1990

Visual surface perception: Surface reconstruction from shading and stereo.

Srinivasan. Raghavan
University of Windsor

NOTICE

AVIS

VISUAL SURFACE PERCEPTION:

SURFACE RECONSTRUCTION FROM SHADING AND STEREO

by

Srinivasan Raghavan

A dissertation

submitted to the Faculty of Graduate studies

through the Department of Electrical Engineering

in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

University of Windsor

©

Windsor, Ontario, Canada

1990

Abstract

Computing shape and depth from stereo vision and shading is one of the important perceptual tasks in early vision. This thesis work is aimed at understanding the computational issues involved in reconstructing a viewed surface in three dimension.

In using stereo vision to compute depth, stereo correspondence between points in the left and right images can be reliably achieved only at points of intensity changes. Owing to the need for computing depth at every point in the image, interpolation of this sparse depth data becomes necessary. Since several surfaces can fit a given sparse grid, an appropriate choice comes from imposing additional constraints. In our work, we propose that the interpolant should not only introduce any additional discontinuities (other than those dictated by the intensity changes), but also that they should preserve those discontinuities. A class of interpolants known as Shepard's surfaces, is shown here to satisfy this constraint. The Shepard's interpolants have been implemented here and the result from testing them on Random DOt Stereograms shows that stereo vision can function alone without any additional visual cues. The natural stereo pair shows that even when intensity changes are sparse the reconstruction preserves the shape although, the interpolant exhibits a tendency to consider spurious stereo matches also as a potential data point.

Besides depth from stereopsis, shape information also becomes important to reconstruct the surface. An important shape cue is available in the smooth shading that an object renders. From the perspective of obtaining shape description (instead of surface normals

alone), we propose a method to compute relative depth, normals and principal curvatures of the surface. Shape information is intrinsic to the surface and is independent of viewer position. Since continuity in normals is ensured through these shape descriptors, the numerical error introduced in the process of reconstruction is shown to be independent of coordinate axes chosen. Our method involves minimization of a global objective function formulated by imposing the following constraints: (i) continuity and integrability of the normals, (ii) minimal deviation from the irradiance values, and (iii) unit normal. Minimizing the objective function with respect to the normals n, relative depth z and shape descriptor A, results in direct computation of all these quantities. In addition, the principal curvatures are shown to be computable from the shape descriptor used here.

Shape and depth information may also be available from other visual cues. We show that depth information at arbitrary set of points can be included as additional set constraints in the shape computing algorithm. Also, any known set of normals can also be exploited to improve the convergence of the shape from shading algorithm, besides smoothly incorporating the additional source of information.

This work, in essence, has resulted in developing a framework that delivers shape information in the form of local curvatures and depth at every point in the image plane.

Dedicated to my parents

whose vision has guided the course of my life

Acknowledgements

There are occasions when one's feeling of gratitude can not be adequately expressed. I find myself in one such situation trying to acnkowledge my indebtedness to all those who, in various ways, contributed for a fruitful tenure as a research student here.

Dr. Shridhar was much more than being a supervisor. He was a mentor and the motivator. His constant attempts to point out the flaws have steered me into the right course. This thesis would not have taken its shape without him. I owe my deep gratitude to him.

As a constant source of encouragement and support, Dr. Ahmadi was always there to offer a helping hand at every needy occasion. I express my sincere appreciation to him and cherish my association with him.

I owe my profound sense of gratitude to Dr.Narasimhamurthi, from whom often I stood benefitted with a rich pool of ideas.

I would like to thank Dr. Jullien, Dr. Watt and Dr.North for their valuable suggestions that led to many improvements.

I would also like to thank Dr. Robert King of the Imperial College at London, for sharing his thoughts as an external examiner. Thanks to Dr.Laveen Kanal for his insightful comments and support.

Thanks are extended to Dr.Sid Ahmed for letting me use his computing facilities for the image transfer.

My friends need a special mentioning here, for without them I would not have found it easy. To avoid a long list I choose to say "THANKS" to every one of them. I can not be any luckier to have such a loving home with my parents, brother and sisters.

Contents

List of Tables and Figures

Notations

$I(x,y)$	Intensity value at the point (x,y)
z_x, z_y	Partial derivatives of z with respect to x and y
l	Light source vector
v	Viewer direction
ρ	Reflectance of the surface
T	Translation operator
Ω	Rotation operator
P	A generic point on a surface
$f(P)$	Representation of a function of P
d_j	Data at point j
f_k	Depth value of a surface at the point P_k
N	Number of points in an image grid
μ	Control parameter of the Shepard's interpolant
S	Representation of a surface in R^3
n	Unit normals on the surface
A	Shape matrix
$\alpha, \beta, \gamma, \sigma, \lambda$	Penalty weights
z_e, z_t	Estimated and true z
n_e, n_t	Estimated and true normal n
$\eta(i,j)$	8-neighborhood of (i,j)
$\nabla^2 G$	Laplacian og Gaussian

Chapter 1

An overview

1 Introduction

This dissertation deals with computational modeling of the two most important early visual tasks - stereo depth perception and shape from shading. Stereo depth computation involves two steps: (i) locating corresponding pairs of feature points in the two images and (ii) interpolating the depth values at the feature points to obtain depth every where. The problem of shape recovery from shading involves computing local orientation, relative depth and curvature information of the surface. The research efforts undertaken here are concerned with stereo data interpolation and shape computing from intensity information.

The major contributions of our study include

(i) suggestion of a stronger constraint to choose the best interpolant for stereo data

(ii) proposing a fast and efficient interpolant called Shepard's Metric scheme for filling the sparse depth map

(iii) laying the emphasis on local shape recovery in the form of curvatures besides estimating the surface normals and

(iv) a robust technique to compute surface normals, curvatures and relative depth from intensity images.

Until very recently, every early visual problem was considered to be modular and independent in nature as suggested by David Marr [Mar 82]. An important concern is raised about this notion in this dissertation. The interplay between various early visual problems is

1

an essential attribute to achieve fast and efficient surface estimates. For instance, rough estimates of shape and depth at occluding boundaries influence considerably the accuracy as well as speed in computing shape at the interior of the surface as found in this work along with several others too [IkH 81, FrC 87]. In addition, the issue of integrating shape and depth from shading and stereo vision has been addressed. This is done by making use of the quick and rough estimates of depth from stereo vision in the shape from shading algorithm developed here.

This chapter is intended to provide an overview of early vision research. Section 2 lays the foundation for the computational approach adopted in our work. Section 3 presents a coarse organization of the vision system. Section 4 provides a comprehensive account of the literature on surface reconstruction. In section 5, the motivation for the present work is laid out. Section 6 concludes with an organization of this dissertation.

2 Computational Vision

One of the goals of early visual perception is depth and shape estimation. The perceived depth and shape information is later used in the higher levels for two purposes: (i) to extract an invariant description of the objects in the scene and (ii) to reconstruct the spatial structure of the scene. The term spatial structure here refers to the spatial arrangement of the objects and their inter-relationships. Discovering this spatial structure in the scene is a very challenging task since the geometry of projecting the three dimensional world into the two dimensional retinae or a camera image

plane results in a massive loss of information content.

Human vision is extremely adept at recovering the surface information through various visual cues present in the scene. Such visual cues include binocular disparity, smooth shading rendered on the objects, motion parallax from moving objects, textural deformations and even the focusing mechanisms employed for visual adaptation. A very promising approach to solve the early visual problems has been developed by David Marr [Mar 82]. We begin with the description of this approach here.

In essence, perceptual tasks could be understood from at least 3 distinct and independent levels. They are :

(i) a computational theory for the specific problem

(ii) an algorithm and a formal representation for the data

(iii) a suitable implementation (neural or machine hardware).

At the first level, the task at hand is thoroughly analyzed to answer questions such as what is the goal of computation, why is it appropriate and what is the logic of the strategy by which it can be carried out? Thus, the goal of the specific perceptual problem is set forth first, followed by an understanding of the physics of that problem. The term physics includes geometry involved, regularities and constraints of the three dimensional world and so forth.

At the second level, the main question to be answered is how this computational theory can be faithfully translated into a methodology or a procedure. Also concerned at this level is the issue of efficient representation of data involved. The choice of representation and the algorithm are closely related to each other. That is, performance of

3

the algorithm can very well depend on the ease with which the representation scheme could help to manipulate data.

The third level incorporates a study of machine hardware or the type of architecture on which the chosen representation and algorithm can be realized. The choice of architecture could have a profound effect on efficiency and economy of resources. For instance, an algorithm that uses only local information from its neighbors can be implemented with a higher level of throughput in a cellular architecture as compared to a general purpose connection machine.

It may be noticed that the three levels described here are not completely decoupled. As an example, the choice of hardware need not be totally independent of the algorithm. Instead, the insight into the neural architecture could very well influence the choice of an algorithm that would run on such a computing system. In fact, a serious question that needs to be answered here is what do we gain by addressing the problem at three distinct levels. The reasoning is as follows.

Sever a l algorithms can be be suggested to implement a computational theory. A comparison of these algorithms should only be done by analyzing how faithfully they implement the computational theory. In the absence of a computational theory, they can only be judged solely on their performance on a few instances of the problem. Such a judgment could be misleading. On the other hand, a deficiency in the computational theory itself can lead to errors in computation. To give an example for this situation consider the computational theory of stereopsis proposed by Marr and Poggio [MaP 79]. They used

4

an invalid assumption that depth varies continuously almost every where. But the assumption is violated at all the edges in the scene. Instead, the assumption that depth varies continuously along the edges is quite valid and needs to be in any algorithm for solving correspondence. Furthermore, separating the implementational issues from those related to computational theory and algorithm does help to avoid micro-level details and hence a better understanding of the problem.

The effectiveness of this approach can be judged by the rigorous computational theories and algorithms developed in early vision [Mar 82]. Unfortunately a principled approach like this could become difficult to handle at higher levels of vision since distinction between the three different levels vanishes. Problems dealt with in this dissertation do yield to this type of analysis and hence this section was devoted for defending this approach.

3 A Hierarchical Framework for Understanding Vision

The massive size of raw sensory input suggests that unless an efficient scheme to abstract useful information is built, it is quite difficult to achieve what human vision does in a few milli seconds. Furthermore, retinal images themselves do not contain the three dimensional information about the viewed object per se. Instead, they only contain some "geometric invariants" of the object. The task of the beginning stages of vision is essentially to recover surface information from the images.

A flow diagram for visual information processing is given in Figure 1.1. The first stage of information processing that occurs at

5
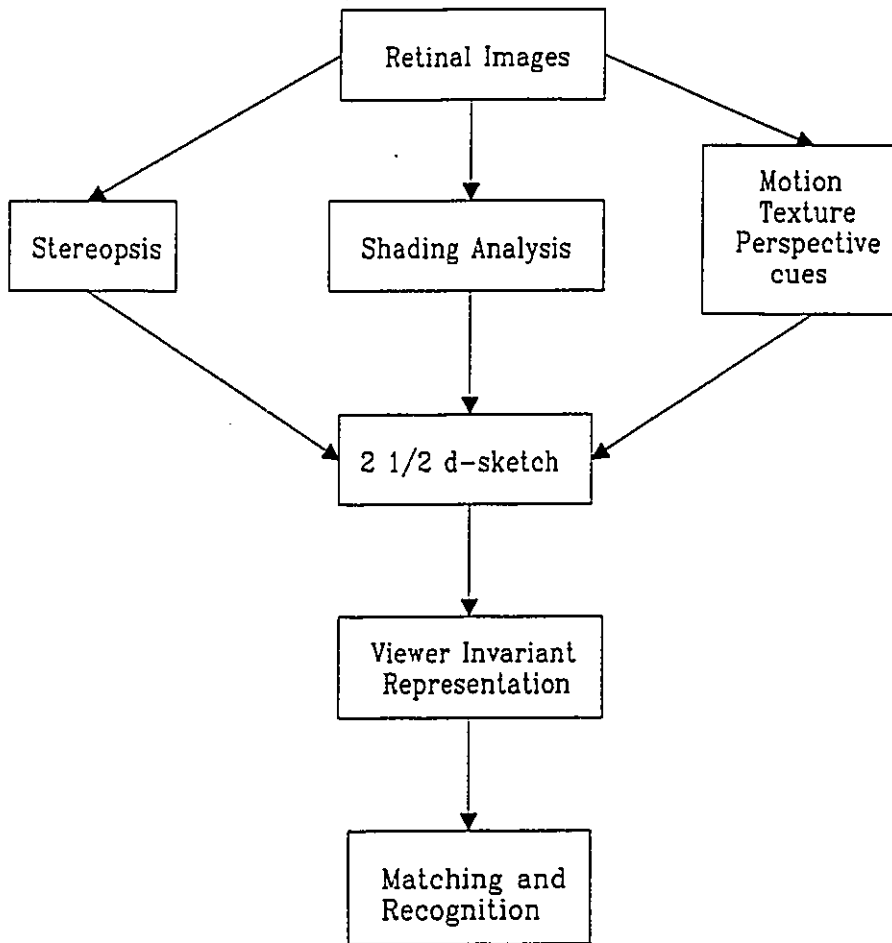
Figure 1.1 Information Flow in Vision

the level of Ganglion cells soon after the incidence of light rays on the retinae, correspond to analyzing brightness values. Called primal sketch, by Marr [Mar 77a, Mar 80], this stage is involved in obtaining two dimensional information such as edges, markings, dots, contrast etc. from the images.

The immediate processes following the extraction of primal sketch are related to three dimensional reconstruction of the scene. There appear to be multiple cues that aid the process of reconstruction. The most important one among them is the binocular disparity. The influence of binocular disparity in vivid depth perception can be understood using the following simple experiments. In the first experiment, a finger or a pencil is held in front of the eyes and the eyes are fixed on this object. A second object is now placed in the field of view and aligned with the first object as viewed from the right eye. Now, if the fingers are viewed from the left eye, the nearer object seems to move to the right as the one of the objects is moved back and forth. The lateral shift introduced here is the binocular or stereo disparity. It can be seen that this shift is inversely proportional to the distance of the object from the eyes. In the second experiment, to see why binocular vision is so important, one of the eyes is closed. Now, the two hands are stretched to any arbitrary length in front of the open eye and their index fingers are allowed to meet each other. It is easy to see that it takes a few trials to get them meet, while with both the eyes open this will happen at the first trial. The effect is more pronounced if sharply pointed objects are used instead of fingers.

The second important cue for reconstruction is the smooth shading available on the objects rendered by the viewing geometry. Thus, photographs of even unfamiliar objects convey their shape under the absence of binocular disparity. Pictures appear vivid because of the shading effect the artists provide. For a non dynamic situation (both viewer as well as the scene are stationary), shading seems to provide a stronger sense of shape [Gib 50] than any other monocular cue. Also, mathematically, it is a totally data driven process. That is, the system does not need to know a priori anything specific about the surface under view.

The remaining cues shown in Figure 1.1 enrich the perception, although they alone are quite weak in supporting surface perception. For instance, perspectivity suggests that points close to the vanishing point should be far away from the viewer. However, size constancy of the objects is assumed and hence a priori knowledge of the size of the object could be necessary often. Motion parallax is quite strongly suggestive of spatial structure but yet requires relative motion between the scene and the viewer. On the same token, texture imposes shape constraints for a given distribution of its primitives [Wit 81]. Nevertheless, naturally found objects are not often rich in texture.

Given the observation that multiple cues are available for three dimensional reconstruction of the viewed surface, the next stage in Figure 1.1 serves as a common medium where all the channels of surface information could be integrated. A specific view of an object is representative of its class of objects only to a certain extent.It is

8

clearly infeasible to store all views of the same object. If some kind of intrinsic information of the object can be extracted from the specific view of the object, then this intrinsic information, due to its invariant nature, can facilitate the process of recognition. However, this computation can be done at the previous stages such that the shape and depth computing processes themselves would directly deliver object intrinsic information. For the sake of clarity, a separate module has been allocated in the block diagram. The ultimate goal of vision being recognition, the stream of visual information flow concludes with a process that associates the collected data with the known characteristics of corresponding object in memory.

The perceptual problems addressed in this thesis are concerned with the first two levels in the visual hierarchy. In the following section a comprehensive literature of various models and computational theories for surface reconstruction in early vision is provided.

4 Literature on Surface Reconstruction

The last two decades have witnessed a strong growth in the computational vision literature especially on visual surface construction. In addition, advances in human vision research have fueled the growth considerably. A comprehensive review of the relevant reported literature is presented here. Individual sections are devoted for stereo vision, shading, motion parallax and texture since these are the most important depth and shape cues in aiding visual surface reconstruction.

4.1. STEREO VISION

In stereo vision, (Figure 1.2) the two eyes slightly converge to a

9
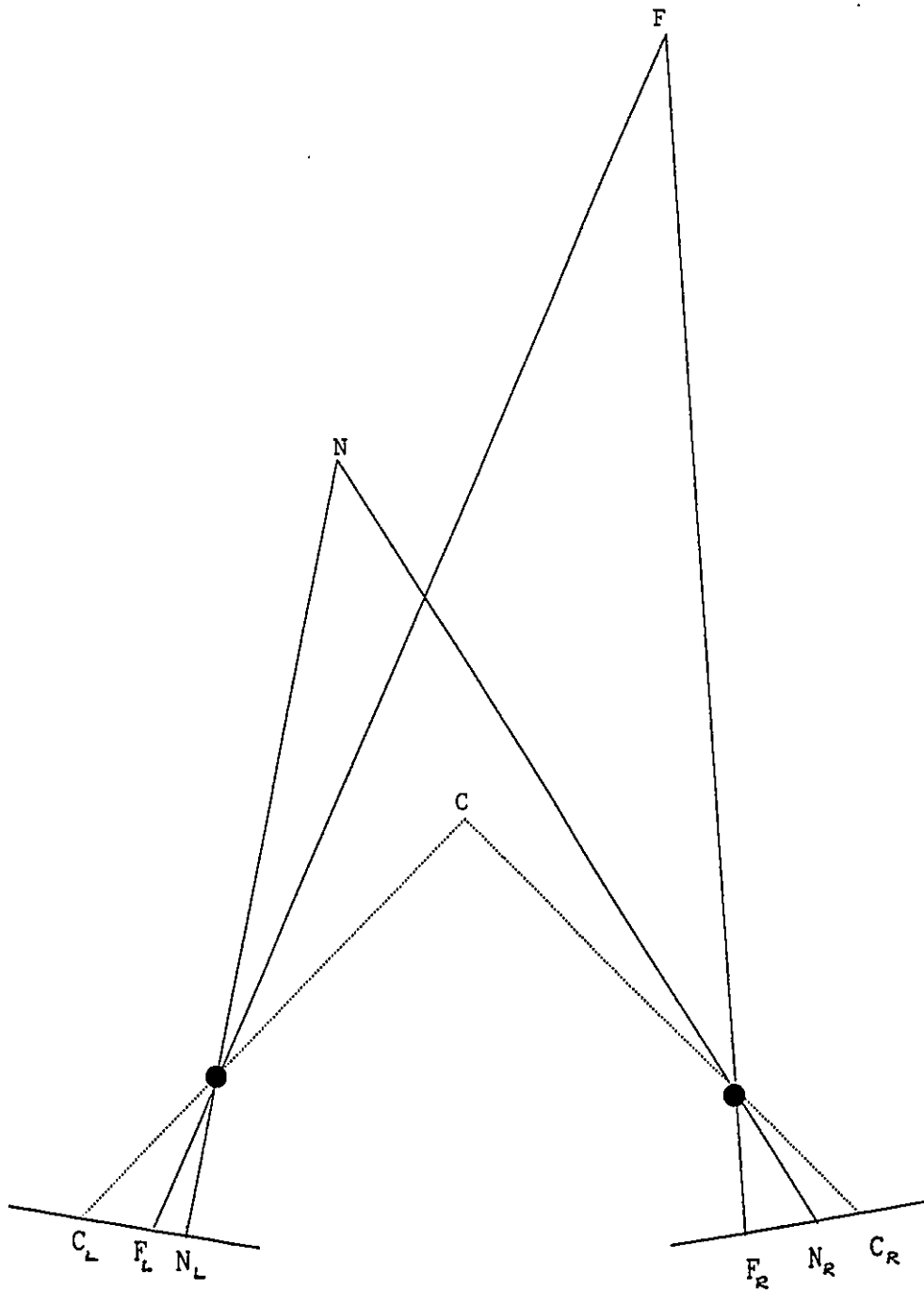
Figure 1.2 Geometry of Stereo Vision

distant point in the visual field so that their optical axes meet at a point, called the fixating point. The image of the fixating point falls on the fovea. The neighboring points in the visual space project to different points on the retinae at some distance away from the image of the fixating point. In general, this distance may not be equal in the two images. The amount of shift a specific point undergoes when projected to corresponding positions in the two images is called its retinal disparity. It can be easily shown from the geometry of the stereo imaging that this disparity is inversely proportional to the distance of the point of interest in space from the imaging plane which can be obtained using a simple triangulation procedure [MaL 82].

4.1.1. Correspondence Problem

Stereopsis involves three steps [PoP 84]. They are: (1) identifying a region or a point in one image (2) locating the corresponding shifted position of this point in the other image (correspondence problem) and (3) measuring the disparity between the corresponding pair to transform it into depth. The second step is the most difficult one. To illustrate the difficulties involved consider the typical stereo imaging geometry shown in Figure 1.2. Let C, F and N be three points lying in the visual field. Let the left and right images for these points be C_L, F_L and N_L, and C_R, F_R and N_R respectively. Clearly, the point C in the left image has potentially at least three matches on the right. Similarly the points F and N also have three potential matches in the right image, of which only one is the correct match. The incorrect matches, referred to as false targets

11

here, pose the need to devise a reliable method to disambiguate the corresponding pairs from all the potential pairs. The false target problem can be shown to be severe as follows. If we have N points on the left image to be matched with N points on the right image, then there are $N^2 - N$ false targets since only N of N^2 pairs are correct pairs. If N is large the false target problem is clearly a serious issue.

Beginning with Wheatstone's work on stereoscope, a great deal of attention has been devoted to the study of stereo vision. The early work in this area has been confined to biological studies until Julesz's work on computer generation of Random Dot Stereograms (RDS) [Jul 60, Jul 71]. To prove that retinal stereo disparity alone is sufficient to provide depth sensation, he constructed a stereogram with a randomly generated texture. The images of the stereo pair differ only in a pre-selected central portion of the image. When fused, the stereogram gives a vivid feeling of depth in the central portion showing that even though the images are monocularly non-sensical, binocularly they offer the percept of two distinct surfaces. Julesz's RDS studies have resulted in his cooperative model of global stereopsis [Jul 71]. This model involves relating the equilibrium achieved by an array of magnetic dipoles to the disparity detectors found in human vision by modeling the stereo fusion process as the interaction between magnetic dipoles. Although, this mechanical model helped understanding several biological phenomena it was far from providing a computational understanding.

The first generation of cooperative neural/computational models

12

for stereo vision were proposed [MaP 76] following the findings of disparity detectors in animal vision by Hubel and Wiesel [HuW 77]. All these models share the concept of cooperative computation in a three dimensional field of disparity detecting elements. The three dimensional structure is a regular arrangement of columns, where each column denotes a specific disparity value or depth in 3-d space. Elements detecting similar disparity values but with input from different areas of the visual field are housed in the same column. Marr and Poggio presented a computational simulation of their model that involves elements with binary activation levels. Although, their model demonstrated that stereo fusion is achievable in case of binary instances such as RDS it proved to be difficult to extend the model for a real life application with multiple gray levels.

The second generation of non-cooperative models for stereopsis were proposed [MaP 79, MaF 81, Gri 85] with solving correspondence between feature points as the main objective. The feature points are the points which reflect strong characteristics such as discontinuity of intensity. The correspondence problem involves finding the lateral translation of an identified feature point in one image to another feature point in the other image. Correspondence is established in a multi-step process, beginning with an initial list of hypothesized pairs. Disambiguation between competing matches is achieved by applying constraints such as continuity of disparity in a local region, compatibility of local contrast, etc. For instance, the Marr-Poggio algorithm [MaP 79] achieves disambiguation by probabilistic means. The input to their algorithm is the set of zero crossings

13

detected by convolving Laplacian of the Gaussian with the images. An important result of their study is that within ω, the channel width employed for Gaussian smoothing, the probability of finding a zero crossing in the other image is less than 0.3 while it is less than 0.7 for 2ω. Thus, if a match is found within ω then its candidacy deserves a consideration. Further disambiguation is achieved by taking into account its contextual information. However, his assumption that disparity vary continuously in a local region is violated across the edges. Since, edges are also captured by zero crossings this assumption is invalid.

A significant contribution to the computational theory of stereo vision, due to Mayhew and Frisby [MaF 81], is that the disparity varies continuously along the zero crossings (and not across the zero crossings). Known as figural continuity constraint, this constraint forms the basis of a powerful theory to solve the correspondence problem. Grimson incorporated the figural continuity constraint in the later version of his algorithm [Gri 85]. It needs to be pointed out here that in order to derive a new set of probabilistic limits he further assumed that the two images are uncorrelated, violating the basic nature of binocular vision. A contour based algorithm for computing correspondence has been proposed by Srinivasan et al [SrR 87], by exploiting the fullest disambiguating power of figural continuity. This method involves employing a hypothesis testing paradigm to distinguish between the correct and incorrect pairing of features. The candidate hypotheses are that (i) a given pair is the actual corresponding pair and (ii) the given pair is an incorrect pair

14

(negation of (i)). A statistical decision rule is invoked to decide in favor of one of the two hypotheses. The method has been tested on natural as well as random dot stereograms. A detailed account of this method and a comprehensive treatment on stereopsis are provided in the Master's thesis of Srinivasan [Sri 87]. It may, however, be noted here that algorithms for the solving correspondence problem deliver depth only at points of intensity changes. In the remainder of this section, we review the work undertaken on interpolation of depth values in the sparse set of intensity changes.

4.1.2. Surface Interpolation

The human vision possesses the tendency to fill-in and interpolate/extrapolate the visual data in many instances of form and surface perception. One such widely known instance is the subjective contours [Mar 82]. In the case of subjective contours (shown in Figure 1.3) the strong contrast at the corners of the square influences the early visual processes to extend (extrapolate) the edge from the corners thus creating an apparent sqaure standing out. Notice that the gray level inside the figure is the same as the one outside, yet the brightness of the region inside the apparent square looks different from the background. Intermittently placed dots have also been known to induce continuous contours along them (interpolation).

While, the afore-mentioned examples explain interpolation / extrapolation in two dimensional vision only, to demonstrate a similar effect in three dimensional vision, Grimson devised an interesting experiment as follows. He created a stereogram to induce the perception of a shape, a cross-section of which is shown in Figure
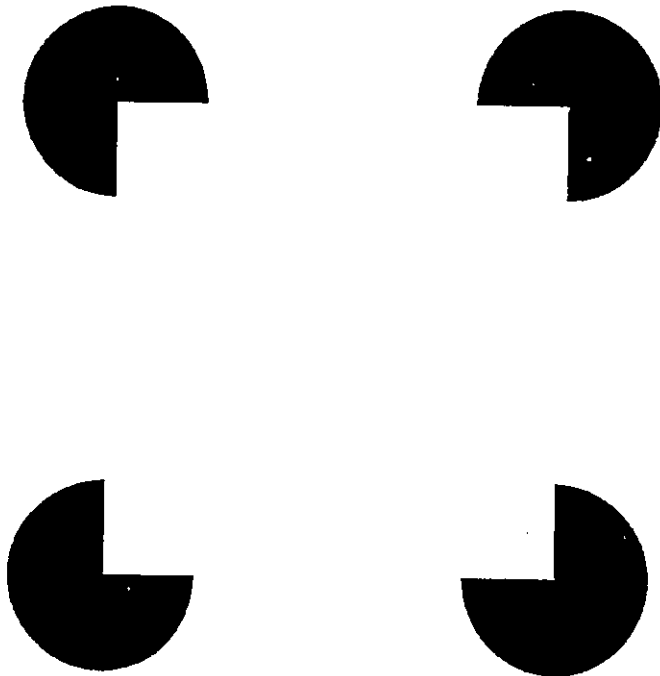
Figure 1.3 Subjective Contours Suggestive of Line Interpolation

————— Perceived Shape

-------------- True Shape
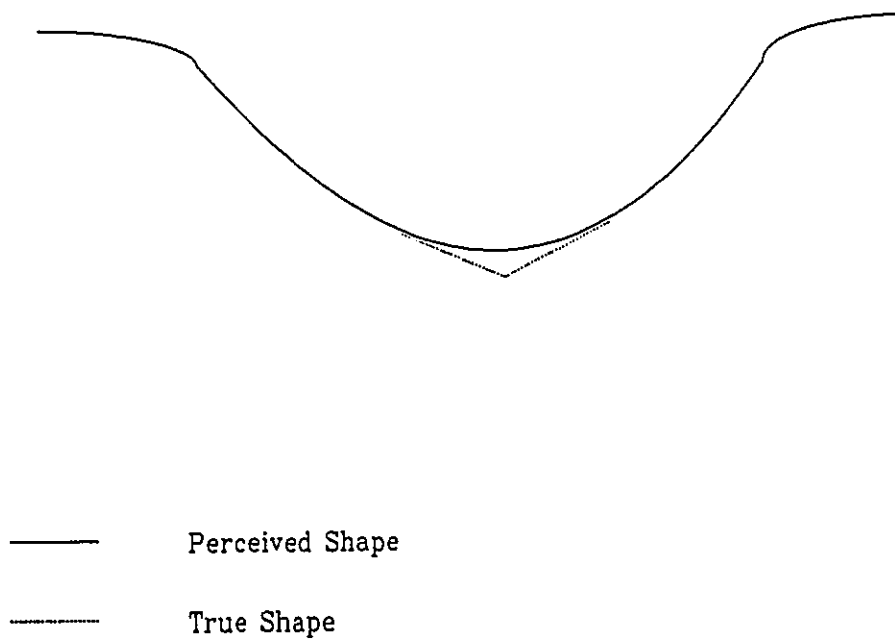
Figure 1.4 Grimson's Experiment on Continuous Nature of Interpolation

1.4. By intentionally leaving a small portion unfilled with disparities, he showed that his subjects experienced the percept of a continuous surface even in the unfilled region. It may be noted that the percept corresponding to the dotted line is also another plausible candidate. Two conclusions can be drawn from this experiment. (i) Stereopsis itself should be supportive of interpolation in surface perception. (ii) The interpolated surface is smooth and continuous. Thus, in order to reconstruct the entire viewed surface, the sparse depth map obtained from matching the corresponding points needs to be filled-in. In contrast with the rich literature on the correspondence problem, relatively very little work has been done on stereo data interpolation.

A computational theory of stereo data interpolation was first proposed by Grimson [Gri 82, Gri 83a]. This theory incorporates an important constraint that the interpolant should not introduce any additional discontinuities other than those present in the form of zero crossings in the depth map. Grimson showed that an interpolant which can satisfy this constraint is yielded by the minimal solution to the functional given below:

$$\phi(f) = \left[\int \int \; (\delta f / \delta x)^2 + 2 \; (\delta f^2 / \delta x \; \delta y) + (\delta f / \delta y)^2 \; dx \; dy \right]^{1/2}$$

----- (1.1)

This functional is a measure of smoothness and minimizing the quadratic variation amounts to fitting the smoothest surface which is at least twice differentiable _every_ _where_. On the other hand, the points at which the depth is available from binocular correspondence could represent actual discontinuities in the local orientation of the

surface, e.g. edges. Thus, the interpolant tends to yield surfaces that appears "rounded" at edges and the discontinuous edges are lost. However, minimizing the quadratic variation introduces better noise immunity, especially noticeable in several instances, as stereo algorithms often generate spurious matches.

Interpolation of sparse data in a two dimensional grid has been considered in many other applications such as geological data extraction [Sch 76]. The nature of data in such applications are different from the the sparse depth map computed in stereopsis, because the data points need not represent edge points and there is no explicit need to preserve the edges. Considerations for the choice of interpolant employed in such cases include computational requirements and visually pleasing appearance. A part of the contribution in this research work is to introduce an important constraint that characterizes the stereo data and to use the constraint for an appropriate choice of the interpolant.

4.2. SHAPE FROM SHADING

The local orientation of the surface under view is in many ways responsible for the irradiance it renders to the viewer. The smooth shading of intensity on a uniform, non-reflective, continuous piece of surface can be attributed to various factors including local surface orientation, light source location and reflectance properties of the surface. Horn, modeled image generation as,

$$I(x,y) = G(z_x, z_y, l, v, \rho) \qquad \text{---- (1.2)}$$

where, $I(x,y)$ is the intensity value at a point (x,y) in the image plane, z_x and z_y are the partial derivatives of the surface at a

corresponding point on the surface, ι is the light source vector, υ is the viewer direction and ρ is the reflectance of the surface. The intensity can be directly related to the surface orientation and the light source alone if some generic assumptions can be made. Typically, in the literature, such assumptions are:

(i) Lambertian and

(ii) illuminated with a single point source at infinity.

Under orthogonal projection the intensity can be shown to be

$$I(x,y) = \rho(\iota \cdot n) \qquad \text{---- (1.3)}$$

where, n is the unit normal involving the partial derivatives as

$$n = [z_x, z_y, -1]^T / \| [z_x, z_y, -1] \| \quad \text{---- (1.4)}$$

The objective of shape from shading methods is to recover the normal n from the I(x,y) and known or unknown light source information.

Some of the important contributions to this area have come from Horn [Hor 75, Hor 77], Ikeuchi and Horn [IkH 81], Pentland [Pen 86], Brooks and Horn [BrH 85] , Lee and Rosenfeld [LeR 85], Frankot and Chellappa [FrC 87] and Ferrie and Levine [FeL 89]. In general, these algorithms can be classified into local or global methods depending on the nature of utilizing the intensity information. Local methods involve using the intensity information or its derivatives only at the point of interest, while global methods incorporate the notion of iterative computation of the solution constrained by global consistency. A taxonomy of the shape from shading algorithms is provided in Figure 1.5.

4.2.1. Local schemes
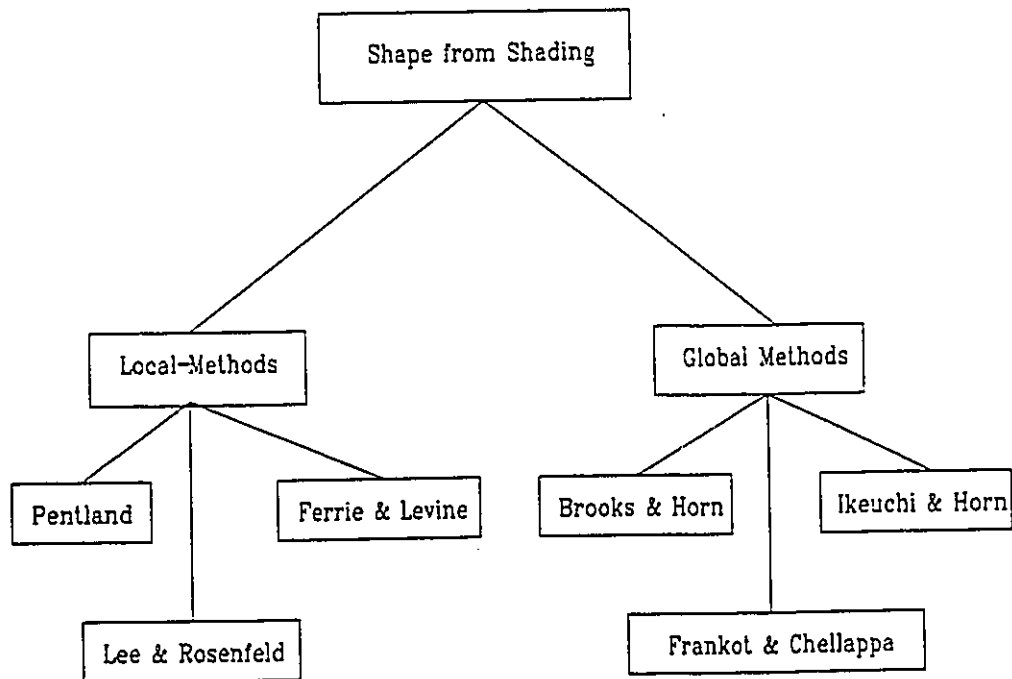
Figure 1.5 A Taxonomy of Shape from Shading Methods

Local methods for computing shape from shading was first proposed by Pentland [Pen 86]. This algorithm was based on an assumption that the surface is locally spherical. He obtained a description of the surface orientation in terms of 'tilt' and 'slant'. Although, his tilt estimator offers a good performance, the slant estimator suffered from loss of accuracy since it specifies only upper and lower limits for the slant quantity. However, Pentland's assumption of local sphericity is strong and restrictive and non-umbilical surfaces are quite common. Lee and Rosenfeld [LeR 87] have proposed a variation of Pentland's approach to improve the estimation of 'slant', by reconstructing the surface in the illumination coordinate system. Ferrie and Levine [FeL 89] have derived conditions under which accurate estimates of 'tilt' and 'slant' may be obtained.

Local methods suffer from the problem of poor noise immunity since they make use of only local information. Thus, noise at a specific point in the image can give rise to inconsistent estimates of orientation when compared to its neighbors. Furthermore, intensity derivatives are used in these methods, making the estimates even less reliable.

4.2.1 Global Methods

Global methods make use of local constraints such as continuity of normals to arrive at global interpretation of the scene. Ikeuchi and Horn [IkH 81] proposed the first global method to compute shape by solving a variational problem as follows: Minimize,

$$e = \iint \{(f_x^2 + f_y^2) + (g_x^2 + g_y^2) + \lambda \, [I(x,y) - R(f,g)]^2\} dx \; dy$$

$$----- \quad (1.5)$$

where, subscripted quantities are the partial derivatives of the respective functions and f and g are the stereographic projections of z_x and z_y given by

$$f = 2 z_x [\text{sqrt}(1 + z_x^2 + z_y^2) - 1] / (z_x^2 + z_y^2)$$

and $g = 2 z_y [\text{sqrt}(1 + z_x^2 + z_y^2) - 1] / (z_x^2 + z_y^2)$

The above minimization problem requires the pre-computation of the reflectance map R(f,g) that denotes the "expected" irradiance quantity for a given surface orientation. In practice, it is difficult to obtain the reflectance map for a given surface since it requires considerable amount of a priori knowledge. Horn and his associates abandoned the method for this reason and years later, Brooks and Horn proposed another global algorithm [BrH 85]. Their iterative scheme involved minimizing a quadratic functional

$$\iint (I - \rho(1.n))^2 + \lambda (n_x^2 + n_y^2) + \mu ||n^2 - 1|| \, dx \, dy \quad \text{--- (1.6)}$$

with respect to n. In the above expression,

λ is a scalar weight that emphasizes a smooth solution

$\mu(x,y)$ is a Lagrange multiplier to ensure that n has unit length

and n_x and n_y are the first partial derivatives of n.

However, it has been pointed out [FrC 87] that the solution based on the above technique may not always yield good results since the estimates of z_x and z_y may not satisfy the integrability constraint. Furthermore, this algorithm experiences considerably long convergence period as found in our implementation.

Frankot and Chellappa, in a recent paper [FrC 87], have proposed a modification of the technique of Brooks and Horn [BrH 85] by incorporating an additional constraint to ensure integrability. In

23

this approach, the authors use a finite sum of Fourier basis functions
to represent surface height z and minimize a distance measure

$$d = \int \int (z_x - \hat{z}_x)^2 + (z_y - \hat{z}_y)^2 dx\, dy \qquad \text{---- (1.7)}$$

with respect to the Fourier coefficients in the representation of z.
$\hat{z}_x$ and $\hat{z}_y$ are the current estimates of z_x and z_y. The authors have
obtained reliable results even when no a priori information on the
occluding boundaries is available. However, the need to use DFT and
IDFT would impose an additional computational burden on the algorithm.

The motivation for the major part of this dissertation was derived
from the current state of the art in shape from shading, in
particular, global techniques.

4.3 STRUCTURE FROM MOTION PARALLAX

Relative motion between an observer and the objects in a scene can
render a vivid feeling of relative depth. Displacement in the images
of moving objects in a monocular series of frames, (e.g. movie) can be
a useful source for computing relative depth or structure of the
scene. In general, the algorithms proposed in the literature to
compute structure from motion can be divided into two approaches,
namely, correspondence based and optic flow based. In the first
approach, a measurement of image displacement (similar to stereo
correspondence) is obtained at a number of points representing strong
features (e.g. point of intensity changes) followed by the computation
of motion parameters and relative depth. In the second approach, the
two dimensional motion of every point (if not, at least densely) in
the image plane is first obtained, followed by the estimation of
motion parameters. In both the approaches, an important assumption is

24

made that the object undergoing motion preserves its rigidity. Motion in 3-d space can be represented as

$$\dot{P} = T + \Omega \times P \qquad \text{---- (1.8)}$$

where, T is the translational velocity and Ω is the rotational velocity and P is the positional vector. The objective of motion estimation is to compute the components of T and Ω. To compute the relative depth, under perspective projection, the image plane coordinates of the projection of a moving point P can be written as,

$$x = X / Z \text{ and } y = Y / Z \qquad \text{------ (1.9)}$$

The projection of P in the image frame, denoted by p, can be seen to have moved to a new location due to the motion of P and thus, its velocity can be shown to be related to $\dot{P}$ by the following.

$$u = \dot{x} = [x\, T_z / Z - T_x / Z] + [xy\, \Omega_x - (1 + x^2)\, \Omega_y + y\, \Omega_z]$$
$$\text{------ (1.10)}$$

$$v = \dot{y} = [y\, T_z / Z - T_y / Z] + [-xy\, \Omega_x + (1 + y^2)\, \Omega_y - x\, \Omega_z]$$
$$\text{------ (1.11)}$$

The dotted quantities here represent the time derivatives. In the above expressions, u and v, the image velocities are known. Once the motion parameters T and Ω are computed, Z can be obtained from (1.10) and (1.11). However, it may be noted here that since Z appears in the denominator as a ratio with translation T and Z is also unknown before computing T, motion parameters can only be estimated up to a constant factor. Thus, the computed Z is a relative depth measure.

4.3.1 Correspondence Based Methods

Correspondence is obtained between successive frames of the imagery in the same way as stereo correspondence between left and

right images is obtained. Ullman [Ull 79] proposed the first correspondence based motion estimation based on an orthographic coordinate system. Although, his model was very simple by solving a set of linear equations, the assumption of orthographic projection in dynamic scene analysis imposes restrictions on the object moving towards or away from camera. Roach and Aggarwal [RoA 79] proposed to solve the problem via perspective imaging coordinates. Their method, however, involved a large search space motivating further refinements of the idea.

Later, Tsai and Huang [TsH 81], investigated the problem by first restricting the instances to locally planar objects. By exploiting the constraints of planar geometry, they showed that a set of "secondary motion" parameters are first obtained followed by the estimation of actual motion parameters by solving a sixth-order polynomial. They refined the idea in their later version to incorporate curved surfaces. An important result established by their work is that given unambiguous correspondences of eight object points in general positions, the motion parameters can be established uniquely from the singular decomposition of the matrix $E = T\Omega$. A parallel result was also derived by Longuet-Higgins LoH 81]. These ideas were extended to straight line correspondences instead of point correspondence, resulting in a better noise stability.

A major disadvantage of the correspondence based methods is obtaining the correspondence itself. Unlike stereo correspondence, motion correspondence suffers from motion blur too. Furthermore, since these algorithms make use of information available only at a few

points in the image, they tend to be very noise sensitive.

4.3.2 Optic Flow Based Methods

The optic flow techniques rely on local spatial and temporal derivatives of the image intensity values, instead of spatial correspondence. Horn and Schunk [HoS 81] showed that a useful constraint can be derived from rigidity and smoothness assumptions. That is, with image velocities u and v,

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t) \quad \text{---- (1.12)}$$

Using this constraint, Bruss and Horn [BrH 81] proposed their least-squares based algorithm to compute the motion parameters. Although the algorithm performed poorly on scenes with edges (hence discontinuous optic flow), their idea paved the way to motivate several others to construct optic flow based structure estimation. In a parallel effort, Jain introduced the notion of using the Focus of Expansion to compute optic flow and demonstrated interesting results [Jai 83].

Successive refinements to the theory came from Adiv [Adi 83], and Negahdaripour and Horn [NaH 87]. Recently, Subbarao [Sub 89] proposed a technique that involves temporal derivatives of the image apart from spatial derivatives that were considered until now. Verri and Poggio [VeP 89] have also recently added their critical evaluation of the usefulness of optic flow to compute structure.

For dynamic scene analysis, motion parallax contributes to much more relative depth and shape information, than cues such as focus adaptation and size constancy.

27

4.4. SHAPE FROM TEXTURE

Regular textures in perspective imagery present a uniform change in the distribution of the texture elements. The changes include slow variation in the density of the texture layout and skewed texture elements. A significantly small attention to recover shape from texture has been found in the literature. This may be partly explained by the fact that regular textures are rare in nature and thus applications are much restrictive.

The models proposed in the literature [Wit 81, DaJ 83] , in general, attempt to extract the basic texture elements first, followed by the estimation of surface orientation from any deviation from the distribution of the texture elements. Variation in the distribution of texture elements are found by evaluating the repetitive occurrence of directional edges. Witkin [Wit 81] proposed a maximum likelihood model for recovering surface orientation by relating regular variations of the texture elements to the slant and tilt of the surface. Davis et al [DaJ 83] improved his technique by presenting two algorithms, one of them based on computing the edge direction histogram and the other based on the maximum likelihood principle. All these methods assume that the edge directions are isotropically distributed.

The crucial problem of all these methods is the extraction of texture elements. Very recently, Blostein and Ahuja [BlA 89] suggested an approach to efficiently identify texture elements by combining this process with extraction of the scene layout. Their method is based on generating a list of candidate texture elements by using a variable spatial scale filter such as $\nabla^2 G$. The best candidate is suggested by a

28

planar surface patch that best describes the expected distribution of the texture elements.

Useful insights into how human vision reconstructs surface orientation from texture have been achieved by the psychophysical studies of Stevens [Ste 83]. However, such ideas are yet to be incorporated into any machine vision algorithm.

5 Motivation for this work

Most of the progress in the area of vision has been realized in understanding early vision. Two reasons emerge to answer why early vision is studied well. First, the recent advancements in single cell electrode methods to analyze characteristics of individual cells in the visual stream have given rise to a much better understanding of early vision than higher levels. Secondly, since early vision allows mathematical modeling of the processes involved, it offers an ease in understanding such processes.

In addition to the reasons provided here for choosing to work with early vision, we will also describe below, the motivational theme for studying stereopsis and shape from shading. Stereo vision has been shown to be one of the most reliable ways to estimate depth. The crux of stereopsis is to identify the corresponding items in the left and right image, followed by extraction of a depth measure. Correspondence can be established without any ambiguities only at points that possess strong features. In general, since these points are only a few in number, the problem here is how to use the depth measurement at these points to compute depth at every other point. This particular problem, although being addressed as scattered data interpolation in general,

29

has a different implication in vision. A computational framework for this problem was provided by Grimson [Gri 82, Gri 83a]. An understanding of his theory reveals that a stronger form of the constraint he imposed on the interpolant is desirable. Our motivation for suggesting a modified computational theory and a faithful implementation for that theory originates here.

With the objective of building the equivalent of a 2 1/2 d sketch, the approach taken here is to model the integrating process of combining stereopsis and a monocular shape computing process. A rich monocular source of shape information is the smooth shading on an object. Since it is a more commonly occurring form of shape cue than any other monocular cue, our goal here is to model shape computing from shading. The current literature on shape from shading is far from being as rich as compared to the literature on stereopsis. The reported work stems from a perspective that de-links shape from shading from any other surface computing process. That is, only after the surface reconstruction is over, the shape inference process is assumed to begin. In contrast, shape from shading should actually use local shape constraints in the process of reconstructing the surface since shape information is intrinsic to the object. Hence our pursuit is to devise a method that uses local shape constraints to arrive at globally consistent estimates, while the computed shape in this process is a very useful byproduct.

In short, from a perspective of developing the 2 1/2 d sketch, it is necessary to view that surface computing processes are not independent. A concern for generating object intrinsic information

remains the motivation for this work.

6 Organization of this dissertation

Beginning with chapter 2, interpolation of sparse data from stereopsis is the subject of chapter 2. In chapter 3, a model for describing shape is presented. An algorithm for computing shape from shading using this local shape descriptor is described in chapter 4. The issue of integrating depth from stereopsis and shape from shading is addressed in chapter 5. The conclusions and details of further work are discussed in chapter 6.

Chapter 2

Surface Interpolation

1 Introduction

The problem of interpolating stereo data needs to be understood with not only pure mathematical considerations, but also with considerations about the visual geometry. Because the constraints about the visual world are best understood at the level of a computational theory, a good starting point here is to analyze a computational theory of interpolation.

2 A computational theory of Surface Interpolation

It may be recalled, from chapter 1, that stereo correspondence can not take place at all points in the image but only at points of intensity changes. In order to reconstruct the surface in its entirety, the sparse stereo data needs to be interpolated to deliver depth at every point. Given the sparse depth map, virtually an infinite number of surfaces can be fitted. This can be illustrated using an example in a two dimensional plane. Consider the set of points shown in Figure 2.1a. Several interpolants can be used to interconnect these points, suggested by the alternatives in Figure 2.1. To distinguish the "best fit" from the rest, we need to employ a few constraints such as smoothness and minimal mean-square deviation from the original data. Such useful constraints form the central part of a computational theory.

2.1 PROBLEM STATEMENT

Let $f(x, y)$ be a function of depth at points $P(x, y)$ defined in the grid (x, y). For convenience, we will order $P(x, y)$ in a set $Q =$
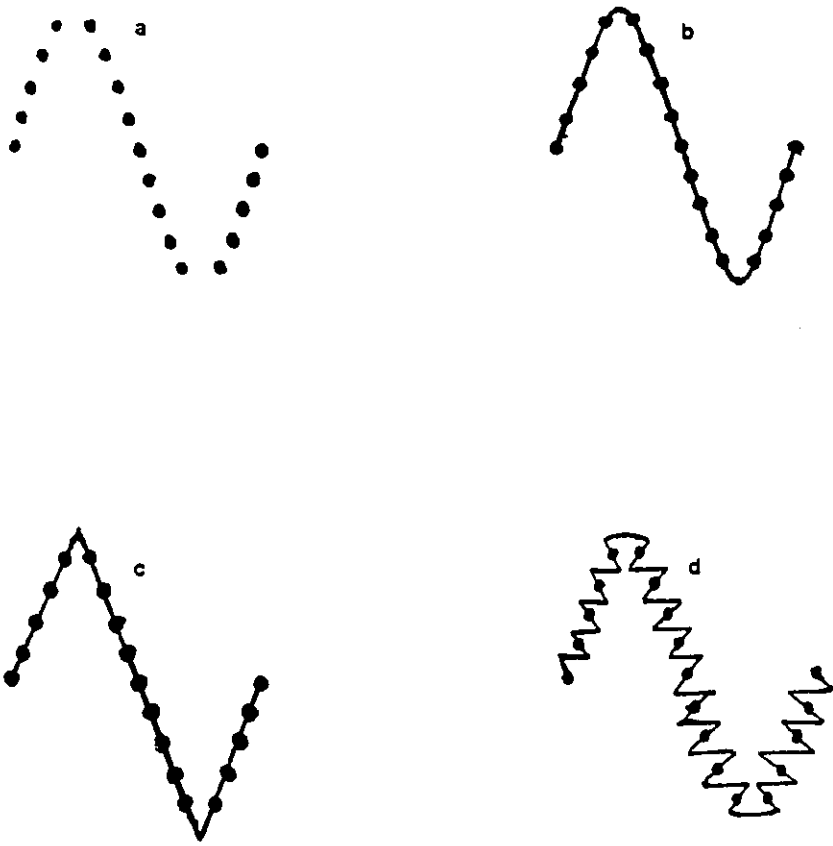
Figure 2.1 A two dimensional example for showing multiple
candidates of interpolants

$\{P_k\}$; $k = 1,N$. Given the known data set $G = \{P_i \ni f(x, y) \mid_{P_i} = f_i\}$ and $G \subseteq Q$, find $f(x, y)$ for all $P_k \in Q$.

2.2 SURFACE CONSISTENCY CONSTRAINT

A few observations on the nature of data points P_k's we have here can be translated to form a very useful constraint for interpolation. Firstly, the sparse set of data points at which we have the stereo disparity are the points of intensity changes. Secondly, the intensity changes can only come from either surface discontinuities or surface albedo changes. This means that since surface discontinuities are truly captured by zero crossings of the $\nabla^2 G$ filtered image, the interpolant should not introduce any additional discontinuities in the surface. Thus, in the two dimensional example given in Figure 2.1, except the choice b every other choice would introduce additional edges. The surface consistency constraint of Grimson [Gri 83b] is based on these observations. It appears here as,

"The absence of zero crossings constrains the possible surface shapes at a given point."

Developed from this constraint was his surface consistency theorem [Gri 83b], which relates the probability of a zero crossing introduced by an interpolant and shape of the surface. Using this theorem he showed that the best surface fit 'f' for the set of points P in W is the minimal solution for the quadtratic variation of the interpolant function. This functional was described in chapter 1 as equation (1) and for convenience it is reproduced here.

$$\phi(f) = \int \int (\delta f / \delta x)^2 + 2 (\delta f^2 / \delta x \, \delta y) + (\delta f / \delta y)^2 \, dx \, dy$$

$$------- \quad (2.1)$$

34

By associating a measure of such quadratic variation with each possible surface, a hyper surface of solutions is generated with a dimension of (N^2+ 1). The minimal point on this hyper surface generated by the functional given by (2.1), corresponds to the desired surface. It should be noted here that the candidate surfaces represented by this functional are ensured to be at least twice differentiable and hence obey the surface consistency constraint. By requiring that the quadratic variation is minimal the intepolant bears a great deal of similarity to least squares regression type of curve fitting of points in two dimension..

2.3 MODIFICATIONS TO GRIMSON'S COMPUTATIONAL THEORY

Our motivation to study the problem of stereo data interpolation towards further enrichment of Grimson's theory stems from the following observations.

(i) The requirement that the interpolant should not introduce any new zero crossing has been rightly captured in Grimson's theory of interpolation. Evidently, if the visual system uses only continuity as a criteria then sharp corners and edges in any percept are unaccountable. However, the surface consistency constraint of Grimson does not explicitly specify whether the existing zero crossings are preserved by the interpolant or not. That is, the discontinuities in the interpolated surface should dictate zero crossings at the same original locations and all those original zero crossings must appear as representatives of discontinuities even after interpolation. Hence, a modified form of surface consistency constraint should read:

" The interpolant must preserve all (and only all) of those zero

35

crossings that correspond to the discontinuities in the surface at the same locations. "

(ii) To satisfy the above constraint, the interpolant should be chosen such that it is at least twice differentiable every where except at the data points (zero crossings), where it should not even be differentiable once but just be merely continuous.

(iii) Computing the depth using this interpolant should also be computationally less expensive and parallely implementable. Because of choosing the smoothest surface, Grimson's quadratic variation algorithm not only fails to satisfy the modified constraint but also is computationally quite expensive (see Section 5.2).

(iv) The interpolant should also facilitate combining any estimates of surface orientation from other cues since our ultimate goal is to build the 2 1/2 d sketch.

Shown in the following section, is a scheme due to Shepard [She 76] that was originally developed for filling scarce data in the domain of geophysical experimentation. Before a detailed treatment of Shepard's surfaces, an informal preview would help to get a better understanding here. Imagine a set of thin poles fixed on a scattered fashion in a grid, representing the depth data at those points. Fitting Shepard's interpolants is metophorized to laying a thin membrane over these poles. The curvature of sagging in the membrane due to gravity can be controlled by adjusting the tension in the membrane. In a similar fashion, the fit of Shepard's surfaces can be adjusted through proper choice of the parameters of the interpolant. The central idea of this scheme, thus seems to be simple, but yet

yields remarkable results as implemented in this work.

3 Shepard's Metric Interpolant

Let $f(P)$ be a function of the point $P = (x,y)$ defined for all the points in a bivariate real plane R^2. Assuming that the points are arranged in an ordered set $\omega = \{P_i\}_{i=1}^{N}$ with all of them being distinct, we shall denote the following quantities. The function value f_i corresponds to point P_i. The Euclidean distance between a generic point P and P_i is d_i. Then the Shepard's function at a generic non data point P is specified as

$$f(P) = \left[\sum_{i=1}^{N} f_i \prod_{i \neq j} d_j\right] / \left[\sum_{i=1}^{N} \prod_{i \neq j} d_j\right] \qquad \text{--- (2.2)}$$

where, $d_i = \| P - P_i \| = \{ (x - x_i)^2 + (y - y_i)^2 \}$

It can easily be seen that $f(P)$ is continuous and interpolates F at all points in ω. The continuity of f results because the d_i's are continuous and the denominator of (2.2) never vanishes, because the points are distinct. The evaluation of (2.2) at point P_k thus gives,

$$f(P) = \left[f_k \prod_{j \neq k} \|P_k - P_j\| \right] / \left[\prod_{j \neq k} \|P_k - P_j\| \right] = f_k$$

since all other terms vanish.

The numerator and denominator of (2.2) have many terms in common. A simpler form is

$$f(P) = \left[\sum_{i=1}^{N} f_i/d_i \right] / \left[\sum_{i=1}^{N} 1/d_i \right] \qquad \text{----(2.3)}$$

Since the P_i's are distinct, 2.2 can also be written as,

$$f(P) = \sum_{i=1}^{N} f_i \, \phi_i \, (P; P_1, P_2, \ldots P_N)$$

where,

37

$$\phi_i(P;P_1,P_2,\ldots P_N) = \prod_{i \neq j} ||P-P_i|| / [\sum_{i=1 j \neq i}^{N} \prod ||P-P_j||]$$

so that $\phi_i(P;P_1,P_2,\ldots P_N)$ are the cardinal basis functions and they possess a very important property that they are almost analytic (and hence infinitely differentiable) everywhere except in the vicinity of data points. This property is the most desirable property in the realm of stereo data interpolation to satisfy the modified version of surface consistency constraint discussed earlier. To strengthen our argument for using Shepard's interpolants for filling stereo data we shall discuss some important properties of this interpolant in the following section.

3.1 PROPERTIES OF SHEPARD'S SURFACES

A considerable amount of study of Shepard's surfaces has been published. This scheme is quite simple but yet possesses very interesting properties [BaD 83, GoW 78]. Any interpolant should satisfy an important requirement, namely, it should be bounded between the infinimum and the supremum of the function values in the data set. That is,

$$\min(f_i) <= f(P) <= \max(f_i) \; \forall \; i \ni 1 <= i <= N.$$

It is quite easy to verify that 2.2 does indeed satisfy this requirement.

Shepard's scheme also possesses an interesting asymptotic behavior. If, for instance, a non-data point is far from a cluster of data points, then the interpolant takes the average value of the cluster since every individual point in the cluster has equal influence over the non-data point. That is,

$$\lim_{d_i \rightarrow \infty} f(P) = 1 / N \sum_{i=1}^{N} f_i \quad \forall \; i \ni 1 <= i <= N.$$

It may be recalled from the previous sections that we could control the amount of sagging of the membrane surface laid over the data points. To see how this can be done, an additional parameter 'μ' is introduced in (2.3)

$$f(P) = \left[\sum_{i=1}^{N} f_i / d_i^{\mu} \right] / \left[\sum_{i=1}^{N} 1 / d_i^{\mu} \right] \quad \text{---}(2.5)$$

The effect of introducing μ can be directly seen to influence the support offered by each data pole. Figure 2.2 shows different instances of interpolation in the one dimensional data set for different values of μ. It is interesting to note that for $\mu > 2$, the surface becomes 'flat' near the data points. For $0 < \mu < 1$ sharp cusps are noticed at the data points. This particular characteristic has been studied well in [Far 86]. Since, for $\mu = 2$ the interpolant resembles the least square quadratic fit but yet not differentiable at data points, we shall use this value.

Finally, since the contribution due to farther points in the grid is quite minimal ($\left[1/d_i^2\right]$ is small), the computational expense incurred in computing them can be avoided by limiting the influence of data points to a confined neighborhood of small radius r. Thus we have,

$$f(P) = \left[\sum_{i=1}^{N} f_i / d_i^{\mu} \right] / \left[\sum_{i=1}^{N} 1 / d_i^{\mu} \right] \quad \forall \; i \ni d_i <= r \quad \text{---}(2.6)$$

In our implementation the local version shown in 2.6 has been used. The critical issue is how "local" the scheme should be such that we do not lose much of the information. Evidently, this question has a
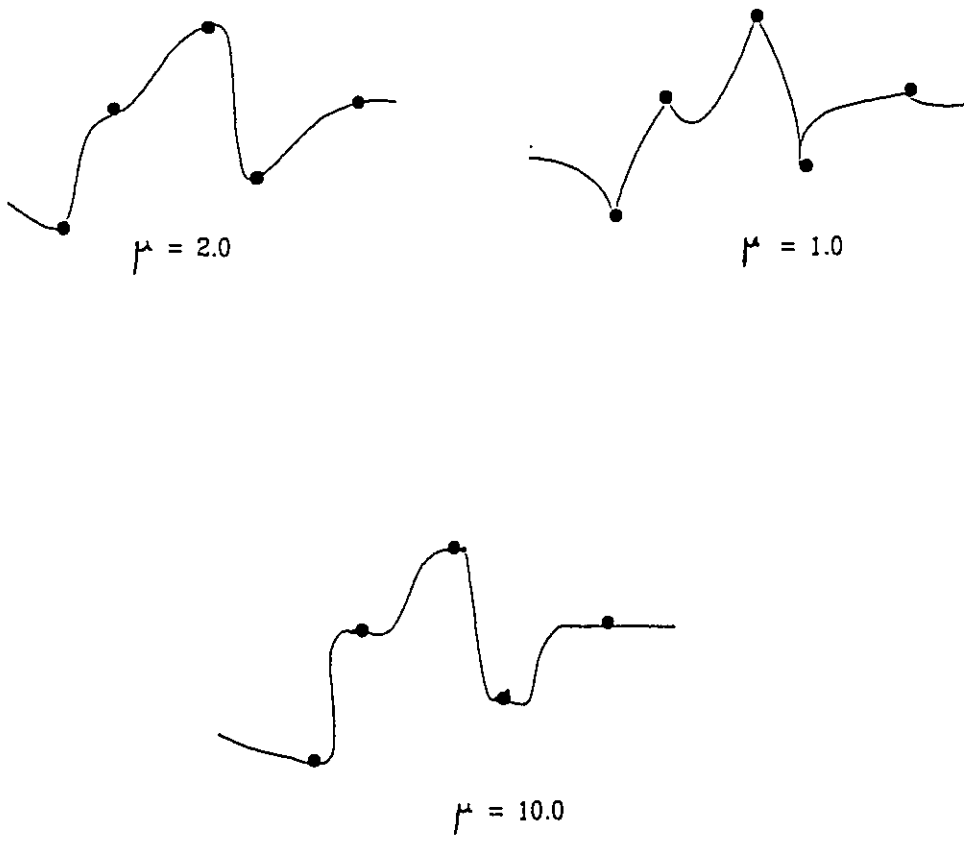
Figure 2.2 Effect of μ in Shepard's Interpolant

direct bearing on the distribution of data points in the grid. A discussion for our choice of r and a justification for this choice are provided in the next section.

All the properties discussed in this section hold good for any other metric other than the Euclidean metric. The metric used in our implementation defines distance between two points in the grid as the maximum difference in x and y coordinates. That is,

$$d = \max(x_p - x_n, \ y_p - y_n)$$

where, $(x_p, \ y_p)$ and $(x_n, \ y_n)$ are the coordinates of the data and non-data points respectively in the image plane.

4 Shepard's Surfaces over Stereo Data

We shall show here how the choice of Shepard's interpolants for stereo data interpolation makes it the appropriate one. It may be recalled from Section 2.3 that the modified version of surface consistency constraint requires that the interpolant should be at least twice differentiable at all non data points while not even differentiable once near the data points. One of the salient features of Shepard's surfaces is that they behave analytic at all non data points. However, the fitted surface is not even differentiable once near the data points, although it appears continuous visually. This particular property satisfies the surface consistency constraint faithfully.

To consider the issue of deciding on how "local" the interpolation neighborhood can be, the notion of spatial frequency channels in early vision is discussed here. Because the information seems to be captured at different levels [Mar 77a], it makes sense to compute zero

crossings of the LoG at different resolutions. Thus, the the filter width or the standard deviation we choose for the Gaussian allows us to extract edge information at a certain scale. Furthermore, the standard deviation σ of Gaussian controls the density of the zero crossings present in the particular channel. Hence σ itself can be taken to be directly related to the width of neighborhood for interpolation. From our experiments it suffices to use one σ as the width of interpolation.

5 Results and discussion

As indicated earlier, since the RDS is one of the rigorous ways to test a stereo vision module, we consider here a 128 X 128 RDS (Figure 2.3 and 2.4) generated with a central square in the middle given a disparity of about 5 pixels. For most humans this is well within the fusion range. After the generation of RDS, the left and right images were convolved with LoG to obtain the intensity changes. The width of Gaussian was chosen to be 5 pixels. This choice is supported by the biological finding that the lowest spatial frequency channel human vision employs corresponds to 4 pixels [Gri 85, MaP 79]. The sparse depth map is generated using a contour based stereo algorithm [SrK 87]. (see Figure 2.5). Figure 2.6 shows the filled depth map as brightness values. Clearly, the edges of square in depth are preserved quite well.

Also, considered here is a scene containing a basketball (Figure 2.7 and 2.8). The camera was fixed with the wall on the background. The result of reconstructed surface is shown in Figure 2.9 as a perspective plot of the surface. It may be noticed that stereo

42
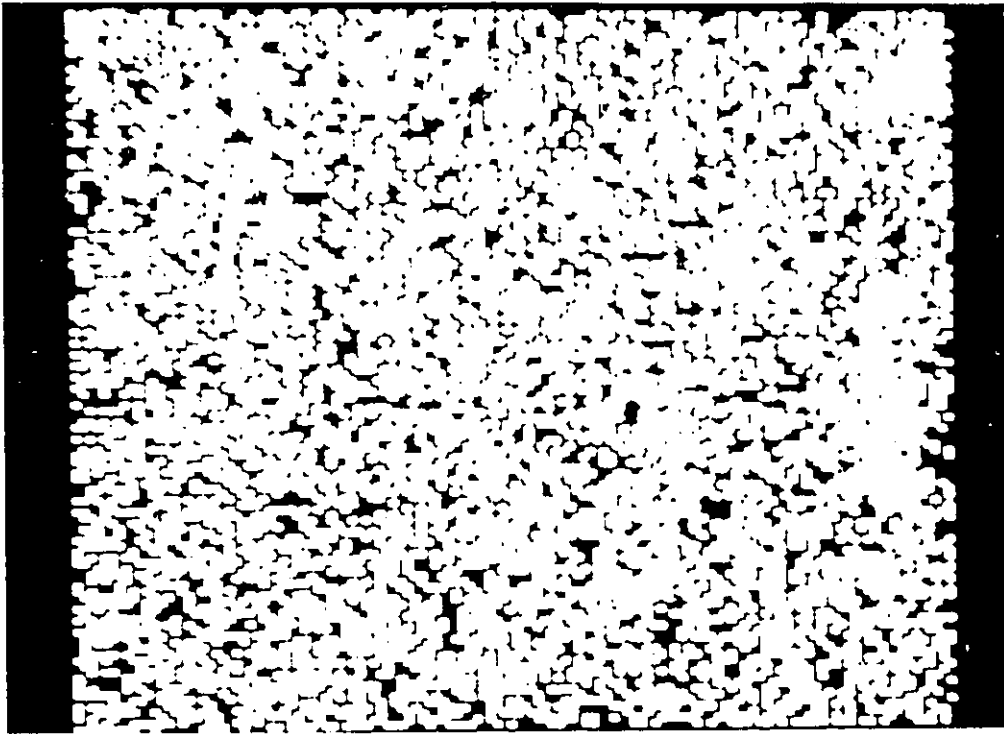
Figure 2.3 Left Random Dot Stereogram
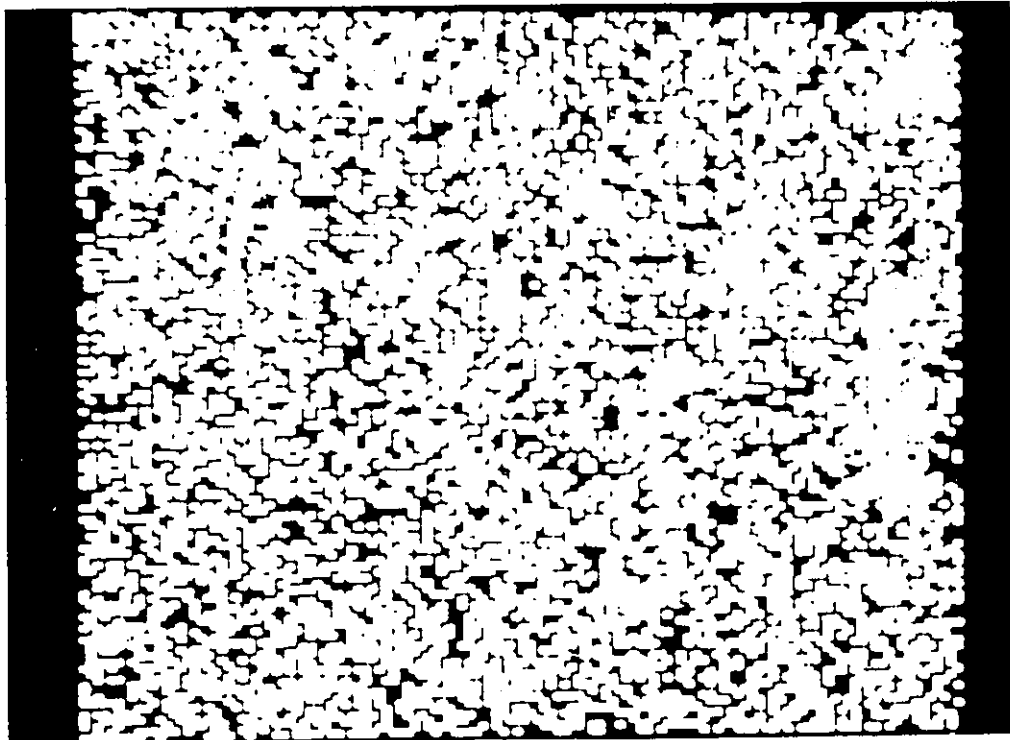
Figure 2.4 Right Random Dot Stereogram

43
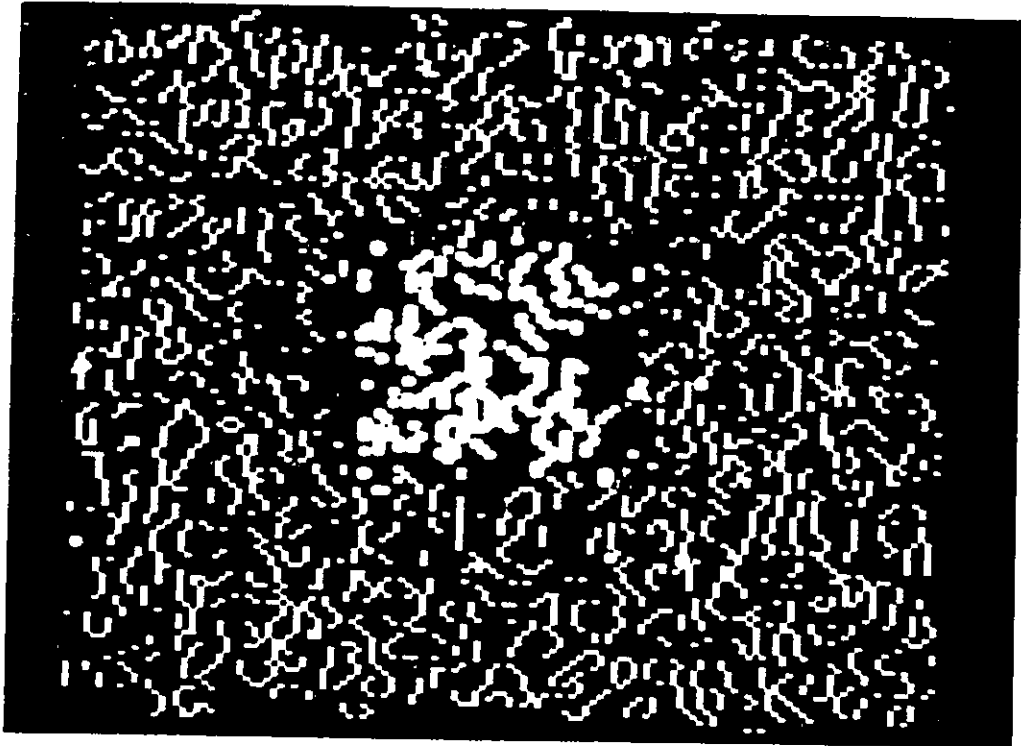
Figure 2.5 Sparse Depth map of the RDS
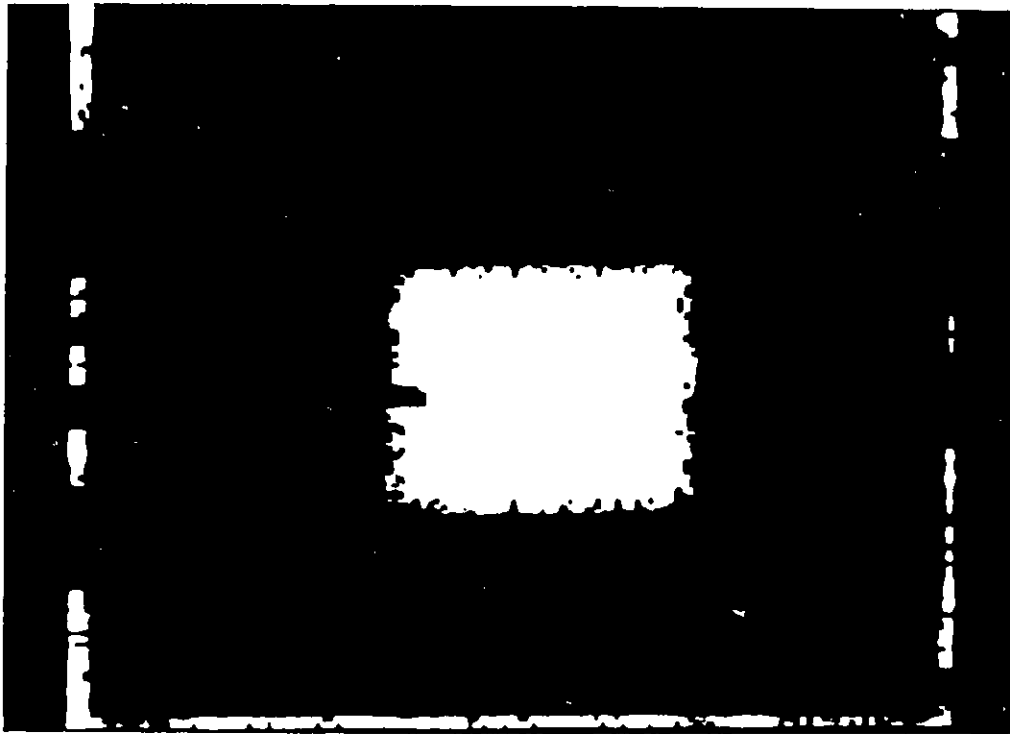
Figure 2.6 Interpolated RDS

44
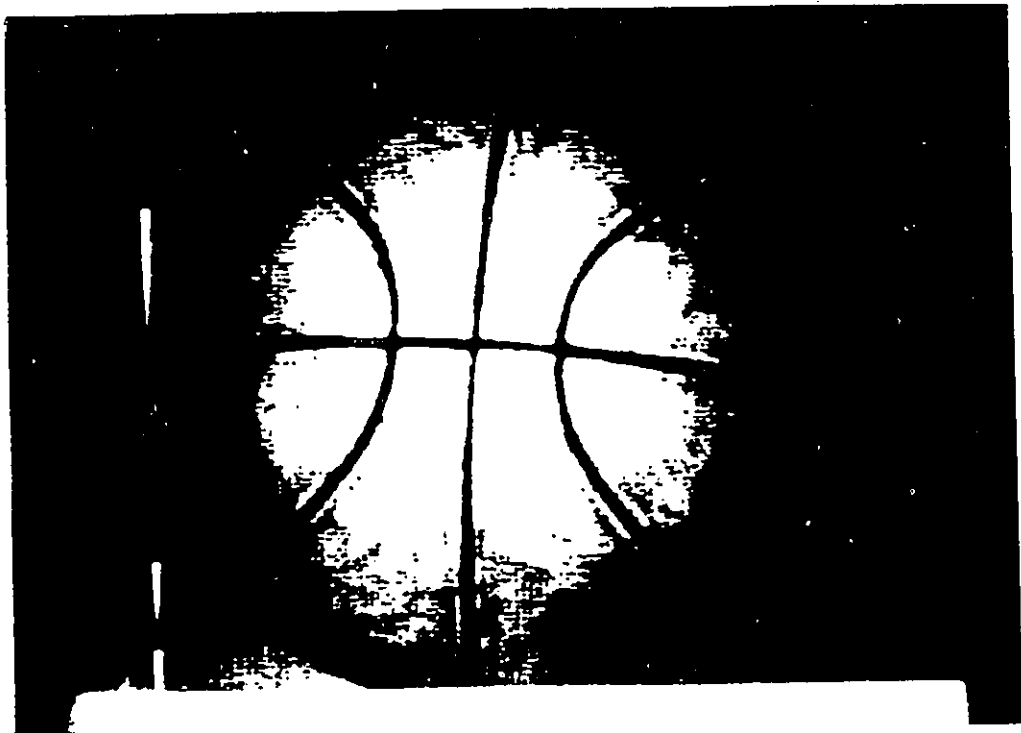
Figure 2.7 Left picture of the basket ball
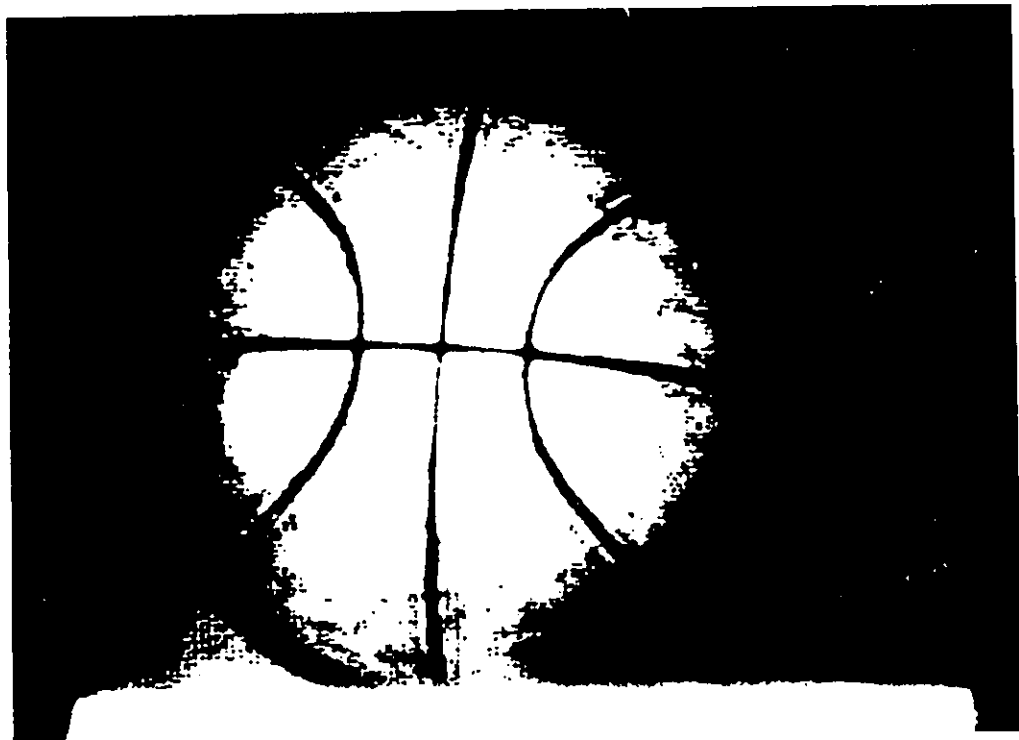
Figure 2.8 Right picture of the basket ball

45
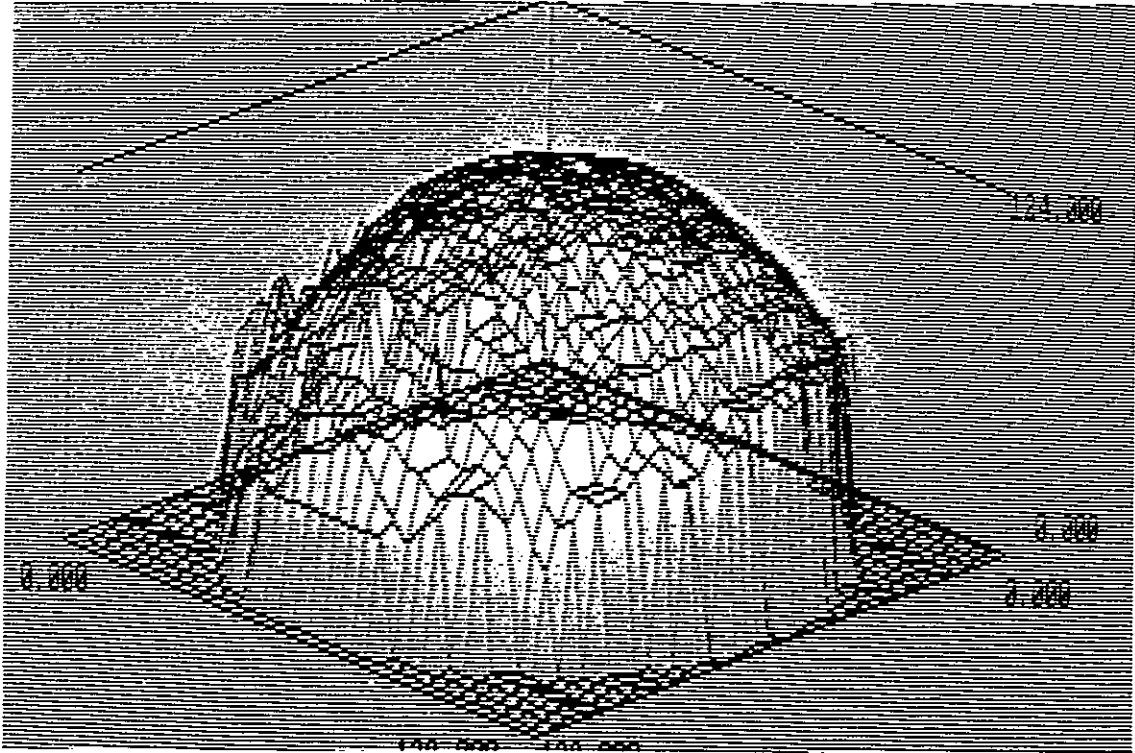
Figure 2.9 Interpolated surface (without the removal of hidden surface)
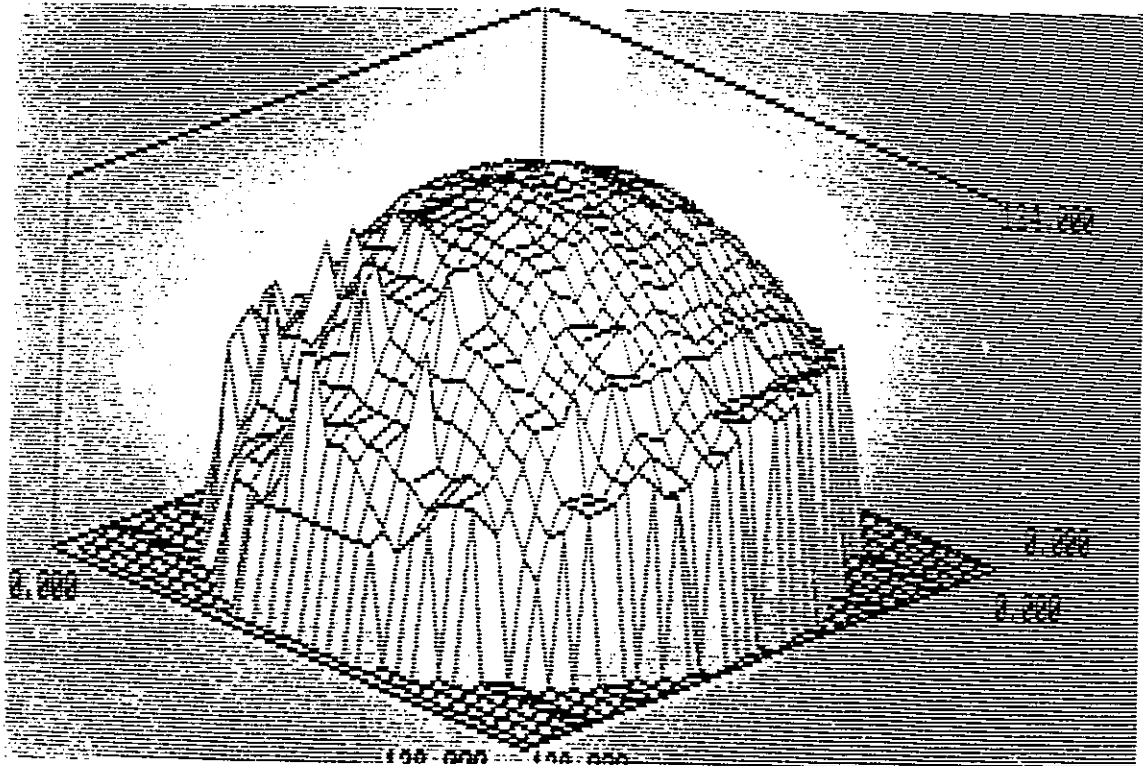
Figure 2.10Interpolated surface (with the hidden surface removed)

46

correspondence algorithms can leave a few spurious matches and these spurious matches indeed cause difficulties as noisy data. Unfortunately, the Shepard's interpolant considers these points also as valid data points resulting in a few thorn like peaks at these points. However, the reconstructed surface does preserve its shape. The reason for this is because, the interpolation scheme is similar to a weighted average scheme. Under noisy data conditions the quadratic fit of Grimson might tend to smooth out the effect of noise rendering the result more visually pleasing.

5.1 COMPUTATIONAL COMPLEXITY

Though it is difficult to obtain a precise count of basic operations, it can however be specified in terms of two parameters as follows. Consider the local form of the interpolation scheme given by expression

$$f(P) = \left[\sum_{i=1}^{N} f_i / d_i^{\mu} \right] / \left[\sum_{i=1}^{N} 1 / d_i^{\mu} \right] \; \forall \; i \ni d_i \leq r \; \text{---(2.6)}$$

In the local scheme N depends on the radius of interpolation and in our case it is the lowest σ used in zero crossing detection. Also, N depends on the density of zero crossings in the image which could vary from scene to scene. Hence, we shall use two factors, α and β, to denote the number of pixels within the given radius, and the density of zero crossings, respectively. With $N = \alpha\beta$, we have, for every nondata point in the image, $(\alpha\beta + 1)$ divisions, $\alpha\beta$ multiplications and $2\alpha\beta$ additions to compute expression (2.6). Since there are $(1 - \beta)$ m^2 nondata points in an m x m frame, we have $(1 - \beta) m^2 (\alpha\beta + 1)$ divisions, $(1 - \beta) m^2 \alpha\beta$ multiplications and $2\alpha\beta (1 - \beta) m^2$ additions.

47

With $\alpha = 5^2 = 25$ and $\beta = 0.4$, the number of divisions involved can be shown to be 6.6 m^2, the number of multiplications 6 m^2 and number of additions as 12 m^2. Evidently, the locality of interpolation plays a key role in the amount of computation; this is reasonable because the denser the information available, the shorter the time it should take to complete the grid without having any need to compute depth at the data points. In general, there are about 25 m^2 basic operations.

5.2 COMPARISON WITH GRIMSON'S APPROACH

The important differences arise both in the computational theory level and the algorithmic implementational level. At the computational theory level, the need to preserve the original discontinuities is totally absent in Grimson's computational theory. This suggests that his minimal quadratic variation fit does tend to smooth out all the edges. At the algorithmic level, the computational requirements for Shepard's scheme is only 25 m^2 basic operations as opposed to 68 m^2 operations per iteration in Grimson's algorithm. The difference stands out, especially, when we notice that the number of iterations Grimson reported is of the order of hundreds, while Shepard's scheme is non-iterative. Furthermore, since there is no need to compute the interpolant value at data points, Shepard's scheme amounts to decrease in the computational time. However, in the minimal quadratic variation fit, the dimension of the hyper surface increases for every additional data point resulting in the increase of computational requirements. In short, for a quick and rough estimate for the depth values, Shepard's scheme is quite superior to Grimson's algorithm. In early vision, this type of quick and rough estimates are more important than slow and

accurate recovery of shape and depth. Such quick estimates are also useable in other shape estimation processes such as shading.

An important aspect in which Grimson's method excels is its better performance over noisy stereo data. Because of its second order differentiable property (every where continuous), it tends to smooth out noise if the mean square level of the noise is comparable to the data. However, since noise is also considered to be a potential data point by Shepard's scheme, it does not suppress the noise problem. Thus, depth estimates tend to be less accurate in such cases. This emerges to be a potential problem, since often stereo algorithms may not deliver 100% correct matches.

6 Summary

Interpolation, both in form and depth, has a definite place in human vision. In depth reconstruction stereopsis, as a primary source of depth, depth information is provided only at points that have strong features (points of intensity changes). Thus, interpolation becomes essential in order to fill the sparse depth map. Although the problem of scattered data interpolation has been studied well outside the realm of stereo vision, Grimson's theory of interpolation is the only attempt ever made to rigorously study this problem. In our work, the weak surface consistency constraint is modified to impose stronger conditions on the choice of interpolant. Since, Grimson's quadratic fit method may not satisfy the stronger form of surface consistency constraint, our choice of Shepard's metric interpolation scheme is defended here. With the depth at every point being available, a part of our objective, namely to reconstruct the surface is achieved.

49

Chapter 3

Shape Analysis

1. Introduction

The ultimate goal of computer vision is to extract some form of invariant description of the objects in order to recognize them. Several possibilities exist to describe visible surfaces of the objects. For example, in computer graphics, standardized descriptions of local patches such as Coon's and Bezier patches [Yor 81] have been used. However, since the shape of an object is invariant under transformations such as translation and rotation it becomes important to devise invariant schemes to describe surfaces by representing them with shape characteristics such as curvatures. A few representatives of earlier attempts on using curvatures specifically for describing shapes of objects include Besl and Jain [BeJ 86], Pentland [Pen 84] and Horn and Ikeuchi [HoI 83]. An impressive demonstration of how Gaussian and mean curvatures provide a convenient form of representing shape is presented by Besl and Jain. Pentland, in his attempt to solve shape from shading, discusses differential geometry of surfaces from a perspective of locally describing a surface as an umbilical patch. Horn and Ikeuchi restrict themselves to description of convex surfaces by a discrete approximation to Gaussian curvatures. Our approach parallels with that of Besl and Jain in the following respect. We begin with the motivation of locally describing shapes towards the goal of formulating a vision problem (in our case, shape from shading) in terms of the shape quantities.

The purpose of this chapter is to provide a comprehensive

treatment of shape description. The motivation is that a local shape descriptor derived in this chapter can be used to formulate the problem of shape from shading directly in terms of quantities representing the shape and surface normals. Section 2 provides a review of differential geometry of surfaces. A polynomial shape fitting approach, resulting in a convenient local shape descriptor, is described in Section 3. Section 4 lists the conclusive remarks on our approach.

2. Differential Geometry of Surfaces

In order to provide the motivation for using a local shape descriptor, a short review of the analytical results in differential geometry of surfaces is presented in this section. We begin with a parametrized description of a surface in 3-d space.

$$S = \{(x,y,z) \in R^3, \ x = f(u,v); \ y = g(u,v); \ z = h(u,v)\}$$

$$\text{---- (3.1)}$$

where $f(u,v)$, $g(u,v)$ and $h(u,v)$ are continuous functions of the parameter pair (u,v). Following the notations adopted by Besl and Jain [BeJ 86], the first and second fundamental forms of describing S are given by,

$$\phi(u,v,du,dv) = dp \cdot dp = [du \ dv] \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix}$$

$$\text{---------- (3.2)}$$

and

$$\psi(u,v,du,dv) = - dp \cdot dn = [du \ dv] \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix}$$

$$\text{---------- (3.3)}$$

51

where,

p is the generic point on the surface, n is the normal at p and, α and β are given by

$$\alpha_{11} = P_u \cdot P_u \ , \ \alpha_{22} = P_v \cdot P_v \ \text{ and } \alpha_{12} = \alpha_{21} = P_u \cdot P_v \ \text{---} \ (3.4)$$

$$\beta_{11} = P_{uu} \cdot n \ , \ \beta_{22} = P_{vv} \cdot n \ \text{ and } \beta_{12} = \beta_{21} = P_{uv} \cdot n \ \text{---} \ (3.5)$$

Intuitively speaking, the first fundamental form (3.2) describes corresponding changes in surface for small changes in parameter space, while the second fundamental form (3.4) describes the corresponding change in normals as well as the surface for small changes in the parameter space.

Equations (3.2) and (3.3) together contain adequate information about the surface. Also, the matrices α and β are very useful quantities here to compute the Gaussian and mean curvatures. A powerful analytical result established in differential geometry relating α and β with curvatures is as follows [Doc 76]. A matrix known as Weingarten mapping matrix can be formed out of α and β, given to be

$$W = [\alpha^{-1}] \ [\beta] \qquad \text{------} \qquad (3.6)$$

and the curvatures are related to W as

Gaussian curvature K = det(W)   ---- (3.7)

and Mean curvature H = 1/2 trace (W)   ---- (3.8)

The existence of α^{-1} is conditioned on the fact that the parametrization functions are continuous. Further details such as dichotomy between principal curvatures and, Gaussian and mean

52

curvatures can be found in [BeJ 86, Doc 76]. The issue of paramount importance here is how to formulate the problem of shape extraction from intensity measurements, so that a shape descriptor similar to W can be directly computable. The next section is specifically aimed at addressing this issue.

3. Shape Characterization

The simplest approach to extracting shape information from measurements is to first obtain an algebraic description of the surface and then compute the shape parameters by differentiation. For example, instead of the implicit representation given by 3.1, we can obtain an explicit characterization

$$z = f(x,y) \quad ---- \quad (3.9)$$

and compute the shape information as a function of the first and second derivatives of the surface. The primary difficulty with such an approach is that these partial derivatives increase as $\tan(\theta)$ where θ is the angle between the normal to the surface and the z axis. Thus, for instance, near the points at which the surface quickly recedes from the viewer the partial derivatives numerically reach extremely high values. This poses some computational difficulties because computing shape near the occluding boundaries requires subtraction of two (nearly equal) numerically large quantities resulting in a loss of accuracy.

A sensible approach to this problem is to formulate the early visual processes in such a way that shape quantities can be directly computed instead of obtaining shape by successively differentiating z. Thus, if a shape fit for the surface normals is first obtained, the

53

surface itself can then be constructed by integrating, resulting in a more numerically robust technique as opposed to differentiation.

Local shape of the surface can be described in many ways. For example, parametrized local patches around an intrinsic axis of the object have been used earlier [Yor 81, MaN 78]. Finding such intrisic frames can be avoided if we could characterize the surface in terms of the principal curvatures. Since principal curvatures are intrinsic to the surface and do not depend on the coordinates used, they seem to be the most appropriate candidates for invariant description of the surface. We shall discuss, in the next section, a shape fitting technique that overcomes the problems associated with the explicit formulation in (3.9) described before.

3.1 SHAPE FITTING AND A LOCAL SHAPE DESCRIPTOR

Let S be a surface and p a point on it. Let n(p) be the unit normal to the surface at the point p. In the shape fitting formulation, we begin with implicitly describing S in the form of a function of its normals as follows.

$$n(p) = f(p) \quad \forall \ p \in S. \qquad \text{----- (3.10)}$$

where, $p = [x_1, \ x_2, \ x_3]^T$ and $n = [n_1, \ n_2, \ n_3]^T$. Then the surface can be constructed by integrating the differential equation

$$n^T dp = n_1 dx_1 + n_2 dx_2 + n_3 dx_3 = 0 \qquad \text{----- (3.11)}$$

where, $[dx_1, \ dx_2, \ dx_3]^T$ is a vector in the tangent space at p. Interestingly, if (3.10) is viewed as a function from $R^3 \text{--}> R^3$ without restricting p to be on S, then (3.10) describes a family of surfaces characterized by (3.11). Any surface in this family is specified by the constant of integration chosen for (3.11). Furthermore, surfaces

of this family share an important virtue, namely, their shape information. This can be seen by examining the derivative of f(p), expressed below as,

$$dn = A \, dp \qquad \text{---- (3.12)}$$

where A is a matrix of partial derivatives of the components of n. Thus, intuitively, A with its second partial derivatives of the surface, contains the shape information. A formal definition of the shape fitting formula in the form of a low order polynomial fit for (3.10) helps to understand this point better. A first order Taylor's approximation of the normal at a point in a generic coordinate system (x_1, x_2, x_3) can be expressed as,

$$n(x_1, x_2, x_3) = n(x_1^0, x_2^0, x_3^0) + A \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ x_3 - x_3^0 \end{bmatrix}$$

where $A = [\delta n / \delta x_1, \delta n / \delta x_2, \delta n / \delta x_3]$.

Or equivalently,

$$n(x_1, x_2, x_3) - n(x_1^0, x_2^0, x_3^0) = A \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ x_3 - x_3^0 \end{bmatrix} \qquad \text{---(3.13)}$$

Now, writing the differential equation of (3.11) in a symmetric form, we have,

$$\frac{1}{2} [n(x_1, x_2, x_3) + n(x_1^0, x_2^0, x_3^0)] \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ x_3 - x_3^0 \end{bmatrix} = 0. \qquad \text{---- (3.14)}$$

Since A contains the shape information (the derivatives of the normal) A may be referred to as 'shape matrix '. In general, the linear fit is adequate for most surfaces found in nature, though extension to quadratic and higher order polynomials is straight forward. In order to understand how this formulation eliminates the disadvantages of explicit schemes, a discussion on the properties of the shape descriptor A is given below.

3.2 PROPERTIES OF THE SHAPE MATRIX 'A'

The major concerns in computing a local shape descriptor are (i) its robust behaviour at points near the occluding boundaries, (ii) the accuracy of shape fit in an arbitrary coordinate system (iii) the ease in recovering curvature information from the descriptor and (iv) its ability to deliver the partial derivatives of the surface that are integrable. We shall address every one of these issues in the following sections.

3.2.1 Robustness and Bounded Behaviour of the Shape Descriptor

One of the serious difficulties with the explicit formulation (3.9) is that the parameters in the right hand side of (3.9), increase without bound as we approach the occluding boundaries. Clearly, if the viewing point (z axis) is moved in such a direction that the occluding boundaries would come in to the center of the visual field, then the unbounded behaviour at those points can be checked. This means that the choice of the coordinate system decides the numerical behaviour of the parameters. In the case of the shape fitting formula in (3.13) the following theorem shows that the norm of A is independent of choice of

56

coordinates.

Theorem 3.1: The euclidean norm of A is invariant under euclidean change of coordinates.

Proof: To demonstrate that A does possess this property, consider euclidean motion imparted to the surface so that its coordinates are changed. Let the quantities in the new coordinates be denoted with $\wedge$, that is,

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = D + R \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad ----(3.15)$$

where $R^T R = I$ and D is a displacement vector, then we have

$$\hat{n}(\hat{p}) = R \hat{n}(p) \text{ and } \hat{n}(\hat{p}') = R \hat{n}(p'). \qquad ---- (3.16)$$

Here, p and p' are neighbors in the original coordinates. It may be noted that euclidean displacement of the surface does not alter the normals. The shape fitting formula (3.13) can now be written as,

$$\hat{n}(\hat{p}) = \hat{n}(\hat{p}') + \hat{A}(\hat{p} - \hat{p}')$$

and hence,

$$\hat{A} = R^T A R. \qquad ---- (3.17)$$

since, R is an orthogonal matrix,

$$|| \hat{A} || = || A ||. \qquad ---- (3.18)$$

Thus, the magnitude of A is independent of the coordinates. Since elements of A parameterize the surface, this result shows that the "size" of the parameters do not depend on the coordinates chosen. $\square$

3.2.2 Sensitivity in an Arbitrary Coordinate System

Just as the way the parameters of the explicit formulation (3.9) exhibit an undesirable behaviour at occluding boundaries, the accuracy of the shape fit at those points suffers too. This is due to the reason that accuracy of any polynomial fit for the surface depends on the magnitude of the partial derivatives of z with respect to x and y. Again, if the viewing point (hence the z-axis) can be moved to bring those points into center of visual field, then accuracy can be improved. Thus, the accuracy depends on the choice of the coordinates used to describe the surface. On the contrary, as the following theorem shows, the accuracy of the shape fit using (3.13) is intrinsic to the surface and does not depend on the choice of the coordinates.

Theorem 3.2: Suppose that the exact formula for the normal vector is given by,

$$n(x_1, x_2, x_3) = n(x_1^0, x_2^0, x_3^0) + A \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ x_3 - x_3^0 \end{bmatrix} + E(x_1, x_2, x_3) \quad ----(3.19)$$

where, E is the error in the fit. Then the euclidean norm of E is invariant under a euclidean change of coordinates.

Proof: After a change in coordinates as euclidean motion is applied by (4.16), the error $\hat{E}$ is given by

$$\hat{E} (\hat{x}_1, \hat{x}_2, \hat{x}_3) = R \ E(x_1, x_2, x_3)$$

Since R is an orthogonal matrix $|| \hat{E} || = || E ||$. □

That is, the error is independent of the coordinates. Thus, instead of obtaining a numerically ill conditioned function z(x,y) and

58

then compute the curvatures by numerically differentiating this function, the shape fitting approach by locally fitting the shape descriptor A, followed by integrating the resulting quantity to compute $z(x,y)$ is insensitive to the choice of coordinates.

3.2.3. Recovering Principal Curvatures

Since principal curvatures prove to be a convenient way to represent local shape, an important requirement of the shape descriptor is to provide an easy means to obtain the curvature map of the surface. In order to establish the result that curvatures are easily computable from the shape matrix A we will first prove the following result, regarding the invariance of the tangential manifold under A. Later, we will show that the eigen values of A restricted to the tangent space represent the principal curvatures.

Theorem 3.3: The tangent space $TM = \{ n \}^{\perp}$ is invariant under A. Equivalently, n is a left eigen vector of A.

Proof: Since $n^T n = 1$, we have $n^T \, dn = 0$. That is, $dn \in TM$. Using (3.12) it can be seen that for any $dp \in TM$, $dn = A \, dp \in TM$. Hence, A maps a vector in TM into a vector in TM.

Since $n^T A \, dp = 0$ whenever $n^T dp = 0$, we have

$$n^T A = \lambda \, n^T \qquad ----- \quad (3.20)$$

That is, n is a left eigen vector of A. □

Since TM is invariant under A, it makes sense to talk of the restriction of A to TM. i.e how A maps vectors in TM. An important property of the restriction of A to the tangent space TM is stated in the next theorem.

Theorem 3.4: The restriction of A to TM is a self-adjoint operator

(i.e the restriction of A to TM is a symmetric matrix). Furthermore, the eigen values of the restriction of A to the TM are the principal curvatures and these are independent of the coordinates chosen.

Proof: Note that the restriction of A to M is dn, the differential of n. The two properties of dn can be found in [Doc 76]. □

Instead of the principal curvatures, the Gaussian (K) and the mean (H) curvatures are preferred since principal curvatures are directional quantities besides being noise sensitive. Furthermore, it can be shown that these values uniquely characterize a given surface [Doc 76, Theorem of Bonnet].

3.2.4. Integrability of Normals

Integrability of the surface normal estimates becomes an issue especially in the shape fitting approach, since eventually the relative depth values need to be obtained by integrating the normals. The theorem 3.4 provides the necessary and sufficient condition for the integrability of (3.11) because the restriction of A to the tangent space is symmetric. However, the theorem is not easy to use since it pertains only to the restriction of A to TM. The surface is completely characterized by the action of A on vectors in TM and its action on vectors outside TM is irrelevant as far as the description of the surface is concerned. Let

$$\overline{A} = A (I - v\,n^T) \qquad\qquad ----- \quad (3.21)$$

where, v is arbitrary. Then for any dp ∈ TM i.e. dp ⊥ n, A p = $\overline{A}$ p. Thus, there are several shape matrices "A" (that only differ on how they behave outside TM) that describe the surface. The following lemma shows that we can restrict the choice of "A" matrices by specifying

how it behaves outside TM.

Lemma 3.1. For a given surface, we can always find a shape matrix A such that n is both a left and right eigen vector.

Proof: In (3.21) let v be chosen as $v = A \, n - \sigma \, n$ where σ is arbitrary. Then

$$\bar{A} \, n = A \, n - A \, n \, n^T n + \sigma \, n \, n^T n = \sigma \, n, \text{ since } n^T n = 1. \; \square$$

Thus if we restrict our attention to only those A's for which n is also a right eigen vector, the condition for integrability is much simpler than the one specified by theorem 4.4.

Theorem 3.5: Suppose that n is both a left and right eigen vector of A. Then a necessary and sufficient condition for (3.11) to be integrable (i.e. cross partial derivatives be equal) is that A be symmetric.

Proof:

Necessity: From theorem 3.4, A has two orthogonal eigen vectors on TM. Since n is also a right eigen vector of A, A has three orthogonal eigen vectors. Hence A is symmetric.

Sufficiency: Suppose that A is symmetric. We have

$$\delta z/\delta x = z_x = -n_1/n_3 \quad \text{and} \quad \delta z/\delta x = z_y = -n_2/n_3$$

Thus,

$$n_3^2 \, \delta^2 z /\delta x \, \delta y = n_2 \, \delta n_3/\delta x - n_3 \, \delta n_2/\delta x \; \text{----} \; (3.22)$$

By using

$$\delta n_3/\delta x = A_{31} - A_{33} \, n_1/n_3$$

$$\delta n_2/\delta x = A_{21} - A_{23} \, n_1/n_3$$

in (3.22) we get

$$n_3^2 \, \delta^2 z /\delta x \, \delta y = - A_{33} \, n_1 n_2/n_3 + n_2 A_{31} + n_1 A_{23} - n_3 A_{21} \; \text{---} \; (3.23)$$

Similarly,

$$n_3^2 \, \delta^2 z \, /\delta y \, \delta x = - A_{33} \, n_1 n_2/n_3 + n_2 A_{13} + n_1 A_{32} - n_3 A_{12} \cdots \quad (3.24)$$

From the symmetry of A it is clear that the two cross partial derivatives are equal. □

It is interesting to note that the proof of sufficiency does not use the fact that n is an eigen vector of A. However, this can easily be established using the result of theorem 3.3 and the fact that for a symmetric matrix any right eigen vector is also a left eigen vector. The importance of the theorem is that this condition is also necessary if we restrict our shape matrices.

It may be noted that A changes continuously for an arbitrary surface, although it remains unchanged for some surfaces such as spheres and cylinders. Suppose that we know A at a point p. Now to compute A its neighbor $\bar{p}$, the projection of A to the tangent manifold at the point $\bar{p}$ is considered. Hence if $\bar{n}$ is the unit normal at the point $\bar{p}$, then the projection of A to the tangent space at $\bar{p}$ is given by,

$$\bar{A}_{proj} = (I - \bar{n} \, \bar{n}^T) \, A \, (I - \bar{n} \, \bar{n}^T) \quad \text{----} \quad (3.25)$$

Using the continuity of the eigen values and eigen vectors of a symmetric matrix to the parameters of the matrix, it is easy to show that the restriction of $\bar{A}$ to the tangent space at $\bar{p}$ is approximately equal to $\bar{A}_{proj}$ to first order terms in (p - $\bar{p}$). Thus, given A matrix at a point and the normals at nearby points, we can compute the shape information at the nearby points by using (3.25).

The invariance of the magnitude of the parameters (A) and accuracy of the fit to change of coordinates is an extremely important

property. For most smooth surfaces occurring naturally and for machined parts, the scene can be viewed as consisting of surfaces with slowly varying shape. In this case, the shape interpolation will have considerably smaller error (truncation error) than any polynomial surface fit. Thus, a segmentation algorithm can be developed using the shape fitting formulation and searching for rapid changes in the shape matrix A. An application of these ideas in CAD based object recognition can be found in [NaJ 88].

4. Summary

To recapitulate, the most convenient form of invariant description of shape characteristics is a map of Gaussian and Mean curvatures. The polynomial shape fitting formulation introduced here provides an easy means to compute these quantities. In short, our approach exhibits the following advantages: (i) The magnitude (in a well defined norm sense) of the parameters that are used in the polynomial fit is independent of the choice of coordinates. (ii) The accuracy of the fit is also independent of the choice of the coordinates. (iii) The shape information (Gaussian curvatures / principal curvatures) is obtained from the eigen values of a symmetric matrix, a numerically robust problem. (iv) The depth map $z(x,y)$ is obtained by integrating a differential equation, thus providing a measure of robustness.

Chapter 4

Computing Shape from Shading

1. Introduction

Human vision is extremely adept at inferring shape from monocular images alone. The occluding contours with smooth shading that the object renders the viewer can serve as a strong clue for inferring their shape. It appears that when the scene is rich in edge details, the human vision system appears to use stereopsis to infer depth at edge locations and then reconstruct the object in three dimension with the aid of shape computing processes. On the other hand, when the visual instance (e.g. viewing a photograph) is devoid of binocular disparity, the system has to rely totally on monocular shape computing processes. Also, binocular vision is unreliable in situations where the object is too far away. Thus, every visual cue appears to contain rich shape and depth information only for a certain visual geometry. In this sense, shading provide shape information for smooth objects with uniform reflective surfaces.

This chapter presents a detailed treatment of the shape from shading problem and a new technique to estimate shape using shape fit methods. The following remarks provide the motivation to investigate the shape from shading problem.

First, current techniques reported in the literature are concerned with estimating the surface normals alone. There is no specific attempt to describe (quantitatively or qualitatively) the nature of the surface at any local point. Presumably, shape inference is a secondary process involving surface fitting strategies.

Second, local shading analysis is quite error prone and noise sensitive in general and hence a global model should be our pursuit.

Third, global models proposed in the literature suffer from one or more of the following difficulties:

(a) Not being able to deliver an integrable surface.

(b) Being Computationally quite expensive.

(c) Treat shape inference process as a secondary one.

Fourth, since shape information is much more important than depth alone for recognition, a method that delivers local shape information becomes necessary.

With the ultimate goal to reconstruct the viewed surface, the problem of computing surface normals is posed here with an implicit emphasis on recovering the relative depth and local shape (curvature) information. The major thrust of this approach is that:

The surface derivatives are recovered first followed by integration of the derivatives to obtain the relative depth.

Based on this notion, a global objective function is derived here in terms of the surface normals, the local shape descriptor described in Chapter 3 and the intensity values. Iterative procedures are suggested to minimize the objective function resulting in fast estimation of shape and relative depth.

This chapter is organized as follows. The imaging geometry and the physical constraints of the shading process are described in Section 2. A precise statement of the problem is provided in Section 3. Section 4 describes the proposed approach using the polynomial shape fitting formulation introduced in Chapter 3. The results of

implementation are reviewed in Section 5. We conclude the chapter with some important remarks in Section 6.

2. Imaging Geometry

The imaging process is complicated involving multiple geometrical phenomena. The intensity at a point on the image plane can be represented by the following expression, with three components:

$$I(x,y) = I_{sp} + I_{re} + I_b \qquad \text{----- (4.1)}$$

where, I_{sp} is the specular component, I_{re} is the reflectance component and I_b is the background intensity component. The primary shading information is contained in the reflectance component, I_{re}. The specular component is responsible for the shiny appearance for glassy objects while the background intensity component illuminates the object from multiple reflections of the neighboring objects. In studying the shape information contained in the basic intensity values, the contribution due to every one of the three components of (4.1) needs to be separated. Since such a separation is not possible, we constrain ourselves with imaging situations where there are no (i) specularities, (ii) multiple reflections from neighboring objects and (iii) the surface is Lambertian. These constraints reduce the image generation due to I_{re}, the reflectance component alone.

For a smooth Lambertian surface (at least twice continuously differentiable), the intensity value at a point is not only a function of the local surface orientation, but also viewing angle, illuminant angle and the surface reflectivity. Thus, for a scene with a single light source,

$$I(x,y) = G(z_x, z_y, \iota, \upsilon, \rho) \text{------} \quad (4.2)$$

where z_x and z_y are the first partial derivatives of the surface, ι is the light source direction, υ is the viewer direction and ρ is the surface reflectivity constant. The reflectance function G can be further simplified to depend only on the local surface orientation if an orthogonal projection, as shown in Figure 4.1, can be assumed.

Under the conditions stated above, an expression for the intensity has been derived as [Hor 77],

$$I(x,y) = \rho(\iota \cdot n) \qquad\qquad ---- \quad (4.3)$$

where, $I(x,y)$ is the intensity at a point, (x,y) in the image plane (Figure 4.1)

$n = [n_1, n_2, n_3]^T$ is the unit normal vector at (x, y, z).

$\iota = [\iota_1, \iota_2, \iota_3]^T$ is the unit light source vector.

Also, the first partial derivatives of the surface can be shown to be,

$$z_x = \delta z/\delta x = -n_1/n_3 . \quad z_y = \delta z/\delta y = -n_2/n_3 . \quad ------(4.4)$$

If the surface is described as $z = f(x,y)$ then the unnormalized normal N is given by

$$N = [z_x, z_y, -1]^T \quad ------ \quad (4.5)$$

The normalized vector is then obtained as

$$n = N / [N^T N]^{1/2} \quad ------ \quad (4.6)$$

A precise statement of the problem based on the above description of the imaging geometry helps to conceive the real issues. To avoid any confusion with the x,y's in three dimension, the image intensities on the two dimensional grid will be denoted by $g(i,j)$ instead of $I(x,y)$.

3. Problem statement

The problem of surface reconstruction and shape determination can be stated as :
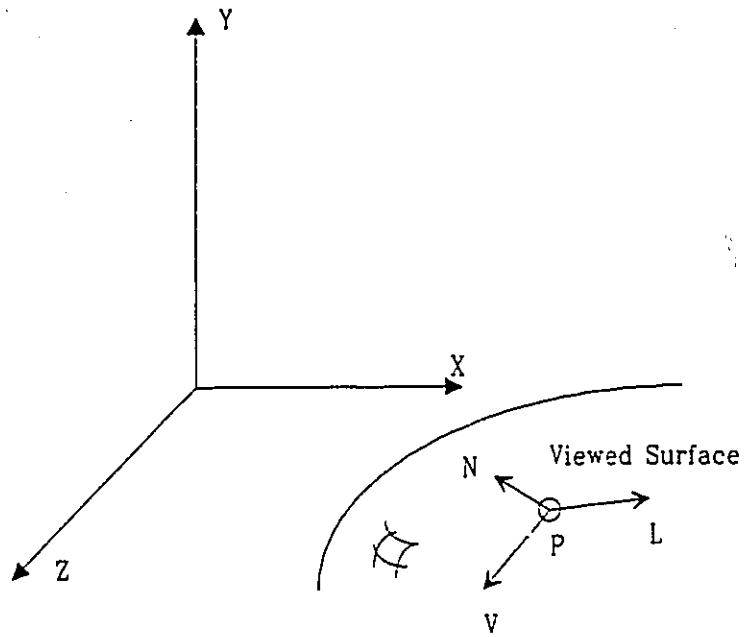
L   The Light Source

N   Normal at P

V   Viewer Direction Parallel to Z–Axis

Figure 4.1 Imaging Geometry for Shading

Given the intensity values, g(i,j) of a surface on the image plane, compute the following quantities.

(i) the normal n for every point in the image plane,

(ii) z(i,j) the image plane distance along the viewer direction (relative depth),

(iii) a shape descriptor that would locally characterize the surface.

Tne solution to this problem is derived from (4.3) under the following assumptions:

(i) the surface is Lambertian

(ii) the illumination is a point source

(iii) the surface is smooth (at least twice continuously differentiable)

(iv) the location of boundaries are known a priori.

While the first two assumptions simplify the model of the imaging process, the last two assumptions are used in formulating the problem in terms of a shape descriptor discussed in the following section.

4. Shape Extraction from Shading via Shape Fitting

A mathematical formulation of the problem of computing shape from shading information, utilizing the shape fitting concepts discussed in Chapter 3, is presented in this section. This formulation also provides a depth map of the scene captured in the image plane. We begin with the polynomial shape fit given by equation (3.13) in Chapter 3. That is, the normals at the generic point (x,y,z) can be related to the normal at the neighbor (x^0, y^0, z^0) as,

$$n(x,y,z) = n(x^0, y^0, z^0) + A \begin{bmatrix} x - x^0 \\ y - y^0 \\ z - z^0 \end{bmatrix} \quad \text{----}(4.7)$$

where $A = [\delta n/\delta x, \ \delta n/\delta y, \ \delta n/\delta z]$, a matrix with columns of partial derivatives of the normal. The crux of our approach is to relate the normals modeled by (4.7) to the intensity measurements in such a way that the resulting objective function can be directly minimized to give normals, the local shape descriptor A and the depth z.

4.1. FORMULATION OF THE MINIMIZATION PROBLEM

In order to derive an objective function to solve for shape, surface normals and depth, the following constraints are imposed here.

(i) The surface normals in the neighborhood of a local point (i,j) should be smooth and modeled by equation (4.7)

(ii) The surface normals should satisfy the following equation.

$$n^T \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = 0. \text{------}(4.8)$$

(iii) The norm of the normals i.e $n^T n = 1$.

(iv) The computed intensity

$$\rho \, \iota^T n(i,j) = g\,(i,\,j)$$

Without any loss of generality, it may be assumed that

$0 <= g\,(i,j) <= 1$ and $\rho = 1$.

All the above requirements can be satisfied in the following formulation.

70

Minimize $J = \sum [J_1(i,j) + J_2(i,j) + J_3(i,j)]$ ---- (4.9)

where,

$$J_1 = (1/N_{i,j}) \sum_{\substack{k,l \\ \in \eta(i,j)}} || n(i,j) - n(k,l) - A \Delta ||^2$$

$$J_2 = (1/N_{i,j}) \beta \sum_{\substack{k,l \\ \in \eta(i,j)}} || ((n(i,j) + n(k,l))/2)^T \Delta ||^2$$

$$J_3 = \gamma (1^T n(i,j) - g(i,j))^2$$

where, β and γ are appropriately chosen weights,

$$\Delta = \begin{bmatrix} i - k \\ j - l \\ z(i,j) - z(k,l) \end{bmatrix}$$

$\eta(i,j)$ is the neighborhood of (i,j)

and $N_{i,j}$ is the number of elements in $\eta(i,j)$.

The optimum estimates of n, z, and A are obtained by minimizing J subject to the unit normal constraint

$$|| n(i,j) ||^2 = 1.$$

In this formulation, J_1 is the smoothness measure for the normals, J_2 ensures that $n^T\Delta$ is small, and J_3 represents the measurement error.

A few direct observations about this formulation are in order here. First, our formulation yields depth information as a direct by-product of the algorithm, while works of Brooks and Horn [BrH 85], Frankot and Chellappa [FrC 87] focus on estimating the normals or the partial derivatives $\delta z/\delta x$ and $\delta z/\delta y$ first and then the depth

information is obtained through integration after ensuring that integrability conditions are satisfied. Second, the requirement that A is symmetric guarantees integrability of $\delta z/\delta x$ and $\delta z/\delta y$ (Theorem 3.5, chapter 3) and integrability is thus implicit in our formulation. Third, by forcing the normals in a neighborhood to satisfy (4.7), the shape matrix A is incorporated into the formulation and thus, the local shape information is also computed along with normals and depth. All these advantages are in addition to the theoretical significance of shape fitting discussed earlier in chapter 3.

4.2. SOLUTION METHODOLOGY

A straight forward approach to minimize (4.9) would be to solve for n, A and z simultaneously. However, for computational efficiency, the problem is divided into two subproblems here.

Problem 1 (P_1) : Given n, z at every point, find A in the least squares sense that fits the interpolating equation (4.9), i.e., minimize $\sum_{i,j} J_1$ with respect to A.(a standard least squares fit).

Problem 2 (P_2) : Given A, find n and z in the least squares that minimizes $\sum_{i,j} J$.

The second phase utilizes the intensity measurement g(i,j) to obtain the estimates of n and z. Although this is a nonlinear least squares fit, an efficient algorithm can be derived by exploiting the sparse structure implicit in the neighborhood dependence.

Because of the equality constraint $\| n(i,j) \|^2 = 1$, the objective function is nonlinear in n but for a fixed n, it is

quadratic in z. This fact can be taken into advantage by partitioning P_2 as,

P_{21} : Given $n(i,j)$ find $z(i,j)$ that minimizes J. P_{22} : Given $z(i,j)$ find $n(i,j)$ that minimizes J.

Thus, the overall structure is as follows:0) Start with initial estimates of n, z and A

1)Repeat

        Repeat

                Solve P_{21}

                Solve P_{22}

        Until convergence

        Solve P_1

Until convergence.

The Problem P_{21} is solved by block Gauss-Seidel approach (for a parallel implementation Gauss iteration would be more appropriate) where two adjacent rows are obtained in one step. The block Gauss-Seidel scheme is seen to be the best way to exploit the sparse nature of the system. An extensive treatment on this scheme may be found in Tarjan [Tar 76]. P_{22} can be solved in a similar fashion by using the Newton - Raphson algorithm coupled with the block Gauss - Seidel scheme. In the test cases run here, it was found out that the standard Gauss - Seidel approach itself was found to be adequate. P_{11} is a linear squares problem that can be solved by standard techniques. We discuss below in detail how the solutions for the three problems P_{11}, P_{21} and P_{22} can be solved.

4.2.1 - Solution for P_{11}- Estimation of Shape Matrix A

73

The algorithm for estimating A in a local region centered at (i,j) assumes that the surface normals n and depth z are known a priori. An examination of the objective function J defined in (4.9) reveals that only J_1 explicitly depends on the shape matrix A. Although A is uniquely defined at each point (i,j) the approach taken here assumes that A is piecewise constant i.e. the shape fitting formula (4.7) is applicable to a region in the surface. The image plane is divided into non-overlapping segments and A is assumed to be constant in each of the regions. Then for each region, J is quadratic in the elements of A. By direct differentiation and imposing the requirement of symmetry, A_s can be obtained as the solution to the Lyapunov equation:

$$A_s N + N A_s = M + M^T \qquad ----- \quad (4.10)$$

where,

$$N = \sum_{\substack{i,j \\ \in \eta_{i,j}}} \sum_{k,l} \Delta \Delta^T \qquad ----- \quad (4.11)$$

$$M = \sum_{\substack{i,j \\ \in \eta_{i,j}}} \sum_{k,l} [n(i,j) - n(k,l)] \Delta^T \quad ----- \quad (4.12)$$

where, (i,j) is over the region under consideration.

4.2.2. Solution for P_{21}- Estimation of z For the Given n

To compute the depth z at every point given the surface normals, it may be observed that J in (4.9) is quadratic in z. Differentiating J with respect to z(i,j) and setting it to zero we obtain

$$\sum_{k,l \in \eta_{i,j}} [A_3^T (A \Delta - n(i,j) + n(k,l)) +$$

74

$$\beta \{(n_3(i,j) + n_3(k,l))\} \{n(i,j) + n(k,l)\} \Delta^T = 0 \;\text{---}\; (4.14)$$

It may be noticed that (4.14) is a sparse system of linear equations. The equation for the point (i,j) involves z(i,j) and z(k,l) where (k,l) belongs to the neighborhood $\eta_{i,j}$. This system can be solved for z by the Gauss-Seidel technique. The sparse nature of this system has been exploited in our implementation by adopting the block Gauss-Seidel method suggested by Tarjan [Tar 76].

4.2.3. Solution for P_{22} - Estimation of n

In order to enforce the requirement that the normals be of unit length, Lagrange multipliers $\lambda(i,j)$ are introduced here and the objective function is augmented as

$$\hat{J} = J \sum_{i,j} \lambda(i,j) \{1 - n(i,j)^T n(i,j)\}^2 \qquad \text{----} \; (4.15)$$

Differentiating with respect to n(i,j) and setting it to zero we obtain

$$\sum_{k,l \in \eta_{i,j}} [n(i,j) - n(k,l) - A\Delta] + \beta \Delta \Delta^T n(i,j) + \gamma(11^T - g(i,j)1)$$
$$= \lambda(i,j) \; n(i,j) \quad \text{----} (4.16)$$

The above equation can be further simplified to

$$n(i,j) + \frac{\gamma}{|\eta_{i,j}|} 1\, 1^T n(i,j) + \sum_{k,l \in \eta_{i,j}} \beta \Delta \Delta^T n(i,j)$$

$$= n_{av}(i,j) + \frac{\gamma}{|\eta_{i,j}|} g(i,j) 1 + \lambda(i,j) n(i,j) \quad \text{----} (4.17)$$

where $n_{av}(i,j)$ is given by

$$n_{av}(i,j) = \frac{1}{|\eta_{i,j}|} \sum_{k,l \in \eta_{i,j}} [n(k,l) + A \Delta \qquad \text{----(4.18)}$$

where $|\eta_{i,j}|$ is the number of elements in $\eta_{i,j}$. It is interesting to interpret (4.17). From (4.7) $n(i,j) \approx n_{av}(i,j)$ to first order terms, and $1^T n(i,j) \approx g(i,j)$ when a good reconstruction of the image is available, and $\Delta^T n(i,j) \approx 0$ to first order terms. Hence from (4.17), $\lambda \approx 0$ to first order terms. This fact can be exploited in solving (4.17), as follows.

Let

$$C = I + \frac{\gamma}{|\eta_{i,j}|} 11^T + \sum_{k,l \in \eta_{i,j}} \beta \Delta \Delta^T \qquad \text{----(4.19)}$$

$$n_0(i,j) = G^{-1} \left[n_{av}(i,j) + \frac{\gamma}{|\eta_{i,j}|} g(i,j)1 \right] \qquad \text{----(4.20)}$$

Then,

$$n(i,j) = n_0(i,j) + \lambda\, G^{-1} n(i,j) \qquad \text{---- (4.21)}$$

It is easy to show that $||G^{-1}|| \leq 1$. Also, (4.21) can be expanded as, (i,j, have been omitted for simplicity)

$$n = n_0 + \lambda\, G^{-1}(n_0 + \lambda\, G^{-1}n)$$
$$= n_0 + \lambda\, G^{-1}n_0 + \lambda^2 G^{-2} (n_0 + \lambda\, G^{-1}n)$$
$$= n_0 + \lambda\, G^{-1}n_0 + \lambda^2 G^{-2} n_0 + \ldots \qquad \text{---- (4.22)}$$

Hence,

$$1 = n^T n = n_0^T n_0 + 2\,\lambda\, n_0^T G^{-1}n_0 + \ldots \qquad \text{---- (4.23)}$$

Neglecting λ^2 and higher order terms, it can be shown that

$$\lambda(i,j) = \frac{1 - n_0^T(i,j)\, n_0(i,j)}{2\, n_0^T(i,j)\, G^{-1}\, n_0(i,j)} \qquad ----(4.24)$$

and

$$n(i,j) = n_0(i,j) + \frac{1 - n_0^T(i,j)\, n_0(i,j)}{2\, n_0^T(i,j)\, G^{-1}\, n_0(i,j)}\, G^{-1} n_0(i,j) \qquad ----(4.25)$$

Since G is a 3 X 3 matrix, (4.25) does not impose any major computational burden.

5. Results and Discussion

The shape computing algorithm, described in the previous sections, was tested on synthetic and real world images. The algorithm has also been tested for its noise tolerance by experimenting with synthetic imagery. In all these cases, the initial set of z, n and A was generated to be a "consistent" set in that appropriate z and A are fitted for a given set of normals.

5.1 Studies with Synthetic Images

In generating synthetic images, Lambertian surfaces, illuminated by a single source, were assumed. Thus, the irradiance equation (4.3) can be used to generate these synthetic images. The following considerations were given to select the type of surfaces generated here.

(i) Surfaces with constant curvatures would help us to find how sensitive the estimated shape information is.

(ii) Surfaces that have both positive as well as negative

77

curvatures that are quickly varying, should help us find how faithfully the shape interpolation algorithm follows undulating slopes.

(iii) Surfaces that undergo self occlusion due to oblique illumination should suggest us how well the algorithm perform at points that reach zero image intensity values very quickly.

To test the algorithm on surfaces with slow varying or constant curvatures, a synthetic sphere was generated with the light source on the z-axis. Figures 4.2 and 4.3 describe the intensity image of the sphere, and its reconstructed shape respectively. Also, tested here was the performance of Brooks and Horn's algorithm (see Figure 4.4). Although visually, the reconstruction appears to be similar in both the cases, at points (near the occluding boundary) where the surface fastly recedes away from the viewer our algorithm performs quite well as compared to the method of Brooks and Horn. The mean square error in normals at these points for the same number of iterations (30 iterations) is 0.00047 in our case and 0.0014 in their case. To test the algorithm in the cases where the intensity values reach zero quickly, (but the points of the surface need not recede from the viewer), the same sphere illuminated from the side (source on the x - axis so that intensity values reach zero rapidly near the ZOY plane) was considered here. Figures 4.5 through 4.6 demonstrate the illuminated part of the sphere and its reconstructed shape in the illuminated region. Finally, for the cases where positive and negative curvatures co-exist, such that the transition from positive to negative curvatures is also quite rapid, we considered an undulating
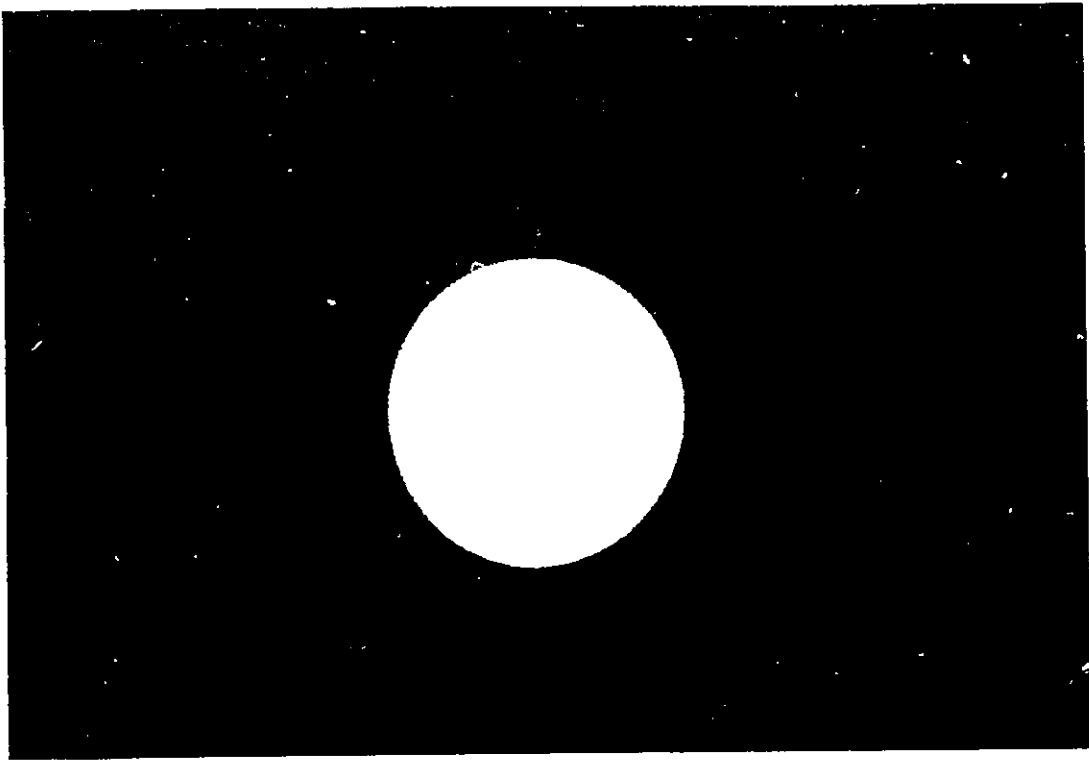
78

Figure 4.2 A synthetic sphere with light source on z axis
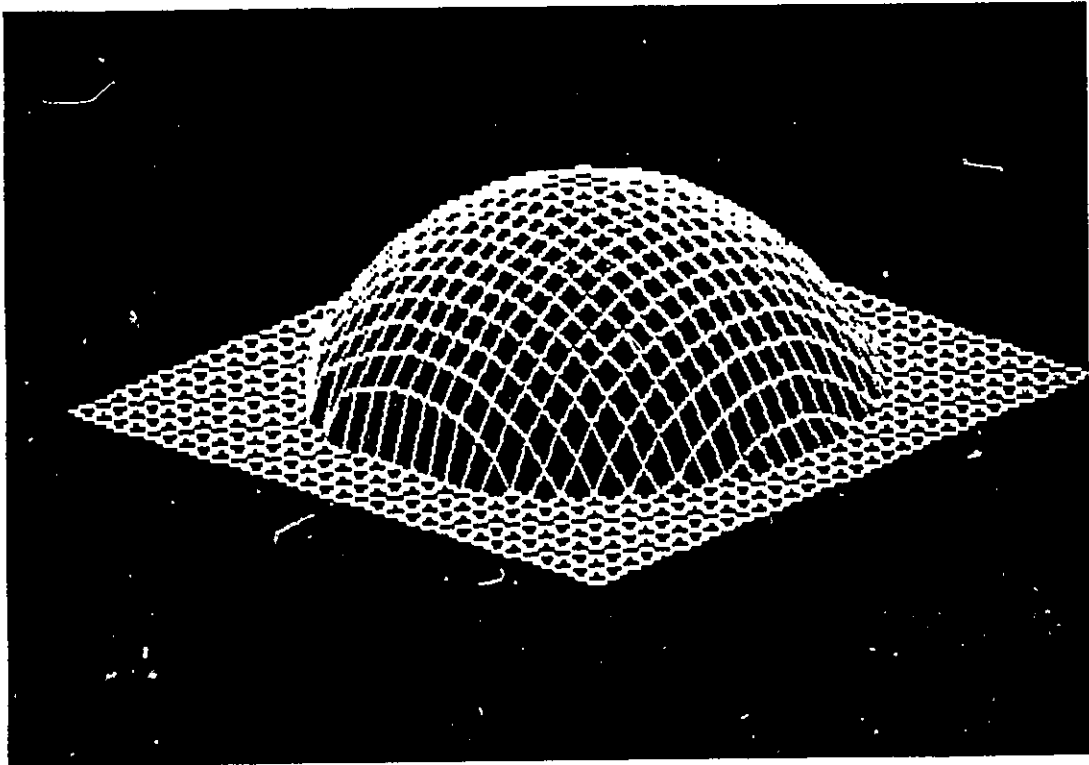

Figure 4.3 Reconstructed surface using the shape-fit approach
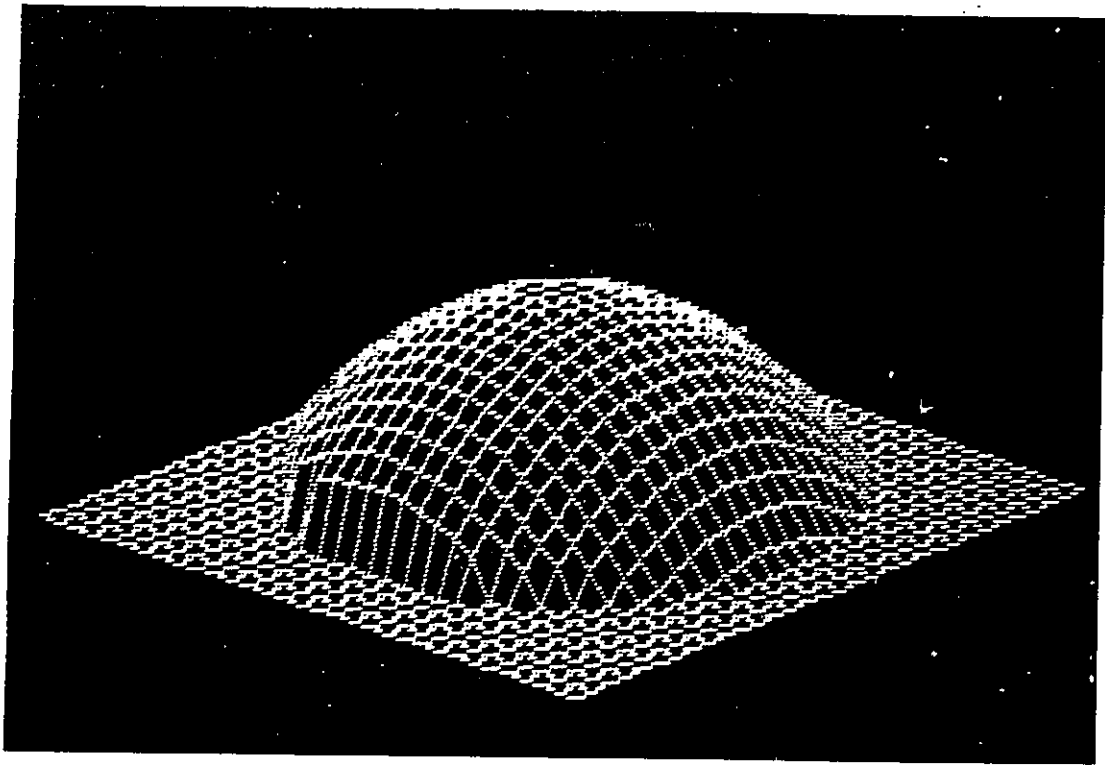
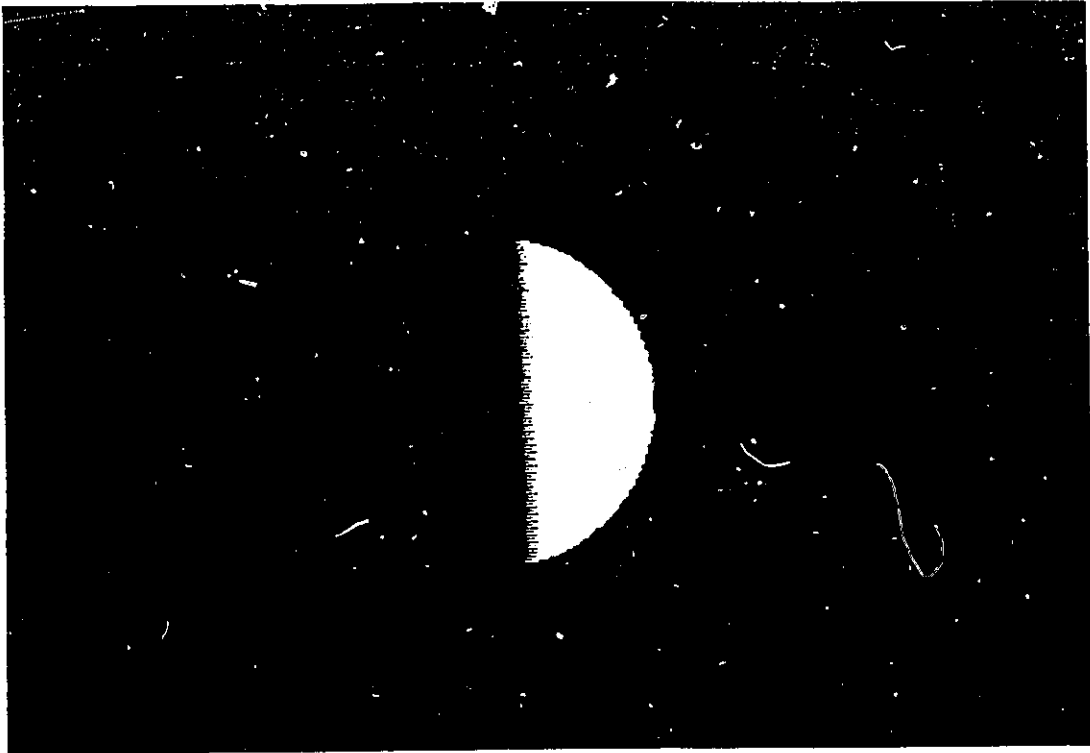Figure 4.4 Reconstructed sphere using Brook's and Horn's Method

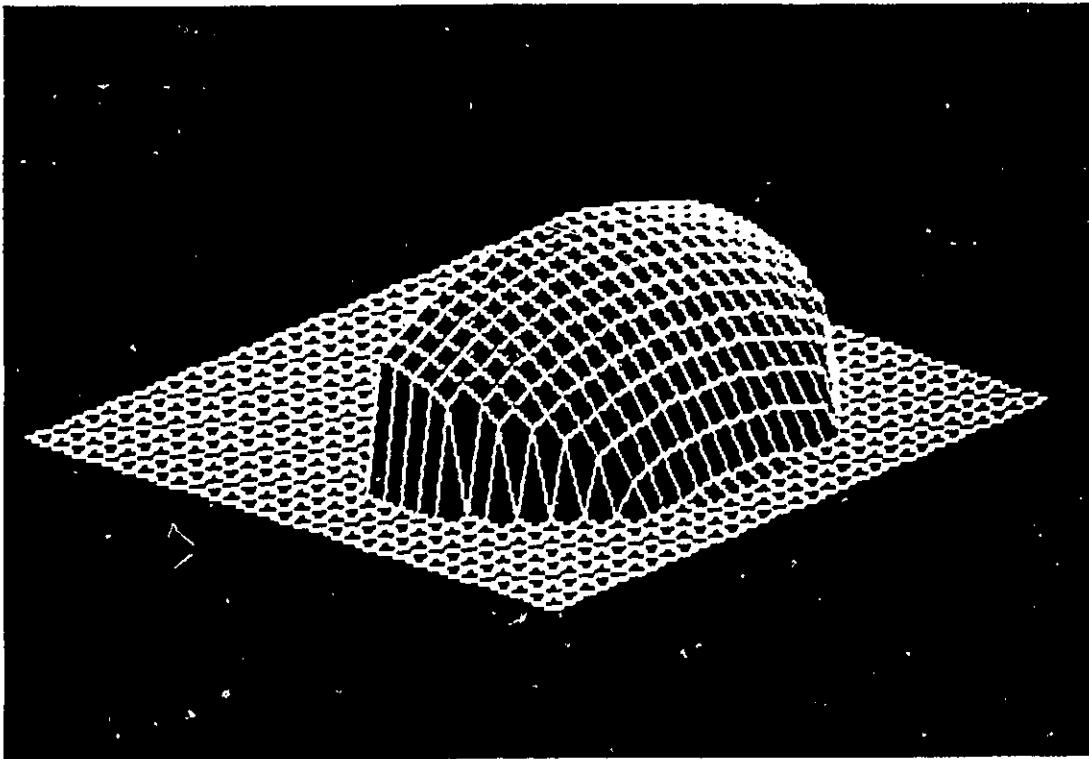Figure 4.5 A synthetic sphere with light source on x axis

Figure 4.6 Reconstructed surface on the illuminated part

81

surface that is circularly symmetric and illuminated from an oblique angle (see Figures 4.7 and 4.8 for its intensity image and its true shape). The ability of the algorithm to preserve the trend of curvatures is shown in Figure 4.9. Noticeably, the reconstructed shape tends to smoothen the areas where the curvature varies very rapidly. However, it should be observed that the reconstruction preserves the undulation primarily due to the inclusion of shape information through shape matrix A. The resulting solution when the shape fit was not included is shown in Figure 4.10. Brooks and Horn's algorithm converged to a horse saddle type of surface in this case (Figure 4.11). Interestingly, when the light source is moved to the z axis (see Figure 4.12), several surfaces are possible candidates that can generate the same image. Running our algorithm in this case, the result was one such possible solution that is convex everywhere (Figure 4.13).

To show the shape information the algorithm delivers, a sphere of radius 30 was inscribed in an image of 32 X 32 and the algorithm was tested in updating A for every 4 X 4 region. The principal curvatures are provided in Table 4.1. It may be noted that except at the border regions, the average curvatures are found to be around 0.03 while the true curvatures are 0.033.

To show the fast convergence, the decrease in average mean square error between true normals and computed normals at every iteration for the sphere is shown in Figure 4.14. In the same graph the convergence for the Brooks and Horn's method is depicted by the dotted curve. Consistently, in all the cases of slow varying surfaces, convergence
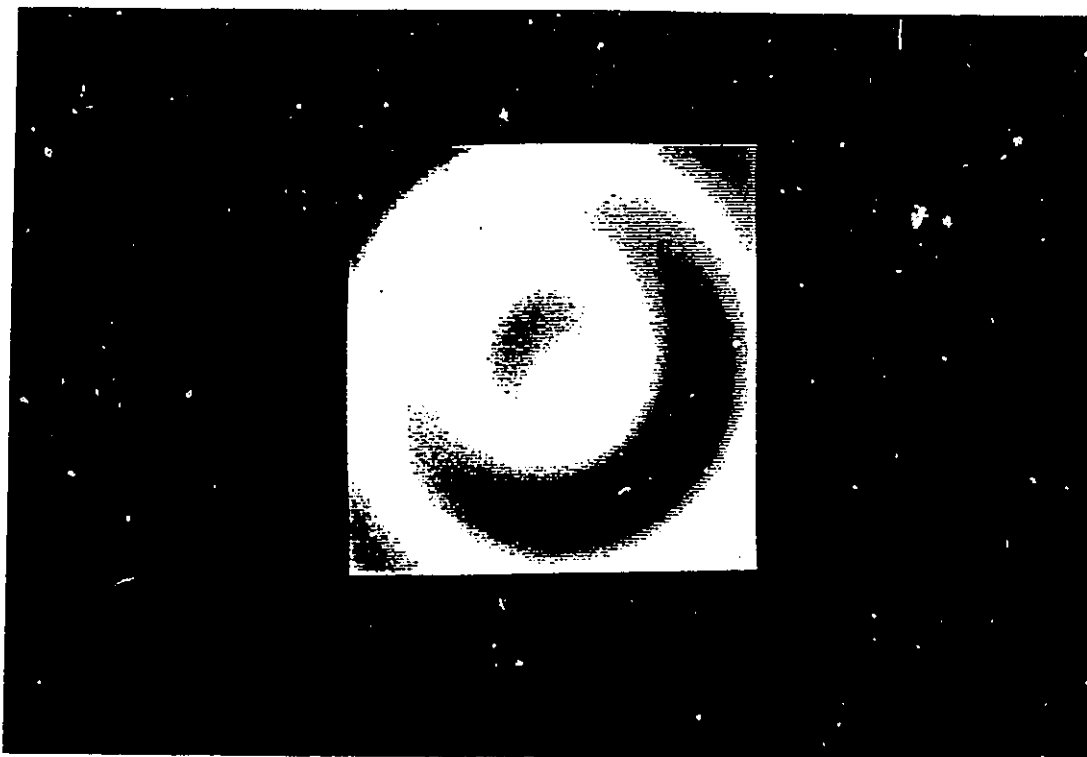
Figure 4.7 A synthetic undulating surface with oblique illumination
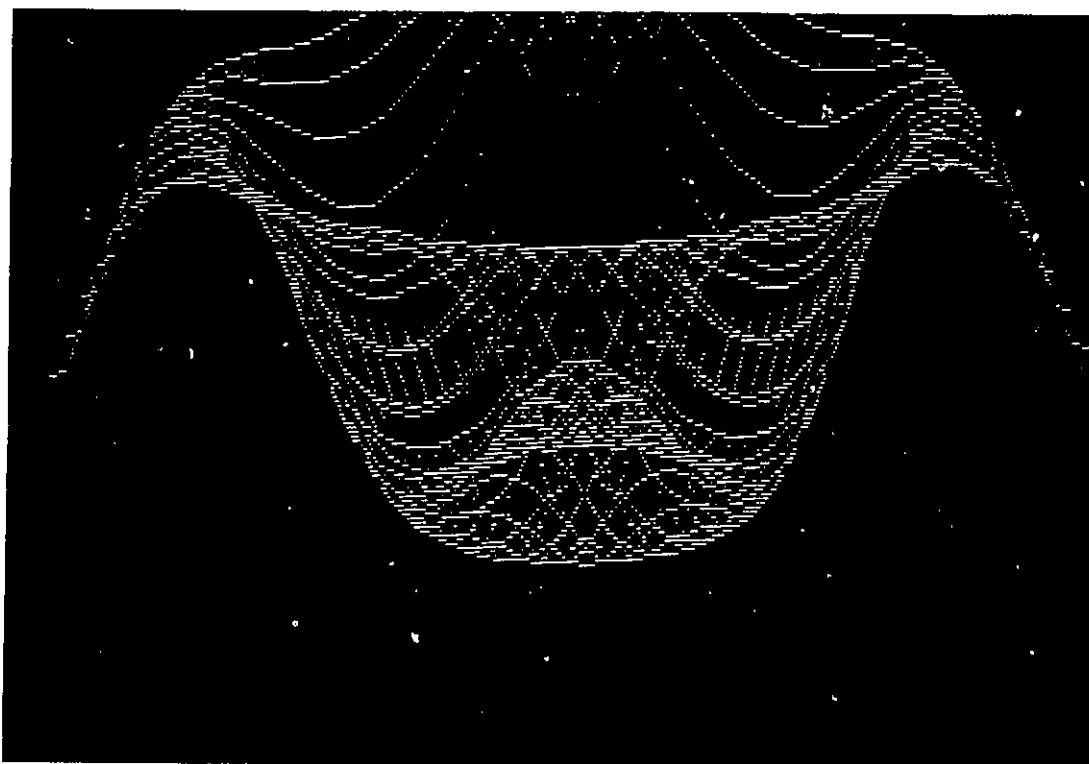
Figure 4.8 The true shape of the undulating surface

83
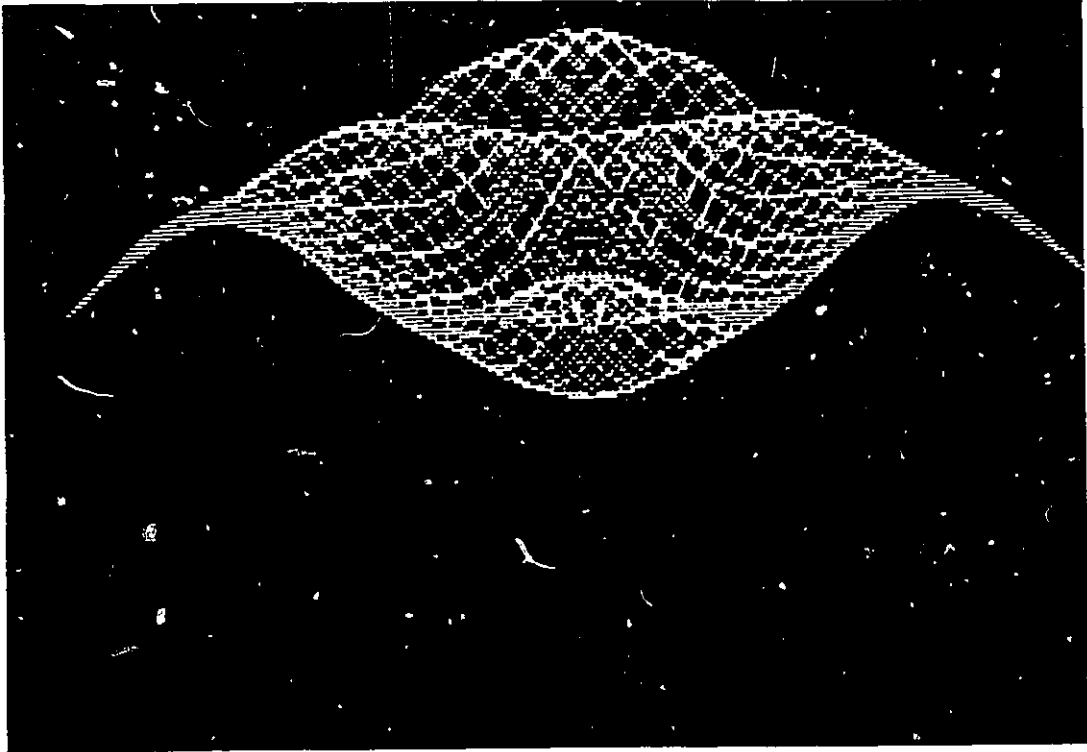
Figure 4.9 Reconstructed surface with shape fit formulation
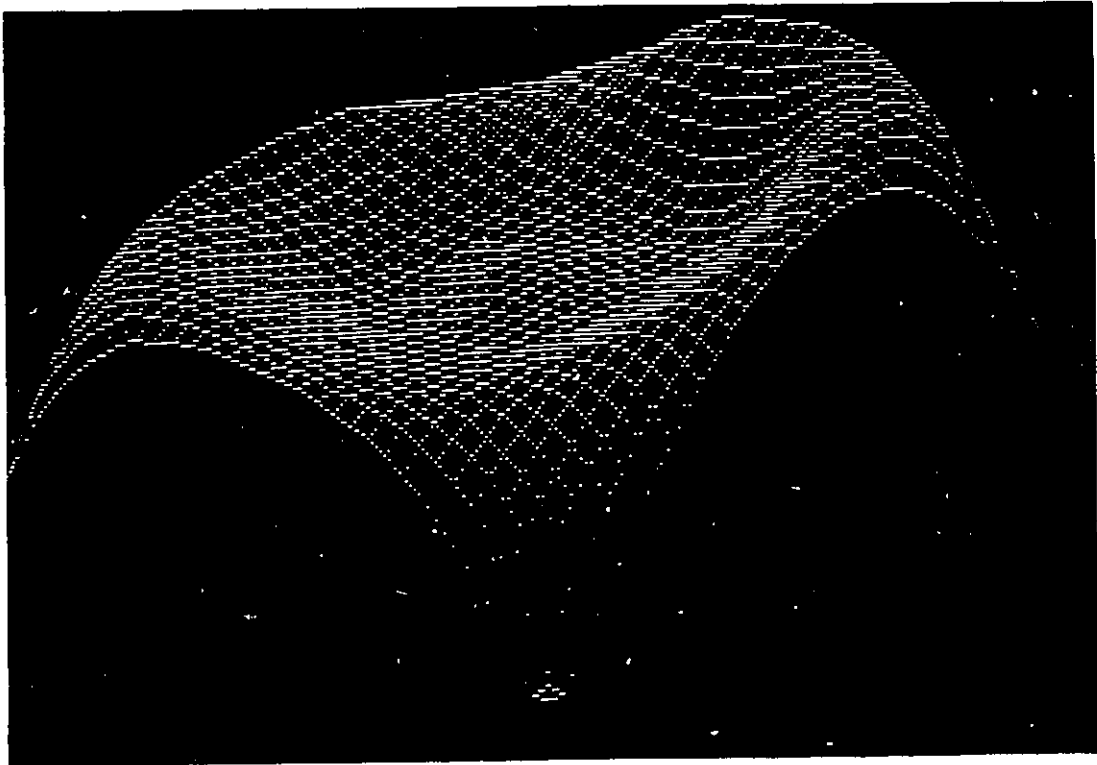
Figure 4.10 Reconstructed surface without inclusion of 'A' matrix

84
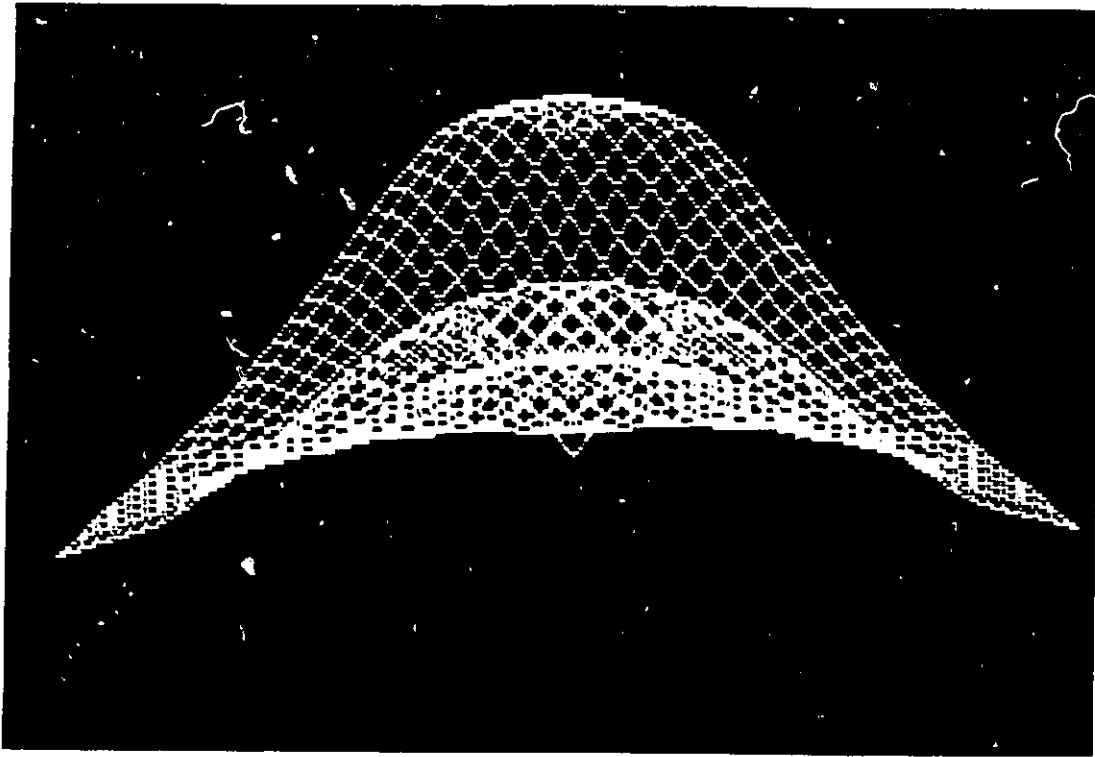
Figure 4.11 Reconstructed surface using Brook's and Horn's method
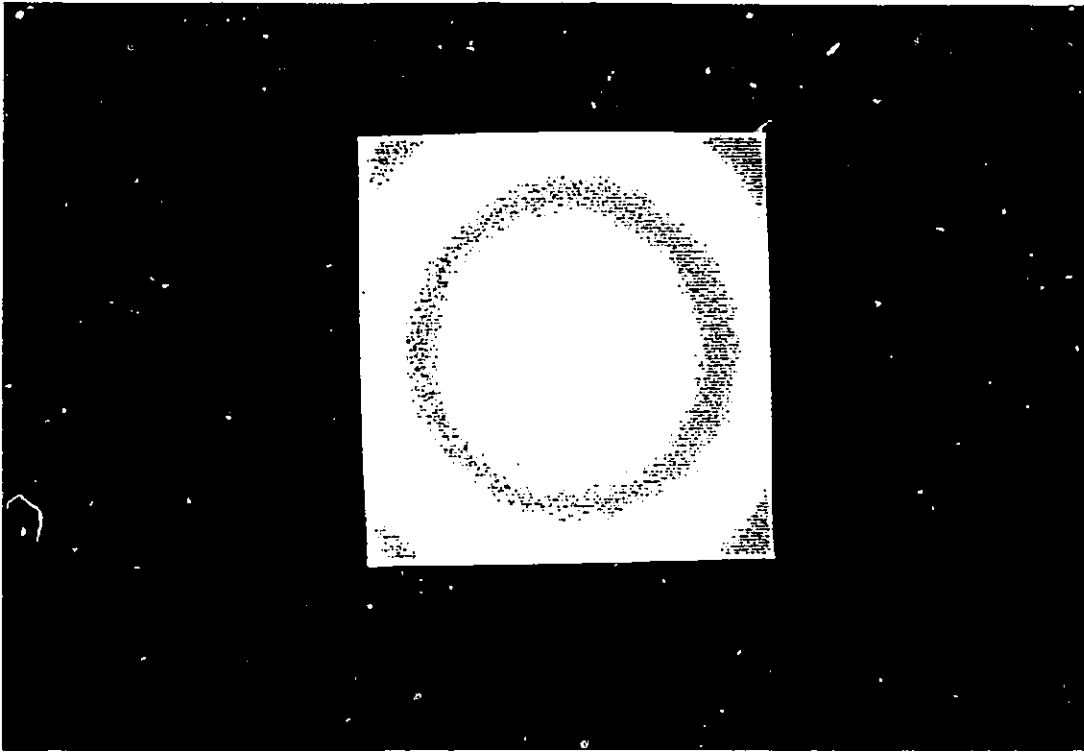
Figure 4.12 Undulating surface with light source on the z axis
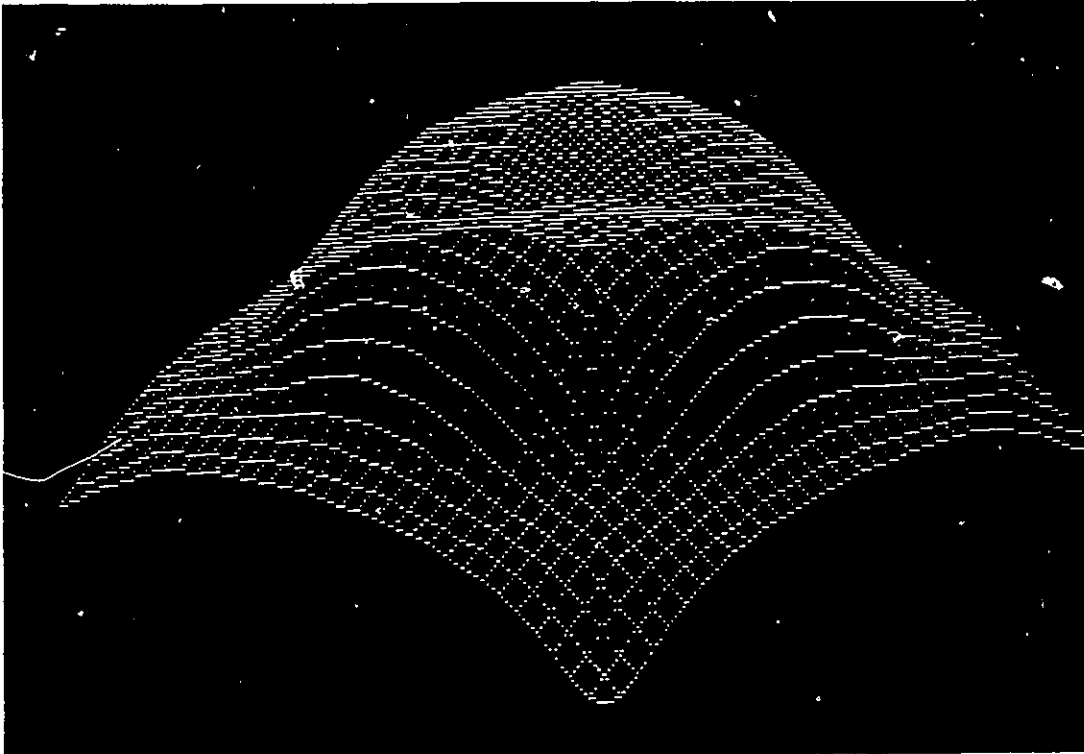
Figure 4.13 One of the possible solutions as converged
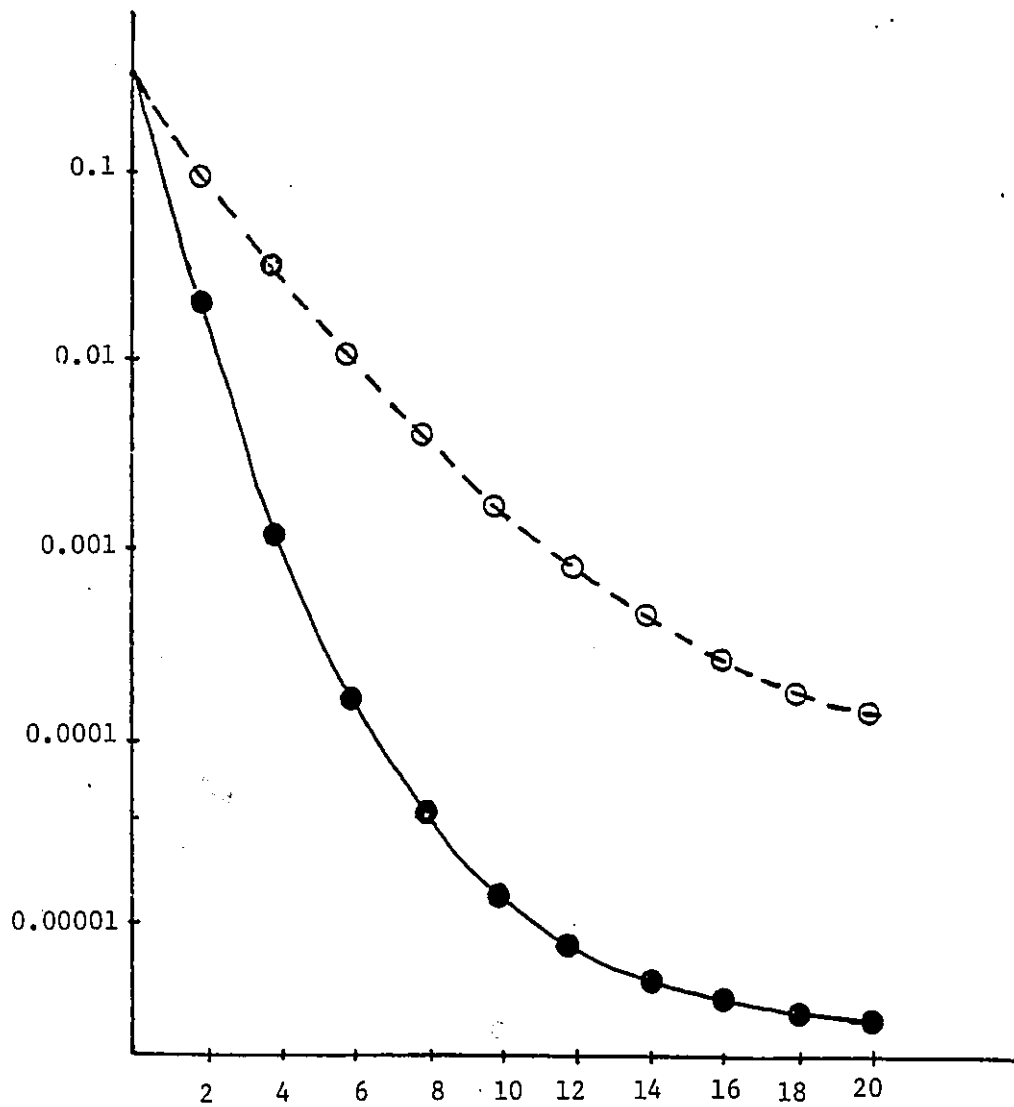
Figure 4.14 Convergence in mean square error of normals on the surface

    ——————   Our algorithm*

    ------   Brooks and Horn's algorithm*

* Comparison done on running the algorithms on the same VAX 11/785 system

-0.029	-0.073	-0.058	-0.051	-0.048	-0.046	-0.043	0.033
-0.012	-0.030	-0.030	-0.032	-0.034	-0.034	-0.035	-0.024
-0.032	-0.053	-0.044	-0.040	-0.041	-0.045	-0.057	0.085
-0.018	-0.031	-0.032	-0.032	-0.032	-0.033	-0.033	-0.031
-0.033	-0.044	-0.038	-0.033	-0.034	-0.039	-0.047	-0.065
-0.023	-0.033	-0.031	-0.031	-0.032	-0.033	-0.033	-0.031
-0.034	-0.042	-0.036	-0.033	-0.034	-0.037	-0.044	-0.053
-0.025	-0.034	-0.032	-0.031	-0.030	-0.033	-0.033	-0.033
-0.035	-0.043	-0.039	-0.033	-0.037	-0.038	-0.045	-0.059
-0.025	-0.034	-0.035	-0.035	-0.033	-0.032	-0.033	-0.034
-0.035	-0.048	-0.041	-0.039	-0.039	-0.042	0.050	0.069
-0.024	-0.034	-0.034	-0.035	-0.035	-0.033	-0.033	-0.034
-0.034	-0.059	-0.049	-0.045	-0.046	-0.051	-0.065	-0.095
-0.020	-0.033	-0.034	-0.035	-0.035	-0.034	-0.032	-0.032
-0.032	-0.040	-0.047	-0.051	-0.056	-0.067	-0.098	-0.099
-0.008	-0.031	-0.032	-0.034	-0.035	-0.034	-0.032	-0.029

Table 4.1 Principal curvatures at the center of each segment

to the final shapes was noticed to occur with in 10 passes. Additional testing performed on planes and cylinders also indicated that when the curvature is zero, the algorithm performed as consistently as spherical surfaces, even though the shape descriptor A has a smaller role to play in convergence in such cases.

5.2 Studies with Real World Images

In experimenting with real world images, two of the images were chosen to observe the potential for reconstructing surfaces with slow varying curvatures and the third image was considered for both its positive and negative curvatures. The first image is that of a basket ball (with its markings removed) lit almost frontally. The light source information provided to the algorithm was a very coarse estimate without any actual measurement of the illumination geometry. Figures 4.15 and 4.16 show the picture of the ball and its reconstruction respectively. The second image was that of a flower pot. Interestingly, while curvature is zero in one of the directions, in the other direction, curvature changes constantly. Figures 4.17 and 4.18 show the picture of the flower pot and its reconstruction respectively. The third image is that of a well curved base of a lamp. The results are shown in Figures 4.19 and 4.20. In all the real world image experiments it was found that the algorithm reached the shape presented here in about 20 passes.

5.3 Studies on Noise Sensitivity

The sensitivity of the algorithm to measurement noise (intensity) can be measured by generating white noise with a range of variances for a synthetic data set. The performance of the algorithm is given

89

Figure 4.15 Photograph of a sherical ball
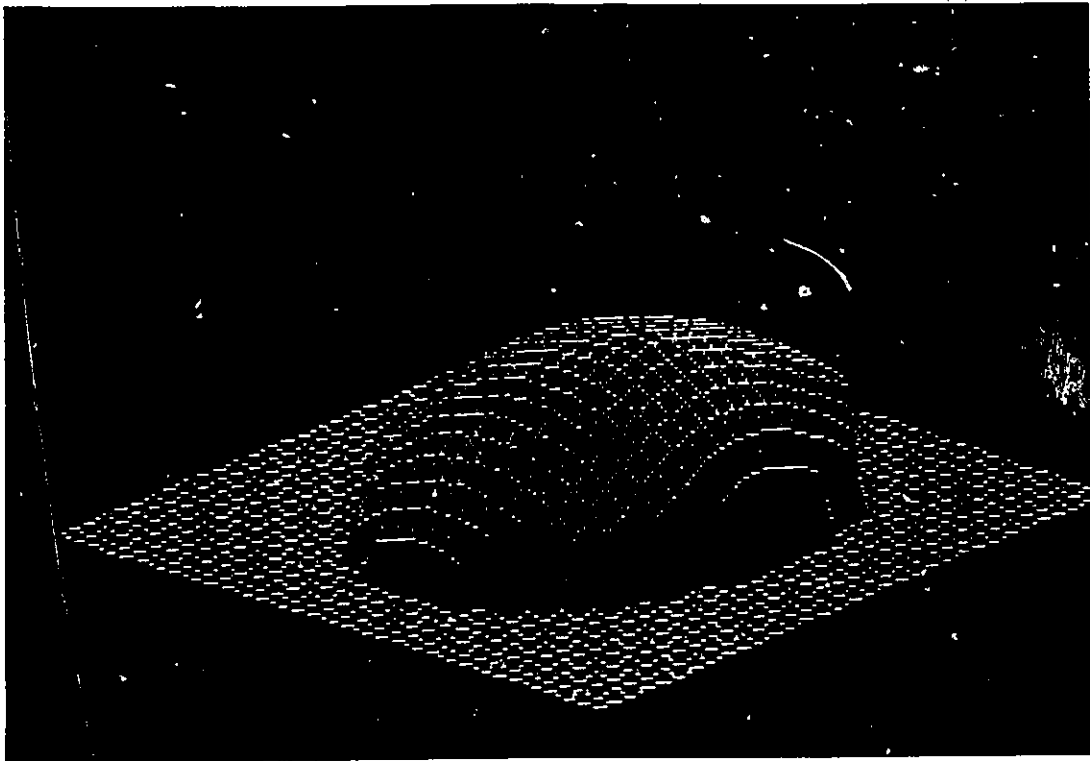
Figure 4.16 Reconstructed shape where intensity was available

Figure 4.17 Image of a flower pot illuminated from an oblique front
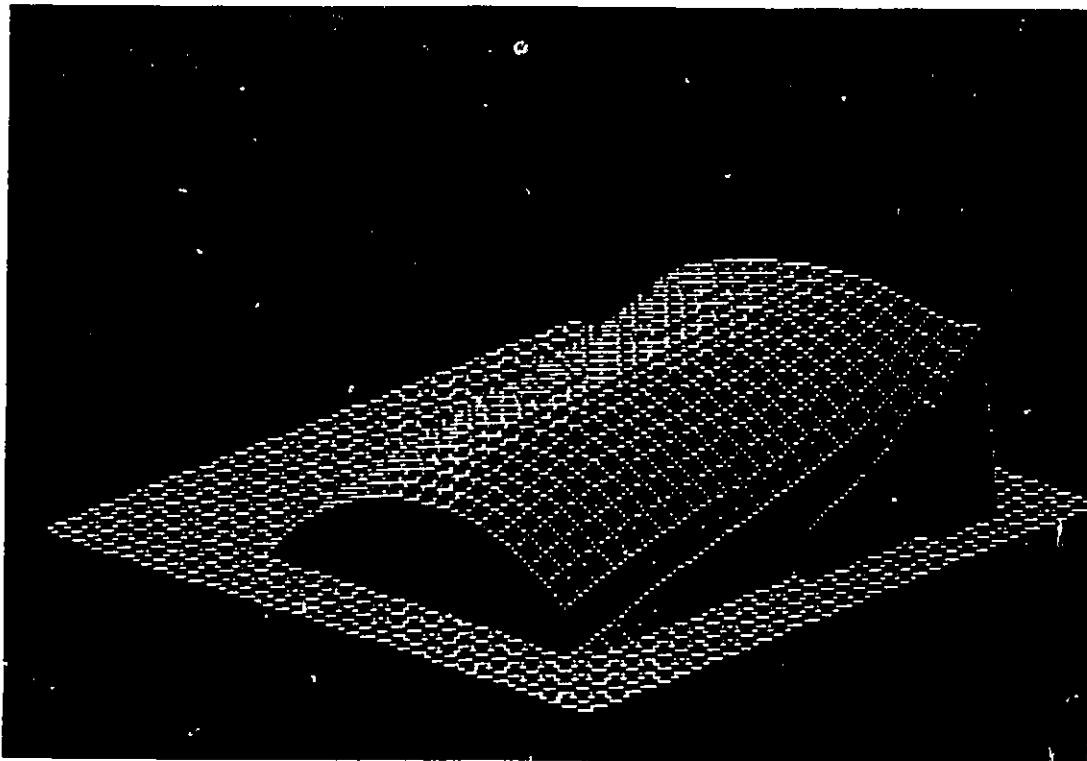
Figure 4.18 Reconstructed shape at the illuminated regions

91
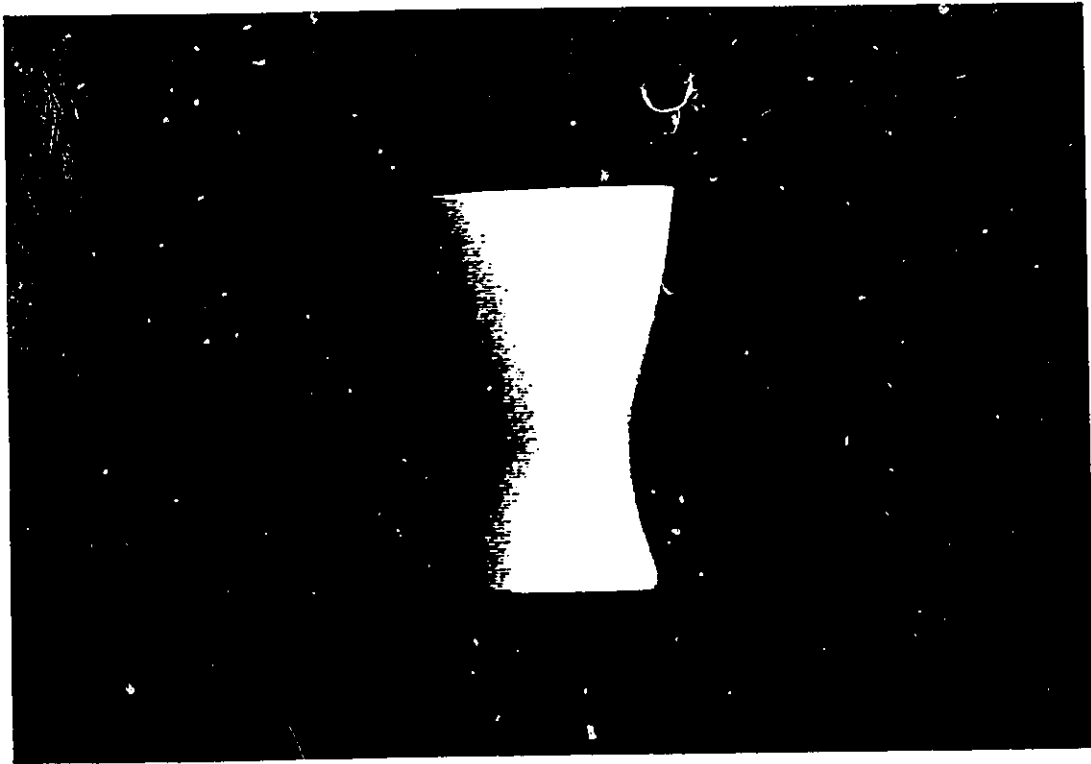
Figure 4.19 Image of the smooth base of a lamp
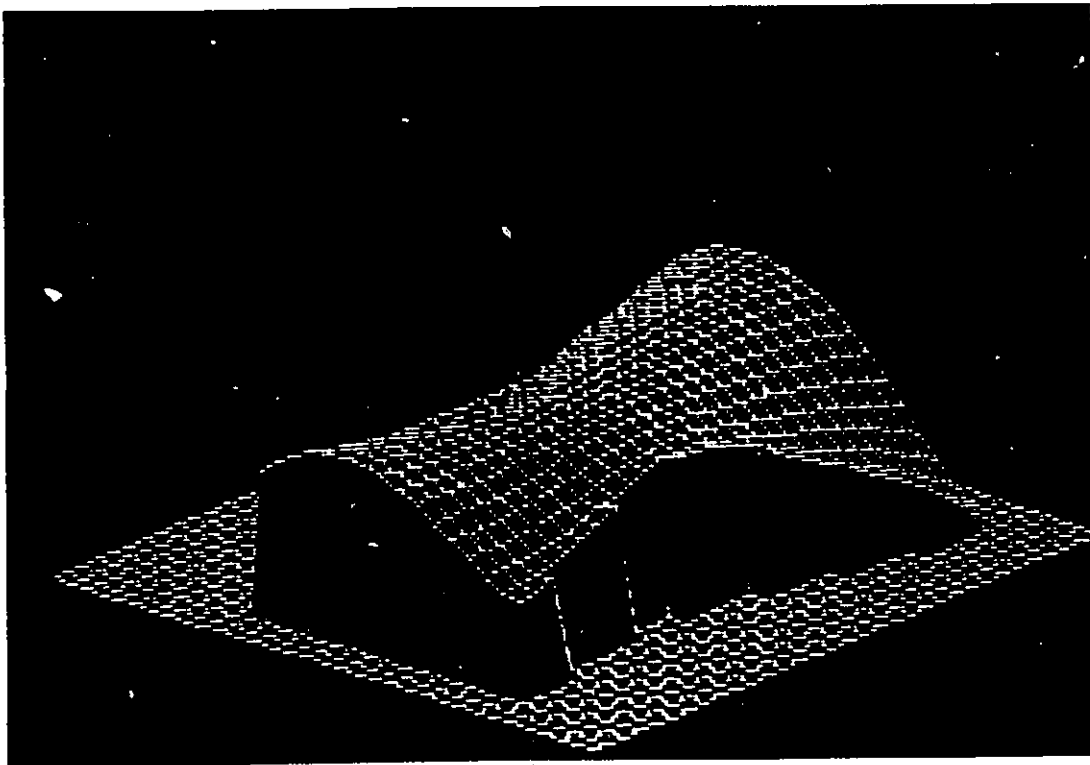
Figure 4.20 Reconstructed shape at illuminated points

here in terms of the error indices defined below.

$$E_z^{(1)} = \sum (z - \hat{z})^2 / S$$

$$E_z^{(2)} = \sum (z - \hat{z})^2 / \sum z^2$$

$$E_n = \sum ||n - \hat{n}||^2 / S$$

where the quantities with the $\hat{}$ are estimated values and S is the total number of pixels in the image. Similar indices are also defined for Gaussian and mean curvatures.

With the conventional definition of Signal to Noise ratio [Pra 78] given as,

$$S/N = -20 \log_{10}(\, sqrt(\sum (I - \hat{I})^2 / \sum I^2))$$

the performance of the algorithm has been observed for a synthetic sphere with radius approximately the size of the image. The error indices are tabulated in Table 4.2. The S/N values used here range from approximately 40 db to 3db.

A graceful degradation of performance occurs as the noise power increases. That is, with a S/N figure of 40 db and higher the algorithm almost reconstructs the original surface. As we lower the S/N ratio we find that although error indices of the normals and the curvatures are not very high, the error indices for z tend to grow larger. However, notice that below 6 db the z error indices are not reliable anymore since at some of the points the intensity values of the corrupted image is below zero (and hence masked from doing any computation). The true indicators in such cases with low signal strength turns out to be the normal and curvature error indices and

93

Error Measures	Signal / Noise Ratio						
	38.90 db	24.92 db	18.90 db	15.37 db	12.88 db	5.83 db	3.33 db
$E_z^{(1)}$ – Error in Z	0.0134	0.0837	0.1840	0.2747	0.3709	0.2612	0.3164
$E_z^{(2)}$ – Error in Z	0.0552	0.2645	0.4425	0.5424	0.6155	0.2962	0.3209
E_n – Error in N	0.0018	0.0052	0.0107	0.0140	0.0166	0.0221	0.0229
E_{gc} – Error in Gaussian Curv.	0.0000035	0.0000075	0.0000334	0.0000433	0.0000436	0.000047	0.000049
E_{mc} – Error in Mean Curv.	0.0000673	0.0000870	0.0002004	0.0002410	0.0002508	0.000289	0.000312

Table 4.2 Noise Sensitivity on a Synthetic Sphere of Radius 30 Units

they do indicate further eccentricity of the converged surface. A cross sectional slice of the converged surface depicting the estimated z-values is shown for different S/N ratios in Figure 4.21.

The following observation is implied from this behaviour. The intensity values for the surface chosen here are slowly varying in the centre (hence small variance) while near the image boundaries the intensity changes quickly. The direct effect of adding white noise in this situation is that the algorithm attempts to fit the best possible smooth surface that satisfies the data at least in a "global sense". Thus, the converged surface tends to get elongated in the center, resulting in a paraboloid with larger foci as the S/N grows (see Figure 4.21). Yet another point to notice is that the smoothness constraints with additional shape information through the shape matrix A appear to stabilize the algorithm to a large extent. We leave this section with the note that this result may be corroborated in human vision by experiments similar to Mingolla and Todd [MiT 86, ToM 83].

5.4 Effect of Initial Conditions and Speed of Convergence

As true with any other iterative procedures, a good set of initial conditions can drive the algorithm to faster convergence. Since the shape of the surface is not a priori known to us, one of the best possible input condition would be the intensity map itself. In case of the lamp-base image experiment, the effect of using intensity map as the initial condition significantly improved the convergence rate (about 12 passes to 8 passes). Further improvements on the speed can be seen when the algorithm is first run on 32 X 32 sampled versions of the original images, as the up-scaled and smoothed version

S/N Ratio of 38.90 db

24.92 db

18.90 db

15.37 db

12.88 db

5.83 db

3.33 db

⊛ Points at which no information is computable.
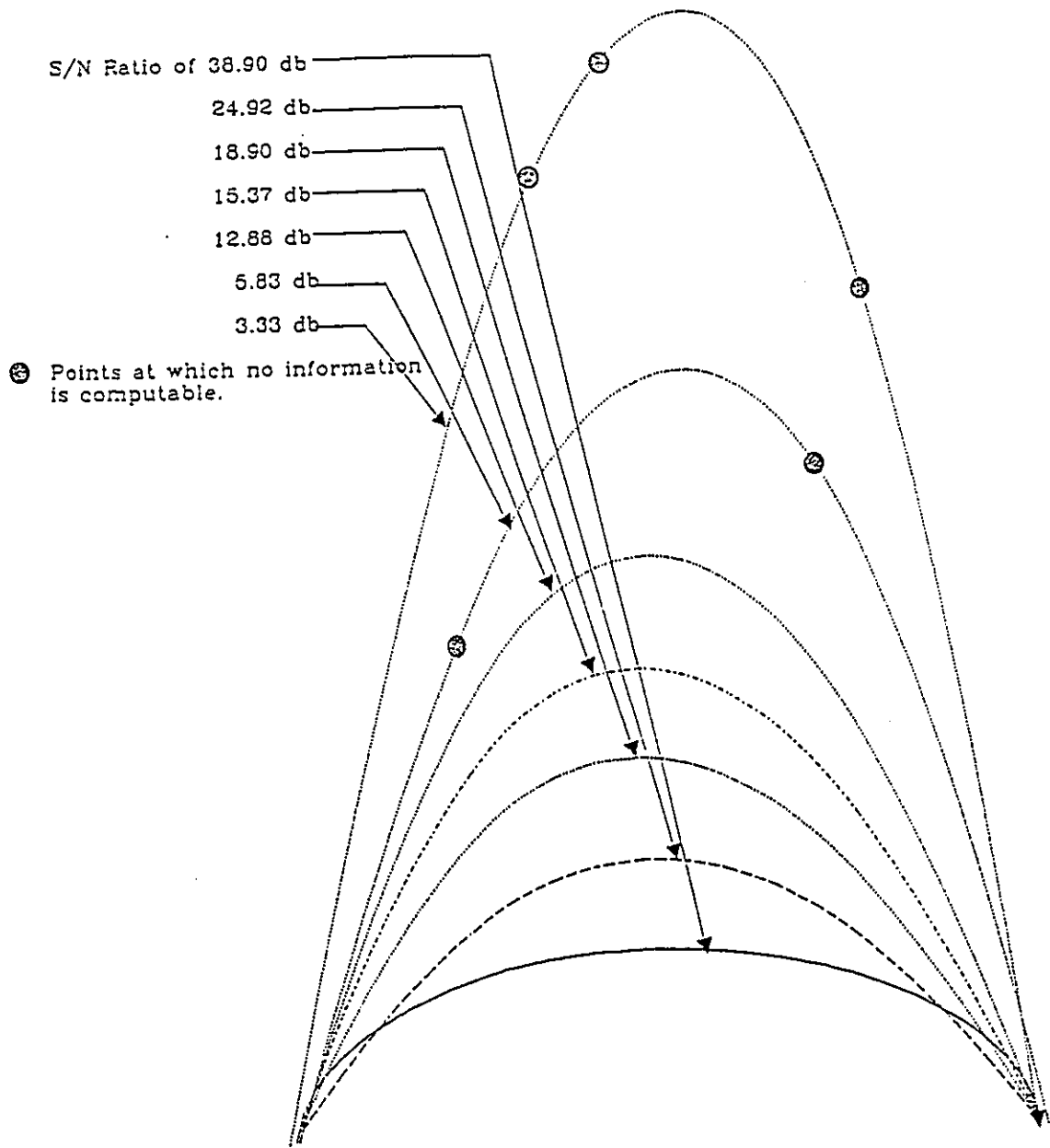
Figure 4.21A Cross Sectional Slice of the Converged Surface
for Different S/N Ratios

of the converged surface on the 32 X 32 version provides a better initial condition. The real world images experimented here were tested in this fashion.

6. Summary

The problem of reconstructing the shape from the intensity information has been a provably difficult one. Since, several surfaces with different viewing conditions can give rise to the same image, it imposes the severity in obtaining a solution. Thus, unless some generic constraints are specified (even then multiple solutions are unavoidable), it becomes difficult to pick the required solution. This work was essentially aimed at building a model to reconstruct the shape with importance given to control error, better speed and obtaining curvatures locally. The crux of the approach is to embed the concept of specifying local shape variation into the method as opposed to the current trend in postponing shape computing to the end of the surface reconstruction problem. A few interesting points here are:

(i) The surface reconstruction as well as shape estimation are achieved simultaneously.

(ii) Depth estimation results as a byproduct even without an additional process of imposing integrability, etc.

(iii) The use of local shape description, via the A matrix, is demonstrated here by showing that curvature information can easily be obtained from this shape descriptor.

(iv) Tests with synthetic and real imagery show that the algorithm converges to robust and accurate estimates of shape, normals and depth.

Chapter 5

Integrating depth and shape

1. Introduction

The richness of information present in every visual cue is significantly tied to the scene. A scene characterized by surfaces with a lot of albedo changes and discontinuities can be exploited quite well by stereo vision. Similarly, if the scene contains objects that are smoothly varying in shape and irradiance then the shape from shading process can deliver reliable results. Each visual cue imposes its own restrictions in order to extract the shape and depth information from the visual data. Thus, the issue here is how visual cues can be combined in a graceful way such that the richness of perception is preserved, no matter which source is dominant.

Only recently, this problem has begun to receive some attention [AlB 88]. An easy approach to this problem is to formulate the early visual tasks in such a way that the available partial information become constraints that need to be satisfied by the shape computing process. If the partial information is in the form of surface derivatives, then the smoothness constraint incorporated in computing shape implicitly uses the partial information to propagate to the neighbor regions. However, if the partial information is in the form of depth at some points, then additional set of constraints must be derived to require that the reconstructed surface preserves the depth at these points. In this chapter, we address the issue of combining any a priori knowledge of depth and shape with the surface reconstruction process from shading.

2. Problem statement

We shall represent the image plane as a grid G. We define the following quantities as,

$$P = \{p_i \mid p_i \in G\}$$
$$D = \{p_j \mid z = z_j \text{ and } p_j \in G\}$$
$$N = \{p_k \mid n = n_k \text{ and } p_k \in G\}$$

and $P \subseteq G$, $D \subseteq G$, and $N \subseteq G$. That is, with known depth at points points P_d in the set D and known normal n at points P_n in the set N, compute n and z for every point p in G. The solution to this problem is obtained using the shape computing algorithm described in Chapter 4. While, depth information is obtainable through stereopsis, normals are available at the occluding boundaries by analyzing the occluding contours [Kov 82, Mar 77b]. In the following section, we show how this additional piece of information can be combined with shape from shading to reconstruct the surface.

3. Shape from Shading as an Integrating Module

An important observation about combining a priori normals and depth is as follows. It is not necessary that depth is available only at points where the normal is available or $D \cap N$ need not be empty. For instance, at places where a surface discontinuity is present, stereopsis can deliver depth, but due to the surface continuity requirement, the shape from shading process can not compute normals at those points. Thus, because of the arbitrary nature of availability of normal and depth, it becomes difficult to implement the 2 1/2 d sketch using the common pool idea of Marr and Nishihara [MaN 78]. Instead, a straight forward approach is to incorporate any known or early

99

available depth and normal in a shape computing process. We describe below first the addition of known depth and later the addition of known normals in aiding the process of surface reconstruction.

3.1 Incorporating Known Depth and Normals in the Shape Fitting Paradigm

The approach adopted here is to append additional constraints in the shape from shading formulation discussed in the previous chapter. We add the additional terms J_4 and J_5 to (4.9) derived in chapter 4, as follows.

Let $J = \sum J_1 + J_2 + J_3 + J_4 + J_5$

where,

$$J_1 = (1/N_{i,j}) \; \alpha \sum_{\substack{k,l \\ \in \eta(i,j)}} \; || \; n(i,j) - n(k,l) - A \, \Delta \; ||^2$$

$$J_2 = (1/N_{i,j}) \; \beta \sum_{\substack{k,l \\ \in \eta(i,j)}} \; || \; \{ \; (n(i,j) + n(k,l))/2 \; \}^T \, \Delta \; ||^2$$

$$J_3 = \gamma \; \{ \iota^T n(i,j) - g(i,j) \}^2$$

----- 5.1.

$$J_4 = \sum_{P_{i,j} \, \in \, D} \sigma \; (z_e(i,j) - z_t(i,j))^2$$

$$J_5 = \sum_{P_{i,j} \, \in \, N} \lambda \; || \; n_e(i,j) - n_t(i,j) \; ||^2$$

where, α, β, γ, σ and λ are appropriately chosen weights,

$$\Delta = \begin{bmatrix} i - k \\ j - l \\ z(i,j) - z(k,l) \end{bmatrix}$$

$\eta(i,j)$ is the neighborhood of (i,j)

$N_{i,j}$ is the number of elements in $\eta(i,j)$

$z_e(i,j)$ represents the estimated depth at points $P_{i,j}$

$z_t(i,j)$ is the depth available from stereopsis at $P_{i,j}$

$n_e(i,j)$ is the estimated normal at occluding contours

$n_t(i,j)$ is the normal available at occluding contours

The optimum estimates of n, z, and **A** are obtained by minimizing J subject to the constraint

$$|| n(i,j) ||^2 = 1. \qquad \text{----- 5.2}$$

as before in chapter 4.

The effect of adding the extra set of above constraints only slightly alters the solution to the minimization problem. Thus, in fitting z for the given set of n and **A**, we obtain the solution by including an additional term with (4.14) as follows.

$$\sum_{k,l \in \eta_{i,j}} [A_3^T(A \Delta - n(i,j) + n(k,l)) +$$

$$\beta \{(n_3(i,j) + n_3(k,l)\} \{n(i,j) + n(k,l)\} \Delta^T +$$

$$\sum_{P_{i,j} \in D} \sigma (z_e(i,j) - z_t(i,j)) = 0 \qquad \text{------ (5.3)}$$

The effect of the extra term is to fix the z in the block Gauss Seidel updation scheme, at the points where depth is known. In incorporating the known normal information, we notice that the continuity constraint plays a key role. The effect of this constraint is to allow the known normals to propogate. The earlier work on combining normals from occluding contours with shape from shading include Ikeuchi and Horn [IkH 81], Frankot and Chellappa [FrC 87] and Brooks and Horn [BrH 85]. Ikeuchi and Horn use normals at the occluding contours as boundary conditions for solving the irradiance equation. However, in [FrC 87], the boundary information is allowed to propagate to other points through the iterative process that updates normals, similar to our method. The advantage in our method is that the derivatives of the normals (A matrix) are also used in the process of propagation.

3.2 Experimentation

In order to demonstrate the potential of our method to integrate depth with shape from shading, a synthetic sphere was created with the occluding boundaries inside the image frame. The depth at the occluding boundaries was simulated to be different from the background. The result of reconstruction is shown in Figure 5.1. It may be noticed that the boundary of the sphere is located well above the background since the depth at the boundaries were specified to be much above the background (about 15 units).

In another experiment the normals at the occluding boundary was specified for the same synthetic sphere and the reconstructed surface
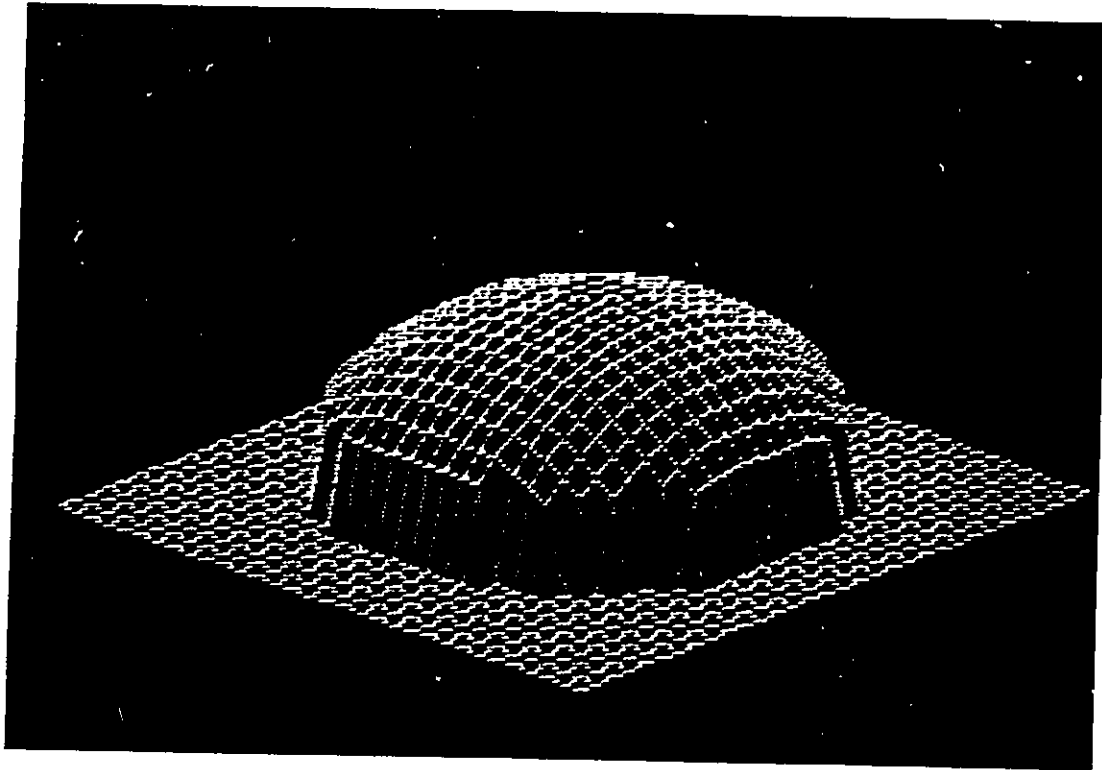
Figure 5.1 Reconstruction of a synthetic sphere with non-zero
background depth information

is shown in Figure 5.2. In comparing this result with the case when normals at occluding boundary was not included (Figure 5.3), little difference is noticeable in the reconstructed surface. However, there appears to be a slight improvement in the numerical figures of the shape information near the boundary (of the order of precision around fifth or higher decimal point). This is not surprising because the role of normals at the occluding boundaries is primarily to guide the convergence to one of the two possible solutions (concave or convex). The noticeable improvement is found to be the increase in speed of convergence for the same error figure. While it took 10 iterations for the reconstruction in Figure 5.3, it took only 7 iterations in the case when normals at the occluding boundary was specified.

In yet another experiment, the stereo pictures of a coffee cup were taken and processed for the stereo depth with $\nabla^2 G$ (large filter width to avoid surface markings) filtered edge maps (see Figures 5.4 and 5.5). Results of integrating the depth and shape are shown in Figure 5.6. Noticeably, the handle of the cup suffers some loss of accuracy since the shading informating in that area was not large enough to drive the convergence to a much smoother solution. A much higher resolution imaging can alleviate this problem to quite a large extent.

Shape information in the form of normals available from other processes such as shape from textures can also be added in a similar fashion. On the same token, depth information from the analysis of motion parallax, can also be integrated using the method described
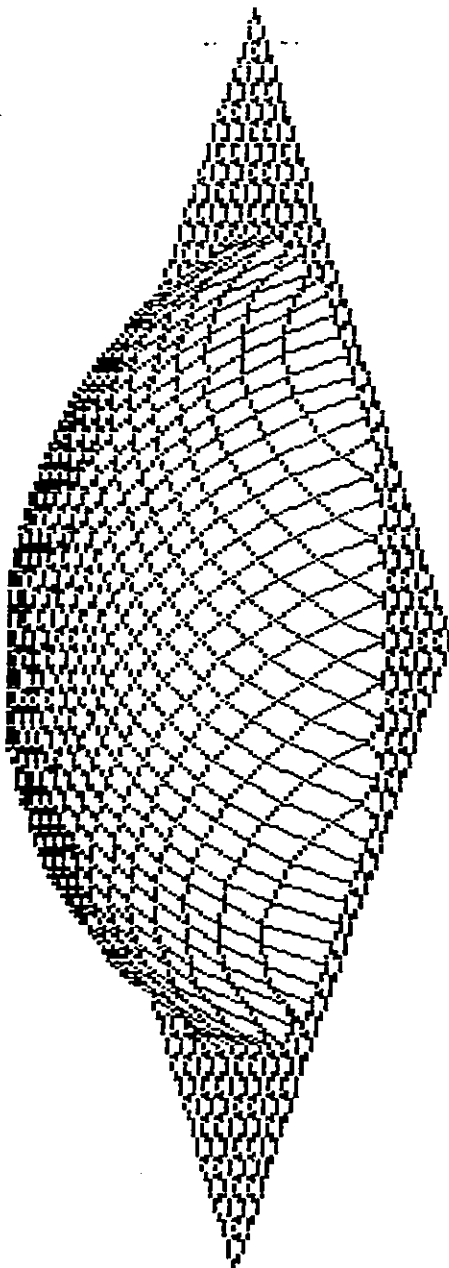
Figure 5.2 Reconstructed sphere with pre-specified normals at the occluding boundaries
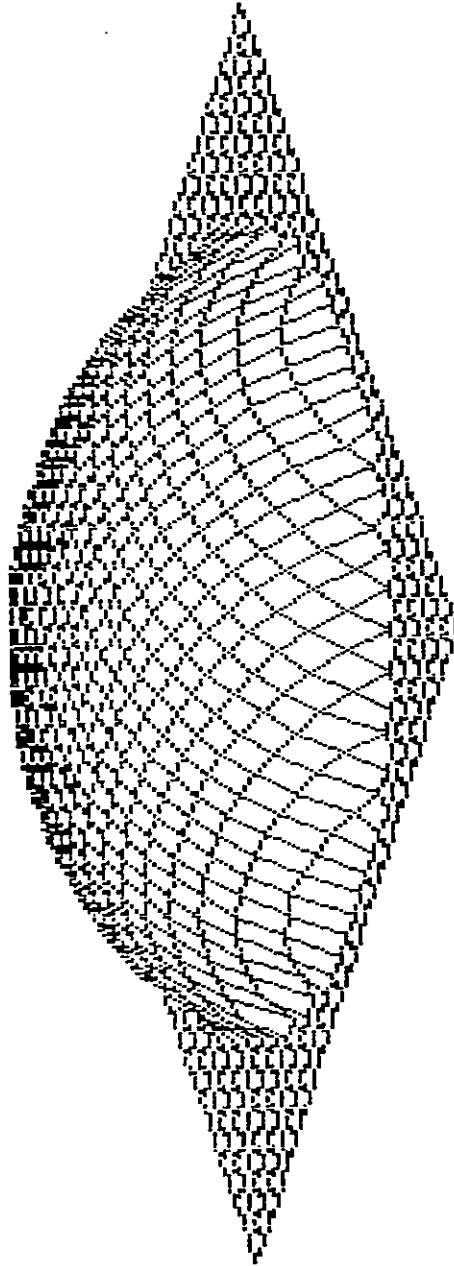
Figure 5.3 Reconstructed sphere with unspecified normals at the occluding boundaries
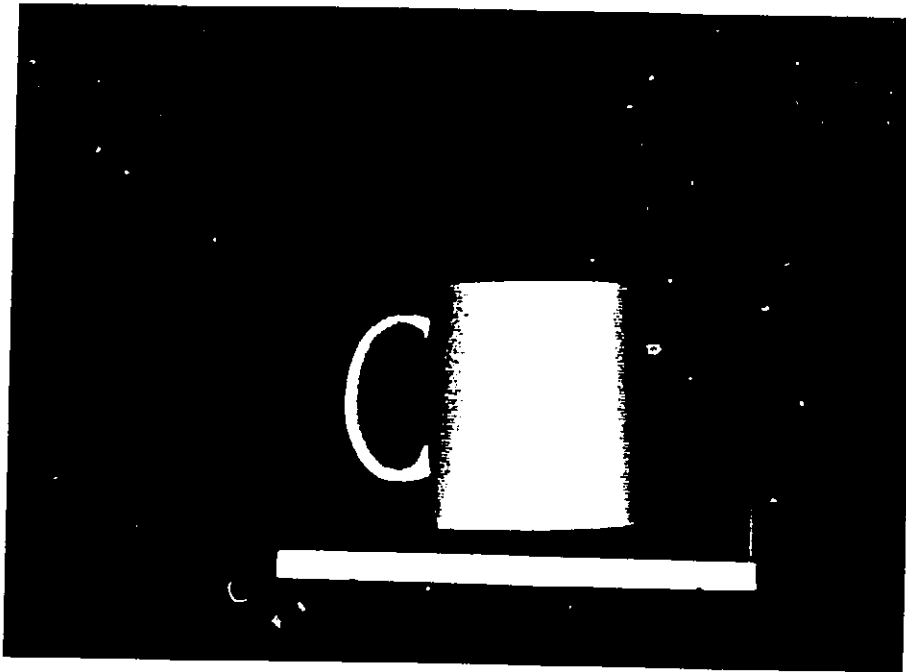
Figure 5.4 Left image of a coffee cup stereo-pair
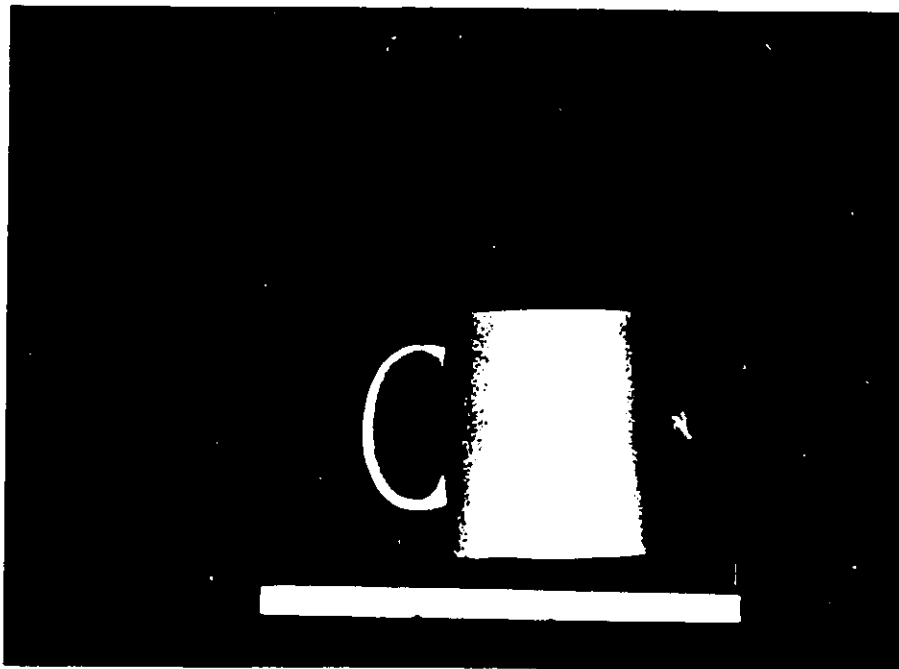
Figure 5.5 Right image of the coffee cup stereo-pair
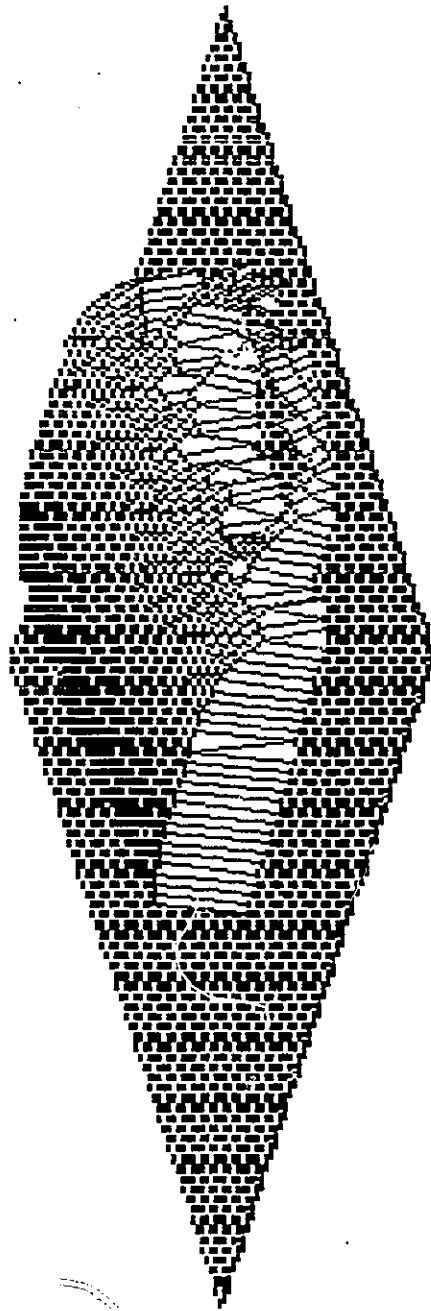
Figure 5.6 Reconstructed coffee cup using shape from shading and depth from stereo

before.

4. Summary

Any given instance of the visual imagery can contain an arbitrarily set of visual cues to reconstruct the surface. A method to integrate shape and depth information from these cues is the focus of this chapter. As an extension to the algorithm for shape form shading presented earlier, depth from stereopsis can be incorporated using an additional set of constraints. Also discussed here is the implicit ability of the algorithm to include any shape information in the form of normals that could be available from other processes.

The primary advantage of using shape from shading as the integrating stage is that it offers robustness and better error immunity due to the inherent nature of shape fit.

Chapter 6

Summary and Conclusions

1. Recapitulation

The principal objective of the research work described in this dissertation was to develop a framework to reconstruct a viewed surface in depth and shape using stereo vision and shading.

To achieve this goal we first investigated the underlying problems in using stereo disparity to compute depth at every point in the image plane. Stereo vision requires matching of corresponding points in the two images and then computing depth at these points. However, correspondence can be established only at points that hold strong features such as points of intensity changes. Since, any given scene can only contain a few intensity changes, our first problem was to find an appropriate scattered data interpolant to obtain depth at all points.

The computational theory of Grimson for surface interpolation was found to possess an inherent difficulty in handling surfaces with sharp discontinuities. A stronger surface consistency constraint that ensures that discontinuities are preserved by the interpolant with out addition of any new zero crossings, was proposed here. A direct implication of this constraint is that the interpolant should not be differentiable at the zero crossings, while they must be at least twice differentiable at every other point. A class of surfaces known as Shepard's metric interpolants are found to possess this property.

The Shepard's interpolants have been implemented here on Random Dot Stereograms as well as natural stereo pairs. The result from

110

testing on RDS shows that stereo vision can function alone without any additional visual cues. The natural stereo pair shows that even when intensity changes are sparse, the reconstruction preserves the shape. In addition, since stereopsis is one of the earliest modules in vision it appears to deliver the quick and coarse estimates of depth that may be used by other surface construction processes. In that sense, Shepard's interpolant proves to be one of the best candidates for the purpose of stereo data interpolation.

Besides depth from stereopsis, shape information also becomes important to reconstruct the surface. An important shape cue is available in the smooth shading that an object renders. Our next goal was to study the problem of computing shape from shading and build a scheme that delivers surface intrinsic information such as local curvatures. The methods proposed in the literature were found to solve this problem from a different perspective altogether. That is, the final objective for these methods is not to compute curvatures but only the normals. Thus, to achieve curvatures the normals have to be differentiated and near the occluding boundaries such a differentiation is highly noise prone.

Instead, in our formulation we begin with local shape descriptors. Interestingly, shape information is intrinsic to the surface and is independent of viewer position. Since continuity in normals is ensured through these shape descriptors, the numerical error introduced in the process of reconstruction is shown to be independent of coordinate axes chosen. By imposing constraints of continuity, integrability of shape descriptor, minimal measurement error and unit normal, a global

objective function is minimized with respect to the normals to obtain the normals. Also, obtained here as a byproduct is the relative depth at every point in the image plane. The principal curvatures are shown to be computable from the shape descriptor used here.

The above method has been tested over a variety of synthetic and natural images. The performance was found to be quite satisfactory in terms of speed, error in reconstruction and robustness amidst the presence of noise.

As it has been stressed from the beginning, our overall concern in this work is to ensure that any shape and depth information that may be available from other visual cues should be easily incorporated into reconstructing the surface. To show that our framework addresses this concern, it was shown earlier that depth information at an arbitrary set of points can be included as additional set constraints in the shape computing algorithm. Also, any known set of normals can also be exploited to improve the convergence of the shape from shading algorithm, besides smoothly incorporating the additional source of information.

In conclusion, the contributions of this work include the following:

(i) Visual surface interpolation has been studied to propose a stronger form of surface consistency constraint and thereby dealing with surface discontinuities with more ease.

(ii) Shepard's metric interpolants are shown to implement this constraint quite faithfully besides being a computationally less expensive scheme.

(iii) The notion of local shape description is emphasized to suggest that shape computing processes can deliver surface information in a more robust fashion.

(iv) Based on the above notion, an algorithm for solving shape from shading has been proposed. Some salient features of the algorithm are

    (a) relative depth is also delivered besides normals

    (b) local principal curvatures are also computed.

(v) Shape and depth from any other source has been shown to be integrated with shape from shading, thereby suggesting that the entire framework can deliver viewer independent surface information from raw visual data.

In the following section we discuss an outline on how this information can be used to advantage in building a complete vision system. We also attempt to describe a few issues that the work has generated for further investigation.

2. Future directions

Possible future work may be sketched at two different levels. At more of a lower level, the following are some important issues that need some attention.

1. Beginning with surface interpolation, even though Shepard's surfaces provide an efficient means to obtain a quick estimate of depth, it is quite sensitive to noise in stereo matching. Some form of weighted smoothing can be done if we a priori know the actual data points. Since it is difficult to distinguish the data points from noise points, one way to get around the difficulty (not necessarily

113

the best way) is to estimate local peaks and set those peaks to an average of its neighbors. Such secondary processes add to the computational problems, there by totally nullifying the advantage of the method itself.

2. The shape from shading algorithm definitely exhibits faster convergence in most of the cases considered here except the undulating surface. Besides the speed of its convergence, the issue of convergence itself needs to be addressed. With relevance to the shape from shading algorithms, the only attempt to establish analytical proof of convergence is that of David Lee [Lee 85]. Similar attempt needs to be undertaken here too. In addition, although estimation of normals and the shape matrix are very fast, the estimation of z for the normals and shape matrix at the current iteration is considerably slow. Speeding up the algorithm is yet another problem that deserves some investigation.

3. A similar model for shape from texture can be constructed by simply replacing the irradiance constraint with a constraint that relates the orientation to the distribution of texture elements.

At a much higher level, two plausible directions are indicated below.

1. Every perceptual task may be analyzed at three different levels, namely computational theory, algorithm and representation, and physical implementation. However, we were only concerned about the first two levels in this work and the issues related to physical implementation are yet to be addressed. Interestingly, the Shepard's interpolant as well as the shape from shading algorithm hold very

strong potential for parallel implementation. With the recent results in analog VLSI [Mea 89] and resistive networks [Hut 89], it is possible to reduce the shape computing problem to an energy or entropy minimization problem. Such networks [HoT 85] have been primarily built for computing scalar quantities from global minimization. The issue here is how these networks can be adopted for our problem of computing the normals. The nonlinear unit normal constraint adds additional complexity to this issue.

2. Referring to the hierarchy of vision described in chapter 1, we find that this work spans from raw visual images to viewer independent curvature information. The following stage is to abstract the curvatures into larger entities that specify regional properties. Since, the ultimate goal is recognition of objects with their spatial relationships preserved, the current day neural nets can be exploited to build associative schemes that learns to recognize. In two dimension the recognition problem has been addressed in various ways. In three dimension, the issue becomes enormously complex that only a very few attempts have been at least partially successful.

What seems like an effortless process, vision continues to pose much tougher challenges than it appeared to. A much deeper understanding is bound to evolve with sustained growth in interest among this vast vision research community.

Glossary

Fusion Range:

- The disparity bound within which human stereo vision can successfully fuse the binocular retinal images.

Fovea:

- The central region of the retina inwhich the resolution of perception is very high in comparison with the periphery.

RDS:

- Random Dot Stereogram; It is created by generating a random texture in the left image, while the right image is a copy of the left image with a pre-identified portion shifted laterally.

LoG or $\nabla^2 G$:

- The Lapalcian of Gaussian; It is a non-directional operator used to obtain intensity changes in the images.

Lambertian Surfaces:

- Surfaces which exhibit the property that the quantity of light emitted in all directions is equal.

Bibliography

[Adi 85] Adiv G., "Determining three-dimensional motion and structure from optical flow generated by several moving objects", IEEE Trans.on PAMI, vol. 7, no.4, pp 384-401, 1985.

[AlB 88] Aloimonos J. and Basu A., "Combining information in low-level vision", University of Maryland - Technical Report, CSTR no. 1947, 1988.

[BaB 81] Baker H.H. and Binford T.O.,"Depth from edge and intensity based Stereo", Proc. Seventh Int.Joint Conf. on Artificial Intelligence, pp 631 - 636, 1981.

[BaF 82] Barnard S.T. and Fischler M.A., " Computational stereo", ACM Computing Surveys, vol. 14, 1982.

[BaT 80] Barnard S.T. and Thompson W.B., "Disparity analysis of images", IEEE Trans. on PAMI, vol 2. 1980.

[BaD 83] Barnhill R.E., Dube R.P. and Little F.F., "Properties of Shepard's surfaces", Rocky Mtn. Jl. of Mathematics, vol 13, pp 365 - 382, 1983.

[BTe 78] Barrow H. and Tenenbaum J.M., "Recovering intrinsic scene characteristics from images", in Computer vision systems, ed. by Hanson A. and Riseman E.,Academic Press, New york, NY, 1978.

[BTe 81] Barrow H. and Tenenbaum J.M., "Computational vision", Proc. IEEE, vol 69, pp 572 - 595, 1981.

[BeJ 86] Besl P.J. and Jain R., "Invariant surface chracteristics for three dimensional object recognition in range images",

Computer Vision, Graphics and Image Processing, vol.33, no.1, pp 33-80, 1986.

[BlA 89] Bolstein D. and Ahuja N., "Shape from texture: Integrating texture element extraction and surface estimation", IEEE Trans. on PAMI, vol. 11, no. 12, pp 1233-1251, 1989.

[BrH 85] Brooks M.J. and Horn B.K.P., "Shape and source from shading", Proc. Int. Joint Conf. on Artificial Intelligence, pp 932 - 936, 1985.

[BHo 81] Bruss A.R. and Horn B.K.P., "Passive Navigation", Computer Vision, Graphics and Image Processing, vol.21, pp 3-20, 1983.

[DaJ 83] Davis L.S., Janos L. and Dunn S.M., "Efficient recovery of shape from texture", vol. 5, no. 5, pp 485-492, 1983.

[Dav 63] Davis P.J., Interpolation and Approximation, Blaisdell Publishing co, 1963.

[Dev 75] Dev P. "Perception of depth surfaces in Random Dot Stereograms: A neuron model", Int. Jl. of Man Machine Studies, vol. 7, pp 511 - 528, 1975.

[Doc 76] Docarma M.P., Differential geometry of curves and surfaces, pp 154 - 155, Prentice Hall, 1976.

[Far 86] Farwig R., "Rate of convergence of Shepard's global interpolation formula", Mathematics of Computation, vol 47, pp 577 - 590, 1986.

[FeL 89] Ferrie F.R. and Levine M.D., "Where and why local shading analysis works", IEEE Trans. PAMI, vol 11, pp 198 - 206, 1989.

[Fra 82] Franke R., "Scattered data interpolation - tests of some

methods", Mathematics of computation, vol 38, pp 181 - 200, 1982.

[FrC 87] Frankot R.T and Chellappa R., "An improved algorithm for the shape from shading problem", Jl. of Indian Institute of Science, vol 66, pp 439 - 446, 1987.

[Gib 50] Gibson J. J., The perception of the visual world, Houghton Miffin, Boston, MA 1950.

[GoW 78] Gordon J. and Wison J.A., "Shepard's method of metric interpolation for bivariate and multivariate interpolation", Mathematics of computation, vol 32, pp 253 - 264, 1978.

[Gri 81a] Grimson W.E.L., From Images to Surfaces: A Computational Study of the human early visual system, MIT press, Cambridge, MA, 1981.

[Gri 81b] Grimson W.E.L., "A computer implementation of a theory of human stereo vision", Phil. Trans. Roy. Soc. Lond., vol B 292, pp 219 - 253, 1981.

[Gri 82] Grimson W.E.L., "A computational theory of visual surface interpolation," Phil. Trans. Roy. Soc. of London, vol B 298, pp 395 - 427, 1982.

[Gri 83a] Grimson W.E.L., "An implementation of a computational theory of visual surface interpolation", Computer vision, Graphics and Image Processing, vol 22, pp 39 - 69, 1983.

[Gri 83b] Grimson W.E.L., "Surface consistency constraints in vision", Computer Vision, Graphics and Image Processing, vol 24, pp 28 - 51, 1983.

[Gri 85] Grimson W.E.L., "Computational experiments with a feature

based stereo algorithm", IEEE Trans. on PAMI, vol 7, pp 17 - 34, 1985.

[HoT 85] Hopfield J. and Tank D., "Neural computation of decisions in optimization problems", Biological Cybernetics, vol.52, pp 141-152, 1985.

[Hor 75] Horn B.K.P., "Obtaining shape from shading information", in Psychology of computer vision, ed. Winston P.H., pp 115 - 155, McGraw Hill, New York, NY, 1975.

[Hor 77] Horn B.K.P., "Understanding image intensities", Artificial Intelligence, vol 8, pp 201 - 231, 1977.

[HoS 81] Horn B.K.P. and Schunk B.G., "Determining optical flow", Artificial Intelligence, vol.17, pp 185-203, 1981.

[HuW 77] Hubel D.H and Wiesel T.N., "Functional architecture of macaque monkey visual cortex", Proc. Roy. of Lond., vol B 198, pp 1 - 59, 1977.

[Hut 89] Hutchinson J., Koch C., Luo J. and Mead C., "Computing motion using analog and binary resisitive networks", IEEE Computer, vol.21, pp 52-63, 1988.

[Ike 81] Ikeuchi K., "Recognition of three dimensional objects using the extended Gaussian image", Proc. Int. Joint Conf. Artificial Intelligence, pp 595 - 600, 1981.

[IkH 81] Ikeuchi K and Horn B.K.P., "Numerical shape from shading and occluding boundaries", Artificial Intelligence, vol 17, pp 141 - 184, 1981.

[Jai 75] Jain R., "Direct computation of the Focus of Expansion", IEEE Trans. on PAMI, vol.5, no. 1, pp 58-64, 1983.

[Jul 60] Julesz B., "Binocular depth perception of computer generated patterns", Bell Tech. Sys. Jl., vol 39, pp 1125 - 1162, 1960.

[Jul 71] Julesz B., "Foundations of Cyclopian perception", The University of Chicago Press, Chicago, 1971.

[Kan 81] Kanade T., "Recovery of three dimensional shape of an object from a single view", Artificial Intelligence, vol 17, pp 409 - 460, 1981.

[Kov 82] Koenderink J.J and van Doorn A.J., "The shape of smooth objects and the way contours end", Perception, vol 11, pp 129 - 137, 1982.

[Lee 85] Lee D., "A provably convergent algorithm for shape from shading", Proc. of DARPA Image Understanding Workshop, Miami, Dec.85.

[LeR 85] Lee C.H. and Rosenfeld A., "Improved methods for estimating shape from shading using the light source coordinate system", Artificial Intelligence, vol 26, pp 125 - 143, 1985.

[LeH 85] Lee S.J., Haralick R.M. and Zhang M.C., "Understanding objects with curved surfaces from a single perspective view of boundaries", Artificial Intelligence, vol.26, no.2, pp 145-170, 1985.

[LoH 80] Longuet Higgins H.C and Prazdny K., "The interpretation of moving retinal images", Proc. Roy. Soc. Lond., vol B 208, 1980.

[LoH 81] Longuet Higgins H.C., "A computer algorithm for reconstructing a scene from two projections", Nature, vol. 293, pp 133-135, Sept. 1981.

[Mar 77a] Marr D., "Early processing of visual information", Proc. Roy. Soc. Lond, vol 197, pp 441 - 475, 1977.

[Mar 77b] Marr D., "Analysis of occluding boundaries", Proc. of Roy. Soc. Lond., vol.B 197, pp 441-475, 1977.

[Mar 80] Marr D., "Visual information processing: The structure and creation of visual representation", Phil. Trans. Roy. Soc. Lond., vol B 290, pp 199 - 218, 1980.

[Mar 82] Marr D., Vision, Freeman and Company, New York, NY, 1982.

[MaH 80] Marr D. and Hildreth E., "Theory of edge detection", Proc. Roy. Soc. Lond., vol B 207, pp 187 - 217, 1980.

[MaN 78] Marr D. and Nishihara H.K, "Representation and recognition of the spatial organization of three dimensional shapes", Proc. Roy Soc. Lond., B 200, pp 269 - 294, 1978.

[MaP 76] Marr D. and Poggio T., "Cooperative computation of stereo disparity", Science, vol 194, pp 283 - 287, 1976.

[MaP 79] Marr D. and Poggio T., "A computational theory of human stereo vision", Proc. Roy. Soc. Lond., vol B - 204, pp 301 - 328, 1979.

[May 82] Mayhew J.E.W, "The interpretation of stereo disparity information: Computation of surface orientation and depth", Perception, vol 11, pp 387 - 403, 1982.

[MaF 81] Mayhew J.E.W and Frisby J.P., "Psychophysical and computational studies towards a theory of human stereopsis", Artificial Intelligence, vol 16, pp 349 - 385, 1981.

[MaL 82] Mayhew J.E.W and Longuet Higgins H.C., "Computational model of binocular depth perception", Nature, vol 297, pp 376 -

378, 1982.

[Mea 89] Mead C., <u>Analog</u> <u>VLSI</u> <u>and</u> <u>Neural</u> <u>Systems</u>, Addison - Wesley, 1989.

[MiT 86] Mingolla E. and Todd J.T., "Perception of solid shape from shading", Bilogical Cybernetics, vol 53, pp 137 - 151, 1986.

[NaH 87] Negahdaripour S. and Horn B.K.P., "Direct Passive Navigation", IEEE Trans. on PAMI, vol.9, no. 1, pp 168-176, 1987.

[NaJ 88] Narasimhamurthi N. and Jain R., "CAD based object recognition: Incorporating metric and topological information", Proc. of SPIE, vol 937, 1988.

[NaS 89] Narasimhamurthi, Srinivasan R., M. Shridhar and M. Ahmadi., "Shape determination from intensity images - A new algorithm", to appear in IEEE Trans. Circuits and Systems.

[Pen 82] Pentland A.P, "Finding the illuminant direction", Jl. of Opt. Soc. of America, vol 72, pp 448 - 455, 1982.

[Pen 84] Pentland A.P., "Local shading analysis", IEEE Trans. PAMI, vol 6, pp 170 - 187, 1984.

[Pen 86] Pentland A.P., "Shading into texture", Artificial Intelligence , vol 29, pp 147 - 170, 1986.

[PoP 84] Poggio G.F and Poggio T., "The anaysis of stereopsis", Ann. Rev. Neurosciences, vol 7, pp 379 - 412, 1984.

[Pra 80] Prazdny K., "Egomotion and relative depth from optical flow", Biological Cybernetics, vol 36, pp 87 - 102, 1980.

[Pra 78] Pratt W., <u>Digital</u> <u>Image</u> <u>Processing</u>, John-Wiley & sons, 1978.

[Ram 88] Ramachandran V.S., "Perceiving shape from shading", Scientific American, pp 76 - 83, August 1988.

[RoA 79] Roach J.W. and Aggarwal J.K., "Computer tracking of objects moving in space", IEEE Trans. on PAMI, vol. 2, no.6, pp 127 - 135, April 1979.

[Sch 76] Schumaker L.L., "Fitting surfaces to scattered data", Approximation Theory, ed Lorentz G., Chui C.. and Schumaker L.L., Academic press, New York, NY, 1976.

[SeL 87] Sejnowski T.J. and Lehky S.R., "Neural network models of visual processing", Short course on computational neuroscience, Society of Neuroscience, 1987.

[She 76] Shepard D., "A two dimensional interpolation function for irregularly spaced data", Proc. ACM Math. Conf., pp 517 - 524, 1976.

[SrK 87] Srinivasan R., Ramakrishnan K.R. and Sastry P.S., "A Contour Based Stereo Algorithm", Proc. of Int. Conf. on Computer Vision, London, 1987.

[Sri 87] Srinivasan R., Computational study of stereopsis: A contour based algorithm for correspondence problem, M.S. thesis, Indian Institute of Science, 1987.

[SrS 87] Srinivasan R., M. Shridhar and M. Ahmadi., "Fitting Shepard's surfaces to stereo data", Proc. SPIE Symp. on Advances in Robotics, Cambridge, MA, 1987.

[SrS 89] Srinivasan R., M. Shridhar and M. Ahmadi., "A neural net for stereo data interpolation", Proc. ICASSP, 1989.

[Ste 81] Stevens K., "The visual interpretation of surface contours", Artificial Intelligence, vol 17, pp 47 - 74, 1981.

[Ste 84] Stevens K., "On gradients and texture gradients", Jl. of

Experimental Psychology, vol.113, no.2, pp 217-220, 1984.

[Sub 89] Subbarao M., "Intrepretation of image flow: A spatio - temporal approach", IEEE Trans. on PAMI, vol. 11, no. 3, pp 266 - 278, 1989.

[Tar 76] Tarjan R.E., "Graph theory and Gaussian elimination", in Sparse matrix computations, edr: Bunch J.R. and Rose D.J., Academic press, 1976.

[ToM 83] Todd J. and Mingolla E., "Perception of curvature and direction of illumination from patterns of shading", Jl. of Experimental Psychology, vol.9, no.4, pp 583-595, 1983.

[TsH 81] Tsai R.Y. and Huang T.S., "Estimating 3-D motion parameters of a rigid planar patch", IEEE Trans. ASSP, vol.29, no.6, pp 1147-1152, Dec. 1981.

[Ull 80] Ullman S. "The interpretation of structure from motion", Proc. Roy. Soc. Lond., vol B 203, pp 405 - 426, 1980.

[Ull 79] Ullman S., Intrepretation of Visual Motion, MIT press, 1979.

[VeP 89] Verri A. and Poggio T., "Motion field and optical flow: qualitative properties", IEEE Trans. on PAMI, vol. 11, no. 5, pp 490-498, 1989.

[WeH 89] Weng J., Huang T.S. and Ahuja N, "Motion and structure from two perspective views: algorithms, error analysis and error estimation", IEEE Trans. on PAMI, vol.11, no.5, pp 451-476, 1989.

[Wit 81] Witkin A. "Recovering surface and shape orientation from texture", Artificial Intelligence, vol 17, pp 17 - 47, 1981.

[Yor 81] York B.W. Shape representation in computer vision, Ph.D

dissertation, Univ. of Massachusettes, Amherst, 1981.

[Zuc 86] Zucker S.W., "Early vision", Computer vision and Robotics Lab, TR 86 - 5R, McGill Univ., 1986.

Vita Auctoris

1962   Born on March 21, Madurai, Tamilnadu, India

1979   Completed secondary education

1984   Graduated from Annamalai University, India, with

            B.E. in Electrical Engineering

1986   Graduated from Indian Institute of Science, India, with

            M.S. in Electrical Engineering

1986 -   Candidate for the degree of Doctor of Philosophy in

            Electrical Engineering, Windsor, Ontario, Canada.