.288-301.

[Bu84] Burns, J.B., A.R. Hanson and E.M. Riseman, "Extracting Straight Lines", Proc. Seventh International Joint Conference on Pattern Recognition, Montreal, August 1984, pp.482-485.

[Ca83] Canny, J.F., "Finding Edges and Lines in Images", Massachusetts Institute of Technology, Master Thesis, June 1983.

[Ch82] Cheng, J.K. and T.S. Huang, "Image Registration by Matching Relational Structures", Proc. Sixth International Joint Conference on Pattern Recognition, 1982, pp.354-356.

[Da75] Davis, L.S., "A Survey of Edge Detection Techniques", Computer Graphics and Image Processing 4, 1975, pp.248-270.

[Dr81] Draper, N.R. and H. Smith, Applied Regression Analysis, 2nd ed., Wiley, New York, 1981.

[Eh78] Ehrich, R.W. and J.P. Foith, "Topology and semantics of intensity arrays", Computer Vision Systems, A. Hanson and E. Risemen (ed.), Academic Press, New York, 1978, pp.111-128.

[Go83] Goshtasby, A., "A Symbolically-assisted Approach to Digital Image Registration with Application to Computer Vision", Michigan State University, Ph.D. Dissertation, 1983.

[Ge77] Gennery, D.B., "A stereo Vision System for an Autonomous Vehicle", Proc. Fifth International Joint Conference on Artificial Intelligence, MIT, Cambridge, Mass., August 1977, pp.576-582.

[Gr81] Grimson, W., From Images to Surfaces, MIT Press, Cambridge, Mass., 1981.

[Gr83] Grimson, W.E.L., "An Implementation of a Computational Theory of Visual Surface Interpolation", Computer Vision, Graphics and Image Processing 22, 1983, pp.39-69.

[Ha80] Haralick, R.M., "Edge and Region Analysis for Digital Image Data", Computer Graphics and Image Processing, Vol. 12, 1980, pp.60-73.

[Ha81] Haralick, R.M. and L.T. Watson, "Facet Model for Image Data", Computer Graphics and Image Processing, Vol. 15, 1981, pp.113-129.

[Ha83] Haralick, R.M. and Y. Yasuoka, "Peak Noise Removal by a Facet Model", .us Pattern Recognition, Vol 16, No. 1, 1983, pp.23-29.

[Ha83] Haralick, R.M, L.T. Watson and T.J. Laffey, "The Topographic Primal Sketch", The International Journal of Robotics Research, Vol. 2, No. 1, 1983, pp.50-72.

[Ha83] Haralick R.M. and J.S. Lee, "The Facet Approach to Optic Flow", Proc. Image Understanding Workshop Washington, D.C., June 1983.

[Ha83] Haralick,R.M., "Ridges and Valleys on Digital Images", Computer Vision, Graphics, and Image Processing, Vol. 22, 1983, pp.28-38.

[Ha84] Haralick, R.M., "Digital step edges from Zero Crossing of Second Directional Derivatives", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, Jan. 1984, pp.58-68.

[Ha83] Harrington, S., Computer Graphics: a Programming Approach, McGraw-Hill, New York, 1983.

[Ho75] Horn, B.K.P., "Obtaining Shape from Shading Information", The Psychology of Computer Vision, P.H. Winston (ed.), McGraw-Hall, New York, 1975, pp.115-155.

[Ho77] Horn, B.K.P., "Understanding image intensities", Artificial Intelligence, Vol. 8, 1977, pp.201-231.

[Ho81] Horn, B.K.P., "Determining Optical Flow", Artificial Intelligence, Vol. 17, 1981, pp.185-203.

[Hu83] Huertas, A., "Using Shadows in the Interpretation of Aerial Images", University of Southern California, ISG Report 104, R. Nevatia (ed.), Oct. 1983, pp.55-84.

[Ik81] Ikeuchi, K. and B.K.P. Horn, "Numerical Shape from Shape from Shading and Occluding Boundaries", Artificial Intelligence, Vol. 17, 1981, pp.141-184.

[Ik81] Ikeuchi, K., "Determining Surface Orientation of Specular Surfaces by Using the Photometric Stereo Method", IEEE Transactions of PAMI, vol. 3, No. 6, 1981, pp.661-669.

[Ke82] Kender, J.R., "A Computational Paradigm for Deriving Local surface Orientation from Local Textural Properties", Proc. IEEE Workshop on Computer Vision, Rindge, NH, August 1982, pp.143-152.

[Kr83] Krusemark, S. and R.M. Haralick, "An Operating System Interface for Transportable Image Processing Software", Computer Vision, Graphics, And Image Processing, Vol. 23, No. 1, July 1983, pp. 42-66.

[Lu82] Lumia, R., "A new connected components algorithm for virtual memory computers", Proceedings IEEE Conference on Pattern Recognition and Image Processing, 1982, pp.560-565.

[Ma76] Marr, D., "Early Processing of Visual Information", Phil. Trans. Royal Society of London, B, 275, 1976, pp.483-534.

[Ma76] Marr, D and T. Poggio, "A Theory of Human Stereo Disparity", Science 194, 1976, pp.283-287.

[Ma77] Marr, D., "Analysis of Occluding Contour", Proc. Royal Society of London, B, 197, 1977, pp.441-475.

[Ma80] Marr, D and E.C. Hildreth, "Theory of Edge Detection", Proc. Royal Society of London, B, 207, 1980, pp.187-217.

[Ma82] Marr, D., Vision: A Computational Investigation into The Human Representation and Processing of Visual Information, Freeman, San Francisco, CA, 1982.

[Me83] Medioni, G.G. and R. Nevatia, "Segment-based Stereo Matching", University of Southern California, ISG Report 104, R. Nevatia (ed.), Oct. 1983, pp.38-54.

[Mu84] Mulgaonkar, P.G., "Analyzing Perspective Line Drawings Using Hypothesis Based Reasoning", Virginia Polytechnic Institute and State University, Ph.D. Dissertation, Aug. 1984.

[Mo77] Morvavec, H.P., "Towards Automatic Visual Obstacle Avoidance", Proc. Fifth International Joint Conference on Artificial Intelligence, MIT, Cambridge, Mass., August 1977, pp.584.

[Ne82] Nevatia, R., Machine Perception, Prenttice-Hall, Englewood Cliffs, New Jersey, 1982.

[Pe82] Pentland, "Local Analysis of the Image: Limitations and Uses of Shading", Proceedings of the Workshop on Computer Vision: Representation and Control, Aug. 1982, pp. 153-161.

[Pe82] Pentland, A., "Local Shading Analysis", SRI International, Technical Note 272, Nov. 1982.

[Pe79] Perkins, W.A., "Region Segmentation of Images by Expension and Contraction of Edge Points", Proceedings of the Sixth International Joint Conference on Artificial Intelligence, Tokyo, Japan, 1979, pp.699-701.

[Po84] Pong, T.C., L.G. Shapiro, L.T. Watson and R.M. Haralick, "Experiments in Segmentation Using a Facet Model Region Grower", Computer Vision, Graphics, and Image Processing, Jan. 1984, pp.1-23.

[Ro65] Roberts, G., "Machine perception of three-dimensional solids", Optical and Electroptical Information Processing, J.T. Tippett, et. al. (eds.), MIT Press, Cambridge, Mass., 1965, pp.179-197.

[Sh83] Shafer, S.A. and T. Kanade, "Using Shadows in Finding Surface Orientations", Computer Vision, Graphics, and Image Processing 22, 1983, pp.145-176.

[Sh83] Shapiro, L.G. and R.M. Haralick, "A hierarchical Relational Model of the Bulkhead", Technical Report, Dept. of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, August 1983.

[Sh83] Shapiro, L.G. and R.M. Haralick, "A hierarchical Relational Model for Automated Inspection Tasks", Proceedings of The International Conference on Robotics, Atlanta, Georgia, March 1984, pp.70-77.

[Sm83] Smith, G.B., "The Relationship between Image Irradiance and Surface Orientation", Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, June 1983, pp.14-19.

[St81] Stevens, K.A. "The Visual Interpretation of Surface Contours", Artificial intelligence, Vol. 17, 1981, pp.47-73.

[Te83] Terzopoulos, D., "Multilevel Computational Process for Visual Surface Reconstruction", Computer Vision, Graphics and Image Processing 24, 1983, pp.52-96.

[Ul79] Ullman, S., The Interpretation of Visual Motion, MIT Press, Cambridge, Mass., 1979.

[Wi81] Witkin, A.P. "Recovering Surface Shape and Orientation from Texture", Artificial intelligence, Vol. 17, 1981, pp.17-45.

[Wo79] Woodham, R.J., "Relating Properties of Surface Curvature to Image Intensity", Proceedings of the Sixth International Joint Conference on Artificial intelligence, Tokyo, Japan, 1979, pp971-977.

[Wo80] Woodham, R.J., "Photometric Method for Determining Surface Orientation from multiple images", Optical Engineering, Vol. 19, No. 1, 1980, pp.139-144.

[Wo81] Woodham, R.J., "Analyzing Images of Curved Surfaces", Artificial intelligence, Vol. 17, 1981, pp.117-140.

[Zu83] Zuniga, O.A. and R.M. Haralick, "Corner Detection using the Facet Model", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition", Washington, D.C., June 1983, pp.30-37.

Appendix A

PERSPECTIVE PROJECTION

Let S(u,v) be the distance of the surface down the z-axis as a function 2-dimensional image coordinates u and v. If the relation between scene coordinates and image coordinates is considered to be perspective projection, then the image intensity as a function of the image coordinates can be written as:

$$I(u,v) = I0 * \frac{a*Sx(u,v) + b*Sy(u,v) - c}{(1 + Sx^2(u,v) + Sy^2(u,v))^{1/2}} + n(u,v) \qquad (2)$$

where

$$Sx(u,v) = \frac{d\ S(u,v)}{dx} \quad \text{and} \quad Sy(u,v) = \frac{d\ S(u,v)}{dy} \ ,$$

n denotes the independent and identically distributed noise, $(a,b,c)^t$ is the unit vector for the illumination direction, and I0 is a constant depends on the surface albedo and the intensity of the light source. Furthermore, a smoothness constraint on the object surface is imposed in the following way. Let R be an arbitrary region over which equation (2) holds. We assume that

$$S(u,v) = \sum_{i=1}^{n} k_i \ h_i(u,v) \qquad \text{for } (u,v) \in R, \quad (3)$$

where hi's are some basis functions and ki's are constants. By using the intensity values $\{I(u,v) \mid (u,v) \in R\}$, we can determine a least squares estimator $\hat{\underline{k}}$ of $\underline{k}$ (note that $\underline{k}$ denotes the vector $[k1,k2,\ldots,kn]^t$).

The least squares approximation problem is to determine $\hat{\underline{k}}$ such that

$$e^2 = \sum_{(u,v) \in R} \left| I(u,v) - I0 * \frac{a*Sx(u,v) + b*Sy(u,v) - c}{(1 + Sx^2(u,v) + Sy^2(u,v))^{1/2}} \right|^2 \qquad (4)$$

is minimized. In order to find the value of $\hat{\underline{k}}$, we need first to determine Sx(u,v) and Sy(u,v). Recall from section 1.1 that the following relations hold for perspective propection:

$$u = \frac{fx}{z} \quad \text{and} \quad v = \frac{fy}{z},$$

where f is the focal length of the camera lens. Let D(x,y) be the distance of the surface down the z-axis as a function of scene coordinates x and y, then

$$D(x,y) = S(u,v) \quad \text{when} \quad u = fx/z \text{ and } v = fy/z.$$

Therefore, under the above condition

$$\frac{d\,S(u,v)}{d\,u} = \frac{d\,D(x,y)}{d\,u} = \frac{d\,D(x,y)}{d\,x} \cdot \frac{d\,x}{d\,u} + \frac{d\,D(x,y)}{d\,y} \cdot \frac{d\,y}{d\,u}$$

and

$$\frac{d\,S(u,v)}{d\,v} = \frac{d\,D(x,y)}{d\,v} = \frac{d\,D(x,y)}{d\,x} \cdot \frac{d\,x}{d\,v} + \frac{d\,D(x,y)}{d\,y} \cdot \frac{d\,y}{d\,v} \quad \text{hold.}$$

In matrix form, the above equations are equivalent to

$$\begin{vmatrix} Su(u,v) \\ \\ Sv(u,v) \end{vmatrix} = \begin{vmatrix} dx/du & du/du \\ \\ dy/dv & dy/dv \end{vmatrix} \begin{vmatrix} Sx(u,v) \\ \\ Sy(u,v) \end{vmatrix} . \qquad (5)$$

Consider now the expression $u = fx/S(u,v)$, by taking the partial derivative of both sides of this expression with respect to u, we obtain

$$1 = \frac{f\,[\,S(u,v)\ dx/du - x\ Su(u,v)]}{S^2(u,v)} .$$

After substituting x by $u*S(u,v)/f$ and some simplifications, we obtain the following expression for dx/du:

$$\frac{dx}{du} = \frac{S(u,v) + u\ Su(u,v)}{f} .$$

Similarly, by differentiating the same expression with respect to v, we obtain

$$\frac{dx}{dv} = \frac{u\ Sv(u,v)}{f} .$$

By repeating the same derivations for the expression $v = fu/S(u,v)$, the expressions for dy/du and dy/dv are found to be

$$\frac{dy}{du} = \frac{v\ Su(u,v)}{f}$$

and

$$\frac{dy}{dv} = \frac{S(u,v) + v\ Sv(u,v)}{f} .$$

By using the above results, we obtain from equation (5)

$$\begin{vmatrix} Sx(u,v) \\ \\ Sy(u,v) \end{vmatrix} = \begin{vmatrix} [S(u,v)+uSu(u,v)]/f & vSu(u,v)/f \\ \\ uSv(u,v)/f & [S(u,v)+vSv(u,v)]/f \end{vmatrix}^{-1} \begin{vmatrix} Su(u,v) \\ \\ Sv(u,v) \end{vmatrix}$$

After inverting the matrix, we obtain

$$\begin{vmatrix} Sx(u,v) \\ Sy(u,v) \end{vmatrix} = \frac{f}{S(u,v)+uSu(u,v)+vSv(u,v)} \begin{vmatrix} Su(u,v) \\ Sv(u,v) \end{vmatrix} . \qquad (6)$$

By using the expressions for Sx(u,v) and Sy(u,v), equation (4) can be rewritten as:

$$e^2 = \sum_{(u,v)\in R} \left| I(u,v) - \frac{I_0 \sum_{i=1}^{n} k_i[(af-cu)h_u^i+(bf-cv)h_v^i-ch_i]}{\left([\sum_{i=1}^{n} k_i(h_i+uh_u^i+vh_v^i)]^2+[f\sum_{i=1}^{n} k_ih_u^i]^2+[f\sum_{i=1}^{n} k_ih_v^i]^2\right)^{1/2}} \right. \qquad (7)$$

where

$$h_u^i = \frac{d\ h_i(u,v)}{du}, \qquad h_v^i = \frac{d\ h_i(u,v)}{dv}$$

and h_i denotes $h_i(u,v)$. Equation (7) is non-linear in the k's. One of the methods for solving this non-linear system is the Marquardt's Method ([Le44],[Ma63]).

Appendix B

TOPOGRAPHIC LABELS ON THE ELLIPSOID

The equation of an elliptical surface is given by:

$$S(x,y) = d - (1^2-m^2x^2-n^2y^2)^{1/2} \quad \text{for } -a \le x \le a \text{ and } -b \le y \le b$$

The first, second and third partials of S are obtained to be:

$$S_x = m^2xA^{-1/2},$$
$$S_y = n^2yA^{-1/2},$$
$$S_{xx} = m^2(1^2-n^2y^2)A^{-3/2},$$
$$S_{yy} = n^2(1^2-m^2x^2)A^{-3/2},$$
$$S_{xy} = S_{yx} = m^2n^2xyA^{-3/2},$$
$$S_{xxx} = 3m^4x(1^2-n^2y^2)A^{-5/2},$$
$$S_{yyy} = 3n^4y(1^2-m^2x^2)A^{-5/2},$$
$$S_{xyy} = S_{yyx} = S_{yxy} = m^2n^2x(1^2-m^2x^2+2n^2y^2)A^{-5/2},$$

and $S_{yxx} = S_{xxy} = S_{xyx} = m^2n^2y(1^2-n^2y^2+2m^2x^2)A^{-5/2}$,

where $A = 1^2 - m^2x^2 - n^2y^2$.

The intensity of the ellipsoid illuminated from directional [a,b,c] is given by:

$$I(x,y) = I0 \left| \frac{am^2x + bn^2y - c(1^2-m^2x^2-n^2y^2)^{1/2}}{(1^2+m^2x^2(m^2-1)+n^2y^2(n^2-1))^{1/2}} \right|$$

The above equation can be rewritten as:

$$I = I_0 \,(a\, N^x + b\, N^y + c\, N^w)$$

where

$$N_x = S_x/(1+S_x{}^2+S_y{}^2)^{1/2},$$

$$N_y = S_y/(1+S_x{}^2+S_y{}^2)^{1/2},$$

and $$N_w = -1/(1+S_x{}^2+S_y{}^2)^{1/2}.$$

The first and second partials of I are given by:

$$I_x = I_0\ (aN_x^x + bN_x^y + cN_x^w),$$

$$I_y = I_0\ (aN_y^x + bN_y^y + cN_y^w),$$

$$I_{xx} = I_0\ (aN_{xx}^x + bN_{xx}^y + cN_{xx}^w),$$

$$I_{xy} = I_{yx} = I_0\ (aN_{xy}^x + bN_{xy}^y + cN_{xy}^w),$$

and $$I_{yy} = I_0\ (aN_{yy}^x + bN_{yy}^y + cN_{yy}^w).$$

If we let $D=1+S_x{}^2+S_y{}^2$, the partials of the N's are found to be:

$$N_x^x = [(1+S_y{}^2)S_{xx} - S_xS_yS_{xy}]D^{-3/2}$$

$$N_y^x = [(1+S_y{}^2)S_{yx} - S_xS_yS_{xy}]D^{-3/2}$$

$$N_y^y = [(1+S_x{}^2)S_{yy} - S_xS_yS_{xy}]D^{-3/2}$$

$$N_x^y = [(1+S_x{}^2)S_{xy} - S_xS_yS_{xx}]D^{-3/2}$$

$$N_x^w = [S_xS_{xx} + S_yS_{xy}]D^{-3/2}$$

$$N_y^w = [S_xS_{xy} + S_yS_{yy}]D^{-3/2}$$

$$N_{xx}^x = [(1+S_y{}^2)S_{xxx}-S_xS_yS_{yxx}-S_xS_{yx}{}^2+S_yS_{xx}S_{yx}]D^{-3/2}- \\ 3[(1+S_y{}^2)S_{xx}-S_xS_yS_{xy}](S_xS_{xx}+S_yS_{yx})D^{-5/2}$$

$$N_{yx}^x = [(1+S_y{}^2)S_{yxx}-S_xS_yS_{xyy}+2S_yS_{xx}S_{yy}- \\ S_xS_{xy}S_{yy}-S_yS_{xy}{}^2]D^{-3/2}- \\ 3[(1+S_y{}^2)S_{xx}-S_xS_yS_{xy}](S_xS_{xy}+S_yS_{yy})]D^{-3/2}$$

$$N_{yy}^x = [(1+S_y{}^2)S_{xyy}-S_xS_yS_{yyy}-S_xS_{yy}{}^2+S_yS_{xy}S_{yy}]D^{-3/2}- \\ 3[(1+S_y{}^2)S_{xy}-S_xS_yS_{yy}](S_xS_{xy}+S_yS_{yy})D^{-5/2}$$

$$\begin{aligned} N^y_{yy} = {} & [(1+S_x{}^2)S_{yyy}-S_yS_xS_{xyy}-S_yS_{xy}{}^2+S_xS_{yy}S_{xy}]D^{-3/2}- \\ & 3[(1+S_x{}^2)S_{yy}-S_yS_xS_{yx}](S_yS_{yy}+S_xS_{xy})D^{-5/2} \end{aligned}$$

$$\begin{aligned} N^y_{xy} = {} & [(1+S_x{}^2)S_{xyy}-S_yS_xS_{yxx}+2S_xS_{yy}S_{xx}- \\ & S_yS_{yx}S_{xx}-S_xS_{yx}{}^2]D^{-3/2}- \\ & 3[(1+S_x{}^2)S_{yy}-S_yS_xS_{yx}](S_yS_{yx}+S_xS_{xx})]D^{-3/2} \end{aligned}$$

$$\begin{aligned} N^y_{xx} = {} & [(1+S_x{}^2)S_{yxx}-S_yS_xS_{xxx}-S_yS_{xx}{}^2+S_xS_{yx}S_{xx}]D^{-3/2}- \\ & 3[(1+S_x{}^2)S_{yx}-S_yS_xS_{xx}](S_yS_{yx}+S_xS_{xx})D^{-5/2} \end{aligned}$$

$$\begin{aligned} N^w_{xx} = {} & (S_xS_{xxx}+S_{xx}{}^2+S_yS_{yxx}+S_{yx}{}^2)D^{-3/2}- \\ & 3(S_xS_{xx}+S_yS_{xy})^2D^{-5/2} \end{aligned}$$

$$\begin{aligned} N^w_{xy} = {} & [S_xS_yS_{yxx}+S_yS_{xyy}+S_{xy}(S_{xx}+S_{yy})]D^{-3/2}- \\ & 3(S_xS_{xx}+S_yS_{yx})(S_xS_{xy}+S_yS_{yy})D^{-5/2} \end{aligned}$$

$$\begin{aligned} N^w_{yy} = {} & (S_xS_{xyy}+S_{xy}{}^2+S_{yy}{}^2+S_yS_{yyy})D^{-3/2}- \\ & 3(S_xS_{xy}+S_yS_{yy})^2D^{-5/2} \end{aligned}$$

After substituting the partials of the S's into the equations for the partials of the N's, we obtain

$$N^x_x = [m^2(l^2-n^2(1-n^2)y^2)]H^{-3/2}$$

$$N^x_y = [m^2n^2xy(1-n^2)]H^{-3/2}$$

$$N^y_y = [n^2(l^2-m^2(1-m^2)x^2)]H^{-3/2}$$

$$N^y_x = [m^2n^2xy(1-m^2)]H^{-3/2}$$

$$N^w_x = [m^2x(m^2l^2+n^2y^2(n^2-m^2)](l^2-m^2x^2-n^2y^2)^{-1/2}H^{-3/2}$$

$$N^w_y = [n^2y(n^2l^2+m^2x^2(m^2-n^2)](l^2-m^2x^2-n^2y^2)^{-1/2}H^{-3/2}$$

$$N^x_{xx} = [3m^4x(1-m^2)(l^2-n^2y(1-n^2))]H^{-5/2}$$

$$N^x_{yy} = [m^2n^2(1-n^2)x][l^2+m^2(m^2-1)x^2-2n^2(n^2-1)y^2]H^{-5/2}$$

$$N^x_{xy} = [m^2n^2(1-n^2)y][l^2-2m^2(m^2-1)x^2+n^2(n^2-1)y^2]H^{-5/2}$$

$$N^y_{yy} = [3n^4y(1-n^2)(l^2-m^2x(1-m^2))]H^{-5/2}$$

$$N^{y}_{xx} = [n^2m^2(1-m^2)y][l^2+n^2(n^2-1)y^2-2m^2(m^2-1)x^2]H^{-5/2}$$

$$N^{y}_{yx} = [n^2m^2(1-m^2)x][l^2-2n^2(n^2-1)y^2+m^2(m^2-1)x^2]H^{-5/2}$$

$$N^{w}_{xx} = \frac{m^2[(l^2-n^2y^2)(3m^4x^2+m^2(l_2-n^2y^2)+n^4y^2)+3m^2n4x^2y^2]}{(l_2-m^2x^2-n^2y^2)^{3/2}H^{3/2}} - \frac{3n^4y^2[n^2(l_2-m^2x^2)+m^4x^2]2}{(l_2-m^2x^2-n^2y^2)^{3/2}H^{5/2}}$$

$$N^{w}_{yy} = \frac{n^2[(l^2-m^2x^2)(3n^4y^2+n^2(l_2-m^2x^2)+m^4x^2)+3n^2m4y^2x^2]}{(l_2-n^2y^2-m^2x^2)^{3/2}H^{3/2}} - \frac{3m^4x^2[m^2(l_2-n^2y^2)+n^4y^2]2}{(l_2-n^2y^2-m^2x^2)^{3/2}H^{5/2}}$$

$$N^{w}_{xy} = \frac{2m^2n^2xy[n^2(l^2-m^2x^2+n^2y^2)+m^2(l^2-n^2y^2+m^2x^2)]}{(l_2-n^2y^2-m^2x^2)^{3/2}H^{3/2}} - \frac{3[m^2n^2xy(m^2(l^2-n^2y^2)+n^4y^2)(m^4x^2+n^2(l^2-m^2x^2))]}{(l_2-n^2y^2-m^2x^2)^{3/2}H^{5/2}}$$

The Hessian we need is

$$H = \begin{vmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{vv} \end{vmatrix}$$

and

$$I_{xx} = I0\ (aN^{x}_{xx} + bN^{y}_{xx} + cN^{w}_{xx})$$

$$I_{xy} = I_{yx} = I0\ (aN^{x}_{xy} + bN^{y}_{xy} + cN^{w}_{xy})$$

$$I_{vv} = I0\ (aN^{x}_{yy} + bN^{y}_{yy} + cN^{w}_{yy})$$

Eigenvalues A satisfy

H - A I is singular

That is, $\det(H - AI) = 0$

$$(I_{xx} - A)(I_{vv} - A) - {I_{xy}}^2 = 0$$

or $I_{xx}I_{vv} - AI_{xx} - AI_{vv} + A2 - I_{xy}^{2} = 0$

or $A2 - (I_{xx}+I_{vv})A + I_{xx}I_{vv} - I_{xy}^{2} = 0$

or

$$A = \frac{I_{xx}+I_{vv}+[(I_{xx}+I_{vv})^{2} - 4(I_{xx}I_{vv}-I_{xy}^{2}))^{1/2}}{2}$$

Or, the more numerically stable formula is

$$A = \frac{I_{xx}+I_{vv}}{2} + [\frac{(I_{xx}-I_{vv})^{2}}{4} + I_{xy}^{2}]^{1/2} .$$

The eigenvector (w) coresponds to the eigenvalue A is then found by solving

$$H\ w = A\ w$$

or
$$\begin{vmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{vv} \end{vmatrix} \begin{vmatrix} w_1 \\ w_2 \end{vmatrix} = \begin{vmatrix} A\ w_1 \\ A\ w_2 \end{vmatrix}$$

Similar results can be obtained for the hyperboloids by following the same deviations.

Appendix C

TOPOGRAPHIC LABELS ON THE PARABOLOID

The equation of an elliptical surface is given by:

$$S(x,y) = d - (mx^2+ny^2)^{1/2} \quad \text{for} \quad -a \leq x \leq a \quad \text{and} \quad -b \leq y \leq b.$$

The first, second and third partials of S are obtained to be:

$S_x = -2mx,$

$S_y = -2ny,$

$S_{xx} = -2m,$

$S_{yy} = -2n,$

$S_{xy} = Syx = 0,$ and

$S_{xxx} = S_{yyy} = S_{xyy} = S_{yxx} = 0.$

The intensity of the paraboloid illuminated from direction [a,b,c] is given by:

$$I(x,y) = \frac{-2amx - 2bny - c}{(1 + 4m^2x^2 + 4n^2y^2)^{1/2}}$$

After substituting the partials of the S's into the equations for the partials of the N's, we obtain

$N^x_x = -2m(1+4n^2y^2)H^{-3/2}$

$N^y_y = -2n(1+4m^2x^2)H^{-3/2}$

$N^x_y = 8mn^2xyH^{-3/2}$

$N^y_x = 8m^2nxyH^{-3/2}$

$$N^x_{xx} = 24m^3x(1+4n^2y^2)H^{-5/2}$$
$$N^x_{xy} = 8mn^2y(1-8m^2x^2+4n^2y^2)H^{-5/2}$$
$$N^x_{yy} = 8mn^2x(1+4m^2x^2-8n^2y^2)H^{-5/2}$$
$$N^y_{yy} = 24n^3y(1+4m^2x^2)H^{-5/2}$$
$$N^y_{xy} = 8m^2nx(1+4m^2x^2-8n^2y^2)H^{-5/2}$$
$$N^y_{xx} = 8m^2ny(1-8m^2x^2+4n^2y^2)H^{-5/2}$$
$$N^w_{xx} = 4m^2(1-8m^2x^2+4n^2y^2)H^{-5/2}$$
$$N^w_{xy} = -48m^2n^2xyH^{-5/2}$$
$$N^w_{yy} = 4n^2(1+4m^2x^2-8n^2y^2)H^{-5/2}$$

As for the ellipsoid, eigenvalues and eigenvectors can then be found directly from the above expressions.