Shape and Image Reconstruction from Focus

A Dissertation Presented

by

Tae-Sun Choi

to

The Graduate School

in Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

in

Electrical Engineering

State University of New York

at

Stony Brook

December 1993

State University of New York
at Stony Brook
The Graduate School

Tae-Sun Choi

We, the dissertation committee
for the above candidate for the
Doctor of Philosophy degree,
hereby recommend acceptance
of this dissertation.

Dr. Murali Subbarao Advisor Department of Electrical Engineering

Dr. John Murray Chairman Department of Electrical Engineering

Dr. Gene Gindi Member Department of Electrical Engineering

Dr. Vaclav Dolezal Outside Member Dept. of Applied Mathematics & Statistics

This dissertation is accepted
by the Graduate School

Abstract of the Dissertation

Shape and Image Reconstruction from Focus

by

Tae-Sun Choi

Doctor of Philosophy

in

Electrical Engineering

State University of New York

at Stony Brook

1993

This thesis presents solutions to the problem of shape recovery and image reconstruction from image focus and stereo. We propose six focus measures which are proved to be theoretically sound under weak assumptions. They include energy maximization of unfiltered, low-pass filtered, high-pass filtered, and band-pass filtered images. One of them –the energy of low-pass filtered image gradient, which has better overall characteristics– is recommended for practical applications.

The image of a scene formed by an optical system such as a lens contains both *photometric* and *geometric* information about the scene. The degree of brightness or radiance and the color of objects in the scene are part of photometric information whereas distance and shape of objects are part of geometric information. Recovering this information from a set of images sensed by a camera is an important problem in computer vision.

Shape-From-Focus (SFF) methods provide one solution to the problem. A new shape-from-focus method was developed based on a concept named *Focused Image Surface* (FIS). The FIS of an object is defined as the surface formed by the set of points at which the object points are focused by a camera lens. The shape of the FIS is determined by searching for a shape which maximizes a focus measure. In contrast with previous literature where the focus measure is computed over the planar image detector of the camera, here the focus measure is computed over the FIS. This results in more accurate shape recovery than the traditional methods. Also, using FIS, a more accurate focused image can be reconstructed from a sequence of images than is possible with traditional methods.

Finally, we extend the proposed SFF method to solve the correspondence problem in stereo-vision. We propose a combined method where a depth map and two focused images obtained by the SFF method are used to solve the correspondence and occlusion problems in stereo-vision efficiently. All the methods presented here have been implemented on an actual camera system, and the experimental results of shape recovery are presented.

To

my parents,

Heeyoul and Seongjin Choi,

and

my beloved wife,

Jaesuk,

and

my lovely daughter,

Eunyoung

Contents

List of Figures

Acknowledgements

I wish to express my deep appreciation to my advisor, Professor Murali Subbarao, for his constant guidance and support during the course of this dissertation. Without his help, this work would never have been possible. I am very grateful to Professor John Murray, Professor Gene Gindi, and Professor Vaclav Dolezal for their constant encouragement and many fruitful discussions.

I am indebted to many of my colleagues at the Computer Vision Laboratory for their collaboration and assistance. The SPARCS prototype camera system used for experiments in this thesis was built by them during the past four years.

Chapters 2 to 5 are based primarily on [56], [57], [58], [59], [60]. This research was supported in part by the National Science Foundation under grant IRI-8821923 and the Olympus Optical Corporation.

Chapter 1

Introduction

1.1 Active Vision

In the human vision system, the eyes focus on an object (accommodation) by adjusting both the pupil's opening size based on its illumination and the lens focal length based on its distance. Also, both eyes turn their direction of view (vergence) to bring the object of interest into the field of view's center. We move our heads for a better view of the object. Therefore visual perception is not passive, but active in the sense of changing the sensor's state [6, 28, 35]. Conceptually these processes can be considered independent modules in active vision, such as focus, stereo, vergence, shading, etc.

In this research we propose the *image focus module* and its application for a cooperative process with stereo. Using blurred images of an object, the module extracts a variety scene information including object distance and radiance (brightness). An object's distance and shape are part of its geometric information. The level of brightness or radiance is a component of photometric information. An important problem in active vision is recovering this information from images sensed by a camera. Many focusing techniques for depth perception have been developed over the past three decades including search depth-from-focus (DFF) methods, Fourier domain approach, spatial domain approach, matrix based method, etc.

The search focus module [36, 57, 59] involves a search of the camera parameter space. The camera parameters include the lens position and the focal length of the lens. This module is accurate but slow; it requires a sequence of image frames (about 10) recorded with different camera parameter settings. In contrast, the nonsearch focus module [15, 61, 65] does not require focusing the object. This method uses two or three blurred images taken with different camera settings. Therefore it is faster than the search focus module but less accurate.

This thesis presents a detailed investigation into this method. The *image focus module* has been studied both as an independent process and as a cooperative process where it interacts with other modules of vision to overcome mutual weaknesses and accomplish efficient perception. Particularly, a combined method with focus and stereo is proposed to solve the correspondence and occlusion problems in stereo-vision and to improve the accuracy in focus.

1.2 Motivation for Research

There are many uses for our research results. The most important of these are the following:

- This work can be used for robust vision systems useful in machine/robotic vision, autonomous vehicle navigation, microscopic imaging, and autofocusing of electronic cameras;
- It can show the potential for integration of the *image focus module* with other modules for shape recovery, particularly for stereo; and
- It can also provide an understanding of the possible role of image focus in human vision.

1.3 Overview

This thesis is organized into following chapters.

In chapter 2, first we introduce our active vision camera system called SPARCS. In chapter 3, we describe two models of image defocusing derived from paraxial geometric optics and diffraction optics [7, 17, 20]. Next, the basic theory of focus is described and a few typical methods for the *image focus module* are explained.

In chapter 4, based on the geometric optics model, we derive a class of focus measures which we prove to be theoretically sound under weak assumptions. We then analyze the side-lobe effect and show how it can be reduced through low-pass filtering. Next we discuss other focusing techniques found in previous literature and include some experimental results.

In chapter 5, we are concerned with the principles and computational methods for recovering the geometry and radiance of an object from its *sensed image.* In computer vision, the sensors are usually planar image detectors such as CCD arrays. Therefore, for curved objects, only certain parts of the image will be focused whereas other parts will be blurred. The traditional SFF methods do not yield accurate shapes or depth-maps of objects. The reason for this is that a focus measure is defined and computed over image frames sensed by a planar image detector. Hence, a new shape-from-focus method is proposed to solve this problem. Next, a theorem and related background will be discussed including some experimental results.

In chapter 6, the proposed SFF method is integrated with the *shape-from-stereo module.* A depth-map and two focused images obtained by the SFF.FIS method are used to solve the correspondence and occlusion problems in stereo vision.

Finally, the summary and conclusions are given in chapter 7.

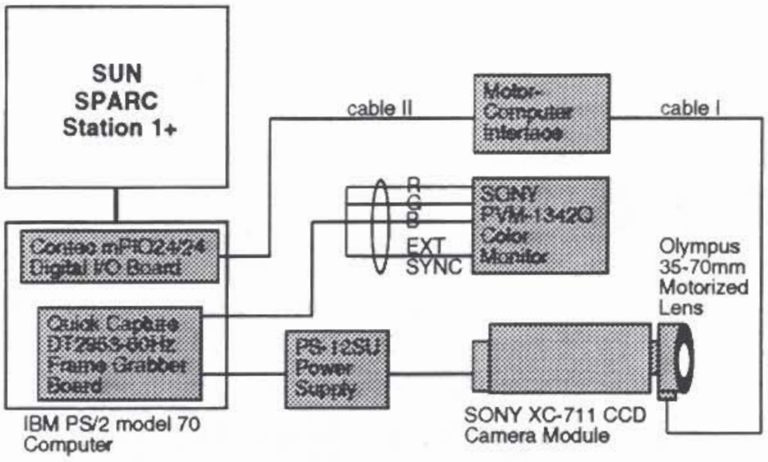
Chapter 2

SPARCS Camera System

In this chapter, we describe our camera system called the Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). SPARCS has been built in our laboratory over the past four years for research and development in the areas of Machine/Robot Vision and Autofocusing of Electronic Camera. In order to investigate various algorithms in these areas, we needed to construct a computer controlled camera system [16] that is able to store sensed images in memory and extract video information from them. Using this information, a computer would then move the motorized lens to the designated position.

2.1 Hardware

A block diagram of the system is shown in Figure 2.1. SPARCS consists of following components.



Stonybrook Passive Autofocusing and Ranging Camera System-SPARCS - is a prototype camera system developed at the Computer Vision Labatory for experimental research in robotic vision, State University of New York at Stony Brook

Figure 2.1: SPARCS.

- SUN SPARC station.
- SONY XC-711 CCD video camera module.
- OLYMPUS OM-System 35-70mm motorized lens.
- Data Translation DT 2953-60Hz QuickCapture frame grabber board.
- SONY PVM-1342Q Trinitron color video monitor.
- Contec mPIO24/24-T 24 channel digital input/output board.
- Chori America Inc. PS-12SU power supply.
- IBM PS/2 model 70 computer.

The IBM PS/2 computer was interfaced with the video camera via two separate channels. The first channel includes the QuickCapture frame grabber hardware. This is used to digitize the incoming video signal and store it. The second channel controls the positioning of the video camera's lens through the use of a parallel digital input/output board and the necessary interface hardware. The interface hardware consists of the motor-computer interface and has been modeled as a black box in Figure 2.1.

In addition to the computer display, a color monitor was added. It is connected directly to the QuickCapture board. The incoming video signal can also be viewed in real time on this monitor.

2.2 Application

The focusing techniques described in this thesis were implemented on the SPARCS system. The focal length of the lens can be varied manually from about 35mm to 70*mm*. The F-number which is defined as the ratio of the focal length f to aperture diameter D can also be set manually to 4, 8, 22, etc. The lens system consists of multiple lenses and focusing is done by moving the front lens forward and backward. The lens **can** be moved either manually or under computer control. To facilitate computer control of lens movement there is a stepper motor with 97 steps, numbered 0 to 96. Step number 0 corresponds to focusing on an object at distance infinity. Step number 96 corresponds to focusing on a nearby object, at a distance of about 50cm from the lens. The motor is controlled by a microprocessor, which communicates with the IBM PS/2 through a digital I/O board (Contec mPIO24/24). Pictures taken by the camera can be displayed in real time on a color monitor (SONY PVM-1342 Q). The images acquired and stored in the IBM PS/2 can be transferred to a SUN workstation.

The camera settings used in the experiments were:

- Focal Length = 35mm;
- F- Number = 4; and
- Camera Gain Control = +6dB.

Recently, a stereo vision system was added to the SPARCS system. In the SPARCS system an additional SONY XC-77 CCD camera and OLYMPUS 35-70 motorized lens were installed. Images from the camera are captured by a video frame capture card of VideoPix , which resides in a SUN SPARC station. These images are processed in the SUN SPARC station. Details for the SPARCS system are described in [70].

Chapter 3

Theory of Focus

3.1 Introduction

Focus is a very important cue for determining the distance of objects from the lens. This chapter provides the necessary background for chapters 4, 5, and 6. It describes two theoretical models which have been considered for focusing techniques in the literature. It then discusses depth-from-focus and depth-from-defocus.

3.2 Camera Model

The camera model used here follows the model proposed in [57].

Image formation in a simple camera is shown in Figure 3.1. We have shown a thin lens model for the optical system, but the analysis here can be easily extended to a thick lens model [17, 18]. In the thin lens model, the first and second principal planes of the optical system coincide and lie at the lens. Let p be a point on a visible surface in the scene and p' be its focused image. The relation between the positions of p and p' is given by the lens formula,

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{3.1}$$

where u is the distance between the first principal plane and the object plane and v is the distance between the second principal plane and the image plane. In this figure, ID is the image detector (CCD array), D is the aperture diameter, and s is the distance between the second principal plane and the image detector. The distance s, focal length f, and aperture diameter D, will be referred to together as *camera parameters* and denoted by $\mathbf{e}$, i.e.

$$\mathbf{e} \equiv (s, f, D). \tag{3.2}$$

3.2.1 Geometric Optics Model

In Figure 3.1, if the object point p is not in focus, it gives rise to a blurred image p'' on the image detector ID. According to geometric optics, the blurred image of p has the same shape as the lens aperture but is scaled by a factor q. This holds irrespective of the position of p on the object plane. Since we have taken the aperture to be circular, the blurred image of p is also a circle with uniform brightness inside the circle and zero outside. This is called a *blur circle.*

Let the light energy incident on the lens from the point p during one exposure period of the camera be one unit. Then, the blurred image of p is the response of the

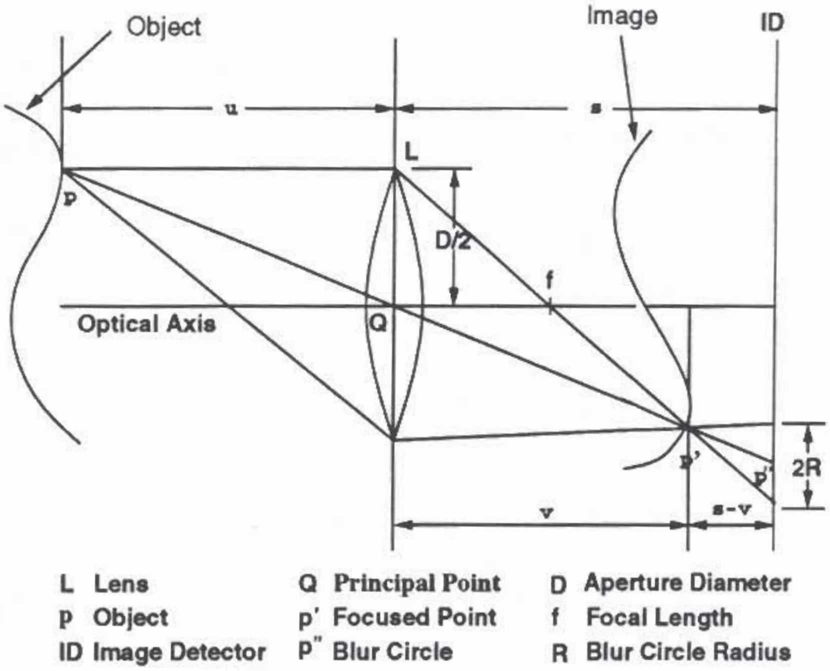


Figure 3.1: Image Formation in a Convex Lens.

camera to a unit point source and hence it is the Point Spread Function (PSF) of the camera system [63, 64]. We will denote this PSF by $h_a(x, y)$.

Let R be the radius of the blur circle and q be the scaling factor defined as $q = 2R/D$. In Figure 3.1, from similar triangles, we have

$$q = \frac{2R}{D} = \frac{s-v}{v} = s\left[\frac{1}{v} - \frac{1}{s}\right]. \tag{3.3}$$

Substituting for $1/v$ from Eq. (3.1) in the above equation, we obtain

$$q = s\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right]. \tag{3.4}$$

Therefore

$$R = q\frac{D}{2} = s\frac{D}{2}\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right]. \tag{3.5}$$

Note that q and therefore R can be either positive or negative depending on whether $s \geq v$ or $s < v$. In the former case the image detector plane is behind the focused image of p and in the latter case it is in front of the focused image of p.

If we assume the camera is a lossless system (i.e., no light energy is absorbed by the camera system) then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_a(x,y)\, dx\, dy \;=\; 1 \tag{3.6}$$

because the light energy incident on the lens was taken to be one unit. Using this and the fact that the blur circle has uniform brightness inside a circle of radius R and zero outside, we get

$$h_a(x,y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

In a practical camera system, if a sequence of images $g_i(x,y)$ are taken at camera parameter settings of $\mathbf{e}_i$ for $i = 1, 2, 3, \cdots$, then the image magnification and mean image brightness may change even though nothing has changed in the scene. For example, moving the lens away from the image detector will increase image magnification and changing the aperture diameter changes mean image brightness. In order to define and compare focus measures for this sequence of images in a correct and consistent manner, the sequence of images must be normalized with respect to these factors first. Normalization with respect to image brightness is carried out by dividing the image brightness at every point by the mean brightness of the image.

Normalization with respect to image magnification is more complicated. It can be done by image interpolation and resampling such that all images correspond to the same field of view [51]. The relation between an original image $g(x,y)$ taken with $s = s_0$ and the corresponding magnification normalized image $g_n(x,y)$ is given by $g_n(x/s_0, y/s_0) = g(x,y)$. However, in most practical applications, the magnification change is less than 3% and can be ignored. This is probably the reason why most of the previous literature fails to mention the magnification correction. But this cannot be overlooked from a theoretical point of view.

In the following discussion we assume that the images have been normalized. Without loss of generality, we assume that both the mean brightness and magnification have been normalized to be 1. After magnification normalization, the radius of the blur circle can be expressed as a function of the camera parameter setting **e** and object distance u as

$$R(\mathbf{e};u) \;=\; \frac{D}{2}\left(\frac{1}{f}-\frac{1}{u}-\frac{1}{s}\right). \qquad (3.8)$$

The Optical Transfer Function (OTF) corresponding to the above PSF (Eq. (3.7)) is

$$H_a(\omega,\nu;\mathbf{e},u) \;=\; 2\frac{J_1(R(\mathbf{e};u)\,\rho(\omega,\nu))}{R(\mathbf{e};u)\,\rho(\omega,\nu)} \qquad (3.9)$$

where ω, ν, and ρ are spatial frequencies specified in radians/unit distance, J_1 is the first order Bessel function, and ρ is the radial spatial frequency

$$\rho(\omega,\nu) \;=\; \sqrt{\omega^2+\nu^2}. \qquad (3.10)$$

Eq. (3.9) explicitly represents the dependence of the OTF on the camera parameter setting $\mathbf{e}$ and the object distance u.

In practice, the image of a point object is not a crisp circular patch of constant brightness as suggested by geometric optics. Instead, due to diffraction, polychromatic illumination, lens aberrations, etc., it will be a roughly circular blob with the brightness falling off gradually at the border rather than sharply. Therefore, as an alternative to the above PSF model [24, 39, 44, 49], a two-dimensional Gaussian is often suggested. It is defined by

$$h_b(x,y) \;=\; \frac{1}{2\pi r^2} e^{-\frac{x^2+y^2}{2r^2}} \tag{3.11}$$

where r is a spread parameter corresponding to the *standard deviation* of the distribution of the PSF. In practice, it is found that r is proportional to R [50, 52], i.e.

$$r \;=\; c\,R \;\text{ for }\; c > 0 \tag{3.12}$$

where c is a constant. It is approximately equal to $1/\sqrt{2}$ in practice [52]. Since the blur circle radius R is a function of $\mathbf{e}$ and u, r can be written as $r(\mathbf{e};u)$. (However, the image of an actual point light source for our camera was quite close to that predicted by geometric optics and was far from a Gaussian.)

The OTF corresponding to the above PSF is (ω, ν in radians/unit dist.)

$$H_b(\omega,\nu;\mathbf{e},u) \;=\; e^{-\frac{1}{2}\rho^2(\omega,\nu)\, r^2(\mathbf{e};u)} \tag{3.13}$$

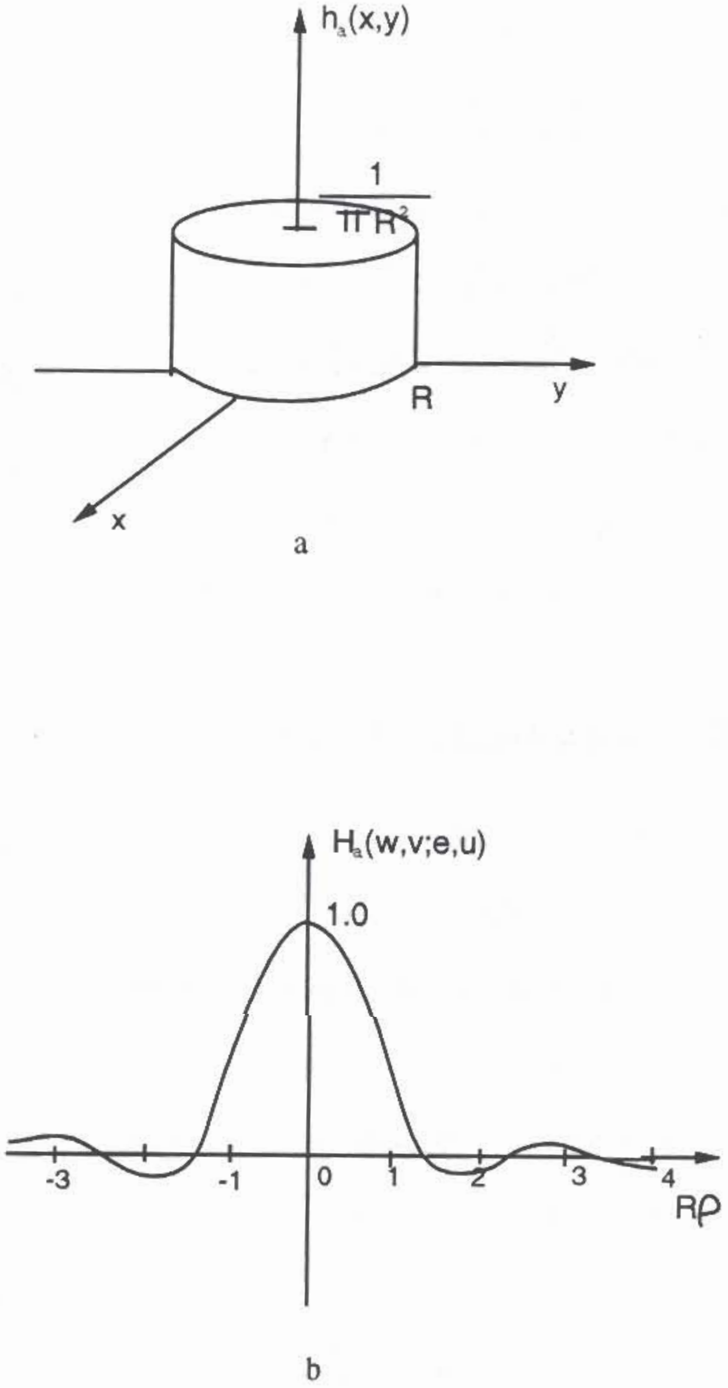


Figure 3.2: Geometric Optics Model, a: Point Spread Function $h_a(x, y)$ b: Optical Transfer Function $H_a(\omega, \nu; \mathsf{e}, u)$.

where

$$r(\mathbf{e};u) \;=\; c\,\frac{D}{2}\left(\frac{1}{f}-\frac{1}{u}-\frac{1}{s}\right). \tag{3.14}$$

Once again, Eq. (3.13) explicitly represents the dependence of the OTF on the camera parameter setting **e** and the object distance u.

In this thesis, two PSF models–Circular Disc and Gaussian– are used to illustrate the problems associated with defining good focus measures. Similar problems arise in the case of a PSF model based on diffraction optics. The focus measures proposed in chapter 4 are applicable to all the three PSF models. In most machine vision applications, Circular Disc is an adequate approximation to the actual PSF.

3.2.2 Diffraction Optics Model

In this section, the Optical Transfer Function of the diffraction optics model is introduced. Most of the results here follow from Hopkins [22], Born and Wolf [7], Subbarao [53, 54], and Ens [14]. Readers are referred to these references for the detailed derivation.

For a defocused aberration free optical system, Hopkins [22] presented the normalized response function as

$$D(s) = \frac{1}{\pi}\int\int_{s} exp(jax)dx\,dy \tag{3.15}$$

where $a = \frac{4\pi}{\lambda}w \mid s \mid$ and $w = \frac{(q-v)\sin^2\alpha}{2}$.

s denotes the region of integration. It is defined by

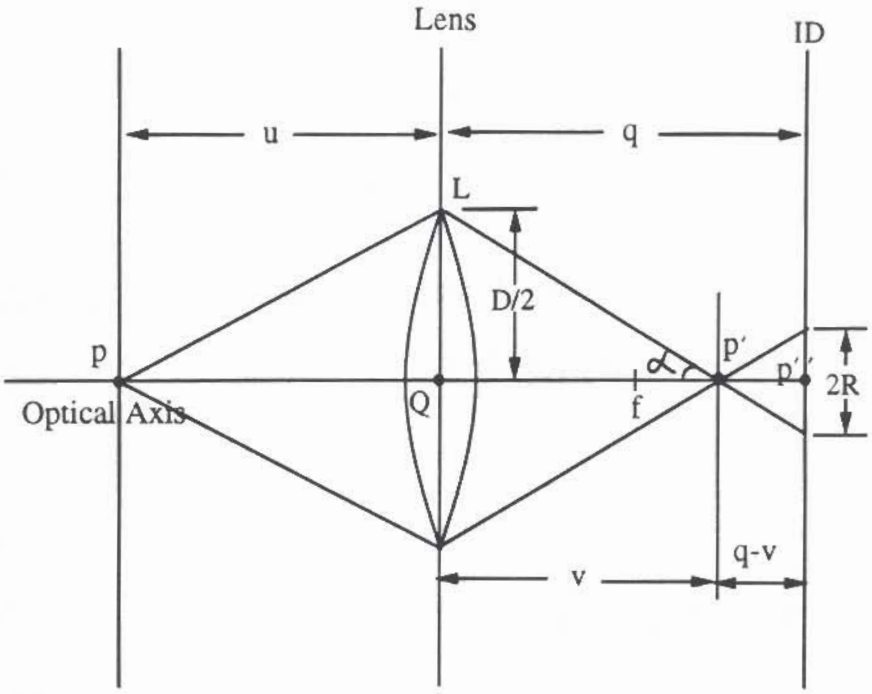


Figure 3.3: Image Formation for the diffraction optics model.

$$s = \frac{\lambda \rho}{\sin \alpha} \tag{3.16}$$

where ρ is the radial spatial frequency and λ is the wavelength.

Because of the symmetry of this region, Eq. (3.15) becomes

$$D(s) = \frac{4}{\pi a} \int_0^{\sqrt{1-(\frac{s}{2})^2}} \sin a \left(\sqrt{1-y^2} - \frac{|s|}{2} \right) dy. \tag{3.17}$$

Hopkins also gives a convenient form for numerical evaluation in terms of Bessel functions as follow.

$$D(s) \;=\; \frac{4}{\pi a} \cos \frac{a\,|s|}{2} [\beta J_1(a) + \frac{1}{2} \sin 2\beta (J_1(a) - J_3(a)) - \frac{1}{4} \sin 4\beta (\tag{3.18}$$

$$J_3(a) - J_5(a)) + \cdots] - \frac{4}{\pi a} \sin \frac{a \mid s \mid}{2} [\sin \beta (J_0(a) - J_2(a)) - \frac{1}{3} \sin 3\beta ($$
$$J_2(a) - J_4(a)) + \frac{1}{5} \sin 5\beta (J_4(a) - J_6(a)) - \cdots]$$

where $\beta = \cos^{-1} \frac{|s|}{2}$.

The corresponding Point Spread Function is obtained by the inverse Fourier transform of Eq. (3.18). See Born and Wolf [7], Subbarao and Lu [63], and Lu [34] for PSF expressions.

3.3 Depth from Focus

The depth-from-focus (DFF) method involves a search of the camera parameter space. The camera parameters include the distance between the lens and the image detector, and the focal length of the lens.

Many DFF methods have been proposed in the literature [19, 23, 25, 27, 31, 36, 57, 59, 69]. In these techniques, a focus measure is defined which is a maximum for the best focused image and it generally decreases as the defocus increases. Therefore, when the camera lens is moved from near to the far end relative to the image detector, typically the focus measure of the target object's image gradually increases, reaches a maximum at the focused lens position, and then decreases gradually thereafter. The problem then is to find the lens position at which the focus measure is a maximum. This is essentially a search of the lens position space.

We proposed a set of DFF techniques based on the paraxial geometric optics model

[57]. Details for our methods are described in chapter 4.

3.4 Depth from Defocus

The depth-from-defocus (DFD) methods do not require a sequence of images, but only a few (about 3) acquired with different camera parameter values. These methods are fast, but less accurate. Their best performance gives an RMS error which is twice that of the DFF methods. Many methods are proposed in the literature including Fourier domain approach [38, 39, 65], spatial domain approach [61], matrix based method [15], pyramid based method [8, 10], etc.

We introduce two typical methods [61, 65] developed by other people in our laboratory. The following parts appeared in a research proposal by Professor M. Subbarao.

3.4.1 Fourier Domain Approach

A brief summary of the Fourier domain approach is given in this section.

The image formation in a camera with variable camera parameters (s, f, D) is shown in Figure 3.1. The normalized radius of the blur circle is given by Eq. (3.5).

For the sake of conceptual simplicity, let the Point Spread Function (PSF) of the camera be a two-dimensional Gaussian:

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.19}$$

where the spread parameter σ is proportional to the blur circle radius R. Therefore we can write

$$\sigma = cR \tag{3.20}$$

where c is a camera constant.

In this camera model, a blurred image g of a planar object at distance u having focused image f is given by the convolution of the PSF h and the focused image f, i.e. $g = h * f$. Therefore, if G and F are Fourier transforms of g and f respectively , we have

$$\begin{aligned} G(\omega,\nu) &= H(\omega,\nu)F(\omega,\nu) & (3.21)\\ &= e^{-\frac{1}{2}(\omega^2+\nu^2)\sigma^2}F(\omega,\nu). & (3.22)\end{aligned}$$

For two blurred images g_1, g_2, taken with two different camera settings $\mathbf{e_1}, \mathbf{e_2}$, we obtain

$$\frac{G_1(\omega,\nu)}{G_2(\omega,\nu)} = e^{-\frac{1}{2}(\omega^2+\nu^2)(\sigma_1^2-\sigma_2^2)} \tag{3.23}$$

or,

$$\sigma_1^2 - \sigma_2^2 = \frac{-2}{\omega^2+\nu^2} ln \frac{\mid G_1(\omega,\nu) \mid}{\mid G_2(\omega,\nu) \mid}. \tag{3.24}$$

Further, from Eqs. (3.5),(3.20) σ_1 and σ_2 can be expressed as

$$\sigma_1 = c\frac{D_1}{2}\left(\frac{1}{f_1} - \frac{1}{u} - \frac{1}{s_1}\right), \tag{3.25}$$

$$\sigma_2 = c\frac{D_2}{2}\left(\frac{1}{f_2} - \frac{1}{u} - \frac{1}{s_2}\right). \tag{3.26}$$

Eliminating $1/u$ from the above realtions [62] we obtain

$$\sigma_1 = \alpha\sigma_2 + \beta \tag{3.27}$$

where

$$\alpha = \frac{D_1}{D_2}, \quad \beta = cD_1\left(\frac{1}{f_1} - \frac{1}{f_2} + \frac{1}{s_2} - \frac{1}{s_1}\right). \tag{3.28}$$

Substitute Eq. (3.27) into Eq. (3.24).

$$(\alpha^2 - 1)\sigma_2^2 + 2\alpha\beta\sigma_2 + \beta^2 = \frac{-2}{\omega^2 + \nu^2} ln \frac{\mid G_1(\omega, \nu) \mid}{\mid G_2(\omega, \nu) \mid}. \tag{3.29}$$

In the above equation σ_2 is the only unknown. The equation is quadratic and therefore σ_2 is easily obtained by solving it. The two solutions result in a two-fold ambiguity. Methods for resolving this ambiguity are discussed in [65]. From the solution for σ_2, the distance u of the object is obtained from Eqs. (3.25),(3.26).

The above discussion illustrates the conceptual feasibility of determining distance from two defocused images. By repeating the above procedure in all image neighborhoods, the depth-map of an entire scene can be obtained from only a few blurred images in parallel. The Gaussian PSF model results in closed-form solution. However, in general, a closed-form solution cannot be obtained for the actual PSF of camera. A numerical method has to be used. Details for the case of arbitrary PSF are presented in [61, 65].

3.4.2 Spatial Domain Approach

In this section we summarize the spatial domain approach to the DFD module. A spatial domain convolution/deconvolution transform (S Transform) was proposed in [55].

For two-dimensional images, under a local cubic polynomial model, it is defined as follows. If a focused image f is blurred by convolution with a circularly symmetric PSF h to result in the blurred image g, then g is the forward S transform of f with respect to the kernel function h, and it is given by

$$g(x,y) = f(x,y) + \frac{h_2}{2}\nabla^2 f(x,y) \tag{3.30}$$

where h_2 is the second moment of h with respect to x or y, i.e.

$$\begin{aligned} h_2 &= \int\int x^2 h(x,y)dx\,dy && (3.31)\\ &= \int\int y^2 h(x,y)dx\,dy && (3.32)\end{aligned}$$

and ∇^2 is the *Laplacian* operator.

The inverse S transform of g with respect to the moment vector $(1,h_2)$ is equal to f and it is defined as

$$f(x,y) = g(x,y) - \frac{h_2}{2}\nabla^2 g(x,y). \tag{3.33}$$

For a Gaussian PSF model used in the previous section, it can be shown that $h_2 = \sigma^2/2$. Therefore, if two blurred images g_1 and g_2 are acquired with different camera settings e_1 and e_2 corresponding to blur parameters σ_1 and σ_2 we obtain

$$\begin{aligned} f &= g_1 - \frac{\sigma_1^2}{4}\nabla^2 g_1 && (3.34)\\ &= g_2 - \frac{\sigma_2^2}{4}\nabla^2 g_2. && (3.35) \end{aligned}$$

From Eqs. (3.27), (3.34), (3.35), and the fact that $\nabla^2 g_1 = \nabla^2 g_2$ we obtain

$$(\alpha^2 - 1)\sigma_2^2 + 2\alpha\beta\sigma_2 + \beta^2 = \frac{4(g_1 - g_2)}{\nabla^2 g_1} \tag{3.36}$$

where α and β are as defined in the previous section. Except for the right hand side, the above relation is similar to Eq. (3.29).

Therefore we have a quadratic equation in σ_2 which can be easily solved. The distance u can be obtained from σ_2 from Eqs. (3.25), (3.26).

Chapter 4

Focusing

4.1 Introduction

Focusing cameras is an important problem in computer vision and microscopy. In this chapter we consider only passive focusing techniques which do not require calibration of the cameras. The techniques considered here necessarily involve a search of the camera parameter space. Therefore we will call these *search focusing techniques.* The camera parameters include the distance between the lens and the image detector, and the focal length of the lens. There are some focusing methods which do not search the camera parameter space [39, 61, 65], but they require accurate calibration of the camera parameters and the corresponding optical transfer function or the point spread function. We first consider the case of focusing a camera to a target object by moving the lens along the optical axis of the lens. The results of this case can be easily extended to focusing a camera by adjusting its focal length, or by moving the target object along the optical axis.

In the previous literature, most definitions of focus measure have been based on heuristics. No proof has been provided about the theoretical soundness of these focus measures, i.e. it has not been proved theoretically that the focus measures have a global maximum for the best focused images. Some definitions of focus measures involve non-linear operators and it is hard to prove or disprove their correctness. Some other definitions are useful only for a very limited class of images. We have not found any systematic and explicit analysis of the image defocusing phenomenon and derivation of a focus measure based on such an analysis. Further, previous literature has ignored an important aspect we call the "side-lobe effect" (explained later) which may give rise to local maxima and cause the global maximum position to be shifted. This makes searching for the position of global maximum difficult and introduces error in the determination of focus position.

In this section, we derive a class of focus measures which we prove to be theoretically sound under weak assumptions, i.e. the focus measures have global maximum for the best focused images. The assumptions correspond to neglecting the side lobes (if they exist) in the optical transfer function of the camera system. We then analyze the side-lobe effect and show how it can be reduced through low pass filtering. Next we discuss a few other focusing techniques from previous literature.

The focus measures derived in this section have been implemented on the SPARCS camera system. A number of experiments were carried out using SPARCS to evaluate the focus measures derived here. The experiments and their results are described. The experimental results show that the focus measures derived here perform well. Based on these results, we find that, one of the focus measures– the *energy of low-pass filtered image gradient* – has the best performence overall. We conclude this section by recommending it for practical applications.

In order to illustrate the theoretical basis of focus measures, we take the optical system to be circularly symmetric around the optical axis and we use a paraxial geometric optics model [17] for image formation. This is a good approximation in practice to actual image formation process modeled by physical optics [52, 63]. However, our analysis itself is applicable to physical optics model also.

4.2 Focus Measures

In this section, we develop a theoretical basis for focus measures. Let $f(x,y)$ be the focused image of a planar object at distance u. The *focused image* $f(x,y)$ at a point (x,y) of a scene is defined as the total light energy incident on the camera aperture (entrance pupil) during one exposure period from the object point along the direction corresponding to (x,y) [64]. We do not know of any previous literature on focusing techniques which gives a precise and correct (we believe) definition of the focused image as we have done here. This definition is essential for a sound analysis of the focusing techniques.

Let $g_i(x, y)$ be a sequence of images of an object recorded for a sequence of camera parameter settings $\mathbf{e}_i$ where

$$\mathbf{e}_i \;=\; (s_i, f, D) \text{ and } s_i < s_{i+1} \text{ OR } \mathbf{e}_i \;=\; (s, f_i, D) \text{ and } f_i < f_{i+1} \tag{4.1}$$

for $i = 1, 2, 3, \cdots$. Here we are considering the variation of only one camera parameter at a time– either the lens position or the focal length because this is the usual mode of operation of almost all cameras. Simultaneous variation of more than one parameter in a random manner can lead to multiple maxima for the focus measure. For the image sequence g_i, the distance u of the object is fixed. A focus measure M computed for an image g_i will be denoted by $M(i)$.

For a planar object perpendicular to the optical axis, the blur circle radius R is a constant over the image of the object (this may not be obvious at first sight, but it can be proved easily). In this case the camera acts as a linear shift invariant system. Therefore, g_i will be equal to the convolution of the focused image $f(x, y)$ with the corresponding point spread function. Convolution in the spatial domain corresponds to multiplication in the Fourier domain. Therfore, if F and G_i are Fourier transforms of f and g_i respectively, we can write

$$G_i(\omega, \nu) \;=\; H_a(\omega, \nu; \mathbf{e}_i, u)\; F(\omega, \nu). \tag{4.2}$$

Substituting for the right hand side of Eq. (4.2) from Eq. (3.9), we obtain

$$G_i(\omega, \nu) \;=\; 2\frac{J_1(R(\mathbf{e}_i; u)\; \rho(\omega, \nu))}{R(\mathbf{e}_i; u)\; \rho(\omega, \nu)}\; F(\omega, \nu). \tag{4.3}$$

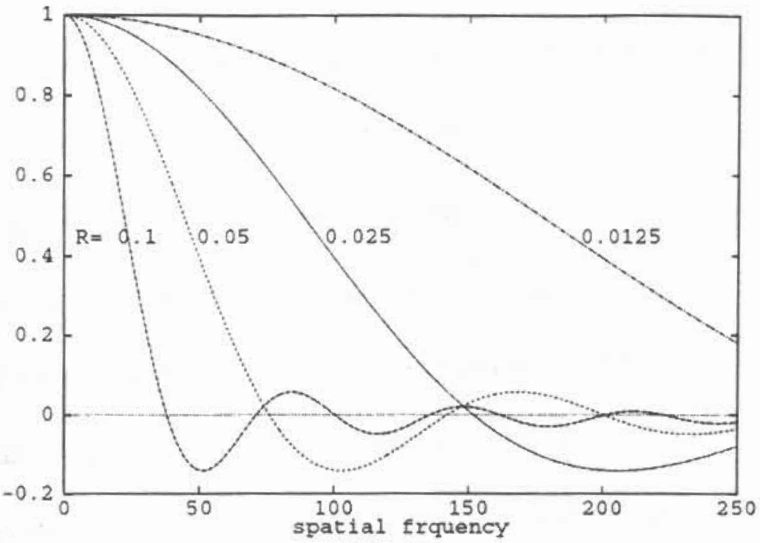


Figure 4.1: Cross sections of OTFs $2\frac{J_1(R\rho)}{R\rho}$.

The left hand side of the above equation can be computed from the recorded images g_i. For a Gaussian PSF model, the expression corresponding to Eq. (4.3) is

$$G_i(\omega,\nu) \;=\; e^{-\frac{1}{2}\rho^2(\omega,\nu)\, r^2(\mathbf{e}_i;u)}\, F(\omega,\nu). \tag{4.4}$$

Given the above defocusing model, the problem now is to define a focus measure which is a maximum for the best focused image among the sequence of images $g_i(x,y)$ and gradually decreases as the image blur increases. Except for the image sequence g_i which forms the input, no other information is assumed to be known such as the actual values of the camera parameters $\mathbf{e}_i$, or the optical transfer function.

Figure 4.1 shows cross sections of the circularly symmetric optical transfer functions H_a for various values of the blur circle radius R. These OTFs are sinc-like with a dominant main lobe followed by side lobes on either side. The OTF magnitude at the origin is 1 and decreases monotonically to zero at $\rho \approx 1.22\pi/R$ at the end of the main lobe. The first side lobe peak is approximately at $1.63\pi/R$ and has a value of about 0.132. The amplitude of the side lobes fall off rapidly (as $\rho^{-1.5}$ [24]) with ρ. If we consider the effect of only the dominant main lobe, we see that the OTFs have the general characteristics of a low pass filter. As the blur circle radius increases (i.e. blur increases), the higher frequencies are attenuated more. In the main lobe, the higher a frequency, the more the attenuation. The area under the main lobe increases with decreasing defocus. If the OTF is assumed to be zero in the side lobes, then an obvious focus measure suggested by the behavior of the main lobe is the volume integral

$$M_0(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |G_i(\omega,\nu)| d\omega d\nu. \tag{4.5}$$

This focus measure is proved to be sound, monotonic, and unimodal with i in Theorem 2 in the appendix assuming the OTF magnitude outside the main lobe to be zero. Theorem 3 proves the same result for the case of a Gaussian OTF. In this case, no assumptions are made as the OTF has no side lobes. Further, the maximum value of this focus measure is shown to be

$$M_{0max} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F(\omega,\nu)| d\omega d\nu \tag{4.6}$$

when the blur circle radius is zero (i.e. when the observed image is in perfect focus).

This focus measure involves the computation of the Fourier transform of the image which is computationally expensive ($O(N^2 \log N)$ for an image of size $N \times N$). Fortunately, we can define other focus measures which can be computed more efficiently. For this purpose, we note that we can use the volume integral of any *function* of $|G_i(\omega,\nu)|$ which increases monotonically with increasing $|G_i(\omega,\nu)|$. This is proved in Theorems 4 and 5 in the appendix. Therefore, literally an infinite number of focus measures can be defined. However, based on computational requirements and noise sensitivity, we select only three measures. *Before we proceed to the three measures, we raise the question as to whether a sound focus measure can be defined which is not a monotonic function of* $|G_i(\omega,\nu)|$. *We believe that any such function would exhibit local maxima for some* $|F(\omega,\nu)|$.

4.3 Three Focus Measures

In this section, we define three focus measures which can be efficiently computed in the spatial domain.

4.3.1 Image Energy

$$M_1(i) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |G_i(\omega,\nu)|^2 d\omega d\nu. \tag{4.7}$$

This does not require the computation of Fourier transform of the image because it can be computed efficiently using the Parseval's theorem:

$$M_1(i) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g_i(x,y)|^2 dx dy. \tag{4.8}$$

According to Theorems 2 and 3 in the appendix, $M_1(i)$ is monotonic, unimodal, and sound.

4.3.2 Energy of image gradient

$$M_2(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\vec{\nabla} g_i(x,y)|^2 \, dxdy. \tag{4.9}$$

Note that

$$\mathcal{F}\left\{\frac{\partial g_i(x,y)}{\partial x}\right\} = j\omega G_i(\omega,\nu) \tag{4.10}$$

where $\mathcal{F}$ denotes the Fourier transform operator. Therefore, from Parseval's theorem,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial g_i(x,y)}{\partial x}\right)^2 dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \omega^2 |G_i(\omega,\nu)|^2 \, d\omega d\nu. \tag{4.11}$$

Similarly,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial g_i(x,y)}{\partial y}\right)^2 dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu^2 |G_i(\omega,\nu)|^2 \, d\omega d\nu. \tag{4.12}$$

Adding the above two relations, we get

$$M_2(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\omega^2+\nu^2\right) |G_i(\omega,\nu)|^2 \, d\omega d\nu. \tag{4.13}$$

Therefore, we see that $M_2(i)$ measures the image energy after $G_i(\omega,\nu)$ has been high pass filtered by a filter L_2 such that

$$\mid L_2(\omega,\nu) \mid = \sqrt{\omega^2+\nu^2} = \rho. \tag{4.14}$$

Therefore, according to Theorems 6 and 7 in the appendix, $M_2(i)$ is a monotonic, unimodal and sound focus measure.

4.3.3 Energy of image Laplacian

$$M_3(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\nabla^2 g_i(x,y)\right)^2 \, dxdy. \tag{4.15}$$

Note that

$$\mathcal{F}\left(\frac{\partial^2 g_i(x,y)}{\partial x^2}\right) = -\omega^2 G_i(\omega,\nu) \tag{4.16}$$

and

$$\mathcal{F}\left(\frac{\partial^2 g_i(x,y)}{\partial y^2}\right) = -\nu^2 G_i(\omega,\nu). \tag{4.17}$$

Therefore

$$\mathcal{F}\left(\nabla^2 g_i(x,y)\right) = -\left(\omega^2+\nu^2\right) G_i(\omega,\nu). \tag{4.18}$$

From Parseval's theorem

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\nabla^2 g_i(x,y)\right)^2\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\omega^2+\nu^2\right)^2 |G_i(\omega,\nu)|^2 d\omega d\nu. \tag{4.19}$$

Therefore, $M_3(i)$ measures the image energy after $G_i(\omega,\nu)$ has been high-pass filtered by a filter L_3 such that

$$\mid L_3(\omega,\nu)\mid = \left(\omega^2+\nu^2\right) \;=\; \rho^2. \tag{4.20}$$

See Figure 4.2 for a plot of $|L_2(\omega,\nu)|$ and $|L_3(\omega,\nu)|$. Therefore, according to Theorems 6 and 7 in the appendix, $M_3(i)$ is a sound, unimodal and monotonic focus measure.

4.4 Analysis of Side Lobe Effect

In the previous section and the appendix, we ignored the presence of side lobes in proving the monotonicity and the correctness of the focus measures. If the focused image of an object has high energy frequency content in the side lobes, then the focus measures discussed earlier may exhibit local maxima. This complicates the task of

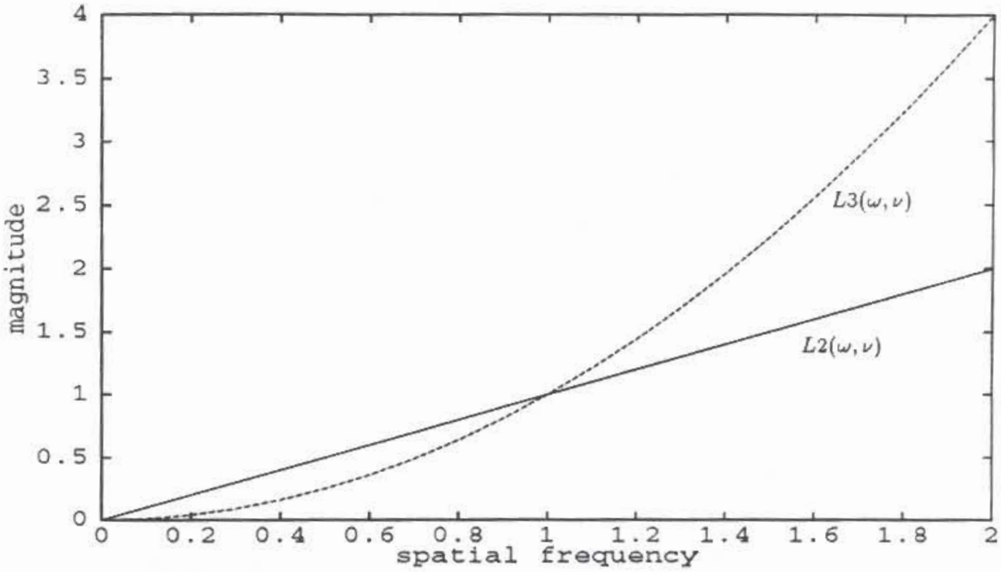


Figure 4.2: Plot of $L_2(\omega, \nu)$ and $L_3(\omega, \nu)$.

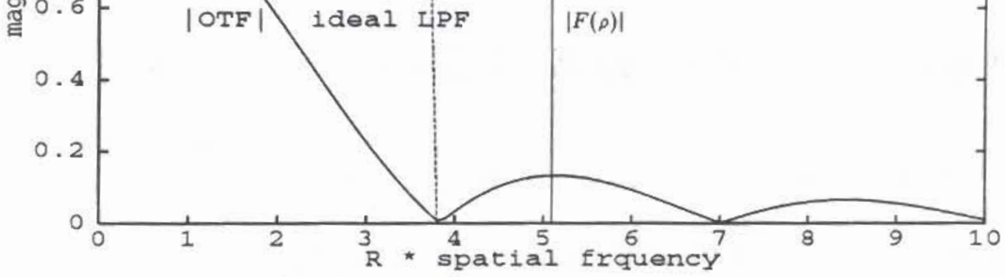


Figure 4.3: OTF and $|F(\rho)|$.

searching for the global maximum. In order to illustrate this, consider the following pathological case: an object whose Fourier spectrum is (see Figure 4.3)

$$|F(\rho)| = \delta\left(\rho - \frac{1.63\pi}{R}\right) \tag{4.21}$$

where δ is the *Dirac delta* function, i.e., the Fourier spectrum is zero everywhere except at the peak of the first side lobe. In this case, there is a local maximum for all the focus measures defined earlier. Both increasing and decreasing the blur circle radius causes the focus measures to decrease.

The above effect of the side lobe can be eliminated by an ideal Low Pass Filter (LPF) which removes the side lobes (see Figure 4.3). The cut off frequency ρ_c of the ideal LPF should be $1.22\pi/R_{max}$ where R_{max} is the maximum blur circle radius expected for the given operating range of the distance of the object. However, this solution for the side lobe effect is too drastic. The disadvantages of this solution are the computational cost of ideal lowpass filtering and the complete elimination of frequency content beyond the cutoff frequency ρ_c of the LPF. Such complete elimination of frequencies will make it impossible to focus on objects with non zero spectrum only for $\rho > \rho_c$. A better solution to the side lobe effect is to attenuate the side lobes relative to the main lobe rather than complete elimination of the side lobes. For this purpose, we suggest a Gaussian lowpass filter. It can be implemented economically by convolving the images in the spatial domain with a Gaussian having a small spread (about σ = 1.5 to 2.5 pixels). Convolution can be implemented efficiently as two one-dimensional operations, first along rows, then along columns, because Gaussian is seperable. If k is the factor by which we wish to attenuate the peak of the first side lobe, the parameter r for the Gaussian low-pass filter can be computed as follows:

$$\begin{aligned} e^{-\frac{1}{2}\rho^2 r^2} &= k && (4.22)\\ \Rightarrow r &= \frac{\sqrt{2}\,\sqrt{-\log\,k}}{\rho} && (4.23)\\ &\approx \frac{\sqrt{2}}{1.63\pi}\,R_{max}\,\sqrt{-\log\,k} && (4.24)\\ r &\approx 0.2762\sqrt{-\log\,k}\;R_{max}. && (4.25)\end{aligned}$$

If $k = 0.25$ and $R_{max} = 5$ pixels, then $r = 1.63$ pixels. The size of the one-dimensional Gaussian filter in this case is about 7.

Another advantage to low-pass filtering the image is the reduction of noise and the attenuation of aliased frequencies near the high end of the spectrum. Aliasing will be a problem if the focused image is not band-limited within the Nyquist frequency band.

The noise magnitude spectrum usually remains the same in all images of the image sequence g_i . In this case , if the effect of the side lobes is neglected, then the noise has no effect on the performance of the focus measures. The focus measures remain monotonic and unimodal. Further, if the effect of side lobes is negligible, then even aliasing has no effect on the performance of the focus measures. However if the frequency content in the side lobes is high due to noise, aliasing, or the focused image, then the focus measures may exhibit local maxima and the global maximum may be shifted. This makes the search for global maximum more difficult and introduces error in the final result. Therefore, the attenuation of side lobes improves the behaviour of the focus measures.

According to Theorems 6 and 7 in the appendix, low pass filtering the image sequence does not affect the soundness, monotonicity, or the unimodality of the focus measures.

4.5 Band-pass filtering

Suppose that the image sequence $G_i(\omega, \nu)$ is filtered by the Gaussian low-pass filter $H_b(\omega, \nu) = e^{-\frac{1}{2}(\omega^2+\nu^2)r^2}$. The Fourier magnitude spectrum [41, 42] of the resulting image sequence is $|H_b(\omega, \nu)|\, |G_i(\omega, \nu)|$. The focus measures $M_1'(i), M_2'(i)$ and $M_3'(i)$ are defined as

$$\begin{aligned}
M_1'(i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\, h_b(x,y) * g_i(x,y)\,|^2 \, dx\, dy && (4.26)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |H_b(\omega,\nu)|^2 \, |G_i(\omega,\nu)|^2 \, d\omega\, d\nu && (4.27)\\
M_2'(i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\, \vec{\nabla}(h_b(x,y) * g_i(x,y))\,|^2 \, dx\, dy && (4.28)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\omega^2+\nu^2)\, |H_b(\omega,\nu)|^2 \, |G_i(\omega,\nu)|^2 \, d\omega\, d\nu && (4.29)\\
M_3'(i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\nabla^2\, (h_b(x,y) * g_i(x,y))\right)^2 dx\, dy && (4.30)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\omega^2+\nu^2)^2\, |H_b(\omega,\nu)|^2 \, |G_i(\omega,\nu)|^2 \, d\omega\, d\nu. && (4.31)
\end{aligned}$$

From the Fourier domain expressions for $M_2'(i)$ above, $M_2'(i)$ can be thought of as the result of first filtering $G_i(\omega, \nu)$ by a filter, $B_2(\omega, \nu)$, having the Fourier Magnitude Spectrum $|B_2(\omega, \nu)| = \sqrt{(\omega^2 + \nu^2)}\; |H_b(\omega, \nu)|$ and then measuring the spectral energy of the resulting image. Similarly, $M_3'(i)$ can be thought of as the result of first filtering $G_i(\omega,\nu)$ by a filter $B_3(\omega,\nu)$ having the Fourier magnitude spectrum $|B_3(\omega,\nu)| = (\omega^2+\nu^2)|H_b(\omega,\nu)|$ and then measuring the spectral energy of the resulting image. A plot of $|B_2(\omega,\nu)|$ and $|B_3(\omega,\nu)|$ are shown in Figure 4.4. We see that both these filters have band-pass characteristics. Experimental results presented later show that these band-pass filters make the focus measures to have sharp peaks while generally retaining monotonicity and unimodality. These band-pass filters have the desirable characteristic of attenuating low frequencies which contribute less to change in the focus measure and attenuating high frequencies affected by side lobes and noise, but emphasizing medium frequencies.

4.6 Discussion of other focus measures

A number of focus measures have been proposed in the literature [25, 26, 27, 31]. Among these, Tenengrad [27, 43, 69] has been found to be the best. More recently, a focus measure based on a modified Laplacian operator (SML operator) is said to perform better than Tenengrad [36].

Tenengrad is a measure of thresholded gradient magnitude. It is similar to $M_2(i)$ except that only those points where the gradient magnitude is greater than a prespecified threshold are used in the calculation. The other points are not used in the calculation of the focus measure. Because of the thresholding operation, a Fourier domain filter analysis of this focus measure is not possible. More importantly, if

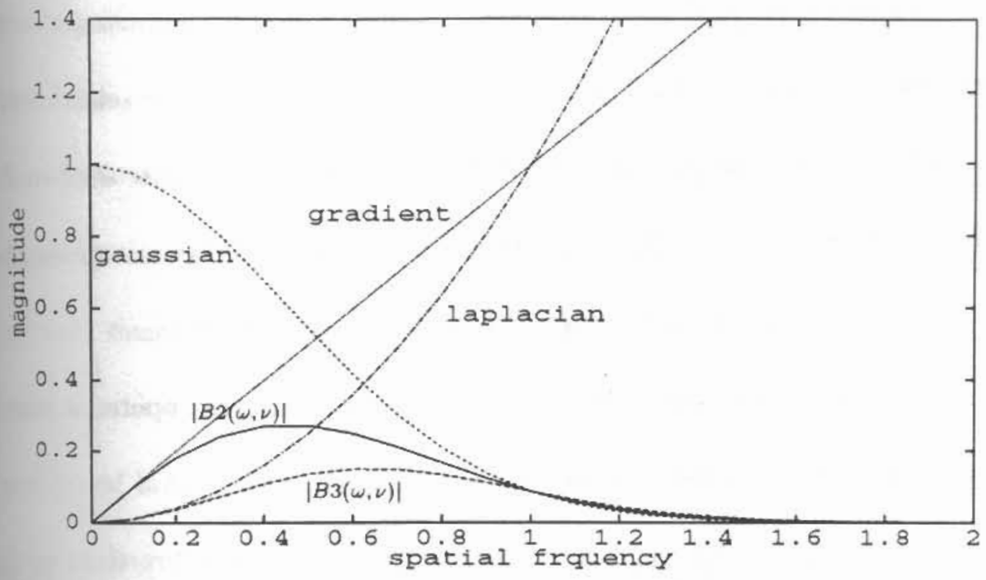


Figure 4.4: The magnitude responses of various fiters.

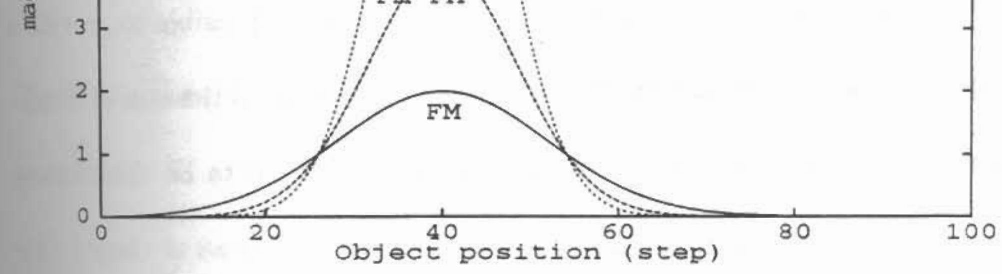


Figure 4.5: Sharpening a blunt focus measure by squaring and cubing.

the threshold is non-zero, this focus measure cannot be proved to be theoretically sound, i.e., it cannot be proved that the global maximum of the focus measure occurs for the best focused image. Moreover, this focus measure involves the selection of a threshold. For these reasons, and the fact that the focus measures discussed in this chapter ($M_1(i)$ to $M_3(i)$ and $M_1'(i)$ to $M_3'(i)$) performed very well in a large number of experiments, we do not recommend Tenengrad in actual applications.

Nayar [36] has proposed a new focus measure based on a new operator named *Sum-Modified-Laplacian* (SML). It is defined as :

$$\nabla^2{}_M\, g_i(x,y) \;=\; \left|\frac{\partial^2 g_i}{\partial x^2}\right| + \left|\frac{\partial^2 g_i}{\partial y^2}\right|. \qquad (4.32)$$

The SML based focus measure SMLF is defined as

$$\text{SMLF} \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\left|\frac{\partial^2 g_i}{\partial x^2}\right| + \left|\frac{\partial^2 g_i}{\partial y^2}\right|\right)\, dx\, dy. \qquad (4.33)$$

SML differs from the usual Laplacian in that the magnitude of the second derivatives are summed instead of their actual values. Therefore it is a non-linear operator and it is more sensitive to zero-mean noise than the Laplacian. The focus measure SMLF is a simple summation of the result of applying the SML operator; it does not involve squaring the integrand. For these reasons, it is not possible to provide a Fourier domain filter analysis of this focus measure. Further, as in the case of Tenengrad, we do not believe that this focus measure can be proved to be theoretically sound.

One of the main reason that Tenengrad and SMLF were said to be better focus measures in the past was that they gave sharper peaks. The sharper the peak produced by a focus measure the better it was thought because the location of the position of the maximum was thought to be more accurate. However it should be noted that a blunt peak of any focus measure can be sharpened by simply squaring the values of the focus measure. In fact, the peak can be sharpened to any desired degree by raising the values to some large positive power. (See Figure 4.5). Therefore, not only the sharpness of the peak, but also the smoothness (or monotonicity) and noise sensitivity are important. Sharpening the peak of a focus measure will magnify and reveal hidden local maxima caused by the side-lobe effect and noise. For bad focus measures, the amplitude of the local maxima will be large where as it will be small for good focus measures, thus reducing the uncertainty in locating the actual global maximum. Focus measures based on second order derivatives such as Laplacian and SML are more noise sensitive than those based on first order derivatives. In view of these observations, we believe that Tenengrad and SMLF have no particular advantage over some of the focus measures considered in this chapter.

4.7 Discrete focus measures

Discrete versions of the focus measures were implemented on SPARCS. The expressions for the $N \times N$ image $g_i(x,y)$ are given here. Each image was first normalized with respect to its mean. Magnification normalization was not done as the change in magnification was less than 2%.

1. *Variance*

Image energy can be computed as $\sum_x \sum_y g_i^2(x,y)$. A closely related focus measur is the grey level *variance*. It is a linear and monotonic function of image energy Variance is computed as

$$\begin{aligned} M_1(i) &= \frac{1}{N^2}\sum_x \sum_y (g_i(x,y) - \mu_i)^2 & (4.34) \\ &= \frac{1}{N^2}\sum_x \sum_y g_i^2(x,y) - \mu_i^2 & (4.35) \\ \text{where } \mu_i &= \frac{1}{N^2}\sum_x \sum_y g_i(x,y). & (4.36) \end{aligned}$$

2. *Energy of image gradient*

This focus measure is computed as

$$\begin{aligned} M_2(i) &= \sum_x \sum_y \left(g_x^2 + g_y^2\right) & (4.37) \\ \text{where } g_x(x,y) &= g_i(x+1,y) - g_i(x,y) & (4.38) \\ \text{and } g_y(x,y) &= g_i(x,y+1) - g_i(x,y). & (4.39) \end{aligned}$$

3. *Energy of Laplacian of the image*

$$M_3(i) = \sum_x \sum_y (g_{xx} + g_{yy})^2 \qquad (4.40)$$

where

$$g_{xx} + g_{yy} = \qquad (4.41)$$

$$\begin{aligned} &-g_i(x-1,y-1) - 4g_i(x-1,y) - g_i(x-1,y+1) \\ &-4g_i(x,y-1) + 20g_i(x,y) - 4g_i(x,y+1) \\ &-g_i(x+1,y-1) - 4g_i(x+1,y) - g_i(x+1,y+1). \end{aligned}$$

4. *Variance of low-pass filtered image*

The image was first low-pass filtered by convolution with a two-dimensional Gaussian. Since the Gaussian is a separable filter, it was implemented as two one-dimensional convolutions, first along rows and then along columns.

$$
\begin{aligned}
g_i'(x,y) &= \sum_p \sum_q h_b(p,q)\; g_i(x-p, y-q) && (4.42)\\
&= \sum_p \sum_q h_b'(p) h_b'(q)\; g_i(x-p, y-q) && (4.43)\\
&= \sum_p h_b'(p) \sum_q h_b'(q) g_i(x-p, y-q). && (4.44)
\end{aligned}
$$

The variance of this smoothed image is computed similar to $M_1(i)$:

$$M_1'(i) = \frac{1}{N^2} \sum_x \sum_y g_i'^2(x,y) - \mu_i'^2 \tag{4.45}$$

where

$$\mu_i' = \frac{1}{N^2} \sum_x \sum_y g_i'(x,y).$$

5. *Energy of low-pass filtered image gradient*

The original image g_i was first low-pass filtered as in the previous case to obtain g_i'. Then $M_2'(i)$ was computed similar to $M_2(i)$ except that g_i' was used in place of g_i.

6. *Energy of Laplacian of the low-pass filtered image*

$M_3'(i)$ was computed similar to $M_3(i)$ except that the low-pass filtered image g_i' was used in place of the original image g_i.

4.8 Experimental Results

Experiments were conducted on a large number of test objects. The results on three of these objects will be presented. The three objects are (i) a cartoon (Figure 4.11), (ii) picture of a face (Figure 4.12), and (iii) a "pin-hole" light source (Figure 4.13). The last object has high spatial frequency content. It was chosen to observe the effect of side lobes on the focus measures. The illumination for the first two objects was about 500 lux.

Each object was placed at different distances (cartoon: 820 mm, face: 950 mm, pin-hole: 1320 mm) in front of the camera and the program was run. SPARCS acquired one image of the object at each lens position. There are 97 lens positions corresponding to 97 steps of the lens stepper motor. From each image thus acquired, a 64×64 subimage of the object was extracted. Due to blurring and spreading of light from point objects, the grey levels at the border of this subimage are affected by image points immediately outside the subimage. This is called the *image overlap problem* [49]. In order to reduce this border effect, the images were multiplied by a two-dimensional Gaussian with a spread parameter of about 1/3 of the image size (about 21).

Then all the 6 focus measures were computed and printed. This procedure was repeated for each of the three objects. The 6 focus measures were normalized to have the same peak values by dividing their values by their maximum values. The results are plotted in Figures 4.6, 4.7, and 4.8. In the plots we see that all focus measures reach a peak at almost the same location. The only difference lies in the sharpness of peaks and the smoothness of the plotted curves.

The percentage change in a focus measure at any given lens position is useful in determining the direction in which the lens should be moved for focusing. Plots of this measure for Object 2 are shown in Figure 4.9 and Figure 4.10. It was computed as

$$\%\ change = \frac{M(i+step) - M(i)}{\frac{1}{2}[M(i+step) + M(i)]} \tag{4.46}$$

where step was set to 5 and $M(i)$ is a focus measure.

In comparing the different focus measures, we use the following criteria (i) monotonicity, (ii) magnitude of slope, and (iii) smoothness. It should be emphasized that good behaviour with respect to these criteria are important both when the images are highly blurred and when the images are almost focused. It is in the highly blurred images that the side lobe effect becomes significant. In the previous literature, the behaviour with respect to highly blurred images has been ignored in the evaluation of focus measures. This aspect is important because, when an object is highly blurred, it is necessary to first determine the direction in which the lens should be moved in order to focus the object. This is done by first computing the focus measure for the current lens position, then moving the lens by a small amount, computing the focus measure again, and comparing the two focus measures. The direction in which the focus measure increases is the direction in which the lens should be moved for focusing. Therefore, if a focus measure is almost flat (i.e. slope is small) or is non-monotonic and noisy, then the direction of lens motion for focusing cannot be determined reliably.

From Figures 4.6, 4.7, 4.8, and from many more experiments not reported here, we make the following observations. The focus measures M_1 and M_1' are smooth but generally exhibit small slopes for both highly blurred and almost focused images. Both are very good focus measures but perhaps not the best, mainly because of small slope. The focus measures M_2 and M_3 measure energy of high-pass filtered images. High-pass filtering amplifies the side lobe effect. Our experiments support this observation based on theory. Both M_2 and M_3 exhibit almost flat and noisy curves for highly blurred images. They exhibit sharp peaks near the focused position, but the peaks are sometimes noisy and rough. The focus measures M_2' and M_3' measure energy of band-pass filtered images. They are found to be generally smooth and monotonic. They exhibit moderately high slopes for both highly blurred and almost focused images. Their peaks are reasonably sharp. The behaviour of M_2' and M_3' are somewhere near the middle of the two extremes represented by M_1 and M_1' on the one end and M_2 and M_3 on the other.

Both M_2' and M_3' are very good focus measures. M_2' is somewhat smoother than M_3'. For this reason, we believe M_2' to be a better focus measure and therefore we recommend it for practical applications.

4.9 Discussion

We have derived 6 focus measures and proved that they are sound under weak assumptions. Experiments show that all of them perform well. By computing more than one focus measure and making judgements based on them, better performance may be obtained at a higher computational cost. In chapter 5, we will discuss the use of the focus measures derived here in depth-from-focus and shape-from-focus algorithms.

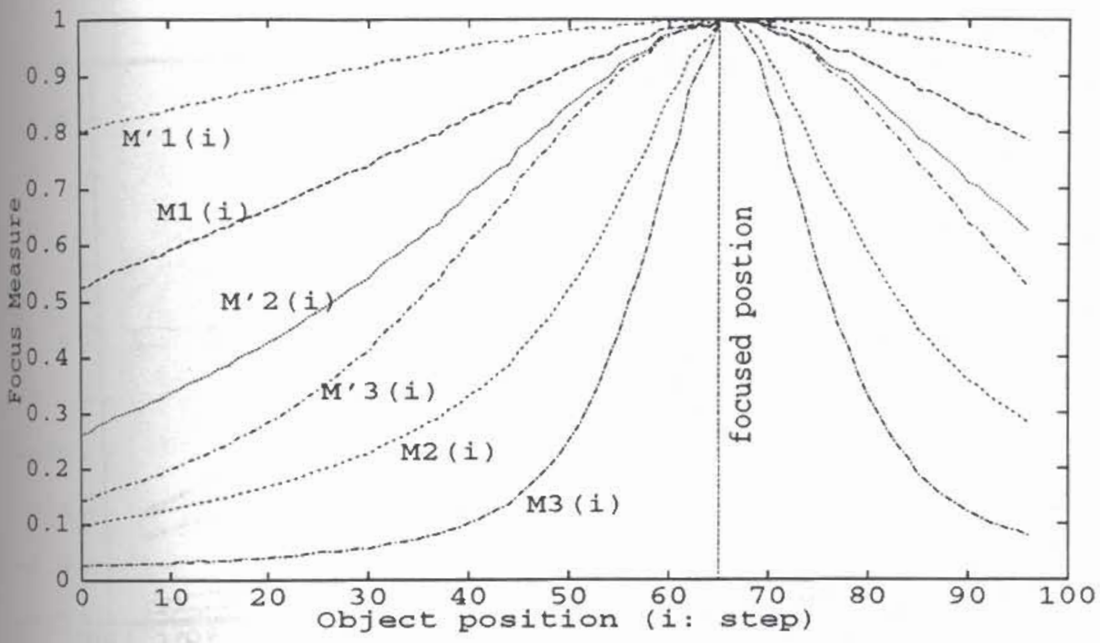


Figure 4.6: Focus measures for Object 1.

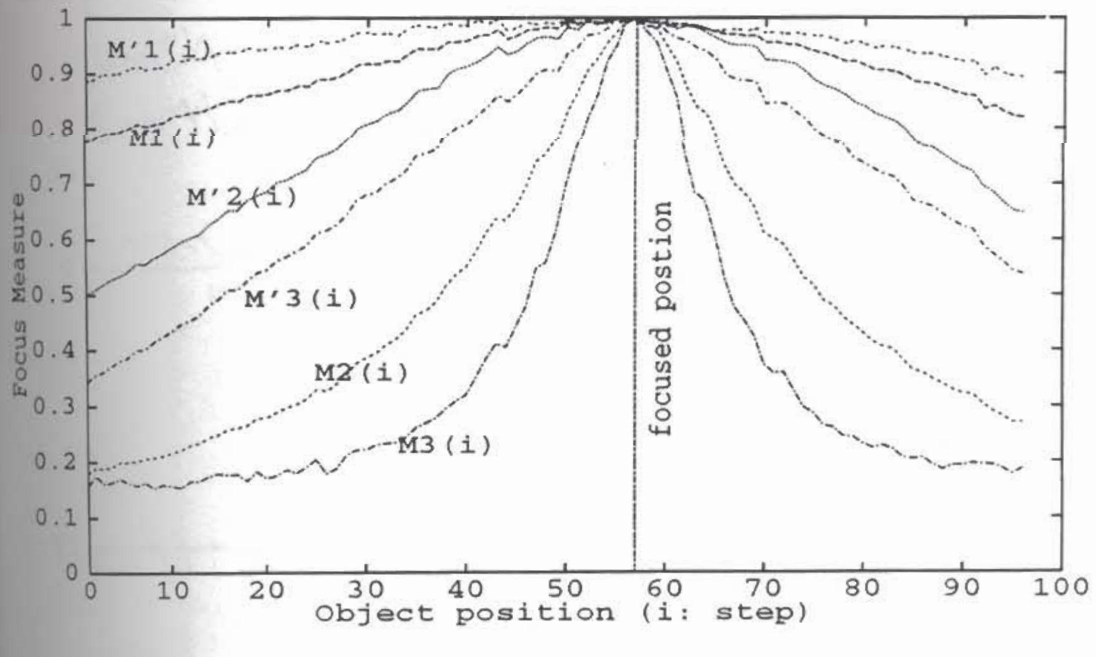


Figure 4.7: Focus measures for Object 2.

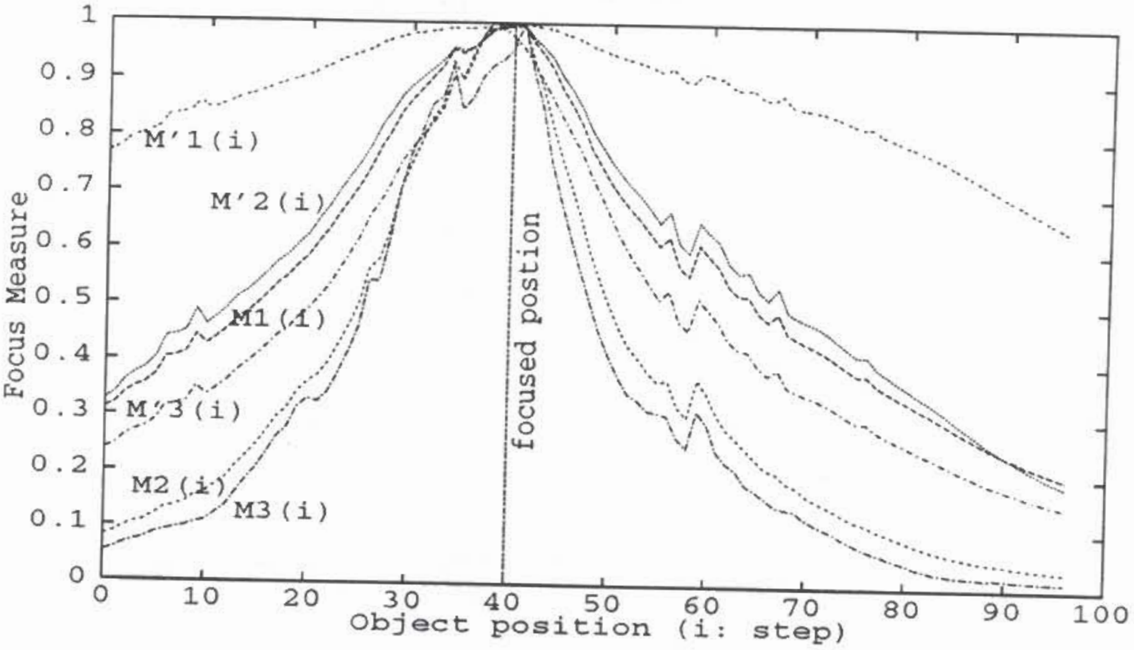


Figure 4.8: Focus measures for Object 3.

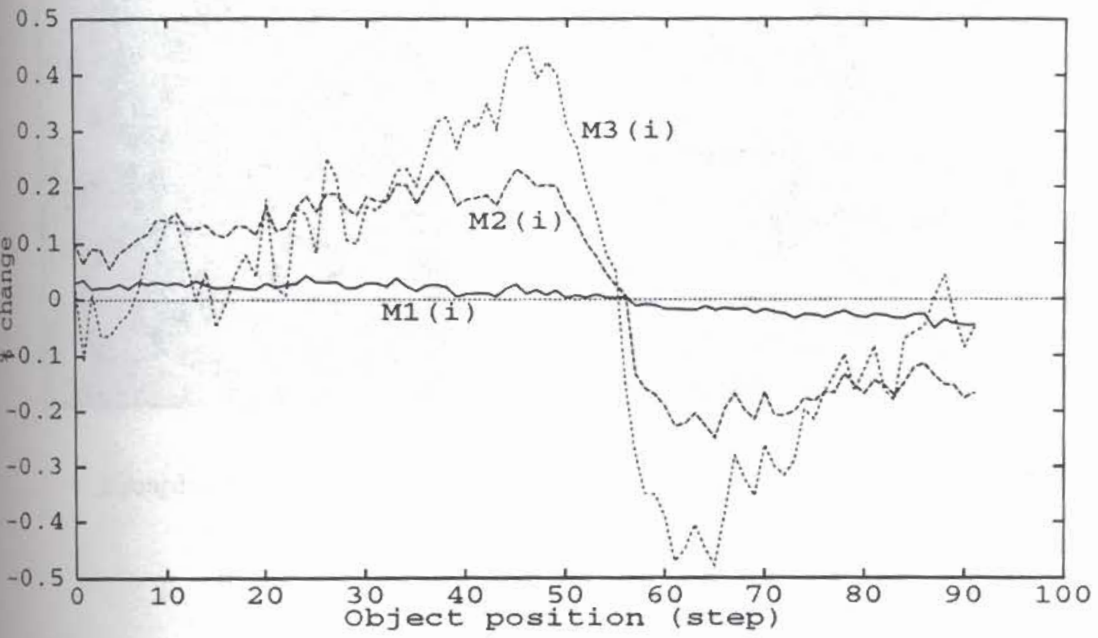


Figure 4.9: % change (M_1, M_2, M_3) for Object 2.

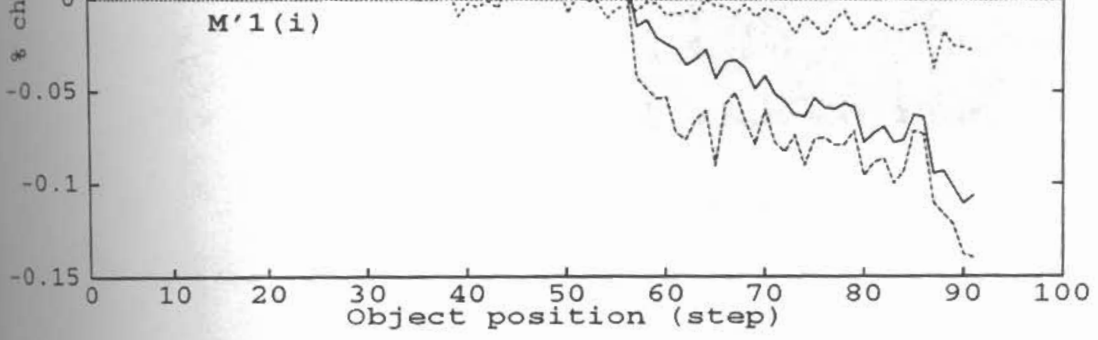


Figure 4.10: % change (M_1', M_2', M_3') for Object 2.

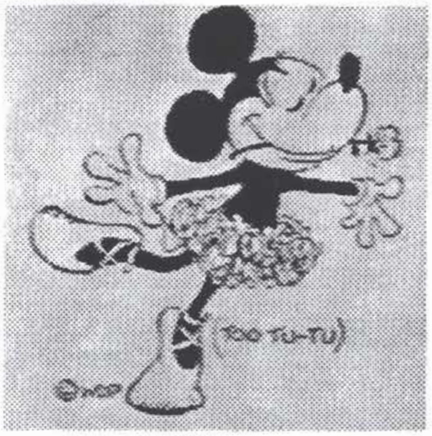


Figure 4.11: Object 1.

Figure 4.12: Object 2.

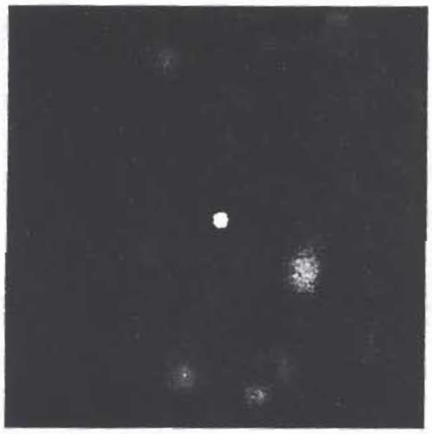


Figure 4.13: Object 3.

Chapter 5

Shape Recovery from Image Focus

5.1 Introduction

This chapter presents a solution to the shape-from-focus problem using a focus measure derived in chapter 4.

For an aberration-free convex lens, (i) the radiance at a point in the scene is proportional to the irradiance at its *focused image* [24], and (ii) the position of the point in the scene and the position of its focused image are related by the *lens formula* Eq. (3.1). Given the irradiance and the position of the focused image of a point, its radiance and position in the scene are uniquely determined. In fact the positions of a point-object and its image are *interchangeable,* i.e. the image of the image is the object itself. Now, if we think of an object surface in front of the lens to be comprised of a set of points, then the focused images of these points define another surface behind the lens (see Figure 3.1). We define this surface to be the *Focused Image Surface* (FIS) and the image irradiance on this surface to be the *focused image*. There is a *one to one correspondence* between FIS and the object surface. The geometry (i.e. the shape) and the radiance distribution of the object surface are uniquely determined by the FIS and the focused image.

In this chapter we are concerned with the principles and computational methods for recovering the geometry and the radiance of an object from its *sensed image*. Note that a sensed image is in general quite different from the focused image of an object. In computer vision, the sensors are usually planar image detectors such as CCD arrays. Therefore, for curved objects, only certain parts of the image will be focused whereas other parts will be blurred. A sensed image will be the focused image only when the shape of the sensor and the shape of FIS match.

In traditional SFF methods (e.g. [23, 27, 31, 36, 43, 56]) a sequence of images are obtained by continuously varying one or both of the following camera parameters: (i) distance between the lens and image detector, and (ii) the focal length. For each image in the sequence, a sharpness measure or focus measure is computed at each pixel using a small (about 15 × 15) image neighborhood around the pixel. At each pixel, that image frame among the image sequence which gives a maximum sharpness measure is determined. The grey level (which is proportional to image irradiance) of the pixel in the image frame thus determined is taken to be the grey level of th focused image for that pixel. The camera parameter values for this image frame ar used to compute the distance of the object point corresponding to the pixel. A simpl measure of sharpness of an image is its grey level variance. Measures based on the energy of derivatives of images are however better suited [56].

The traditional SFF methods do not yield accurate shapes or depth-maps of objects. The main reason for this is that a focus measure is defined and computed over image frames sensed by planar image detectors. The focus measure at each pixel in an image frame is computed using a small window around the pixel. This corresponds to a piecewise constant approximation of the object shape in the window (see Figure 5.2(a)). Because of this approximation, the focused image reconstructed from the image sequence will be an approximation to the actual focused image.

The fundamental contribution of our method is the idea that focus measures should be computed over the FIS using pixels lying on the FIS in the image sequence rather than over image frames where the pixels lie on a plane. Maximization of focus measures computed over FIS avoids the piecewise constant approximation of object shape found in the traditional SFF methods. The computational implementation of this idea involves two steps. The first step is essentially to estimate an approximate FIS using one of the traditional SFF methods. The second step is to refine this approximate estimate by searching for an FIS shape which maximizes a focus measure computed over pixels lying on the FIS. The search is local and therefore computationally efficient. At present, our implementation corresponds to a piecewise planar (or linear) approximation of object shape as opposed to piecewise constant approximation. However, our implementation algorithm can be easily extended to higher order approximation at the cost of additional computation.

Our SFF algorithm named SFF.FIS has been implemented on the SPARCS camera system . A number of experiments were carried out using SPARCS to evaluate our SFF algorithm. The experiments and their results are described. The experimental results show that our algorithm performs well.

In this chapter we are mainly concerned with SFF methods which give dense and accurate depth-maps, and which do not require a detailed knowledge of the camera characteristics. These methods require a sequence of image frames (about 10 to 30) recorded with different camera parameter settings. However, there are methods [15, 30, 33, 39, 46, 47, 48, 61, 65] which do not require a sequence of images, but only a few (about 2 or 3) acquired with different camera parameter values. These methods are very fast (about 10 times), but less accurate (their best performance gives a Root Mean Square (RMS) error which is twice that of the SFF methods). These fast methods are known as Depth-from-Defocus (DFD) methods whereas the SFF methods considered here are known as Depth-from-Focus (DFF) methods. Clearly, DFD methods can be used first to obtain a rough estimate of shape and then DFF or SFF methods can be used to refine the rough estimate to obtain a more accurate estimate of shape.

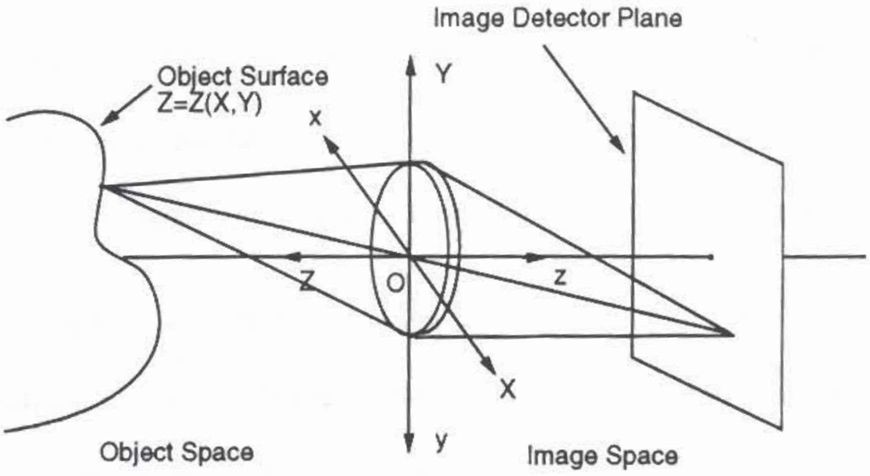


Figure 5.1: Object and Image Coordinate Systems.

We first consider the case of recording the image sequence by moving the image detector (or lens) along the optical axis of the lens. The results of this case can be easily extended to that of obtaining the image sequence by adjusting the focal length of the lens. When the image detector of a camera is moved from one end to the other, typically the focus measure in an image window gradually increases, reaches a maximum at the FIS, and then decreases gradually thereafter. The problem then is to find the image detector position at which the focus measure is a maximum. This is essentially a search of the image detector position space.

5.2 Relation between object surface and FIS

Figure 5.1 shows a right handed Object Space Coordinate System (OSCS) (X, Y, Z) with its origin O at the optical center of a convex lens. The visible surface of any object in the object space (scene) can be expressed as $Z = Z(X, Y)$. We assume that

$$Z(X, Y) \geq f \tag{5.1}$$

because any object point closer than focal length f will produce only a virtual image but no real image.

The image formed by a convex lens is inverted with respect to the object. Therefore, for convenience, we define a left handed Image Space Coordinate System (ISCS) (x, y, z) (see Figure 5.1) with its origin at O. The axes of the ISCS point in the direction exactly opposite to that of the corresponding axes of the OSCS. For any point (X, Y, Z) in OSCS, let its focused image in the ISCS be (x, y, z). The points (X, Y, Z) and (x, y, z) form a conjugate pair of points. Using the properties of similar triangles, it is easy to show that

$$\frac{x}{z} = \frac{X}{Z} \quad and \quad \frac{y}{z} = \frac{Y}{Z}. \tag{5.2}$$

For an object surface $Z(X, Y)$, the corresponding Focused Image Surface or FIS can be denoted by $z(x, y)$. It follows from the lens formula (3.1) that

$$\frac{1}{f} = \frac{1}{Z(X, Y)} + \frac{1}{z(x, y)}. \tag{5.3}$$

The above relation can be used to obtain the shape of an object from its FIS. It is helpful to recall here a few well-known results. A point at $Z = \infty$ comes to focus at $z = f$, a point at $Z = 2f$ comes to focus at $z = 2f$, and a point at $Z = f$ comes to focus at $z = \infty$.

Although the above relation between the shape of an object $Z(X,Y)$ and its FI $z(x,y)$ is quite simple, the following theorem is particularly interesting.

Theorem 1 *The Focused Image Surface of a planar object is also planar.*

Proof Let

$$Z = PX + QY + R \tag{5.4}$$

be a planar object where P and Q are respectively the slopes of the surface along X and Y axes respectively and R is the intercept along the Z axis. We have

$$Z - PX - QY = R \tag{5.5}$$

$$\Longrightarrow 1 - P\frac{X}{Z} - Q\frac{Y}{Z} = \frac{R}{Z}. \tag{5.6}$$

(5.7)

Using relations (5.2),(3.8), the above equation can be written as

$$1 - P\frac{x}{z} - Q\frac{y}{z} = R\left(\frac{1}{f} - \frac{1}{z}\right). \tag{5.8}$$

Rearranging terms in the above relation, we obtain

$$z = px + qy + r \tag{5.9}$$

where

$$c = \frac{f}{R-f}, \quad p = -cP \quad q = -cQ \quad \textit{and} \quad r = cR. \tag{5.10}$$

Therefore we have shown that the FIS of a planar object $Z(X,Y) = PX + QY + R$ is also planar and is given by $z(x,y) = px + qy + r$. The relation between the object surface parameters P, Q, R and FIS parameters is given by Eq. (5.10).

The fact that the FIS of a planar object is also planar can be used to conclude the following: (i) the FIS of a polyhedral object is also polyhedral (with the same number of planar faces), (ii) the FIS of a straight line is also a straight line, and (iii) the FIS of a surface which can be generated by sweeping a straight line in 3-D object space is also a surface which can be generated by sweeping a straight line in 3-D image space. The first result follows directly from the theorem. The second result can be proved by noting that the intersection of two planes in object space is a straight line and the FIS of the straight line is the intersection of the FISs of the two intersecting planes which are themselves planar. The third result follows from the second result. A consequence of the third result is that the FIS of a cone is a "distorted cone", and the FIS of a cylinder is a "distorted cylinder".

An important significance of the above theorem in SFF methods is that if the shape of an object can be approximated well by a piecewise planar surface then the shape of the corresponding FIS can also be approximated well by a piecewise planar surface.

5.3 FM operator

There are many focus measures which perform well when used in SFF methods [27, 36, 56]. Any one of these could be used in our algorithm described next. Two simple examples of reasonably good focus measures are energy of low-pass filtered image gradient and energy of Laplacian of the image. In our implementation we have chosen the energy of Laplacian of the image. To reduce the computational time, it is computed by first applying the Laplacian operator, squaring the resulting values at every pixel, and then summing the values in the window of interest.

Before the focus measure of an image is calculated, the image is first normalized with respect to brightness. This is done by dividing the grey level of each pixel by the mean grey level of the whole image.

5.4 SFF Algorithm

Conceptually, our Shape-from-Focus (SFF) algorithm can be described as follows. The image detector is first moved to $z = z_0$. A sequence of images $g(i, j, k)$ are recorded by moving the image detector to positions $z_i = z_0 + i * \delta$ where δ is a small displacement, for $i = 0, 1, 2, \cdots, I-1$, $j = 0, 1, 2, \cdots, J-1$, and $k = 0, 1, 2, \cdots, K-1$. Usually, $z_0 = f$. J and K are the number of rows and columns respectively in each image frame and I is the number of image frames (see Figure 5.2). We can think of this image sequence as an image volume. In this image volume, our problem is to find the set of pixels which lie on the focused image surface (FIS) of the object. For surfaces with a slope of up to about 1.0, for any given row j and column k, there is only one pixel which lies on the FIS. The image frame number i to which this pixel belongs depends on (j, k) and therefore it can be expressed as a function $i(j, k)$. The grey level of this pixel is $g(i(j, k), j, k)$. The relation between the row number j and the y coordinate is $y = (j - j_c)d$ where $j_c = J/2$ is the row index of the center row and

d is the distance between two rows of pixels on the image detector array. Similarly, the relation between the x coordinate and the column index k is $x = (k - k_c)d$ where $k_c = K/2$ is the column index of the center column and d is the distance between two columns of pixels on the image detector array.

The shape of the FIS can be determined from the function $i(j, k)$ which gives the frame number of the pixel lying on the FIS for any given (j, k). The focused image $F(j, k)$ of the object is obtained from the image sequence and the function $i(j, k)$ as

$$F(j, k) = g(i(j, k), j, k). \tag{5.11}$$

In order to find the function $i(j, k)$ which specifies the FIS, we use the fact that the focus measure of $F(j, k)$ (or $g(i(j, k), j, k)$) is a maximum over all possible functions. Since a search for a function is computationally expensive, a two phase procedure is used in our implementation.

In the first phase of the algorithm, a rough estimate of FIS is estimated using a traditional SFF method as follows. A small set (about 10) of N image frames g_n at regular intervals of I/N are selected from the original image sequence g_i. For each selected image frame, focus measures are computed in small field-of-views or image windows of size $M \times M$ (value of M varies from 3 to 15). For a given field of view, the corresponding image windows in different image frames shift and change magnification from one image frame to the other. Ideally, this should be taken into

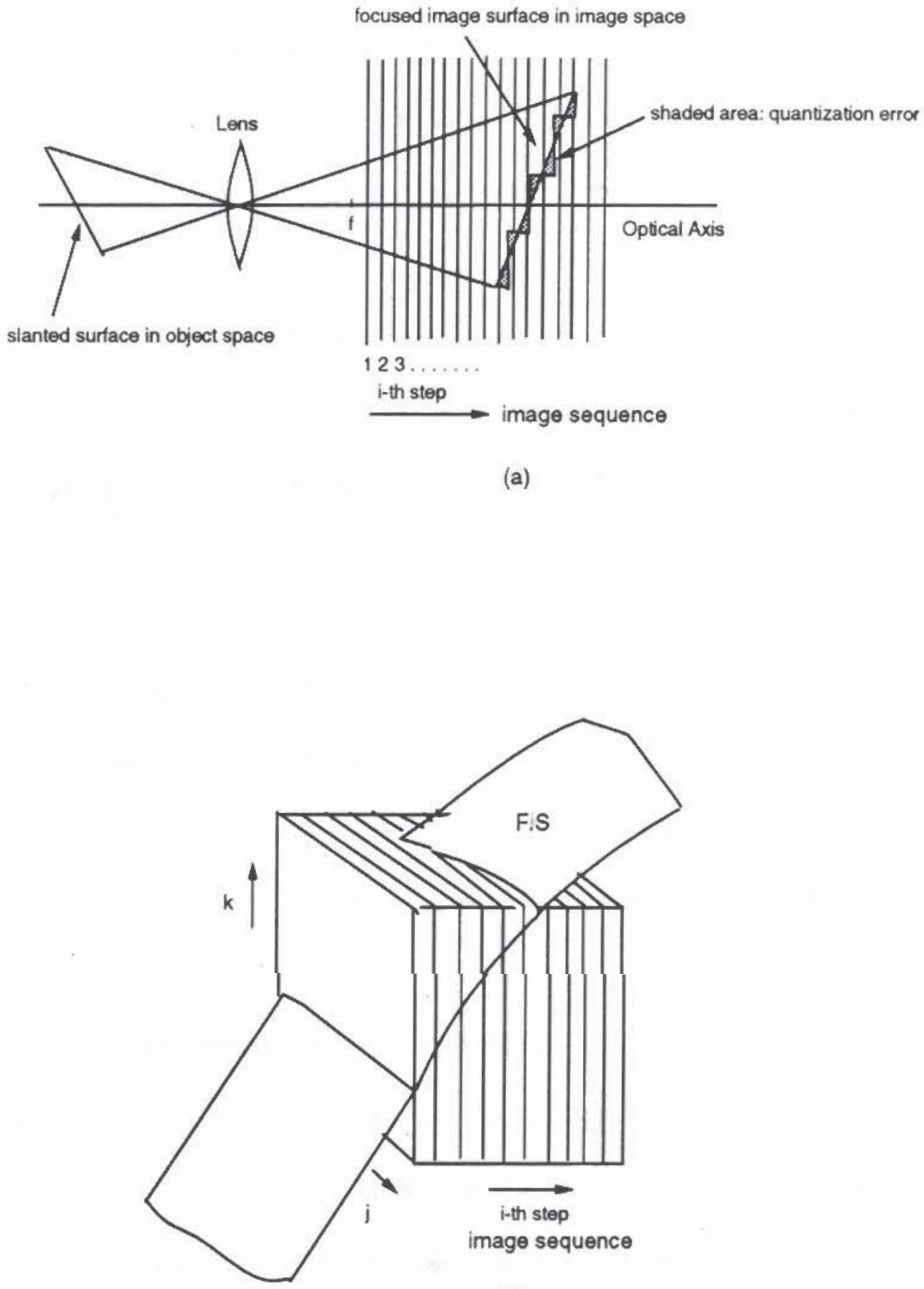


Figure 5.2: Direct Fitting on the FIS (Focused Image Surface).

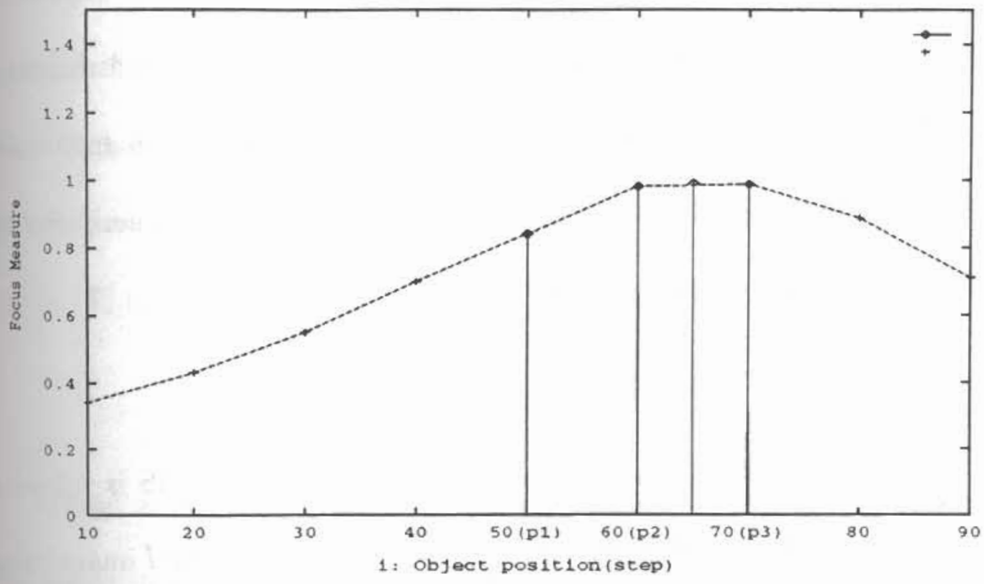


Figure 5.3: Focus Measure for a target pixel (j,k) and Polynomial Interpolation using a second order polynomial.

account, but in practice this effect may be neglected as the shift and magnification changes are usually small.

Usually overlapping windows are used at intervals of about $M/4$. Non-overlapping windows may also be used. For each window (or field-of-view), the image frame number for which the focus measure is a maximum over all the image frames is determined. The image frame numbers thus found gives a very rough estimate of the FIS. This estimate can be improved through a local interpolation scheme. For example, a low (quadratic or higher) order polynomial can be fitted to a few data points (focus measure as a function of image frame number) around the image frame with the maximum focus measure, and the location of the local maximum of the polynomial can be taken as the improved estimate of the focus location. Usually a quadratic or a cubic polynomial fit gives satisfactory results. A Gaussian function fit is used by some researchers [36]. Having obtained a rough estimate of the FIS, an approximate estimate of the slope of the FIS can be obtained by computing partial derivatives (or finite differences) along x and y directions.

In the second phase of our algorithm, the initial estimate of FIS is refined as follows. In this phase, the entire original image sequence g_i containing I image frames is used. For every window in which the FIS was estimated in the first step, a small cubic volume (about the size of $M \times M \times M$) image space is considered in the image sequence. The volume is centered at the initial estimate of FIS in that window. Now in this volume, a search is made for a planar surface which is closest to the actual FIS by maximizing the focus measure computed over the planar surface. The initial estimates of position and orientation of the FIS are used as starting values during the search. A brute-force or a simple gradient ascent search can be used. In our implementation, in order to increase robustness against noise, the focus measure was summed over the estimated FIS and two more surfaces parallel to it where one was about one image frame closer to camera and the other was about one image frame away from the camera.

An outline of the algorithm implemented by us can be summarized as follows.

- In every field of view, do the steps below.

- Read as input the initial estimates of the position and orientation parameters of the FIS. Let i_0 be the initial estimate of position, and p_0, q_0 be the initial estimates of the slopes of the FIS along the x and y axes respectively. Then the initial estimate of the FIS is described by

$$i(j,k) = i_0 + p_0 j + q_0 k \tag{5.12}$$

- Read as input the size of the search space for position and orientation. Let the search space for position be the range $i_{min} = i_0 - \delta_i$, to $i_{max} = i_0 + \delta_i$, for x-slope be the range $p_{min} = p_0 - \delta p$, to $p_{max} = p_0 + \delta p$, and for y-slope be the range $q_{min} = q_0 - \delta q$, to $q_{max} = q_0 + \delta q$. If $i_{min} \leq 1$ set $i_{min} = 1$ and if $i_{max} \geq I$ set $i_{max} = I$. Take the value of maximum allowable slope MAXSLOPE for the surface along each axis to be about 1.0. If $p_{min} \leq -$MAXSLOPE set $p_{min} = -$MAXSLOPE and if $p_{max} \geq$ MAXSLOPE set $p_{max} =$ MAXSLOPE. Similarly, if $q_{min} \leq -$MAXSLOPE set $q_{min} = -$MAXSLOPE and if $q_{max} \geq$ MAXSLOPE set $q_{max} =$ MAXSLOPE. Read as input the searching step sizes $\delta_i, \delta_p, \delta_q$ for i, p, q, respectively.

- Read image window size $2S+1$.

```
$maxsum = 0;$

$i_p = i_0;\ p_p = p_0;\ q_p = q_0;$

for $i_m = i_{min}$ to $i_{max}$ in steps of $\delta_i$ do /* search position space
```

```
        for $p_m = p_{min}$ to $p_{max}$ in steps of $\delta_p$ do /* search x-slope
            for $q_m = q_{min}$ to $q_{max}$ in steps of $\delta_q$ do /*search y-slope

    /* compute focus measure on a candidate FIS */

                $sum$=0;

                for $j = -S$ to $S$ do
                    for $k = -S$ to $S$ do
                        $l = \lfloor i_m + j * p_m + k * q_m \rfloor$;
                        $sum = sum + \sum_{m=l-1}^{m=l+1} (Laplacian[m][j][k])^2$;
                    end /* $k$ loop
                end /* $j$ loop
                if($sum > maxsum$)
                then $i_p = i_m$; $p_p = p_m$; $q_p = q_m$; $maxsum = sum$;
            end /* $q_m$ loop
        end /* $p_m$ loop
    end /* $l_m$ loop
```

i_p, p_p, and q_p are the position and slopes of the estimated FIS. In the above algorithm, $Laplacian[m][j][k]$ is the image Laplacian computed at [m,j,k].

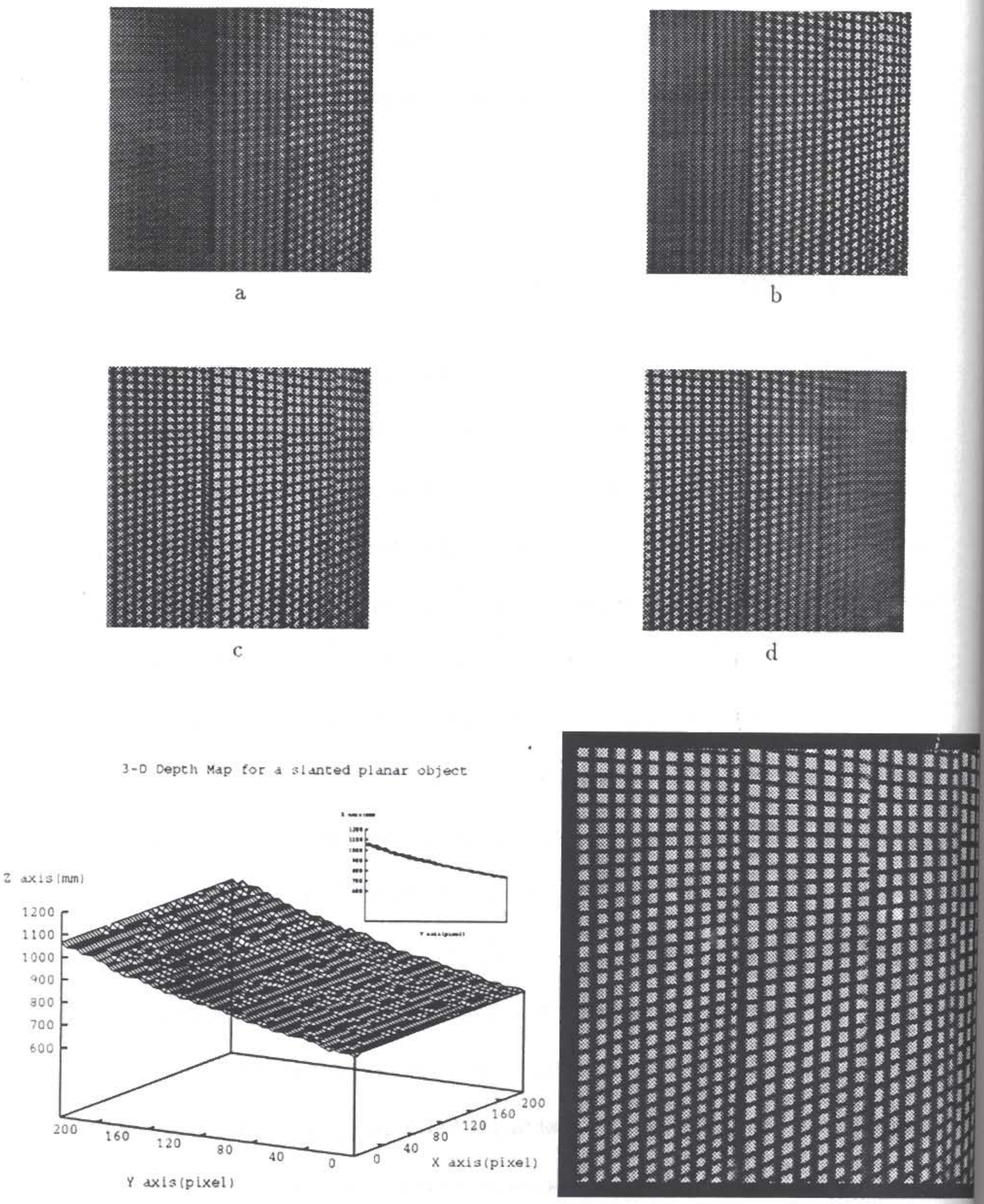
The estimated FIS is described by $i(j,k) = i_p + p_p * j + q_p * k$ in the image window being processed. At the cost of additional computation, it is also possible to search and obtain piece-wise quadratic and higher order approximations to the FIS.

5.5 Experimental Results

The SFF algorithm described above and the corresponding traditional method were implemented on the SPARCS camera system. The methods were similar in all respects except that the focus measure was computed over estimated FIS in the case of our SFF method whereas in the traditional method the focus measure was computed over image frames in image windows. These implementations were used to compare the improvement obtained by our SFF algorithm with the traditional method.

Here we present the results for two objects: (i) a slanted planar object (Figure 5.4), and (ii) a cone object of length about 79 inches and base diameter of about 15 inches (Figure 5.5). It is found that the the improvement in accuracy for the slanted planar object is marginal whereas it is significant for the cone object. This indicates that FIS improves the accuracy substantially in the case of curved objects.

The illumination for the two objects was about 600 lux. Image size was 256×256. In order to reduce electronic noise, for a fixed lens position five image frames were time averaged. The image sequence contained 97 image frames, one for each lens step position of the stepper motor. The absolute displacement between two consecutive image frames was about 0.03 mm and the distance between pixels was about 0.013



a

b

c

d

e

f

Figure 5.4: A slanted planar object.

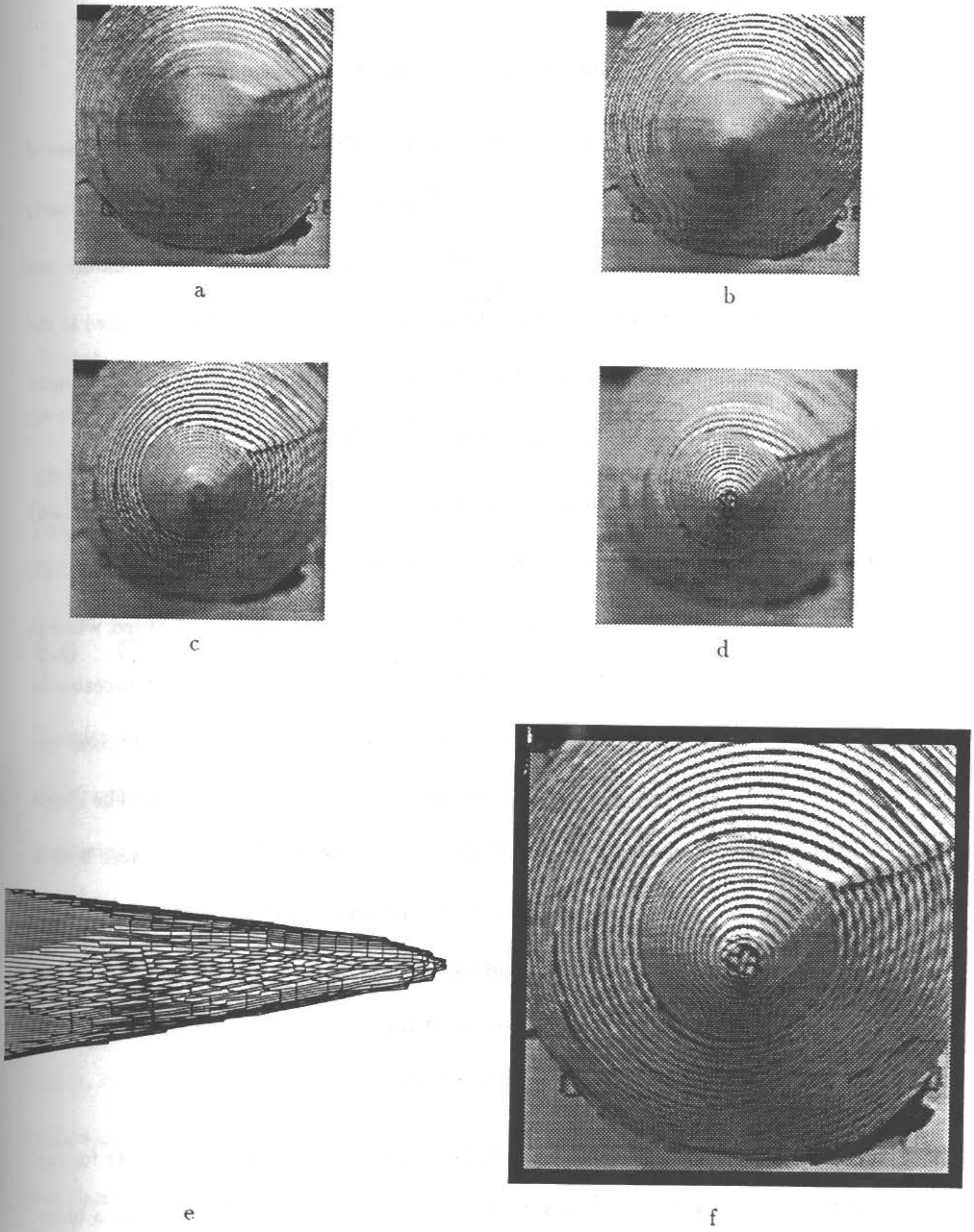


Figure 5.5: A cone object.

mm. The window size for computing focus measures was 15 × 15.

An initial estimate of FIS was obtained by computing the sum of the square of image Laplacian for 9 image frames equally spaced apart (about 10 frames apart) in the original image sequence. The position of the maximum focus measure was first improved by a quadratic interpolation scheme using three points centered at the maximum point. A typical plot of the 9 values along with the position of maximum position obtained using interpolation is shown in Figure 5.3.

Figure 5.4 shows the results for the slanted planar object. Figs. 5.4(a) to 5.4(d) show the image frames recorded when the lens position was at motor steps 20, 40, 60, and 80. In each of these frames, only one part of the image is focused whereas the other parts are blurred by varying degrees. This is particularly noticeable in Fig. 5.4(d) where the closer part of the object on the left is focused whereas the blur increases gradually towards right as the object distance increases. The shape or depth-map recovered by our SFF algorithm is shown in Fig. 5.4(e). Here both a 3-D surface plot and a smaller side view of the plot are shown for clarity. The results in this case are close to the actual shape except in regions where there is insufficient contrast. The reconstructed focused image of the object is shown in Fig. 5.4(f). We see that all parts of the image are in sharp focus.

Figure 5.5 is similar to Figure 5.4 except that the results in this case are for the cone object. In Fig. 5.5(e) we see that the recovered shape of the cone has a blunt tip rather than a sharp tip. This is due to the piece-wise planar approximation.

	Traditional Method	SFF.FIS
RMS Error(Slanted Planar Object)	0.40 step	0.37 step
RMS Error(Simulated Cone Object)	2.22 step	1.41 step
Number of test points	3136	3136

Table 1. RMS Error.

Except in areas where there is insufficient grey level variance, the shape recovered is good. Figure 5.5(f) shows the reconstructed focused image of the cone object. In comparison with the image frames shown in Fig. 5.5(a) to 5.5(d), the reconstructed image appears focused everywhere.

5.6 Comparison with Traditional Method

In order tc compare our SFF method with the traditional method we carried out experiments on both real data and simulated data. Simulated data was necessary in comparing the results of the two methods with accurate ground truth. For real data, we did not have adequate facilities to accurately measure the ground truth.

Figures 5.8 and 5.9 show the results on real data for the slanted planar object obtained using the traditional method and our SFF method respectively. In these figures, it is difficult to visually compare the two results. Since the accurate ground truth was not available, we fitted planar surfaces to the data (we knew that the object was planar but did not accurately know its position and orientation) using a least-square error minimization approach. Then the root-mean-square (RMS) error was computed between the fitted planes and the data. The results are shown in Table 1. We see that the RMS error for the traditional method is 0.40 lens steps (out of 97 steps) whereas it is 0.37 lens steps for the traditional method. Therefore the improvement in accuracy in this case is marginal.

Figures 5.10 and 5.11 show the results on real data for the cone object. As expected, the tip of the cone is sharper in the case of our SFF method whereas it is blunt in the case of the traditional method. The surface of the cone in Figure 5.11 is a little smoother than in Figure 5.10. The actual cone object was made of cardboard and it was distorted. Therefore we could not fit a cone to the data in order to compare the accuracy of the two methods as we did above in the case of the planar object.

In order to do a rigorous quantitative comparison of our SFF method with the traditional method we did the following experiment on simulation data. A camera simulation software called Active Vision Simulator (AVS) [34, 63] was used to generate 97 images of a cone object corresponding to 97 lens positions. A paraxial geometric optics model was used for image formation in the computation of the blurred images. For the simulation software, the input camera parameters (focal length, aperture, pixel size, etc.) were set to be the values of our actual camera. Figure 5.14 shows the input depth-map of the cone. Figures 5.12 and 5.13 show the results of the traditional method and our SFF method respectively. The RMS errors between these two results and the ground truth in Figure 5.14 were computed. For the traditional method the RMS error was 2.22 lens steps out of 97 steps whereas the RMS error was 1.41 lens

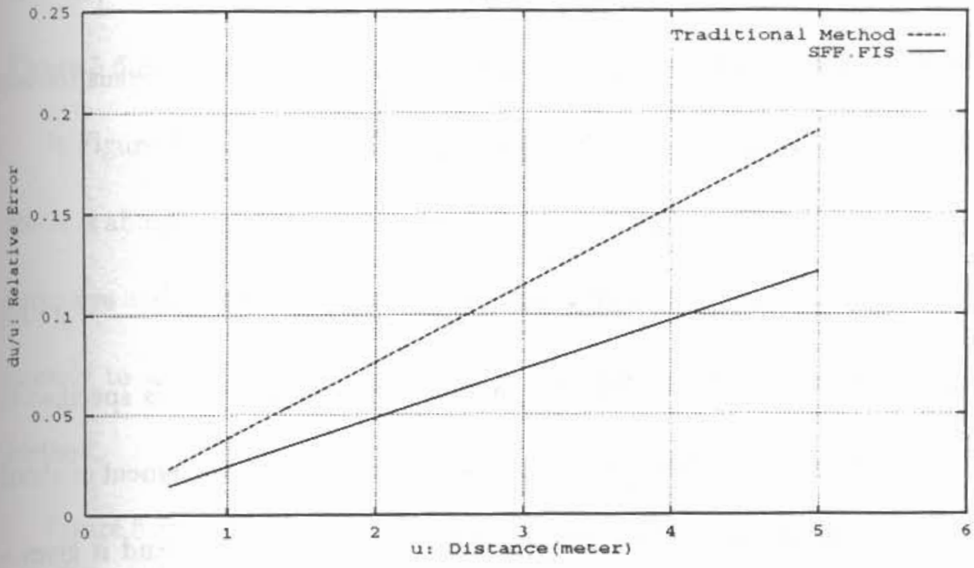


Figure 5.6: Relative Error.

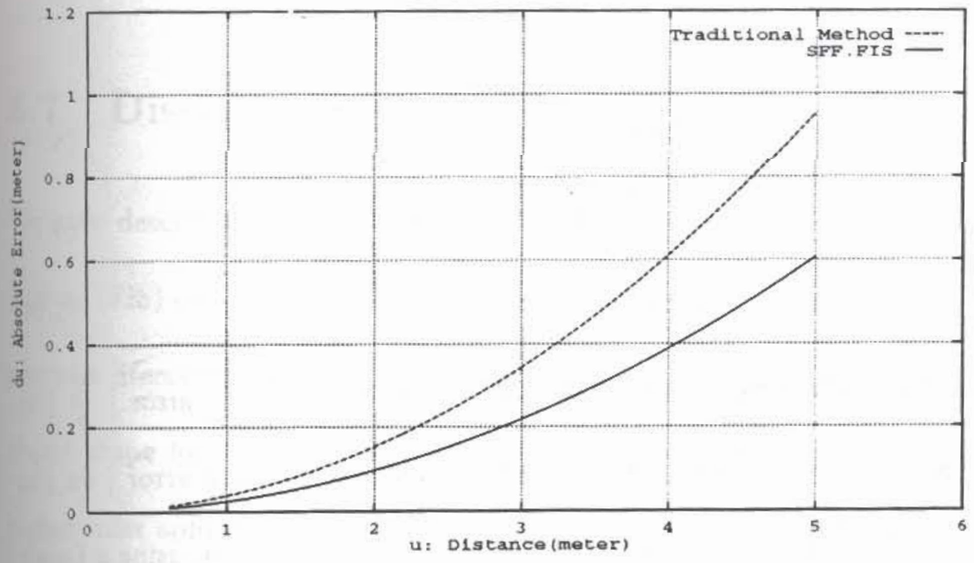


Figure 5.7: Absolute Error.

steps (out of 97 steps) for our SFF method. Therefore, in this case, the traditional method has about 1.5 times the error for our SFF method.

The relationship between the reciprocal of the object distance 1/u versus the lens step number is almost linear and can be expressed as

$$1/u = ax + b \tag{5.13}$$

where x specifies lens position. For our camera, the lens position is specified in terms of a motor step number where each step corresponds to a displacement of about 0.03mm. The RMS errors mentioned above are for the lens position and it gives a good indication of the performance of the method for application in camera systems. In order to compute the error in terms of object distance, we have to consider the error differentials in Eq. (5.13).

$$\begin{aligned} \mid \delta(1/u) \mid &= a \mid \delta x \mid && (5.14) \\ \Rightarrow \mid \frac{\delta u}{u} \mid &= a \mid \delta x \mid u && (5.15) \\ \Rightarrow \mid \delta u \mid &= a \mid \delta x \mid u^2. && (5.16) \end{aligned}$$

From the above relations we see that the relative (percentage) error $\mid \frac{\delta u}{u} \mid$ in actual distance u increases linearly with distance, and the absolute error $\mid \delta u \mid$ in actual distance increases quadratically with distance. For our camera, using a Depth from Focus method the constants were found to be $a = 0.0172$ and $b = -0.1143$ [57, 66].

Setting $| \delta x |$ to be the RMS error of 2.22 steps for the traditional method and 1.41 steps for the new method respectively, a plot of relative error $| \frac{\delta u}{u} |$ is shown in Figure 5.6 and a plot of the absolute error is shown in Figure 5.7.

In Figure 5.6 we see that for the new method the percentage error in distance at 0.6 meter is about 1.45% and increases linearly to about 12.13% at 5 meter distance. This compares well with the resulting error of about 2.29% at 0.6 meter and increasing linearly to about 19.1% at 5 meter distance that is obtained with the traditional method.

Figure 5.7 shows that for the new method, absolute error increases quadratically from 8.7 mm at 0.6 meter to about 606 mm at 5 meter distance. The corresponding numbers for the traditional method are 13.7 mm at 0.6 meter and about 955 mm at 5 meter distance.

5.7 Discussion

We have described a new Shape-from-Focus method based on the Focused Image Surface (FIS) of objects. Unlike piece-wise constant approximation described in the previous literature, our method corresponds to piece-wise planar approximation of object shape for SFF methods. It can also be used for piece-wise quadratic and higher order approximations to FIS, at the cost of additional computation.

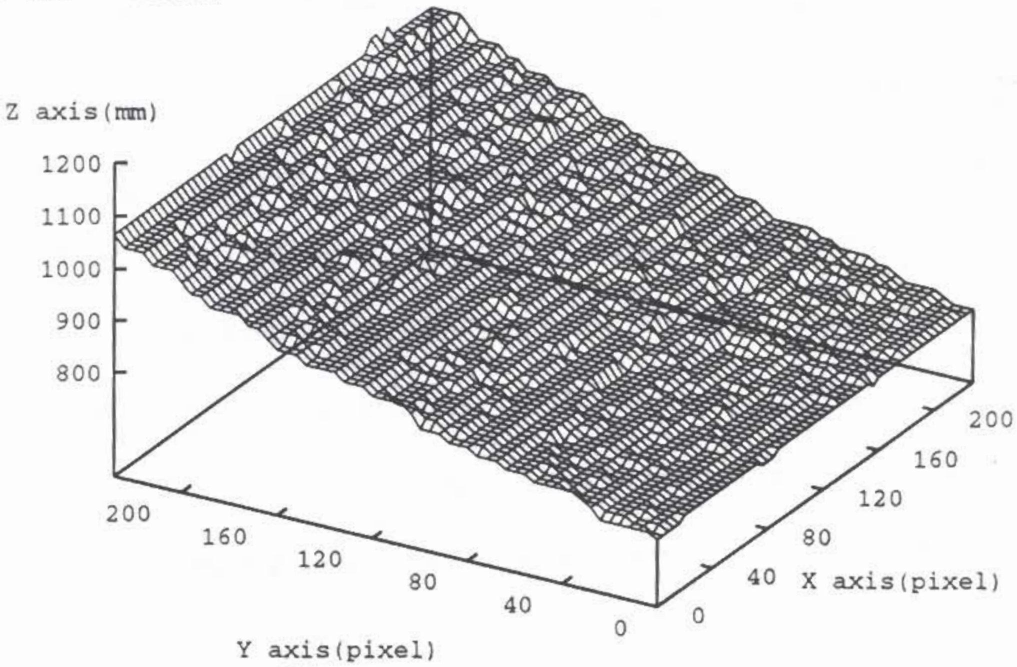


Figure 5.8: 3-D Depth Map for a slanted planar object by the traditional method.

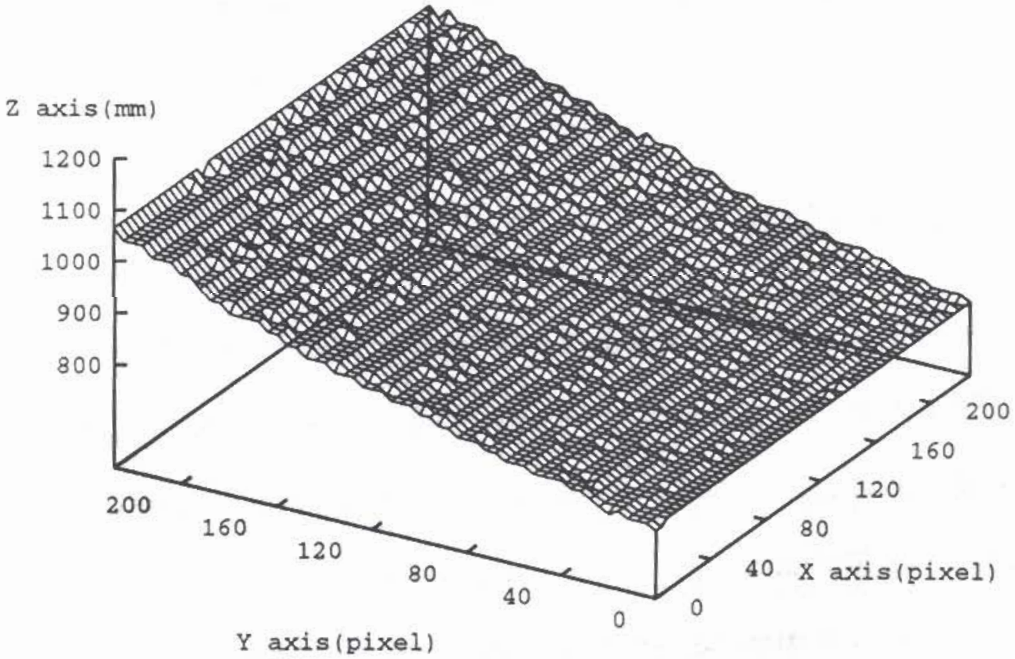


Figure 5.9: 3-D Depth Map for a slanted planar object by the new method (SFF.FIS).

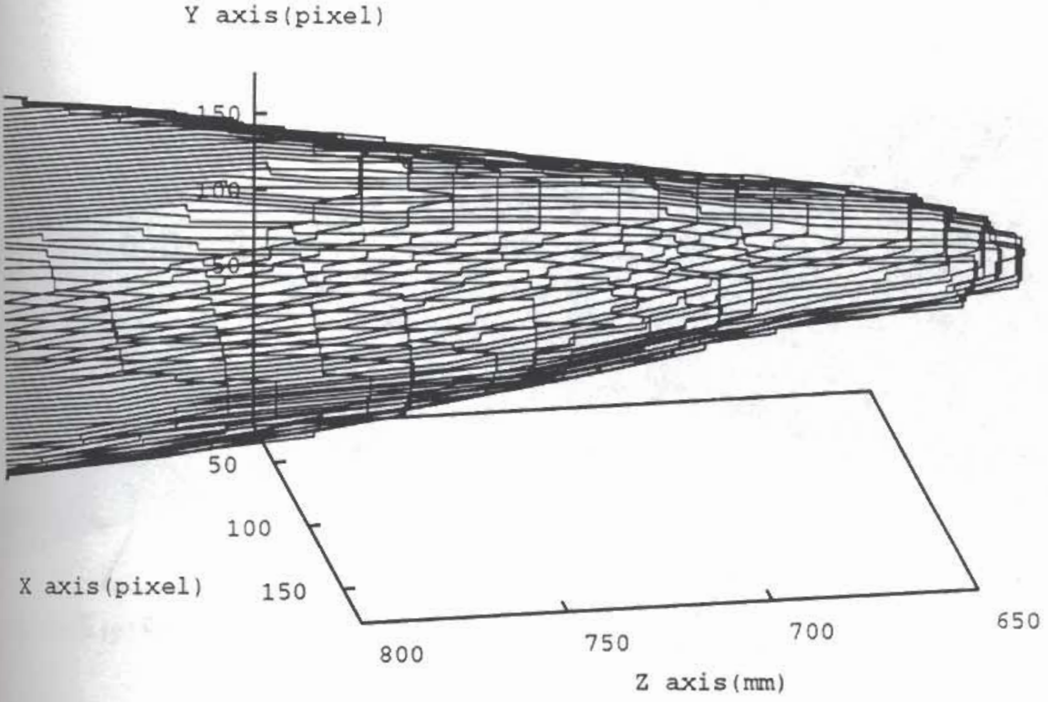


Figure 5.10: 3-D Depth Map for a cone object by the traditional method.

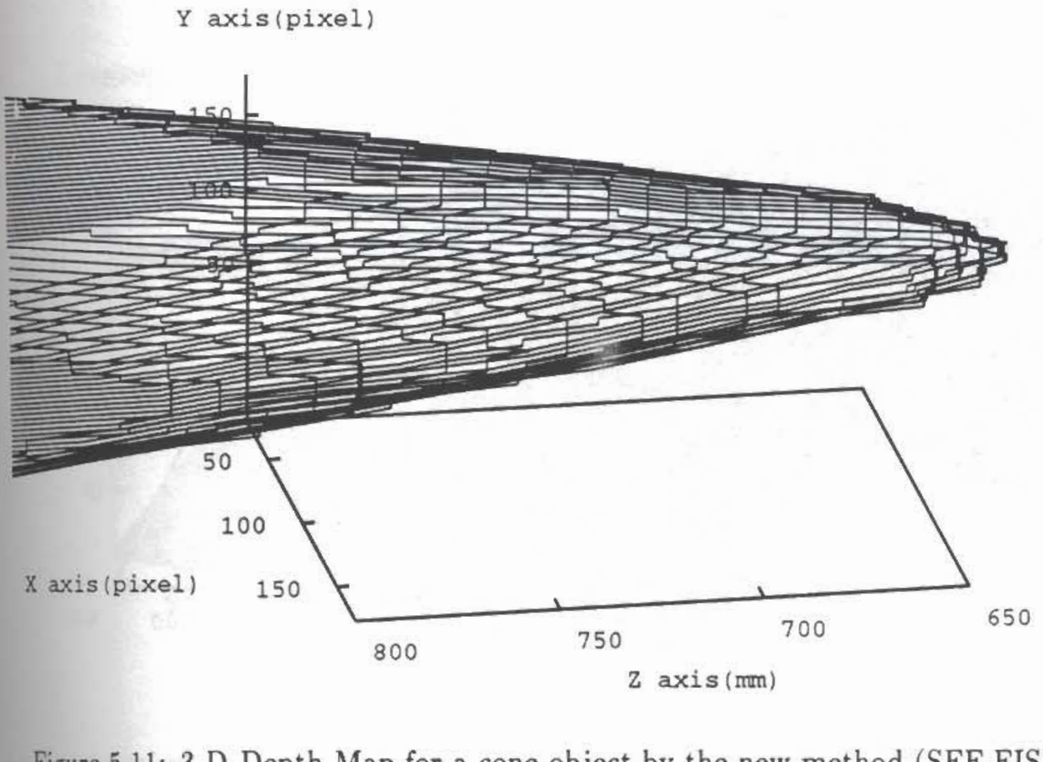


Figure 5.11: 3-D Depth Map for a cone object by the new method (SFF.FIS).

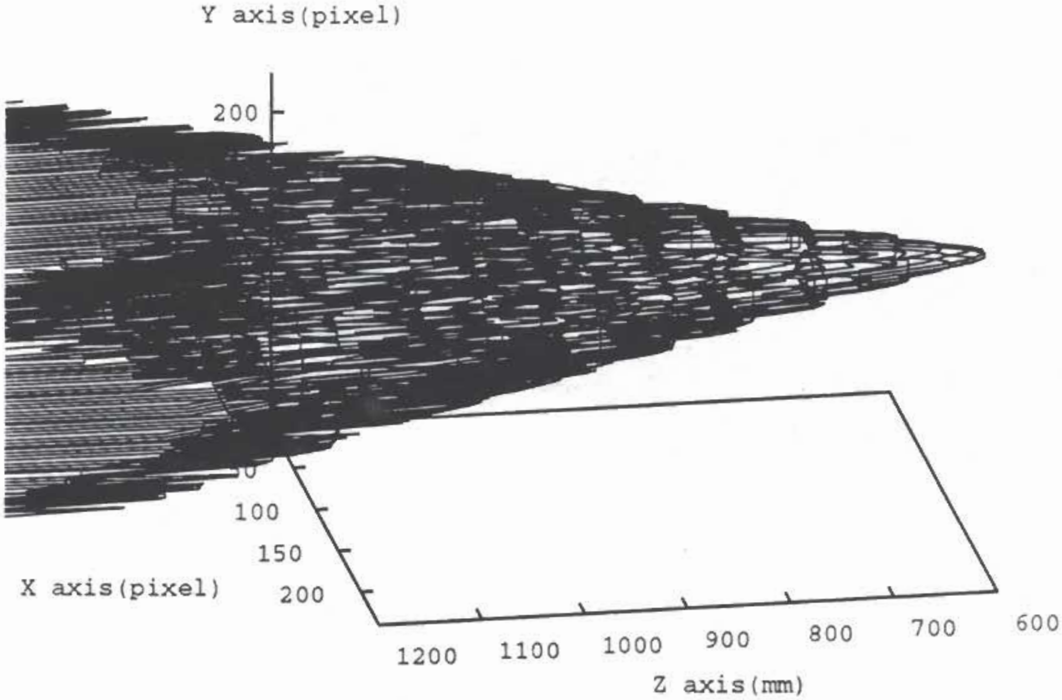


Figure 5.12: 3-D Depth Map for a simulated cone object by the traditional method.

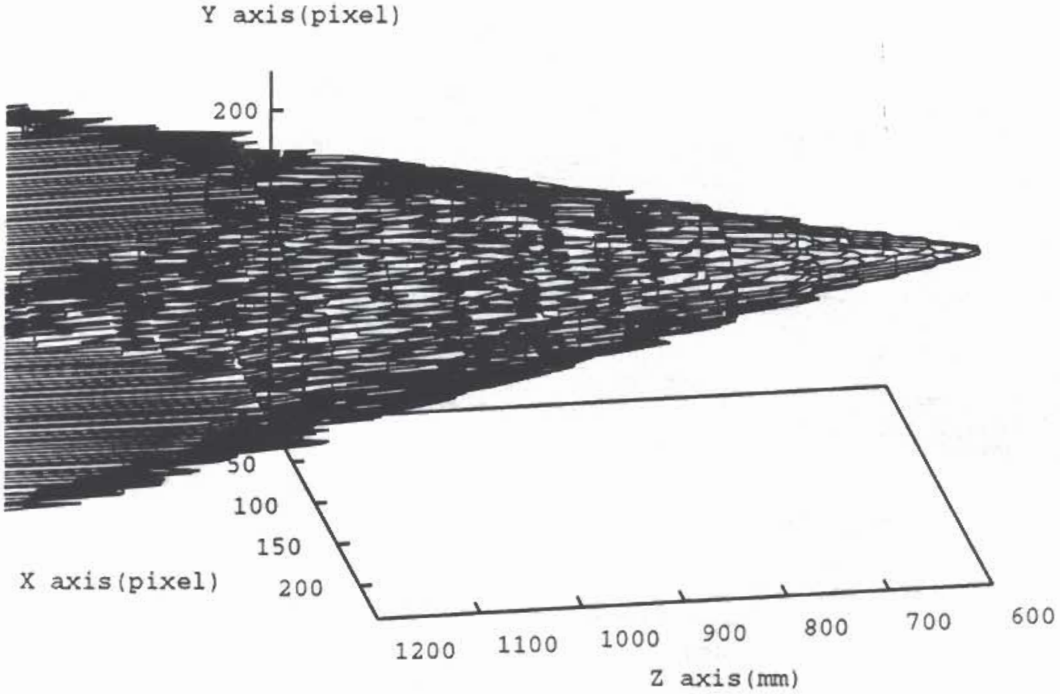


Figure 5.13: 3-D Depth Map for a simulated cone object by the new method (SFF.FIS).

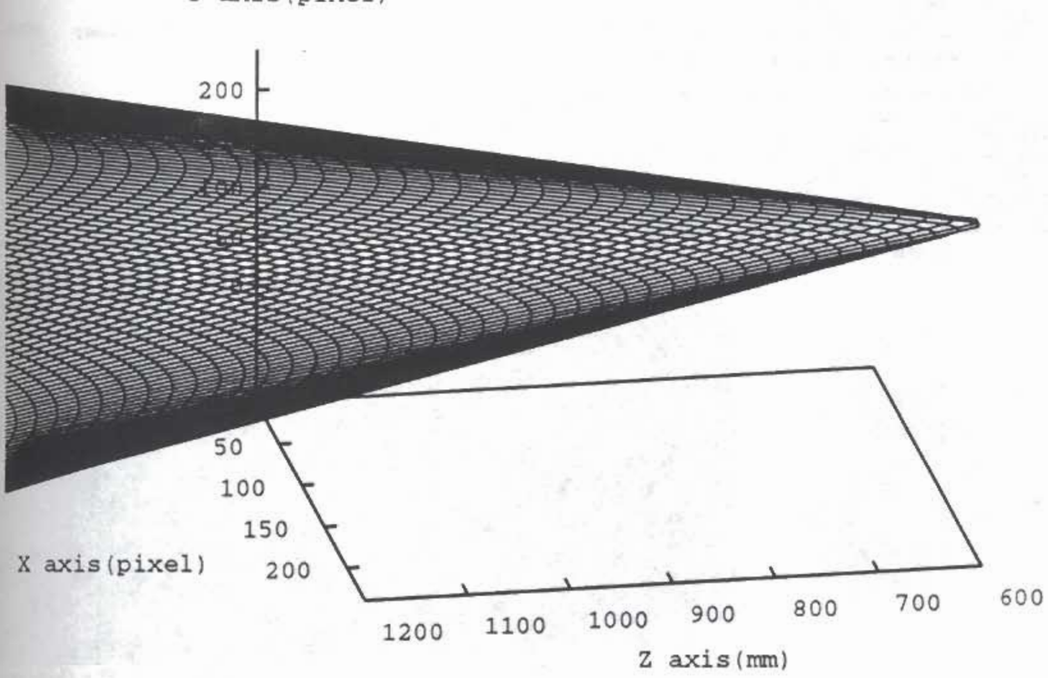


Figure 5.14: A simulated cone object.

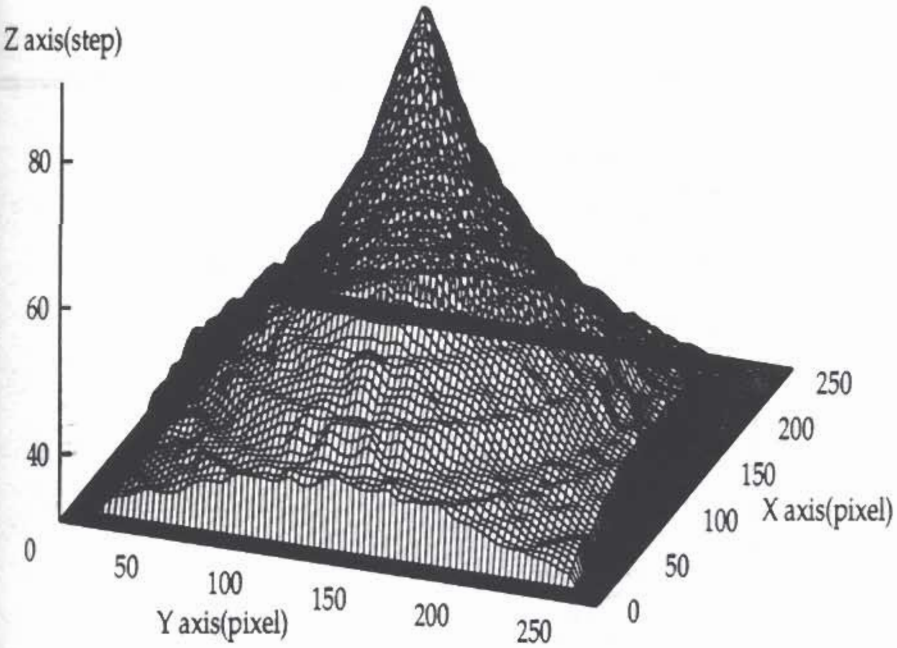
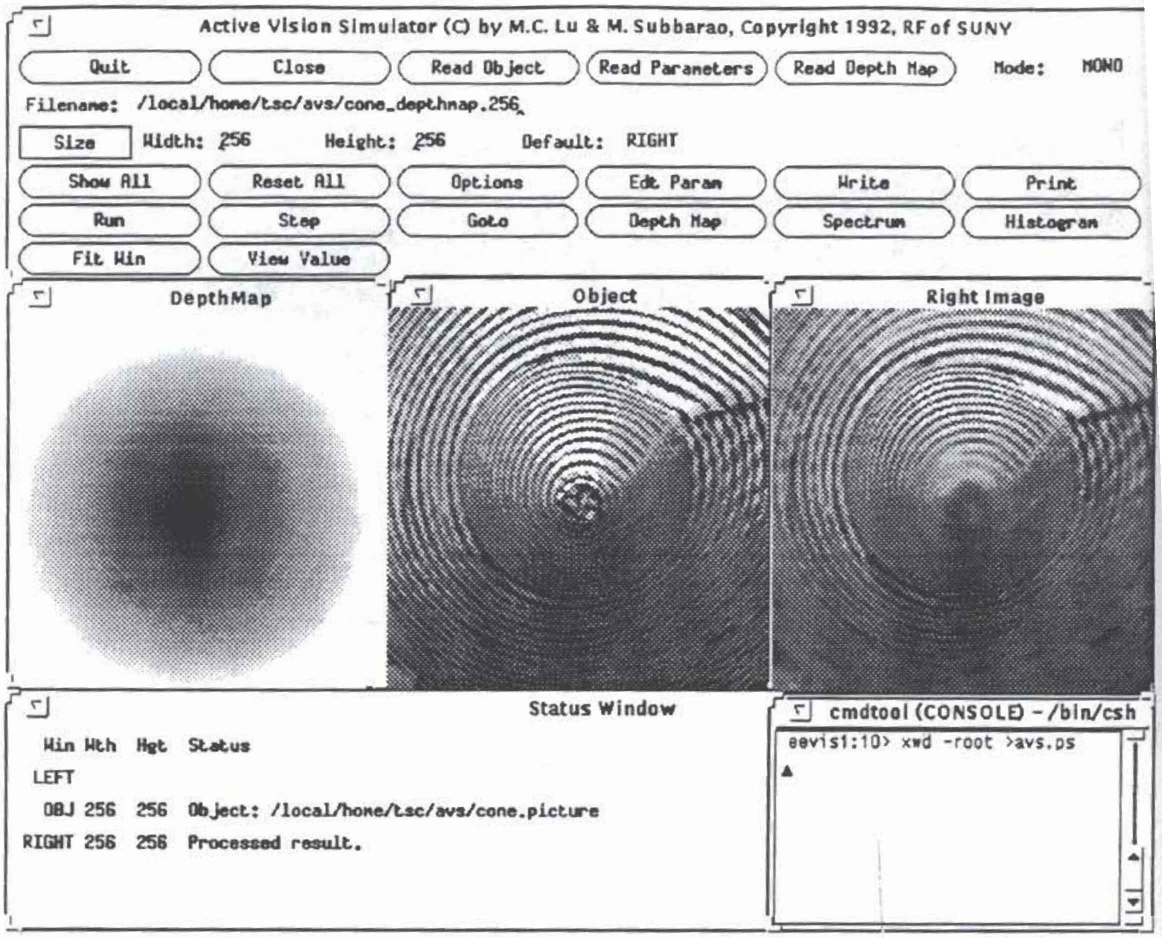


Figure 5.15: 3-D Depth Map in lens step for a simulated cone object by the new method (SFF.FIS).



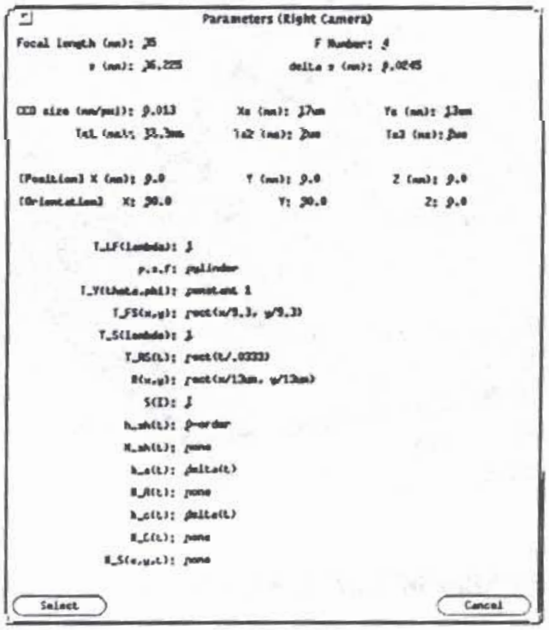
Mode: MONO

Filename: /local/home/tsc/avs/cone_depthmap.256

Width: 256 Height: 256 Default: RIGHT

```
Win Wth  Hgt  Status
LEFT
 OBJ 256  256  Object: /local/home/tsc/avs/cone.picture
RIGHT 256 256  Processed result.
```

```
eevis1:10> xwd -root >avs.ps
```



Focal length (mm): 35
F Number: 4
s (mm): 36.225
delta s (mm): 0.0245
CCD size (mm/pxl): 0.013
Xs (mm): 17mm
Ys (mm): 13mm
Ts1 (ms): 33.3ms
Ts2 (ms): 2us
Ts3 (ms): 2us
[Position] X (mm): 0.0
Y (mm): 0.0
Z (mm): 0.0
[Orientation] X: 90.0
Y: 90.0
Z: 0.0
T_LF(lambda): 1
p.s.f: cylinder
T_V(theta,phi): constant 1
T_FS(x,y): rect(x/9.3, y/9.3)
T_S(lambda): 1
T_RS(t): rect(t/.0333)
R(x,y): rect(x/13um, y/13um)
S(I): I
h_sh(t): 0-order
H_sh(t): none
h_a(t): delta(t)
H_A(t): none
h_c(t): delta(t)
H_C(t): none
N_S(x,y,t): none

Figure 5.16: Active Vision Simulator (AVS).

Chapter 6

Combining Shape from Focus and Stereo

6.1 Introduction

In chapter 5, we discussed a new shape-from-focus method based on the FIS. This chapter will show how the shape-from-focus method is integrated with stereo.

Many integration methods [2, 3, 11, 12, 28] for focus, stereo, shading, and motion have been proposed to solve the problem of shape recovery.

Krotkov [28] developed a cooperative sensing process, in which focus and stereo ranging verify the results of each other. His method identifies false matches with focusing. He used a maximum likelihood estimation (MLE) method to integrate focus and stereo ranges. Das and Ahuja [2, 11, 21] proposed a coarse-to-fine surface reconstruction method from stereo and focus for large scenes having large depth ranges. Their method is designed to achieve integration through iteration of the following three steps: visual target selection, fixation, and stereo reconstruction.

In this chapter, the objective of the research is to work on the integration of two individual modules for shape recovery. In particular, we are interested in combining the depth information from focus and stereo. Shape-from-focus methods give 3-D depth information from a sequence of blurred images. Stereo methods give 3-D depth information from two images taken by cameras which are slightly shifted along the horizontal axis, in the case of parallel optical axes of two cameras. The image of an object in the left image is shifted with respect to the image in the right image. This disparity is inversely proportional to distance. The problems with stereo are correspondence and occlusion. In contrast, the SFF methods don't have these problems. Therefore a combined method is proposed. The basic idea of this method is that a depth-map and two focused images obtained by a SFF method are used to solve the correspondence problem in stereo-vision.

The SFF method used here is SFF.FIS presented in chapter 5. In this method, a focus measure was computed over estimated FIS searched by a piecewise planar surface. Once one pair of ranges (z_f, z_s) are obtained from focus and stereo, two ranges are integrated using maximum likelihood estimation method.

Our combined algorithm called SFFS has been implemented on the SPARCS camera system. The experiments and their results are described. The experimental results show that SFFS reduces the number of false matches.

The organization of the rest of this chapter is as follows. Section 6.2 describes the theoretical background and accuracy of range estimation for stereo. Section 6.3 discusses the SFFS algorithm combined with focus and stereo. In section 6.4, the experimental results obtained by our camera system will be given. Finally, we discuss the running speed of our algorithm in section 6.5.

6.2 Theoretical Background of Stereo

In this section we are concerned with the principles and accuracy of range estimation for shape-from-stereo.

6.2.1 Shape from Stereo

Stereo method is an important solution for 3-D shape recovery using binocular cue. The distance from the lens to a point on the object surface is determined by finding the disparity of a conjugate pair in the left and right images.

Given an object point $p = (x, y, z(x, y))$, let (x_l', y_l') and (x_r', y_r') be the image coordinates in the left and right images in Figure 6.1 respectively. In the case where the optical axes of two cameras are parallel and perpendicular to the baseline, the object distance $z(x, y)$ is described as follows.

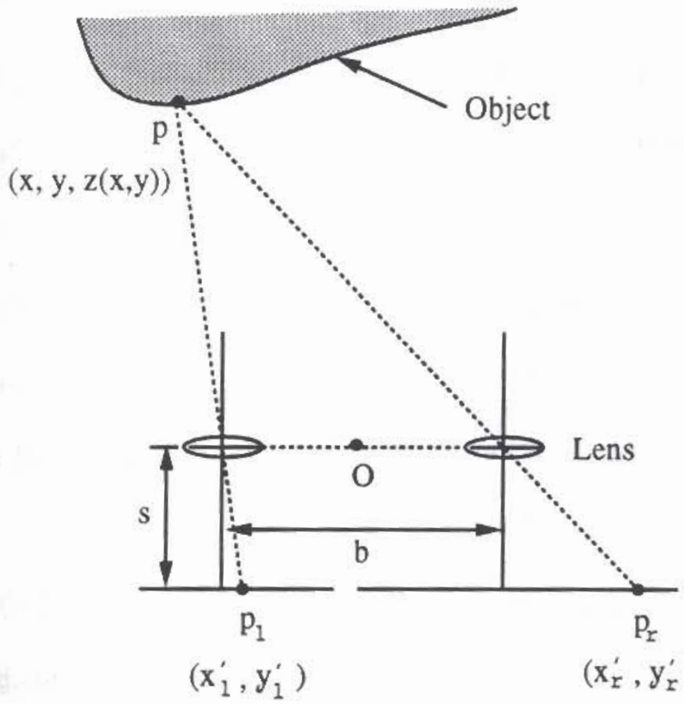


Figure 6.1: Simple lens geometry for stereo vision.

$$
\begin{aligned}
z(x,y) &= \frac{bs}{x'_l - x'_r} \\
&= \frac{bs}{D_s} \\
&\equiv z_s
\end{aligned}
\tag{6.1}
$$

where b is the baseline distance between two cameras, s is the distance from the lens to the image detector, and D_s is the disparity of two conjugate points.

The disparity between two conjugate points p_l and p_r is given by $x'_l - x'_r$. Since the disparity is inversely proportional to distance, the absolute error in measuring the disparity increases quadratically with respect to distance [24].

The conjugate point p_r of the right image corresponding to a point p_l of the left image is found on the corresponding epipolar line by using a matching method. The epipolar lines can be found by the lens geometry or by a set of known markers on the object surface through camera calibration [68]. Wang and Pavlidis [71] used a registration method for finding epipolar lines in the left and right images. For each scan line of the left image, the method searched for the corresponding scan line of the right image such that the penalty of a criterion function is minimized in the given range.

Once the epipolar lines are registered, conjugate points are found by searching with a matching algorithm such as the edge-based method, the correlation method, and grey-level matching [24].

The edge-based method uses the difference in grey-level values across the edge for finding conjugate points. This method doesn't work well with blurred or smoothed images. The correlation method searches for similar brightness patterns on the epipolar line to find matched points. There are difficulties with small spatial variation and scaling of corresponding patches with this method. Similarly, grey-level matching matches two corresponding patches in the left and right images. Okutomi and Kanade [37] proposed the Sum of Squared Difference (SSD) method which belongs to grey-level matching. In the SSD method the sum of square-difference values are simply added over the given small window (see Figure 6.2).

6.2.2 Accuracy of Range Estimation

Stereo vision is a more accurate method than mono vision. The relative error in stereo is determined by the following formula.

$$\begin{aligned} D_s &= \frac{sb}{z} && (6.2) \\ \Rightarrow \delta D_s &= sb\frac{\delta z}{z^2} && (6.3) \\ \Rightarrow \frac{\delta z}{z} &= \frac{\delta D_s}{sb} z. && (6.4) \end{aligned}$$

In the DFF method, the lens step x is inversely proportional to the distance $z(= z_f)$. The relative error in focus is derived by

$$\begin{aligned} \frac{1}{z} &= ax + b && (6.5) \\ \Rightarrow \frac{\delta z}{z} &= a\delta x z. && (6.6) \end{aligned}$$

From the above relations, we see that the relative (percentage) error $\mid \frac{\delta z}{z} \mid$ in actual distance z increases linearly with distance. For our camera, using the Depth from Focus method, the constants were found to be a=0.0172 and b= – 0.1143 per meter [66]. δx can be set at the RMS error of 1.41 step which was obtained from the SFF.FIS method for a curved object using our camera system (see Table 1).

If we assume that stereo vision has sub-pixel accuracy in disparity, the value of δD_s can be chosen to be about 0.5 pixel. Therefore the accuracy ratio of focus and stereo is as follows.

$$\begin{aligned} \frac{\frac{\delta z}{z}\big|_{focus}}{\frac{\delta z}{z}\big|_{stereo}} &= a\delta x z\frac{sb}{\delta D_s z} \qquad (6.7) \\ &\approx 13.1 \end{aligned}$$

where $s \approx 35mm$, $a = 0.0172 \times 10^{-3}$, $b = 100mm$, $\delta D_s = 0.5 \times 0.013mm$, and $\delta x = 1.41$ step.

From above ratio we see that stereo vision gives a more accurate result than mono vision methods by about 13 times. However, stereo methods suffer from false matches with smoothed or blurred images. Its accuracy can be improved with focused images rather than blurred images.

6.3 Integration of Shape from Focus and Stereo

In this section, we discuss our combined method with focus and stereo.

A depth-map and two reconstructed images for the left and right sides were obtained from the SFF.FIS algorithm. A sequence of images $g(i,j,k)$ are recorded by moving the image detector to positions $z_i = z_0 + i * \delta$ where z_0 is an initial position and δ is a small displacement, for $i = 0, 1, 2, \cdots, I-1$, $j = 0, 1, 2, \cdots, J-1$, and $k = 0, 1, 2, \cdots, K-1$. Usually, $z_0 = f$. J and K are the number of rows and columns respectively in each image frame and I is the number of image frames (see Figure 5.2).

The two reconstructed images were used as two images for the left and right sides in stereo. The images are almost completely in focus, which improves the accuracy of matching. The focused images were normalized by dividing by the average values of a 256 × 256 image respectively to eliminate the different illumination effects in grey level.

The epipolar lines are obtained as follows. Four small circle markers are attached on the object surface at four corners of a square. Then the left and right images of the object are recorded. There are two horizontal lines and two vertical lines which connect two adjacent markers in the left and right images respectively. For each horizontal line of the left image, the corresponding horizontal line of the right image is determined. Using this relation, for each scan line $left_i$ of the left image, the corresponding scan line $right_i$ of the right image is computed. These scan lines are registered in two given 256 by 256 images for the left and right images.

6.3.1 Matching

With normalized images, the corresponding epipolar lines are registered for matching the conjugate points. A search is made for the image points in the precomputed range (i_0-MINDISPARITY$\leq i_m < i_0$+MAXDISPARITY) on the corresponding epipolar line. This, in our experiments, saves computation by about 60% as compared to the uncombined stereo method. i_0 is an initial value obtained from the SFF method and i_m is a search variable in pixel. MINDISPARITY and MAXDISPARITY are the minimum and maximum allowable disparities from the initial value i_0. The values

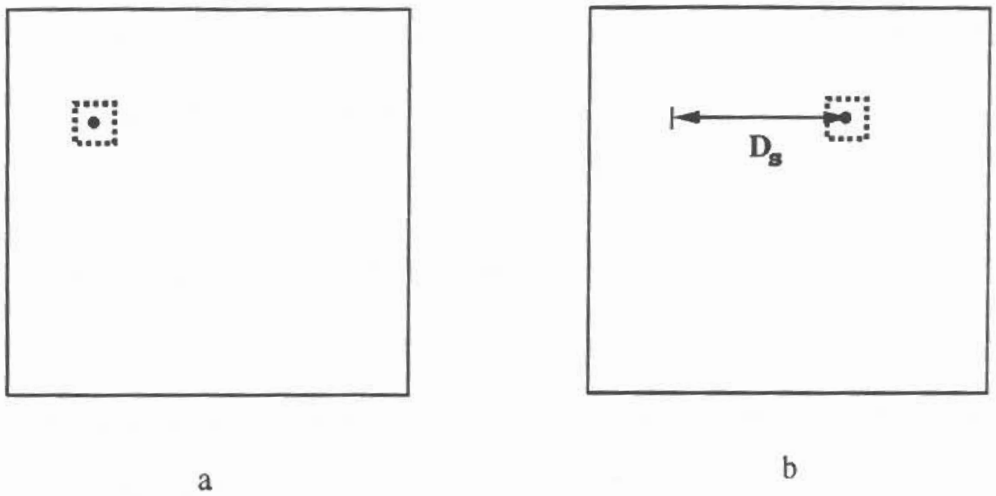
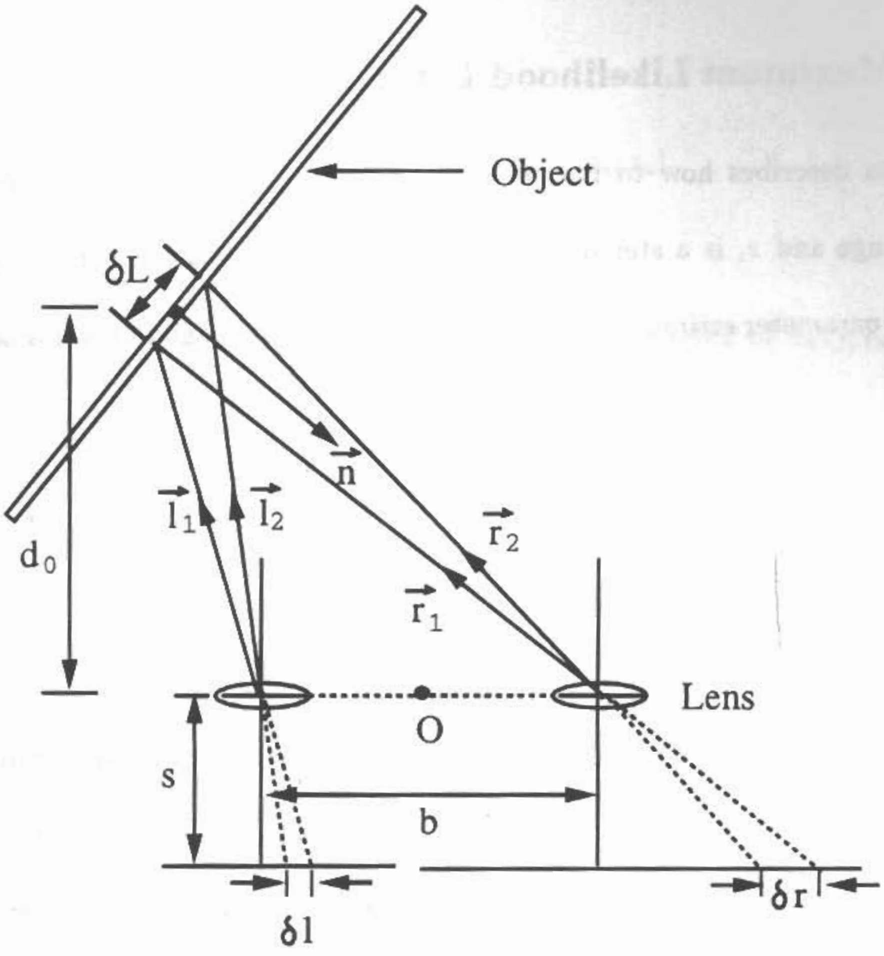


Figure 6.2: The Sum of Squared Difference (SSD) matching where D_s is a disparity of a pair of conjugate points, a: left image b: right image.

were taken to be 25 pixels (about 10 % of the whole search space on the epipolar line). The SSD method was used for matching conjugate points in the left and right images on the epipolar lines (see Figure 6.2).

In our combined method, a matching measure is computed over a small window (about 21 × 21) by the SSD method in two reconstructed images after normalizing. The problem here is how to determine a scale factor [32] for the corresponding window, particularly in the horizontal direction. In fact, the size of the corresponding window for a conjugate point is not exactly the same and is dependent on its surface orientation ($\vec{n}$) and the viewing direction ($\vec{l}$ or $\vec{r}$) from the principal plane of lens. The ratio of the two window sizes in the left and right images is $cos\theta_l$ to $cos\theta_r$ where $cos\theta_l$ is $\vec{n} \cdot \vec{l}$ and $cos\theta_r$ is $\vec{n} \cdot \vec{r}$ (see Figure 6.3).



igure 6.3: Scaling effect: $\frac{\delta l}{\delta r} = \frac{\cos\theta_l}{\cos\theta_r}$ where $\cos\theta_l = \vec{n}\cdot\vec{l}$, $\cos\theta_r = \vec{n}\cdot\vec{r}$, and all the ectors are unit vectors ($\vec{l_1} \simeq \vec{l_2} \simeq \vec{l}$ and $\vec{r_1} \simeq \vec{r_2} \simeq \vec{r}$ where $\delta l, \delta r \ll d_0$).

Therefore the scale factor in a given small window is determined by the lens geometry. This adjustment of the scale factor reduces error caused by the improperly scaled window size. Finally, searching for a minimum of SSD values is done by limiting the search range of matching on the epipolar lines.

6.3.2 Maximum Likelihood Estimation

This section describes how to integrate one pair of ranges (z_f, z_s) where z_f is a focusing range and z_s is a stereo range for one physical distance. This is a typical problem of parameter estimation [28] in the case of two *i.i.d.* random variables z_f, z_s, as follows.

$$z_f = z_{f\mu} + n_f \tag{6.8}$$

$$z_s = z_{s\mu} + n_s \tag{6.9}$$

where the observation errors n_f, n_s are assumed as Gaussian random variables with zero mean and variance σ_f^2, σ_s^2 (N(0,σ_f), N(0,σ_s)). Since $z_{f\mu}$ and $z_{s\mu}$ are unknown constants as a real range at each pixel for the same object, the two constants are identical, i.e. $z_{f\mu} = z_{s\mu} = z_\mu$. Each observation z_f or z_s consists of z_μ plus a Gaussian random variable n_f or n_s respectively.

The joint pdf of z_f and z_s is described by

$$f_{z_f z_s}(z_f, z_s; z_\mu) = f_{z_f}(z_f; z_\mu) f_{z_s}(z_s; z_\mu) \tag{6.10}$$

$$\equiv \quad L(z_f, z_s; z_\mu) \tag{6.11}$$

where $L(z_f, z_s; z_\mu)$ is called the likelihood function.

The value of z_μ that maximizes $L(z_f, z_s; z_\mu)$ is called the maximum likelihood estimator $\hat{z_\mu}$, that is,

$$L(z_f, z_s; \hat{z_\mu}) \quad \geq \quad L(z_f, z_s; z_\mu) \tag{6.12}$$

for $\forall\ z_\mu \in$ positive real number.

Since the log function is monotonic, the maximum value of $L(z_f, z_s; z_\mu)$ is also that of $\log L(z_f, z_s; z_\mu)$.

$$\begin{aligned} L(z_f, z_s; z_\mu) &= \frac{1}{2\pi\sigma_f\sigma_s} exp(-\frac{(z_f - z_\mu)^2}{2\sigma_f^2} - \frac{(z_s - z_\mu)^2}{2\sigma_s^2}). && (6.13) \\ log L(z_f, z_s; z_\mu) &= -\frac{(z_f - z_\mu)^2}{2\sigma_f^2} - \frac{(z_s - z_\mu)^2}{2\sigma_s^2} + log\frac{1}{2\pi\sigma_f\sigma_s}. && (6.14) \end{aligned}$$

Next, differentiate both sides of Eq. (6.14) with respect to z_μ and set the result equal to zero:

$$\frac{\partial log L(z_f, z_s; z_\mu)}{\partial z_\mu} \quad = \quad 0. \tag{6.15}$$

This yields

$$\hat{z_\mu} = \frac{\sigma_s^2 z_f + \sigma_f^2 z_s}{\sigma_s^2 + \sigma_f^2}. \tag{6.16}$$

Therefore the MLE of z_μ is the weighted sum of z_f and z_s and the weights are determined using the criterion that the likelihood function is a maximum. In section 6.2.2, the accuracy ratio of focus and stereo was found to be 13.1. The ratio can be used to find the RMS error in the case of stereo by dividing by 13.1 ($\sigma_s = \frac{\sigma_f}{N}$ where N=13.1).

The numerical value of $\hat{z_\mu}$ used in this chapter was found from Eq. (6.16) by

$$\begin{aligned}
\hat{z_\mu} &= \frac{(\frac{\sigma_f}{N})^2 z_f + \sigma_f^2 z_s}{(\frac{\sigma_f}{N})^2 + \sigma_f^2} && (6.17)\\
&= \frac{z_f + N^2 z_s}{1 + N^2} && (6.18)\\
&\approx 0.006 z_f + 0.994 z_s && (6.19)
\end{aligned}$$

where N= 13.1. Since $\hat{z}_f$ and z_s are random variables, $\hat{z_\mu}$ is also a random variable with mean z_μ and variance $\sigma_{\hat{z_\mu}}(N(z_\mu, \sigma_{\hat{z_\mu}}))$.

$\sigma^2_{\hat{z_\mu}}$ is derived by

$$\begin{aligned}
\sigma^2_{\hat{z_\mu}} &= \left(\frac{\sigma_s^2}{\sigma_s^2 + \sigma_f^2}\right)^2 \sigma_f^2 + \left(\frac{\sigma_f^2}{\sigma_s^2 + \sigma_f^2}\right)^2 \sigma_s^2 && (6.20)\\
&= \frac{\sigma_s^2 \sigma_f^2}{\sigma_s^2 + \sigma_f^2} && (6.21)\\
&= \frac{\sigma_f^2}{1 + \frac{\sigma_f^2}{\sigma_s^2}} && (6.22)\\
&= \frac{\sigma_s^2}{1 + \frac{\sigma_s^2}{\sigma_f^2}} && (6.23)\\
&< \sigma_f^2 \text{ or } \sigma_s^2, && (6.24)
\end{aligned}$$

(obviously $\frac{\sigma_f^2}{\sigma_s^2}$ *or* $\frac{\sigma_s^2}{\sigma_f^2} > 0$).

In the case of ratio 13.1, $\sigma^2_{\hat{z}_\mu}$ is $0.006\sigma_f^2$ or $0.994\sigma_s^2$, which shows that our combined method is more accurate in the sense of minimum variance. In fact the value σ_s is just a minimum variance of the observed range by stereo, without considering mismatches due to textures and edges, etc. on the object surface. Therefore the RMS error obtained by uncombined stereo methods is expected to be much greater than σ_s while the RMS error by the combined stereo method is close to σ_s.

6.3.3 SFFS Algorithm

An outline of the SFFS algorithm is summarized as follows.

- Find a depth-map and two reconstructed images for the left and right images using the SFF.FIS algorithm;

- Normalize the two reconstructed images by dividing by their average values respectively;

- Register the epipolar lines for matching conjugate points. For each scan line j of the left image, search for the corresponding scan line epipolar[j] of the right image using small circle markers on the object surface;

- Find conjugate points on the epipolar lines in the two reconstructed images. The detailed search algorithm by the SSD method is described as follows;

1. In every field of view, do the steps below.

2. Read as input the initial estimates of the disparity and scale factor of the given

small window(21 × 21) for finding conjugate points. Let i_0 be the initial estimate o the disparity, and p_0 be the initial estimate of the scale factor. Then the criterion function of the SSD algorithm is described by

$$SSD = \sum \sum_{j,k \in WINDOW} (f_l[j][k] - f_r[epipolar[j]][i_m + \lfloor p_m * k \rfloor])^2 \qquad (6.25)$$

where $f_l[j][k]$ and $f_r[j][k]$ are grey levels of a pixel (j,k) in the left and right images respectively, i_m is a position parameter, and p_m is a scale factor parameter.

3. Read as input the size of the search space for disparity and scale factor. Let the search space for disparity be the range i_{min}=i_0-MINDISPARITY, to i_{max}= i_0 + MAXDISPARITY, and the scale factor range be p_{min}=p_0 $-\delta p$, to p_{max}=p_0+δp. Take the values of minimum and maximum allowable disparities MINDISPARITY and MAXDISPARITY. Read as input the searching step sizes δ_i, δ_p for i, p, respectively.

4. Read image window size $2S$+1.

```
$minsum = 10^6$;

$i_p = i_0$; $p_p = p_0$;


for $i_m = i_{min}$ to $i_{max}$ in steps of $\delta_i$ do /* search position space

        for $p_m = p_{min}$ to $p_{max}$ in steps of $\delta_p$ do /* search scale factor

  /* compute matching measure of a candidate conjugate point */


            $sum$=0;
```

```
            for $j = -S$ to $S$ do
                for $k = -S$ to $S$ do
                    $sum = sum + (f_l[j][k] - f_r[epipolar[j]][i_m + \lfloor p_m * k \rfloor])^2$ ;
                end /* $k$ loop
            end /* $j$ loop
            if($sum < minsum$)
            then $i_p = i_m$; $p_p = p_m$; $minsum = sum$;
    end /* $p_m$ loop
end /* $i_m$ loop
```

i_p and p_p are the disparity and scale factor for the conjugate points; and

- Finally integrate one pair of ranges (z_f, z_s) obtained from focus and stereo.

$$MLE\ of\ z_\mu = \frac{\sigma_s^2 z_f + \sigma_f^2 z_s}{\sigma_s^2 + \sigma_f^2}.$$

6.4 Experimental Results

The proposed method was implemented on the SPARCS camera system. Here we present the result for a slanted planar object (Figures 6.7 and 6.8). The illumination

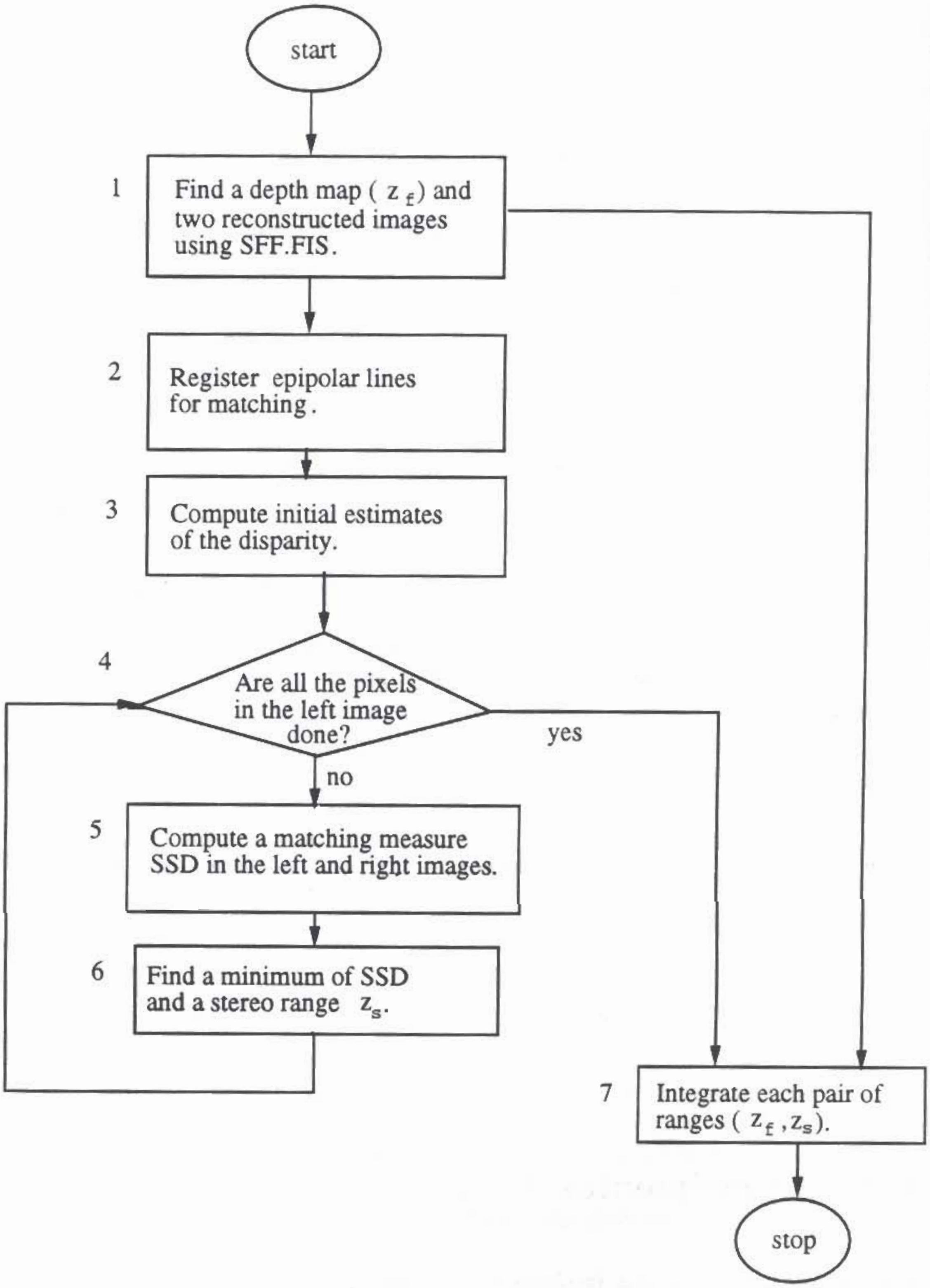


Figure 6.4: Flow Chart for SFFS algorithm.

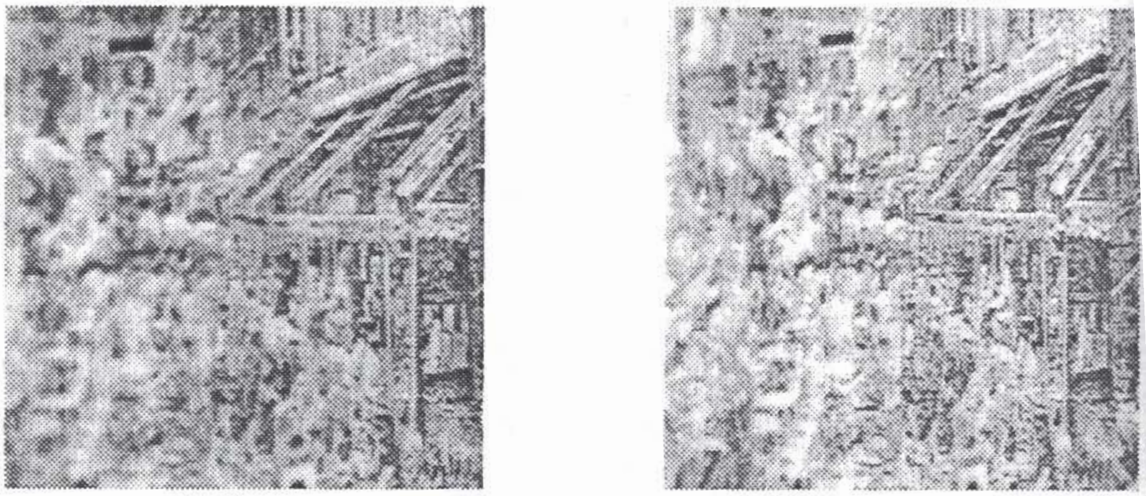
for the object was about 450 lux. Image size was 256 × 256. In order to reduce electronic noise, five image frames were time averaged for a fixed lens position.

The image sequence contained 97 image frames, one for each lens step position of the stepper motor. The window size for computing focus measures was 21 × 21.

An initial estimate of FIS was obtained by computing the sum of the square of image Laplacian for 9 image frames equally spaced (about 10 frames apart) in the original image sequence. The position of the maximum focus measure was first improved by a quadratic interpolation scheme using three points centered at the maximum point. The initial estimate of FIS is refined by searching for a maximum focus measure in the small cubic image volume space [59].

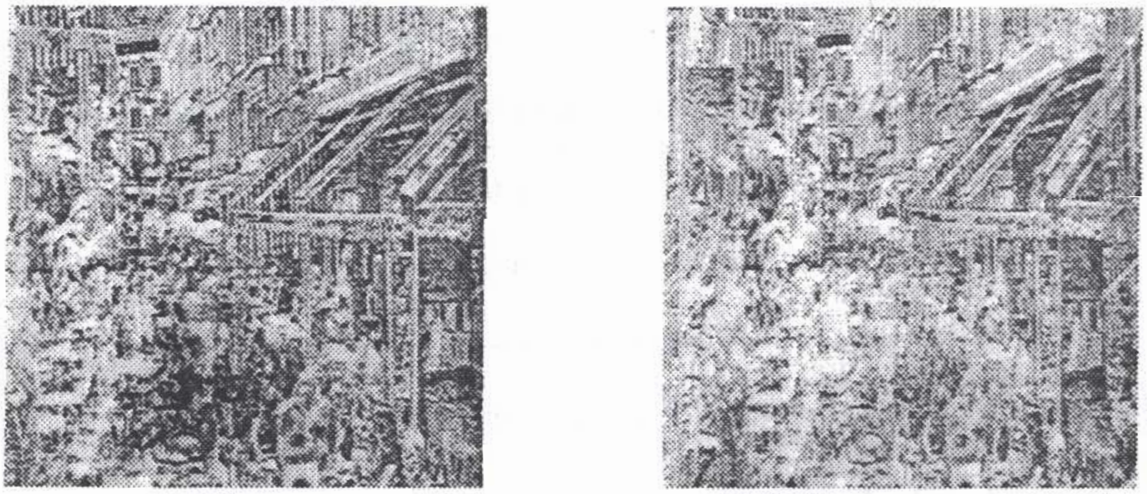
In the second phase of our algorithm, the results obtained by SFF.FIS limit the search range on the epipolar lines and compute the scale factor for the corresponding window size of conjugate point at each pixel. The optical axis of two cameras is almost in parallel. Four position markers using a circle are used to determine epipolar lines through calibration. The matched epipolar lines are registered for matching. Matching is done by the SSD method in the limited search range on the epipolar lines. The distance was determined by the stereo formula (Eq. (6.1)).

Next, one pair of ranges (z_f, z_s) are integrated at each pixel. The results for the slanted planar object are shown in Figures 6.7 and 6.8. The shape recovered by the uncombined stereo method is shown in Figure 6.7. The depth-map has a lot of noise, particularly in the regions where there is insufficient contrast. Figure 6.8 shows the



a b

Figure 6.5: One pair of images taken by a stereo camera system at lens step 60 for a slanted planar object, a: left image b: right image (shifted by 220 pixels to the left side).



a b

Figure 6.6: Reconstructed Images for a slanted planar object, a: left image b: right image (shifted by 220 pixels to the left side).

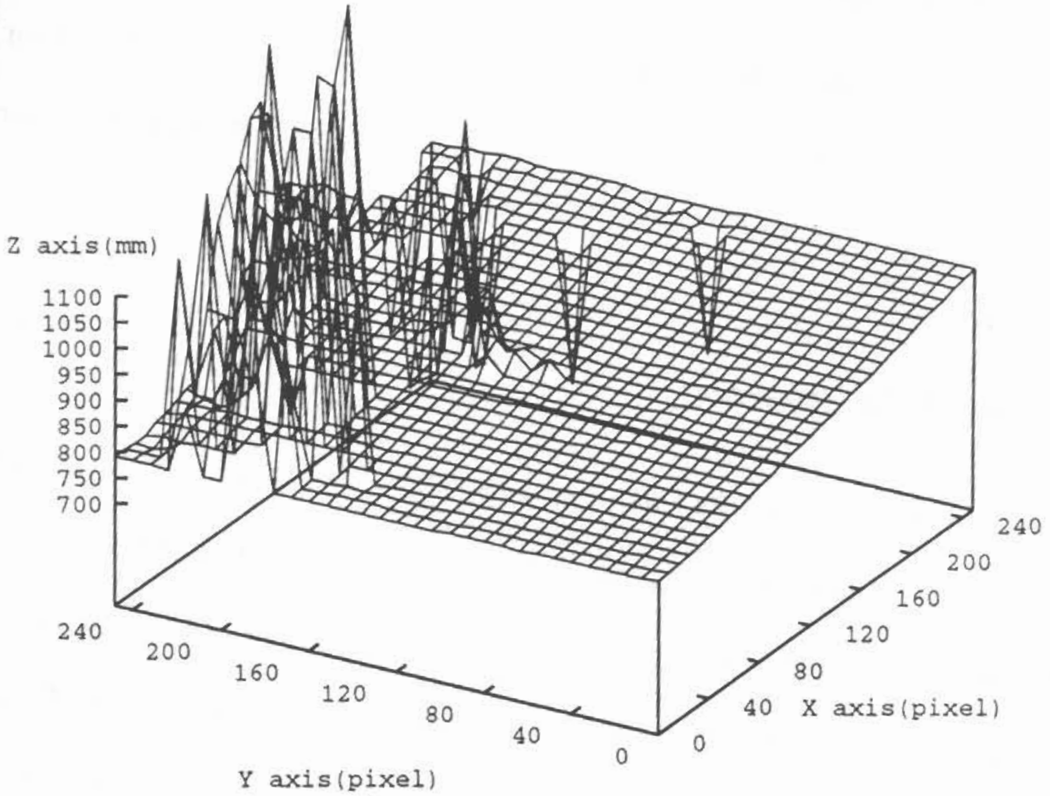


Figure 6.7: 3-D Depth Map for a slanted planar object by the uncombined stereo method.

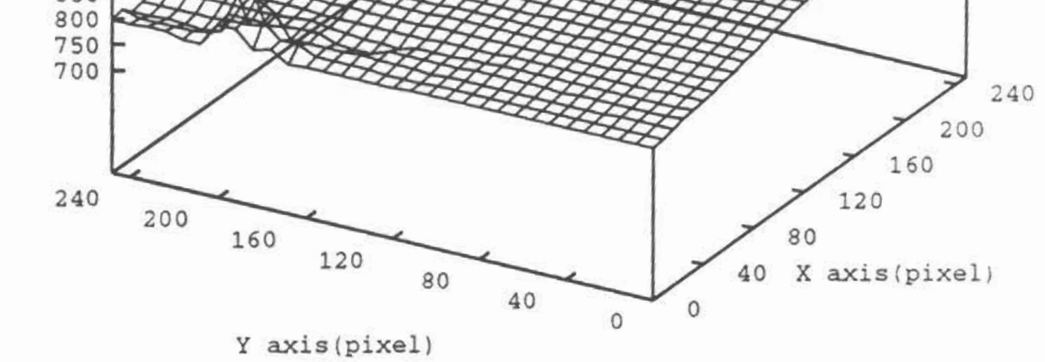


Figure 6.8: 3-D Depth Map for a slanted planar object by the combined method (SFFS).

recovered shape by the SFFS algorithm. Many false matches have been reduced in the regions with insufficient contrast. As a result, the depth-map surface is reasonably smooth.

6.5 Discussion

This section describes the execution of the SFFS algorithm. The time used for running our algorithm was 32 minutes CPU time on a SUN SPARC station. Focusing range takes 20 minutes and stereo range takes 12 minutes. On the other hand, the execution time for uncombined stereo range was 84 minutes CPU time. Therefore the computation time was reduced by about 60 % using the SFFS method, as compared to the uncombined stereo method. However, the running time can be reduced further if we optimize the search algorithm in the image volume space. In fact, SFFS requires a large number of repeated computations to determine focus measures at every pixel by sequential processing. This problem can be solved by pipelined parallel processing [29] using a parallel machine such as Transputer and Hypercube.

Chapter 7

Summary and Conclusions

7.1 Summary of the dissertation

In this thesis, we have presented a study on the image focus module for shape recovery and image reconstruction.

1. **Camera Model**

Two models of image defocusing were described: a paraxial geometric optics model and a diffraction optics model. Since the diffraction optics model uses the wave theory of light through a lens, it requires much computational time.

On the other hand, the geometric optics model is simple and requires less computational time. Also, its result is similar to the diffraction optics model. Therefore it is used as an adequate alternative to the diffraction model.

2. Image Focus Module

The concept of depth from focus involves a search of the camera parameter space. This technique is passive and does not require camera calibration. Next, the concept of depth from defocus involves determining the distance from two or three defocused images. This method is also passive and requires accurate calibration of the camera parameters.

3. Focusing

A set of DFF techniques are derived from the paraxial geometric optics model as follows.

- Image Energy;
- Energy of image gradient;
- Energy of image Laplacian;
- Variance of low-pass filtered image;
- Energy of low-pass filtered image gradient; and
- Energy of low-pass filtered image Laplacian.

If an image has a high spatial frequency content in the side lobes, then focus measures may exhibit local maxima. The side lobe effect can be eliminated by a Gaussian low-pass filter which removes the side lobes. The cut off frequency ρ_0 of the Gaussian filter should be $1.63\pi/R_{max}$ where R_{max} is the maximum blur circle radius.

The three objects tested for the proposed focusing techniques were:

- a cartoon
- picture of a face
- a "pin-hole" light source

The last object has a high spatial frequency content and was chosen to observe its effect.

4. Shape Recovery from Image Focus

Shape-from-Focus methods provide a solution to the shape recovery problem. A new SFF method called SFF.FIS was proposed for more accurate shape recovery. The basic idea of this method is that focus measures are computed over the focused image surface.

The computational implementation of this idea involves two steps: The first, estimate an approximate FIS using one of the traditional SFF methods; Then refine the FIS by searching for an FIS shape to maximize the focus measure computed over the pixels lying on it. The search is local and therefore computationally efficient.

This method requires a sequence of image frames recorded with different camera parameter settings. This method is more accurate than traditional methods. However, its running time is slow because there are many pixels which require the same numerical operations to compute a focus measure.

The tested objects are a slanted plane and a cone of length about 79 inches and base diameter of about 15 inches.

5. Combining Shape from Focus and Stereo

In this method, a depth map and two focused images were obtained by the SFF.FIS algorithm to solve the correspondence problem in stereo-vision.

Next, conjugate points on the epipolar lines were found in two reconstructed images. Finally, one pair of ranges $(z_f,\ z_s)$ obtained from focus and stereo were integrated using maximum likelihood estimation method.

7.2 Conclusions

We have developed a set of DFF techniques based on the paraxial geometric optics model. They include energy maximization of unfiltered, low-pass filtered, high-pass filtered, and band-pass filtered images. It was shown that in the presence of high spatial frequencies, noise, and aliasing, focusing techniques based on band-pass filters perform well. Theoretical and experimental results suggest that one of the focus measures M_2' has better overall characteristics. Therefore it is recommended for use in practical applications. At some additional computational cost, better performance may be obtained by computing two or more focus measures and making judgements based on multiple measures rather than a single measure.

Next, a new Shape-from-Focus method has been proposed based on maximizing focus measures computed directly over the Focused Image Surface (FIS). This method corresponds to a piece-wise planar approximation of the object's shape as opposed to a piece-wise constant approximation adopted by SFF methods in the previous literature. We have experimentally demonstrated the effectiveness of our SFF method on real-world examples. The method here can be extended to obtain piece-wise quadratic and higher order approximations to FIS at the cost of additional computation. The RMS errors between the SFF.FIS method and the traditional method were computed. For the traditional method the RMS error was 2.22 lens steps out of 97 steps whereas the RMS error was 1.41 lens steps (out of 97 steps) for our SFF method. Therefore the SFF.FIS method is more accurate than the traditional method by about 36%.

Finally, we have described a combined method with focus and stereo for shape recovery. A depth map and two focused images obtained by the SFF.FIS algorithm are used to solve the correspondence problem efficiently in stereo vision. The experimental results indicate that our SFFS method gives much more accurate results than the uncombined stereo method. The number of false matches have been reduced dramatically in the regions with insufficient contrast and the recovered surface is reasonably smooth.

Appendix A

Focus Measures for Geometric Optics Model

In this appendix, we use the following notation and relations.

$f(x, y)$ is the focused image of an object and $F(\omega, \nu)$ is its Fourier transform where ω, ν and $\rho = \sqrt{\omega^2 + \nu^2}$ are spatial frequencies expressed in radians/unit distance. $g_1(x, y)$ and $g_2(x, y)$ are two normalized images of the object recorded by a camera with camera parameters $\mathbf{e}_1$ and $\mathbf{e}_2$. R_1 and R_2 are the normalized blur circle radii corresponding to g_1 and g_2 respectively. $H_1(\rho)$ and $H_2(\rho)$ are the camera OTFs corresponding to g_1 and g_2 respectively. $G_1(\omega, \nu)$ and $G_2(\omega, \nu)$ are the Fourier transforms of g_1 and g_2 respectively. Defocusing is a convolution operation and the following relations hold:

$$G_1(\omega,\nu) = H_1(\rho)F(\omega,\nu) \tag{A.1}$$

$$G_2(\omega,\nu) = H_2(\rho)F(\omega,\nu). \tag{A.2}$$

Definition of the focus measure $M_0(i)$:

$$M_0(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid G_i(\omega,\nu) \mid d\omega d\nu \text{ for } i = 1,2. \tag{A.3}$$

$\alpha(x)$ is a monotonically increasing function such that if $0 \leq x_1 < x_2$ then $\alpha(x_1) < \alpha(x_2)$.

Definition of the focus measure $M_0'(i)$:

$$M_0'(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \alpha(\mid G_i(\omega,\nu) \mid) d\omega d\nu \text{ for } i = 1,2. \tag{A.4}$$

$L(\omega,\nu)$ is the transfer function of a filter.

Definition of the focus measure $M_0''(i)$:

$$M_0''(i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid L(\omega,\nu) \mid \alpha(\mid G_i(\omega,\nu) \mid) d\omega d\nu. \tag{A.5}$$

The first theorem shows that $M_0(i)$ is a sound focus measure assuming that the OTF is zero outside the main lobe. It increases monotonically as the blur circle radius decreases (i.e. the image blur decreases) and reaches a maximum when the image is in best focus (i.e. the blur circle radius is zero). The Theorem also shows that $M_0(i)$

has a unique maximum and therefore has no local maxima. In this theorem, the paraxial geometric optics model is used for the OTF, except that the magnitude of the OTF outside its main lobe is taken to be zero.

Theorem 2 *If*

$$H_1(\rho) = \begin{cases} \frac{2J_1(R_1\rho)}{R_1\rho} & \textit{for } 0 \leq R_1\rho \leq 1.22\pi \\ 0 & \textit{elsewhere} \end{cases} \tag{A.6}$$

$$H_2(\rho) = \begin{cases} \frac{2J_1(R_2\rho)}{R_2\rho} & \textit{for } 0 \leq R_2\rho \leq 1.22\pi \\ 0 & \textit{elsewhere} \end{cases} \tag{A.7}$$

and

$$\mid R_2 \mid \; > \; \mid R_1 \mid \; > \; 0 \tag{A.8}$$

then

(i) $M_0(1) > M_0(2)$ *and*

(ii)

$$\lim_{|R_1| \to 0} M_0(1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mid F(\omega, \nu) \mid d\omega d\nu.$$

Proof (i) $\mid R_2 \mid \; > \; \mid R_1 \mid$

$\Longrightarrow 1 > \mid H_1(\rho) \mid > \mid H_2(\rho) \mid \geq 0$ for $0 \; < \; \rho \; \leq \; \frac{1.22\pi}{R_1}$

$\Longrightarrow \mid H_1(\rho) \mid\mid F(\omega, \nu) \mid \; > \; \mid H_2(\rho) \mid\mid F(\omega, \nu) \mid$

$\Longrightarrow \mid G_1(\omega, \nu) \mid \; > \; \mid G_2(\omega, \nu) \mid$

$\Longrightarrow$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mid G_1(\omega, \nu) \mid d\omega d\nu \; > \; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mid G_2(\omega, \nu) \mid d\omega d\nu$$

$\Longrightarrow M_0(1) \ > \ M_0(2)$

(ii)

$$\lim_{|R_1|\to 0} M_0(1) = \lim_{|R_1|\to 0} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid H_1(\rho) \mid\mid F(\omega,\nu) \mid d\omega d\nu$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid F(\omega,\nu) \mid d\omega d\nu$$

because

$$\lim_{|R_1|\to 0} H_1(\rho) = 1.$$

Theorem 3 is the same as Theorem 2, except that the OTF corresponds to a Gaussian. In this case the OTF has no side lobes and therefore no assumptions are made. The proof of this theorem is similar to that of Theorem 2.

Theorem 3 *If*

$$H_1(\rho) = e^{-\frac{\rho^2 r_1^2}{2}}, \quad H_2(\rho) = e^{-\frac{\rho^2 r_2^2}{2}} \tag{A.9}$$

$$r_1 = cR_1, \ \ r_2 = cR_2 \qquad c \textit{ is a proportionality constant} \tag{A.10}$$

and

$$\mid R_2 \mid \ > \ \mid R_1 \mid \ > \ 0 \tag{A.11}$$

then

(i) $M_0(1) \;>\; M_0(2)$ *and*

(ii)

$$\lim_{|R_1| \to 0} M_0(1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\, F(\omega, \nu) \,|\, d\omega d\nu.$$

$M_0'(i)$ is a focus measure defined as the volume integral of a monotonically increasing function of $|G(\omega, \nu)|$. Typical examples of the monotonically increasing function $\alpha(x)$ are $\alpha(x) = x^2$ or $\alpha(x) = x^n$ for $n > 0$. Theorem 4 shows that $M_0'(i)$ has properties similar to $M_0(i)$. With minor exceptions, the proof of this theorem is similar to Theorem 2.

Theorem 4 *If the conditions (A.6) to (A.8) in Theorem 2 are satisfied, then*

$$(i) M_0'(1) \;>\; M_0'(2) \tag{A.12}$$

and

$$(ii) \lim_{|R_1| \to 0} M_0'(1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \alpha(|\, F(\omega, \nu) \,|) d\omega d\nu. \tag{A.13}$$

Proof Similar to Theorem 2; the main step to be noted is that

$|\, G_1(\omega, \nu) \,| \;>\; |\, G_2(\omega, \nu) \,|$

$\Longrightarrow \alpha(|\, G_1(\omega, \nu) \,|) \;>\; \alpha(|\, G_2(\omega, \nu) \,|).$

Theorem 5 *If conditions (A.9) to (A.11) are satisfied, then relations (A.12) and (A.13) will be true.*

Theorem 6 shows that a focus measure which works correctly for an image sequence will also work correctly if the entire image sequence is filtered by the same filter $L(\omega,\nu)$. Therefore, band pass filtering the image sequence will not affect the monotonicity or the location of the maximum of a focus measure.

Theorem 6 *If conditions (A.6) to (A.8) are satisfied, then*

$$(i) M_0''(1) \ > \ M_0''(2) \tag{A.14}$$

and

$$(ii) \lim_{|R_1|\to 0} M_0''(1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid L(\omega,\nu) \mid\mid F(\omega,\nu) \mid d\omega d\nu. \tag{A.15}$$

Proof Similar to the proof of Theorem 2; the main step to be noted is that

$\mid G_1(\omega,\nu) \mid \ > \ \mid G_2(\omega,\nu) \mid$

$\Longrightarrow \mid L(\omega,\nu) \mid\mid G_1(\omega,\nu) \mid \ > \ \mid L(\omega,\nu) \mid\mid G_2(\omega,\nu) \mid .$

Theorem 7 *If conditions (A.9) to (A.11) are satisfied, then relations (A.14) and (A.15) will hold.*

Proof Similar to Theorem 6.

Bibliography

[1] A. Abbott and N. Ahuja, "Surface reconstruction by dynamic integration of focus, camera vergence, and stereo", *Proceedings of Second International Conference on Computer Vision*, pp. 532-543, December 1988.

[2] N. Ahuja and A. Abbott , "Active Stereo: Integrating Disparity, Vergence, Focus, Aperture, and Calibration for Surface Estimation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-15, No. 10, pp. 1007-1029, October 1993.

[3] J. Aloimonos, I. Weiss, and A. Bandopadhyay, "Active Vision", *International Journal of Computer Vision*, pp. 333-356, 1988.

[4] A. Asada and J. Slotine, *Robot Analysis and Control*, John Wiley & Sons, 1986.

[5] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, 1982.

[6] R. Bajcsy, " Active perception vs. passive perception", *Proceedings of Workshop on Computer Vision*, pp. 55-59, Bellaire, MI, October 1985.

[7] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Oxford, Sixth Edition, 1980.

[8] P. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code", *IEEE Transactions on Communication*, Vol. COM-31, pp. 532-540, 1983.

[9] C. T. Chen, *Linear System Theory and Design*, Holt, Rinehart and Winston, 1984.

[10] T. Darrell and K. Wohn, "Pyramid based depth from focus", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, pp.504-509 , June 1988.

[11] S. Das and N. Ahuja, "Integrating Multiresolution Image Acquisition and Coarse-to-fine Surface Reconstruction from Stereo", *Proceedings of Third International Conference on Computer VIsion*, pp. 361-365, December 1990.

[12] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo-A Review", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 6, pp. 1489-1510. December 1989.

[13] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.

[14] J. Ens, "An Investigation of Methods for Determining Depth from Focus". Ph.D. Dissertation, University of British Columbia, 1992.

[15] J. Ens and P. Lawrence, "An Investigation of Methods for Determining Depth from Focus", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-15, No. 2, pp. 97-108, February 1993.

[16] F. Fuma, E. Krotkov, and J. Summers, "The Pennsylvania Active Camera System", Grasp Lab. 62, MS-CIS-86-15, University of Pennsylvania, February 1986.

[17] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics,* John Wiley & Sons, 1978.

[18] J. Goodman, *Introduction to Fourier Optics,* McGraw-Hill Book Co., 1968.

[19] P. Grossmann, "Depth from focus", *Pattern Recognition Letters* , Vol. 5, pp. 63-69, January 1987.

[20] E. Hecht, *Optics,* Addison-Wesley Publishing Co., 1987.

[21] W. Hoff and N. Ahuja, "Extracting Surfaces from Stereo Images: An Integrated Approach", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp.284-294 , 1987.

[22] H. H. Hopkins, "The frequency response of a defocused optical system", Proc. Royal Society of London, A231, pp. 91-103, 1955.

[23] B. K. P. Horn, "Focusing", Artificial Intelligence Memo No. 160, MIT, May 1968.

[24] B. K. P. Horn, *Robot Vision,* McGraw-Hill Book Company, 1986.

[25] R. A. Jarvis, "Focus optimization criteria for computer image processing", Microscope 24, pp.163, 1976.

[26] R. A. Jarvis, "A perspective on range finding techniques for computer vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 2, pp. 122-139, March 1983.

[27] E. Krotkov, "Focusing", *International Journal of Computer Vision*, 1, pp. 223-237, 1987.

[28] E. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo,* Springer-Verlag, New York, 1989.

[29] S. Y. Kung, *VLSI Array Processors,* Prentice Hall, 1988.

[30] S. Lai and C. Fu, "A Generalized Depth Estimation Algorithm with a Single Image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-14, No. 4, pp. 405-411, April 1992.

[31] G. Ligthart and F. Groen, "A Comparison of Different Autofocus Algorithms", *International Conference on Pattern Recognition*, pp. 597-600,1982.

[32] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-10, No. 5, pp. 713-720, September 1988.

[33] H. Longuet-Higgins, "Multiple interpretations of a pair of images of a surface", Proc. Royal Society of London, A418, pp. 1-15, 1988.

[34] M. C. Lu, "Computer Modeling and Simulation Techniques for Computer Vision Problems", Ph.D. Dissertation, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, February 1993.

[35] D. Marr, *Vision,* Freeman, San Francisco, CA, 1982.

[36] S. Nayar, "Shape from Focus System for Rough Surfaces", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 302-308, June 1992.

[37] M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.63-69, 1991.

[38] A. P. Pentland, "Depth of scene from depth of field", *Proceedings of Image Understanding Workshop*, pp. 253-259, April 1982.

[39] A. P. Pentland, "A new sense for depth of field", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 4, pp. 523-531, July 1987.

[40] A. P. Pentland, T. Darrell, M. Turk, and W. Huang, "A simple, real-time range camera", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, pp. 256-261, June 1989.

[41] R. A. Roberts and C. T. Mullis, *Digital Signal Processing*, Addison-Wesley Publishing Co.,1987.

[42] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Vol. 1, Academic Press, 1982.

[43] J. F. Schlag, A. C. Sanderson, C. P. Neuman, and F. C. Wimberly, "Implementation of automatic focusing algorithms for a computer vision system with camera control", CMU-RI-TR-83-14, Robotics Institute, Carnegie-Mellon University, 1983.

[44] W.F. Schreiber, *Fundamentals of Electronic Imaging Systems*, Springer-Verlag, New York, 1986.

[45] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1986.

[46] M. Subbarao, "Direct Recovery of Depth-map", Tech. Report No. 87-02, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350, Feruary, 1987.

[47] M. Subbarao, "Direct Recovery of Depth-map II: A New Robust Approach", Tech. Report No. 87-03, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350, March, 1987.

[48] M. Subbarao, "Direct Recovery of Depth-map I: Differential methods", *Proceedings of the IEEE Computer Society Workshop on Computer Vision,* pp. 58-65, December 1987.

Tech. Report No. 87-02, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350, Feruary, 1987.

[49] M. Subbarao, "Parallel depth recovery by changing camera parameters", *Second International Conference on Computer Vision,* Florida, USA, pp. 149-155, December 1988.

[50] M. Subbarao, "Efficient depth recovery through inverse optics", Editor: H. Freeman, *Machine Vision for Inspection and Measurement,* Academic press, Boston, pp. 101-126, 1989.

[51] M. Subbarao, "Computational methods and electronic camera apparatus for determining distance of objects, rapid autofocusing, and obtaining improved focus images", U.S. patent application serial number 07/373,996, June 1989 (pending).

[52] M. Subbarao, "Determining distance from defocused images of simple objects", Tech. Report No. 89.07.20, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

[53] M. Subbarao, "On the Depth Information in the Point Spread Function of a Defocused Optical System", Tech. Report No. 90.02.07, Computer Vision Lab-

oratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

[54] M. Subbarao, "Optical transfer function of a diffraction-limited system for polychromatic illumination", *Applied Optics*, Vol. 29 No. 4, pp. 554-558, February 1990.

[55] M. Subbarao, "Spatial Domain Convolution/Deconvolution Transform", Tech. Report No. 91.07.03, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

[56] M. Subbarao, T. Choi, and A. Nikzad, "Focusing Techniques", *OE/BOSTON '92, SPIE conference Proceedings*, Vol. 1823, Boston, pp.163-174 , November 15-20,1992.

[57] M. Subbarao, T. Choi, and A. Nikzad, "Focusing Techniques", *the Journal of Optical Engineering*, Vol. 32 No. 11, pp. 2824-2836, November 1993.

[58] M. Subbarao and T. Choi, "A New Method for Shape from Focus", *SPIE's International Symposium on Optical Tools for Manufacturing and Advanced Automation*, Vol. 2064(08), September 7-10, 1993.

[59] M. Subbarao and T. Choi, "Accurate Recovery of Three-Dimensional Shape from Image Focus", (*submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*).

[60] M. Subbarao and T. Choi, *Method and Apparatus for Accurate 3-D Shape Measurement and Image Reconstruction using Image Focus*, U.S. patent application, July 1993.

[61] M. Subbarao, and G. Surya, "Application of Spatial-Domain Convolution / Deconvolution Transform for Determining Distance from Image Defocus", *OE/BOSTON '92, SPIE conference Proceedings*, Vol. 1822, Boston, November 15-20,1992.

[62] M. Subbarao and N. Gurumoorthy, "Depth Recovery from blurred edges", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, pp. 498-503 , June 1988.

[63] M. Subbarao, and M. C. Lu, "Computer Modeling and Simulation of Camera Defocus", Tech. Report No. 92.01.16, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992.

[64] M. Subbarao and A. Nikzad, "A model for image sensing and digitization in machine vision", *OE/BOSTON '90, SPIE conference*, Boston, November 1990.

[65] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing: A Practical Approach", Tech. Report No. 92.01.17, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992. (An abridged version of this appears in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp.773-776 , June 15-18,1992.

[66] M. Subbarao and T. Wei, "Fast Determination of Distance and Autofocusing from Image Defocus: A New Fourier Domain Approach", (*submitted to IEEE Transactions on Robotics and Automation.*)

[67] R. Y. Tsai, "Multiframe Image Point Matching and 3-D Surface Reconstruction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 2, pp. 159-174, March 1983.

[68] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses", *IEEE Journal of Robotics and Automation* , Vol. RA-3(4), pp. 323-344, August 1987.

[69] J. M. Tenenbaum, "Accommodation in Computer Vision", Ph.D. Dissertation, Stanford University, November 1970.

[70] B. Topielski, "A Computer Controlled Lens Positioning System for Autofocusing", Tech. Report No. 90.11.17, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992.

[71] Y. P. Wang and T. Pavlidis, "Optimal Correspondence of String Subsequences", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-12, No. 11, pp. 1080-1087, November 1990.