Cooperative Shape From Shading and Stereo for 3D Reconstruction

COOPERATIVE SHAPE FROM SHADING AND STEREO FOR 3D RECONSTRUCTION

By

JEFF FORTUNA, B.Eng

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Engineering

McMaster University

April 2001

MASTER OF ENGINEERING (2001) MCMASTER UNIVERSITY

(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: Cooperative Shape From Shading and Stereo for 3D Reconstruction

AUTHOR: Jeff Fortuna
B.Eng

SUPERVISOR: David Capson

NUMBER OF PAGES: x, 75

Abstract

This thesis presents a survey of techniques to obtain the depth component from two-dimensional (2D) images. Two common techniques - stereo and shape from shading are examined here. Their performance is compared with an emphasis on noting the fundamental limitations of each technique. An argument is presented which suggests an adjustment of the paradigm with which stereo and shape from shading have been treated in three-dimensional vision. The theoretical development of the stereo and lighting models is followed by experiments illustrating use of these models for a variety of objects in a scene. A comparison of the results provides a motivation for combining them in a particular way. This combination is developed, and its application is examined. Using the model that is consistent for both shape and lighting, significant improvement over either stereo or lighting models alone is shown.

Acknowledgements

I would like to thank Dr. David Capson, my supervisor, for his support and guidance with this research. I also thank my fellow researchers in the lab and everyone else who had patience with me while I completed this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Basic Techniques

Stereo is one of the fundamental techniques for determining depth in a scene from 2D image representations. Our own visual system uses stereopsis extensively in determining a depth representation of a scene [20]. The central issue in determining the depth of a point in 3D space from a pair of images is the correspondence problem. In order to calculate the depth of the point by triangulation from the image pair, the same point must be identified in both images. Typically, block correlations are computed across images to find these correspondences [13]. To narrow down the search regions, the geometry of the camera setup is exploited. The disparity (distance between matching points) directly determines the depth value at this point. Unfortunately because of the discrete nature of the pixels in digitized images, for typical stereo arrangements, generally only a few levels of depth can be calculated accurately in a scene. A further problem arises when the object contains featureless surfaces. The correlation operation is destined to produce an incorrect match due to a lack of features to correlate. These issues will be examined in some detail.

The geometry of the relationship of stereo camera images is called the epipolar

geometry. An absolute depth measurement can be obtained from a calibrated camera setup with knowledge of the Essential matrix [16]. In the uncalibrated case, the projective geometry of the setup produces the Fundamental matrix [17]. These matrices provide the relationship between a point on one of the images, and its relative position on the other. The remainder of the discussion of stereo will consider the uncalibrated case. It is possible to determine the epipolar geometry with 7 point matches between the stereo image pairs. Generally, more matches are used. There are a multitude of issues involved in the robust calculation of the Fundamental matrix characterizing the epipolar geometry [35]. Without a robust determination of the epipolar geometry, the correspondence calculation for the depth map becomes intractable.

Another fundamental technique for calculating a depth map is that of using shape from shading. Once again, there is a biological motivation for using this idea [23]. Shape from shading algorithms involve computing the surface normal of an object in accordance with a reflectance function for the surface. Most commonly, Lambertian surfaces are considered. A Lambertian surface is one where a differential radiance falling on the surface is reflected at a value proportional to the cosine of the angle between the normal of the surface and the illumination direction. Shape is recovered from the surface by globally minimizing a cost function on a constraint such as smoothness, or by locally applying constraints on properties such as surface shape, smoothness and brightness.

The Lambertian model of reflectance is a very simple one and very few surfaces exhibit purely Lambertian reflectance. Most surfaces exhibit some degree of specular reflection. Shadows and radiation interchange between objects in the scene further complicate the picture. In addition to this, the lighting scenario in a scene is typically much more complex than a single light where all beams arrive in parallel from one direction as is assumed by the Lambertian reflectance model.

The shape from shading technique produces a smoothly varying depth map with

a local gradient that can be determined to a resolution dependent on the number of brightness quantization levels. Since the overall depth is a result of the integration of the local gradients, the number of overall depth levels represented is very high. This is in sharp contrast to the stereo correspondence method. Unfortunately, the shape from shading models generally consider constant albedo surfaces. For a great many objects, this assumption is untrue. On the other hand, the stereo correspondence method is albedo independent. The opposing nature of these considerations will be illustrated.

Given the critical problems cited with stereo and shape from shading, if success is to be made in building three-dimensional worlds from vision, a more complete and consistent computational theory utilizing both stereo and shading in a complementary way is required. In this thesis, experiments will be conducted with the intent of being illustrative of the fundamental flaws in the traditional stereo and shading lines of thought. A method will be devised to illustrate the co-operation of lighting and stereo and the technique will be examined with some examples that would produce poor results with shape from shading or stereo alone.

1.2 Literature Survey

1.2.1 Calibration and Epipolar Geometry

The landmark paper for a discussion of the determination of the intrinsic and extrinsic parameters of a camera is [31]. In this paper, a review of previous techniques is given, and a new technique is presented. Breaking the problem into a two stage process, a radial alignment constraint is used to solve for extrinsic parameters using linear equations.

In [1], the process of rectification is described in detail. Many rectification techniques exist. One such was described by [6]. From extrinsic and intrinsic parameters and the camera projection matrices, it was shown with respect to these matrices, how a rectification projection matrix can be determined for each camera.

Determining the geometry of a stereo pair from an uncalibrated setup was examined in great detail by [17]. The fundamental matrix, and its geometric interpretation, is examined thoroughly. There is a comparison of results from using linear and non-linear methods to estimate this matrix. In the process, the numerical issues in this computation are highlighted. To resolve some of the issues raised, [35] illustrates a robust technique for calculating the fundamental matrix. Additionally, [35] is a review of the uncertainty in computing the epipolar geometry, serving as a summary of current methods for fundamental matrix estimation.

1.2.2 Depth From Stereo

The idea of using random dot stereograms as tests for researching human stereopsis was introduced in [12]. From these stereograms, it was proven that human vision does not require high-level object descriptions such as surfaces or shapes, to determine depth from binocular views. Thus, simple point by point matching is a very important component to stereo vision. Mayhew and Frisby's paper [20] extends this idea to show that humans relate this point by point disparity information to the construction of a 'binocular raw primal sketch' (BRPS). The raw primal sketch is defined as using intensity change information in the image to recognize primitive features, such as edge segments, terminations, or blobs. These primitives are then combined into larger tokens to produce a 'full primal sketch'. A biological basis for this idea can be found in [11]. The BRPS assertions are tested with computational experiments in Mayhew and Frisby's work. They also published a collection of papers describing the process of building 3D models from stereopsis [21] in which an important correspondence

method, the PMF (Pollard, Mayhew, and Frisby) algorithm, was highlighted [27]. David Marr [18] explored this idea of the binocular raw primal sketch in his search for a consistent computational theory of human vision. Both stereo and shading were considered from the viewpoint of finding some unification of vision theories.

The first paper on using dynamic programming to reconstruct a surface from a stereo image pair was [7] Some of the earliest work on determining depth from stereo computer images was illustrated in [15]. Starting with [2], a great number of papers have been published on both area and feature based stereo. A central issue to all of this work is the correspondence problem. Since this problem is essentially one of pattern recognition, the standard techniques from this discipline have been applied to stereo vision. Correspondences can be computed by correlations using intensity values, spectral values, or a variety of other features, such as histograms [29]. Another important work was [22], describing stereo matching with an adaptive window. This paper details many of the issues of windowed correlation based correspondence.

1.2.3 Shape From Shading

Recovering shape from the intensity of pixels in an image requires knowledge of the interaction between surface materials, light sources, and surface orientation. The early reflection models of the late 60's came from the computer graphics community. [34] introduced a simple model that shaded points on a surface with an intensity inversely proportional to the distance of the point from the light source. [9] used a propagation technique to calculate the depth along a characteristic strip. These strips propagate outward from singular points where the initial surface curve is approximated. The most commonly used reflection model used in computer graphics to date is described in [24]. This model takes into account both the diffuse and the specular reflections of a surface. In the graphics context, the intensity of a facet vertex is calculated with this model, given the surface normal, the lighting direction and the surface's

reflection properties. With the introduction of the reflectance map in [10] the diffuse part of Phong's reflection model is locally used to derive the reflectance of an object for each pixel intensity in an image. [23] and [30] used a linear approximation of the reflectance function to find the depth of an object at each point .

1.2.4 Combined Stereo And Shading

From [3], many attempts have been made to combine stereo and lighting. Most of the methods use simplified lighting models with the assumption of Lambertian surfaces and constant albedo. This places a severe restriction on the number of applications of these methods. A case in point is [5], combining stereo and shading by considering stereo to provide low frequency spatial information and shading to provide high frequencies. Additionally, much work has been done using stereo to provide the initial and boundary conditions for the shape from shading model. More recently, [14] describes a more complex environment using Phong shading but requires photometric characteristics. Photometric shape from shading eliminates the albedo dependence from the lighting model but requires views of the object under different lighting conditions.

1.3 Outline Of Thesis

In Chapter 2, the theoretical background of camera geometry, stereo vision, and stereo viewing geometry will be provided. Chapter 3 will provide a detailed discussion of lighting models and their application to determining shape from shading. Chapter 4 provides a framework for combining the stereo and lighting theory which can be applied consistently to determine the three-dimensional structure of a scene. In Chapter 5, stereo, shape from shading and combined stereo and lighting are considered in turn, with experiments to show the issues involved in each technique. Chapter 6 provides

a summary of the main results from Chapter 5, and describes possible improvements to the combined stereo and lighting models.

Chapter 2

Stereo Vision

2.1 Introduction

Using two cameras (commonly denoted a stereo pair) for computing the depth of a scene is a process called *triangulation.* Given a point in one of the camera's images and the corresponding point in the other image, the depth of the point in the scene can be calculated from the distance (or disparity) between the two points. The geometry of such a setup is shown in Figure 2.1

The three-dimensional structure of the scene is projected onto the two cameras giving each one a slightly different 2-D view. In general then, finding a corresponding point from one image to the next is a 2-D search problem. There is a geometric constraint called the *epipolar geometry*, that can be exploited to reduce the 2-D search to a 1-D search. If this geometry is known, a process of *rectification* can be employed, whereby the images are warped so that the corresponding points lie on a common line (typically horizontal) in both images.

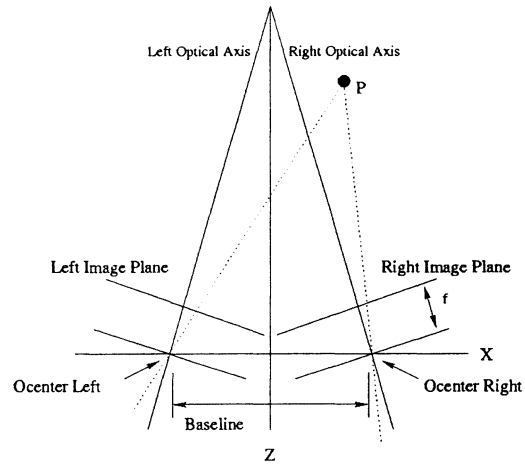

Figure 2.1: Stereo Geometry

2.2 Camera Model

Using a pinhole model for a camera, each is represented by its optical center **C** and its image plane. A point **M** in 3-D space is projected onto the camera's image plane through a projection defined by the linear transformation **P** in projective (homogeneous) co-ordinates.

$$\mathbf{m} = \mathbf{PM} \tag{2.1}$$

where

$$\mathbf{m} = \begin{bmatrix} u \\ v \\ s \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

The matrix **P** in 2.1 can be decomposed into the product $\mathbf{P} = \mathbf{A}(\mathbf{I} \mid \mathbf{0})\mathbf{D}$ where

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{2.2}$$

and

$$\mathbf{A} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.3}$$

D is the extrinsic camera parameter matrix, which describes the rotation **R** and translation **t** of the camera with respect to the world co-ordinate frame. **A** is the intrinsic camera parameter matrix which describes the transformation of the scene image onto the camera's image plane. α_u and α_v are the focal lengths in horizontal and vertical pixels respectively. u_0 and v_o are the co-ordinates of the *principal point*, which is the point on the image plane intersected by the optical axis.

2.3 Epipolar Geometry

The general relationship between two cameras observing a common point in a scene is shown in Figure 2.2.

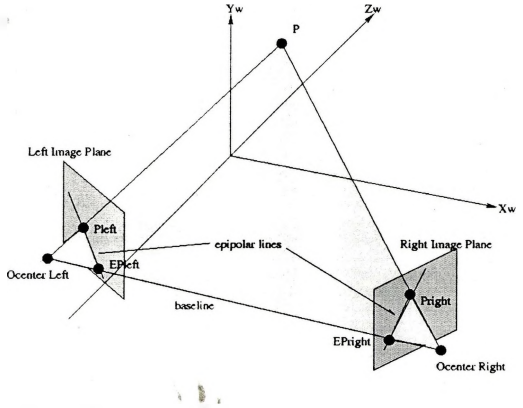

Figure 2.2: Epipolar Geometry

The optical centers of the left and right cameras are denoted C and C' respectively. The optical center of the left camera projects to a point e' on the right image plane. This point is called the right epipole. Similarly, the optical center of the right camera projects to a point e on the left image plane, and this point is the left epipole. The scene point M projects onto points m and m' on the left and right image planes respectively. The plane containing the optical centers and the scene point M cuts the left and right image planes at the lines l.

For cameras with intrinsic matrices $\mathbf{A}$ and $\mathbf{A}'$, the rotation between the two image planes is described by the rotation matrix $\mathbf{R}$ and the planes are translated by a vector $\mathbf{t}$. The epipolar constraint arises from the fact that m, m', C and C' lie in the same plane. This leads to the equation

$$\mathbf{m}^T\mathbf{F}\mathbf{m}' = 0 \tag{2.4}$$

with

$$\mathbf{F} = \mathbf{A}^{-T}\mathbf{t}x\mathbf{R}\mathbf{A}'^{-1}$$

see [8] for a derivation of 2.4. The matrix $\mathbf{F}$ is called the *Fundamental Matrix*. This matrix provides a linear transformation between a point $\mathbf{m}$ on one image and a line (called the *epipolar line*) in the other image, since

$$\mathbf{l}'_{\mathbf{m}} = \mathbf{F}\mathbf{m} \tag{2.5}$$

Thus, once this matrix is known, only the line $\mathbf{l}'_{\mathbf{m}}$ must be searched to find a correspondence across images.

2.4 Estimating the Fundamental Matrix

In an uncalibrated system, the Fundamental Matrix $\mathbf{F}$ can be estimated from point correspondences. If we have 8 or more point matches, we can use least-squares to solve

$$\min_{\mathbf{F}} \sum (\mathbf{m}_i^T \mathbf{F} \mathbf{m}_i')^2 \tag{2.6}$$

If we let the u and v co-ordinates of each point correspondence

$$\mathbf{u} = [u_1 u_2, v_1 u_2, u_2, u_1 v_2, v_1 v_2, v_2, u_1, v_1, 1]^T$$

and

$$\mathbf{f} = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T$$

we can write the minimization as

$$\min_{\mathbf{f}} \|\mathbf{U}\mathbf{f}\|^2$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

To transform the problem into one of classical minimization, we impose a constraint on the norm of $\mathbf{f}$ giving

$$\min_{\mathbf{f}} \|\mathbf{U}\mathbf{f}\|]^2 \quad \text{subject to} \quad \|\mathbf{f}\| = \mathbf{1} \tag{2.7}$$

The solution is the eigenvector of $\mathbf{U^T U}$ associated with the smallest eigenvalue. Unfortunately, this method is very sensitive to noise even if the size of data set is large. This is for a couple of reasons. First, it has been shown in [17] and [35] that the matrix $\mathbf{F}$ depends on 7 parameters only. Thus, the rank of F is only 2 and this criterion is not satisfied with the linear approach. Secondly, the minimization is not

normalized, the result being a bias in the location of the epipoles. Further, if there are any false matches or points that are incorrectly localized, the final estimate of **F** will be completely incorrect. The SVD can be used to reduce the rank of the estimate of **F** to two and there exist robust techniques for removing data outliers. The details of these methods are described in [35] and will not be repeated here.

2.5 Rectification

If the intrinsic and extrinsic parameters of the camera setup can be determined directly with a calibration technique as described in [31] or through physical measurement, a transformation of each image plane can be found that makes the epipolar lines parallel to one of the image axes. This process is called rectification. A detailed discussion of this is given in [6]. The process of image rectification is shown in Figure 2.3

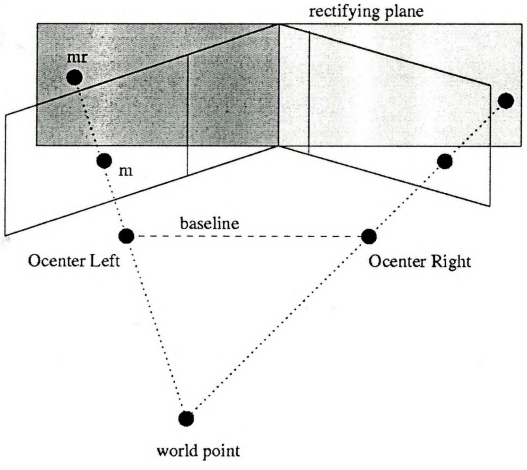

Figure 2.3: Rectification

If we devise a rectification that does not alter the optical center of the camera, we can find some rectification projection matrix $\mathbf{P}_r$ so that

$$\mathbf{m}_r = \mathbf{P}_r \mathbf{P}^{-1} \mathbf{m} \tag{2.8}$$

for each camera, where $\mathbf{m}$ is the original image point and $\mathbf{m}_r$ is the rectified image point. We can write each camera's rectification projection matrix as

$$\mathbf{P}_{r1} = \left[\begin{array}{c|c} \mathbf{a}_1^T & a_{14} \\ \mathbf{a}_2^T & a_{24} \\ \mathbf{a}_3^T & a_{34} \end{array}\right] \quad \mathbf{P}_{r2} = \left[\begin{array}{c|c} \mathbf{b}_1^T & b_{14} \\ \mathbf{b}_2^T & b_{24} \\ \mathbf{b}_3^T & b_{34} \end{array}\right] \tag{2.9}$$

The planes $\mathbf{a}_3\mathbf{w} + a_34$ and $\mathbf{b}_3\mathbf{w} + b_34$ represent the focal planes of the rectification projection matrices. If we subject those matrices to the following constraints, we can obtain a unique rectifying projection matrix for each image.

Common focal plane:

$$\mathbf{a}_3 = \mathbf{b}_3 \quad \text{and} \quad \mathbf{a}_{34} = \mathbf{b}_{34} \tag{2.10}$$

Optical centers equal to those of the original projections:

$$\mathbf{P}_{r1} \begin{bmatrix} \mathbf{c}_1 \\ 1 \end{bmatrix} = 0 \qquad \mathbf{P}_{r2} \begin{bmatrix} \mathbf{c}_2 \\ 1 \end{bmatrix} = 0 \tag{2.11}$$

Vertical coordinate of a 3D point projected onto the rectifying image plane must be equal in both images:

$$\mathbf{a}_2 = \mathbf{b}_2 \quad \text{and} \quad \mathbf{a}_{24} = \mathbf{b}_{24} \tag{2.12}$$

Choose rectifying focal planes to be parallel to the intersection of the two original focal planes:

$$\mathbf{a}_3^T (\mathbf{f}_1 \wedge \mathbf{f}_2) = 0 \tag{2.13}$$

where $\mathbf{f}_1$ and $\mathbf{f}_2$ are the third rows of the original camera one and camera two projection matrices respectively.

The rectified image planes should be orthogonal:

$$\mathbf{a}_1^T \mathbf{a}_2 = 0 \qquad \mathbf{b}_1^T \mathbf{b}_2 = 0 \tag{2.14}$$

Setting the two principal points to (0,0):

$$\mathbf{a}_1^T \mathbf{a}_3 = 0 \quad \mathbf{a}_2^T \mathbf{a}_3 = 0 \quad \mathbf{b}_1^T \mathbf{a}_3 = 0 \tag{2.15}$$

The horizontal and vertical focal lengths occur at the intersections of the focal plane and the horizontal and vertical planes of the axes of the retinal coordinates:

$$\alpha_u = \|\mathbf{a}_1 \wedge \mathbf{a}_3\| \qquad \alpha_u = \|\mathbf{a}_2 \wedge \mathbf{a}_3\| \tag{2.16}$$

from this we get

$$\begin{aligned} \|\mathbf{a}_1\|^2 \|\mathbf{a}_3\|^2 &= \alpha_u^2 \\ \|\mathbf{a}_2\|^2 \|\mathbf{a}_3\|^2 &= \alpha_v^2 \\ \|\mathbf{b}_1\|^2 \|\mathbf{a}_3\|^2 &= \alpha_u^2 \end{aligned} \tag{2.17}$$

Defining a scale factor:

$$\|\mathbf{a}_3\| = 1 \qquad \|\mathbf{b}_3\| = 1 \tag{2.18}$$

Combining all of this, we get a set of four systems of equations, which can be solved using the SVD for each system in turn.

2.6 Correspondence and Depth

Finding corresponding points across a stereo pair of images for each point in the images produces a map of the depth of each pixel in the scene. For each point **I**(m,n) in one of the images, an $M \times N$ window around this point is compared with a similar sized window surrounding each candidate point in the other image. The point with the minimum mean square error between windows is considered to be the match. The mean square error function, in general, is

$$\mathbf{MSE}(m,n) = \sum_{i=m}^{m+M-1} \sum_{j=n}^{n+N-1} (\mathbf{I}_2(i,j) - \mathbf{W}(i-m, j-n))^2 \tag{2.19}$$

where $\mathbf{I}_2$ is the second image, $\mathbf{W}$ is an $M \times N$ sub-image of $\mathbf{I}_1$ centered at (m,n). This MSE function is computed for *each* pixel in $\mathbf{I}_1$. The distance between the point in $\mathbf{I}_1$ and the point with the minimum MSE in $\mathbf{I}_2$ is called the *disparity* of these corresponding points. For rectified images, or images where the epipolar geometry has been found, the candidate points lie in a line on the other image, as described above, so the MSE function becomes a function of one variable only. This greatly reduces the computational complexity of MSE. Further, there is a maximum disparity that the geometry permits. Thus, the MSE function only need be calculated for values of the variable up to the disparity limit.

Chapter 3

Shape From Shading

3.1 Introduction

In order to exactly recover the shape of an object from the intensity of light at each point on its surface, knowledge of the surface materials and light sources must be obtained. Given this knowledge, and physical laws and models of the interaction of light with surfaces, three dimensional features can be extracted from the intensity image of a scene.

Underlying the description of light interaction with surfaces are the laws governing the reflection. In general, we can describe some fundamental relationships for quantifying radiometric (measure of radiant energy) quantities. Scene (reflected) radiance, the reflectance distribution function, and the relationship between radiation exchanging patches provide the basis for the reflection models and reflectance maps used in computer vision and graphics. These relationships are simplified, or special cases are considered, to form the commonly used models.

3.2 Radiance

The radiant intensity of a surface with area A_2, given that it receives radiant power Φ in watts is defined as

$$E = \frac{d\Phi}{dA_2} \tag{3.1}$$

and is expressed in watts per square meter. The radiant intensity of the surface, into a solid angle Ω_1 is defined as

$$I = \frac{d\Phi}{d\Omega_2} \tag{3.2}$$

and is expressed in watts per steradian. The radiant power emitted by a surface, expressed with respect to the solid angular density of the radiant power is described by the relation

$$L_1 = \frac{d^2\Phi}{dA_1 cos(\Theta_1) d\Omega_1} \tag{3.3}$$

and is called the radiance L. Radiance is described in $Wm^{-2}sr^{-1}$ (watts per square meter per steradian) where Θ_1 is the angle between the surface normal and the emitted radiation direction, A_1 is the surface area, Φ is the radiant power, Ω_1 is the solid angle over which radiation is emitted.

3.3 Radiosity

Given to surface patches that are exchanging radiation, we can find a relationship to describe the radiation exchange between the patches. This relationship is called the *form factor* between the patches. The energy leaving the differential area dA_i reaching dA_j is given by

$$dE_i dA_i = I_i \frac{dA_i cos(\Theta_i) dA_j cos(\Theta_j)}{r} \tag{3.4}$$

where I_i is the reflected intensity of the ith patch. Figure 3.1 illustrates the geometry of the two differential patches. The definition of form factor between two elemental

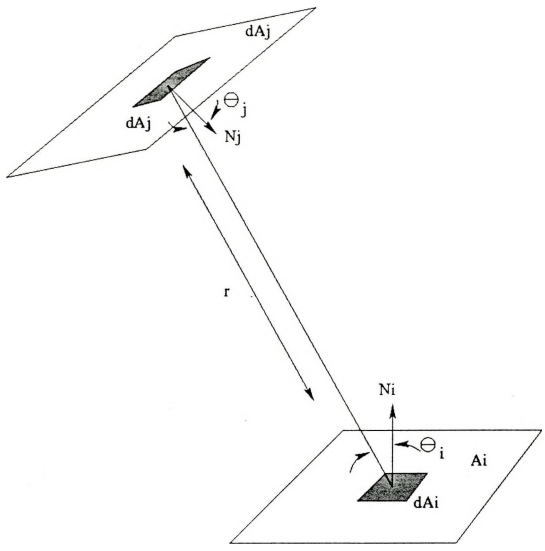

Figure 3.1: Form Factor Geometry

areas dA_i and dA_j is

$$FdA_idA_j = \frac{\text{Radiative energy reaching } dA_j \text{ from } dA_i}{\text{Total energy leaving } dA_i \text{ in all directions}} \tag{3.5}$$

If the radiant intensity is independent of direction (a Lambertian surface), the total energy, E_idA_i leaving d_A in all directions is given by integrating over the hemisphere centered on dA. Thus, $E_i = I_i\pi$ and

$$FdA_idA_j = \frac{I_idA_icos(\Theta_i)dA_jcos(\Theta_j)}{I_idA_i\pi r^2} = \frac{cos(\Theta_i)dA_jcos(\Theta_j)}{\pi r^2} \tag{3.6}$$

Integrating over A_j, and integrating over A_i (taking the area average, since the intensity I_i is independent of direction and position over A_i), we get

$$F_{ij} = \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{cos(\Theta_i)cos(\Theta_j)}{\pi r^2}dA_jdA_i \tag{3.7}$$

If we set up an equation that relates energy reflected from a patch to any emitted energy plus the incoming energy from other patches, we can state that Radiosity × area = emitted energy + reflected energy or

$$B_idA_i = E_idA_i + \rho_i\int_j B_jdA_jF_{d_AjdAi} \tag{3.8}$$

If we discretize the surfaces in the scene into n patches, we have

$$B_i A_i = E_i A_i + \rho_i \sum_{j=1}^{n} B_i F_{ij} A_j \tag{3.9}$$

Since the energy exchange depends only on the relative geometry of the patches, $F_{ij} = F_{ji}\frac{A_j}{A_i}$ we get the basic radiosity relationship

$$B_i = E_i + \rho_i \sum_{j=1}^{n} B_j F_{ij} \tag{3.10}$$

This equation exists for EACH patch, so we can set up n simultaneous equations in n unknown B_i values.

3.4 Reflectance Distribution Function

The bidirectional reflectance distribution function (BDRF) is used to quantitatively describe the reflection characteristics of a surface. The BDRF describes how bright a differential surface dA appears when it is illuminated and viewed from general directions. By definition, the BDRF is the ratio of the reflected differential radiance dL_1 in the direction of the viewer, and the differential irradiance dE_2 in the direction of the illumination

$$f_r(\Theta_2, \phi_2; \Theta_1, \phi_1) = \frac{dL_1(\Theta_2, \phi_2; \Theta_1, \phi_1; E_2)}{dE_2(\Theta_2, \phi_2)} \tag{3.11}$$

where f_r is expressed in sr^{-1}. The geometry of the BDRF is shown in figure 3.2

3.5 Lambertian Reflectance

If the reflecting surface is perfectly diffuse, that is it appears equally bright when observed from all directions and no absorption nor transmission of radiation takes place, then it is considered to be a Lambertian reflector. The properties of this type

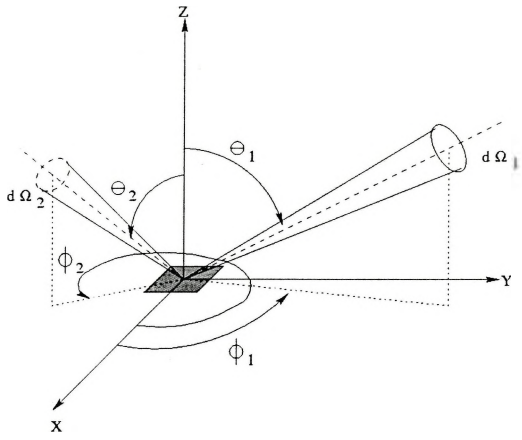

Figure 3.2: Geometry of the bidirectional reflectance distribution function

of reflectance are summarized below.

1. The reflected radiance does not depend on direction and is constant, i.e. $L_1(\Theta_1, \phi_1) = const.$
2. The BDRF is constant, i.e. $f_r(\Theta_2, \phi_2, \Theta_1, \phi_1) = f_r = const.$
3. The radiant emittance M is equal to the irradiance E_2, i.e. $M = E_2$.

Expressing the radiant emittance M as an integral of the reflected radiance L_1 over the visible hemisphere

$$\int_{\Omega_1} L_1 d\Omega_1 = L_1 \pi = E_2 \tag{3.12}$$

we get

$$f_r = \frac{L_1}{E_2} = \frac{1}{\pi} \tag{3.13}$$

The reflected radiance from a Lambertian reflector can be described as a differential radiance proportional to the cosine of the angle between the normal at the illumination point and the illumination direction. Assuming that the reflectance function is as described in 3.13, we get

$$L_1 = \frac{E_0}{\pi} \cdot \rho \cdot cos(\Theta_2) \tag{3.14}$$

where ρ (*albedo*) is a term describing the overall fraction of radiation which is reflected from the surface, given that some fraction of the incoming radiation is absorbed by the surface. If we omit the constant scale factor and describe the reflection geometry by a surface normal n and the illumination direction s we can write a simple expression for the reflectance map of a Lambertian reflector

$$R_n(\mathbf{n}) = E_0 \cdot \rho \cdot cos\angle(\mathbf{n}, \mathbf{s}) \tag{3.15}$$

This equation is often written as

$$R_n(\mathbf{n}^\circ) = E_0 \cdot \rho \cdot \mathbf{n}^{\circ T}\mathbf{s}^\circ \tag{3.16}$$

with $\mathbf{n}^\circ$ and $\mathbf{s}^\circ$ as normalized vectors

3.6 Perfect Specular Reflection

If we assume that a surface reflects light in a single direction if the incident radiation is confined to a single direction, we can define a reflection model similar to that of a Lambertian reflector. The most commonly used model for this type of reflection was developed by Phong [24]. Given the geometric relationship shown in Figure 3.3, he described specular reflection as

$$R(\mathbf{n}^\circ, \mathbf{h}^\circ) = k \cdot (\mathbf{n}^{\circ T}\mathbf{h}^\circ)^n \tag{3.17}$$

where $\mathbf{n}^\circ$ is the normalized surface normal, $\mathbf{s}^\circ$ is the normalized illumination direction, $\mathbf{v}^\circ$ is the normalized viewer direction $\mathbf{h}^\circ$ is the normalized halfway vector and n is an exponent representing the smoothness of the surface (larger n for smoother surfaces).

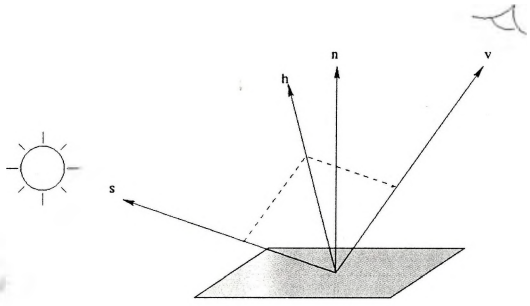

Figure 3.3: Geometry of Reflection

Chapter 4

Combined Shape From Shading And Stereo

4.1 Introduction

The objective of the combination of stereo derived shape meshes and lighting models is to create a consistent model for integrating stereo and lighting information. In this way, stereo and lighting cooperate to produce a single depth map which was derived from a lighting model which takes into account the object's overall shape. Conversely, the stereo model could be modified to be consistent with the scene lighting conditions.

The application of stereo algorithms to a pair of images produces a cloud of points in 3D space. Shading algorithms generally are applied to the normals of surface patches. Thus, to use stereo in the context of lighting, we need to convert a point cloud representation to surface patches. In general, this is not easy to do. For the case where objects can be easily separated from the background and can be segmented from other objects, the problem of determining a surface mesh is simplified. For the purposes of this work, all objects are considered to be seperable from the background and only one object will be examined in each scene.

For each patch on the mesh, we can define a surface normal and albedo, as shown in Figure 4.1. This allows for the computation of the reflectance function of the mesh under the same lighting conditions that the image was taken. If the stereo mesh is rendered under lighting that assumes constant albedo, any differences from the measured intensity provides an estimate of the surface albedo.

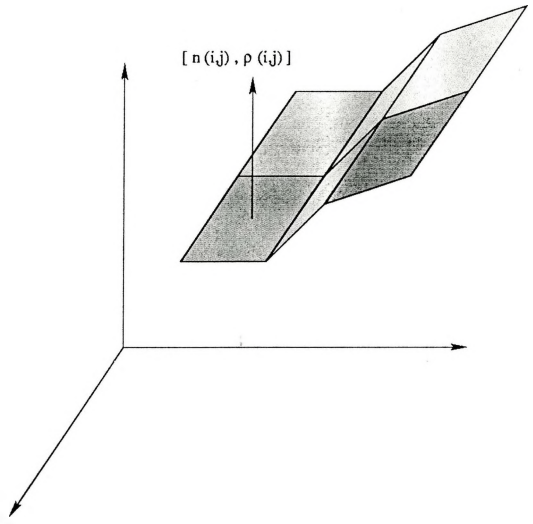

Figure 4.1: Lighting The Stereo Mesh

Once we have an albedo estimate, we can proceed to calculate the surface shape from one of many shape from lighting techniques. These methods essentially solve for the surface normal given the surface gradient and integrate the normals across the surface to find the depth. Regardless of the technique employed, the lighting model employed to describe the reflectance given the surface normal provides the constraints for the surface shape. For the combined stereo and lighting algorithm described herein, a Lambertian lighting model is assumed.

4.2 A Stereo - Derived Mesh

Given a cloud of points in three space, the first task in fitting a surface to these points is to find interpolating B-spline curves in both the x and y direction. B-spline curve fitting is a common practice, and the theoretical development will not be included here. If we consider a single direction, x, we can write

$$x(u_p) = \sum_{i=0}^{m} \mathbf{P_{xi}} B_i(u_p) = D_{xp} \tag{4.1}$$

where D_{xp} is the x component of the data point. This defines a system of m-1 equations in m+1 unknowns. Adding two more control points, we can define m+1 equations in m+1 unknowns:

$$\begin{bmatrix} B_0(u_2) & \cdots & B_m(u_2) \\ \vdots & & \vdots \\ B_0(u_2) & \cdots & B_m(u_2) \end{bmatrix} \begin{bmatrix} \mathbf{P}_{x0} \\ \vdots \\ \mathbf{P}_{xm} \end{bmatrix} = \begin{bmatrix} D_{x2} \\ \vdots \\ D_{xm+2} \end{bmatrix} \tag{4.2}$$

On a segment by segment basis, the B-spline can be converted to Bezier curves by

$$\begin{bmatrix} \mathbf{P}_0 & \mathbf{P}_1 & \mathbf{P}_2 & \mathbf{P}_3 \end{bmatrix} = \mathbf{B}^{-1}\mathbf{B}_s \begin{bmatrix} \mathbf{Q}_0 & \mathbf{Q}_1 & \mathbf{Q}_2 & \mathbf{Q}_3 \end{bmatrix} \tag{4.3}$$

where: $\mathbf{P}$ are the Bezier control points $\mathbf{Q}$ are the B-spline control points $\mathbf{B}$ is the Bezier matrix $\mathbf{B}_s$ is the B-spline matrix

This equation is applied on a segment by segment basis for both the x and y direction. The result for multiple segment B-splines is shown in Figure 4.2

At this point the boundaries for the Bezier patches have been defined. All that remains is to define the inner control points. Several options exist for doing this, the simplest of which considers that the boundary curves of the patch should be more or

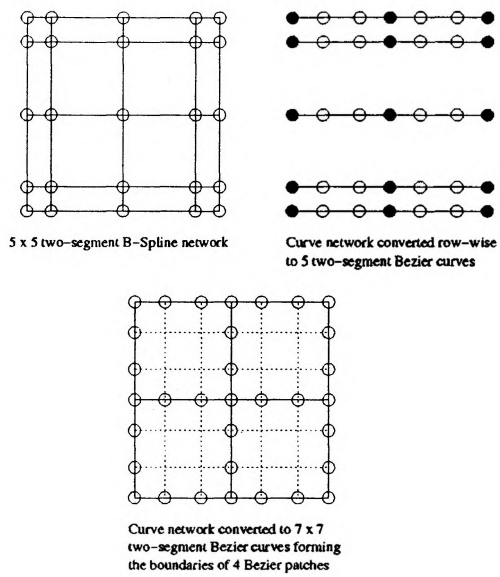

Figure 4.2: Converting B-spline Curve Network To Bezier Patches

less translations of each other. In this case, the inner points can be determined from the boundary points.

4.3 Lighting Model

For a more general class of objects where the albedo varies across the object but the surface is still lambertian, we can use a slightly more complex lighting model

$$R_n(\mathbf{n}^o(i,j)) = E_0 \cdot \rho(i,j) \cdot \mathbf{n}^{oT}(i,j)\mathbf{s}^o \tag{4.4}$$

where i and j are the discretized projected locations on the surface where the normals and the albedo are measured.

In order to apply this model, it is necessary to have an estimate of the surface albedo at each point. For the albedo estimate to be consistent with the orientation of the surface normal, the pixel intensity of the projected surface must be equal to the intensity as calculated by the reflectance function.

$$E(i,j) = R_n(\mathbf{n}^o(i,j)) = E_0 \cdot \rho(i,j) \cdot \mathbf{n}^{oT}(i,j)\mathbf{s}^o \tag{4.5}$$

4.4 Stereo - Derived Albedo Estimate

If an estimate of the overall shape of the object can be obtained by stereo, the surface normals, reflectance function and albedo can be calculated. It is extremely important to have a stereo estimate that contains little noise. Step changes in surface height due to the limited resolution of stereo depth maps would create wildly inaccurate surface normals. Additionally, incorrect depth measurements due to a lack of features in block correlation algorithms create an unsuitable estimate of the normals. Both of these issues can be resolved by using a stereo map determined by features that are determined by an edge map of the object. Depth information is only calculated at these edge points. This minimizes correlation errors due to a lack of features. Secondly, the resulting sparse depth map is fitted to a surface using the spline or bezier surface technique described above. This provides a more continuous and smooth surface.

4.4.1 Applying The Albedo Estimate

Care must be taken in applying the albedo estimate to the reflectance function model. The stereo derived normal used to calculate the albedo will not be perfectly accurate. Since this albedo term is used directly to find the surface orientation under the shape from shading algorithm, most of the surface detail provided by the shape from shading algorithm would be swamped by the albedo term. One simple and effective way to alleviate this problem is to apply a varying amount of the albedo term depending on how closely the shape from shading depth result matches with the stereo result. In this way, the stereo result becomes the ground truth, and the lighting model

is adjusted accordingly. The albedo estimate can be scaled by a correction factor $E_{correction}$ according to

$$\mathbf{E_{correction}} = \frac{(\mathbf{Z_{sfs}} - \mathbf{Z_{stereo}})}{max(\mathbf{Z_{sfs}} - \mathbf{Z_{stereo}})} \tag{4.6}$$

We can then apply the albedo estimate to the lighting model by

$$E(i,j) = E_0 \cdot ((E_{correction}\rho(i,j)) + (1 - E_{correction})) \cdot \mathbf{n^o}^T(i,j)\mathbf{s^o} \tag{4.7}$$

then

$$E(i,j) = E_0 \cdot (E_{correction}E_{Albedo}(i,j) + (1 - E_{correction})E_{surface}(i,j)) \tag{4.8}$$

The first term on the right hand side is the contribution of the albedo to the intensity, and the second term is the effect of surface tilt alone. To employ shape from shading with albedo independence, we divide out the albedo from our albedo estimate in the first term according to

$$E(i,j) = E_0 \cdot (E_{correction}\frac{E_{Albedo}(i,j)}{\rho(i,j)} + (1 - E_{correction})E_{surface}(i,j)) \tag{4.9}$$

Chapter 5

Experiment

5.1 Introduction

To exemplify all of the issues involved in obtaining three dimensional surface representations of general objects, some ideal and non-ideal objects were examined. The algorithms employed to illustrate stereo and shape from shading were fundamental, utilizing no prior knowledge of the scene. Dense stereo maps were obtained by simple block correlation, and shape from shading was obtained using the method described in [30]. Specularity was handled by modifying the Lambertian reflectance model to follow the Phong shading model.

Obtaining a consistent stereo and lighting model required the following innovations. First the stereo map was only computed at edge features, and cubic spline surface interpolation was employed to provide a continuous depth. It was critical to calculate stereo in this way. The stereo model is used as a basis for the shaded surface, and large errors in the stereo result will destroy the validity of the lighting model. Dense stereo correlation is guaranteed to produce large errors where there are no features to correlate. Edge-based stereo ensures that there are features to correlate.

The second important innovation was to compute a reflectance map from the stereo result. In this way the surface albedo can be estimated. This estimate can then be applied to the lighting model.

The third innovation involved using both the stereo and lambertian lighting depth maps to give a cue as to how much of the surface albedo estimate to apply to the lighting model.

5.2 Stereo

The process of computing an accurate stereo map is complicated by the calculation of point to point correspondences between two images which are disparate views of the same scene. Further, given a discrete CCD imaging element, the resulting stereo images are pixelized. Because of this, there is a maximum resolution to which the depth can be measured.

5.2.1 Rectification and the Fundamental Matrix

To illustrate the process of rectification and the calculation of the fundamental matrix, $\mathbf{F}$ was calculated in two ways, first from the point correlations using the SVD to solve a set of linear equations and second from the camera projection matrices calculated from the knowledge of the camera geometry. Additionally, $\mathbf{F}$ was measured from point correspondences and calculated from projection matrices for rectified images to show that the epipolar geometry has been corrected to be horizontal.

5.2.2 Unrectified F From Point Correspondences and the Camera Projection Matrix

The stereo images in Figure 5.2 show the left and right view of a block taken over a baseline of 800 mm with each camera rotated 24 degrees in the y axis, according to Figure 5.1.

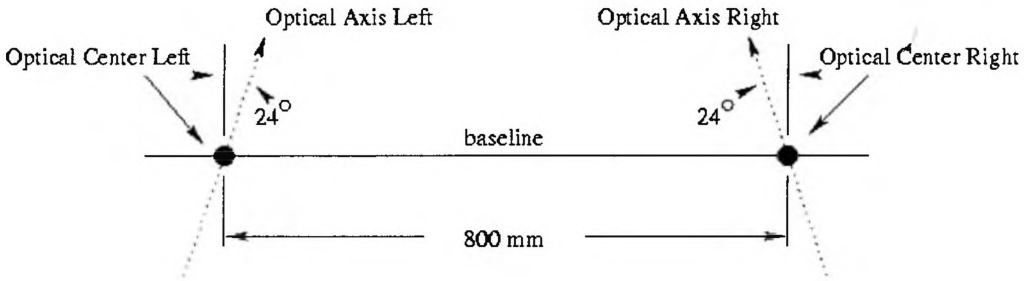

Figure 5.1: Geometry of Stereo Camera Setup

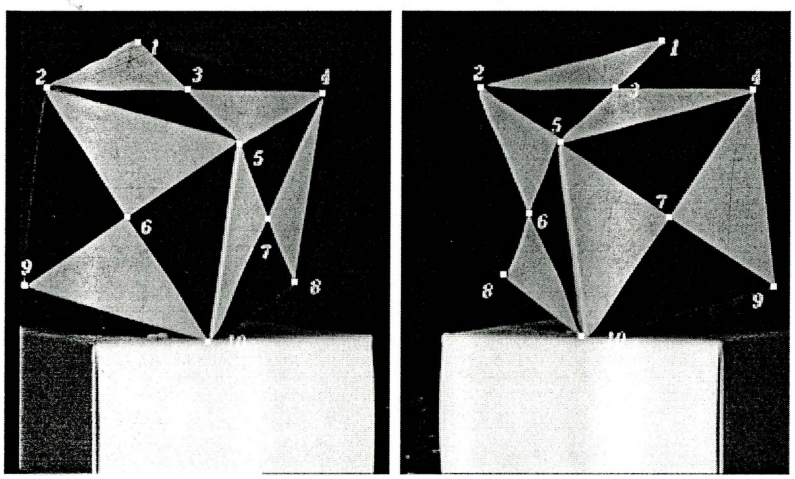

Figure 5.2: Left and Right Stereo Images of Block (800 mm Baseline)

The numbered points on each image were matched by hand, and the SVD was used to solve the linear equations formed by each match, under the constraint $\min_{\mathbf{F}} \sum(\mathbf{m}_i^T \mathbf{F} \mathbf{m}_i')^2$. The resulting epipolar lines are shown in Figure 5.3.

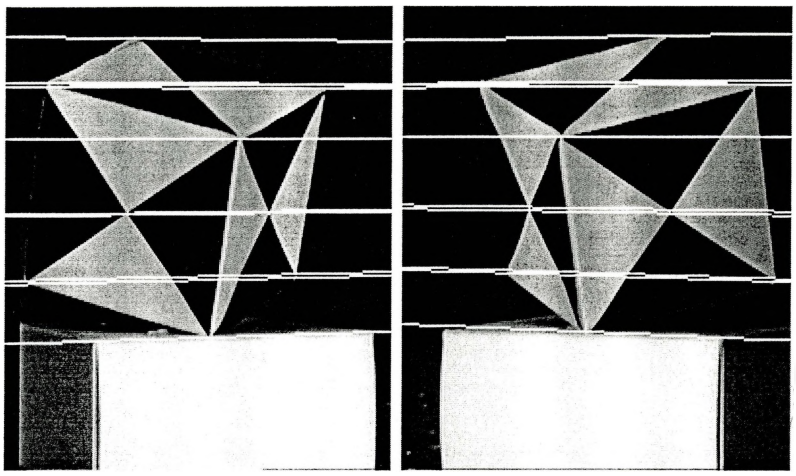

Figure 5.3: Left and Right Epipolar Lines of Block From Manual Matching

F was also calculated using the geometric constraint $\mathbf{F} = [\mathbf{P1C}]_x\mathbf{P1P2}^+$ where **P1** and **P2** are the projection matrices for the left and right cameras respectively. **P1** and **P2** were calculated using the manufacturer's specifications for the camera and the lens. For the camera and lens used, $FocalLength = 25mm$, $XPixelSize = \frac{4.8}{320}mm$, $YPixelSixe = \frac{3.6}{240}mm$, and the principal points $u_0 = 160$, $y_0 = 120$.

The resulting epipolar lines for the calculated F are shown in Figure 5.4.

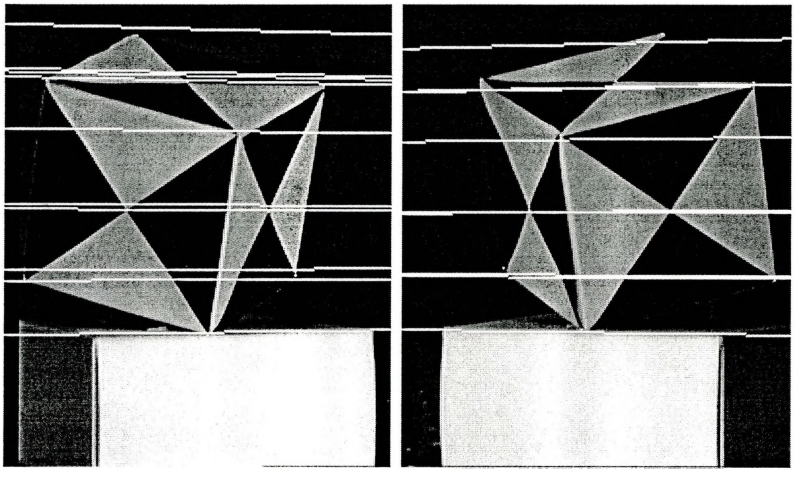

Figure 5.4: Left and Right Epipolar Lines of Block For Calculated F

5.2.3 Rectified F From Point Correspondences and the Camera Projection Matrix

When the block images were rectified according to the algorithm described by [6], the resulting images are shown in Figure 5.5.

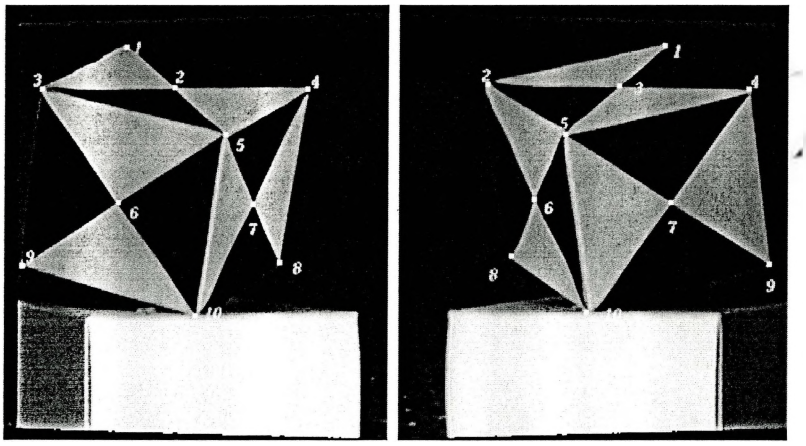

Figure 5.5: Left and Right Stereo Rectified Images of Block (800 mm Baseline)

Again the numbered points on each image were matched by hand and the SVD was used to solve the linear equations formed by each match. The resulting epipolar lines are shown in Figure 5.6.

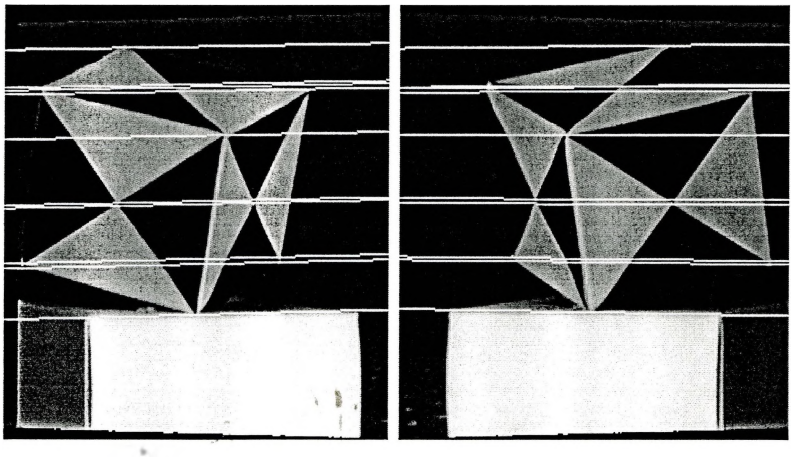

Figure 5.6: Left and Right Rectified Epipolar Lines of Block From Manual Matching

The resulting epipolar lines for the calculated F of the rectified images are shown in Figure 5.7.

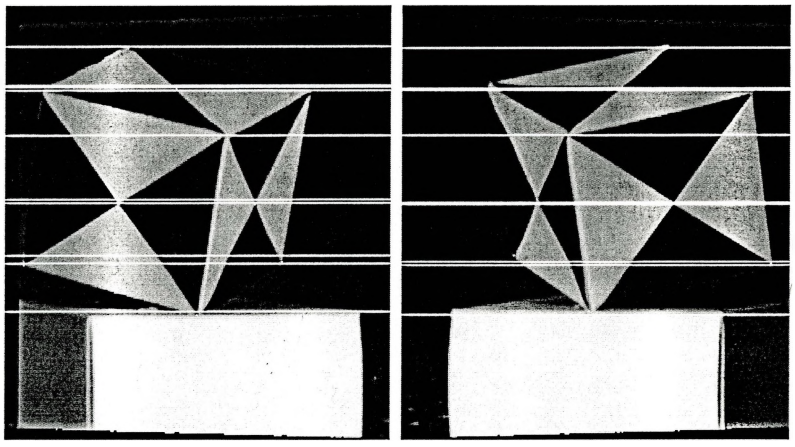

Figure 5.7: Left and Right Rectified Epipolar Lines of Block For Calculated F

When any of the points are incorrectly matched as in the example shown in Figure 5.8, the epipolar line calculation is completely wrong. Points one and six were matched together for this example.

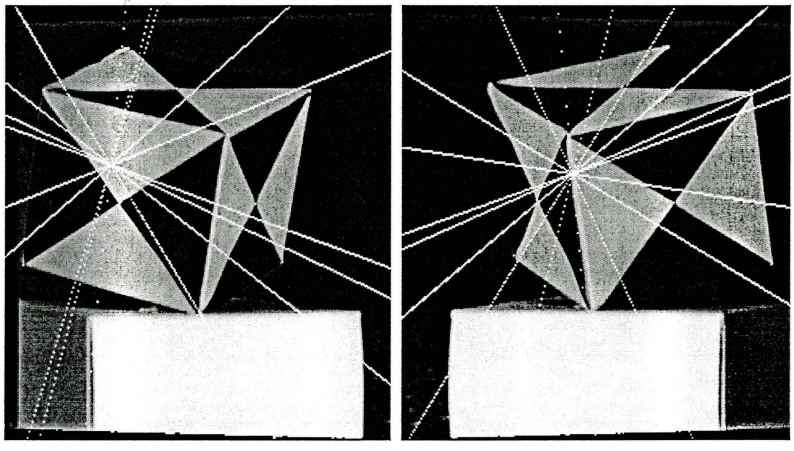

Figure 5.8: Epipolar Lines For Rectified Block Images - One Incorrect Match

5.2.4 Summary Of Rectification And Fundamental Matrix Results

The process of finding the epipolar geometry of a stereo camera setup is critical for the practicability of finding correspondences between images. Unfortunately, determining epipolar geometry is a difficult process. For all the examples given, the point matches were done manually. This process is difficult to automate with 100 percent accuracy, particularly if the views are widely disparate. Additionally it was shown that because of the sensitivity of the system of equations used to solve for F, even a single incorrect point match can produce incorrect epipolar geometry.

When the camera projection matrices are known, the calculation of epipolar lines does not require point matches. However, it now requires knowledge of the intrinsic and extrinsic parameters of the camera. The difference between the epipolar geometry calculated by point correspondences and of that calculated from the projection matrices was indeed quite small. Since all of the experiments were done with an uncalibrated setup, there is no exact epipolar geometry information available, making it impossible to quantitatively compare the results. As illustrated in our examples however, even approximate knowledge of the parameters can produce usable epipolar geometry information.

5.2.5 Theoretical Stereo Resolution of Practical Sensors

The depth resolution of discrete CCD sensors in a stereo configuration is determined by the distance between the two cameras (the baseline). In order to maintain parallel geometry (without rectification), the cameras can undergo only horizontal translation. This limits the magnitude of the baseline distance, since both cameras must still see the same object. In order to increase this distance, the cameras can be angled with respect to each other. Both cases are illustrated here, to show the depth resolution for

each. Additionally, the theoretical resolution will be calculated, given the baseline, from the camera projection equations. Since block based correlation was used, the surface to be measured was painted with a texture of random black and white dots, ensuring maximum uncorrelation from block to block.

5.2.6 Low Resolution, Short Baseline Surface Measurement

With a baseline of 80 mm, the stereo images of a can at a distance of 1 m are shown in Figure 5.9.

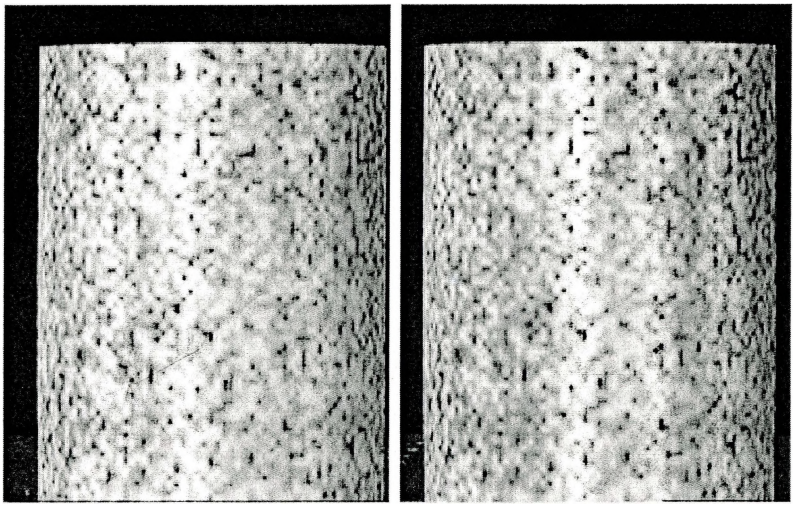

Figure 5.9: Left and Right Stereo Images of Cylinder (80 mm Baseline)

The resulting depth map obtained from these stereo images is shown in Figure 5.10

From Figure 5.10 it is seen that there is a maximum disparity of 12 at a depth of 1 m. From the geometry of the can, over a 60 pixel change in horizontal distance, (from the horizontal center of the can toward the left or right) z changes 16.9 mm for the 100 mm wide can. From the figure, there is a 3 pixel change in disparity over this change in z. Thus, the measured resolution is 16.9 mm / 3 pixels = 5.63 pixels / mm. Similar measurements were made at different baselines to obtain the measured

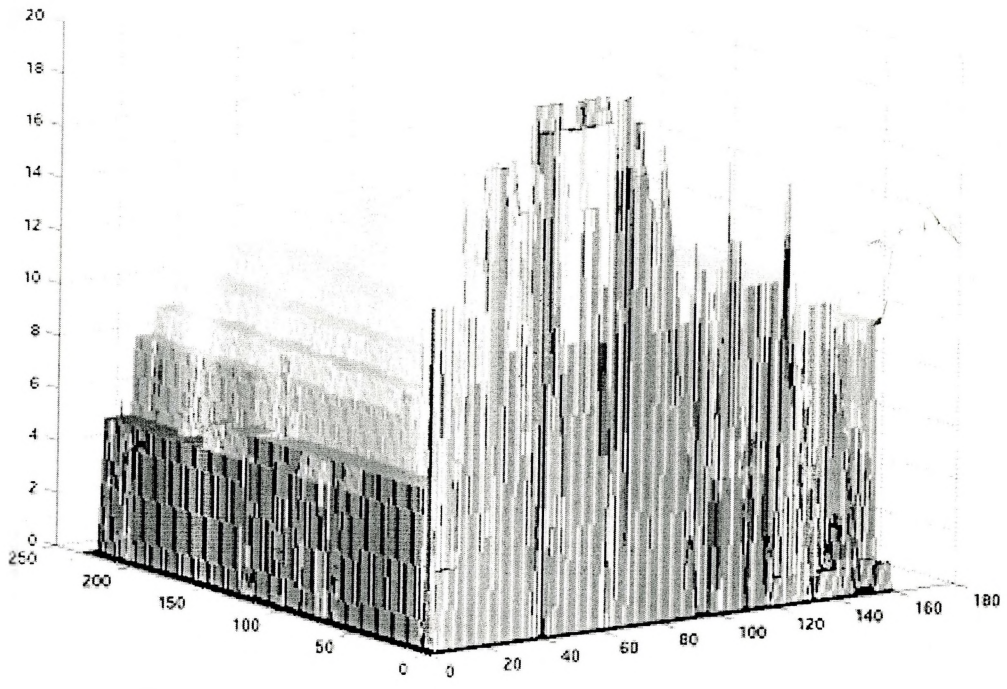

Figure 5.10: Surface Depth of Cylinder (80 mm Baseline)

resolutions in Table 5.1. The camera's CCD has a resolution of approximately 67 pixels / mm

5.2.7 High Resolution, Long Baseline Surface Measurement

With a baseline of 800 mm, the stereo images of a can at a distance of 1 m are shown in Figure 5.11.

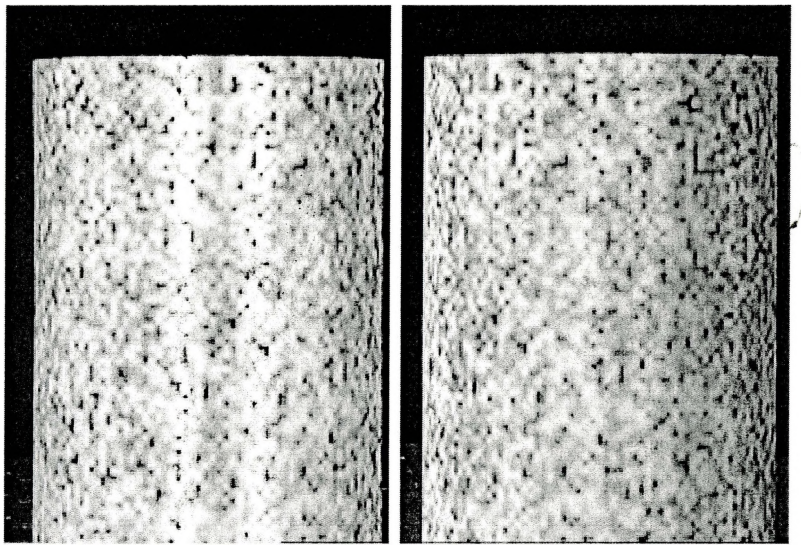

Figure 5.11: Left and Right Stereo Images of Cylinder (800 mm Baseline)

The resulting depth map obtained from these stereo images is shown in Figure 5.12.

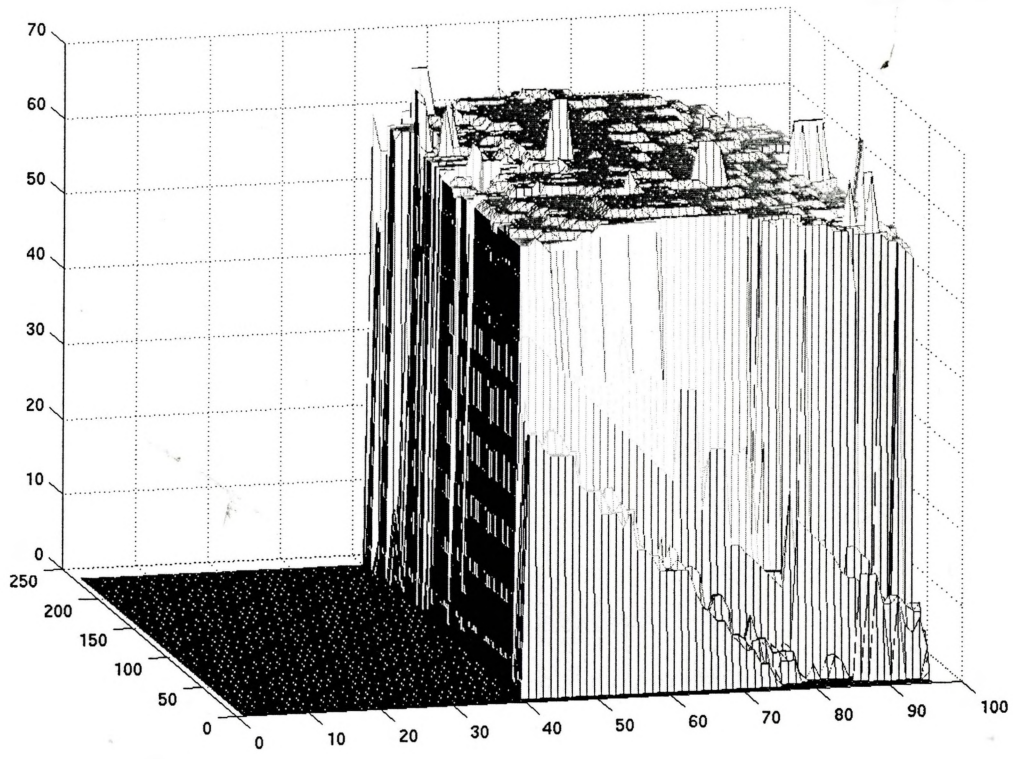

Figure 5.12: Surface Depth of Cylinder (800 mm Baseline)

5.2.8 Summary Of Theoretical Resolution Results

From theory, $d = \frac{fb}{z}$ and $\frac{\partial d}{\partial z} = -\frac{fb}{z^2}$. The theoretical and measured resolutions are shown below.

Baseline (mm)	Theoretical (mm / pixel)	Measured (mm / pixel)
40	14.92	8.43
80	7.46	5.63
200	2.98	2.41
400	1.49	2.10
600	.995	1.69
800	.746	.526

Table 5.1: Theoretical and Measured Resolution

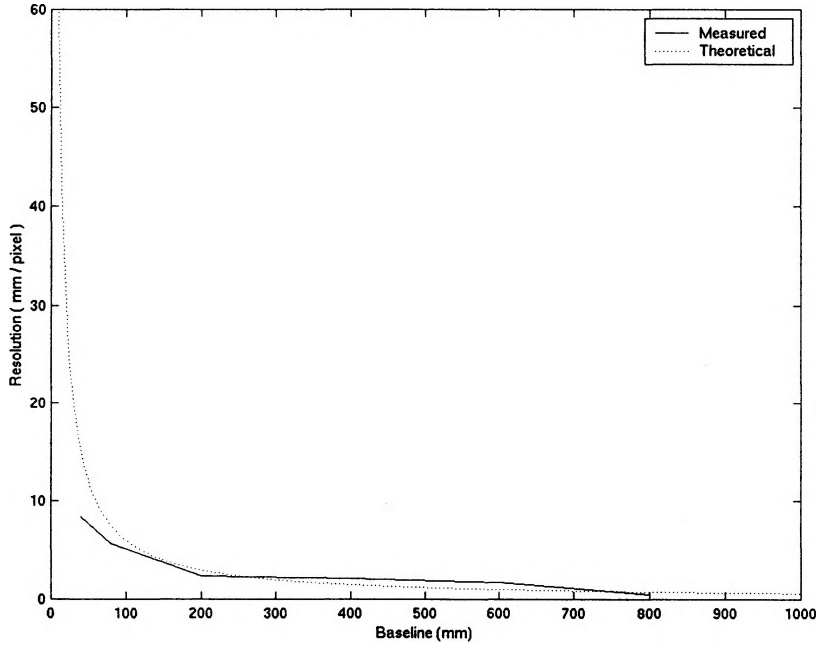

Figure 5.13: Theoretical and Measured Resolution at 1m Distance

The short baseline stereo produced a measured depth resolution (which agreed with the theoretical resolution of the sensor) of about $7\frac{mm}{pixel}$. Clearly, it is difficult

to obtain smooth surface information at this coarse of a resolution. Increasing the baseline distance by ten - fold increased the measured and theoretical resolution by a factor of about 10, to $.7\frac{mm}{pixel}$. While this resolution is quite usable for a number of applications, an 800 mm baseline produces disparate images, making the correlation process much more difficult and error prone, as seen in the numerous surface errors in the long baseline case.

5.3 Shape From Shading

Shape from shading algorithms exploit the simple fact that the surface normal can be recovered by solving differential equations involving the surface gradient in the x and y directions. The approach chosen here was a local linerization of the reflectance function $E(i,j) = R_n(\mathbf{n}^o(i,j)) = E_0 \cdot \mathbf{n}^{oT}(i,j)\mathbf{s}^o$. The equations were solved using a Jacobi iterative scheme.

5.3.1 Ideal Lambertian Reflectance

Shape calculated under the Lambertian reflectance assumption produces a smooth surface with a point by point depth resolution equal to the resolution of the gradient map. However, since the reflectance map is locally linearized, the method has no validity over discontinuities in the surface. As such, the surface must be continuous. The example given illustrates the application of Tsai and Shah's algorithm over the same surface that was used for the stereo example, except that the random dot pattern was removed, and replaced with a Lambertian surface. The lighting was controlled so that the surface was spot-lit directly from the front, as shown in Figure 5.14

The same surface that was imaged with stereo was repainted to act as a Lambertian reflector. The image to which the shape from shading algorithm was applied is shown in Figure 5.15

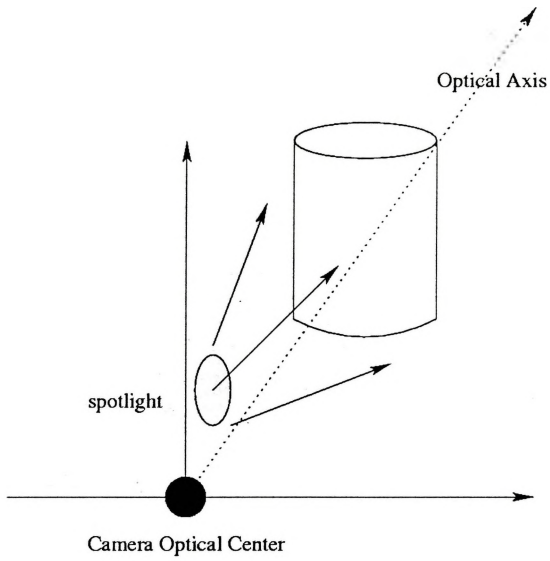

Figure 5.14: Lambertian Reflectance Test Setup

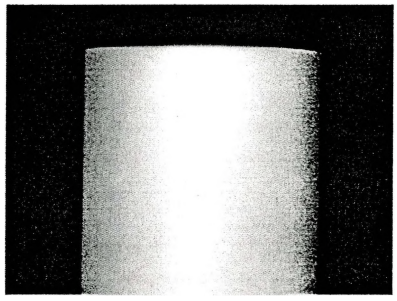

Figure 5.15: Lambertian Reflectance Surface

The resulting depth map, after two iterations of the algorithm is shown in Figure 5.16

5.3.2 Summary Of Lambertian Reflection Results

From the observed depth map, the depth resolution of this technique is very high. Each local gradient calculation can be determined using the full range of the greyscale information. As such, each point could have a surface normal calculated to a resolution of 1 in 256 (for the case of 8 bit greyscale images). These normals are integrated

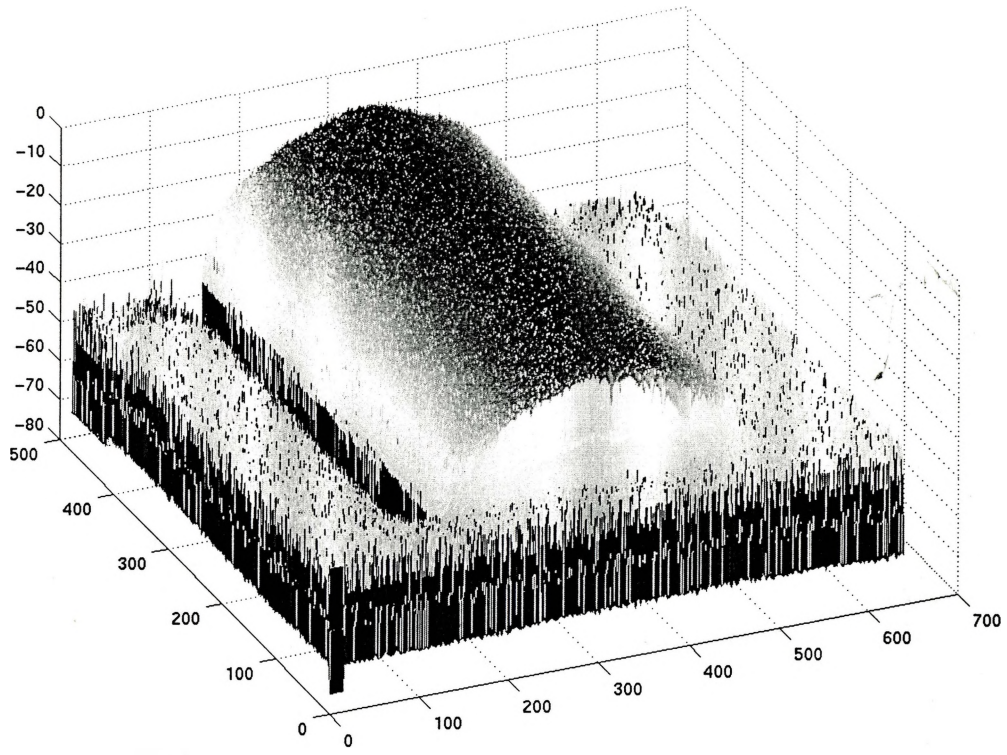

Figure 5.16: Lambertian Surface Depth

across the entire surface pixel by pixel. This high resolution depth is in sharp contrast to the stereo case, which is limited by the sensor's resolution. Unfortunately, very few objects in reality exhibit purely Lambertian reflectance, and as will be seen later, without modification, a Lambertian surface model applied to a non-Lambertian surface is not valid.

5.3.3 Specularity

If the reflectance function is modified to fit a model for describing surfaces that do not reflect light evenly in all directions, such surfaces can be treated in the same way as the Lambertian case. For the example used here, the reflectance function was modified to follow the Phong shading model, $\mathbf{n}^{o^T}\mathbf{s}^o + k \cdot (\mathbf{n}^o\mathbf{h}^o)^n$. For the case where $\mathbf{s}^o = \mathbf{v}^o$, the halfway vector, as described in Figure 3.3 is equal to $\mathbf{s}^o$ the model reduces to $\mathbf{n}^{o^T}\mathbf{s}^o + k \cdot (\mathbf{n}^{o^T}\mathbf{s}^o)^n$. A reflective surface was imaged, and diffuse only (Lambertian), and diffuse with specularity (Phong) lighting models were compared. The object imaged is shown in Figure 5.17.

Figure 5.17: Specular Surface

When only diffuse reflectance was included in the model, the resulting depth map is shown in Figure 5.18.

If the model is modified to employ Phong shading, the resulting depth map is shown in Figure 5.19

5.3.4 Summary Of Specularity Results

Applying the Phong shading model to handle specular surfaces is a significant step forward in building a lighting model general enough to deal with a wide class of real world objects. The strong dependence of the reflection component and the viewing direction, handled with the exponential term, provided a simple way of handling

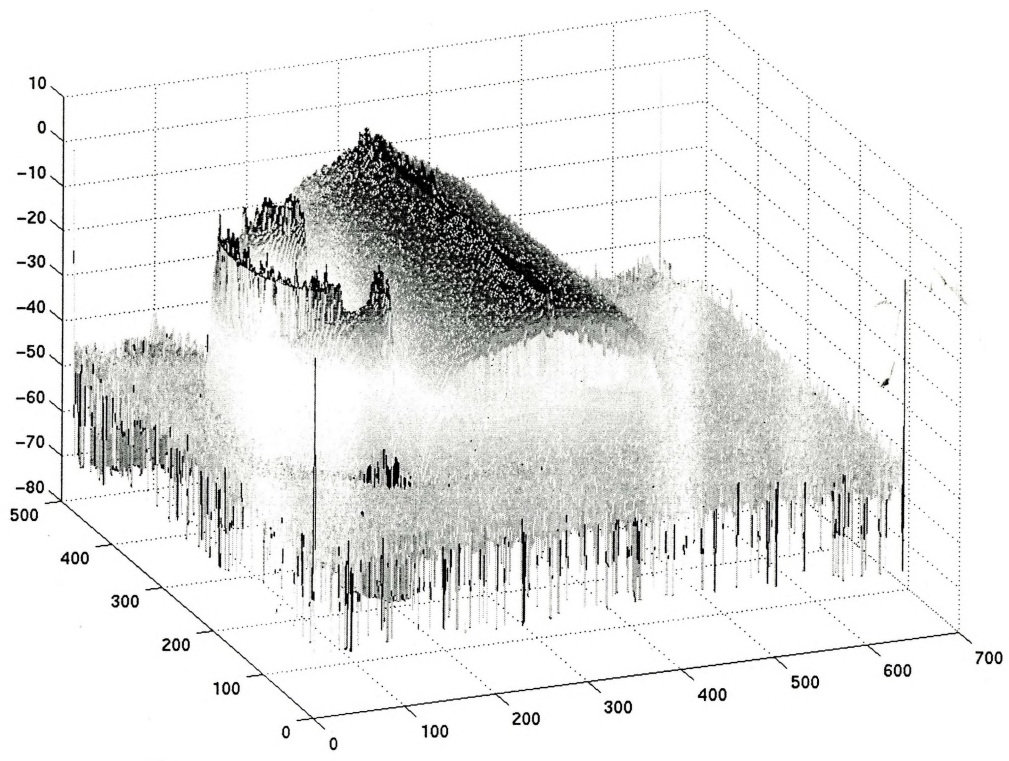

Figure 5.18: Lambertian Model Surface Depth

reflective surfaces in isolation. Unfortunately, reflective surfaces interact with their environment in the light that they reflect onto other nearby surfaces. This important consideration will come to light again when we try to develop a more complete strategy for dealing with objects in a general scene.

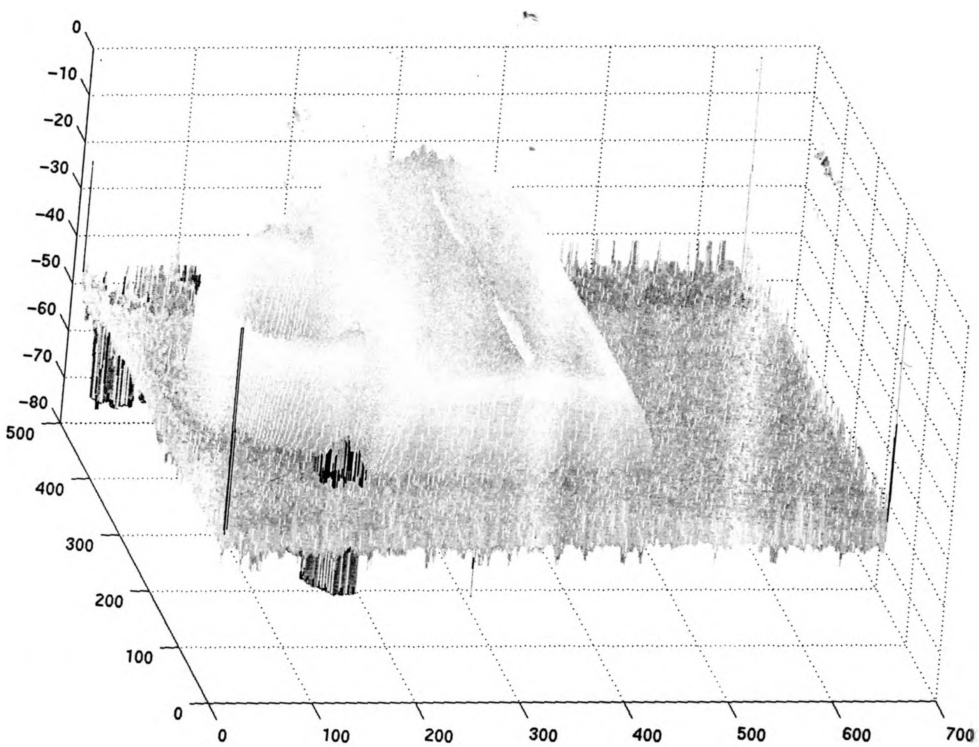

Figure 5.19: Phong Model Surface Depth

5.4 Consistant Stereo and Lighting

As a simple illustration of the process of making lighting and stereo information consistent, Surfaces which would produce incorrect depth results under either stereo or lighting based algorithms were examined. The combined algorithm first computes a stereo map of the salient features (edges) of the object. This surface is cubic spline interpolated to find an approximate surface shape. Lighting is then projected onto the surface, from the direction of the known lighting source. The difference between the observed pixel intensity and the surface reflectance map gives an estimate for the pixel by pixel surface albedo. At the same time, the surface is examined using a simplified (Lambertian) reflectance model. This will produce incorrect results, because it assumes constant albedo. The amount of error in the depth map between the stereo depth result and the shape from shading depth map is used to determine the amount of the albedo term that will be employed at each pixel. The resulting steps are summarized below.

Step 1: Find an edge map for the left and right images.

Step 2: Calculate correspondences at edge pixels only.

Step 3: Interpolate the sparse depth map with a cubic spline surface fit.

Step 4: Project lighting onto the surface using $R(\mathbf{n}) = \mathbf{n}^T\mathbf{s}$.

Step 5: Find $\rho(i,j) = \frac{E(i,j)}{R(i,j)}$.

Step 6: Calculate a depth map using Lambertian reflectance model.

Step 7: Compare the difference in the depth maps, and calculate a correction factor over the range [0,1] for minimum to maximum depth difference.

Step 8: Recalculate the depth map using the modified lighting model and the correction factor by adjusting the measured pixel intensity according to $E(i,j) = E_{correction} \cdot \frac{E(i,j)}{\rho(i,j)} + (1 - E_{correction}) \cdot E(i,j)$. This implements the lighting model $E(i,j) = E_0 \rho(i,j) \cdot \mathbf{n}^{oT}(i,j)\mathbf{s}^o$ with variable amounts of ρ for each pixel.

Three examples will be presented that show how a consistent lighting and shape model can disambiguate albedo variation for lighting and improve depth resolution from the stereo result.

A cylinder with a variable albedo comprised the first example. Here, the stereo result alone would produce a low resolution, discontinuous depth estimate and the shape from shading algorithm alone would fail with the variable albedo. A second example where even the modified lighting model is incorrect is illustrated next. Specifically, surface discontinuities which cannot be handled by the process used to determine the depth map under shape from shading, are included. As a final example, an object which features a smooth surface was examined, and the general visual result of the scheme with the original intensity map reprojected onto the surface, was observed.

5.4.1 Example 1 - Variable Albedo

A Lambertian surface with a variable albedo, shown in Figure 5.20, was used to illustrate how stereo and lighting can be made to describe a consistent view of an object. The object was tilted to ensure that the simple block matching stereo algorithm did not fail on horizontal edges.

A zero cross edge detector was employed to find features for the block matching stereo algorithm. These images are shown in Figure 5.21.

After calculating correspondences, and interpolating the surface using a cubic spline surface fit, the resulting surface is shown in Figure 5.22. Only part of the

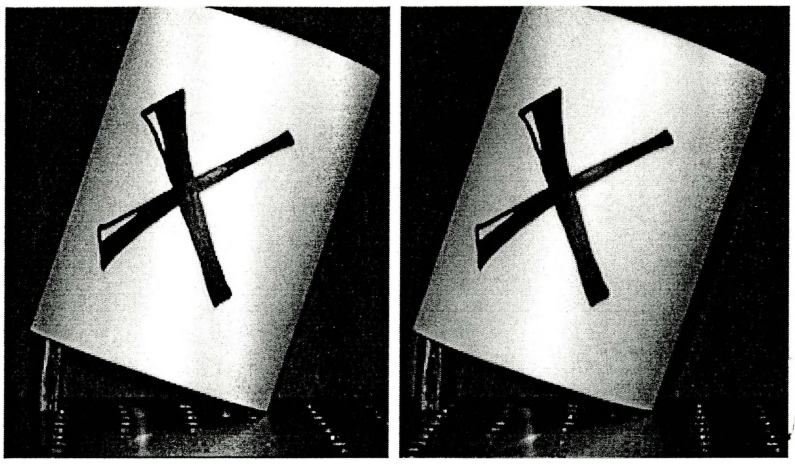

Figure 5.20: Variable Albedo Surface Left and Right Stereo Images, 80 mm Baseline

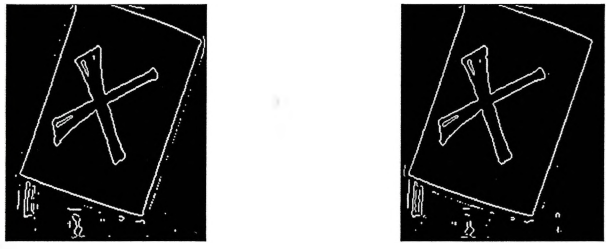

Figure 5.21: Step 1 - Edge Map Left and Right Stereo Images

surface was fit, since the spline surface fit can only provide a continuous surface, and thus cannot handle discontinuities. As such, surface information did not exist past the object boundaries.

When light was projected onto the surface, the resulting reflectance map is shown in Figure 5.23. The projected light was from the same direction as the lighting under which the images were taken.

After the removal of changes in the intensity with respect to the surface orientation, what is left is the surface albedo, as shown in Figure 5.24. Notice that the surface albedo is somewhat independent of the surface tilt, as the effect of the depth induced reflectance has been removed.

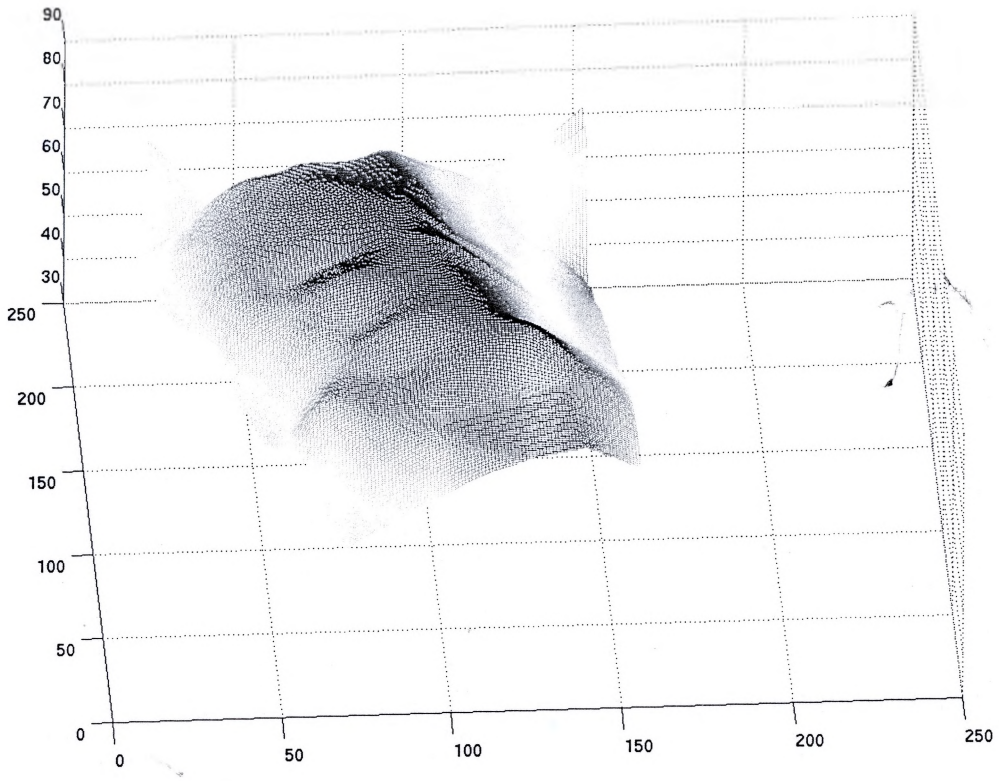

Figure 5.22: Step 3 - Depth Map From Stereo

Applying the linearized shape from shading algorithm to the right stereo image, we produce a depth map as shown in Figure 5.25. Since there is a strong intensity gradient due to the 'X' painted on the surface, the surface depth at that point follows the contour provided by that gradient, producing a large dip in the surface.

Comparing the stereo and lighting calculated depth maps, a normalized correction factor based on the difference, is shown in Figure 5.26. The correction only applies over the range where the stereo information is valid. If there is no correction factor information, the shape from shading computed depth is used.

The correction factor increases gradually from the minimum error of the depth maps, such as over the white part of the surface, to a maximum over the middle of

Figure 5.23: Step 4 - Reflectance Map From Stereo

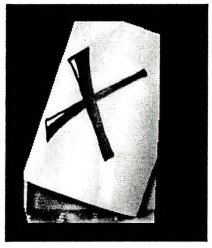

Figure 5.24: Step 5 - Surface Albedo

the black part of the 'X'. This is the slope of the intensity gradient from the original image. This slope, which is of finite value, will cause the application of the correction factor to be gradual over the region of the 'X'.

Once the original image intensities are adjusted to reflect the change in surface albedo, the resulting intensity image is shown in Figure 5.27. Once again, only where valid stereo information is present is the image intensity modified. Here, the result of this gradual application of the correction factor is clearly visible. The boundary regions of the 'X' are overly dark, with lightening increasing towards the center.

Finally, the resulting depth map from using the corrected image intensity is shown in Figure 5.28. Again, the effect of this gradual application of intensity is visible. The surface still has a small dip over the region where the correction factor was insufficient to correct for the variance in albedo.

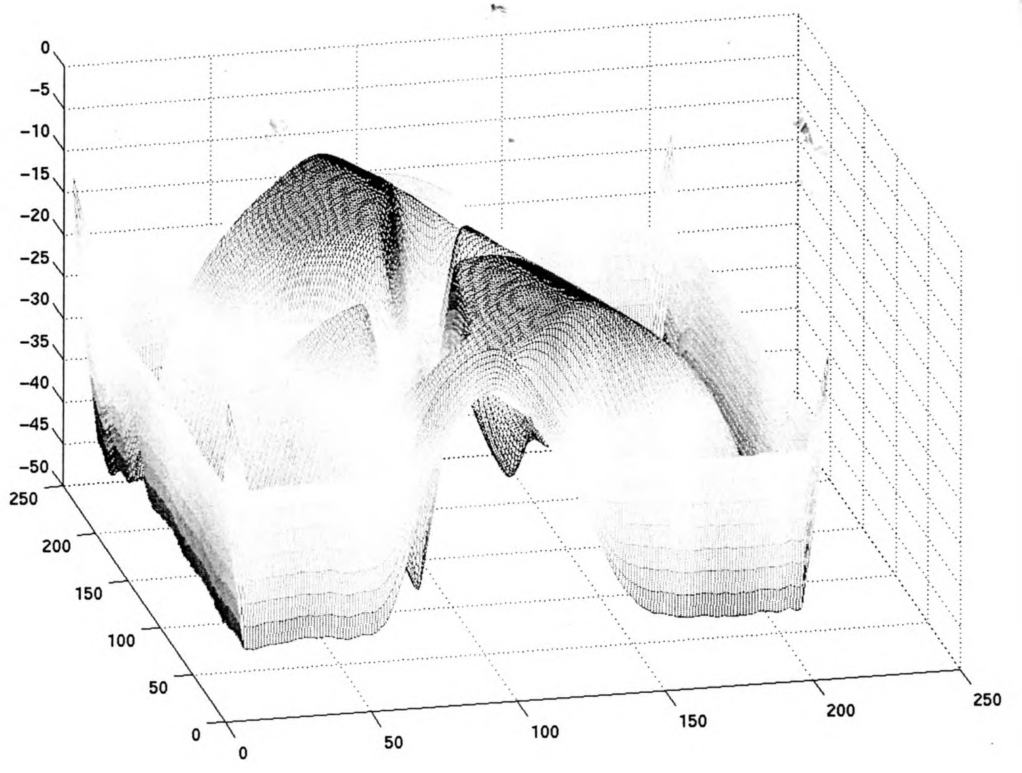

Figure 5.25: Step 6 - Surface Depth From Shape From Shading

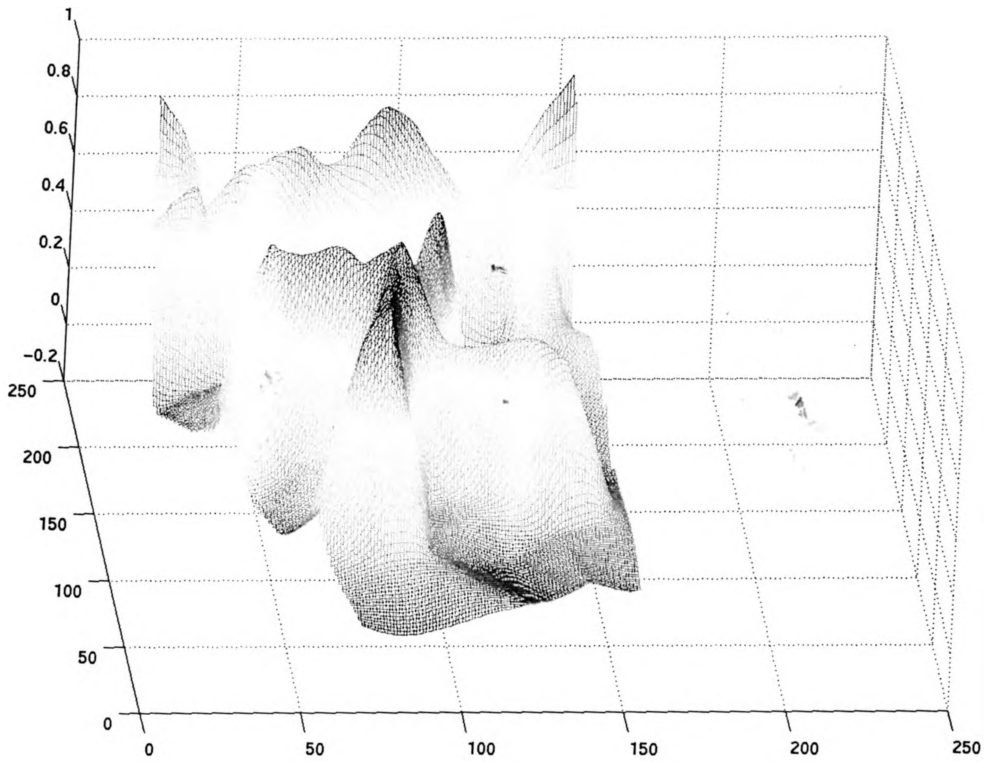

Figure 5.26: Step 7 - Correction Factor

Figure 5.27: Step 8 - Albedo Corrected Image

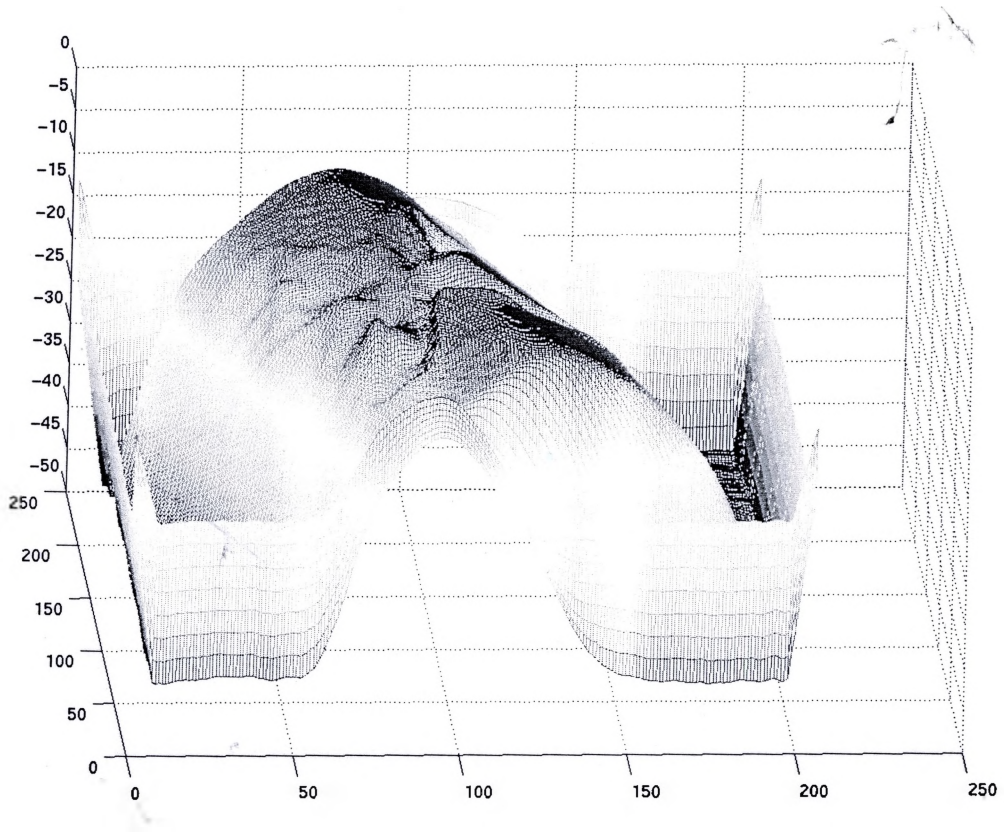

Figure 5.28: Step 8 - Consistant Stereo And Lighting Depth Map

5.4.2 Example 2 - Surface Discontinuities

As an example of the application of an incorrect algorithm to the solution of the lighting model equation, an object with many surface discontinuities was observed. This clearly illustrates the need to understand the physical nature of the problem at hand, and to apply an appropriate solution. The object examined is shown in Figure 5.29.

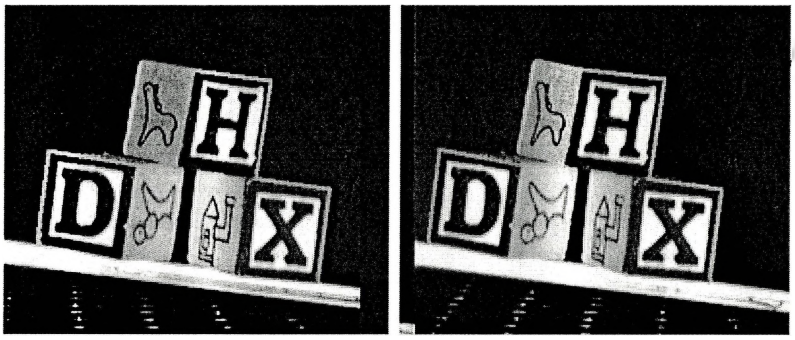

Figure 5.29: Discontinuous Surface Left and Right Stereo Images, 80 mm Baseline

Applying edge based stereo to this object produced a result as shown in Figure 5.30. The stereo result, while lacking detail, due to the low resolution, short baseline method, still produced a good overall shape of the objects in the scene.

When the shape from shading algorithm was applied to this object, the resulting surface depth is shown in Figure 5.31. This depth map bears little or no resemblance to the original object. Additionally, there is not just a simple correspondence between variance in albedo and errors in surface depth. The major problems occur at surface discontinuities, where there is no correspondence between intensities before and after the discontinuities.

The difference between the two depth maps produced a correction factor as shown in Figure 5.32. When this correction factor was applied to the image intensity, the corrected image intensity appears in Figure 5.33. Most of the correction factor is being applied at the surface discontinuities. Unfortunately, this correction will be applied

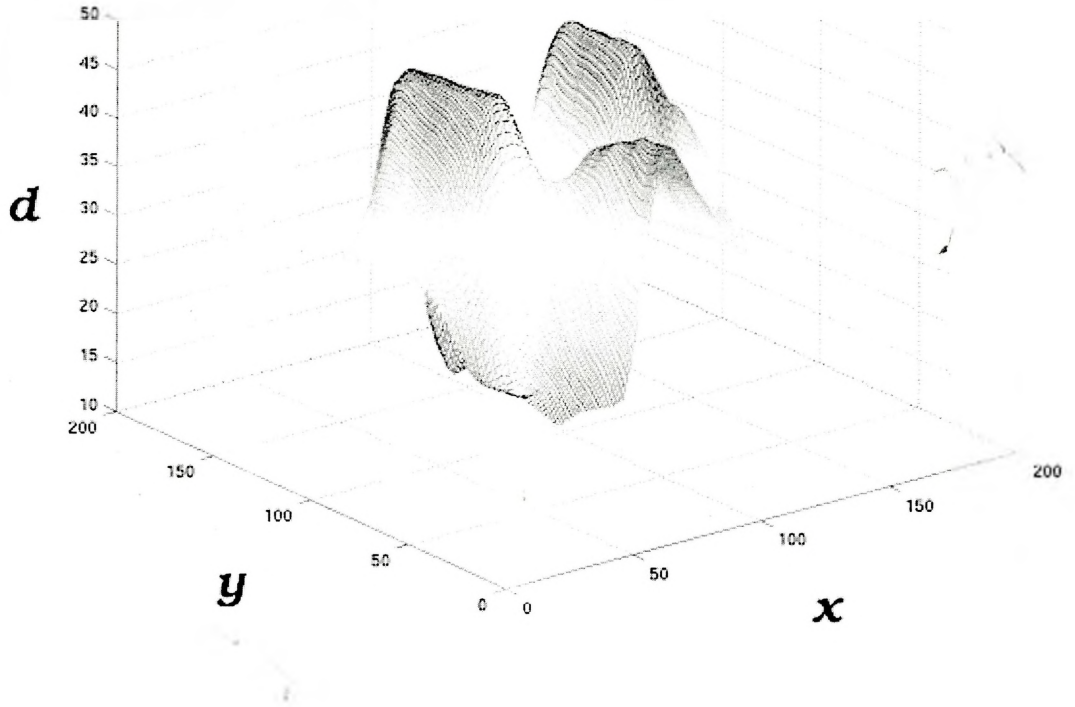

Figure 5.30: Stereo Depth Map

on the basis of surface albedo, and this is not the reason for the high correction value. Because of this, the application of the albedo based intensity correction will do nothing to correct the shape from shading depth errors.

Once the corrected image intensity was used for the shape from shading, the final depth map result is shown in Figure 5.34. Over the continuous surfaces of the objects, such as the front part of the blocks, the correction did serve to smooth the effect of the albedo variations due to the block lettering. Over the rest of the depth map however, the results are very poor. This provides a motivation for applying a surface by surface lighting model that does not depend on a continuous surface assumption.

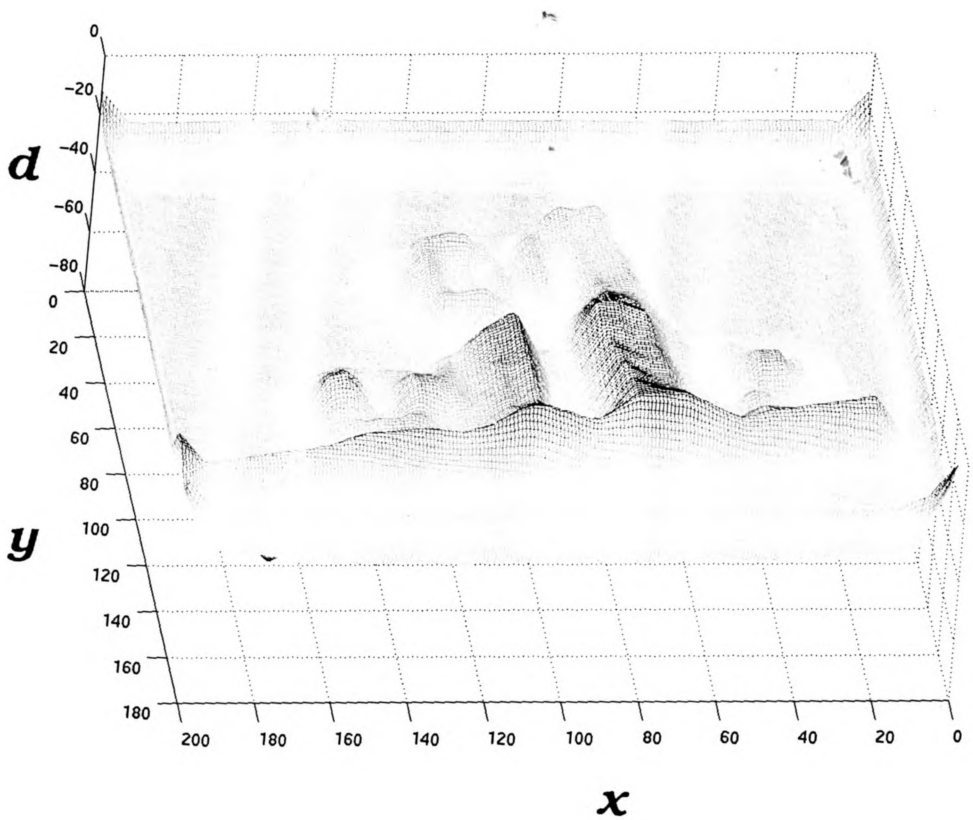

Figure 5.31: Shape From Shading Depth Map

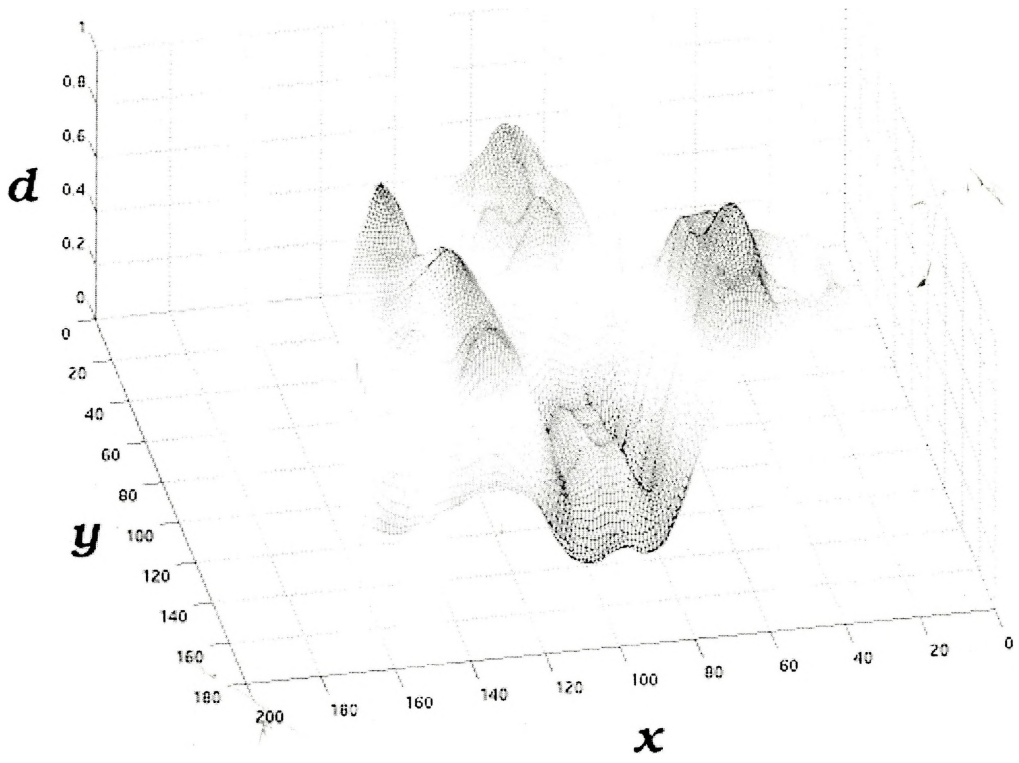

Figure 5.32: Discontinuous Surface Correction Factor

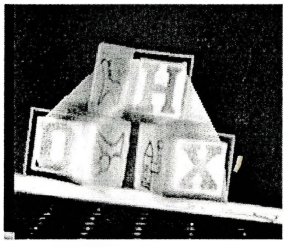

Figure 5.33: Corrected Intensity Image

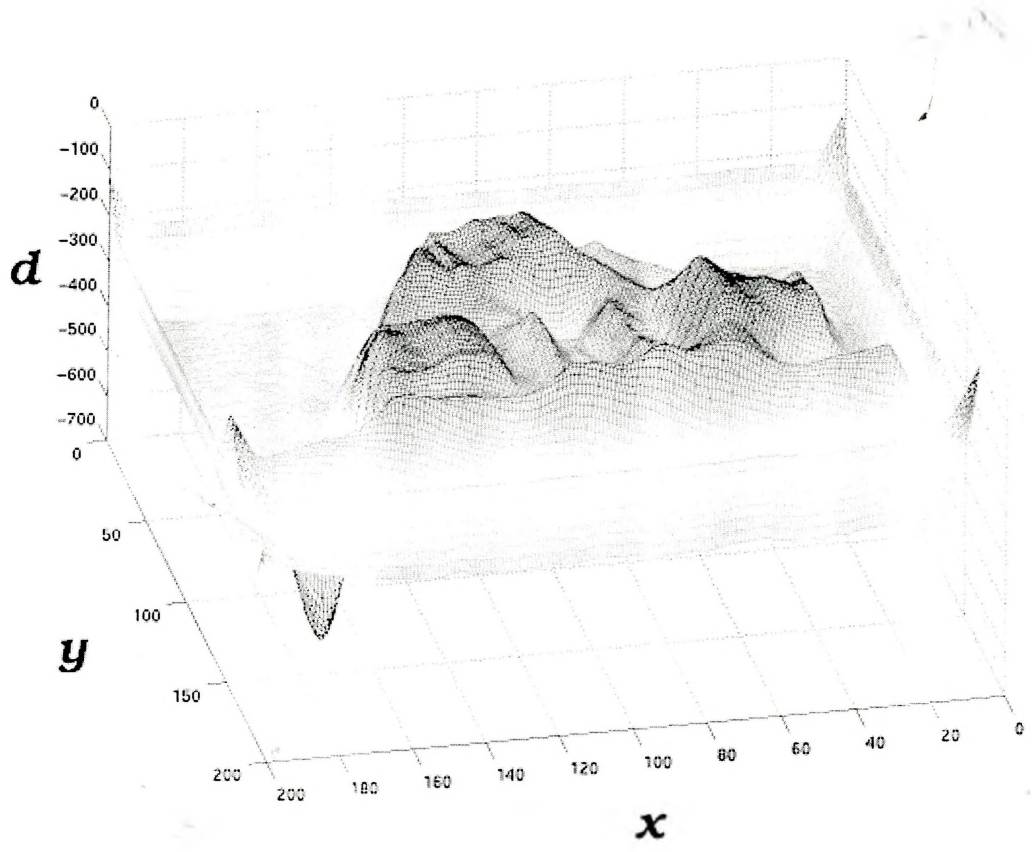

Figure 5.34: Stereo And Lighting Depth Map

5.4.3 Example 3 - A Common Object

As a final example, an object for which it would be difficult to obtain a depth map by stereo was examined. The stereo and lighting results were integrated, and the resulting depth map was remapped with the original texture for aesthetic appeal. Additionally, lighting was projected on the final surface result to illustrate the difference between the imaged shadowing and the reconstructed depth map shadows.

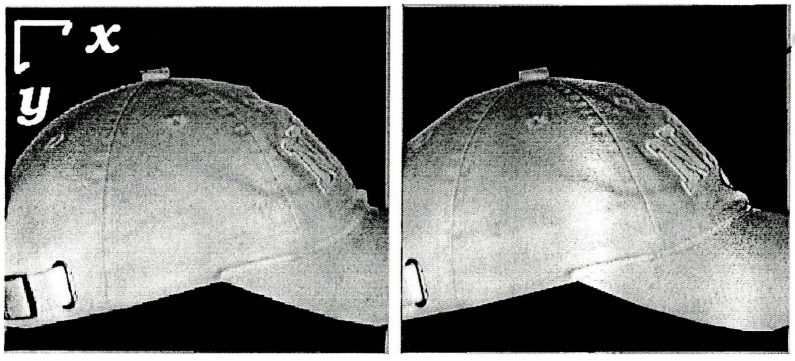

Figure 5.35: Left and Right Stereo Images

The shape from shading and stereo depth maps are shown in Figures 5.36 and 5.37.

Reprojecting the lighting in the same direction as the original lighting map, with the combined stereo and lighting depth map produced a result as shown in Figure 5.38. This result can be compared with Figure 5.39 which shows the original intensity image projected onto the constructed surface. In a great many cases, the small deviations and folds in the fabric are illustrated in both the synthesized depth map and the original image. This image would normally be very difficult to apply stereo techniques to, but the edge based method, with spline surface fitting produced a reliable result.

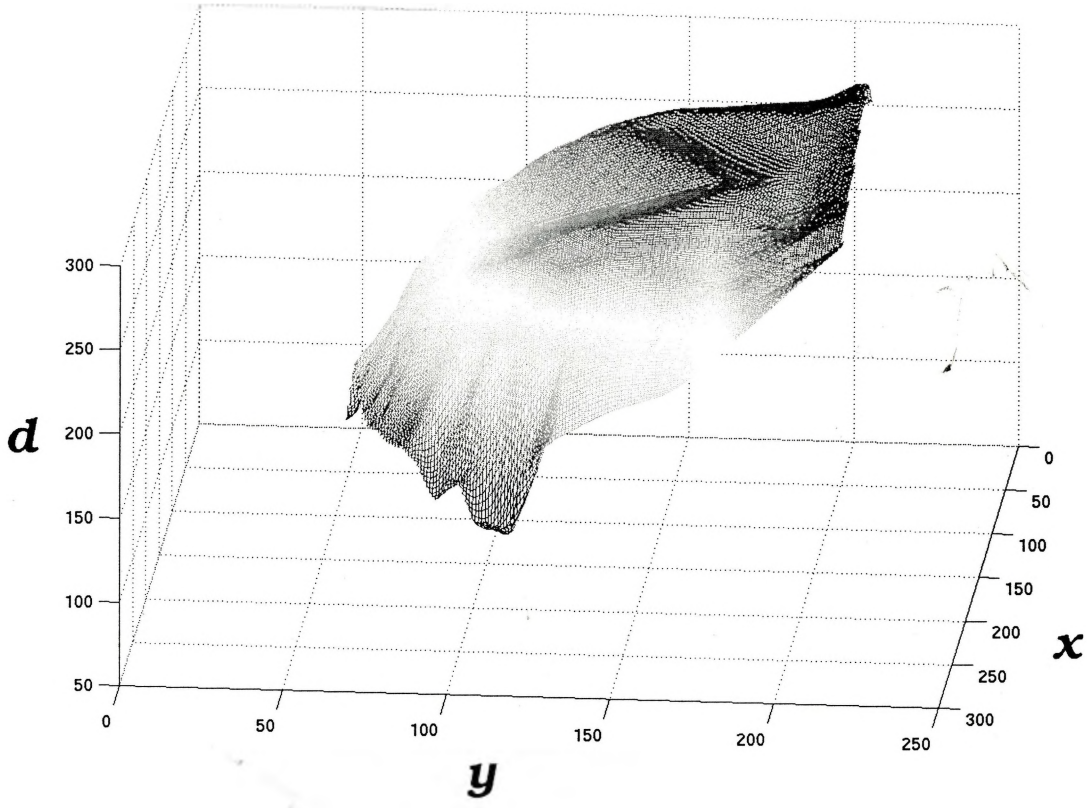

Figure 5.36: Stereo Depth Map

5.4.4 Summary Of Consistant Stereo And Lighting Results

It was shown in the first example in this section, that by modifying the Lambertian surface assumption to include a variable albedo, surfaces with such albedos could be described by a simple reflectance map equation that included the albedo at each point on the surface. Stereo can be used to provide this albedo estimate by giving an initial surface estimate, which can be used to estimate the surface albedo. Care must be taken in the application of this albedo information, however. Direct application will simply drive the lighting model to converge on a surface defined by the stereo result. An appropriate mixture of the two must be used. Thus, the weighting scheme,

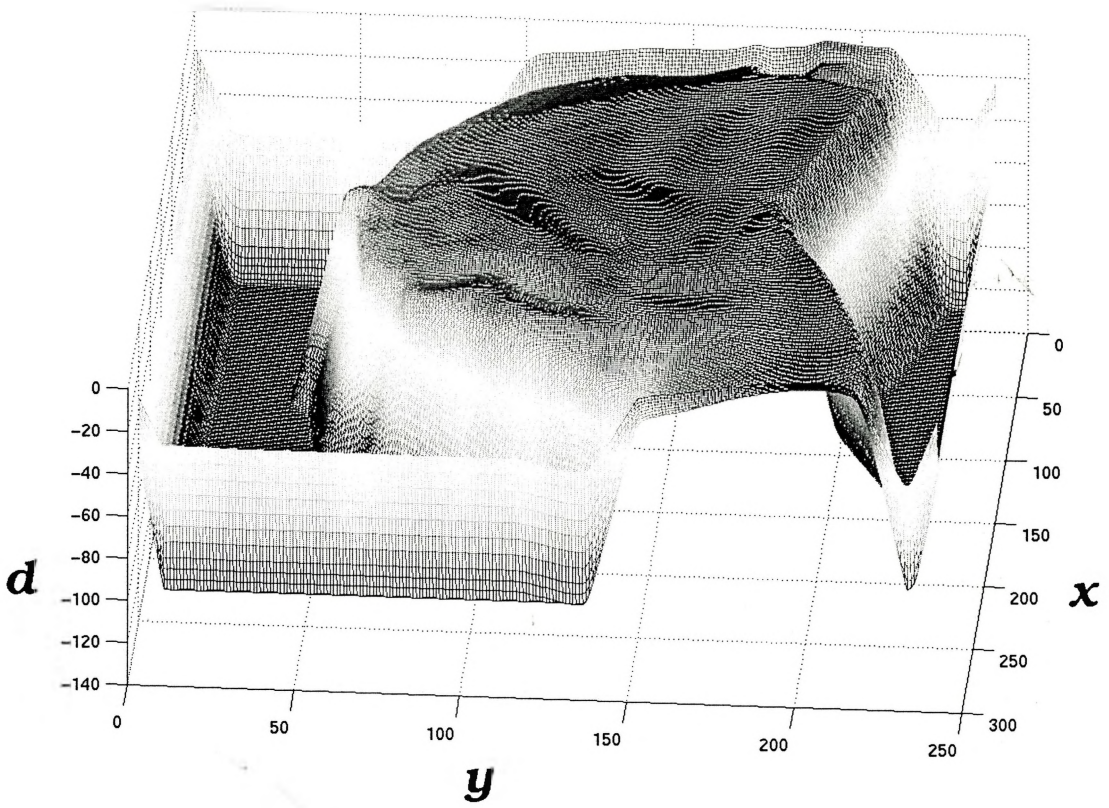

Figure 5.37: Shape From Shading Depth Map

based on the depth map difference, with the stereo as the ground truth, when stereo information was available, was chosen. One side effect from this weighting scheme, however, is that the correction factor is applied gradually, depending on the difference between the two depth maps, and there will be regions where the albedo correction is insufficient to adjust for the variations in albedo.

Applying the linearized shape from shading algorithm to a discontinuous surface was shown to be a flawed approach. The model simply does not fit the reality. Over regions of surface continuity, the results were still usable.

The final example typifies an application of the consistent stereo and lighting technique to a common object that was not idealized in any way for the method,

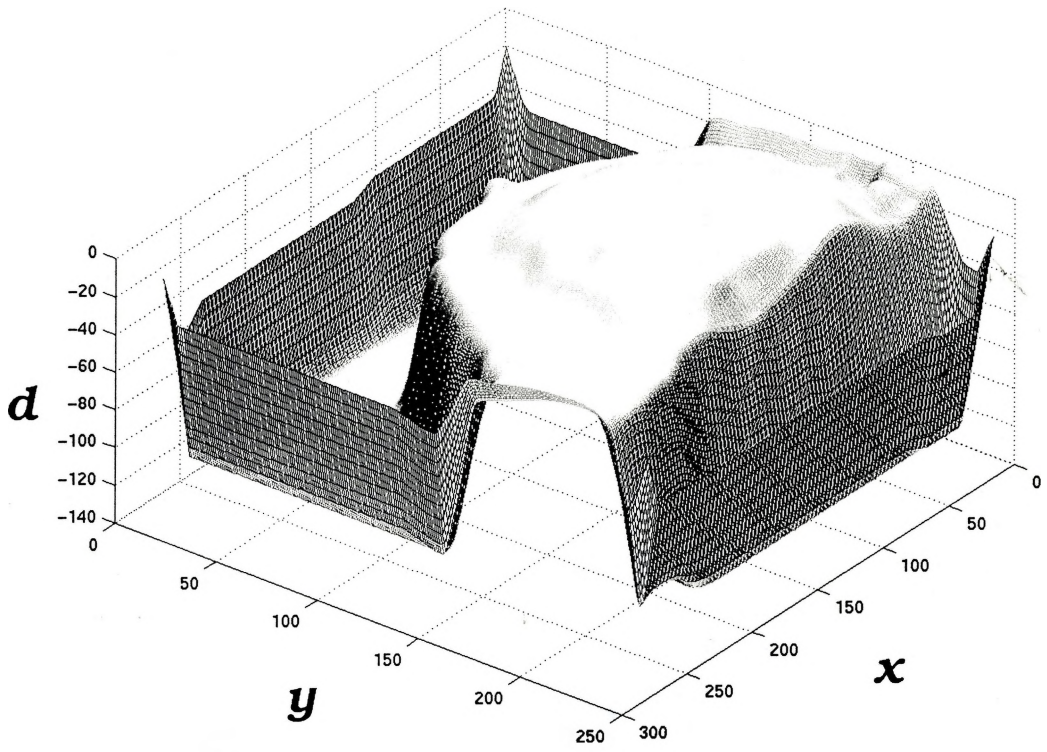

Figure 5.38: Stereo And Shape From Shading Depth Map With Reprojected Lighting

except that it did not contain surface discontinuities. When the texture was remapped onto the surface, the overall visual result was quite convincing. The micro features, such as the stitching, were absent from the albedo - independent reconstructed depth map. Some of the features were simply too small to be described with either stereo or shape from shading, and some of the features were of variable albedo, which would not be on the albedo - independent map in any case.

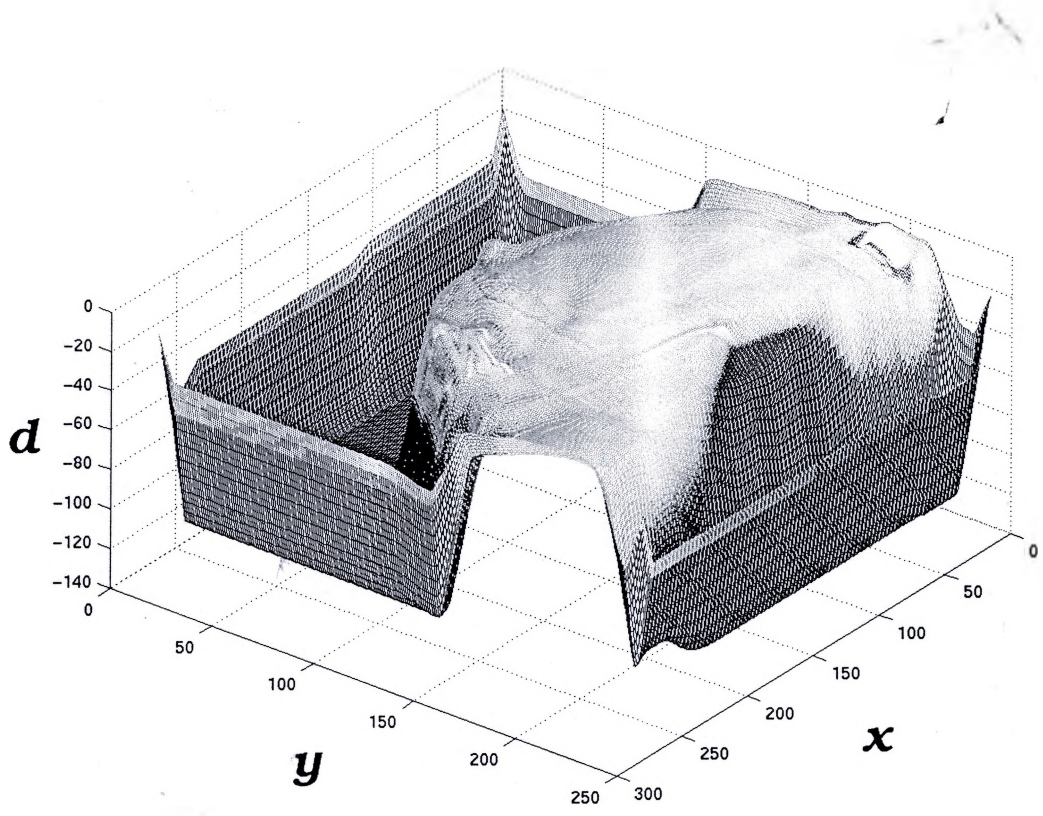

Figure 5.39: Texture Mapped

Chapter 6

Discussion And Future Work

6.1 Discussion

6.1.1 Epipolar Geometry And Area Based Stereo

Before the process of point correlation can continue, knowledge of the epipolar geometry must be obtained. This constrains the search for matches to a linear one and makes the problem computationally tractable. As was shown, finding this geometry is relatively simple if knowledge of the camera geometry is known. Finding the epipolar geometry is quite difficult if point matches must be used, since one wrong match will produce a poor estimate of the epipolar lines, due to the sensitivity of the system of equations used to solve for the fundamental matrix.

If a stereo depth map is produced with CCD sensors of average resolution and a short baseline, the resulting depth resolution allows for very few levels of depth when imaging objects of a scale on the order of a few tens of centimeters at a distance of 1 meter. If the baseline is extended to about a meter, the depth resolution allows for sub - millimeter resolution at a distance of a meter. However, baselines of this width allow for a very narrow field of overlap between left and right sensors, due

to the widely disparate views. Additionally, these disparate views make matching difficult, since the view of a feature in one image can be significantly warped in the other image, due to perspective.

The result of these issues is that stereo alone cannot produce smooth, error free depth maps with average resolution sensors. Additionally there can be great difficulty determining epipolar geometry, unless the geometry of the imaging system is known a-priori.

6.1.2 Shading and Lambertian and Specular Reflectance

When a simple lighting model, such as Lambertain reflectance, is applied to a Lambertian surface, a very high resolution depth map can be obtained. If the model is modified slightly, to accommodate the presence of specularity, by using the Phong shading model, specular surfaces can produce quite accurate depth maps. However, if these models are applied to surfaces whose reflection characteristics do not behave as the model describes, the resulting depth map contains large errors.

6.1.3 Combining Stereo And Lighting

If the albedo of a surface is not constant over the entire surface, then the lighting model must take into account this albedo dependence. As a simple example of how to use stereo and lighting together in one consistent view, a variable albedo surface was observed under a Lambertian lighting model and stereo was used to provide an estimate of the albedo. The new albedo dependent Lambertian reflectance was used to calculate a final depth map. Edge based stereo, instead of dense, block based correlation was used. This ensured a minimum of errors by fitting a cubic spline surface over depth points that were found at features where there was a high probability of finding correspondences.

This lighting model was far too simple to be able to handle surface discontinuities, since the point by point surface depth is dependent on the intensity gradient, which is incorrect across a surface boundary. However, a more complex lighting model, as will be described later, could handle this condition.

To summarize, the reason for combining stereo and lighting was to ensure that the stereo and lighting results agreed with each other. Although it did allow the application of lighting based depth calculation over albedo dependent surfaces, this was not the primary point of the combined stereo and lighting experiments. The main idea that this work illustrates is that stereo or lighting based depth methods separately are insufficient and inaccurate models for describing the depth of a scene. They will produce results that are inconsistent with each other unless a method is employed to ensure that they are consistent.

6.2 Future Work

6.2.1 Improving The Lighting Model

In review, it was shown that there are a variety of problems with the depth map provided solely by area-based stereo. Specifically, the resolution of the sensor only gives a few levels of depth in a scene given a typical stereo geometry. This produces a very coarse approximation to the actual range of a point in the scene. Further, many points in the scene have few or no identifiable features, making correlation with points in another view impossible. The occurances of these problems is due to the application of the incorrect model for the computation of image disparities. The general model of using triangulation for depth assumes an accurate measure of the position of correspondences between images. Using a discrete sensor with relatively low resolution and attempting to find correspondences with blocks over the

entire image is in complete violation of the triangulation model. It is thus unwise to continue to apply a flawed model to the stereo process. Below, a more complete and robust model for stereo is described.

The problems that were observed with shape from shading were also shown to be the result of applying an incomplete and inaccurate model to the lighting scenario of the scene. The assumption of Lambertian reflectance is incorrect for all but a special few objects. The integration of the point reflectances across a surface also fails when the surface has discontinuities. Accurate and robust lighting models can be constructed for a complete scene even with multiple objects interacting provided that an estimate of the surface orientation is known a-priori. As such using an incomplete lighting model to try to estimate the surface orientation has met with little success. The lighting model has been applied too early in the visual process.

The direct combination of stereo and shape from shading using frequency based methods, or other ad-hoc schemes are all limited by the fundamental issues present in either of the methods separately. This discussion leads to the formulation of another way of looking at a computational theory of depth from stereo and illumination information, which is consistent for both shape and lighting in the scene. Stereo is used as a low-level process, and provides a three-dimensional mesh outline, similar to the wire frame models that are used in 3D CAD. From this mesh, surfaces are fit by a co-operative process. First, Bezier patches are constructed from the array of points provided by stereo. With this estimate of the form factor geometry, one can first solve for the reflectance of each patch, employing the process of successive refinement for radiosity. For each iteration, there are two important variables for each patch, with respect to the patch from which we are "shooting" light energy. The first variable is the form factors of the patch, and the second is the reflectance. In general, given the appropriate multivariate optimization process, it is possible to provide an optimum solution for the patch reflectances and the form factors, since the form factors have

an initial estimate from the stereo geometry. Each iteration of the process provides an updated estimate of all of the patch reflectances and the form factor terms one patch at a time.

The completeness and accuracy with which radiosity can describe scene radiance is evident from viewing graphics built with this technique. Similarly, taking into account diffuse and specular reflection when estimating patch reflectivity proves to be a great improvement over the Lambertian reflectance assumption.

6.2.2 Improving Low-Level Vision With Stereo

The lowest level part of the process of determining the depth of scene should include the use of stereo. However, area-based stereo, which produces a dense depth map is unsuitable, as stated above. Stereopsis has been studied extensively in the human visual system, and it has been shown by [18] and [20] that zero-crossings, detected by a LOG (or an approximation to this, a difference of gaussians (DOG) play a critical role. In this light, edge based stereo provides an excellent basis for the low-level part of depth calculation. In very recent work, [36] revisited the work of Marr and Poggio. In this work, they use Marr and Poggio's co-operative algorithm, where each cell can offer excitatory or inhibitory responses to neighboring cells. However, [36] reverts back to using stereo to provide a continuous, dense depth map, which, again, is unsuitable.

To eliminate the problems of epipolar geometry and widely disparate views, the sensors used in stereo calculation should be of higher resolution than the standard medium resolution imaging hardware that is in use. As a point of comparison, the human eye, in the foveal region, has a density of approximately $150,000\frac{pixels}{mm^2}$. This corresponds to a resolution of about 337,000 imaging elements (cones) in an area of about $1.5mm^2$. A standard 640 by 480 pixel CCD is over an area of about $4.8mm$ by $3.6mm$ ($17.28mm^2$). This gives a resolution of about $1000\frac{pixels}{mm^2}$. Such high resolution

sensors would allow for very short baselines, greatly simplifying the matching process. While such sensors are somewhat expensive, they are commercially available.

6.2.3 Building 3D Models With Visual Assistance

Computer graphics modeling relies heavily on surface description (usually with Spline or Bezier surface fitting) and lighting models. Generally, 3D models are separately described for each object in the scene. If vision is to be used to assist the process of building models, there needs to be some way of separating objects in a scene. Object segmentation is a high level visual function, requiring some prior knowledge of the object geometry to successfully segment it from its surroundings. While the surface fit stereo result with consistent lighting does not answer the question of how to segment objects, they offer object models that could be used as a starting process for the segmentation process. Again, the idea here would be to make the scene segmentation consistent with both the depth result and the rendered result. If the processes of stereo, lighting, and segmentation are treated cohesively, the resulting model will be much more accurate, and will apply over a much wider set of conditions than with any one of the models independently.

Bibliography

[1] N. Ayache. *Artificial Vision For Mobile Robots: Stereo Vision and Multisensory Perception.* MIT Press, Cambridge, Massachusetts, 1991.

[2] S. Barnard and W. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 2(4):333–340, 1980.

[3] A. Blake, A. Zisserman, and G. Knowles. Surface descriptions from stereo and shading. *Image And Vision Computing*, 3(4), 1985.

[4] M. Cohen and J. Wallace. *Radiosity And Realistic Image Synthesis.* Morgan Kaufmann Publishers, Inc., 1993.

[5] J. Cryer, P. Tsai, and M. Shah. Combining shape from shading and stereo using human vision model. Technical report, Department Of Computer Science, University of Centra Florida Technical Report, 1996.

[6] A. Fusiello, E. Trucco, and A. Verri. Rectification with unconstrained geometry. In *Proceedings Of The British Machine Vision Conference*, pages 400–409. BMVA Press, September 1997.

[7] G. Gimel'farb, V. Marchenko, and V. Rybak. An algorithm for automatic identification of identical sections on stereo pairs of photographs. *Cybernetics*, 8(2), 1972.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry In Computer Vision.* Cambridge University Press, 2000.

[9] B. K. P. Horn. *Shape From Shading: A Method For Obtaining The Shape Of A Smooth Opaque Object From One View.* PhD thesis, MIT, 1975.

[10] B. K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8(2):201–231, 1977.

[11] D. Hubel and T. Wiesel. Brain mechanisms of vision. *Scientific American*, 241(3):150–162, 1979.

[12] B. Julesz. *Foundations Of Cyclopean Perception.* University Chicago Press, 1971.

[13] Reinhard Klette, Karsten Schluns, and Andreas Koschan. *Computer Vision - Three Dimensional Data From Images.* Springer-Verlag Singapore Pte. Ltd., 1998.

[14] H. Lange. Advances in the cooperation of shape from shading and stereo. In *Proceedings of the Second International Conference on 3D Imaging and Modeling*, Ottawa, Canada, October 1999.

[15] M. Levine, D. O'Handley, and G. Yagi. Computer determination of depth maps. *Computer Graphics And Image Processing*, 2:131–150, 1973.

[16] H. C. Longuet-Higgens. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–138, 1981.

[17] Q.-T. Luong and O.D. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. Technical report, I.N.R.I.A Report, 1995.

[18] D. Marr. *Vision: A Computational Investigation Into The Human Representation and Processing of Visual Information.* W. H. Freeman and Company, 1982.

[19] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194, 1976.

[20] J. Mayhew and J. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.

[21] J. Mayhew and J. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.

[22] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. *International Journal Of Computer Vision*, 7(2):143–162, 1992.

[23] A. Pentland. Shape information from shading: A theory about human perception. In *Second International Conference On Computer Vision*, pages 404–413, Tampa, FL, December 1988. Computer Society Press.

[24] B.-T. Phong. Illumination for computer-generated pictures. *Comm. ACM*, 18(6), June 1975.

[25] L. Piegl and W. Tiller. *The Nurbs Book.* Springer-Verlag, 1997.

[26] M. Pilu. Uncalibrated stereo correspondence by singular value decomposition. Technical Report HPL-97-96, HP Laboratories Bristol Technical Report, August 1997.

[27] S. Pollard, J. Mayhew, and J. Frisby. A stereo correspondence algorithm using a disparity gradient constraint. *Perception*, 14:449–470, 1985.

[28] D. Samaras, D. Metaxas, P. Fua, and Y. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 480–487, Hilton Head Island, South Carolina, 2000.

[29] Y. Shirai and Y. Nishimoto. A stereo method using disparity histograms of multiresolution channels. In *Proc. 3rd International Symposium on Robotics Research*, pages 27–32, France, 1985.

[30] P. Tsai and M. Shah. Shape from shading using linear approximation. *Image And Vision Computing Journal*, 12(8):487–498, 1994.

[31] R. Y. Tsai. A versatile camera calibration technique for hign-accuracy machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal Of Robotics And Automation*, pages 323–344, 1997.

[32] A. Watt. *3D Computer Graphics.* Addison-Wesley, 1993.

[33] A. Watt and M. Watt. *Advanced Animation And Rendering Techniques.* Addison-Wesley, 1992.

[34] C. Wylie, G. Romney, D. Evans, and A. Erdhal. Halftone perspective drawings by computer. In *Proceedings of the Fall Joint Computer Conference*, pages 49–58, 1967.

[35] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. Technical Report 2927, I.N.R.I.A Report, June 1996.

[36] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo occlusion detection. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 22(7):675–684, 2000.