

# **Analysis and Application of Autofocusing and Three-Dimensional Shape Recovery Techniques based on Image Focus and Defocus**

A Dissertation Presented

by

Jenn-Kwei Tyan

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 1997

Copyright © by  
Jenn-Kwei Tyan  
1997

State University of New York  
at Stony Brook

The Graduate School

Jenn-Kwei Tyan

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

---

Muralidhara Subbarao, Advisor, Associate Professor  
Department of Electrical Engineering

---

Hang-Sheng Tuan, Professor, Chairman  
Department of Electrical Engineering

---

Nam Phamdo, Assistant Professor  
Department of Electrical Engineering

---

Vaclav Dolezal, Professor  
Department of Applied Mathematics & Statistics

This dissertation is accepted by the Graduate School.

---

Graduate School

**Abstract of the Dissertation**  
**Analysis and Application of Autofocusing**  
**and Three-Dimensional Shape Recovery**  
**Techniques based on Image Focus and**  
**Defocus**

by

Jenn-Kwei Tyan

Doctor of Philosophy

in

Electrical Engineering

State University of New York at Stony Brook

1997

Autofocusing and three-dimensional (3D) shape recovery techniques based on image focus and defocus are analyzed and their practical applications are demonstrated. Focus measures based on summing the absolute values of image derivatives are used by many researchers in the past. We first investigated the unsoundness of those focus measures. We also argued that energy of the Laplacian of the image is

a good focus measure and is recommended for use in practical applications. Important properties of the Laplacian focus measure are investigated. Application of the Laplacian focus measure to 3D microscopy is demonstrated.

The optimal focus measure for a noisy camera in passive search based autofocusing (AF) and depth-from-focus (DFF) applications depends not only the camera characteristics but also the image of the object being focused or ranged. In the early stage of this research, a new metric named *Autofocusing Uncertainty Measure* (AUM) was defined which is useful in selecting the most accurate focus measure from a given set of focus measures. AUM is a metric for comparing the noise sensitivity of different focus measures. In the later stage of this research, an improved metric named *Autofocusing Root-Mean-Square Error* (ARMS error) was defined. Explicit expressions have been derived for both AUM and ARMS error, and the two metrics are shown to be related by a monotonic expression. The theories are verified by experiments as well as computer simulations.

Another ranging method, depth-from-defocus (DFD) using the Spatial Domain Convolution/Deconvolution Transform Method (STM), is an useful technique for autofocusing and 3D shape recovery. The noise sensitivity analysis of STM is investigated. A theoretical treatment of this problem provides

the accuracy check of STM in the presence of noise which has been done only under experimental observation. The derived theoretical results and supporting experimental results are presented.

Finally, the integration of DFF and DFD methods is developed for fast and accurate 3D shape recovery. This has been demonstrated successfully with experiments on a prototype camera system.

To my parents, wife and son

# Contents

<b>Abstract</b>	<b>iii</b>
<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xv</b>
<b>Acknowledgements</b>	<b>xvi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation	1
1.2 Literature Review	4
1.2.1 Depth from Focus (DFF)	4
1.2.2 Depth from Defocus (DFD)	6
1.2.3 Combining with Stereo Ranging	9
1.3 Dissertation Overview	9
<b>2 Shape from Focus and its Application to 3D Microscopy</b>	<b>14</b>
2.1 Introduction	14
2.2 Focused and Defocused Images	16
2.3 Model of Focus Measures	19



2.3.1	Review of Focus Measures . . . . .	22
2.3.2	Unsound Focus Measures . . . . .	24
2.3.3	The Laplacian Filter . . . . .	28
2.4	Shape from Focus (SFF) . . . . .	33
2.5	Implementation . . . . .	37
2.5.1	Microscope Camera System . . . . .	37
2.5.2	Experimental Results . . . . .	39
2.6	Summary and Conclusion . . . . .	49
<b>3</b>	<b>Noise Sensitivity Analysis of Depth from Focus . . . . .</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.2	Autofocusing Algorithm . . . . .	54
3.3	A Metric for Focus Measures . . . . .	57
3.3.1	Autofocusing Uncertainty Measure (AUM) . . . . .	59
3.4	ARMS Error . . . . .	66
3.4.1	Relation between AUM and ARMS Error . . . . .	69
3.5	Noise Sensitivity Analysis . . . . .	70
3.6	Experiments . . . . .	77
3.6.1	Computer Simulation Experiments . . . . .	80
3.7	Conclusion . . . . .	89
<b>4</b>	<b>Noise Sensitivity Analysis of Depth from Defocus . . . . .</b>	<b>91</b>
4.1	Introduction . . . . .	91
4.2	Camera Model and Defocused Image . . . . .	93
4.3	STM: A Spatial Domain Approach to DFD . . . . .	95

4.4	Noise Sensitivity Analysis of STM . . . . .	97
4.5	Experiments . . . . .	102
4.5.1	Experiments with Synthetic Images . . . . .	102
4.5.2	Experiments on Real Images . . . . .	107
4.6	Conclusion . . . . .	108
<b>5</b>	<b>Integration of Depth from Focus and Defocus . . . . .</b>	<b>114</b>
5.1	Introduction . . . . .	114
5.2	Camera System . . . . .	116
5.3	Camera Calibration . . . . .	118
5.3.1	Lens Step vs Focusing Distance . . . . .	119
5.3.2	Blur Parameter vs Focused Lens Step . . . . .	121
5.3.3	Thresholds vs System Noise . . . . .	124
5.3.4	Other Considerations . . . . .	126
5.4	The Integration of DFF and DFD . . . . .	126
5.4.1	Autofocusing . . . . .	127
5.4.2	3D Shape Recovery . . . . .	129
5.5	Conclusion . . . . .	136
<b>6</b>	<b>Summary, Conclusions and Future Research . . . . .</b>	<b>138</b>
6.1	Overview . . . . .	138
6.2	Summary of Depth from Focus and Defocus . . . . .	139
6.3	Conclusions . . . . .	140
6.4	Future Research . . . . .	142

<b>A</b>	<b>Related Derivations in Chapter 3</b>	<b>146</b>
A.1	Variance of Focus Measures	146
A.2	Gradient Focus Measure	150
A.3	Another Approach for RMS Error in DFF	152
A.4	Wave Optics for PSF	155
<b>B</b>	<b>Related Derivations in Chapter 4</b>	<b>157</b>
B.1	The Variance of STM	157
	<b>Bibliography</b>	<b>160</b>

## List of Figures

2.1	Image Formation in a Convex Lens . . . . .	16
2.2	The OTF of a blurred camera attenuates high frequencies. The attenuation effect increases with increasing value of the blur circle radius $R$ . . . . .	19
2.3	Model of Image Sensing . . . . .	21
2.4	Model of Focus Measure. The FMF for gradient along the $x$ and $y$ axis are $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ respectively and the FMF for the Laplacian is $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ . . . . .	21
2.5	Frequency responsed of various filters . . . . .	23
2.6	A simple 1D example. Based on $\int \left  \frac{dg(x)}{dx} \right  dx$ , it is seen that $\gamma_{11} = \gamma_{12} = \gamma_{13} = \gamma_{14}$ regardless how the PSF is changed. If focus measure is computed based on $\int \left  \frac{dg(x)}{dx} \right ^2 dx$ , according to the Schwarz inequality we can obtain $\gamma_{11} > \gamma_{12} > \gamma_{13} > \gamma_{14}$ . . . . .	26
2.7	Blurring does not change the image where the function is locally linear. . . . .	29
2.8	Blurring changes the image where the function is locally quadratic. . . . .	29
2.9	Shape from focus . . . . .	35

2.10 Block diagram . . . . .	42
2.11 Automated SFF microscope system . . . . .	42
2.12 Some defocused images of a mustard seed . . . . .	43
2.13 Reconstructed image of a mustard seed . . . . .	44
2.14 Depth map of the mustard seed . . . . .	44
2.15 Some defocused images of a metal scratch . . . . .	45
2.16 Reconstructed image of a metal scratch . . . . .	46
2.17 Depth map of the metal scratch . . . . .	46
2.18 Some defocused images of a BGA ball . . . . .	47
2.19 Reconstructed image of a BGA ball . . . . .	48
2.20 Depth map of the BGA ball . . . . .	48
3.1 Autofocusing Algorithm . . . . .	54
3.2 Definition of AUM at the focused position $S_f$ . . . . .	63
3.3 Comparison of two focus measures $\gamma_1$ and $\gamma_2$ at the focused position . . . . .	63
3.4 Definition of AUM at a position $S_0$ far from the focused position $S_f$ . . . . .	64
3.5 Comparison of two focus measures $\gamma_1$ and $\gamma_2$ at a position $S_0$ far from the focused position $S_f$ . Note: $\sigma_2 > \sigma_1$ but $AUM2 < AUM1$ , therefore $\gamma_2$ is better than $\gamma_1$ . . . . .	64
3.6 quadratic polynomial interpolation . . . . .	65
3.7 linear interpolation . . . . .	65
3.8 The verification of variance of focus measures in SPARCS system . . . . .	82

3.9	The verification of ARMS errors in SPARCS system . . . . .	83
3.10	The verification of ARMS errors in computer simulation . . . . .	84
3.11	Noise characteristics in 1-D . . . . .	85
3.12	Textured images . . . . .	86
3.13	Simulation records . . . . .	87
3.14	ARMS vs Noise . . . . .	88
3.15	AUM vs Noise . . . . .	88
4.1	The flow chart for experimental verification . . . . .	109
4.2	Step edge focused at step 0 and defocused images at lens steps 10,40,70 . . . . .	110
4.3	Cubic polynomial focused at step 90 and defocused images at lens steps 10,40,70 . . . . .	110
4.4	Gray-level in edge image . . . . .	110
4.5	Gray-level in polynomial image . . . . .	110
4.6	Test images taken at step 50 . . . . .	113
5.1	Stonybrook Vision System (SVIS) . . . . .	116
5.2	Autofocusing Experiment . . . . .	119
5.3	Lens Step vs 1/Focusing Distance . . . . .	123
5.4	Blur Parameter vs Focused Lens Step . . . . .	123
5.5	Test Objects for Experiments . . . . .	132
5.6	Reconstructed Focused Image of a Prism Object . . . . .	134
5.7	Depth Map of the Prism Object . . . . .	134
5.8	Reconstructed Focused Image of a Cone Object . . . . .	134

5.9	Depth Map of the Cone Object . . . . .	134
5.10	An Inclined Plane without Projected Pattern . . . . .	135
5.11	Poor Depth Map of the Inclined Plane . . . . .	135
5.12	Reconstructed Focused Image of an Inclined Plane (with Projected Pattern) . . . . .	135
5.13	Good Depth Map of the Inclined Plane . . . . .	135

## List of Tables

3.1	Experimental results . . . . .	86
4.1	Step edge with noise std. 1.0 . . . . .	111
4.2	Step edge with noise std. 2.0 . . . . .	111
4.3	Step edge with noise std. 3.0 . . . . .	111
4.4	Cubic polynomial with noise std. 0.3 . . . . .	112
4.5	Cubic polynomial with noise std. 0.6 . . . . .	112
4.6	Cubic polynomial with noise std. 0.9 . . . . .	112
4.7	The object A with noise std. 1.6 . . . . .	113
4.8	The object B with noise std. 1.6 . . . . .	113
5.1	Autofocusing results. The correct mean was taken to be the mean of DFF when computing standard deviation for all three methods. . . . .	133



## Acknowledgements

I would like to express my sincere gratitude to my advisor Professor Muralidhara Subbarao, for his continuous guidance and precious advice during these years of my research. I am grateful to Professor Han-Sheng Tuan, Nam Phamdo and Vaclav Dolezal for serving on my thesis committee and for their valuable suggestions and comments.

Many thanks to all my friends who made my stay enjoyable at Stony Brook ever since I came. My special gratefulness would extend to my fellow students Mr. Yen-Fu Liu and Mr. Ta Yuan for their generous help with my research in the Computer Vision Lab. I also have to thank Ms. M. Krause for her assistance during my graduate studies.

This work was sponsored in part by the Olympus Optical Corporation and the National Science Foundation. This support is gratefully acknowledged.

Finally, my deepest appreciation and love go out to my parents, my wife Yu-Chuan and my son Geng-Ray. I thank them for their patient support and constant encouragement.

# Chapter 1

## Introduction

### 1.1 Motivation

The recovery of three-dimensional (3D) scene information from two-dimensional (2D) images is a fundamentally important problem for machine vision inspection and measurement tasks. The 3D scene information consists of *photometric information* (i.e. scene radiance along different directions of view) and *geometric information* (i.e. the distance of visible surfaces along different viewing direction). The 2D image formed by an optical lens contains both photometric and geometric information. The research on this problem in computer vision so far, has resulted in a variety of sensors and algorithms [24, 5] that can be broadly classify into two categories: active and passive. Active techniques such as point triangulation and laser radar produce relatively reliable depth maps, and have been applied to many industrial applications. However, when

the environment cannot be controlled, as in the case of distant objects in large-scale scenes, active methods prove impractical. As a consequence, passive techniques are desirable.

Passive sensing methods, such as stereo and structure from motion, rely on algorithms that establish local correspondences between two or more images. From the resulting disparity estimates or motion vectors, the depths of points in the scene are computed. The process of determining correspondence is widely acknowledged as being computationally expensive. In addition, the above techniques suffer from the occlusion or missing part problems. In fact, it is not possible to compute depths of scene points that are visible in only one of the images. Other passive techniques are based on image focus analysis. *Depth From Focus* (DFF) uses a sequence of images taken by changing the focus setting of the imaging optics in small steps. For each pixel, the focus setting that maximizes image contrast is determined. In turn, this can be used to compute the depth of the corresponding scene point and provide highly accurate measurement

In contrast, *Depth From Defocus* (DFD) uses only two images with different optical settings. The objects need not be focused. The relative defocus in the two images can be used to determine 3D depth. The focus level in the two images can be varied by changing focus setting of the lens, by moving the image detector with respect to the lens, or by changing the aperture size. Depth from defocus is not confronted with the above mentioned missing part and correspondence problems. This

makes it an attractive method for depth measurement.

The goal of the research presented here is to investigate and develop robust techniques for determining the distance of objects in a scene and reconstructing the focused images based on image focus and defocus analysis. In machine vision applications, a typical CCD video camera system used in passive techniques involve several sources of signal noise such as optical noise or electronic noise. As a result of signal noise, no two images, regardless of how carefully the image acquisition environment is controlled, will have identical grey level values for each pixel position. Thus, the errors in depth measurement are inevitable. The performance of DFF and DFD due to noise effects has been evaluated through experimental observation. The methods for evaluation usually need large samples and many trials. In this approach, the evaluation becomes subjective and the experiments are time consuming. The noise sensitivity analysis of DFF and DFD is an important problem and is the topic of this research. The analysis presented in this thesis provides efficient and accurate solution for uncertainty in depth measurement. Another important goal is to implement a fast and accurate autofocusing and 3D shape recovery algorithm. Since DFF and DFD have some advantages and disadvantages, an integrated method with DFF and DFD is proposed. The theories developed here are verified by experiments and computer simulations.

## 1.2 Literature Review

Using image focus and defocus analysis to obtain depth information from the observed scene has been addressed by many researchers. Here we briefly summarize the previous work in this area.

### 1.2.1 Depth from Focus (DFF)

The basis of DFF methods is the focus measure which measures the focus quality. One goal of this research is how to evaluate and develop effective focus criteria. In 1968, Horn proposed an automatic focusing algorithm by using the Fourier transform and analyzing the frequency spectrum of the image [21]. In principle, this was a good approach but not practical, because the computed information was computationally expensive. Another approach which was easy to calculate, explored by Tenenbaum in 1970 [76], examined the effect of defocusing an edge image. He developed the gradient magnitude as a focus measure to be maximized.

After Tenenbaum, several researchers proposed different focus measure functions based on spatial domain analysis. In 1974, Muller and Buffington [37] implemented *squared gradient*, *Laplacian* and *signal power* as focus measures for automatic focusing. Javis, two years later [23] proposed three more focus measures: *histogram entropy*, *grey level variance* and *sum-modulus-difference*. He found that all three focus

measures were promising in focus optimization. Of the three, the *sum-modulus-difference* formulated from image gradient was the least ambiguous for the image tested.

Ligthart and Groen in 1982 [35] performed a concise comparison of most of focus measure functions. They concluded that the *squared gradient magnitude* and *grey level variance* methods gave the best results. One year later, Schlag *et al.* [52] evaluated five automatic focusing algorithms, and found the *magnitude of the gradient* to be best.

In the eighties, Krotkov [30, 29] compared the performance of different focus measures with several images. He found that *magnitude of the gradient* with the kernel of the Sobel operator was superior to the others in monotonicity about the peak and robustness in the presence of noise.

Darrell and Wohn [9] developed a new depth from focus method that obtained an image of a scene at different lens positions. Then using Laplacian and Gaussian pyramids to isolate frequency ranges, they were able to interpolate depth to various points in the scene.

More recently, Nayar [40] proposed the *sum-modified-Laplacian* focus function to measure the relative degree of focus between images. He also developed a Gaussian model that describes focus measure variations due to defocusing. This model is used to interpolate between a finite number of focus measures to obtain accurate depth estimates.

Subbarao and Choi [60, 61] derived a class of focus measures based on band-pass filtering analysis. They assumed that the side lobes in the optical transfer function of the camera system can be neglected and the

focus measure has global maximum for the best focused images. The derived focus measures were evaluated experimentally, and concluded that the *energy of low pass filtered image gradient* had the best performance overall. In addition, they proposed a new concept named *Focused Image Surface* (FIS). The shape of an object was determined by searching for the maximum of focus measure over the FIS rather than the planar image detector of the camera. This results in more accurate shape recovery than the traditional methods.

### 1.2.2 Depth from Defocus (DFD)

If the level of defocus could be measured throughout the images, then a few images would be sufficient to determine the distance of all objects in the scene. This forms the basis of DFD methods. The degree of defocus can be characterized by the spread parameter of the *Point Spread Function* (PSF) which corresponds to the blur parameter of the image. How these spread parameters can be solved for and mapped to depth estimate are the central problem in DFD.

Pentland [45] in 1982 introduced the work by assuming the PSF as a Gaussian model and the object as a sharp step edge. He used the Laplacian to measure the width of the blurring edge. The depth recovery was then computed from the spread parameter of the Gaussian. In 1988, Subbarao and Gurumoorthy [62] removed the assumption about the PSF so that any rotationally symmetric function could be used. Based on the

image of step edge, they presented the *Line Spread Function* (LSF) to characterize the camera system, and showed that the spread parameter of LSF is related to the distance of the object.

In order to apply to arbitrary scenes, Pentland's technique [46] was extended in 1987. He used two images of the scene, one taken with a pinhole aperture and the other with a wide aperture. He estimated the signal power in one or more frequency bandwidths in each image, then two values were compared to a lookup table to produce an estimate of range. The main disadvantage of his algorithm was to adopt the pin-hole camera which is sensitive to signal noises.

Assuming that the point spread function was a Gaussian, Subbarao derived the equations which related changes in each of these parameters to the Gaussian spread parameter and hence the distance to the scene. These parameters were allowed for both *infinitesimal* change [54] and *finite* change [55]. In a follow-up paper [56] this was also extended to other point spread functions.

In the early nineties, Ens and Lawrence [10] provided a depth from defocus method based on a matrix regularization method. They estimated a convolution matrix, which is convolved with one of the images to produce the other image. The matrix corresponds to the relative blur between the two images. Once the matrix is computed, it can be mapped to depth estimates. This method produced accurate depth maps, but the iterative nature of the convolution matrix made it computationally expensive.



Subbarao and Surya [66, 75] proposed the *S-Transform Method* (STM) and applied it to depth from defocus. They modeled the image as a third-order polynomial in small image region in the spatial domain, and generalized the spread parameters to be the square root of the second central moment of the PSF. A simple and elegant expression was derived by them (see Chapter 4 for details). This method produces reasonable depth estimates in 48x48 image regions. The overall RMS error of 2.3 percent in lens position was reported. However, it does not yield depth maps with high spatial resolution that are needed when depth variations in the scene are significant.

In the same period, Subbarao and Wei performed the theoretical and experimental work for a more general solution in the Fourier domain [57, 56, 73]. They removed Pentland's constraint of one image being formed by a pin-hole camera and gave a closed form solution assuming a Gaussian PSF. The spread parameter is measured from the ratio of the one-dimensional Fourier coefficients of the two images. They implemented their method on an actual camera system and reported a RMS error of 3.5 percent in lens position, over a large number of experiments. A relatively large window size of about 128x128 pixels was used for computation.

Since the frequency characteristics of general scenes are unpredictable, it is difficult to obtain accurate depth estimates. Recently, Xiong and Shafer [87] implemented the *moment filter* to compensate the response of the PSF changing with image frequency and compute depth

in a least-squares sense using all dominant frequencies of the image. Nayar *et al.* [41] developed the focus range sensor using active illumination pattern to constrain the dominant frequencies in the scene. Their methods maximized the accuracy and robustness of depth from defocus.

### 1.2.3 Combining with Stereo Ranging

Due to the weakness of stereo ranging, several researchers have already begun to combine it with depth from focus or defocus for a robust vision systems. Krotkov and Kories have integrated their depth from focus system with a stereo ranging system [29, 31]. Abbott and Ahuja [1] have sought to dynamically integrate focus, camera vergence and stereo. Choi and Wei [8, 81] also proposed a simple method to integrate depth from focus or defocus with stereo method in their dissertations, respectively. Recently, Subbarao *et al.* [71] proposed a new algorithm which combined depth from focus, defocus and stereo in a monocular camera with stage motion system for 3D shape recovery and focused image reconstruction.

## 1.3 Dissertation Overview

This dissertation is based on our research works [67, 68, 69, 70, 71, 72, 78] and is organized as follows.

Chapter 2 discusses the shape from focus methods for microscope objects. We first review the image formation process and describe defocused images as processed versions of focused images. The concepts of focused and defocused images give us the background for DFF and DFD techniques. Then the theory of focusing is summarized and the model of focus measure is introduced. Some of the sound focus measures which has been proposed and commonly used in the past for focusing analysis are also surveyed. In particular, we do not consider focus measures based on summing the absolute values of derivatives of the image. Such filters can be proved to be unsound through our counter examples. In addition, we argue that energy of the Laplacian of the image is a good focus measure to use, because it has some important desirable characteristics based on a spatial domain analysis. We address those characteristics in detail and recommend it for use in most scenes. If the scene information contains low contrast, the depth measurement based on DFF becomes unreliable. A thresholding technique is then developed to eliminate image regions susceptible to noise effects. Based on the shape from focus algorithms, we implemented a fully automated shape from focus system for recovering 3D shape and reconstructing focused image of microscopic objects in our computer vision laboratory. The description of this system is given and the experiments with several microscopic objects are presented. The experimental results demonstrate the value of the application of image focusing theory in microscopy going beyond the conventional sample inspection.

Evaluating the performance of focus measures is an important issue in DFF research. In chapter 3, a method is described for selecting the optimal focus measure with respect to grey-level noise from a given set of focus measures in passive autofocus and DFF applications. The method is based on two new metrics that have been defined for estimating the noise-sensitivity of different focus measures. The first metric – the *Autofocusing Uncertainty Measure* (AUM)– was developed in the early stage of this research. It is introduced here in order to illustrate some underlying concepts on the relation between the grey-level noise and the resulting error in lens position for autofocus. The second metric– *Autofocusing Root-Mean-Square Error* (ARMS error)– is an improved metric closely related to AUM. AUM and ARMS error metrics are based on a theoretical noise sensitivity analysis of focus measures, and they are related by a monotonic expression. We validate our theory on AUM and ARMS error through actual and simulation experiments. It is shown that the theoretically estimated and experimentally measured values of the standard deviation of a set of focus measures are in agreement. In comparison, all known prior work on comparing the noise sensitivity of focus measures have been a combination of subjective judgement and experimental observations. Our results show that for a given camera the optimally accurate focus measure may change from one object to the other depending on their focused images. Therefore selecting the optimal focus measure from a given set involves computing all focus measures in the set. However, if computation needs to be minimized, as mentioned

previously, the Laplacian of the image is superior to the others and is suggested for use in practical applications.

Chapter 4 provides a theoretical treatment of the noise sensitivity analysis of depth from defocus using the Spatial-Domain Convolution/Deconvolution Transform Method (STM). STM is a useful technique for 3D vision. Unlike DFD based on frequency domain approach, STM does not require information about the actual form of the camera's PSF and involves simple local operations to measure defocus level in the spatial domain on only two images recorded with different camera parameters (e.g. by changing lens position or changing aperture diameter). The theory we derived fills an important gap in the current research literature wherein the noise sensitivity analysis of STM is limited to experimental observations. Given the image and noise characteristics, an expression for the Root Mean Square (RMS) error is derived in lens position for focusing an object. This RMS error is useful in estimating the uncertainty in depth obtained by STM. The results of computer simulation experiments for different noise levels and real camera experiments for one noise level are presented. The experiments validate the theoretical results.

The integration of DFF and DFD with experiments on the prototype vision system is provided in Chapter 5. DFF algorithms involve the search for a set of camera parameters corresponding to the sharpest image of the scene. Thus, a large number of images recorded with different camera parameters setting are required in DFF methods, which

limits the speed in computer vision application. In contrast, DFD methods need only 2 or 3 blurred images with different degrees of defocus for depth estimate, which can be processed faster than DFF methods. However, accuracy-wise, DFF methods are better than DFD methods. Since DFF and DFD methods have some pros and cons, autofocusing and 3D shape recovery can be faster and more accurate by integrating DFF and DFD methods. We combine DFD methods using STM with DFF methods on a vision system named SVIS. SVIS is a new camera system built in our computer vision lab. It has no lens data available. Thus, a robust calibration procedure has been developed for SVIS. We provide the implementation details and experimental results.

Finally, a summary, conclusions and future research are presented in Chapter 6.

## **Chapter 2**

# **Shape from Focus and its Application to 3D Microscopy**

## **2.1 Introduction**

Several techniques for three dimensional (3D) shape recovery of microscopic objects have been developed in the past such as interferometry and confocal microscopy. In this chapter, we develop the shape from focus (SFF) technique that uses image focus analysis to recover 3D shape of microscopic objects.

The SFF needs to record a large sequence of image frames of a 3D scene with different camera parameters (e.g. focal length or/and lens to image detector distance). In each image frame, different objects in the scene will be blurred by different degrees depending on their distance from the camera lens. Each object will be in best focus in only one image frame in the image sequence. The entire image sequence is processed by using the focus measure filter to find the best focused image of each

object in the 3D scene. The distance of each object in the scene is then found from the camera parameters that correspond to the image frame that contains the best focused image of the object.

The SFF approach can be applied in a microscope equipped with a CCD camera for the object in micrometer range. In contrast to camera system for the macroscopic object, a sequence of image frames of the microscopic object are acquired by moving the object itself (translational stage) rather than by changing camera parameter settings. All camera parameters are fixed during the image acquisition and processing. Due to different image acquisition approach, the distance of each object in the scene is measured by a relative value rather than a absolute value.

Next, we review the image formation process and show that a defocused imaging system plays the role of a low-pass filter. The theory of focus measure is also reviewed. The focus measures considered in this chapter are only those which have been proved to be sound [60] based on the effect of the OTF main lobe. The focus measures based on summing the absolute values of the derivatives of the image [76, 30, 39] are unsound. We provide the proof through counter examples. The Laplacian operator is a good focus measure and is used in our experiments. Some important properties about the Laplacian operator are provided. We conclude this chapter with several results produced by the automated SFF microscope system.



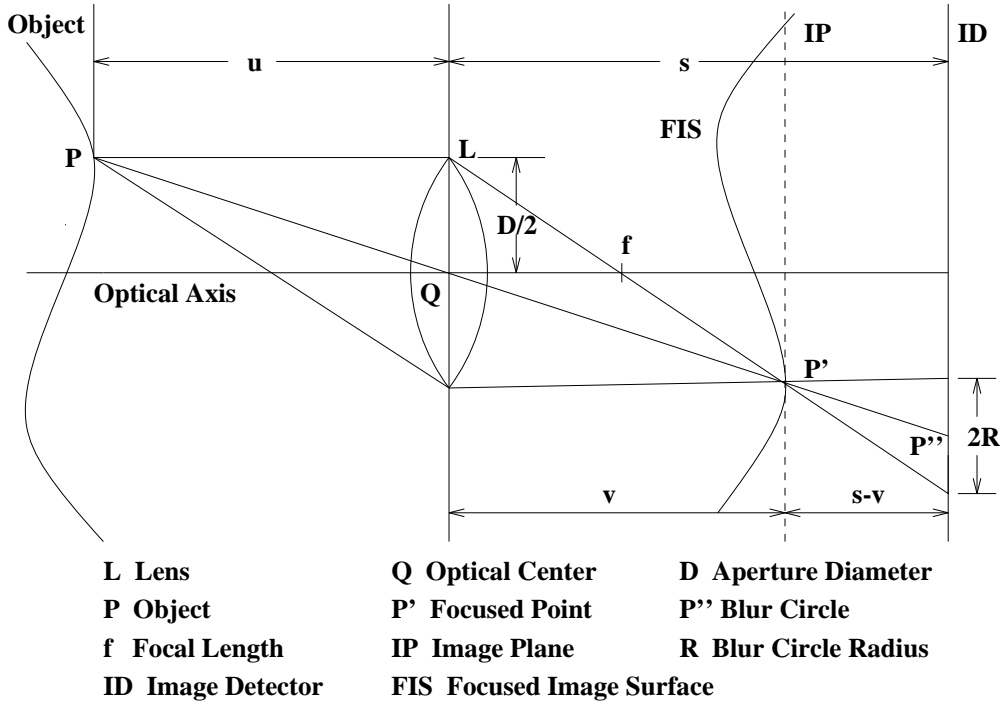


Figure 2.1: Image Formation in a Convex Lens

## 2.2 Focused and Defocused Images

The image of a scene sensed by an optical system consists of the *geometric* and *photometric* information about the scene. The distance and radiance of objects are associated with geometric and photometric information, respectively [21]. According to paraxial-geometric optics, the basic image formation process is shown in Fig. 2.1. For an aberration-free convex lens, light beams that are radiated by the object point  $P$  and intercepted by the lens are refracted by the lens to converge at a point  $P'$  on the image plane. The distance of the point  $P$  can be determined by the well-known lens formula

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \quad (2.1)$$

where  $f$  is the focal length,  $u$  is the distance between the object plane and the lens plane, and  $v$  is the distance between the image plane and the

lens plane. If the image detector (CCD array) coincides with the image plane, a clear or *focused image*  $f(x, y)$  is sensed by the image detector. The radiance of the point P is proportional to the irradiance at its focused image. In a 3D scene, each point on the object surface corresponding to a single point on the image surface is according to the above relationship. The image surface is known as *Focused Image Surface* (FIS).

If, however, the image detector does not coincide with the image plane and is displaced from it by a distance  $s - v$ , the light received from the object point P by the lens is distributed over a circular patch on the image detector. Thus, the point object P is blurred or *defocused* on the image detector (see Fig. 2.1) and it is imaged as a blur circle P" of radius  $R$ . This blurred image  $h(x, y)$  is the Point Spread Function (PSF) of the camera system. The PSF can be modeled using physical or geometric optics [21, 46, 56, 57]. In a small image region if the imaged object surface is (approximately) a plane normal to the optical axis, then the PSF is the same for all points on the plane. Then the *defocused image*  $g(x, y)$  in the small image region on the image detector is equal to the convolution of the focused image  $f(x, y)$  and the PSF  $h(x, y)$ :

$$g(x, y) = h(x, y) * f(x, y) \quad (2.2)$$

where  $*$  denotes the convolution operator.

From Fig. 2.1, it is seen that a defocused image of the object can be obtained in several ways: by changing the distance  $s$  between the lens and the image detector, by changing the distance  $u$  between the lens and the objects, by changing the aperture diameter  $D$ , or/and by varying the

focal length  $f$ . Note that defocusing is observed for both positive and negative image detector displacements.

From the above discussion, given the irradiance and the position of the focused image of a point, its geometric and photometric information in the scene can be uniquely determined. In traditional shape from focus methods (e.g. [30, 39, 60, 67]), the focused image is obtained by searching a sequence of defocused images which are obtained by continuously varying camera parameter settings (either  $s$  or/and  $f$ ). Thereafter, a criterion function called focus measure is computed for each image in the sequence to determine whether the image is focused or not. A focus measure is computed at each pixel (i.e. each direction of view) in a small (about  $15 \times 15$ ) image neighborhood around the pixel. At each pixel, that image frame among the image sequence which has the maximum focus measure is found by a optimized search technique such as a binary search, Fibonacci search [25] or combining search (see later Chapter 4). The grey level (which is proportional to image irradiance) of the pixel in the image frame thus found gives the grey level of the focused image for that pixel. The values of  $s$  and  $f$  for this image frame are used to compute the distance of the object point corresponding to the pixel. Since there is an one to one correspondence between FIS and the object surface, the search procedure can be repeated for every pixel in the image frame. Then the 3D shape and the focused image of the object can be obtained. Therefore, DFF methods involve a search for the values of  $s, u$  or/and  $f$  that results in a maximum focus measure and these methods require the

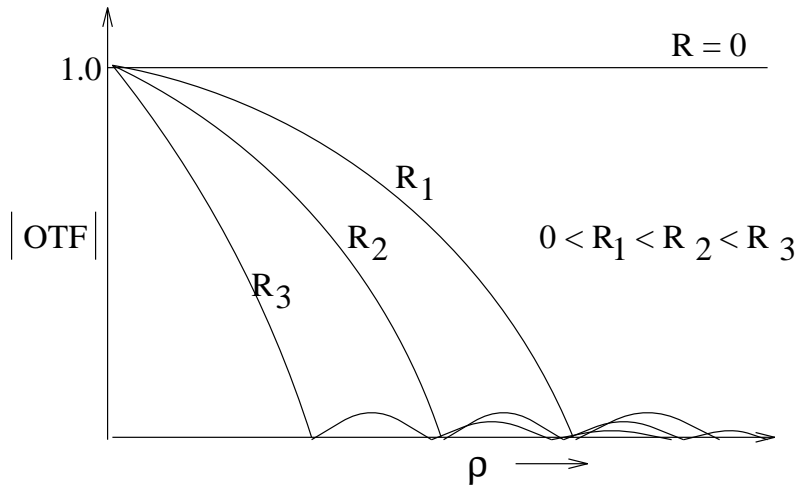


Figure 2.2: The OTF of a blurred camera attenuates high frequencies. The attenuation effect increases with increasing value of the blur circle radius  $R$ .

acquisition and processing of a large number of images.

## 2.3 Model of Focus Measures

As mentioned previously, the position of the focused image for each pixel in the image frame is determined by searching the maximum focus measure among a sequence of images. The accuracy of autofocus and 3D shape measurement using the image focus analysis technique depends on the particular focus measure that is used. Here we summarize some relevant results about theory and model of focus measure based on geometric optics. A detailed discussion of this topic can be found in several papers including [60, 30, 23, 76].

Let us analyze the defocusing process in the frequency domain  $(\omega, \nu)$ . If  $G(\omega, \nu)$ ,  $F(\omega, \nu)$ , and  $H(\omega, \nu)$  are the Fourier transforms of  $g(x, y)$ ,  $f(x, y)$ , and  $h(x, y)$ , respectively, we can express Eq.(2.2) as:

$$G(\omega, \nu) = H(\omega, \nu) \cdot F(\omega, \nu) \quad (2.3)$$

The OTF  $H(\omega, \nu)$  corresponding to  $h(x, y)$  is circularly symmetric and its cross section has a form similar to the sinc function (see Fig. 2.2). For a focused image, the first zero crossing of the OTF is very far from the zero spatial frequency and this zero crossing moves closer to the zero spatial frequency as the blur increases. Therefore the effect of blurring is to attenuate higher frequencies. The attenuation increases monotonically within the main lobe of the sinc-like OTF. The effect of the main lobe on the computed focus measures usually dominates that of the side lobes. This model of image sensing is summarized in Fig. 2.3. A more detailed model of image sensing is described in [64].

A general focus measure is modeled as follows (see Fig. 2.4). First the image for which the focus measure needs to be computed is normalized for brightness by dividing the image by its mean brightness. Then it is convolved with a focus measure filter (FMF). Then the energy (sum of squared values) of the filtered image is computed. This energy is the focus measure. Most FMFs correspond to filters that emphasize (or amplify) high frequencies. This seems appropriate since blurring has the opposite effect, i.e. high frequency attenuation.

The focus measures modeled here cover most of the focus measures that have been used by many researchers so far [60] except those based

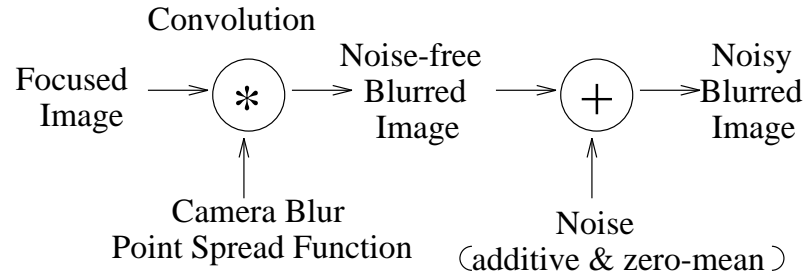


Figure 2.3: Model of Image Sensing

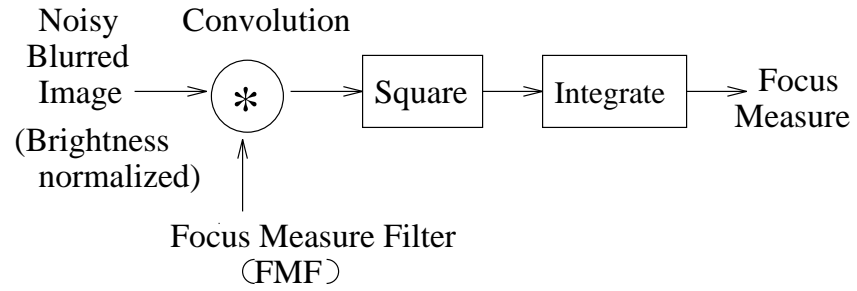


Figure 2.4: Model of Focus Measure. The FMF for gradient along the  $x$  and  $y$  axis are  $\frac{\partial}{\partial x}$  and  $\frac{\partial}{\partial y}$  respectively and the FMF for the Laplacian is  $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ .

on sum of absolute values of image derivatives [76, 30, 39]. We will see in the later section that such filters are proved to be unsound through counter examples.

### 2.3.1 Review of Focus Measures

In this section we list some general and sound focus measures. The symbol  $\gamma$  is used to represent focus measures.

#### 1. Gray Level Variance

This focus measure is computed as

$$\gamma_1 = \frac{1}{A} \int \int (g(x, y) - \mu)^2 dx dy \quad (2.4)$$

where  $\mu$  is the mean value of the gray level in the image region  $A$ . This focus measure is the grey level variance of the image [23]. It is linearly related to the integral of the power spectrum of the intensity distribution.

#### 2. Gradient Magnitude Squared

The integral of gradient magnitude squared is defined as [52, 76, 7]

$$\gamma_2 = \int \int \left[ \left( \frac{\partial g(x, y)}{\partial x} \right)^2 + \left( \frac{\partial g(x, y)}{\partial y} \right)^2 \right] dx dy \quad (2.5)$$

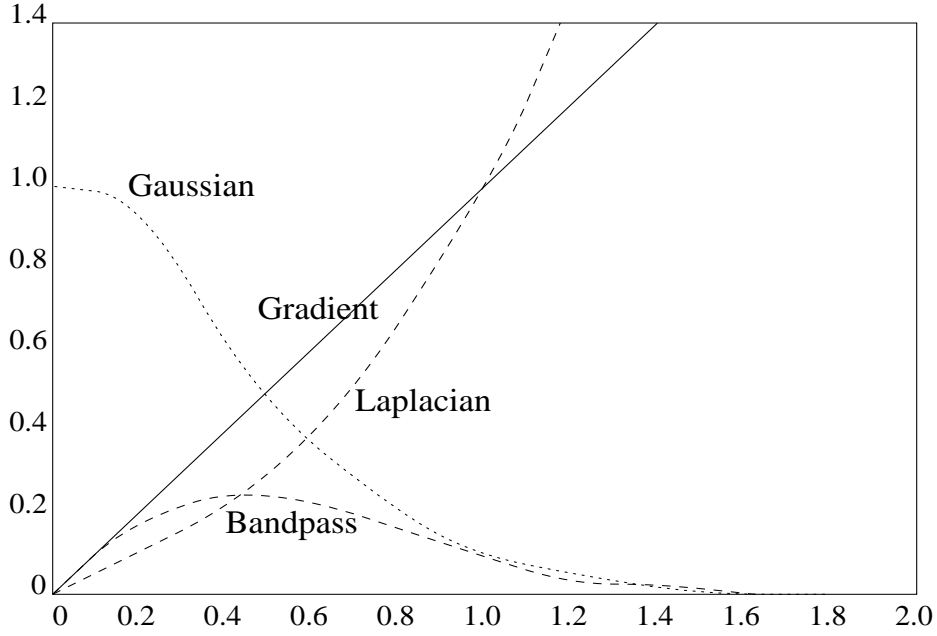


Figure 2.5: Frequency response of various filters

### 3. Laplacian Squared

The focus measure based on Laplacian filter [9, 37] is defined as

$$\gamma_3 = \iint \left( \frac{\partial^2 g(x, y)}{\partial x^2} + \frac{\partial^2 g(x, y)}{\partial y^2} \right)^2 dx dy \quad (2.6)$$

The Laplacian is a high pass filter which computes the second order derivative of the image.

### 4. Bandpass Filtering

Bandpass filtering as a focus measure is obtained by smoothing the images first with a low pass filter such as a Gaussian and then applying one of the above high pass filters [60]. One example is

$$\gamma_4 = \iint \left[ \frac{\partial^2}{\partial x^2} (h_b(x, y) * g(x, y)) + \frac{\partial^2}{\partial y^2} (h_b(x, y) * g(x, y)) \right]^2 dx dy \quad (2.7)$$

where  $h_b$  is a Gaussian low-pass filter. Fig. 2.5 shows those focus mea-



asures corresponding to the filtering characteristics exhibiting in the frequency spectrum.

### 2.3.2 Unsound Focus Measures

Some researchers in the past [76, 23, 30, 39] have used focus measures based on integral of the absolute value of image derivatives such as *sum modulus difference* and *sum modified laplacian*. The soundness of these focus measures has not been investigated in the past. In this section this class of focus measures are shown to be unsound by constructing a class of images for which these focus measures do not exhibit a maximum for the focused image.

First we consider the case of a one-dimensional signal and then indicate how it can be extended to two-dimensional images.

A PSF  $h(x)$  has the following properties that

$$h(x) \geq 0 \text{ for all } x \quad (2.8)$$

and

$$\int_{-\infty}^{\infty} h(x) dx = 1 \quad (2.9)$$

Now consider an arbitrary function  $p(x)$  with properties  $p(x) \geq 0$  for all  $x$  and  $\int_{-\infty}^{\infty} p(x) dx = 1$ . Let  $\gamma_i$  be a focus measure defined as the integral of the absolute value of the  $i$ -th derivative of a blurred signal  $g(x)$ , i.e.

$$\gamma_i = \int_{-\infty}^{\infty} \left| \frac{d^i g(x)}{dx^i} \right| dx \quad (2.10)$$

Let  $g(x) = h(x) * f(x)$  where  $f(x)$  is the focused signal corresponding to  $g(x)$ . Therefore

$$\begin{aligned}\gamma_i &= \int_{-\infty}^{\infty} \left| \frac{d^i}{dx^i} (h(x) * f(x)) \right| dx \\ &= \int_{-\infty}^{\infty} \left| h(x) * \frac{d^i f(x)}{dx^i} \right| dx\end{aligned}\quad (2.11)$$

Let  $p(x)$  be the solution of the differential equation

$$\frac{d^i f(x)}{dx^i} = A p(x) \quad (2.12)$$

where  $A$  is any positive constant, i.e.  $A > 0$ . We have

$$\begin{aligned}\gamma_i &= A \int_{-\infty}^{\infty} |h(x) * p(x)| dx \\ &= A\end{aligned}\quad (2.13)$$

The last step of the derivation above can be justified by the following well-known result in the probability theory. If  $h(x)$  and  $p(x)$  are the probability density functions of two independent random variables  $x_h$  and  $x_p$  respectively, then the probability density function of their sum, i.e.  $x_h + x_p$ , is  $h(x) * p(x)$ ,  $h(x) * p(x) > 0$  for all  $x$ , and

$$\int_{-\infty}^{\infty} h(x) * p(x) dx = 1 \quad (2.14)$$

Therefore, for a focused signal  $f(x)$  given by the solution of Eq. (2.12), the focus measure  $\gamma_i$  remains the same no matter what  $h(x)$  is. Therefore  $\gamma_i$  is not sound in that it does not have a maximum when  $h(x) = \delta(x)$  (*dirac delta function*). Two particular examples are  $p(x) = \delta(x)$  for  $\gamma_1$  and  $\gamma_2$  where  $f(x)$  will be  $A \cdot u(x)$  and  $A \cdot \text{ramp}(x)$  respectively (see Fig. 2.6). Note that  $u(x)$  is the unit step function and  $\text{ramp}(x)$  is the unit ramp

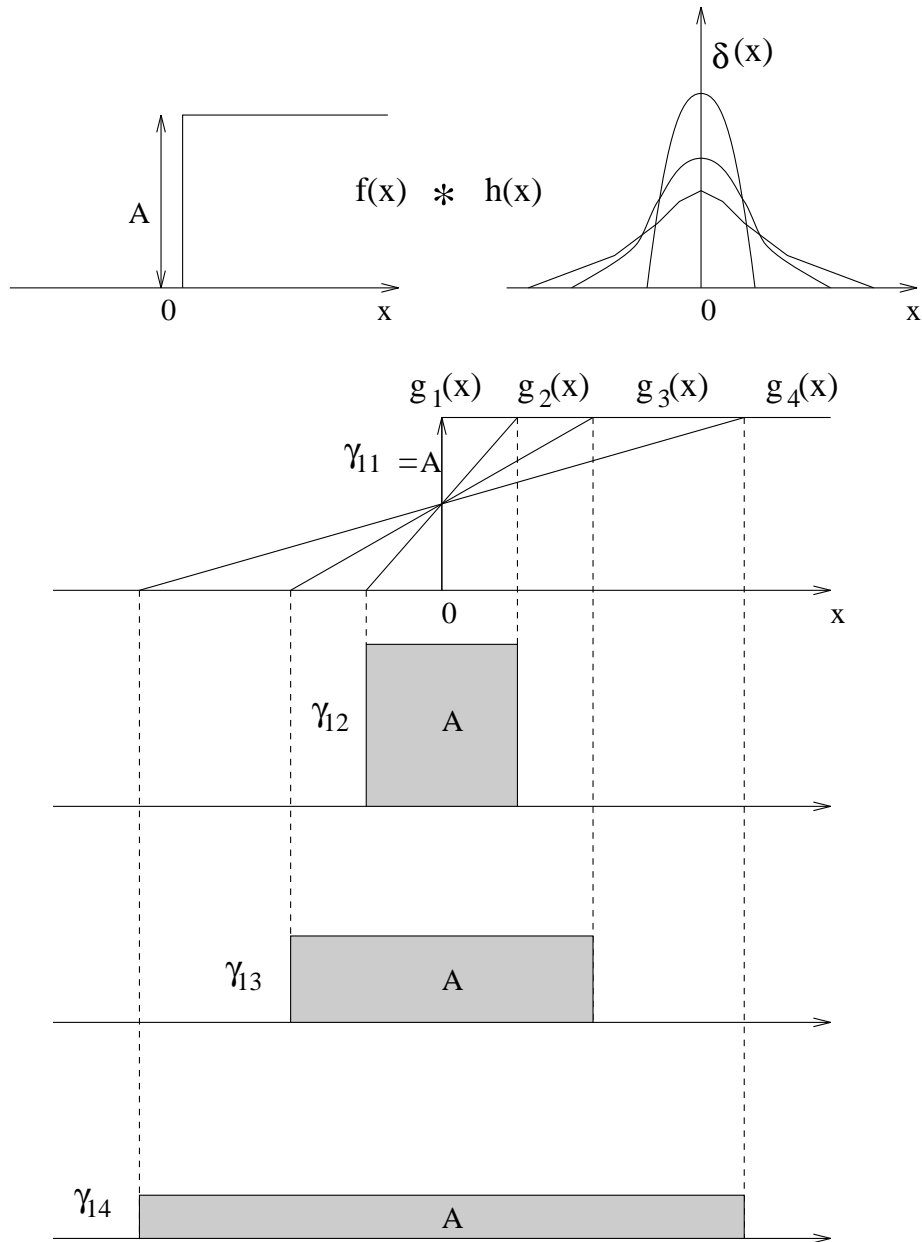


Figure 2.6: A simple 1D example. Based on  $\int \left| \frac{dg(x)}{dx} \right| dx$ , it is seen that  $\gamma_{11} = \gamma_{12} = \gamma_{13} = \gamma_{14}$  regardless how the PSF is changed. If focus measure is computed based on  $\int \left| \frac{dg(x)}{dx} \right|^2 dx$ , according to the Schwarz inequality we can obtain  $\gamma_{11} > \gamma_{12} > \gamma_{13} > \gamma_{14}$ .

function defined as

$$ramp(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.15)$$

The one-dimensional case above can be easily extended to two-dimensional images by assuming that the image changes along only one of the dimensions and is a constant along the other dimension. A more general counter example can be constructed as follows: Let

$$g(x, y) = 0 \quad \text{if } |x| > B \text{ or } |y| > B \quad (2.16)$$

$$\gamma_i = \gamma_{xi} + \gamma_{yi} \quad (2.17)$$

$$\gamma_{xi} = \frac{1}{(2B)^2} \int_{-B}^B \int_{-B}^B \left| \frac{\partial^i g(x, y)}{\partial x^i} \right| dx dy \quad (2.18)$$

$$\gamma_{yi} = \frac{1}{(2B)^2} \int_{-B}^B \int_{-B}^B \left| \frac{\partial^i g(x, y)}{\partial y^i} \right| dx dy \quad (2.19)$$

Let the PSF be separable

$$h(x, y) = h_x(x) \cdot h_y(y) \quad (2.20)$$

and

$$\frac{\partial^i f(x, y)}{\partial x^i} = A_x p_x(x) \quad (2.21)$$

$$\frac{\partial^i f(x, y)}{\partial y^i} = A_y p_y(y) \quad (2.22)$$

where  $p_x(x)$  and  $p_y(y)$  have the properties of a probability density function (i.e. they are always positive and they integrate to unity). Therefore

$$\begin{aligned}
\gamma_{xi} &= \frac{1}{(2B)^2} \int_{-B}^B \int_{-B}^B \left| \frac{\partial^i}{\partial x^i} ((h_x(x) \cdot h_y(y)) * f(x, y)) \right| dx dy \\
&= \frac{1}{(2B)^2} \int_{-B}^B \int_{-B}^B |(h_y(y) \cdot h_x(x)) * (A_x \cdot p_x(x))| dx dy \\
&= \frac{A_x}{(2B)^2} \int_{-B}^B \int_{-B}^B \left| \left\{ \int_{-B}^B h_y(y) dy \right\} \{h_x(x) * p_x(x)\} \right| dx dy \\
&= A_x
\end{aligned} \tag{2.23}$$

Similarly we can obtain  $\gamma_{yi} = A_y$ . Hence the proof.

### 2.3.3 The Laplacian Filter

The Laplacian FMF has some desirable properties which makes it suitable for practical applications. It is sound, simple, rotationally symmetric, and is a linear filter (unlike the gradient magnitude). In addition it has two additional properties under the geometric optics model of image formation not shared by other filters based on image derivatives. The two properties correspond to the two cases of a noise free image: (i) when the image does not change due to blurring as is the case when the grey level variation is linear, the value of the focus measure is zero, and (ii) when the image does change as is the case when the grey level variation is a polynomial of second or higher degree, the focus measure is non-zero (see Figs. 2.7 and 2.8). When the grey level variation in a noise free image is linear, then the image does not change due to blurring

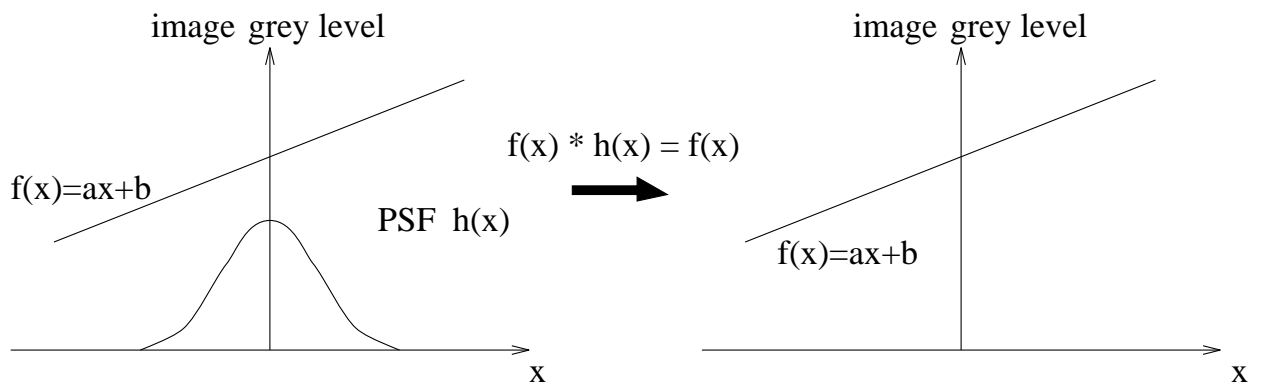


Figure 2.7: Blurring does not change the image where the function is locally linear.

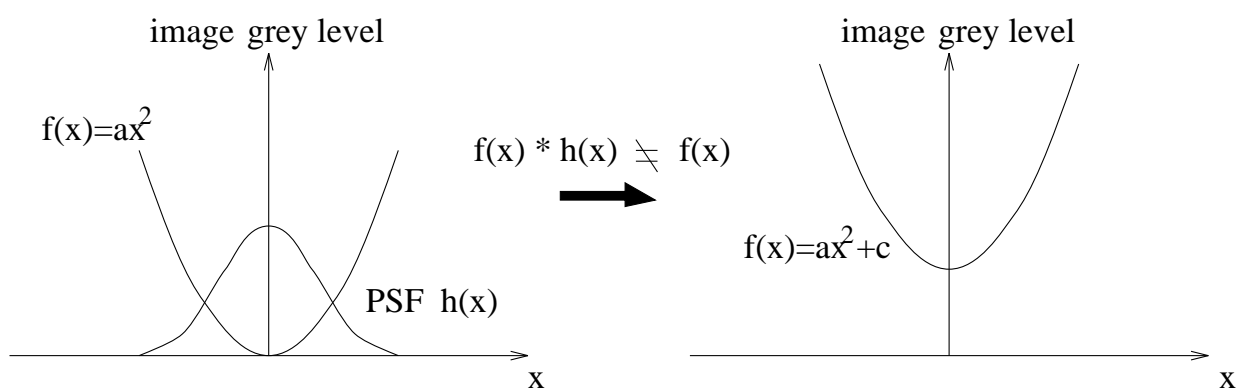


Figure 2.8: Blurring changes the image where the function is locally quadratic.

under geometric optics model of image formation . The two properties of the Laplacian based focus measure are proved as following:

First it is shown that: (i) the image does not change due to blurring when the grey level variation is linear, and in this case, (ii) the focus measures based on grey level variance and energy of the image gradient are in general non-zero, but the focus measure based on the Laplacian is zero.

It is shown in [66] that when an image  $f(x, y)$  given by a third order polynomial is blurred to obtain  $g(x, y)$  by a rotationally symmetric PSF  $h(x, y)$ , we have

$$g(x, y) = f(x, y) + \frac{h_{2,0}}{2} \nabla^2 f(x, y) \quad (2.24)$$

where  $h_{2,0}$  is the second moment of  $h(x, y)$  defined by

$$h_{2,0} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 h(x, y) dx dy \quad (2.25)$$

Eq. (2.24) can be used to obtain an expression for  $g(x, y)$  when  $f(x, y)$  is linear by setting the coefficients of second and third order terms in  $x$  and  $y$  to be zero. Therefore, if

$$f(x, y) = a_0 + a_1 x + a_2 y \quad (2.26)$$

we have

$$g(x, y) = f(x, y) \quad (2.27)$$

because

$$\nabla^2 f(x, y) = 0 \quad (2.28)$$

Eq. (2.27) proves that the image does not change due to blurring when the grey level variation is linear (Eq. (2.26)). The brightness normalized

image  $g_n(x, y)$  is obtained by dividing  $g(x, y)$  by its mean in an interval  $-A \leq x \leq A, -A \leq y \leq A$ . We have

$$g_n(x, y) = \frac{a_0 + a_1x + a_2y}{a_0} \quad (2.29)$$

It can be easily verified that, in general, the focus measures of  $g_n(x, y)$  computed in the interval  $-A \leq x \leq A, -A \leq y \leq A$ , is non-zero for image grey level variance, and energy of image gradient along  $x$  and  $y$  directions, but the Laplacian based focus measure is zero since

$$\nabla^2 g_n(x, y) = 0 \quad (2.30)$$

Next it is shown that: (i) when the image is a second order polynomial (quadratic), it changes due to blurring, (ii) the Laplacian based focus measure of the image is non-zero, and (iii) focus measures based on third and higher order derivatives are zero. Let

$$f(x, y) = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 \quad (2.31)$$

Setting the coefficients of third order terms in  $x$  and  $y$  to be zero in Eq. (2.24), we obtain

$$g(x, y) = a_0 + (a_3 + a_5) h_{2,0} + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 \quad (2.32)$$

After normalizing the image by dividing it by its mean we have

$$g_n(x, y) = \frac{g(x, y)}{a_0 + (a_3 + a_5) h_{2,0} + \frac{A^2}{3} (a_3 + a_5)} \quad (2.33)$$

The Laplacian of the normalized image can be expressed as

$$\nabla^2 g_n(x, y) = \frac{2(a_3 + a_5)}{a_0 + (a_3 + a_5) \frac{A^2}{3} \left(1 + \frac{3h_{2,0}}{A^2}\right)} \quad (2.34)$$



Eq. (2.33) proves that the image in general changes due to blurring. Eq. (2.34) shows that the Laplacian based focus measure is in general non-zero. In Eq. (2.33), we see that third and higher order derivatives of  $g_n(x, y)$  are all zero. This proves that focus measures based on third and higher order derivatives are zero, even though the image has changed due to blurring. If  $g(x, y)$  is not normalized for brightness, then  $f(x, y)$  has to be at least a 4-th order polynomial in order for the Laplacian focus measure to change for  $g(x, y)$  as compared with  $f(x, y)$ .

Therefore the first property is not shared by the grey-level variance and image gradient focus measures corresponding to zeroth and first order derivatives respectively. FMFs based on third and higher order derivatives do not share the second property because when the grey level variation is only quadratic, the corresponding focus measures are zero even though the images change due to blurring. This is because the third and higher order derivatives of a quadratically varying grey level image is zero everywhere.

The first property makes it easy to test the hypothesis whether a given image region has the necessary “contrast” so that autofocusing is possible. The confidence level of the hypothesis that the “contrast signal” is indeed present in a given image region and therefore autofocusing is possible can be estimated from the quantity:

$$\beta = \frac{\gamma - \gamma_\eta}{\text{std}(\gamma_{noise})} \quad (2.35)$$

where  $\gamma$  is the computed focus measure,  $\gamma_\eta$  is the expected value of the focus measure if only noise was present, and  $\text{std}(\gamma_{noise})$  is the standard

deviation of the focus measure if only noise was present. If the focus measure due to noise alone  $\gamma_{noise}$  has a normal distribution with mean  $\gamma_\eta$  and standard deviation  $\text{std}(\gamma_{noise})$ , then  $\beta$  must be 3.0 or more for 99.9% confidence that some contrast information is present in the image region for autofocusing.

In a sense the Laplacian eliminates all unnecessary information (the constant and linear components of grey level variation in the focused image and therefore their contribution to the focus measure) and retains all necessary information (the quadratic and higher level components in the focused image and therefore their contribution to the focus measure). Therefore we recommend the Laplacian FMF for autofocusing and depth-from-focus in practical applications. In the presence of high noise levels, we recommend the Laplacian of the Gaussian (LOG) filter [17] as the FMF.

## 2.4 Shape from Focus (SFF)

Next we describe the algorithm of microscope-based shape and focused image recovery. Fig. 2.9 shows a rough surface of unknown object placed on the translational stage. The object is magnified using the objective lens of the microscope and imaged by a standard CCD camera. The degree of focus in the image is varied by displacing the translational stage (i.e. the object) in the vertical direction with respect to a fixed configuration of camera system and image detector. This is different with

the camera system for macroscopic objects where camera parameters were changed to acquire defocused images [74, 81].

Let us consider the surface patch,  $\Delta_p$ , that lies on the unknown surface of the object.  $\Delta_p$  is assumed a planar surface corresponding to a small window size in the image. Each image contains many surface patches. If the stage is moved towards the focal plane, the image of  $\Delta_p$  will gradually increase in its degree of focus and will be perfectly focused when  $\Delta_p$  lies on the focal plane. Further movements of the patch  $\Delta_p$  will again increase the defocusing of its image. If the stage motion and image acquisition are continuous process, a sequence of images  $N$  can be obtained as the stage is moved in increments of  $\delta$  and an image is acquired at each stage position. To automatically detect the focus position, a focus measure operator is applied on  $\Delta_p$  in the sequence of images. At  $\delta$ , we can search that image frame among the image sequence which has the maximum focus measure. If the position associated with that image frame is found, the height or standard offset of  $\Delta_p$  is computed with respect to the arbitrary selected position where we usually define the first image taken. In fact, the distance of  $\Delta_p$  from focal plane, image detector, or any other coordinate systems can be determined with respect to the imaging system. This procedure may be applied independently to all surface patches to obtain the shape of the entire surface. Besides, the reconstructed image for the object can be synthesized from those focused patches in a set of images.

An accurate shape recovery from image focus analysis depends on

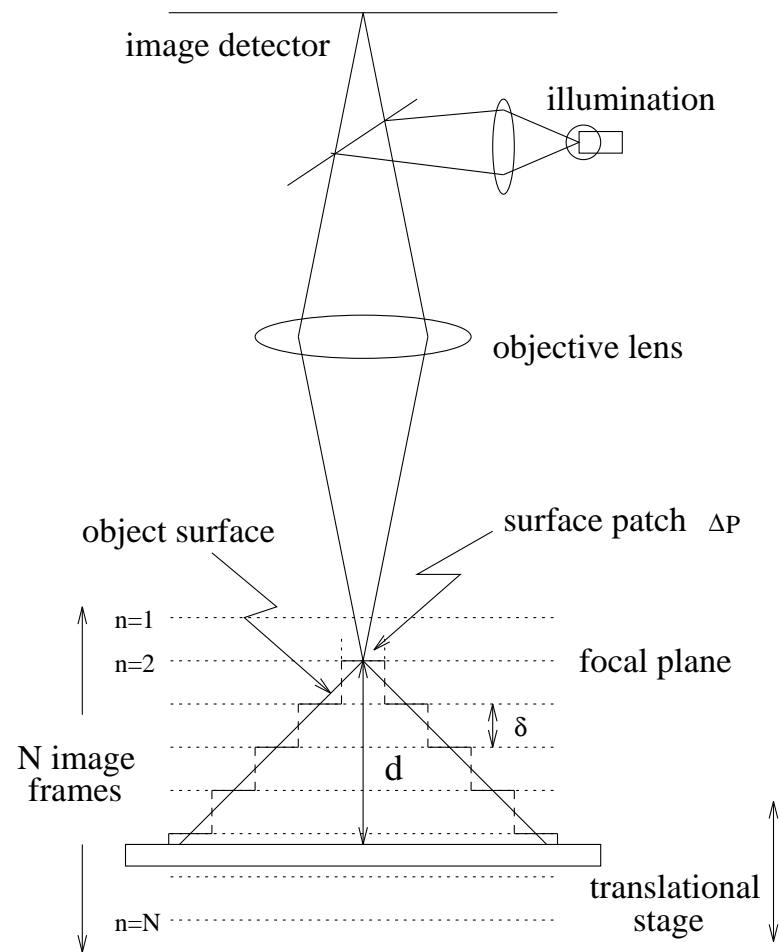


Figure 2.9: Shape from focus

that the spatial surface variation of an object. There are two trade-offs in the shape from focus algorithm. The first, if  $\delta$  is displaced between the two consecutive images and a set of images  $N$  is obtained, the total displacement that is  $(N - 1) \cdot \delta$  has to be greater than a maximum depth  $d$  of the object so that the shape of an object can be completely covered in an image volume. Usually, we want to achieve the best shape recovery results that  $\delta$  is as small as possible (i.e. increasing resolution in field of depth), but it will sacrifice the computational efficiency which needs to process a more number of images. The acceptable choice in the depth resolution depends on surface slope, depth range and algorithm itself. Typically, frame number  $N$  from 30 to 40 and  $\delta$  from 1  $\mu m$  and 10  $\mu m$  are suitable for most of microscopic objects in micro-meter range. In practice, due to the behavior of the focus measure, if the interpolation function such as Gaussian or quadratic is used to localize the peak about several largest focus measures in computing depth estimates, it can benefit in increasing depth resolution but remains the same acquired image frames. Note that if a sequence of defocused images are obtained by changing the camera parameters, the resolution of depth will depend on the resolution of camera parameters.

The second trade-off is the window size versus the spatial resolution in field of view (FOV). As focus measure is computed for each image in a small window size, the window size corresponding to surface patch has to be big enough in order to contain some spatial frequency information. The focus measure is more sensitive to noise and image overlap problem

as if the window size is smaller , and it can result to the wrong depth estimate. The larger window size has more scene information is especially needed for those images with low contrast region. The compromise depends on signal-to-noise ratios and image details. The suggested window size is about 16x16 pixels based on the experiments for the most of well-textured image. An improved method showing more accurate shape recovery has been proposed in [61] which the focus measure should be computed over the focused image surface (FIS) rather than the planar surface.

## **2.5 Implementation**

### **2.5.1 Microscope Camera System**

SFF described above was implemented on a fully automated camera system for the shape recovery of microscopic objects. The system was built over the last three years in our Computer Vision Laboratory. Figs. 2.10 and 2.11 show a block diagram and a photograph of the system respectively. A Olympus BH-2 model microscope is used to image the objects. Objects can be magnified using objective lenses with 5x, 10x, 20x, 50x and 100x magnification. The object can be illuminated using bright field illumination where light energy is focused on the object by the

same lenses that are used to magnify the object. A SONY XC-711 color CCD camera mounted on the microscope is capable of imaging a total of 768x493 pixels with each pixel 13  $\mu m$  in size. The pictures from the camera module through an RGB signal output are captured by a frame grabber board (QuickCapture DT55 of Data Translation). The DT55 board resided in a personal computer with an Intel Pentium 166 MHz processor is able to digitize images with 640x480 square pixel resolution and 256 gray levels. The captured pictures are processed and stored in the computer. In addition to the computer display, a color monitor (SONY PVM-1342Q) which is connected directly to the DT55 board is added. The incoming video signal then can be viewed in real-time on this monitor.

Instead of moving the lens system, the autofocus is done by moving the microscope stage upward and downward. A MD-2 stepper motor of Arrick Robotics Inc. is attached to the focus adjustment knob of the microscope to perform the motorization, and its driver system is connected to the parallel printer port of the computer. The position of the stage can be computer controlled with a resolution and accuracy up to 0.5  $\mu m$ . The shape from focus algorithm can be programmed and executed on the computer. The system has been tested to recover the shape of a variety of industrial, metallic as well as mineral samples.

## 2.5.2 Experimental Results

In this section we present and describe experimental results produced by the above microscope camera system. Due to various spatial resolution and surface reflectance between the objects, we need to adjust the appropriate illumination and magnification lens for each object placed on the microscope stage. Several parameters such as total required image numbers  $N$ , stage displacement  $\pm\delta$  and threshold value also need to be determined and provided into the shape from focus program, where the sign of  $\delta$  represents the direction of the stage motion which we define the stage downward with the positive sign and *vice versa*. The program then automatically increments/decrements the stage position, digitizes and processes an 640x480 image for each position. In the experiments, each image was obtained after four time-averaged as well as image brightness was normalized with respect to its mean value. In fact, the magnification between defocused images depends on their distance from the scene. We may assume the magnification factor for small displacement  $\delta$  to be constant. The Laplacian focus measure and the quadratic interpolation technique are then used in searching the best focus position from a sequence of images. The window size of focus measure is 16x16 which resolves an image with a 40x30 depth map. There are two monitors served for display: the analog monitor receives video signal from camera and displays a real-time picture taken at each stage position, and the computer monitor shows an updated focused image right after the processing. The complete process including image acqui-



sition, stage motion, and computation of depth map and focused image takes a total of about 40 seconds for a sequence of 30 defocused images.

The three samples of the application of shape from focus to 3D microscopy were conducted. In the experimental results, we use step number to represent the recorded image position and define one step increment corresponding to move the stage about  $1\text{ }\mu\text{m}$ . Step number 0 is defined as the position where the first image is obtained. Each increment step is then referred to step number 0. For instance, step 30 stands for the stage moved down about  $30\text{ }\mu\text{m}$  from step 0 and so on. The first sample is a mustard seed. Fig. 2.12 (a) to (f) shows some defocused images of the mustard seed taken at different stage position. As we see, each image frame in the image sequence is mostly blurred by varying degrees in different regions, and only focused in some regions corresponding to those parts of the object which are in the focal plane of the microscope. Shape from focus program is used to identify these focused regions. A total of 30 images were obtained with stage position increments of  $8\text{ }\mu\text{m}$  for this sample. Figs. 2.13 and 2.14 show the reconstructed image and 3D shape profile under the 10x magnification and the  $0.8\text{ mm} \times 0.6\text{ mm}$  field of view.

The second experiment was conducted to recover the 3D shape from a scratch defect on the metallic surface. The sample shows its several defocused images in Fig. 2.15 (a) to (f). In this case, a total of 36 image were obtained with using stage position increments of  $4\text{ }\mu\text{m}$ . The optical setup is the same as the previous experiment. Though some region of

the scratch surface produces images with low contrast, the reconstructed image and 3D shape profile showing in Figs. 2.16 and 2.17 generated by the system are accurate and detailed.

The shape from focus program has also been applied to a variety of industrial samples. Some modern VLSI chips have a Ball Grid Array (BGA) as a pin-out connectors. Each connector or “ball” needs to be inspected for any defects in its roughly hemi-spherical shape. Several defocused images of one BGA ball obtained at different stage positions show in Fig. 2.18 (a) to (f). For this sample, we used a very low illumination to reduce the specular reflection on the top of the ball surface. However, it made the boundary of an image dimmer. A threshold value 6.0 was selected to classify the region with very weak texture as background and a zero depth value was assigned to the background region. This allows to remove a few scattered erroneous depth estimates that result from the lack of texture in some image region. This thresholding number is based on per pixel of focus measure (i.e. normalized to window size) and is obtained from an empirical data. The entire 40 image sequence using stage position increments of  $2\text{ }\mu\text{m}$  is processed to produce the 3D shape of the ball and also its image which is focused everywhere (see Figs. 2.19 and 2.20). In this case, the BGA ball is magnified by 20x lens and viewed in the  $0.4\text{ mm} \times 0.3\text{ mm}$  field of view. Unlike the previous experiments, a repeatability test of ball height can be carried out on this sample by autofocusing the image regions at the substrate position and then at the ball top position through 20 trials. The mean

Figure 2.10: Block diagram

Figure 2.11: Automated SFF microscope system

Figure 2.12: Some defocused images of a mustard seed

Figure 2.13: Reconstructed image of a mustard seed

Figure 2.14: Depth map of the mustard seed

Figure 2.15: Some defocused images of a metal scratch

Figure 2.16: Reconstructed image of a metal scratch

Figure 2.17: Depth map of the metal scratch

Figure 2.18: Some defocused images of a BGA ball



Figure 2.19: Reconstructed image of a BGA ball

Figure 2.20: Depth map of the BGA ball

and standard deviation of ball height from these 20 trials were found about  $51.8 \text{ } \mu\text{m}$  and  $1.67 \text{ } \mu\text{m}$ , respectively.

## 2.6 Summary and Conclusion

In this chapter we have presented shape from focus method for recovering 3D shape and reconstructing the focused image of objects under a microscope. The algorithm itself is based on searching the focus position which we use the Laplacian operator to identify and distinguish it in the image sequence. The refined depth estimate is computed by a quadratic interpolation. We also provided some spatial characteristic analysis about the Laplacian operator and recommended it for use in practical application. The automated microscope system built by us performs the shape from focus technique proved to work well. The system is applicable to objects that are up to a few hundred microns in size. We are able to have the results in a robust, repeatable way and within reasonable time delays.

In the current system, the experiments were conducted on high textured opaque objects. For the objects with non-textured surface, or the surface with specular reflectance, this technique is difficult to work well. Better illumination techniques are needed to facilitate the measurements for those objects with no high frequency contents in their image. For instance, projecting a texture patterns on non-textured surface, or

to utilize an uniform illuminator for shinning surface. The shape from focus method is therefore directly applicable. Our results demonstrate that the shape from focus method is a powerful technique for the microscopic objects.

## **Chapter 3**

# **Noise Sensitivity Analysis of Depth from Focus**

## **3.1 Introduction**

Experimental evaluations of different focus measures have been reported in [30, 35, 52, 39, 60, 76]. So far there has not been any theoretical treatment of the noise sensitivity of focus measures. In the existing literature, all known work have been a combination of experimental observations and subjective judgement. In this chapter we show that the noise sensitivity of a focus measure depends not only on the noise characteristics but also on the image itself. The optimally accurate focus measure for a given noise characteristics may change from one object to the other depending on its image. This makes it difficult to arrive at general conclusions from experiments alone.

For a given camera and object, the most accurate focus measure can be selected from a given set through experiments as follows. For each focus measure, the object is autofocused several times, say 10, starting with an arbitrary default lens position. The mean of the 10 focused positions and their standard deviation are an estimate of the correct focused position and root-mean-square (RMS) error respectively. The focus measure with the minimum estimate of RMS error is taken to be the optimal. In practical applications such as consumer video cameras or digital still cameras, it is desirable to find the best focus measure from a given set by autofocusing only once. It is quite undesirable to repeat 10 or several trials.

If one has a detailed and accurate information on the focused image of the object to be focused and the camera characteristics such as its OTF, noise behaviour, and camera parameters, then it would be possible to estimate the RMS error theoretically with only one trial. However such information is rarely available in practical applications.

In the absence of such detailed and accurate information, we propose two new metrics named *Autofocusing Uncertainty Measure* (AUM) and *Autofocusing Root-Mean-Square Error* (ARMS error) both of which can be computed with only one trial of autofocusing. In DFF applications, AUM and ARMS error can both be easily translated into uncertainties in depth using Eq. (2.1). The key assumption underlying the definition of AUM and ARMS error is that the mean value of focus measures are locally smooth with respect to lens position (e.g. quadratic near the

peak). AUM and ARMS error metrics are general and applicable to any focus measure satisfying the local smoothness assumption. The analysis here deals with focusing errors caused only by grey-level noise and not other factors such as non-front parallel surfaces [28]. The analysis here shows that the autofocusing noise sensitivity of a focus measure depends on the image of the object to be autofocused in addition to the camera characteristics.

For an object with unknown focused image, finding the optimally accurate focus measure involves computing all the candidate focus measures at a set of lens positions and computing AUM/ARMS error for each of the lens positions. Then the lens is moved to the focused position estimated by the optimal focus measure (which has minimum AUM/ARMS error). Usually the number of candidate focus measures that should be considered for good performance is only a few (about 3). Also, almost all focus measures require only a modest amount of computing. Therefore selecting the optimal focus measure from a candidate set comes at a small computational cost. However, if it is necessary to use minimal computing in autofocusing by using the same focus measure for all objects, then as mentioned in Chapter 2 we argue that the energy of the image Laplacian is a good focus measure to use because it is shown to have some important desirable characteristics based on our spatial domain analysis.

Note that AUM and ARMS error metrics are applicable to those focus measures which are based on sum of absolute values of image derivatives,

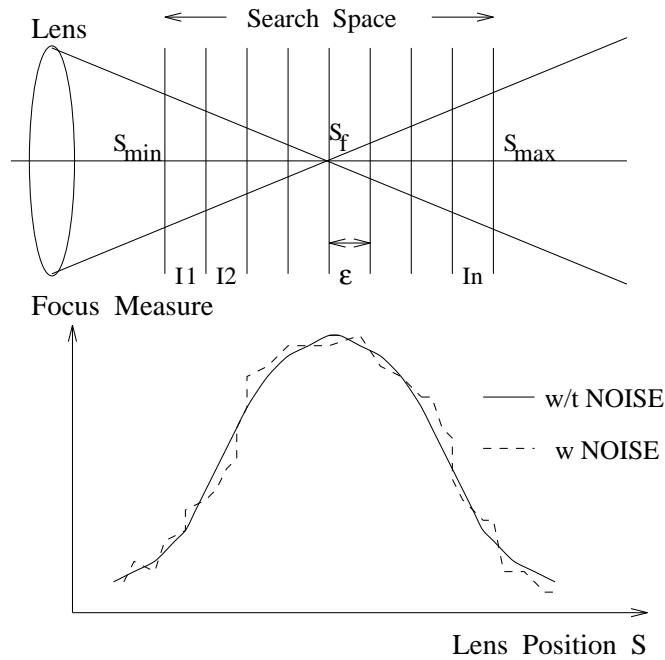


Figure 3.1: Autofocusing Algorithm

but we have not carried out a complete analysis of them here. Since they have been proved to be unsound [67, 60] based on the effect of the OTF main lobe. These unsound focus measures may be optimal for some scenes, but for some other scenes they can give incorrect results even in the absence of all noise. Such filters are proved to be unsound through counter examples in Chapter 2.

## 3.2 Autofocusing Algorithm

We have introduced the concept of DFF method in the previous chapter. Generally speaking, DFF method is a “*search focusing technique*” which necessarily involves a search of the camera parameter space. In this section, we proposed an autofocus algorithm to assist this search

to be done in a rapid way.

In a typical passive autofocus application such as a consumer video camera, autofocus is done by moving the lens with respect to the image detector in a narrow range of about a few millimeters or about one tenth of an inch. Let this range be  $[s_{min}, s_{max}]$  (see Fig. 3.1). A typical value for the limits of the range is  $s_{min} = f$  and  $s_{max} = 1.1f$  where  $f$  is the focal length of the lens. Within the range limits, the problem is to find the lens position  $s_f$  where the image in a specified part of the image detector is in best focus. Due to the limited depth-of-field caused by diffraction effects, the change in the best focused image is indistinguishable by the image detector when the lens is moved in a small range of size  $\epsilon$  around the best focused position  $s_f$ . Therefore there is no benefit in achieving autofocus accuracy better than  $\pm\epsilon/2$ . We only need to move the lens to an arbitrary position in the range  $[s_f - \epsilon/2, s_f + \epsilon/2]$ . Typically  $\epsilon$  is about one part in 200 of  $s_{max} - s_{min}$ . Therefore the range  $[s_{min}, s_{max}]$  can be divided into  $n$  intervals  $I_i = [s_{min} + i \cdot \epsilon, s_{min} + (i + 1) \cdot \epsilon]$  for  $i = 0, 1, 2, \dots, n - 1$  with  $s_{max} = s_{min} + n \cdot \epsilon$ . It is sufficient to compute the focus measures at only one point in each of these intervals during autofocus.

In real-time autofocus applications, the bottle-neck is not the computational time but the time taken for the mechanical motion of the lens to move from one position to the other. Therefore it is important to minimize this time at the cost of some additional computation. Search algorithms such as Fibonacci search and binary search are optimal com-



putationally but not necessarily in the time consumed in lens motion. Also, in consumer applications like hand-held video cameras, it is undesirable for the lens to oscillate between extreme focus and defocus conditions rapidly. People find it uncomfortable. It is desirable for the image to gradually come to focus with only minor overshoots near the focused condition.

Based on the above discussion we propose the following algorithm for autofocusing. First the focus measure is computed at the current lens position and the lens is moved by about  $10\epsilon$  to another position. The focus measure is again computed. The sign of the change in the two focus measures is used to determine the direction in which the lens should be moved. Then a sequential search is begun by moving the lens in steps of about  $(n/8)\epsilon$  in the correct direction until the focus measure decreases for the first time. Then a binary search is initiated in the interval containing the last three lens positions until the search interval has been narrowed to about  $10\epsilon$ . Then a quadratic or a Gaussian is fitted to three or more points which are about  $5\epsilon$  apart to find the focused position. Note that, according to geometric optics, the focus measure curve will be symmetric about the focus position  $s_f$ . Also, shifting focus position  $s_f$  will shift the curve by the same amount with only small change in its shape. This algorithm combines sequential search, binary search, and interpolation, to minimize the lens motion. Additional improvement can be obtained if more information is available about the particular application.

As an alternative to the above algorithm, one may use a depth-

from-defocus algorithm [73, 66] when possible to obtain an estimate of the focus position, and then refine this estimate by computing a focus measure at several points near the estimated position, fitting a curve to the points, and finding the position of the curve maximum. The initial estimate is improved by this method because depth-from-defocus methods are less accurate than the search based method above. After improving the initial estimate of the depth-from-defocus method as here, one can compute AUM/ARMS error for the focus measure used in refining the initial estimate.

### **3.3 A Metric for Focus Measures**

A metric is needed for comparing the noise sensitivity of focus measures both at the focused lens position and at an arbitrary lens position. In Section 3.1, an experimental method was described briefly for finding the focus measure with minimum RMS autofocusing error. The method involves repeatedly autofocusing a given object. Performance of a focus measure at an arbitrary lens position is of interest for the following reason. In practical applications it will be necessary to determine the direction in which the lens should be moved from an arbitrary initial lens position for autofocusing. The desired direction is the direction in which the best focused lens position is located (Fig. 3.1). This direction is found by computing the focus measure at the current lens position

and at another position a small distance away. The direction in which the focus measure increases is the direction in which the lens should be moved. This method will give the correct direction for any sound focus measure in the absence of noise (because the focus measure increases monotonically), but in the presence of noise, some focus measures may be more prone to give the wrong direction than others. The best focus measure for this purpose can be once again determined experimentally. For a given camera system, object, object distance, and lens position, the sign of the finite differences of the focus measure is used to find the direction in which the lens should move in a large number of trials. The percentage of times the correct direction is found is a measure of noise sensitivity. It will be seen later that the best focus measure depends both on the camera PSF and the image of the object. For a given focus measure, the RMS error will change with the camera PSF and the image of the object.

In practical autofocus applications, autofocus of an object has to be done in a few seconds or less in only one trial. Trials cannot be repeated physically to determine the best focus measure. Therefore we need a theoretical metric that can be computed in only one trial of autofocus. The metric should ideally require as little information about the camera system and the object as possible. It should also be simple and not require much computational resources. These desirable characteristics motivate the metric proposed here.

### 3.3.1 Autofocusing Uncertainty Measure (AUM)

First we introduce AUM as a metric for focus measures to illustrate some underlying concepts. Later we introduce the ARMS error which is based on weaker assumptions than AUM. At any lens position  $s_0$  (see Fig. 3.2), each focus measure  $\gamma$  is associated with a probability density function  $p(\gamma(s_0))$ , an expected value (mean)  $E\{\gamma(s_0)\}$ , and a standard deviation  $\text{std}\{\gamma(s_0)\}$ . However, the focus measure with the minimum standard deviation is not necessarily the best because we are not interested in the accuracy of the focus measure itself, but in the corresponding mean lens position and its standard deviation. Estimating the standard deviation of the lens position requires a knowledge of the function that relates the expected value of the focus measure to the lens position (see Fig. 3.2). This function depends on the camera PSF as a function of camera parameters and the focused image of the object. In the absence of accurate information about the camera PSF and the object, the function is estimated in a desired interval through sampling and interpolation. For example, near the maximum, the focus measure may be computed at 3 to 5 nearby lens positions and a smooth function such as a quadratic polynomial or a Gaussian is fitted. The assumption is that the computed values of the focus measure are (nearly) the expected values of the focus measure. This assumption will be removed later in defining the ARMS error.

Referring to Fig. 3.2, the AUM at the maximum of the focus measure

$\gamma$  is defined as follows:

$$\bar{\gamma}_f = \bar{\gamma}(s_f) \quad (3.1)$$

$$\sigma = \bar{\gamma}(s_f) - \bar{\gamma}(s_1) = \bar{\gamma}(s_f) - \bar{\gamma}(s_2) \quad (3.2)$$

$$\text{AUM} = s_2 - s_1 \quad (3.3)$$

where  $\sigma$  is the standard deviation of the focus measure. In order to compute AUM, we need to know  $\sigma$ . In the next section we derive a general formula that can be used to estimate  $\sigma$  as a function of the image and its noise level. Further we need to know the shape of the curve  $\bar{\gamma}(s)$  near the peak. As discussed earlier, the position of  $\bar{\gamma}_f$  and the function  $\bar{\gamma}(s)$  near  $\bar{\gamma}_f$  are estimated by fitting a curve (quadratic or Gaussian) to a few points (at least 3) near the maximum. Intuitively, AUM is a measure similar to the RMS error in lens position that can be determined through repeated trials.

Fig. 3.3 shows a typical comparison of two focus measures. The maximum values of the two focus measures have been normalized to be the same. We see that although  $\sigma_2 > \sigma_1$ ,  $\text{AUM}_2 < \text{AUM}_1$ , implying that  $\gamma_2$  is better than  $\gamma_1$ .

Referring to figure 3.6, focus measure  $\bar{\gamma}$  is modeled to be locally quadratic in a small interval of size  $2\delta$  with respect to lens position near the focused position:

$$\bar{\gamma}(s) = as^2 + bs + c \quad (3.4)$$

Let the focus measure be given at three arbitrary positions which are  $\delta$  apart. Without loss of generality, let the three positions be  $s_- = -\delta$ ,  $s_0 = 0$ , and  $s_+ = +\delta$ . Let  $\bar{\Gamma}_- = \bar{\gamma}(s_-)$ ,  $\bar{\Gamma}_0 = \bar{\gamma}(s_0)$  and  $\bar{\Gamma}_+ = \bar{\gamma}(s_+)$ . Near the focused position,  $\bar{\Gamma}_0 > \bar{\Gamma}_-$  and  $\bar{\Gamma}_0 > \bar{\Gamma}_+$ . Solving for the coefficients of the quadratic expression, we obtain

$$\begin{aligned} a &= \frac{\bar{\Gamma}_+ + \bar{\Gamma}_- - 2\bar{\Gamma}_0}{2\delta^2} \\ b &= \frac{\bar{\Gamma}_+ - \bar{\Gamma}_-}{2\delta} \\ c &= \bar{\Gamma}_0 \end{aligned} \quad (3.5)$$

Let  $s_f$  be the lens position where the focus measure becomes the maximum and  $\bar{\Gamma}_f = \bar{\gamma}(s_f)$ . At  $s_f$ , the derivative of  $\bar{\Gamma}$  vanishes. Therefore we obtain

$$\begin{aligned} s_f &= \frac{-b}{2a} \\ &= \frac{\delta}{2} \frac{(\bar{\Gamma}_+ - \bar{\Gamma}_-)}{(2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-)} \end{aligned} \quad (3.6)$$

Substituting the above equation in (3.4) we obtain

$$\bar{\Gamma}_f = -\frac{b^2 - 4ac}{4a} \quad (3.7)$$

Given that  $AUM = s_2 - s_1$ , we obtain  $s_1$  and  $s_2$  as the roots of the equation

$$\begin{aligned} \bar{\Gamma}(s) &= \bar{\Gamma}_f - \sigma \\ &= as^2 + bs + c \end{aligned} \quad (3.8)$$

Thus solving the above equation, we obtain

$$AUM = \left| \frac{\sqrt{b^2 - 4a(c + \sigma - \bar{\Gamma}_f)}}{a} \right| \quad (3.9)$$

Substituting with Eqs. (3.5) and (3.7) in the above equation yields

$$\text{AUM} = 2\delta \left( \frac{2\sigma}{2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-} \right)^{\frac{1}{2}} \quad (3.10)$$

At a position far away from the focused lens position  $s_f$ , AUM is defined as in Fig. 3.4. This is somewhat similar to that near the peak:

$$\sigma = (\bar{\gamma}(s_2) - \bar{\gamma}(s_1))/2 \quad (3.11)$$

$$\text{AUM} = s_2 - s_1 \quad (3.12)$$

Once again,  $\sigma$  is computed from the known noise characteristics and the image. The shape of the focus measure curve is estimated by a linear (2 points) interpolation using the values of the focus measure at  $s_-$  and  $s_+$  that are  $\delta$  apart. Without loss of generality, let  $s_- = -\delta/2$  and  $s_+ = +\delta/2$  and the focus measures at these points be  $\bar{\Gamma}_-$  and  $\bar{\Gamma}_+$  respectively (see Fig. 3.7). The linear model yields the expression

$$\frac{s - s_-}{s_+ - s_-} = \frac{\bar{\Gamma} - \bar{\Gamma}_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \quad (3.13)$$

The above equation can be rewritten as:

$$s = \delta \left( \frac{\bar{\Gamma} - \bar{\Gamma}_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \right) - \frac{\delta}{2} \quad (3.14)$$

We obtain  $s_1$  and  $s_2$  by solving

$$\bar{\Gamma}(s) = \frac{\bar{\Gamma}_+ + \bar{\Gamma}_-}{2} \pm \sigma \quad (3.15)$$

where  $\sigma$  is the standard deviation of the focus measure. Using equation (3.14) and solving for AUM, we obtain

$$\begin{aligned} \text{AUM} &= |s_1 - s_2| \\ &= \frac{2\delta\sigma}{|\bar{\Gamma}_+ - \bar{\Gamma}_-|} \end{aligned} \quad (3.16)$$

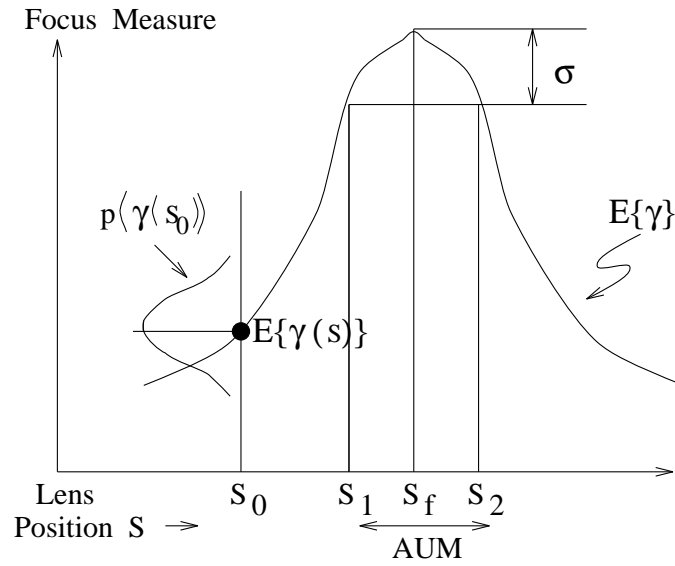


Figure 3.2: Definition of AUM at the focused position  $S_f$ .

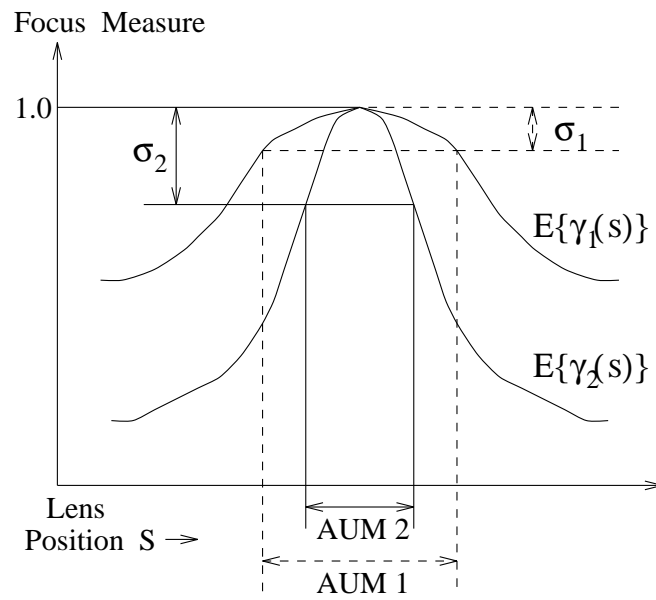


Figure 3.3: Comparison of two focus measures  $\gamma_1$  and  $\gamma_2$  at the focused position



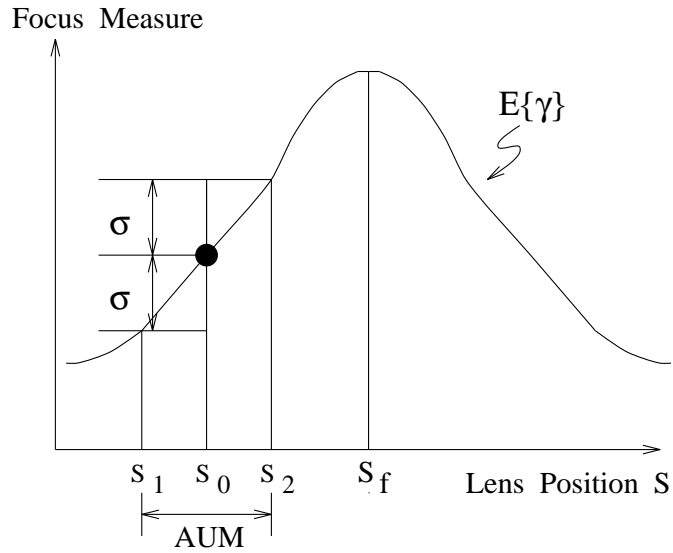


Figure 3.4: Definition of AUM at a position  $S_0$  far from the focused position  $S_f$ .

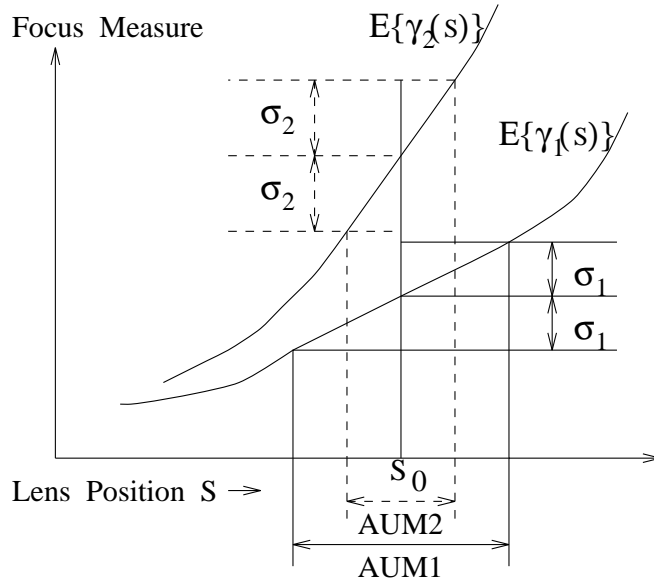


Figure 3.5: Comparison of two focus measures  $\gamma_1$  and  $\gamma_2$  at a position  $S_0$  far from the focused position  $S_f$ . Note:  $\sigma_2 > \sigma_1$  but  $AUM2 < AUM1$ , therefore  $\gamma_2$  is better than  $\gamma_1$ .

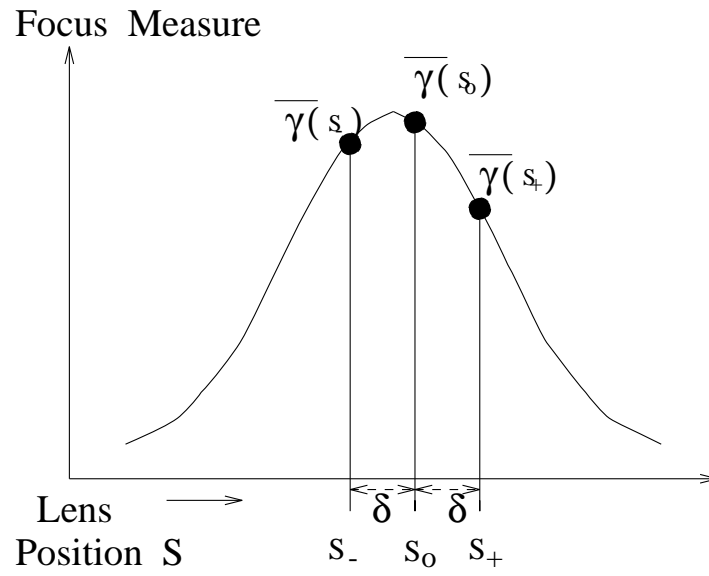


Figure 3.6: quadratic polynomial interpolation

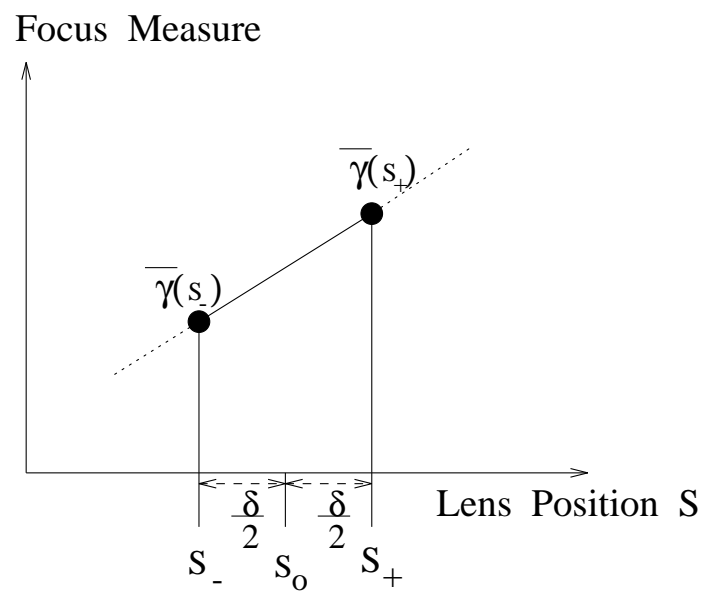


Figure 3.7: linear interpolation

Fig. 3.5 shows a comparison of two focus measures far away from the focused position. Once again We see that although  $\sigma_2 > \sigma_1$ ,  $AUM_2 < AUM_1$ , implying that  $\gamma_2$  is better than  $\gamma_1$ .

### 3.4 ARMS Error

In this section, an explicit expression for the *Autofocusing Root-Mean Square Error* (ARMS error) is derived, which is based on weaker assumptions than AUM. An exact expression for the RMS error depends on the Optical Transfer Function (OTF) of the camera and the Fourier spectrum of the focused image. Deriving such an exact expression is complicated because of the nature of the camera's OTF and the variability of the Fourier spectrum of the focused image for different objects. Further, usefulness of such an expression in practical applications is limited since all the information necessary to evaluate the expression (e.g. OTF and camera parameters) may not be available. However, an approximate expression that is very useful in practical applications can be derived under some weak assumptions. The assumption we use is that *the expected value of the focus measure is locally smooth with respect to lens position*. We model this local smoothness by a quadratic polynomial, but the analysis here can be extended to other models (e.g. cubic or Gaussian). However such extensions do not appear to offer significant advantages compared to the quadratic model in practical applications.

We are interested in the RMS value of  $s_{max}$ . For this reason, the focus measure  $\Gamma_i$  will be expressed as the summation of their expected value  $\bar{\Gamma}_i$  and their noise component  $n_i$ :

$$\Gamma_i = \bar{\Gamma}_i + n_i \quad \text{for } i = -, 0, +. \quad (3.17)$$

In this case we obtain a set of equations similar to Eqs (3.4) to (3.7) with the difference that  $\bar{\Gamma}_i$  are replaced by  $\Gamma_i$ , therefore we obtain

$$\begin{aligned} s_{max} &= \frac{\delta}{2} \left( \frac{\Gamma_+ - \Gamma_-}{2\Gamma_0 - \Gamma_+ - \Gamma_-} \right) \\ &= \frac{\delta}{2} \left( \frac{\bar{\Gamma}_+ - \bar{\Gamma}_- + n_+ - n_-}{2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_- + 2n_0 - n_+ - n_-} \right) \\ &= \frac{\delta}{2} \left( \frac{\bar{\Gamma}_+ - \bar{\Gamma}_-}{2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-} \right) \left( 1 + \frac{n_+ - n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \right) \left( 1 + \frac{2n_0 - n_+ - n_-}{2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-} \right)^{-1} \end{aligned} \quad (3.18)$$

Near the focused position we have  $\bar{\Gamma}_0 > \bar{\Gamma}_+$  and  $\bar{\Gamma}_0 > \bar{\Gamma}_-$ . Therefore, if the signal to noise ratio is sufficiently large, we have

$$|2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-| \gg |2n_0 - n_+ - n_-| \quad (3.19)$$

We obtain  $s_{max} \approx s'_{max}$  where

$$s'_{max} = \bar{s}_{max} \left( 1 + \frac{n_+ - n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \right) \quad (3.20)$$

Note: we cannot assume that  $|\bar{\Gamma}_+ - \bar{\Gamma}_-| \gg |n_+ - n_-|$  because, near the focused position,  $\bar{\Gamma}_+$  and  $\bar{\Gamma}_-$  may be nearly equal. Simplifying the expression for  $s'_{max}$  we obtain

$$s'_{max} = \bar{s}_{max} + \frac{\delta}{2} \left( \frac{n_+ - n_-}{2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-} \right) \quad (3.21)$$

Now the ARMS error is defined as the standard deviation of  $s'_{max}$ , i.e.

$$\begin{aligned} \text{ARMS error} &= \frac{\delta}{2} \frac{1}{(2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-)} \cdot \text{std}(n_+ - n_-) \\ &= \frac{\delta}{2} \frac{1}{(2\bar{\Gamma}_0 - \bar{\Gamma}_+ - \bar{\Gamma}_-)} \cdot (\sigma_+^2 + \sigma_-^2)^{\frac{1}{2}} \end{aligned} \quad (3.22)$$

where  $\sigma_+$  and  $\sigma_-$  are the standard deviations of the focus measures  $\Gamma_+$  and  $\Gamma_-$  respectively.

For a lens position away from the maximum focused position, the above expression for ARMS error will not be valid since the assumption in Equation (3.19) will not be valid. In this case we find that  $\Gamma_- < \Gamma_0 < \Gamma_+$ , the local linear model for the focus measure will be better than the local quadratic model. The ARMS error for this case is defined based on focus measures at only two lens positions (rather than three) that are  $\delta$  apart. Without loss of generality, let the two positions be  $s_- = -\delta/2$  and  $s_+ = +\delta/2$  and the focus measures at these points be  $\Gamma_-$  and  $\Gamma_+$  respectively (similar to Fig. 3.7). The linear model yields the expression

$$\frac{s - s_-}{s_+ - s_-} = \frac{\Gamma - \Gamma_-}{\Gamma_+ - \Gamma_-} \quad (3.23)$$

The above equation can be rewritten as:

$$s = \delta \left( \frac{\Gamma - \Gamma_-}{\Gamma_+ - \Gamma_-} \right) - \frac{\delta}{2} \quad (3.24)$$

Once again, we express  $\Gamma_+$  and  $\Gamma_-$  as  $\Gamma_+ = \bar{\Gamma}_+ + n_+$  and  $\Gamma_- = \bar{\Gamma}_- + n_-$  where  $\bar{\Gamma}_+$  and  $\bar{\Gamma}_-$  are the expected values and  $n_+$  and  $n_-$  are the noise components.

Now the ARMS error is defined as the standard deviation of  $s'$  where  $s'$  is the solution of  $\Gamma(s) = \frac{\bar{\Gamma}_+ + \bar{\Gamma}_-}{2}$ . Solving this equation we obtain

$$\begin{aligned} s' &= \frac{\delta}{2} \left( \frac{\bar{\Gamma}_+ - \bar{\Gamma}_- - 2n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_- + n_+ - n_-} \right) - \frac{\delta}{2} \\ &= \frac{\delta}{2} \left[ \left( 1 - \frac{2n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \right) \left( 1 + \frac{n_+ - n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \right)^{-1} - 1 \right] \end{aligned} \quad (3.25)$$

Assuming  $|\bar{\Gamma}_+ - \bar{\Gamma}_-| \gg |n_+ - n_-|$  and  $|\bar{\Gamma}_+ - \bar{\Gamma}_-| \gg |2n_-|$ , we obtain

$$\begin{aligned} s' &\approx \frac{\delta}{2} \left[ 1 - \frac{2n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} - \frac{n_+ - n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} - 1 \right] \\ &\approx \frac{\delta}{2} \left( \frac{n_+ + n_-}{\bar{\Gamma}_+ - \bar{\Gamma}_-} \right) \end{aligned} \quad (3.26)$$

Hence, the ARMS error would be

$$\text{ARMS error} = \text{std}(s') = \frac{\delta \left( \sigma_+^2 + \sigma_-^2 \right)^{\frac{1}{2}}}{2|\bar{\Gamma}_+ - \bar{\Gamma}_-|} \quad (3.27)$$

### 3.4.1 Relation between AUM and ARMS Error

Comparing the expressions for AUM and ARMS error from equations (3.10) and (3.22) we find

$$\begin{aligned} \frac{\text{AUM}^2}{\text{ARMS}} &= 16\delta \frac{\sigma}{\sqrt{\sigma_+^2 + \sigma_-^2}} \\ &\approx 8\sqrt{2}\delta \quad \text{if } \sigma_+ \approx \sigma_- \approx \sigma \end{aligned} \quad (3.28)$$

The ratio of the square of AUM and ARMS error is a constant. Therefore AUM and ARMS error are monotonically related. If we re-define AUM so that instead of using equation (3.8) we obtain  $s_1$  and  $s_2$

by solving

$$\Gamma(s) = \Gamma_{max} - \sigma^2 \quad (3.29)$$

then we find that AUM and ARMS error are linearly related for a given focus measure and focused image.

For a lens position far away from the focused position, comparing the expressions (3.16) and (3.27) for AUM and ARMS error yields

$$\begin{aligned} \frac{\text{AUM}}{\text{ARMS}} &= \frac{4\sigma}{\sqrt{\sigma_+^2 + \sigma_-^2}} \\ &\approx 2\sqrt{2} \quad \text{if } \sigma_+ \approx \sigma_- \approx \sigma \end{aligned} \quad (3.30)$$

For this case, they are linearly related.

### 3.5 Noise Sensitivity Analysis

In this section we derive expressions for the *expected value* (mean) and *variance* of the focus measures modeled in Chapter 2. These are useful in computing the *standard deviation*  $\sigma$  of the focus measure and its AUM/ARMS error.

Let  $f(m, n)$  be the blurred noise free discrete image and  $\eta(m, n)$  be the additive noise. The noisy blurred digital image recorded by the camera is

$$f_\eta(m, n) = f(m, n) + \eta(m, n) \quad (3.31)$$

The noise  $\eta(m, n)$  at different pixels are assumed to be independent, identically distributed random variables with zero mean and standard

deviation  $\sigma_n$ . This  $\sigma_n$  can be easily estimated for a camera by imaging a uniformly bright object and computing the standard deviation of the grey level distribution. The images are assumed to be of size  $(2N+1) \times (2N+1)$  and focus measure filter (FMF)  $a(i, j)$  of size  $(2M+1) \times (2M+1)$ . Without loss of generality, the filtering operation will be represented by the *moving weighted sum* (MWS) operator instead of the usual *convolution* operator. MWS is correlation and is equivalent to convolution if, for example, the FMF is rotated by 180 degrees about its center by assigning  $a(-i, -j)$  to  $a(i, j)$ . Denoting the MWS operator by  $\star$  it is defined by

$$a(i, j) \star f_\eta(m, n) = \sum_{i=-M}^M \sum_{j=-M}^M a(i, j) f_\eta(m+i, n+j) \quad (3.32)$$

In the remaining part of this paper we shall use the following convention to simplify notation. A double summation will be abbreviated with a single summation as:

$$\sum_{m=-N}^N \sum_{n=-N}^N = \sum_{m,n}^N \quad \text{and} \quad \sum_{i=-M}^M \sum_{j=-M}^M = \sum_{i,j}^M \quad (3.33)$$

Let  $g(m, n)$  be the image obtained by filtering the noisy blurred image  $f_\eta(m, n)$  with the FMF  $a(i, j)$ :

$$g(m, n) = a(i, j) \star f_\eta(m, n) = F(m, n) + \mathcal{N}(m, n) \quad (3.34)$$

where

$$F(m, n) = a(i, j) \star f(m, n) \quad (3.35)$$

$$\mathcal{N}(m, n) = a(i, j) \star \eta(m, n) \quad (3.36)$$



The focus measure  $\gamma$  is defined as

$$\begin{aligned}\gamma &= \frac{1}{(2N+1)^2} \sum_{m,n}^N g^2(m,n) \\ &= \gamma_{signal} + \gamma_{noise} + \frac{2}{(2N+1)^2} \sum_{m,n}^N F(m,n) \mathcal{N}(m,n)\end{aligned}\quad (3.37)$$

where  $\gamma_{signal}$  and  $\gamma_{noise}$  are defined by:

$$\gamma_{signal} = \frac{1}{(2N+1)^2} \sum_{m,n}^N F^2(m,n) \quad (3.38)$$

$$\gamma_{noise} = \frac{1}{(2N+1)^2} \sum_{m,n}^N \mathcal{N}^2(m,n) \quad (3.39)$$

Now the expected value of the focus measure  $E\{\gamma\}$  is (note that the expectation operator  $E$  is linear and commutes with summation):

$$E\{\gamma\} = \gamma_{signal} + \gamma_{\eta} + \frac{2}{(2N+1)^2} \sum_{m,n}^N F(m,n) E\{\mathcal{N}(m,n)\} \quad (3.40)$$

where

$$\gamma_{\eta} = E\{\gamma_{noise}\} \quad (3.41)$$

Since we assume  $\eta(m,n)$  is zero mean, the last term of equation (3.40) will vanish. Now the second term can be written as

$$\gamma_{\eta} = \frac{1}{(2N+1)^2} \sum_{m,n}^N \sum_{i_1,j_1}^M \sum_{i_2,j_2}^M a(i_1,j_1) a(i_2,j_2) E\{\eta(m+i_1,n+j_1) \eta(m+i_2,n+j_2)\} \quad (3.42)$$

In the above equation, if  $i_1 \neq i_2$  or  $j_1 \neq j_2$ , then, since noise in different pixels are independent and zero mean, we obtain

$$\begin{aligned}& E\{\eta(m+i_1,n+j_1) \eta(m+i_2,n+j_2)\} \\ &= E\{\eta(m+i_1,n+j_1)\} E\{\eta(m+i_2,n+j_2)\} \\ &= 0\end{aligned}\quad (3.43)$$

However, if  $i_1 = i_2$  and  $j_1 = j_2$ , then

$$E\{\eta^2(m + i_1, n + j_1)\} = \sigma_n^2 \quad (3.44)$$

Therefore, we get

$$\gamma_\eta = \frac{1}{(2N + 1)^2} \sum_{m,n}^N \sum_{i,j}^M a^2(i, j) \sigma_n^2 = A_n \sigma_n^2 \quad (3.45)$$

where

$$A_n = \sum_{i,j}^M a^2(i, j) \quad (3.46)$$

Therefore

$$E\{\gamma\} = \gamma_{signal} + \gamma_\eta = \gamma_{signal} + A_n \sigma_n^2 \quad (3.47)$$

The above equation is a fundamental result. It shows that the expected value of the focus measure is a sum of two components— one due to signal alone and another due to noise alone. Therefore, if a focus measure is computed on a set of images for autofocusing, the effect of noise is to increase the computed focus measure by the same value on average for all images. The reason for this is that while the image signal changes in blur level with lens position, the noise characteristics of the camera remains the same. Therefore, the average increase in focus measure due to noise does not change the location of the focus measure peak. It is the variance of the focus measure that changes the location of the focus measure peak and therefore introduces error in autofocusing.

Now consider the variance of the focus measure:

$$Var\{\gamma\} = E\{\gamma^2\} - (E\{\gamma\})^2 \quad (3.48)$$

From Eq.(3.37), noting that  $E\{\mathcal{N}(m, n)\} = 0$  and  $E\{\mathcal{N}^2(m_1, n_1)\mathcal{N}(m_2, n_2)\} = 0$ , we obtain

$$E\{\gamma^2\} = \gamma_{signal}^2 + 2\gamma_{signal}\gamma_\eta + E\{\gamma_{noise}^2\} + \frac{4}{(2N+1)^4} \times \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N F(m_1, n_1)F(m_2, n_2)E\{\mathcal{N}(m_1, n_1)\mathcal{N}(m_2, n_2)\} \quad (3.49)$$

From Eqs. (3.47 , 3.48, 3.49) we obtain

$$Var\{\gamma\} = E\{\gamma_{noise}^2\} - \gamma_\eta^2 + \frac{4}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N F(m_1, n_1)F(m_2, n_2)E\{\mathcal{N}(m_1, n_1)\mathcal{N}(m_2, n_2)\} \quad (3.50)$$

Note that  $E\{\gamma_{noise}^2\} - \gamma_\eta^2$  is the variance of  $\gamma_{noise}$  which is independent of signal. Therefore , equation (3.50) can be written as:

$$Var\{\gamma\} = Var\{\gamma_{noise}\} + \frac{4}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N F(m_1, n_1)F(m_2, n_2)E\{\mathcal{N}(m_1, n_1)\mathcal{N}(m_2, n_2)\} \quad (3.51)$$

The equation above shows that the variance of a focus measure depends on the image signal in addition to noise level. Further simplification of the above expression is presented in Appendix A.1. The formula presented there can be applied directly in practical applications. Now we consider 3 examples to illustrate the application of the formula. In these examples, the noise will be modeled as Gaussian. For a zero mean Gaussian random variable  $\eta$  with standard deviation  $\sigma_n$  we have [44]  $E\{\eta^4\} = 3\sigma_n^4$ . This result will be used in the following examples.

## 1. Gray Level Variance

The image is normalized by subtracting the mean grey value from the grey level of each pixel. The focus measure filter in this case is

$$a(i, j) = \begin{cases} 1 & \text{if } i = j = 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.52)$$

Using the formula (A.16) for variance derived in Appendix A.1 we obtain

$$Var\{\gamma\} = \frac{2\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^N f^2(m, n) \quad (3.53)$$

## 2. Gradient Magnitude Squared

For gradient squared along x-axis

$$a_x(i, j) = [-1 \quad 1] \quad (3.54)$$

Substituting  $a(i, j)$  above in Eq. (A.16) for variance in Appendix A.1 we obtain:

$$Var\{\gamma_x\} = \frac{12\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A_x(i, j) * f(m, n)]^2 \quad (3.55)$$

where

$$A_x(i, j) = [-1 \quad 2 \quad -1] \quad (3.56)$$

For gradient squared along y-axis

$$a_y(i, j) = [-1 \quad 1]^T \quad (3.57)$$

In the same manner, we have

$$Var\{\gamma_y\} = \frac{12\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A_y(i, j) * f(m, n)]^2 \quad (3.58)$$

where

$$A_y(i, j) = [-1 \quad 2 \quad -1]^T \quad (3.59)$$

A cross item  $Var\{\gamma_{xy}\}$  is generated by the effect of noise on  $\gamma_x$  and  $\gamma_y$  which are not independent (see Appendix A.2 for more details). The expression of  $Var\{\gamma_{xy}\}$  can be derived as

$$Var\{\gamma_{xy}\} = \frac{8\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A_x(i, j) * f(m, n)] [A_y(i, j) * f(m, n)] \quad (3.60)$$

Therefore, combining above equations, we obtain

$$\begin{aligned} Var\{\gamma\} &= Var\{\gamma_x\} + Var\{\gamma_y\} + Var\{\gamma_{xy}\} \\ &= \frac{24\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A_x(i, j) * f(m, n) + \\ &\quad A_y(i, j) * f(m, n)]^2 \end{aligned} \quad (3.61)$$

### 3. Laplacian

The discrete Laplacian is approximated by

$$a(i, j) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3.62)$$

Substituting this  $a(i, j)$  into the formula (A.16) for variance we obtain

$$Var\{\gamma\} = \frac{1352\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A(i, j) * f(m, n)]^2 \quad (3.63)$$

where

$$A(i, j) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & -8 & 2 & 0 \\ 1 & -8 & 20 & -8 & 1 \\ 0 & 2 & -8 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (3.64)$$

## 3.6 Experiments

In the first set of experiments, Eq. (3.51) for the variance of focus measures was verified as follows. The autofocusing algorithm described earlier was implemented on a system named Stony Brook Passive Autofocusing and Ranging Camera System (SPARCS) [60]. In SPARCS, a 35 mm focal length lens is used. The lens is driven by a stepper motor that can move the lens to 97 different step positions. The standard deviation of the camera noise was estimated by imaging a flat and uniformly bright object and then computing the grey level variance of the recorded image. The noise characteristics of the camera is shown in Figure 3.11 by plotting the grey level profile of one row of pixels at two different illumination levels (300 and 500 lux). Three objects labeled A,B, and C (see Fig. 3.12) were used in the experiments.

An object was placed in front of the camera, and for some fixed lens

position, 10 images of size  $32 \times 32$  of the object were recorded. These images slightly differed from each other due to electronic noise. A given focus measure was computed for each of the 10 images. The standard deviation of the resulting 10 focus measures was then computed. This was the experimentally determined standard deviation of the focus measure. The theoretical estimation of the standard deviation of the focus measure was computed using equation (3.51). For this purpose, the standard deviation of the noise was obtained as mentioned earlier using a flat uniformly bright object. The noise-free image needed in equation (3.51) was obtained by averaging 4 noisy images of the object. The procedure for this experiment is shown in Fig. 3.8. Table 3.1 shows the experimentally computed and theoretically estimated standard deviations of different focus measures. We see that the two values are close thus verifying Equation (3.51).

In the next experiment, the objects A, B, and C, were autofocused using the algorithm described in section 3.2. In each case, the experimental and theoretical ARMS error were computed (the unit is lens steps). Near the focus position, images were recorded at 3 positions  $s_-$ ,  $s_0$  and  $s_+$  which were 5 steps apart. At each position, 10 images were recorded, and using these the mean and the standard deviation of the focus measure there were computed. Then the theoretically estimated ARMS error was computed using Eq. (3.22). The same data was used to compute 10 experimental focus positions using Eq. (3.18). The standard deviation of these 10 positions was the experimental ARMS error. Fig.

3.9 shows the procedure for experimental verification. The resulting values are shown in the last two columns of Table 3.1. We see that they are very close. These values also indicate the relative autofocusing accuracy of the three focus measure filters— grey level variance, gradient magnitude squared, and Laplacian squared. The measured noise standard deviation was 0.95 (grey level units) for the camera, and the SNR for the three objects were 35 dB, 28 dB and 20 dB respectively.

Three main conclusions can be drawn from the experimental results. First, for a given object (i.e. fixed image content), ARMS error decreases with increasing signal-to-noise ratio (SNR). This implies that low contrast objects and noisy cameras have more autofocusing error. Second, the focus measure with minimum standard deviation is not necessarily the focus measure that gives minimum error in autofocusing. Third, best focus measure could be different for different objects depending on both image content and noise characteristics; SNR alone cannot be used to determine the best focus measure. For example, the best focus measure for the objects with SNR 35 dB and SNR 28 dB are the Laplacian squared, but for the object with SNR 20 dB, the best focus measure is gradient magnitude squared. The gray level variance performed very poorly for object C and the autofocusing was totally unreliable due to the absence of a well defined peak. This is indicated by the N/A entries in the table.



### 3.6.1 Computer Simulation Experiments

Experiments similar to the ones above were carried out on simulated image data. The purpose of these experiments was to further verify our theoretical results. In these experiments, unlike in the previous experiments, the noise-free image data and precise characteristics of noise were known accurately. Therefore we expected a closer agreement than previous experiments between theoretically estimated and experimentally determined values of the standard deviation of focus measures. This expectation was satisfied thus verifying our theory more accurately. In addition, unlike the previous experiments, the simulation experiments were carried out at many different levels of noise rather than at only one level of noise. The theory was verified to be correct at all noise levels. The test object shown in Fig. 3.13.a was added with various levels of zero-mean Gaussian random noise to get a set of noisy images. At each noise level, the mean and the standard deviation of the focus measure were computed using 10 noisy images. Then the standard deviation of the focus measures were estimated theoretically using Eqs. (3.53), (3.61) and (3.63). The above procedure is shown in Fig. 3.10. The plots in Fig. 3.13 from (b) to (d) show that the experimental and theoretical standard deviation are in close agreement at all noise levels for all three focus measures.

Another experiment similar to the second experiment for real data described earlier was conducted to verify Equation (3.22) for ARMS error on simulation data as follows. The focused image of a planar object

normal to the optical axis was used as input to a program that models image sensing in a CCD video camera. The program was a modified version of the Image Defocus Simulator (IDS) developed by Lu [64]. The IDS program was modified to improve the accuracy of blurred images computed for small degrees of blur. The improved accuracy was achieved by increasing the sampling rate and by using a wave optics model [20] of the camera's PSF (see Appendix A.4). A sequence of blurred images were generated corresponding to different lens positions in the SPARCS camera system in our laboratory.

Three images near the focused position were selected from the image sequence generated above and a specified level of zero-mean Gaussian random noise was added to these. Then the focused position was computed using Eq. (3.18). The above step was repeated 10 times, and the standard deviation of the resulting 10 values of the focused positions was calculated to obtain experimental value of the ARMS error. Then the ARMS error was estimated theoretically using Eq. (3.22). The process above was repeated for various noise levels and three different focus measures. The results are plotted in Fig. 3.14. We see that the two ARMS error are in good agreement. In a similar manner, a plot of the two AUMs for various noise levels are shown in Fig. 3.15. The monotonic relation between ARMS error and AUM are also demonstrated from those two plots.

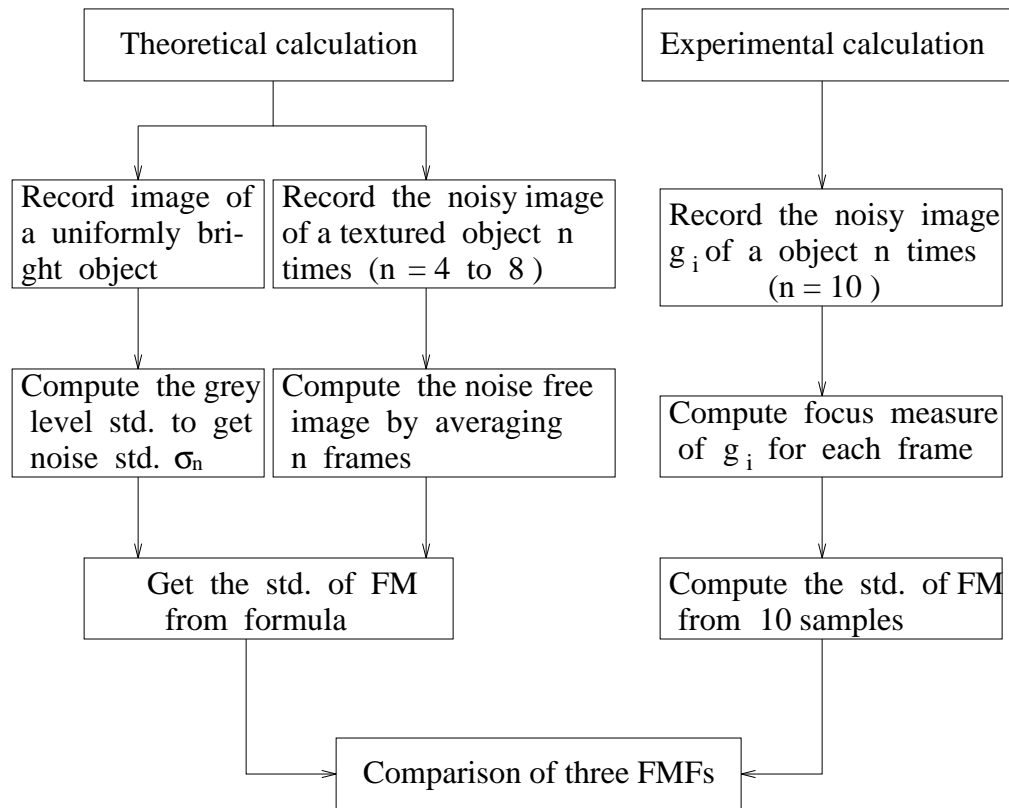


Figure 3.8: The verification of variance of focus measures in SPARCS system

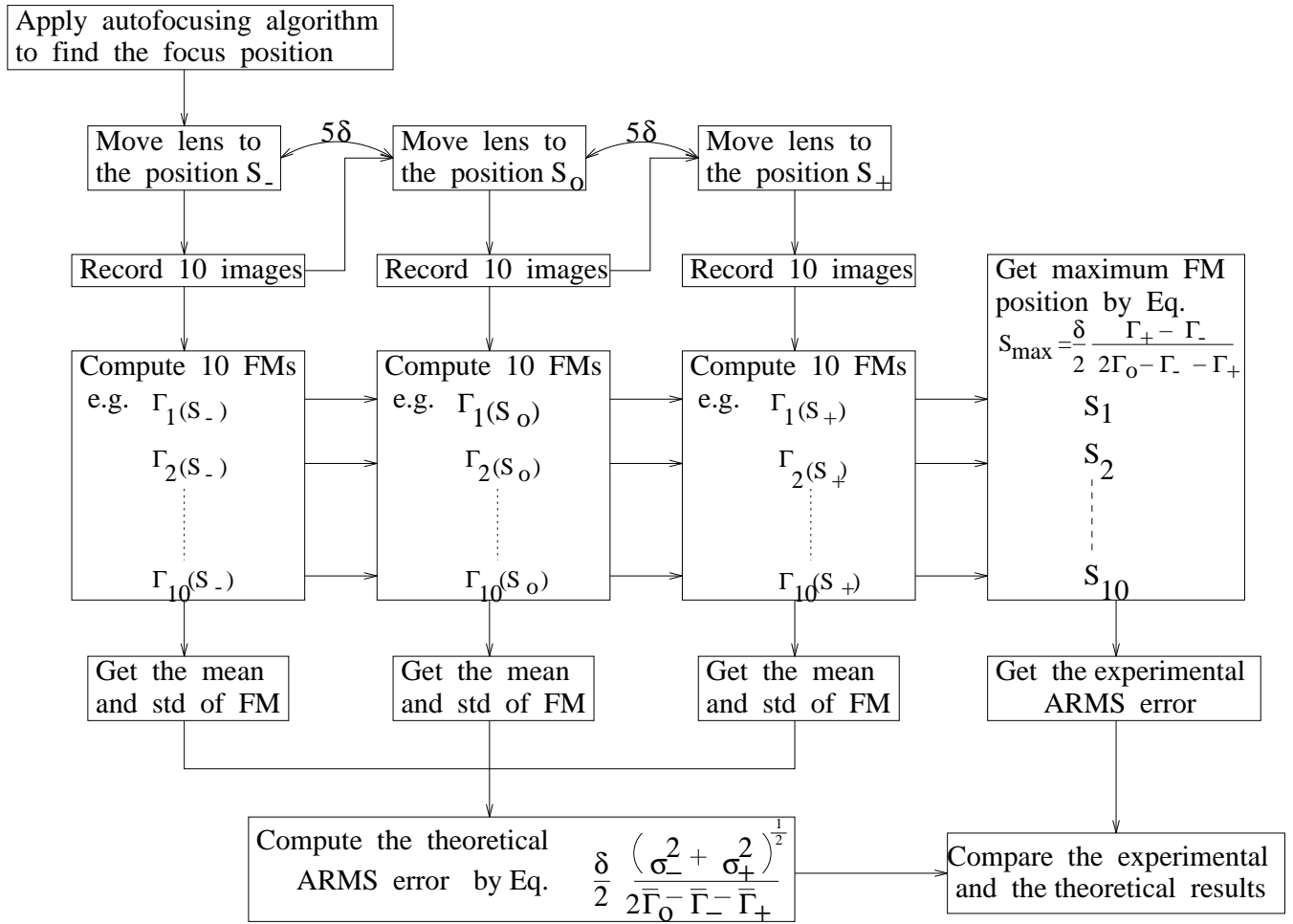


Figure 3.9: The verification of ARMS errors in SPARCS system

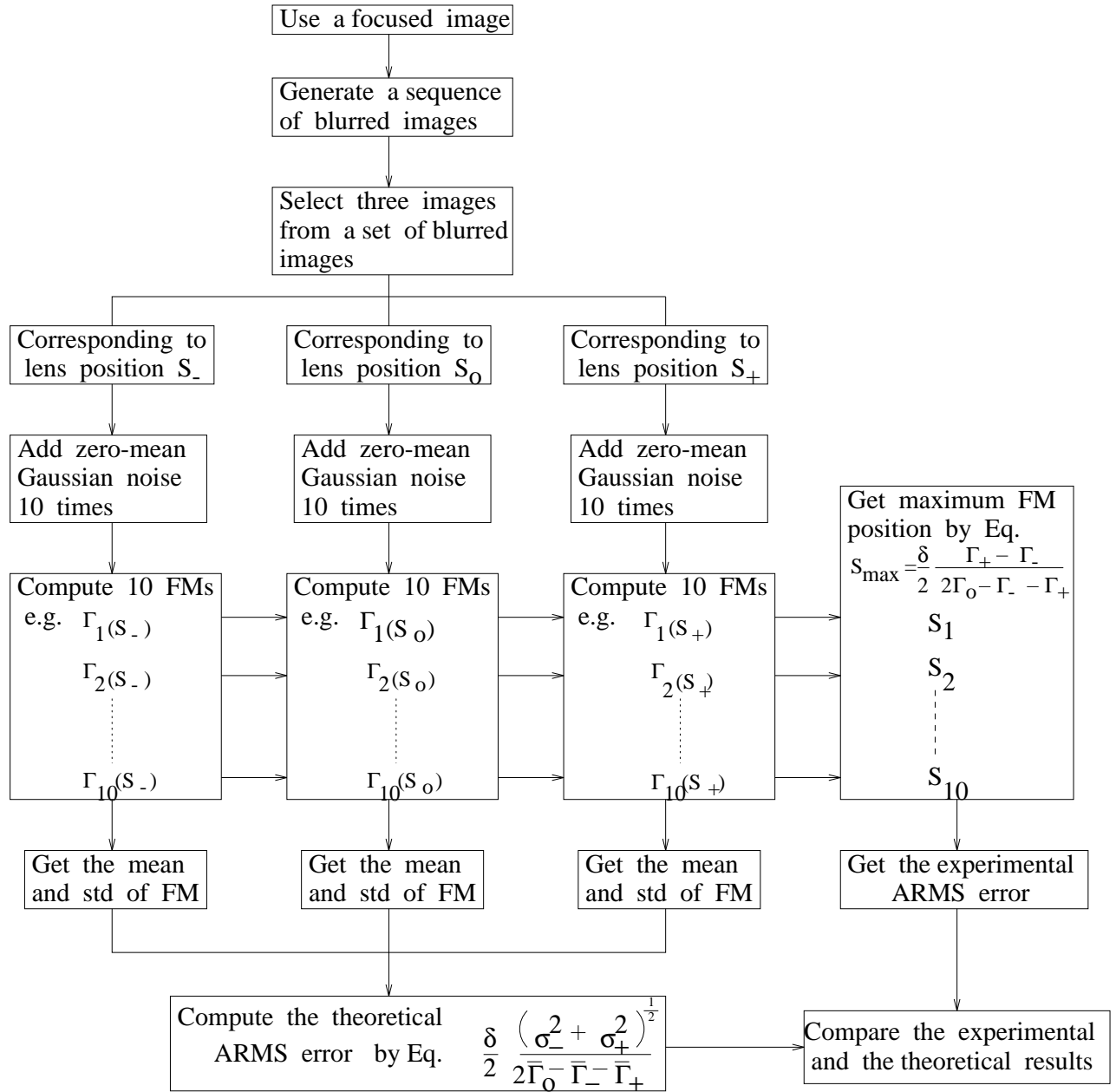


Figure 3.10: The verification of ARMS errors in computer simulation

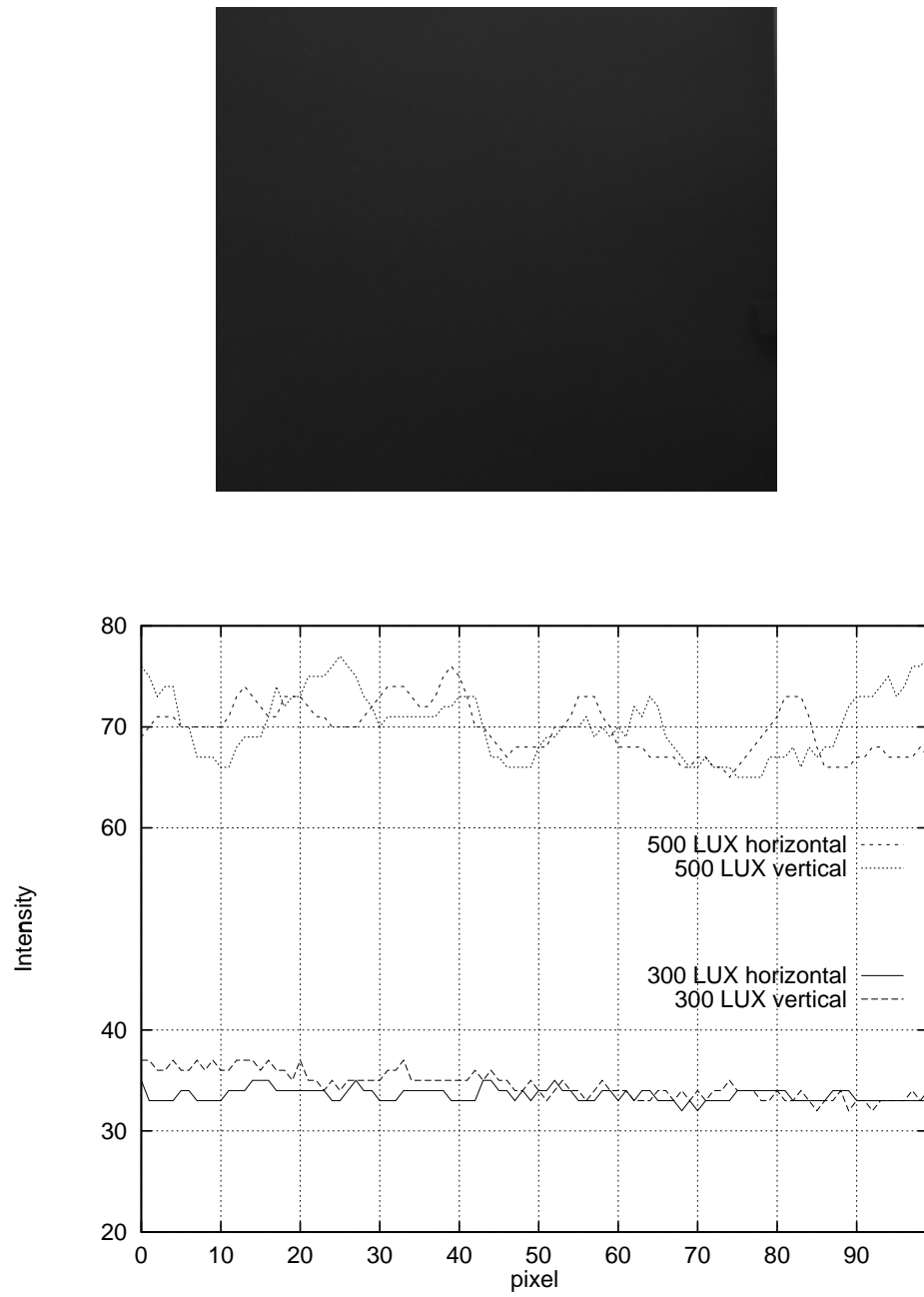


Figure 3.11: Noise characteristics in 1-D

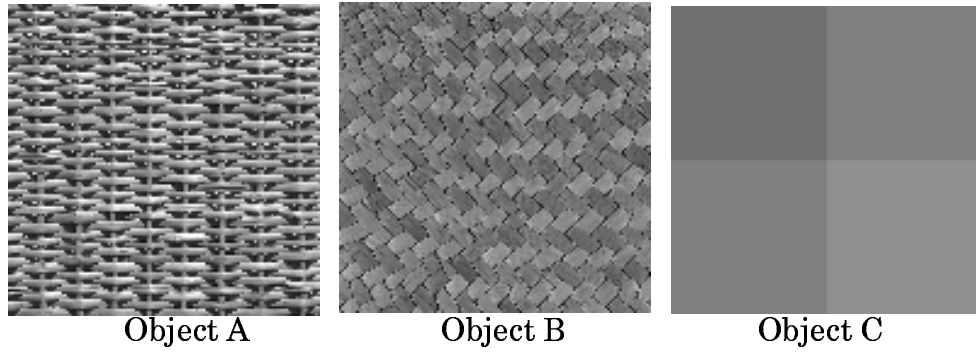


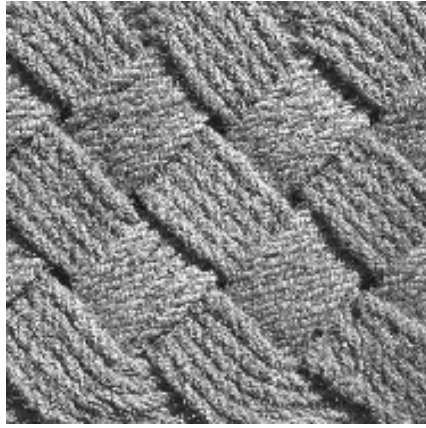
Figure 3.12: Textured images

Object A SNR: 35dB	Theoretical std of FM	Experiment std of FM	Theoretical ARMS	Experiment ARMS
Laplacian	18.92	17.27	0.020	0.018
Gradient	5.87	6.31	0.023	0.024
Variance	1.82	2.13	0.025	0.028

Object B SNR: 28dB	Theoretical std of FM	Experiment std of FM	Theoretical ARMS	Experiment ARMS
Laplacian	3.71	4.05	0.044	0.043
Gradient	1.06	1.25	0.048	0.049
Variance	0.85	1.02	0.10	0.11

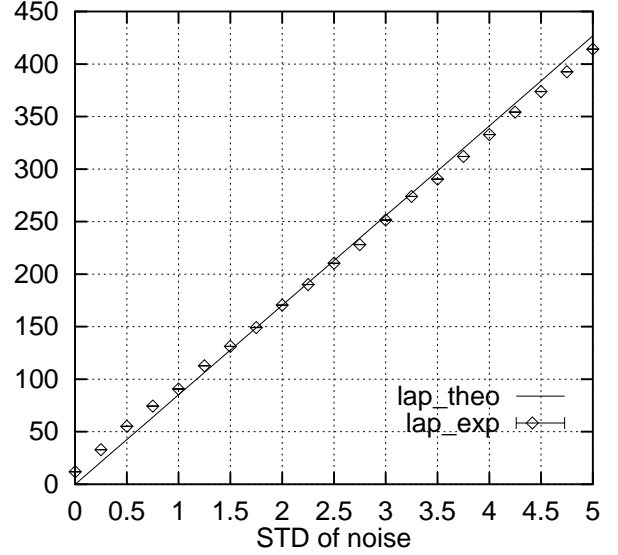
Object C SNR: 20dB	Theoretical std of FM	Experiment std of FM	Theoretical ARMS	Experiment ARMS
Laplacian	1.67	1.37	0.09	0.10
Gradient	0.32	0.46	0.06	0.07
Variance	N/A	N/A	N/A	N/A

Table 3.1: Experimental results



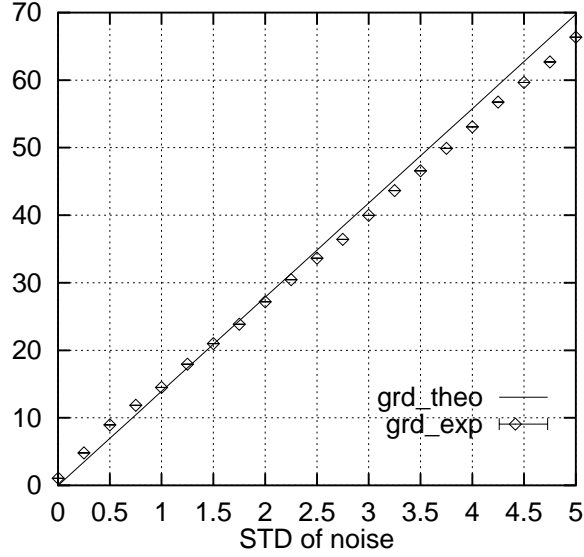
Test object  
(a)

STD of Lap. FM



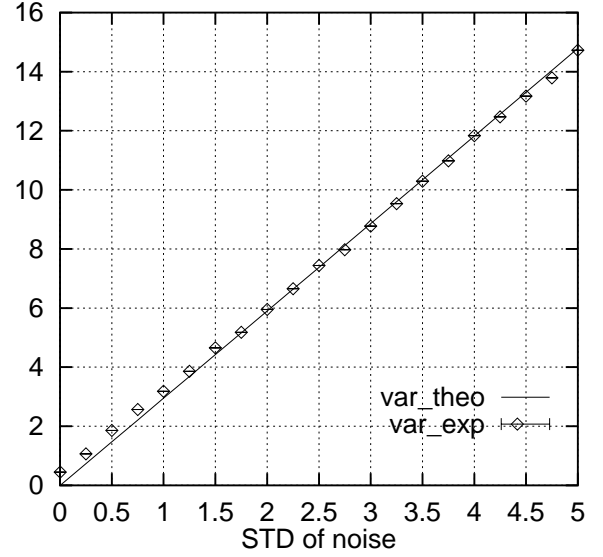
(b)

STD of Grd. FM



(c)

STD of Var. FM



(d)

Figure 3.13: Simulation records



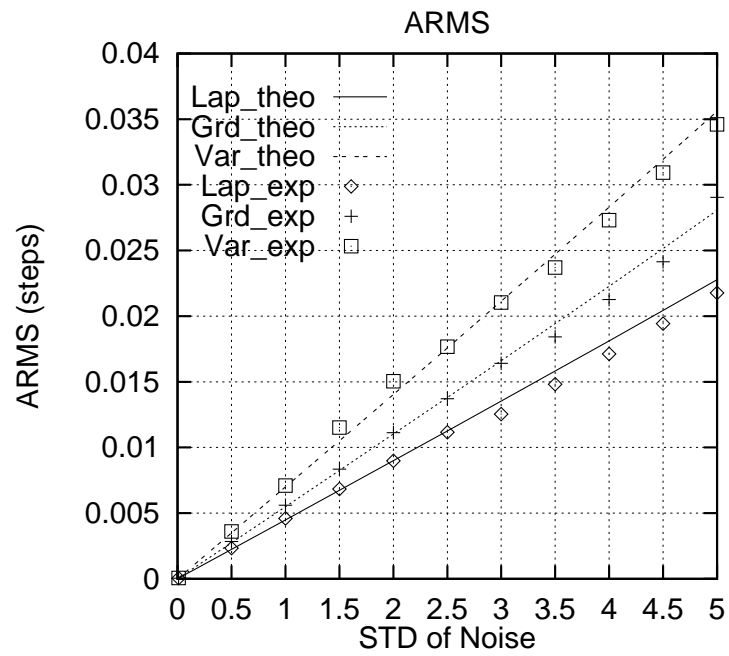


Figure 3.14: ARMS vs Noise

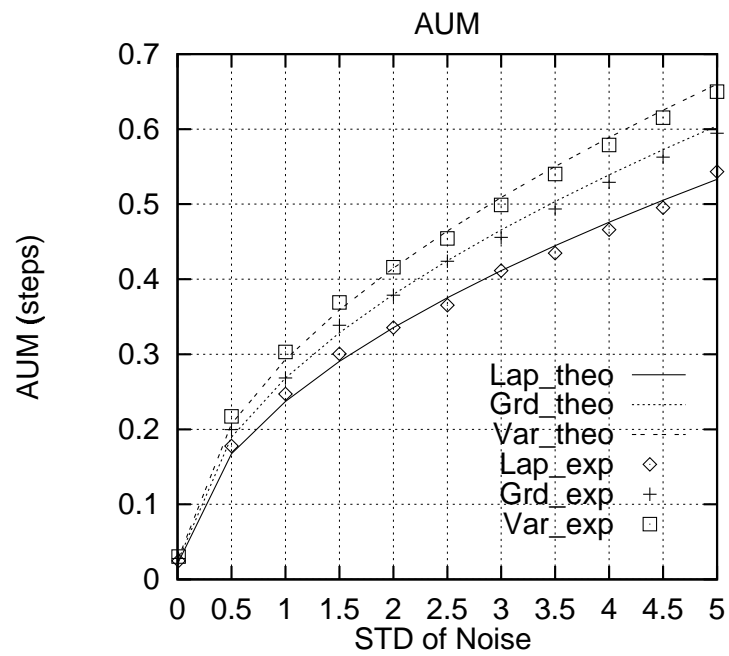


Figure 3.15: AUM vs Noise

### 3.7 Conclusion

ARMS error has been defined as a metric for selecting the optimal focus measure for autofocusing with respect to grey-level noise from a given set of focus measures. It is based on the assumption of local smoothness of focus measures with respect to lens position. ARMS error can be applied to any focus measure whose variance can be expressed explicitly as a function of grey-level noise variance. Such an expression has been derived for a large class of focus measures that can be modeled as the energy of filtered images. Equations 3.47 and A.16 for the mean and variance respectively of a focus measure along with Equations 3.22 and 3.27 for ARMS error completely specify the dependence of autofocusing error on both grey-level noise and image content. These equations can be used to estimate the autofocusing accuracy of different focus measures, and the one with minimum error can be selected for application. In applications where computation needs to be minimized by computing only one focus measure, we recommend the use of the Laplacian as the focus measure filter. Laplacian has some desirable properties such as simplicity, rotational symmetry, elimination of unnecessary information and retaining of necessary information.

This work can be extended in several ways. First, explicit expressions for the variance of other focus measures such as sum of absolute

values of image derivatives could be derived so that ARMS error can be used to estimate their autofocusing accuracy. Second, in the definition of ARMS error, the local smoothness of focus measures could be modeled by an  $n$ -th order polynomial or a continuous function with  $n + 1$  parameters instead of the quadratic polynomial. In this case, the focus measure will need to be measured at  $n + 1$  points (instead of 3 for the quadratic model) in local intervals of interest. Third, deriving an optimal focus measure filter for a given image and noise level remains to be investigated. One way to approach this problem would be to first minimize noise in images using an optimal noise-reduction filter such as Wiener filter or matched filter [50, 77] and then applying the Laplacian focus measure filter. This involves additional computation, but it will improve autofocusing accuracy in most cases.

## **Chapter 4**

# **Noise Sensitivity Analysis of Depth from Defocus**

### **4.1 Introduction**

Three-dimensional (3D) shape recovery through image defocus analysis has been called Depth-from-Defocus (DFD). In DFD methods, unlike Depth-from-Focus (DFF) methods, it is not required to focus an object in order to find the distance of the object. DFD has been investigated by many researchers [10, 46, 73, 56, 57, 66, 32]. In image defocus analysis or DFD, a few images (about 2-3) of a 3D scene are acquired by a camera with different degrees of defocus. The distance of an object is directly estimated from a measure of the level of defocus in the images. DFD methods do not involve searching for camera parameters (e.g. lens position or focal length) in contrast with Depth-from-Focus methods which

require a large number of images recorded with different camera parameters, and searching for the parameters that result in the sharpest focused image [30, 39, 60, 67].

In DFD, a method based on a new *Spatial Domain Convolution/Deconvolution Transform* was proposed by Subbarao and Surya [66, 59]. This method, named S Transform Method or STM, involves simple *local operations* in the spatial domain on only two images. The operations correspond to extracting depth information by cancelling the effect of focused image of an object through local deconvolution. In STM, the acquisition of two images can be taken with changing different camera parameters such as lens position, focal length and aperture diameter. Both images can be arbitrarily blurred and neither of them needs to be a focused image. STM is faster in comparison with DFF methods, but less accurate. In Chapter 3, we have provided a theoretical and experimental treatment of noise sensitivity of DFF methods [67]. In this chapter we address the noise sensitivity of the DFD approach based on STM. A theoretical treatment of this problem is presented, and a method is developed for estimating the Root-Mean-Square (RMS) error for STM. In the existing literature, noise sensitivity of DFD methods is limited to experimental observations only. In this context, the theory developed in this chapter is new, and it facilitates estimating errors under various levels of noise and image contrast (as measured by its Laplacian). In order to verify our theory easily and precisely, we use sampled images of a blurred step edge and a cubic polynomial function in simulation

experiments. In addition, experiments conducted on real images are implemented. In our analysis, random noise is assumed to be additive, zero-mean, and uncorrelated. Our analysis shows that the RMS error of STM in terms of the focused lens position of an object is linearly related to the noise standard deviation.

Next section describes the camera model and defocused image. Section 4.3 provides a summary of STM approach to DFD. The noise sensitivity analysis of STM is presented in Section 4.4, and experimental results in Section 4.5.

## 4.2 Camera Model and Defocused Image

As mentioned in Chapter 2, the image formation in a camera with variable camera parameters  $(s, f, D)$  is shown in Fig. 2.1. The camera parameter setting is denoted by  $\mathbf{e}_i$  where  $\mathbf{e}_i = (s_i, f_i, D_i)$  is a vector representing the  $i$ -th camera parameter setting. The relationship between the focused lens position  $v$ , the focal length of the lens  $f$ , and the distance of the object  $u$ , is given by the well-known lens formula (see Eq. 2.1). In Fig. 2.1, there is the radius  $R$  of blur circle which is a size of blurred image on the image detector as if the object point is not in focus. According to paraxial geometric optics, it can be shown using the lens formula and similar triangles that the radius  $R$  of the blur circle can be expressed as a function of the camera setting and object distance  $u$  as

$$R = \frac{Ds}{2} \left( \frac{1}{f} - \frac{1}{u} - \frac{1}{s} \right) \quad (4.1)$$

Image magnification varies with the distance  $s$  between the lens and the image detector in Fig. 2.1. Therefore the magnification is normalized corresponding to some standard distance  $s_0$ . After this normalization, the blur circle radius  $R'$  becomes

$$R' = \frac{Ds_0}{2} \left( \frac{1}{f} - \frac{1}{u} - \frac{1}{s} \right) \quad (4.2)$$

The blur circle corresponds to the Point Spread Function (PSF) of the camera. Let the PSF be denoted by  $h(x, y)$ . If the object surface is approximated to be planar and parallel to the image detector plane in a small image region, then the observed image  $g(x, y)$  is the convolution of the focused image  $f(x, y)$  with the PSF  $h(x, y)$  (also see in Chapter 2). Circularly symmetric PSFs are parameterized in terms of a *spread parameter*  $\sigma$  which is the *standard deviation* of distribution of  $h(x, y)$ . It is defined as

$$\sigma^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + y^2) h(x, y) dx dy \quad (4.3)$$

It can be found that  $\sigma$  is proportional to the blur circle radius  $R'$ . The constant of proportionality is dependent on the optics. In most practical cases except as  $\sigma$  is very small, it is shown that

$$\sigma \approx \frac{R'}{\sqrt{2}} \quad (4.4)$$

is a good approximation [62, 58]. Combining Eqs (4.2) and (4.4),  $\sigma$  is the function of camera parameters  $(s, f, D)$  and the distance of the object  $u$  which can be expressed as

$$\sigma = mu^{-1} + c \quad (4.5)$$

where

$$m = -\frac{Ds_0}{2\sqrt{2}} \quad c = \frac{Ds_0}{2\sqrt{2}} \left[ \frac{1}{f} - \frac{1}{s} \right] \quad (4.6)$$

It is shown that  $\sigma$  linearly depends on inverse distance  $u^{-1}$  for a given camera setting. Note that the sign of  $\sigma$  depends on the direction of defocusing observed on the image detector.

If there are two camera parameter settings  $e_1, e_2$ , we have

$$\sigma_1 = \frac{D_1 s_0}{2\sqrt{2}} \left( \frac{1}{f_1} - \frac{1}{u} - \frac{1}{s_1} \right) \quad , \quad \sigma_2 = \frac{D_2 s_0}{2\sqrt{2}} \left( \frac{1}{f_2} - \frac{1}{u} - \frac{1}{s_2} \right) \quad (4.7)$$

Eliminating  $1/u$  from the above two relations we obtain

$$\sigma_1 = \alpha \sigma_2 + \beta \quad (4.8)$$

where

$$\alpha = \frac{D_1}{D_2} \quad , \quad \beta = \frac{D_1 s_0}{2\sqrt{2}} \left( \frac{1}{f_1} - \frac{1}{f_2} + \frac{1}{s_2} - \frac{1}{s_1} \right) \quad (4.9)$$

### 4.3 STM: A Spatial Domain Approach to DFD

The central problem in depth from defocus is to measure the amount of blur in a small region of an image. The solution for the depth of the object is to find out the relation from the blurred (observed) image and the parameter (i.e. radius) of blur circle. STM method based on such a concept provides a useful solution. In this section we provide a very brief summary of STM. Full details of this method can be found in [66, 74]. A new spatial-domain convolution/deconvolution transform (S Transform) is defined in [59]. The definition of the transform for the general case is



quite complicated. However, a special case of the transform suffices for image defocus analysis. This case turns out to be particularly simple.

For two-dimensional images, under a local cubic polynomial model, the transform 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:

$$g(x, y) = f(x, y) + \frac{\sigma^2}{4} \nabla^2 f(x, y) \quad (4.10)$$

where  $\nabla^2$  is the *Laplacian* operator. The *inverse S transform* of  $g$  with respect to the moment vector  $(1, \sigma^2/2)$  is equal to  $f$  and it is defined as

$$f(x, y) = g(x, y) - \frac{\sigma^2}{4} \nabla^2 g(x, y) \quad (4.11)$$

In the general case, neither the object function  $f(x, y)$  nor the PSF  $h(x, y)$  (i.e. blur circle) is known, resulting in two unknown function and hence requiring two images for a solution. Let two blurred images  $g_1$  and  $g_2$  are acquired with different camera parameter settings  $\mathbf{e}_1$  and  $\mathbf{e}_2$  corresponding to blur parameters  $\sigma_1$  and  $\sigma_2$ , we obtain

$$f(x, y) = g_1(x, y) - \frac{\sigma_1^2}{4} \nabla^2 g_1(x, y) \quad (4.12)$$

$$f(x, y) = g_2(x, y) - \frac{\sigma_2^2}{4} \nabla^2 g_2(x, y) \quad (4.13)$$

Eliminating  $f(x, y)$  from the above two relations, and using Eq. (4.8), and the fact that  $\nabla^2 g_1 = \nabla^2 g_2$  (which can be proved using Eq. (4.10) where  $f(x, y)$  is a cubic polynomial), we obtain

$$(\alpha^2 - 1)\sigma_2^2 + 2\alpha\beta\sigma_2 + \beta^2 = \frac{8(g_1 - g_2)}{\nabla^2 g_1 + \nabla^2 g_2} \quad (4.14)$$

where  $\alpha$  and  $\beta$  are as defined in the previous section.

In the above equation  $\sigma_2$  is the only unknown. The equation is quadratic and therefore  $\sigma_2$  is easily obtained by solving it. The two solutions result in a two-fold ambiguity. Methods for resolving this ambiguity are discussed in [73, 66, 74, 81]. From the solution for  $\sigma_2$ , the distance  $u$  of the object is obtained from Eq. (4.7).

The above discussion illustrates the conceptual feasibility of determining distance from two defocused images. Repeating the above procedure in all image neighborhoods, the depth-map of an entire scene can be obtained from only two blurred images in parallel. In fact, in order to minimize the image overlap problem and make the method robust in the presence of noise, Eq. 4.14 can be modified by squaring first and then integrating over a small region around the point  $(x, y)$ . The best estimate of  $\sigma_2$  is then taken by the mode of the histogram distribution [74].

## 4.4 Noise Sensitivity Analysis of STM

In this section we derive expressions for the *expected value* (mean) and *variance* of the focused lens position  $v$  in the lens formula (Fig. 2.1) for STM. The mean and variance will be expressed in terms of the camera parameters  $e_i$  and the noise variance  $\sigma_n^2$ . These expressions are useful in computing the root-mean-square (RMS) error for STM in estimating the focused lens position  $v$  due to noise. We start the derivation from the

right hand side of Eq. (4.14). Note that we do not consider the integrated version of STM here [74].

Let  $g_k$  be the blurred noise free discrete images,  $\eta_k$  be the additive noise for  $k = 1, 2$ . The two noisy blurred digital images recorded by the camera are

$$g_{\eta_k} = g_k + \eta_k \quad \text{for } k = 1, 2 \quad (4.15)$$

The noise  $\eta_k$  at different pixels are assumed to be independent, identically distributed (i.i.d.) random variables with zero mean and standard deviation  $\sigma_n$ .

In STM the assumption of local cubic polynomial model for the focused image is relaxed by using a set of discrete image smoothing filters [36, 66]. The recorded images are convolved with a smoothing kernel  $L_o$  in this case as

$$g'_{\eta_k} = g_{\eta_k} * L_o \quad (4.16)$$

where  $*$  denotes convolution. Another filter  $L_2$  is used for estimating the second order image derivatives for computing the Laplacian of the image. Denoting the image Laplacian by  $g''_{\eta_k}$ , we have

$$g''_{\eta_k} = g_{\eta_k} * L_2 \quad (4.17)$$

Using Eqs. (4.15), (4.16) and (4.17), we can rewrite the right hand side of Eq. (4.14) as

$$\begin{aligned} G &= 8 \frac{g'_{\eta_1} - g'_{\eta_2}}{g''_{\eta_1} + g''_{\eta_2}} \\ &= 8 \frac{(g_1 * L_o - g_2 * L_o) + (\eta_1 * L_o - \eta_2 * L_o)}{(g_1 * L_2 + g_2 * L_2) + (\eta_1 * L_2 + \eta_2 * L_2)} \end{aligned} \quad (4.18)$$

In the above equation, there are two components in both the numerator and the denominator where one component depends on image signal only (but not noise) and the other component on noise only. Let  $P$  and  $Q$  denote the numerator and the denominator respectively in the above equation, and let  $P = \overline{P} + P'$  and  $Q = \overline{Q} + Q'$ , where

$$\overline{P} = g_1 * L_o - g_2 * L_o \quad (4.19)$$

$$P' = \eta_1 * L_o - \eta_2 * L_o \quad (4.20)$$

$$\overline{Q} = g_1 * L_2 + g_2 * L_2 \quad (4.21)$$

$$Q' = \eta_1 * L_2 + \eta_2 * L_2 \quad (4.22)$$

Now we can write

$$\begin{aligned} G &= 8 \frac{P}{Q} = \frac{\overline{P} + P'}{\overline{Q} + Q'} \\ &= 8 \frac{\overline{P}}{\overline{Q}} \left( 1 + \frac{P'}{\overline{P}} \right) \left( 1 + \frac{Q'}{\overline{Q}} \right)^{-1} \end{aligned} \quad (4.23)$$

Assuming  $\overline{Q} \gg Q'$  and  $\overline{P} \overline{Q} \gg P' Q'$  (which will be true when the signal-to-noise ratio is sufficiently large), we obtain

$$G \approx 8 \frac{\overline{P}}{\overline{Q}} \left( 1 + \frac{P'}{\overline{P}} - \frac{Q'}{\overline{Q}} \right) \quad (4.24)$$

Note that we cannot assume that  $\overline{P} \gg P'$  because, the two images may have similar blur level although they correspond to quite different camera parameter settings, and therefore  $\overline{P}$  may be close to zero.

Now the expected value of  $G$  is

$$E\{G\} = 8 \frac{\overline{P}}{\overline{Q}} + \frac{8}{\overline{Q}} E\{P'\} - \frac{8\overline{P}}{\overline{Q}^2} E\{Q'\} \quad (4.25)$$

Since we assume  $\eta_k$  to be zero mean, the last two terms of the above equation will vanish. (Note that the expectation operator  $E$  is linear and commutes with summation.) Therefore we obtain

$$E\{G\} = 8 \frac{\bar{P}}{\bar{Q}} \quad (4.26)$$

This result shows that the expected value  $E\{G\}$  depends only on the signal, but not noise.

Now we consider the variance of  $G$ . Using Eqs (4.24) and (4.26), and noting that  $E\{P'\} = E\{Q'\} = 0$ , we obtain

$$\begin{aligned} Var\{G\} &= E\{G^2\} - (E\{G\})^2 \\ &= E\left\{64 \left( \frac{\bar{P}}{\bar{Q}} + \frac{P'}{\bar{Q}} - \frac{\bar{P}Q'}{\bar{Q}^2} \right)^2\right\} - \left(8 \frac{\bar{P}}{\bar{Q}}\right)^2 \\ &= \frac{64}{\bar{Q}^2} E\{P'^2\} + \frac{64\bar{P}^2}{\bar{Q}^4} E\{Q'^2\} - \frac{128\bar{P}}{\bar{Q}^3} E\{P'Q'\} \end{aligned} \quad (4.27)$$

In the Appendix B.1, it is shown that the term  $E\{P'Q'\}$  vanishes, and the term  $E\{P'^2\}$  and  $E\{Q'^2\}$  will depend on the standard deviation  $\sigma_n$  of the noise and the coefficients of the filters. Denoting the coefficients of  $L_0$  filter by  $a_0(i, j)$  and the coefficients of  $L_2$  filter by  $a_2(i, j)$ , we obtain

$$Var\{G\} = \mathcal{A} \sigma_n^2 \quad (4.28)$$

where

$$\mathcal{A} = \frac{128}{\bar{Q}^2} \sum_{i,j}^M a_0^2(i, j) + \frac{128\bar{P}^2}{\bar{Q}^4} \sum_{i,j}^M a_2^2(i, j) \quad (4.29)$$

We see that  $\mathcal{A}$  is independent of noise and therefore  $Var(G)$  is proportional to noise variance  $\sigma_n^2$ .

Next we consider the left hand side of Eq (4.14). We limit our analysis to the case where the aperture diameter is not changed (i.e.  $D_1 = D_2$ ), but the lens position and/or the focal length are changed (i.e.  $f_1 \neq f_2$  and/or  $s_1 \neq s_2$ ). In this case, the quadratic equation reduces to a linear equation because  $\alpha = 1.0$ . Therefore we get the unique solution:

$$\sigma_2 = \frac{G - \beta^2}{2\beta} \quad (4.30)$$

From Eq (4.7) and the lens formula, we derive an approximate linear expression for focused lens position  $v$ :

$$\begin{aligned} \sigma_2 &= \frac{Ds_0}{2\sqrt{2}} \left[ \frac{1}{f_2} - \frac{1}{u} - \frac{1}{s_2} \right] \\ &= \frac{Ds_0}{2\sqrt{2}} \left[ \frac{1}{v} - \frac{1}{s_2} \right] \\ &\approx \frac{D}{2\sqrt{2}s_0} (s_2 - v) \end{aligned} \quad (4.31)$$

where we assumed  $s_2 v \approx s_0^2$ . This approximation is valid for most camera systems in machine vision.

Now substituting for  $\sigma_2$  using the above equation into Eq. (4.30), we obtain

$$\begin{aligned} v &= -\frac{2\sqrt{2}s_0}{D} \left( \frac{G}{2\beta} - \frac{\beta}{2} - \frac{Ds_2}{2\sqrt{2}s_0} \right) \\ &= \mathcal{B}G + \mathcal{C} \end{aligned} \quad (4.32)$$

where

$$\mathcal{B} = -\frac{\sqrt{2}s_0}{\beta D} \quad \text{and} \quad \mathcal{C} = s_2 + \frac{\sqrt{2}s_0\beta}{D} \quad (4.33)$$

$\mathcal{B}$  and  $\mathcal{C}$  depend only on the camera parameters. Thus, the expected value and the variance of  $v$  are obtained by combining Eqs. (4.28) and

(4.32) as:

$$E\{v\} = 8\mathcal{B}\frac{\overline{P}}{\overline{Q}} + \mathcal{C} \quad \text{and} \quad Var\{v\} = \mathcal{A}\mathcal{B}^2\sigma_n^2 \quad (4.34)$$

The above equations show that the standard deviation of focused lens position  $v$  is linearly related to the standard deviation of noise. Given the noise standard deviation  $\sigma_n$ , the camera parameters, and the defocused images, we can directly compute the RMS error for STM using the above formula. In addition, we can use this equation to select pixels with good signal-to-noise ratio in order to obtain reliable depth estimates.

## 4.5 Experiments

### 4.5.1 Experiments with Synthetic Images

The validity of Eq. (4.34) for the expected value and variance of the focused lens position  $v$  was verified through simulation experiments. The experiments were carried out for two objects, one a step edge and another a cubic polynomial. The blurred images of the test objects were obtained by simulating a camera system similar to the one used in the original implementation of STM reported in [66]. The parameters of the camera system were— focal length 35 mm, F-number 4, and pixel (CCD) size 0.013 mm X 0.013 mm. The distance  $s$  (see Fig. 2.1) between the lens and the image detector was assumed to be varied by a stepper

motor with each step corresponding to a displacement of 0.030 mm. The distance  $s$  and the focused lens position  $v$  are expressed in terms of the step number of the stepper motor with step 0 corresponding a distance of focal length  $f$ . Therefore when the lens was at step 0, objects at infinity were focused on the image detector and at step 90 objects at about 49 cm were focused. For convenience, we specify the distance of objects in terms of the lens step number for which it is in focus.

In the first experiment, we used a step function as an input object. A vertically oriented step edge is defined as

$$f(x, y) = \begin{cases} A & \text{for } x > 0 \\ \frac{A+B}{2} & \text{for } x = 0 \\ B & \text{for } x < 0 \end{cases} \quad (4.35)$$

where  $A$  and  $B$  represent the uniform brightness for two regions. We chose  $A = 176$  and  $B = 80$  and an image of size  $64 \times 64$  in our experiment. The edge is a vertical line located in the middle of the image. For a given distance of this object from the camera, the corresponding blur circle radius  $R$  can be calculated using the camera parameters. Next the blurred image sensed by the camera can be computed by convolving the focused image with the point spread function corresponding to a blur circle radius of  $R$ . In the case of a step edge, as an alternative to numerical convolution, an analytical expression can be obtained for the blurred image and the expression can be evaluated and sampled at



discrete points. This latter approach was followed in our experiments. It can be shown [22] that the defocused image of the step edge defined above will be:

$$g(x, y) = B + (A - B) \left[ \frac{1}{2} + \frac{1}{\pi} \left( \sin^{-1} \left( \frac{x}{R} \right) + \frac{x}{R} \sqrt{1 - \left( \frac{x}{R} \right)^2} \right) \right] \quad \text{for } |x| < R \quad (4.36)$$

where  $R$  is the radius of the blur circle. Two defocused images  $g_1$  and  $g_2$  were computed for two different lens positions of step number 10 and 40 respectively (all other camera parameters were left unchanged). A zero-mean Gaussian random noise was then added to both defocused images  $g_1$  and  $g_2$ .

According to Eq. (4.30) we only need to compute the value of  $\sigma_2$  at one pixel  $(x, y)$  in the image to obtain an estimate of the distance. This pixel should be chosen such that it is blurred in both  $g_1$  and  $g_2$  and the signal to noise ratio is sufficiently high in both  $g_1$  and  $g_2$  as required by Eq. (4.24). Also the images should be such that their Laplacians should be roughly the same in order for the local cubic polynomial model assumption used by STM to be applicable. Further the degree of blur should not be too high in either of the two images (the blur circle diameter should not exceed about 12 pixels) so that the error due to the image overlap problem [66] remains low. In the experiments, when one of the image was blurred too much, then a third image  $g_3$  with a lower level of blur was used in place of the highly blurred image. More specifically, the images  $g_1$  and  $g_2$  were computed for lens position of step 10 and 40 respectively. When the focused position  $v$  predicted by these two images was more than step

50, then  $g_1$  was considered to be blurred too much and therefore a third image  $g_3$  corresponding to lens position at step 70 was computed and used. Fig. 4.2 shows three defocused images for lens positions at steps 10, 40 and 70, respectively when the object focused position was at step 0. The corresponding widths of the blur edge for the three images were 6, 24 and 40 pixels respectively (Fig. 4.4).

We compute  $G$  in Eq. (4.14) from two defocused images, and the camera constant  $\beta$  is computed from a knowledge of the camera parameters. Then an estimate of the focused lens position  $v$  is calculated from Eq. (4.32). This procedure for estimating distance  $v$  is repeated 20 times at a given noise level and camera parameter setting. The experimental mean and standard deviation of  $v$  are calculated from these 20 trials. The theoretical mean and standard deviation of  $v$  is computed based on Eq. (4.34). Fig. 4.1 illustrates the above procedures. Table 4.1, 4.2 and 4.3 show the results of the experimental and theoretical computations at three noise levels for various distances of the object (from focused position of step 0 to step 90). These tables also show for each case the signal-to-noise ratio defined by

$$SNR = 10 \log_{10} \left( \frac{\overline{Q}^2}{\frac{1}{N} \sum_{i=0}^N Q_i'^2} \right) \quad (4.37)$$

where  $\overline{Q}$  and  $Q'_i$  are as defined in Eqs (4.21) and (4.22). A comparison of results in the tables shows that the experimental and theoretical results are in close agreement thus verifying the theory. In particular, as predicted by theory, we see that the standard deviation of  $v$  (i.e. the RMS error of focused lens position) is linearly related to noise standard

deviation.

In the second experiment, an object having an image brightness corresponding to a cubic polynomial was used. The procedure was similar to the first experiment. The coefficients of the cubic polynomial were taken to be those that modeled the step edge in the first experiment. The coefficients were obtained by using a least squares error fitting technique [49]. The resulting image was

$$f(x, y) = a_3x^3 + a_2x^2 + a_1x + a_0 \quad (4.38)$$

where  $a_3 = -0.003$ ,  $a_2 = 0.313$ ,  $a_1 = -5.957$  and  $a_0 = 101.031$ . In this case, by convolving  $f(x, y)$  with the PSF of the camera corresponding to a blur circle of radius  $R$ , it can be shown that the blurred image is given by

$$g(x, y) = a_3x^3 + a_2x^2 + \left( \frac{3R^2}{4}a_3 + a_1 \right)x + \frac{R^2}{4}a_2 + a_0 \quad (4.39)$$

The blurred images  $g_1$  and  $g_2$  are computed by discrete sampling of the above function for two different blur circle radii  $R_1$  and  $R_2$  respectively. These are then filtered with  $L_0$  and  $L_2$  filters to obtain smoothed images and their Laplacians. In this experiment, unlike the previous one, almost any pixel can be selected for computing  $v$ , and a third image is not needed. The results are good even for highly blurred cases. However, as mentioned before, we use the third image just to reduce the variance of  $v$ . Three sampled images computed for lens positions at step 10, 40 and 70, when the object is focused at step 90, are shown in Fig. 4.3. The blur circle radii corresponding to these images are respectively 22, 13 and 5 pixels respectively. Fig. 4.5 shows their gray-level profile. Table 4.4, 4.5

and 4.6 show the experimental and theoretical results. Once again, we see that the results are in a good agreement.

### 4.5.2 Experiments on Real Images

Images of real scenes were taken by using our SPARCS camera system [81]. The camera parameters were set to the same as those synthetic images used in the previous experiments. The defocused images of the object obtained by moving the lens causes the variations in image brightness and in image magnification, even though nothing has changed in the scene. Therefore, normalization with respect to image brightness was carried out by dividing the image brightness at every point by the mean brightness of the image. Also, only those pixels in the central region of images resulting in the least magnification change were chosen for the experiments. Detail magnification correction can be found [82, 71].

For each of the two focus settings, 20 images were averaged to get one image with high signal-to-noise ratio. The noise of camera system was estimated by using a flat and uniformly bright object. The gray level variance of this image was computed for the standard deviation of the system noise. In the case of 500 LUX illumination, the standard deviation of the system noise was found about 1.6. The above two averaged defocused images, an estimated standard deviation of the system noise and camera parameters were provided for the theoretical calculation in Eqs. (4.34). Those recorded 20 pairs of images also were used to compute

20 focus positions by Eq. (4.32). The mean and standard deviation of these 20 trials give the experimental results. Figure 4.6 shows the focus images of the test object A and B both taken at step position 50. Again, a third image was used in place of one of the images blurred too much. Unlike the previous experiment, three blurred images used for computing focus position were taken at step position 15, 45 and 75. Tables 4.7 and 4.8 show the experimental and theoretical results. The experiments conducted on these real images also verify our theory.

## 4.6 Conclusion

Eq. (4.34) provides a method for estimating the uncertainty in the focused position  $v$  for an object as a function of the camera parameters, noise level, and image signal. This can be used to select those pixels that yield reliable estimates of depth of objects and ignore the unreliable pixels in the application of STM. As expected, the reliability of pixels increases with increasing value of image Laplacian at those pixels but decreases with increase in noise standard deviation. If an object is planar, or has a known shape form (e.g. spherical), then Eq. (4.34) can be used to combine the depth information provided by different pixels in some optimal manner to infer the actual shape parameters of the object.

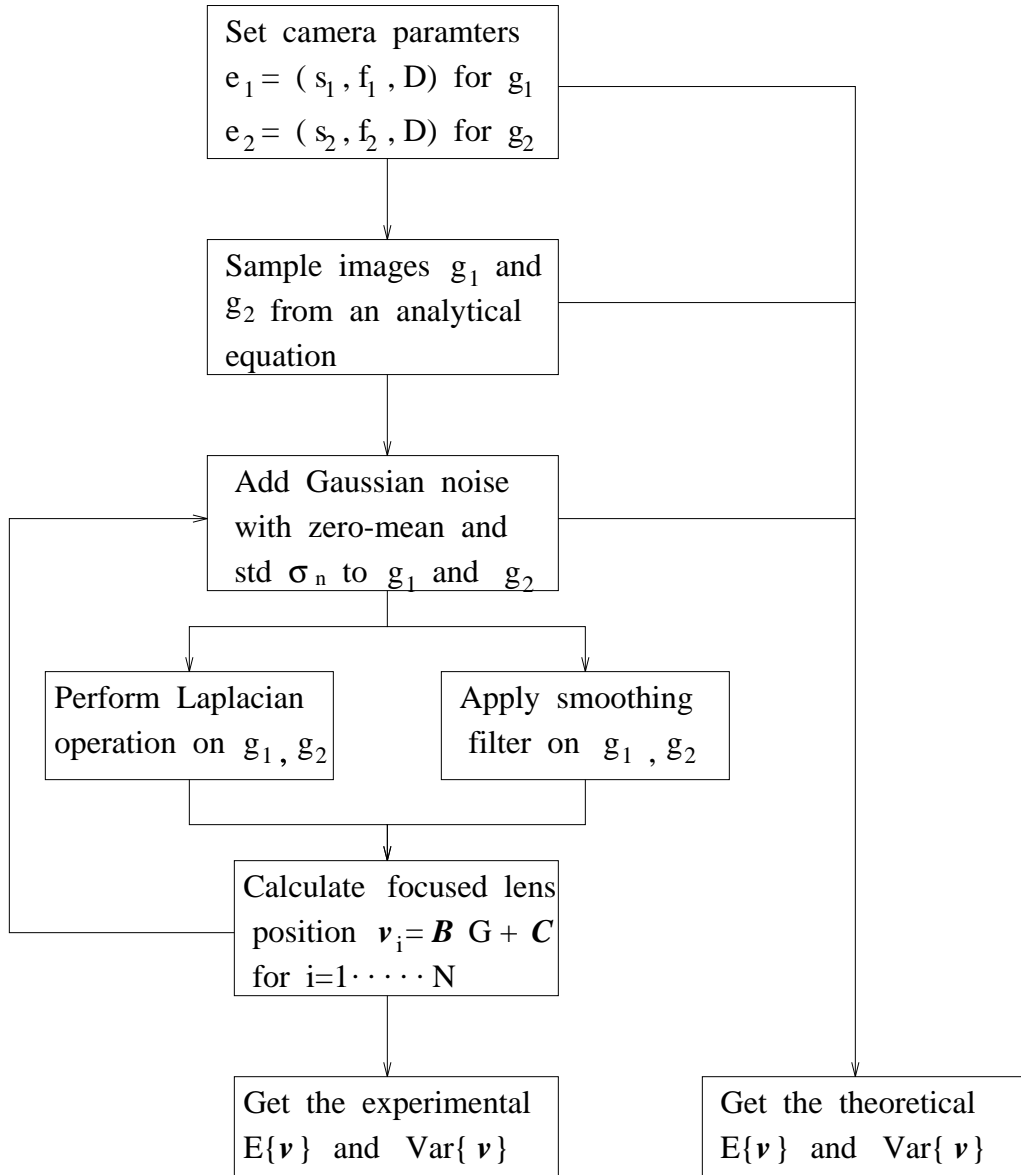


Figure 4.1: The flow chart for experimental verification



Figure 4.2: Step edge focused at step 0 and defocused images at lens steps 10,40,70

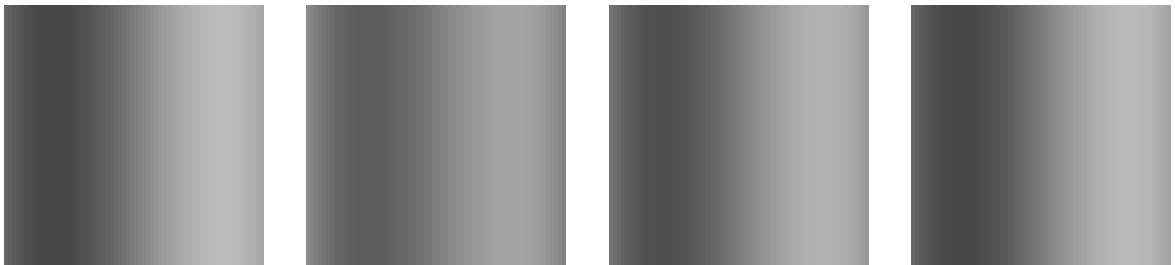


Figure 4.3: Cubic polynomial focused at step 90 and defocused images at lens steps 10,40,70

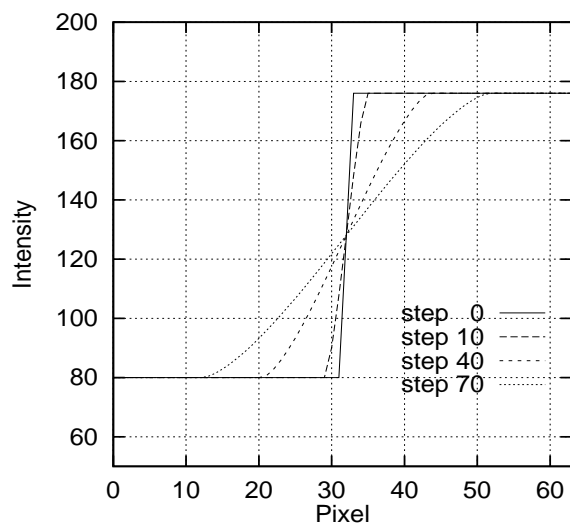


Figure 4.4: Gray-level in edge image

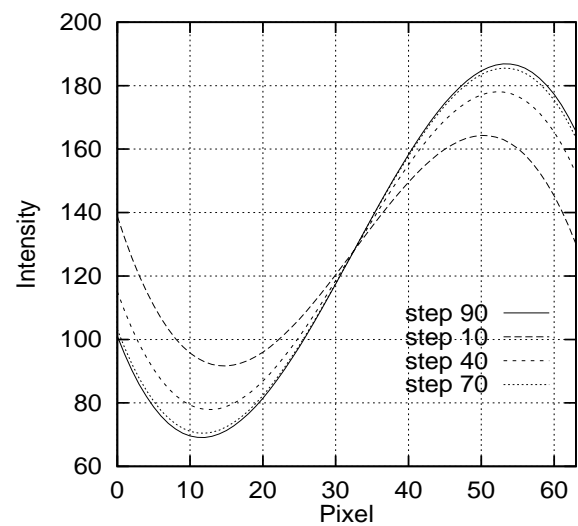


Figure 4.5: Gray-level in polynomial image

Object Distance	Lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
0	28.47	-0.80	-0.80	0.00	0.78	0.98	-0.20
10	33.54	9.25	9.24	0.01	0.34	0.48	-0.14
20	33.29	19.72	19.74	-0.02	0.41	0.52	-0.11
30	34.35	30.21	30.15	0.06	0.29	0.32	-0.03
40	30.82	40.49	40.48	0.01	0.42	0.48	-0.06
50	31.84	51.09	51.08	0.01	0.41	0.35	0.06
60	32.12	60.16	60.11	0.05	0.30	0.25	0.05
70	29.49	69.08	69.12	-0.04	0.41	0.41	0.00
80	28.78	79.78	79.77	0.01	0.57	0.65	-0.08
90	29.94	89.05	89.31	-0.26	0.65	0.87	-0.22

Table 4.1: Step edge with noise std. 1.0

Object Distance	Lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
0	22.45	-0.80	-0.89	0.09	1.56	2.03	-0.47
10	27.52	9.25	9.21	0.04	0.68	0.99	-0.31
20	27.28	19.72	19.76	-0.04	0.83	1.05	-0.22
30	28.33	30.21	30.10	0.11	0.58	0.64	-0.06
40	24.80	40.49	40.49	0.00	0.84	0.98	-0.14
50	25.82	51.09	51.05	0.04	0.82	0.77	0.05
60	26.10	60.16	60.07	0.09	0.61	0.58	0.03
70	23.47	69.08	69.18	-0.10	0.82	0.81	0.01
80	22.76	79.78	79.92	-0.14	1.14	1.29	-0.15
90	23.92	89.05	89.64	-0.59	1.29	1.69	-0.40

Table 4.2: Step edge with noise std. 2.0

Object Distance	lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
0	18.93	-0.80	-1.07	0.27	2.34	3.05	-0.71
10	23.99	9.25	9.16	0.09	1.02	1.52	-0.50
20	23.75	19.72	19.76	-0.04	1.24	1.60	-0.36
30	24.81	30.21	30.04	0.17	0.88	0.97	-0.09
40	21.28	40.49	40.52	-0.03	1.26	1.50	-0.24
50	22.29	51.09	51.02	0.07	1.23	1.02	0.21
60	22.57	60.16	60.02	0.14	0.92	0.75	0.17
70	19.94	69.08	69.25	-0.17	1.24	1.24	0.00
80	19.24	79.78	80.34	-0.56	1.72	1.98	-0.26
90	20.39	89.05	90.04	-0.99	1.94	2.75	-0.81

Table 4.3: Step edge with noise std. 3.0



Object Distance	Lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
0	30.98	-0.98	-1.09	0.11	0.61	0.74	-0.13
10	31.84	9.62	9.52	0.10	0.49	0.51	-0.02
20	30.79	20.05	20.09	-0.04	0.42	0.42	0.00
30	29.44	30.29	30.43	-0.14	0.43	0.48	-0.05
40	29.56	40.38	40.54	-0.16	0.49	0.56	-0.07
50	30.42	50.30	50.36	-0.06	0.60	0.64	-0.04
60	32.11	60.04	60.10	-0.06	0.44	0.44	0.00
70	29.04	69.64	69.71	-0.07	0.50	0.51	-0.01
80	27.31	79.08	79.22	-0.14	0.60	0.71	-0.11
90	28.88	89.37	89.63	-0.26	0.72	0.91	-0.19

Table 4.4: Cubic polynomial with noise std. 0.3

Object Distance	Lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
0	24.96	-0.98	-1.24	0.26	1.22	1.49	-0.27
10	25.82	9.62	9.39	0.23	0.99	1.04	-0.05
20	24.77	20.05	20.15	-0.10	0.85	0.84	0.01
30	23.42	30.29	30.59	-0.30	0.86	0.95	-0.09
40	23.53	40.38	40.75	-0.37	0.99	1.14	-0.15
50	24.39	50.30	50.48	-0.18	1.21	1.31	-0.10
60	26.09	60.04	60.18	-0.14	0.89	0.89	-0.00
70	23.02	69.64	69.81	-0.17	1.01	1.03	-0.02
80	21.28	79.08	79.44	-0.36	1.21	1.41	-0.20
90	22.86	89.37	89.73	-0.36	1.44	1.79	-0.35

Table 4.5: Cubic polynomial with noise std. 0.6

Object Distance	Lap. img SNR (db)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
0	21.44	-0.98	-1.44	0.46	1.83	2.21	-0.38
10	22.30	9.62	9.24	0.38	1.48	1.59	-0.11
20	21.25	20.05	20.19	-0.14	1.27	1.27	0.00
30	19.89	30.29	30.77	-0.48	1.28	1.48	-0.20
40	20.01	40.38	40.98	-0.60	1.48	1.75	-0.27
50	20.87	50.30	50.65	-0.35	1.81	2.00	-0.19
60	22.56	60.04	60.26	-0.22	1.34	1.34	-0.00
70	19.49	69.64	69.95	-0.31	1.51	1.57	-0.06
80	17.76	79.08	79.73	-0.65	1.81	2.14	-0.33
90	19.33	89.37	90.07	-0.70	2.17	2.59	-0.42

Table 4.6: Cubic polynomial with noise std. 0.9

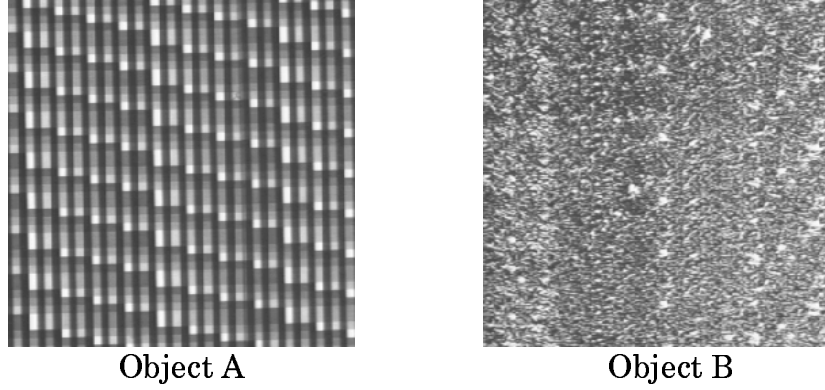


Figure 4.6: Test images taken at step 50

Object Distance	Lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
10	20.46	8.99	8.43	0.56	2.20	2.94	-0.74
20	30.99	19.88	20.20	-0.32	1.52	1.80	-0.28
30	33.85	30.15	30.26	-0.11	0.65	0.82	-0.17
40	32.95	40.32	40.72	-0.40	0.96	1.32	-0.36
50	28.21	49.41	49.20	0.21	0.78	1.04	-0.26
60	36.46	61.04	61.52	-0.48	0.49	0.68	-0.19
70	34.89	69.98	69.64	0.34	0.85	1.48	-0.63
80	22.52	78.39	78.99	-0.60	1.77	2.60	-0.83
90	17.02	87.45	87.89	-0.44	2.60	3.77	-1.17

Table 4.7: The object A with noise std. 1.6

Object Distance	Lap. img SNR (dB)	Theo. $E\{v\}$	Exp. $E\{v\}$	Diff $\delta$	Theo. $Std\{v\}$	Exp. $Std\{v\}$	Diff $\delta$
10	23.74	9.31	9.59	-0.28	1.62	2.07	-0.45
20	27.12	20.36	20.52	-0.16	0.70	1.02	-0.32
30	31.80	29.85	29.75	0.10	0.38	0.48	-0.10
40	39.49	40.62	40.42	0.20	0.49	0.76	-0.27
50	37.51	50.25	50.30	-0.05	0.44	0.64	-0.20
60	32.41	60.21	60.03	0.18	0.31	0.44	-0.13
70	35.33	69.93	69.67	0.26	0.63	0.97	-0.34
80	23.70	78.68	78.88	-0.20	1.06	1.71	-0.65
90	18.16	88.53	88.92	-0.39	1.82	2.78	-0.96

Table 4.8: The object B with noise std. 1.6

## **Chapter 5**

# **Integration of Depth from Focus and Defocus**

## **5.1 Introduction**

In this chapter, we present a method for integrating DFD and DFF proposed in [71]. Depth from Focus (DFF) and Depth from Defocus (DFD) each has some advantages and disadvantages. The integration of DFF and DFD can provide an useful technique for fast and accurate autofocusing and three-dimensional (3D) shape recovery of objects.

As seen in Chapter 2, DFF methods process a large sequence of image frames of a 3D scene in order to reconstruct a focused image and find the depth map of an object. DFD methods require processing only a few images (about 2-3) which is less computational than DFF methods. In addition, only a few images are sufficient to determine the distance

of all objects in a scene using the DFD methods, irrespective of whether the objects are focused or not. The two main disadvantages of the DFD methods are (i) they require accurate camera calibration for the camera characteristics (a blur parameter as a function of camera parameters), and (ii) they are less accurate than DFF methods.

Therefore, an DFD method can be combined with an DFF method to reduce the number of images acquired and processed but attain the same accuracy as DFF. First DFD method is used to obtain a rough depth-map. This requires acquiring and processing only 2 or 3 images. Then DFF is applied to a short sequence of images which are acquired with camera parameters so that only objects in the rough depth-map range estimated by DFD are focused. Acquiring and processing of image frames that correspond to focusing objects that are far away from the depth-map estimated by DFD is avoided. This saves image acquisition and processing time of unnecessary image frames. The accuracy of the depth-map obtained will be the same as that of DFF. In comparison with a bare DFF, the combined DFD-DFF may save much time in the best case when all objects in a scene are at the same distance, but in the worst case when objects in scene are at all possible distances, the combined DFD-DFF will take a bit more time than the bare DFF. In a typical application, the combined method can be expected to save some modest time.

The technique described in this chapter is implemented on a camera system named Stonybrook Vision System (SVIS). Methods for calibrating



Figure 5.1: Stonybrook Vision System (SVIS)

the camera system and results of experiments on SVIS are presented.

## 5.2 Camera System

The integration of DFD and DFF was implemented on a camera system named Stonybrook Vision System or SVIS (see Fig. 5.1). SVIS is a vision system built over the last 1 year in the Computer Vision Laboratory, State University of New York at Stony Brook. SVIS consists of a digital still camera (DELTIS VC 1000 of Olympus Co.). S-Video signal from the camera is digitized by a frame grabber board (Matrox Meteor Standard board). All processing is done on a PC (Intel Pentium, 200 MHz.). The camera is mounted on a linear motion stage driven by a stepper motor (X-9 stage and MD-2 stepper motor system of Arrick Robotics Inc.). This stage provides the capability to develop a multiple-baseline

stereo system using only a single lens configuration. The stepper motor that moves the camera is controlled through a parallel printer port on the PC. Focusing of the camera is done by a motorized lens system inside the camera. The lens motor is controlled from a serial (RS-232) communication port on the PC. The Matrox frame grabber installed in the PC is used to record  $640 \times 480$  size monochrome images with 8 bits/pixel. A user friendly windows software interface has been developed to control the whole system under MS Windows 95 OS. It includes convenient controls for manipulating (i) the lens system, (ii) digitizer board, (iii) linear motion stage on which the camera is mounted, and (iv) all the application programs. In addition, an overhead projector is used to project a high contrast pattern onto 3D objects that have low contrast.

The camera lens system has separate controls for zooming and focusing. Zooming can be varied from a focal length of 10.2 mm (WIDE mode) to 19.6 mm. (TELE mode). The experimental results reported in this chapter were carried out with a zoom focal length of 19.6 mm (TELE mode). Focusing is done by driving a stepper motor that controls lens position with respect to the image sensing CCD in the camera. The stepper motor has step positions ranging from 0 to 170. Objects at infinity are focused when the lens stepper motor is at step number 0, and objects close by (about 25 cm.) are focused when the lens position is at step 170. Since each lens step number corresponds to focusing objects at some unique distance, we often use this corresponding step number to specify the distance of objects. If an object is said to be at a distance of step X, it

means that the distance of the object is such that the object would be in best focus if the lens is moved to step number  $X$ . Specifying distance of objects in terms of lens step numbers is particularly convenient in DFF and DFD. In SVIS, each step corresponding to a displacement of lens is about 15  $\mu\text{m}$ .

### 5.3 Camera Calibration

The internal parameters of camera system such as focal length, distance between lens and image detector, aperture diameter etc. were not known accurately. Therefore SVIS had to be calibrated with respect to several important factors. The first factor needing calibration was the relation between the distance of objects from the camera and the corresponding lens step number at which the objects would be in best focus. The second factor needing calibration was a blur parameter needed for image defocus analysis. In addition, two thresholds  $T_1$  and  $T_2$  were calibrated for DFF and DFD respectively to identify and ignore regions of low contrast with respect to noise (i.e. low signal-to-noise ratios). Finally, some considerations related camera calibration are discussed. Calibration procedures with respect to each of these factors were carried out as follows.

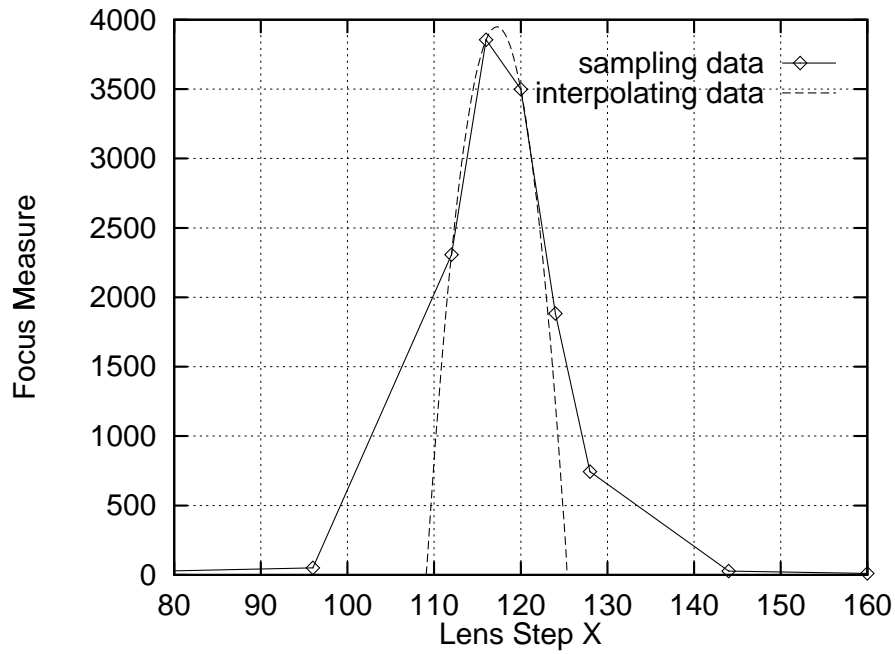


Figure 5.2: Autofocusing Experiment

### 5.3.1 Lens Step vs Focusing Distance

The lens position in step number and the corresponding focused distance was obtained using an autofocus algorithm as follows. A large planar high contrast object was placed normal to the camera's optical axis at a known distance from the camera. The camera was then autofocused by maximizing a focus measure. The lens position in step number that resulted in a maximum focus measure was found by a binary search type of algorithm (see also details in Chapter 3).

In order to have this search algorithm more robust and faster, we began the sequential search by moving the lens from lens position either 0 or 175 in every 32 steps incremented/decremented. It was stopped until the focus measure decreased its value more than 20% of previous one. Then a binary search was begun in the interval of 32 steps after



sequential search. After the search intervals was narrowed down to 4 steps, the quadratic interpolation was applied to three points with 4 steps apart around the maximum to obtain the location of actual peak of focus measure. In this way, it was proved to reduce the required image frames as well as avoid the false search due to noise. As the focused step position was found, the distance of the object from the camera was recorded. Fig. 5.2 shows some sample positions used in this search algorithm where the object was found at focused position 117. This procedure was repeated for many different distances of the object corresponding to roughly 5 step intervals for the focused lens position.

Several objects used for this experiment are shown in Fig. 5.5a to 5.5e. Each image used was the central 128x128 region. The focus measure was the sum of square of Laplacian of image grey-level. The relationship between the reciprocal of object distance  $1/u$  versus the step number  $X$  is almost linear and can be expressed as

$$\frac{1}{u} = aX + b \quad (5.1)$$

where  $a = 0.049$  and  $b = -4.234$  if we fit those data for focal length at TELE mode end. If at WIDE mode end, the constant  $b$  was found -4.108. This calibration data was used in finding the distance of object points given the focused lens step number for the object points (see Fig. 5.3). The experimental results only recorded the ranging distance from step 90 to step 170.

### 5.3.2 Blur Parameter vs Focused Lens Step

The DFD used in our implementation is based on STM mentioned in the previous chapter. In DFD, only one camera parameter, the lens position (step number) was varied in acquiring the two needed images. All other parameters (focal length and aperture diameter) were nearly constant. In this case ( $\alpha = 1$  in Eq. 4.14) we find that a blur parameter  $\sigma_2$  (which is proportional to the diameter of blur circle) is linearly related to a quantity  $G'$  that can be computed from the two recorded images by:

$$\sigma_2 = \frac{G' - \beta^2}{2\beta} \quad (5.2)$$

Note that  $G'$  is considered here as the integrated version of STM rather than the version in Eq. 4.30 [66]. The camera constant  $\beta$  in the above equation is a function of the two camera parameter settings at which the two images are recorded. It can be computed if the camera parameters are known. Since they were not known, it was determined experimentally as follows.

The DFD[66] was implemented with two images recorded at lens positions of step 120 and 155. An object was placed at such a distance that it was focused when the lens was at step 120. In this case the blur parameter for the image recorded at step 120 is zero, but the blur parameter  $\sigma_2$  for the image recorded at step 155 is  $-\beta$ . Therefore  $\beta$  is obtained directly by computing the square root of  $-G'$  (a quantity that can be computed from the two observed images) and the sign of  $\beta$  is negative (Note that in this case the sign of  $G'$  will be negative). This

method yielded  $\beta = -2.0$  for focal length at WIDE mode end and  $\beta = -4.0$  at TELE mode end.  $\beta$  can also be computed by placing an object at a distance corresponding to the focused lens position 155. In this case, the blur parameter for the second image is zero, but that for the first image is equal to  $\beta$ . Therefore  $\beta$  can be estimated as the negative value of the square root of  $G'$ .

Another calibration table is needed that relates the blur parameter of an object in the second image (recorded at step 155) to the focused lens step number of the object. This table is created as follows. First an object is placed at a given distance and the DFF method is used to autofocus the object by a binary search for the maxima of a focus measure. The lens step number that autofocuses the object is recorded. Then the DFD method is applied and the blur parameter  $\sigma_2$  is calculated. This procedure was repeated for several different objects at the same distance and the average  $\sigma_2$  and the average focus step number were recorded. This gives one entry of the calibration table. This procedure was repeated for several different object distances at roughly regular intervals in terms of focused lens step number (about 5). The gap between each entry in the table were filled by linear interpolation with respect to lens step number and blur parameter. The resulting data is shown in Fig. 5.4.

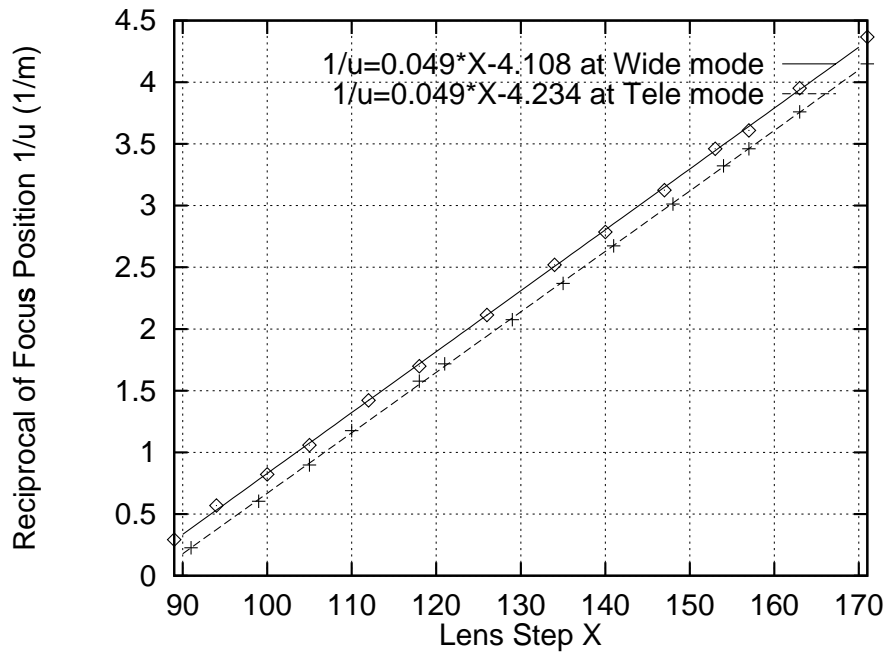


Figure 5.3: Lens Step vs 1/Focusing Distance

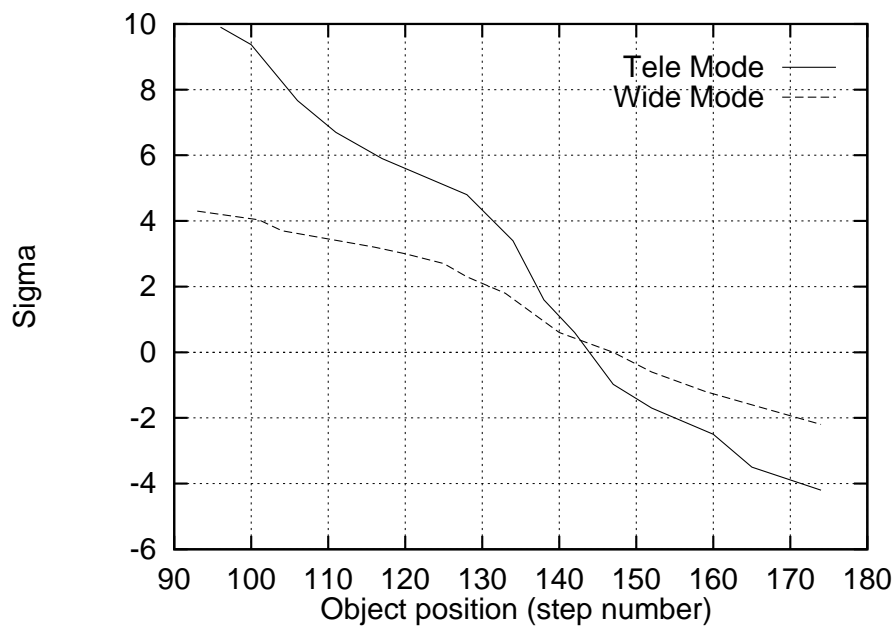


Figure 5.4: Blur Parameter vs Focused Lens Step

### 5.3.3 Thresholds vs System Noise

Both of DFF and DFD methods compute the depth based on a small image window. When a image window does not have enough contrast information, those methods are susceptible to noise leading to erroneous results. There are two thresholds  $T_1$  and  $T_2$  required for focus measure in DFF and Laplacian image in DFD, respectively.

In order to estimate those two values, we placed an planar object of near constant brightness (see Fig. 5.5f) at different focused distances from the camera. The illumination was controlled about 500 Lux. The focus measure operator, the sum of square of the Laplacian, was applied on this image within many 16x16 non overlapping regions. The mean of these regions was found to be 6.28 and the standard deviation from this mean value was 0.45. Based on the confidence level of the Laplacian focus measure defined in Chapter 2, a threshold of this mean plus three times its standard deviation was chosen for reliable results (see Eq. 2.35). Therefore, we obtained

$$\frac{1}{N_1^2} \sum^{N_1} (\nabla g_i)^2 \geq T_1 \quad (5.3)$$

where  $T_1 = 8.63$  used for each image  $g_i$  in the region  $N_1 \times N_1$  for one depth estimate. Any focus measure below this value, the depth for that point belongs to background.

In Chapter 3, we have given the expression for computing variance

of focus measures. It can be seen that, by using a larger window size,  $T_1$  can be lowered resulting in more reliable depth estimates. However, a larger window size will reduce the resolution of the depth-map. Hence there is a trade off between window size for focus measure computation and the resolution of the depth map.

The second threshold  $T_2$  is estimated for the two Laplacian images used in STM. In Chapter 4, we have verified that the reliable depth estimates depend on the high signal-to-noise ratios of the Laplacian images. In this case, we took two blurred images  $g_1$  and  $g_2$  of low contrast object at lens position 120 and 150. The mean value computed from many 9x9 regions of sum of square of two Laplacian images was found 3.52 and the standard deviation from this mean value was 4.18. The same definition as  $T_1$  which is the mean plus three times the standard deviation, we obtained

$$\frac{1}{N_2^2} \sum^{N_2} (\nabla g_1 + \nabla g_2)^2 \geq T_2 \quad (5.4)$$

where  $T_2 = 16.1$ . The depth is not countable if the computed value is less than  $T_2$ . Note that those thresholds were computed as focal length set at WIDE mode end which usually are higher results than at TELE mode end. We picked those higher values for using in any condition of focal length as well as for both focused and defocused images.

### 5.3.4 Other Considerations

The image magnification variation due to change in camera parameter settings for autofocus is another important factor needing calibration. This is because correspondence between different image regions is needed between image frames recorded with different lens positions. This facilitates comparison of focus measures computed in different image frames to find the image frame in which a given image region is in best focus. Several researchers have investigated the effects of magnification change and proposed some solutions such as image warping or telecentric optics [82, 83, 9, 79]. In our experiments, the approach of detailed calibration to constant magnification image is described in [71], which is based on spatial interpolation and resampling techniques.

If SVIS is used for stereo vision, calibration is needed to establish a relation between object distance and the corresponding stereo disparity. A careful calibration procedure for stereo ranging can also be found in some literatures [19, 71, 29, 43].

## 5.4 The Integration of DFF and DFD

The integration of DFF and DFD for rapid passive autofocus and 3D shape recovery were implemented on the SVIS. The following describes the details of this integration and shows some experimental results.

### 5.4.1 Autofocusing

In the chapter 3 we proposed the autofocusing algorithm based on searching maximum focus measure among the image sequence. The search procedure combines sequential search, binary search and quadratic interpolation to minimize the lens motion and the computation time. In general, this algorithm purely relies on image focus analysis (i.e. DFF) and needs to process about 12 to 18 image frames in order to accomplish autofocusing for a object.

The following an alternative autofocusing method by combining with DFD algorithm is implemented. First DFD method is used to obtain an estimated distance. This requires to process only 2 images. Then the DFF autofocusing algorithm using binary search and quadratic interpolation is applied within  $\pm 10$  steps of the estimated position rather than searching the whole lens position. This save image acquisition and processing time of unnecessary image frames that correspond to focusing objects far away from the estimated distance. The accuracy of ranging obtained will be the same as that of DFF, but the required image frames reduce to about 5 to 8.

The autofocusing experiments using DFD, DFF and a combined method (i.e. DFD+DFF) were conducted on the 5 test objects (see Fig. 5.5 a-e). Only the center region of image is processed and the size of region is 48x48. The object illumination was controlled at about 400



Lux. The focal length of camera was set at condition of TELE mode end. In addition, a look-up table associated with focal length at TELE mode end for DFD is used. The experimental results are tabulated in Table 5.1. The first column of the table shows object distance. Columns 2 to 4 correspond to the results of DFD, DFF and DFD+DFF. Each column with two subcolumns in the table shows the mean focus position and the standard deviation in step number for 5 test objects. The last two rows show the overall RMS error and required image number for each method, respectively. Note that we use the results from DFF as a benchmark. The calculation of the standard deviations for DFD and DFD+DFF were respect to the mean value of DFF rather than the mean value of themselves. The RMS error for DFD is 3.14 steps out of about 70 steps. This corresponds to about 4.48% RMS error in the lens position for autofocus-ing. The RMS error for DFF and DFD+DFF methods are 0.66 and 0.94 steps out of 70 steps which correspond to about 0.94% and 1.34% RMS errors in autofocus-ing, respectively. As expected, The combined method maintains the similar accuracy to DFF but requires less image frames. It is noted that the entry of DFD in the table has poorer performance at far from camera position, which because two images acquired at lens positions of step 120 and 155 become highly blurred. In this case the results of ranging were unreliable so that DFF needs to search at least within  $\pm 10$  steps of the estimated position in order to tolerate this uncertainty. In fact, if the focused lens position of the object is between at step 120 and at step 155, we found the combined method can further

reduce the required images to only 5. That is, 2 defocused images using in STM plus 3 images (with 4 or 5 steps apart) for quadratic interpolation near that estimated position. In this way the refined result is still able to maintain similar accuracy as the other.

### 5.4.2 3D Shape Recovery

First a rough depth-map is obtained using DFD[66]. One estimate of depth is obtained in each image region of size  $48 \times 48$ . The two needed images were recorded at lens steps 120 and 155. Four image frames were time-averaged to reduce noise. The magnification of the images were normalized using the magnification table. The DFD was applied to images of size  $432 \times 432$ . Thus a coarse depth-map array of size  $9 \times 9$  was obtained. At this stage, the depths were expressed in terms of the lens step number that focuses objects at that depth. In each image region, the actual depth-map at higher resolutions was assumed to be within  $\pm 10$  steps of the estimated depth-map step number. Using this initial depth-map, a higher resolution depth-map of size  $27 \times 27$  (one estimate in  $16 \times 16$  image region) was obtained using DFF. The lens step numbers for which image frames needed to be recorded and processed in DFF were determined using the following algorithm. The purpose of this algorithm is to acquire and process only those image frames near the estimated depth-map values. This avoids processing unnecessary image frames in which all image regions are highly blurred.

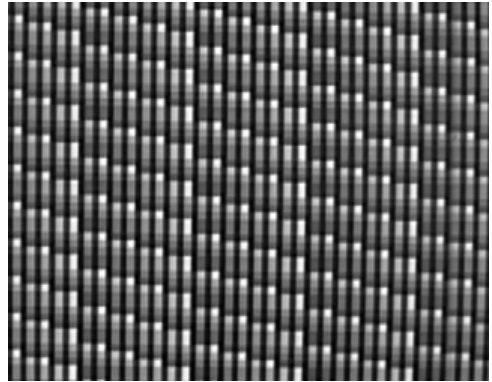
The estimated depth values (in step numbers) is first quantized to multiples of DEL (=1) steps. Then, for each quantized value  $s$  that occurs in the depth-map,  $2\delta + 1$  ( $\delta = 1$ ) lens step positions at  $s + i * DEL$  for  $i = 0, \pm 1, \pm 2, \dots, \pm numdel$  are marked. If any of these steps are outside the range of minimum and maximum step positions, they are discarded. Then image frames are recorded at each marked lens position. All images are normalized with respect to the magnification corresponding to the step position at which they are recorded. Then, in the resulting image sequence, focus measures are computed in image regions of size  $16 \times 16$ . The step number where the focus measure is a maximum in each image region is determined. These image regions with maximum focus measures are synthesized to obtain a focused image of the entire scene. Further, the maximum focus measure and the two focus measures in the preceding image frame (DEL steps below) and the succeeding image frame (DEL steps above) are taken. A local quadratic curve is fitted to the three focus measures (the center one being the maximum) and the position of the maximum of the curve is computed. This position is taken as an improved estimate of the depth-map. If the focus measure for the preceding or succeeding image frame is not available, then this last step is not performed. From the previous experiments, DFF has roughly 1% RMS errors. If the depth map is filtered with an median filter, we found that the depth error can be reduced by a factor of 1/3.

Experiments were conducted on three of the objects— prism, cone and inclined plane, and the results are presented here. The program

parameters were set to operate well for objects in the depth range of 30cm. to 80 cm. from the camera. In the experiments, those objects were placed about 60 cm distance from the camera. In general, the magnification correction data obtained by calibration was sufficiently accurate for DFF, but not for DFD. The results of DFD was satisfactory at the center of the image where the effect of magnification change was minimum, but the results were not satisfactory away from the center. This was compensated by allowing for error in DFD and applying DFF in a larger range than necessary, thus increasing the number of images and computing used. It should be noted that shape of objects is obtained by looking for changes in scene depth-map. Therefore the percentage error in depth-map will be much less than that in shape. The focused image and 3D shape of a prism recovered from DFD/DFF are shown in Figs. 5.6 and 5.7 as well as the results of a cone object are shown in Figs. 5.8 and 5.9. For inclined plane object, there are many regions with almost constant intensity in the image. The reliable depth map is sparse and lots of points belong to background after thresholding. If a high contrast pattern is projected on the object, we obtain a much more dense and reliable depth map. Results for the inclined plane object are shown in Figs. 5.10 to 5.13. In general, each object takes about 40 seconds on the Pentium PC for completing shape recovery tasks.



a



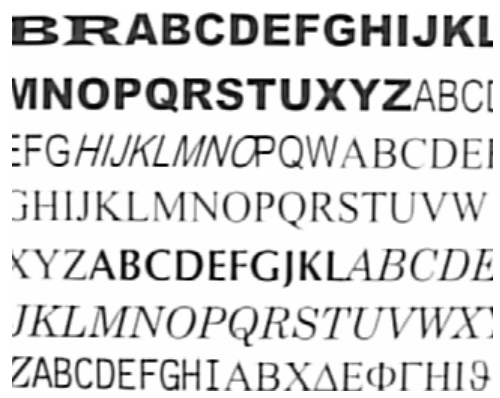
d



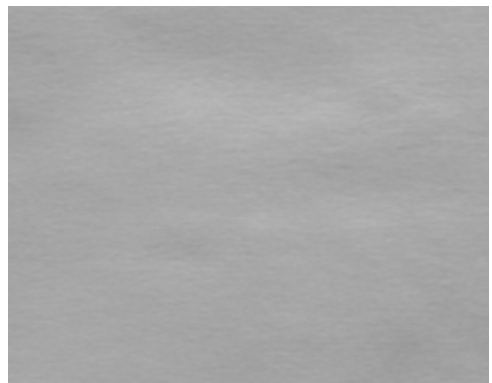
b



e



c



f

Figure 5.5: Test Objects for Experiments

Object Distance (cm)	DFD		DFF		DFD+DFF	
	mean	std.	mean	std.	mean	std.
26.3	162.00	3.55	165.00	0.63	164.60	0.89
30.8	152.00	1.66	152.00	0.50	151.80	0.58
37.4	141.20	1.48	141.00	1.09	141.00	0.89
49.2	129.20	1.82	129.80	0.75	129.60	0.75
62.8	120.00	2.78	118.20	0.40	118.00	0.66
85.3	107.60	4.15	110.60	0.49	110.80	0.77
130.6	106.40	4.27	103.00	0.00	103.20	0.45
270.5	95.19	3.85	94.40	0.80	95.80	1.87
RMS Error	3.14		0.66		0.94	
Acquired Images	2		12-18		5-8	

Table 5.1: Autofocusing results. The correct mean was taken to be the mean of DFF when computing standard deviation for all three methods.

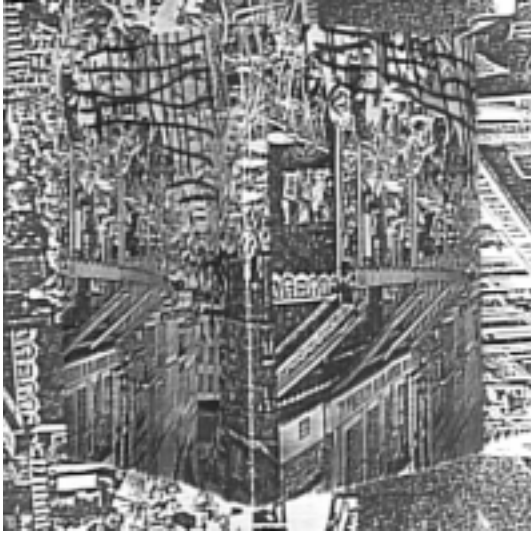


Figure 5.6: Reconstructed Focused Image of a Prism Object

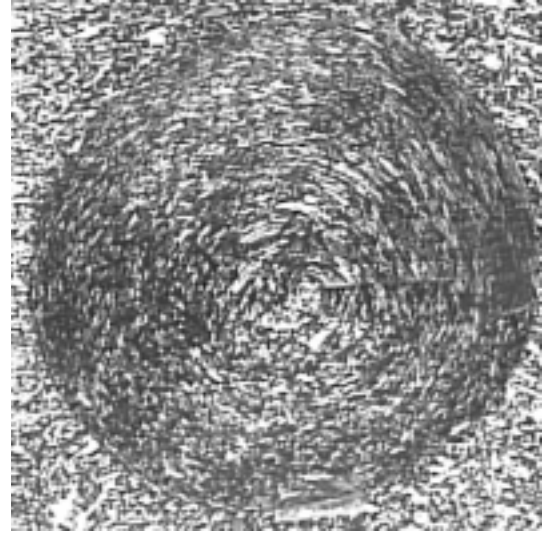


Figure 5.8: Reconstructed Focused Image of a Cone Object

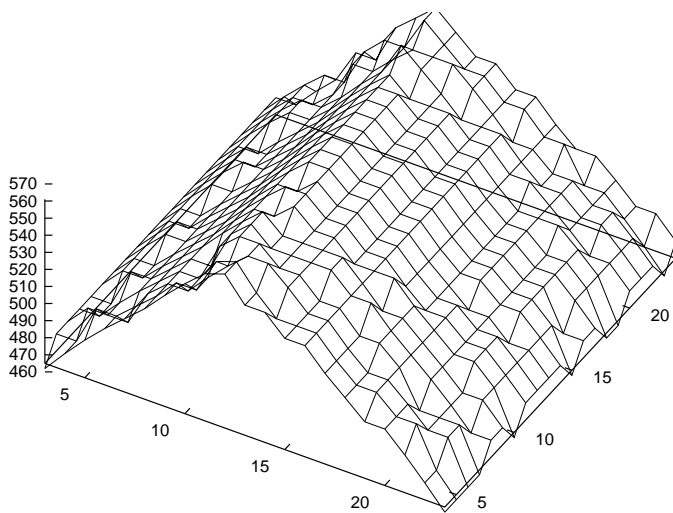


Figure 5.7: Depth Map of the Prism Object

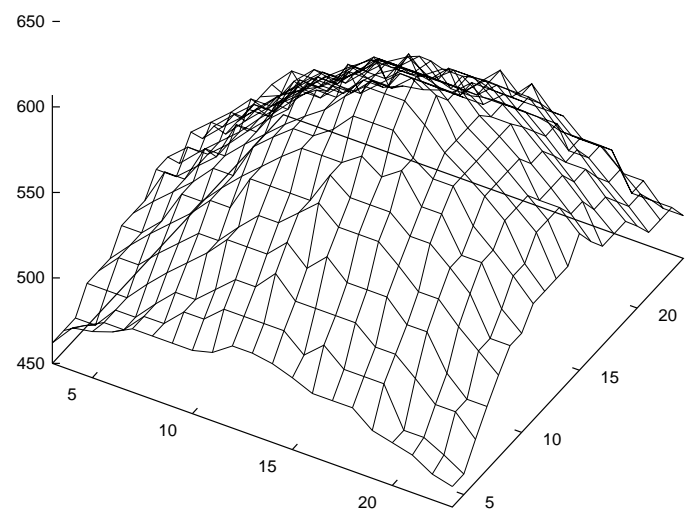


Figure 5.9: Depth Map of the Cone Object

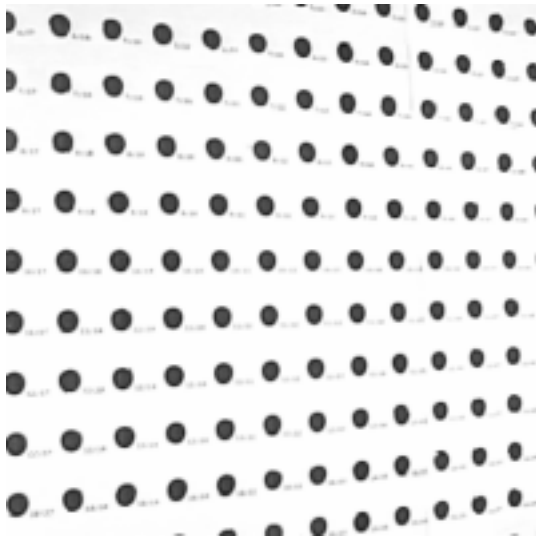


Figure 5.10: An Inclined Plane without Projected Pattern

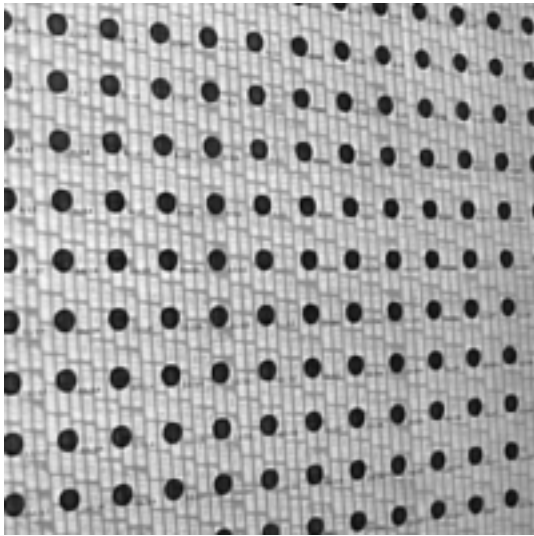


Figure 5.12: Reconstructed Focused Image of an Inclined Plane (with Projected Pattern)

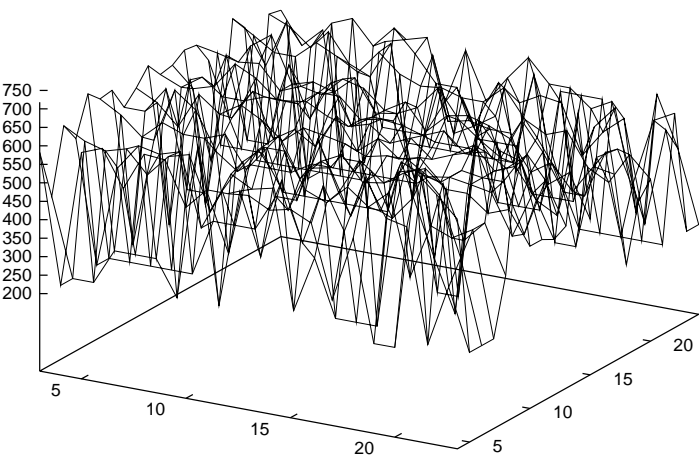


Figure 5.11: Poor Depth Map of the Inclined Plane

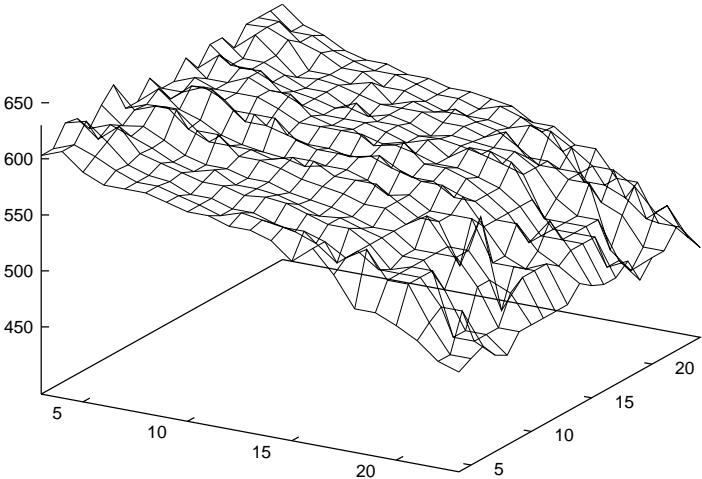


Figure 5.13: Good Depth Map of the Inclined Plane



## 5.5 Conclusion

In this chapter, we integrated DFF and DFD methods for recovering the 3D shape of objects. The integrated method has been implemented on a vision system—SVIS— and the conceptual feasibility of integrating DFD and DFF was demonstrated. The main advantage of the method is the reduction in required image frames but retain the same accuracy as DFF method. It is found that careful calibration of the camera system with respect to several factors are very important in building a successful vision system for integrating depth from defocus and focus. Details on these calibration methods are presented.

In some vision systems, this method can be further combined with stereo vision, because DFD and DFF methods are less accurate than stereo vision in providing the depth-map of a scene. However, unlike stereo vision, DFF and DFD do not suffer from the *correspondence* and *occlusion* problems. It is well-known in stereo vision that the matching process for correspondence is computationally intensive since all permissible values of stereo disparity need to be searched. In addition, sometimes there is a possibility that no matching points are present because of occlusion. In order to simplify the stereo correspondence and occlusion detection problems, DFD and DFF can provide a rough depth-map to reduce the searching range of stereo disparity for correspondence

matching. Besides, false matches due to occlusion are also reduced. Therefore, stereopsis yields a more reliable and accurate ranging and 3D shape of objects. The detailed technique for integrating DFD and DFF with stereo vision is described in [71].

## **Chapter 6**

### **Summary, Conclusions and Future Research**

#### **6.1 Overview**

The following three sections are provided in this chapter. Section 6.2 gives a summary of depth from focus and defocus methods which describes their pros and cons. The conclusions based on our research works are presented in the section 6.3 which outlines the contribution of this dissertation. In section 6.4, we indicate some opportunities for extensions and future research.

## 6.2 Summary of Depth from Focus and Defocus

The Depth from Focus (DFF) technique uses a search for the sharpest focus position over a sequence of images taken at different lens focus settings or camera positions. The main advantages of this technique as compared to the others such as stereo vision, optical flow and shading, there is no *correspondence* and *occlusion* problems involved. Since the depth is locally calculated in small image region, the computational costs incurred in DFF are lower than the others. The further speedup can be utilized through parallel processing and hardware implementation. In addition, DFF working on single imaging system is also suitable for some applications such as microscope. Consequently, DFF can serve a valuable method in machine vision for depth recovery. However, the limitation of this technique was that it required the large number of images obtained sequentially by adjusting the lens setting. During the image acquisition and process time, there is no scene motion permitted. It is then difficult to apply for moving objects or real-time applications.

Alternatively, the Depth from Defocus (DFD) approach attempts to model the blurring process in a local image region as a function of depth. Two images are taken with a difference in camera parameters, resulting in a change in blur over small corresponding regions in the two images. The depth is then measured from the blur change. In general, DFD has most of advantages of DFF. The potential advantage of DFD over DFF

is speed, because image acquisition time is the eventual limiting factor. For a general scene, however, DFF has greater applicability, since the only assumption made about the blurring process is that a quantitative measure of the defocus can be minimized at the position of sharpest focus. This reveals the problem of DFD which is less accurate than DFF for most of cases.

## 6.3 Conclusions

In this dissertation, we have investigated image focus and defocus analysis for passive autofocusing and three-dimensional shape recovery. Our research so far has addressed the question of how to select the best focus measure for autofocusing in the presence of noise. We have shown that a class of focus measures (those based on summing the absolute values of image derivatives) used by many researchers in the past are unsound. These focus measures should not be used since the advantage of computational savings attributed to them is marginal and is clearly outweighed by their unsoundness. In the absence of application specific information, we have argued that the Laplacian focus measure filter should be used. The properties of the Laplacian filter and its advantages have been investigated. An automated microscope shape-from-focus system has been developed. The system has been used to recover dense and accurate depth maps for several microscopic objects. The system is ap-

plicable to variety of industrial inspection tasks.

We have developed a model for focus measure filters, and provided a theoretical noise sensitivity analysis of these filters. Explicit expressions have been derived for the mean and variance of the focus measures computed using the filters. Based on this analysis, we have defined a metric, the ARMS error, for comparing the relative performance of different focus measures. ARMS error is useful in selecting the best focus measure from a given set of focus measures. We have verified our theoretical results experimentally. The experiments have included those using an actual camera system and those using computer simulated data.

Noise sensitivity analysis for depth from defocus using Spatial-Domain Convolution/Deconvolution Transform Methods (STM) is also given. Thorough theoretical treatment is provided for the accuracy check of STM in the presence of noise which has been only observed in experimental. We derived explicit expressions using relationships among the depth, camera parameters and the amount of blurring in images for the root-mean-square (RMS) error in autofocusing. The analysis is helpful in presenting the depth errors for a given set of images and noise characteristics. We verified the theory using both synthetic scenes and real scenes, and found that the experimental results confirm the theoretical results. Note that the version of STM considered here is a little different than the one in Surya's thesis [74]. No integration and mode selection are done here.

The DFD and the DFF methods have been combined to obtain the

speed advantage of DFD and the accuracy of DFF. We first used DFD to obtain a rough estimate of distance, then DFF was applied in a narrow interval around this estimated distance by acquiring a few more images and searching for the focus distance. An efficient camera calibration procedures was developed to obtain camera constants such as  $\beta$  and look-up tables to relate  $\sigma_2$  to object distance  $u$  and lens position  $s$  for focusing the object. We have presented the experimental results obtained from SVIS camera system.

## 6.4 Future Research

Some of the problems extended from the research work that we suggest as future research are outlined next.

### 1. Optimal Focus Measure Solution

Our work so far provides a method for selecting the best focus measure filter from a given set of focus measure filters. The best filter depends both on image content and noise characteristics. At present, there is no known method to *derive* the best focus measure filter for a given sequence of noisy blurred images. Although the Laplacian has been shown to be a good filter, it should be possible to design a better filter using additional information in the form of the given sequence of noisy blurred

image. For example the noisy image could be first smoothed to reduce noise and then a focus measure filter could be applied. The parameters of the smoothing filter need to be determined. The operations of smoothing and applying the focus measure filter could be combined.

## **2. Different Models for ARMS Error**

Our expression for ARMS error is based on a local quadratic model for the focus measure. It may be useful to explore other models, such as a Gaussian, or a model that is dynamically determined based on the blurred images processed until the current moment. The obtained solution will be helpful to apply our analysis of autofocusing method to the shape from focus and focused image reconstruction problem.

## **3. Error Sensitivity Analysis of Defocus Module**

Error sensitivity analysis of autofocusing methods based on defocus module using Fourier frequency domain approach is an important problem. Our analysis has so far been restricted to the method based on spatial domain approach (i.e. STM). Preliminary analysis and experimental results about this issue are presented in [81]. A more thorough error analysis, both theoretical and experimental, will be our interest. The various noise sources such as MTF data, camera parameters, CCD array need to be modeled.



#### **4. Speed Up Implementation**

Development of parallel algorithms for the image focus and defocus analysis are of interest in real-time applications such as robot/vehicle navigation or machine vision inspection. Further, just as special purpose ASIC and DSP hardware have been investigated by many researchers for edge detection, stereo vision and optical flow computation [16, 27, 47, 12], we would like to explore hardware implementation of the DFF and DFD algorithms for fast 3D shape and focused image recovery.

#### **5. Optimal Camera System Design**

The accuracy of autofocus and 3D shape recovery using the DFF method depends on many things such as the camera parameters, image content, and noise characteristics. For example, smaller focal length lenses (e.g. microscope lenses) are useful for depth recovery at shorter ranges and longer focal length lenses (e.g. binocular/telescope lenses) are useful at longer ranges. For a given specifications on the range of operation and accuracy of depth recovery, a camera system with some particular values for the camera parameters performs optimally. It is important to develop a method for determining the parameters of such a camera system, taking into account the available technology (e.g. CCD pixel size).

## 6. Stereo Based Microscope System

In current microscope system, we only use image focus analysis based on single lens configuration for 3D shape recovery. The similar concept on SVIS system using a single camera for stereo vision can be implemented on microscope system. We can use a stepper motor mounted on a translational stage of microscope to drive the stage with a linear motion on X-Y plane. The camera is still mounted such that its optical axis is perpendicular to linear motion of the stage. The right and left images for stereo disparity analysis are obtained by moving the stage to different positions and recording images. We are interested in integrating the stereo vision with DFF and DFD methods to provide fast and accurate 3D shape measurement on microscope system.

## Appendix A

### Related Derivations in Chapter 3

#### A.1 Variance of Focus Measures

The term  $E\{\gamma_{noise}^2\}$  can be shown to be equal to

$$\frac{1}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M \sum_{i_3, j_3}^M \sum_{i_4, j_4}^M \left( \prod_{k=1}^4 a(i_k, j_k) \right) E\{\eta(m_1 + i_1, n_1 + j_1) \eta(m_1 + i_2, n_1 + j_2) \eta(m_2 + i_3, n_2 + j_3) \eta(m_2 + i_4, n_2 + j_4)\} \quad (\text{A.1})$$

The above expression can be evaluated by considering the following cases.

Case 1: Among the four factors involving  $\eta$ , at least one of them is different from the other three. In this case the entire term evaluates to zero.

Case 2: All the four factors involving  $\eta$  are the same. Here the conditions  $C_1$  and  $C_2$  will both be true where

$$C_1 : (i_1 = i_2) \ \& \ (j_1 = j_2) \ \& \ (i_3 = i_4) \ \& \ (j_3 = j_4) \quad (\text{A.2})$$

$$C_2 : (m_2 = m_1 + i_1 - i_3) \ \& \ (n_2 = n_1 + j_1 - j_3) \quad (\text{A.3})$$

Therefore, the term becomes

$$\begin{aligned} & \frac{E\{\eta^4\}}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{i_1, j_1}^M \sum_{i_3, j_3}^M a^2(i_1, j_1) a^2(i_3, j_3) \\ &= \frac{E\{\eta^4\}}{(2N+1)^4} \sum_{m_1, n_1}^N \left( \sum_{i_1, j_1}^M a^2(i_1, j_1) \right)^2 \\ &= \frac{E\{\eta^4\}}{(2N+1)^2} A_n^2 \end{aligned} \quad (\text{A.4})$$

Case 3: Any two pairs among the four factors involving  $\eta$  are equal but not all four are equal. Here we consider two subcases.

Case (a): Condition  $C_1$  is true and condition  $C_2$  is false. In this case we get

$$\begin{aligned} & \frac{\sigma_n^4}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N \sum_{i_1, j_1}^M \sum_{i_3, j_3}^M a^2(i_1, j_1) a^2(i_3, j_3) \\ & - \frac{\sigma_n^4}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{i_1, j_1}^M \sum_{i_3, j_3}^M a^2(i_1, j_1) a^2(i_3, j_3) \\ &= \sigma_n^4 A_n^2 \left[ 1 - \frac{1}{(2N+1)^2} \right] \end{aligned} \quad (\text{A.5})$$

Case (b): Define conditions

$$\begin{aligned} C_3 : & (m_1 + i_1 = m_2 + i_3) \ \& \ (n_1 + j_1 = n_2 + j_3) \\ & \& \ (m_1 + i_2 = m_2 + i_4) \ \& \ (n_1 + j_2 = n_2 + j_4) \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} C_4 : & (m_1 + i_1 = m_2 + i_4) \ \& \ (n_1 + j_1 = n_2 + j_4) \\ & \& \ (m_1 + i_2 = m_2 + i_3) \ \& \ (n_1 + j_2 = n_2 + j_3) \end{aligned} \quad (\text{A.7})$$

From the above conditions we deduce respectively

$$C_5 : (i_1 - i_3 = i_2 - i_4) \ \& \ (j_1 - j_3 = j_2 - j_4) \quad (\text{A.8})$$

$$C_6 : (i_1 - i_4 = i_2 - i_3) \ \& \ (j_1 - j_4 = j_2 - j_3) \quad (\text{A.9})$$

Let  $Q$  be a boolean variable with value 1 if the following condition is true and zero otherwise:

$$Q :: (C_5 \text{ OR } C_6) \ \& \ \text{NOT } C_1 \quad (\text{A.10})$$

In this case we obtain

$$\frac{\sigma_n^4}{(2N+1)^2} \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M \sum_{i_3, j_3}^M \sum_{i_4, j_4}^M Q \cdot \left( \prod_{k=1}^4 a(i_k, j_k) \right) \quad (\text{A.11})$$

Now consider

$$\begin{aligned} & \frac{4}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N F(m_1, n_1) F(m_2, n_2) E\{\mathcal{N}(m_1, n_1) \mathcal{N}(m_2, n_2)\} \\ = & \frac{4}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M F(m_1, n_1) F(m_2, n_2) a(i_1, j_1) a(i_2, j_2) \\ & E\{\eta(m_1 + i_1, n_1 + j_1) \eta(m_2 + i_2, n_2 + j_2)\} \end{aligned} \quad (\text{A.12})$$

The term involving  $\eta$  is non zero only when  $m_1 + i_1 = m_2 + i_2$  **and**  $n_1 + j_1 = n_2 + j_2$ . Introducing the change of variables  $m = m_1 + i_1 = m_2 + i_2$  **and**  $n = n_1 + j_1 = n_2 + j_2$  for  $-2(M+N)-1 \leq m, n \leq 2(M+N)+1$ , therefore we get

$$\begin{aligned} & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m, n}^{M+N} \left[ \sum_{i_1, j_1}^M a(i_1, j_1) F(m - i_1, n - j_1) \right] \left[ \sum_{i_2, j_2}^M a(i_2, j_2) F(m - i_2, n - j_2) \right] \\ = & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m, n}^{M+N} [a(i, j) * F(m, n)]^2 \\ = & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m, n}^{M+N} [A(i, j) * f(m, n)]^2 \\ = & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m, n}^{M+N} F'^2(m, n) \end{aligned} \quad (\text{A.13})$$

where  $*$  represents the convolution operator and

$$A(i, j) = a(i, j) * a(-i, -j) \quad (\text{A.14})$$

$$F'(m, n) = A(i, j) * f(m, n) \quad (\text{A.15})$$

Combining all the results from eqs. (3.45, A.4, A.5, A.11 and A.13), we get

$$\begin{aligned} Var\{\gamma\} = & \frac{A_n^2 E\{\eta^4\}}{(2N+1)^2} - \frac{A_n^2 \sigma_n^4}{(2N+1)^2} + \frac{\sigma_n^4}{(2N+1)^2} \times \\ & \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M \sum_{i_3, j_3}^M \sum_{i_4, j_4}^M Q \cdot \left( \prod_{k=1}^4 a(i_k, j_k) \right) + \frac{4\sigma_n^2}{(2N+1)^2} \gamma'_{signal} \end{aligned} \quad (\text{A.16})$$

where

$$\gamma'_{signal} = \frac{1}{(2N+1)^2} \sum_{m, n}^{M+N} F'^2(m, n) \quad (\text{A.17})$$

In the above equation, the first three terms do not depend on the image signal. They can be computed and prestored. Among these three terms, the first two can be computed manually, but the third term may need a small computer program to evaluate. The last term in the above equation depends on the image being processed. Exact computation of this term requires knowledge of the noise-free image which is not possible. However the value of the term can be approximated using the noisy image  $g(m, n)$ . The approximation is valid for high signal to noise ratio. Therefore we have

$$\begin{aligned} Var\{\gamma\} \approx & \frac{A_n^2 E\{\eta^4\}}{(2N+1)^2} - \frac{A_n^2 \sigma_n^4}{(2N+1)^2} + \frac{\sigma_n^4}{(2N+1)^2} \times \\ & \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M \sum_{i_3, j_3}^M \sum_{i_4, j_4}^M Q \cdot \left( \prod_{k=1}^4 a(i_k, j_k) \right) + \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m, n}^{M+N} (A(i, j) * g(m, n))^2 \end{aligned} \quad (\text{A.18})$$

## A.2 Gradient Focus Measure

The accurate derivation for the variance of focus measure using gradient magnitude squared filter should consider a cross item generated through x direction and y direction. If we start with our definition from equation (3.37), the total focus measure  $\gamma$  would be sum of along x-axis  $\gamma_x$  and y axis  $\gamma_y$  components and it becomes:

$$\begin{aligned}\gamma &= \gamma_x + \gamma_y \\ &= \gamma_{sx} + \gamma_{nx} + \frac{2}{(2N+1)^2} \sum_{m,n}^N F_x(m,n) \mathcal{N}_x(m,n) \\ &\quad + \gamma_{sy} + \gamma_{ny} + \frac{2}{(2N+1)^2} \sum_{m,n}^N F_y(m,n) \mathcal{N}_y(m,n)\end{aligned}\quad (\text{A.19})$$

where the suffix  $sx$  and  $nx$  mean the focus measure component contributed in signal and noise along x axis separately as so do  $sy$  and  $ny$ .

If the noise characteristics are the same in both of direction, the expected value of focus measure in the equation (3.47) is able to be applied here as:

$$E\{\gamma\} = \gamma_{sx} + \gamma_{sy} + 4\sigma_n^2 \quad (\text{A.20})$$

However, a cross item appear as we take square to both of  $\gamma$  and  $E\{\gamma\}$ . If we define:

$$F_x \mathcal{N}_x = \frac{2}{(2N+1)^2} \sum_{m,n}^N F_x(m,n) \mathcal{N}_x(m,n) \quad (\text{A.21})$$

as well as

$$F_y \mathcal{N}_y = \frac{2}{(2N+1)^2} \sum_{m,n}^N F_y(m,n) \mathcal{N}_y(m,n) \quad (\text{A.22})$$

and substituting to eq. (A.19), then the variance will be written from eq. (3.48) as:

$$\begin{aligned}
Var\{\gamma\} &= Var\{\gamma_x\} + Var\{\gamma_y\} + Var\{\gamma_{xy}\} \\
&= E\{(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x)^2 + (\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y)^2 \\
&\quad + 2(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x)(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y)\} - \\
&\quad \{(\gamma_{sx} + 2\sigma_n^2)^2 + (\gamma_{sy} + 2\sigma_n^2)^2 + 2(\gamma_{sx} + 2\sigma_n^2)(\gamma_{sy} + 2\sigma_n^2)\} \quad (\text{A.23})
\end{aligned}$$

where

$$Var\{\gamma_x\} = E\{(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x)^2\} - (\gamma_{sx} + 2\sigma_n^2)^2 \quad (\text{A.24})$$

and

$$Var\{\gamma_y\} = E\{(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y)^2\} - (\gamma_{sy} + 2\sigma_n^2)^2 \quad (\text{A.25})$$

The  $Var\{\gamma_x\}$  and the  $Var\{\gamma_y\}$  are exactly same as we derived in eqs. (3.55) and (3.58). Thus, the  $Var\{\gamma_{xy}\}$  is a cross term and equals to:

$$2E\{(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x)(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y)\} - 2(\gamma_{sx} + 2\sigma_n^2)(\gamma_{sy} + 2\sigma_n^2) \quad (\text{A.26})$$

Because of  $E\{\mathcal{N}_x\} = E\{\mathcal{N}_y\} = 0$ , the above equation can be simplified and expanded as:

$$\begin{aligned}
Var\{\gamma_{xy}\} &= 2E\{F_x\mathcal{N}_x F_y\mathcal{N}_y\} \\
&= \frac{8}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N F_x(m_1, n_1) F_y(m_2, n_2) E\{\mathcal{N}_x(m_1, n_1) \mathcal{N}_y(m_2, n_2)\} \\
&= \frac{8}{(2N+1)^4} \sum_{m_1, n_1}^N \sum_{m_2, n_2}^N \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M a_x(i_1, j_1) a_y(i_2, j_2) F_x(m_1, n_1) F_y(m_2, n_2) \\
&\quad E\{\eta(m_1 + i_1, n_1 + j_1) \eta(m_2 + i_2, n_2 + j_2)\} \quad (\text{A.27})
\end{aligned}$$



The term  $E\{\eta(m_1 + i_1, n_1 + j_1)\eta(m_2 + i_2, n_2 + j_2)\}$  is non zero only when  $m = m_1 + i_1 = m_2 + i_2$  **and**  $n = n_1 + j_1 = n_2 + j_2$  **for**  $-2(M + N) - 1 \leq m, n \leq 2(M + N) + 1$ , **therefore we have**

$$\begin{aligned}
 Var\{\gamma_{xy}\} &= \frac{8\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} \left[ \sum_{i_1,j_1}^M a_x(i_1, j_1) F_x(m - i_1, n - j_1) \right] \times \\
 &\quad \left[ \sum_{i_2,j_2}^M a_y(i_2, j_2) F_y(m - i_2, n - j_2) \right] \\
 &= \frac{8\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [a_x(i, j) * F_x(m, n)] [a_y(i, j) * F_y(m, n)] \\
 &= \frac{8\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A_x(i, j) * f(m, n)] [A_y(i, j) * f(m, n)]
 \end{aligned} \tag{A.28}$$

where

$$A_x(i, j) = a_x(i, j) * a_x(-i, -j) \tag{A.29}$$

$$A_y(i, j) = a_y(i, j) * a_y(-i, -j) \tag{A.30}$$

, they are the same as Eq. (3.56) and Eq. (3.59) respectively.

### A.3 Another Approach for RMS Error in DFF

In chapter 3, we proposed two metrics - AUM and ARMS error for evaluating the performance of focus measures. During this research, we investigated another approach based on the density function of focus measure itself to compute RMS error in autofocusing. The theory is derived as follows:

Since autofocusing is done by moving the lens with respect to the image detector in a narrow range, the distance from focal length to image detector can be discretized as lens positions  $s_0, s_1, \dots, s_N$ . At each lens position, a focus measure  $\gamma_{s_i}$  is associated with a probability density function  $P(\gamma_{s_i})$  with an expected value  $\eta_{s_i}$  and a standard deviation  $\sigma_{s_i}$ . If we assume the density function of focus measure is a normal distribution for any lens positions, it can be expressed as:

$$P(\gamma_{s_i}) = \frac{1}{\sigma_{s_i} \sqrt{2\pi}} \exp\left\{-\frac{(\gamma_{s_i} - \eta_{s_i})^2}{2\sigma_{s_i}^2}\right\} \quad (\text{A.31})$$

For a particular position  $s_k$ , it is possible to be the focus position only if it has a maximum focus measure as compared to focus measures at all of lens positions. Therefore, the probability of maximum focus measure location can be written as:

$$\begin{aligned} & \tilde{P}(\mathbf{s}_k = \text{focus position}) \\ &= P_0(\gamma_{s_k} > \gamma_{s_0}) P_1(\gamma_{s_k} > \gamma_{s_1}) \cdots P_N(\gamma_{s_k} > \gamma_{s_N}) \\ &= \prod_{s_i=s_0, s_i \neq s_k}^{s_N} P_i(\gamma_{s_k} > \gamma_{s_i}) \end{aligned} \quad (\text{A.32})$$

Consider one function with two random variables  $\gamma_{s_k}$  and  $\gamma_{s_i}$ , we obtain another random variable  $z$

$$P_i(z > 0) = P_i(\gamma_{s_k} > \gamma_{s_i}) \quad (\text{A.33})$$

where

$$z = \gamma_{s_k} - \gamma_{s_i} \quad (\text{A.34})$$

If focus measure computed at two different position is independent, then the joint density function of  $z$  is equal to the convolution of the

density function  $P(\gamma_{s_k})$  and  $P(\gamma_{s_i})$  Hence, we can obtain

$$P_z(z) = \frac{1}{2\pi\sigma_{s_i}\sigma_{s_k}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(z + \gamma_{s_i} - \eta_{s_k})^2}{2\sigma_{s_k}^2} - \frac{(\gamma_{s_i} - \eta_{s_i})^2}{2\sigma_{s_i}^2}\right\} d\gamma_{s_i} \quad z \in R \quad (\text{A.35})$$

the above integral can be simplified as:

$$P_z(z) = \frac{1}{2\pi\sigma} \exp\left\{-\frac{(z - \eta)^2}{2\sigma^2}\right\} \quad (\text{A.36})$$

where

$$\begin{aligned} \eta &= \eta_{s_k} - \eta_{s_i} \\ \sigma &= \sqrt{\sigma_{s_k}^2 + \sigma_{s_i}^2} \end{aligned} \quad (\text{A.37})$$

The density function of  $z$  shows itself still a normal distribution with mean  $\eta$  and standard deviation  $\sigma$ . Thus, the probability for all  $z > 0$  can be integrated from its density function.

$$\begin{aligned} P_i(z > 0) &= \int_0^{\infty} P_z(z) dz \\ &= \frac{1}{2\pi\sigma} \int_0^{\infty} \exp\left\{-\frac{(z - \eta)^2}{2\sigma^2}\right\} dz \end{aligned} \quad (\text{A.38})$$

Applying the Gaussian function integral [44] and changing with variables, it yields

$$\begin{aligned} P_i(z > 0) &= G\left(\frac{\eta}{\sigma}\right) \\ &= G\left(\frac{\eta_{s_k} - \eta_{s_i}}{\sqrt{\sigma_{s_k}^2 + \sigma_{s_i}^2}}\right) \end{aligned} \quad (\text{A.39})$$

where the function  $G$  is expressed in terms of the *error function* defined as

$$\begin{aligned} erf(x) &= \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left\{-\frac{y^2}{2}\right\} dy \\ &= G(x) - \frac{1}{2} \end{aligned} \quad (\text{A.40})$$

then equation ( A.32) can be obtained

$$\tilde{P}(\mathbf{s}_k = \text{focus position}) = \prod_{s_i=s_0, s_i \neq s_k}^{s_N} G\left(\frac{\eta_{s_k} - \eta_{s_i}}{\sqrt{\sigma_{s_k}^2 + \sigma_{s_i}^2}}\right) \quad (\text{A.41})$$

Consequently, the expected value and variance can be computed from the density function  $\tilde{P}(\mathbf{s}_k)$

$$\begin{aligned} E\{\mathbf{s}_k\} &= \sum_{s_k=s_0}^{s_N} \left[ s_k \times \prod_{s_i=s_0, s_i \neq s_k}^{s_N} G\left(\frac{\eta_{s_k} - \eta_{s_i}}{\sqrt{\sigma_{s_k}^2 + \sigma_{s_i}^2}}\right) \right] \\ \text{Var}\{\mathbf{s}_k\} &= \sum_{s_k=s_0}^{s_N} \left[ (s_k - \bar{S})^2 \times \prod_{s_i=s_0, s_i \neq s_k}^{s_N} G\left(\frac{\eta_{s_k} - \eta_{s_i}}{\sqrt{\sigma_{s_k}^2 + \sigma_{s_i}^2}}\right) \right] \end{aligned} \quad (\text{A.42})$$

where  $E\{\mathbf{s}_k\} = \bar{S}$ . The RMS error in autofocusing derived here is helpful to select the most accurate focus measure from all of the focus measures in a set. However, the above equation needs to be computed by obtaining the mean and standard deviation of focus measures at all of lens positions. For practical usage, it is valid to approximate the above equation by using the range from +5 steps to -5 steps of maximum focus measure position. As compared to AUM and ARMS error, this approach is obviously more complex and more computational.

## A.4 Wave Optics for PSF

The PSF used in our simulation experiment is based on a formula from the Optical Transfer Function (OTF) of a defocused optical system derived by Hopkins in his classic paper [20]. A simplification version of

Hopkins' OTF is defined as:

$$H(\rho, \Delta) = \frac{4}{\pi} \int_{\rho}^1 \sqrt{1-t^2} \cos [2\pi\Delta\rho(t-\rho)] dt \quad (\text{A.43})$$

where the focus defect  $\Delta$  related to camera parameter settings is defined as:

$$\Delta \simeq \frac{2R^2}{\lambda} \left( \frac{1}{f} - \frac{1}{u} - \frac{1}{v} \right) \quad (\text{A.44})$$

where  $\lambda$  is the wavelength of the incident light. The corresponding PSF,  $h(r, \Delta)$ , can be obtained by the inverse Fourier-Bessel Transform given by

$$h(r, \Delta) = 2\pi \int_0^{\infty} H(\rho, \Delta) J_0(2\pi\rho r) \rho d\rho \quad (\text{A.45})$$

where  $J_0$  is the Zeroth Order Bessel function of the first kind and  $r$  is the reduced radial distance on the image detector plane.

A comparative study of PSF between wave optics and geometric optics has been carried out earlier by Lee [33]. Note that here the PSF using wave optics is slightly different with Lu who implemented the model in the IDS [64].

## Appendix B

### Related Derivations in Chapter 4

#### B.1 The Variance of STM

In this section we discuss the derivation of Eq (4.28) by simplifying Eq (4.27). Let the smoothing filter  $L_0$  and the Laplacian filter  $L_2$  both be of size  $(2M + 1) \times (2M + 1)$ . Let the coefficients of the two filters be  $a_0(i, j)$  and  $a_2(i, j)$  respectively. Also, let  $\sum_{i,j}^M$  denote  $\sum_{i=-M}^M \sum_{j=-M}^M$ . Then filtering with  $L_0$  and  $L_2$  are convolution operations defined respectively by

$$\eta_k * L_0 = \sum_{i,j}^M a_0(i, j) \eta_k(m - i, n - j) \quad (\text{B.1})$$

$$\eta_k * L_2 = \sum_{i,j}^M a_2(i, j) \eta_k(m - i, n - j) \quad (\text{B.2})$$

There are three terms—  $E\{P'^2\}$ ,  $E\{Q'^2\}$  and  $E\{P'Q'\}$ — that need to be discussed in Eq (4.27). We start from the term  $E\{P'^2\}$ . Using Eqs.

(4.20, 4.22) we obtain

$$\begin{aligned}
 E\{P'^2\} &= E\{(\eta_1 * L_o - \eta_2 * L_o)^2\} \\
 &= E\{(\eta_1 * L_o)^2\} + E\{(\eta_2 * L_o)^2\} - 2E\{(\eta_1 * L_o)(\eta_2 * L_o)\}
 \end{aligned} \tag{B.3}$$

In the above equation, the first term can be written as

$$E\{(\eta_1 * L_o)^2\} = \sum_{i_1, j_1}^M \sum_{i_2, j_2}^M a_0(i_1, j_1) a_0(i_2, j_2) E\{\eta_1(m - i_1, n - j_1) \eta_1(m - i_2, n - j_2)\} \tag{B.4}$$

If  $i_1 \neq i_2$  or  $j_1 \neq j_2$ , then, since noise in different pixels are independent and zero mean,

$$\begin{aligned}
 &E\{\eta_1(m + i_1, n + j_1) \eta_1(m + i_2, n + j_2)\} \\
 &= E\{\eta_1(m + i_1, n + j_1)\} E\{\eta_1(m + i_2, n + j_2)\} \\
 &= 0
 \end{aligned} \tag{B.5}$$

However, if  $i_1 = i_2$  and  $j_1 = j_2$ , then

$$E\{\eta^2(m - i_1, n - j_1)\} = \sigma_n^2 \tag{B.6}$$

Therefore, we get

$$E\{(\eta_1 * L_o)^2\} = \sum_{i, j}^M a_0^2(i, j) \sigma_n^2 \tag{B.7}$$

Similarly, for the second term  $E\{(\eta_2 * L_o)^2\}$  in Eq (B.3), we obtain the same result as above. For the remaining term we have

$$2E\{(\eta_1 * L_o)(\eta_2 * L_o)\} = 2(E\{\eta_1\} * L_o)(E\{\eta_2\} * L_o) = 0 \tag{B.8}$$

Thus, we obtain

$$E\{P'^2\} = 2 \sum_{i,j}^M a_0^2(i,j) \sigma_n^2 \quad (\text{B.9})$$

The derivation for  $E\{Q'^2\}$  is similar to that for  $E\{P'^2\}$  above, and therefore we have

$$E\{Q'^2\} = 2 \sum_{i,j}^M a_2^2(i,j) \sigma_n^2 \quad (\text{B.10})$$

Now consider the term  $E\{P'Q'\}$ ,

$$\begin{aligned} E\{P'Q'\} &= E\{(\eta_1 * L_o - \eta_2 * L_o)(\eta_1 * L_2 + \eta_2 * L_2)\} \\ &= E\{(\eta_1 * L_o)(\eta_1 * L_2)\} + E\{(\eta_1 * L_o)(\eta_2 * L_2)\} \\ &\quad - E\{(\eta_2 * L_o)(\eta_1 * L_2)\} - E\{(\eta_2 * L_o)(\eta_2 * L_2)\} \quad (\text{B.11}) \end{aligned}$$

Using arguments similar to those in simplifying Eqs. (B.4) and (B.8), we obtain

$$E\{(\eta_1 * L_o)(\eta_1 * L_2)\} = E\{(\eta_2 * L_o)(\eta_2 * L_2)\} = \sum_{i,j}^M a_0(i,j) a_2(i,j) \sigma_n^2 \quad (\text{B.12})$$

as well as

$$E\{(\eta_1 * L_o)(\eta_2 * L_2)\} = E\{(\eta_2 * L_o)(\eta_1 * L_2)\} = 0 \quad (\text{B.13})$$

All the terms cancel out or vanish and result in  $E\{P'Q'\}$  to be zero. Therefore, combining all the results from Eqs. (B.9, B.10 and B.11), we can rewrite Eq (4.27) as

$$Var\{G\} = \frac{128}{Q^2} \sigma_n^2 \sum_{i,j}^M a_0^2(i,j) + \frac{128 \overline{P}^2}{Q^4} \sigma_n^2 \sum_{i,j}^M a_2^2(i,j) \quad (\text{B.14})$$



## Bibliography

- [1] A. L. Abbott and N. Ahuja, "Surface Reconstruction by Dynamic Integration of Focus, Camera Vergence and Stereo", *Second Intl. Conf. Computer Vision*, IEEE Computer Society, pp. 532-543, Dec. 1988.
- [2] S. Allegro, C. Chanel and J. Jacot, "Autofocus for Automated Microassembly under a Microscope", *IEEE International Conference on Image Processing*, vol. 2, pp. 677-680, Lausanne, Switzerland, September 1996.
- [3] N. Asada, H. Fujiwara and T. Matsuyama, "Edge and depth from focus", *Proc. of Asian Conference on Computer Vision*, pp. 83-86, November 1993.
- [4] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Oxford, Sixth Edition (1980).
- [5] L. H. Bieman, "Survey of design considerations for 3-D imaging systems", *Proc. of SPIE Conference on Optics, Illumination, and Image Sensing for Machine Vision III*, Vol. 1005, pp. 138-144, 1988.

- [6] V. M. Jr. Bove, "Entropy-based depth from focus", *Journal of Optical Society of American A*, vol 10, pp. 561-566, April 1993.
- [7] J. Brenner *et al.*, "An Automated Microscope for Cytologic Research - A Preliminary Evaluation", *Journal of Histochemistry and Cytochemistry*, vol. 24, no. 1, pp. 100-111, 1976.
- [8] T. Choi, "Shape and Image Reconstruction from Focus", Ph.D. Thesis, Dept. of Electrical Engg., SUNY at Stony Brook, 1993.
- [9] T. Darrell and K. Wohn, "Pyramid based depth from focus", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 504-509, June 1988.
- [10] J. Ens and P. Lawrence, "A Matrix Based Method for Determining Depth from Focus", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 600-609, June 1991.
- [11] J. Ens, "An Investigation of Methods for Determining Depth from Focus", Ph.D. Dissertation, University of British Columbia, 1992.
- [12] Olivier Faugeras *et al.*, "Real Time Correlation-Based Stereo: Algorithm, Implementations and Applications", *Rapport Technique 2013*, INRIA, August 1993.
- [13] M. Gokstorp, "Computing depth from out-of-focus blur using a local frequency representation", *Proceedings of IEEE International Conference on Pattern Recognition*, pp. 153-158, October 1994.

- [14] B. Girod and S. Scherock, "Depth from Defocus of Structured Light", *Proceedings of SPIE on Optics, Illumination and Image Sensing for Machine Vision*, Vol. 1194, November 1989.
- [15] P. Grossman, "Depth from Focus", *Pattern Recognition Letters*, 9(1), pp. 63-69, Jan. 1987.
- [16] A. Gruss, S. Tada and T. Kanade, "A VLSI smart sensor for fast range imaging", *Proceedings of ARPA Image Understanding Workshop*, pp. 977-986, April 1993.
- [17] R. M. Haralick and L. G. Shapiro, "*Computer and Robot Vision*", Vol. 1, Addison-Wesley Publishing Co., Inc, 1992.
- [18] M. Herman, T. Kanade, and S. Kuroe, "Incremental Acquisition of a Three-Dimensional Scene Model from Images", *IEEE Trans. Pattern Analysis and Mach. Intell.*, PAMI-6 (3), pp. 331-340 (May 1984).
- [19] W. Hoff and N. Ahuja, "Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection" *IEEE Trans. Pattern Analysis and Mach. Intell.*, Vol 11, No 2, pp. 121-136 (Feb. 1989).
- [20] H. H. Hopkins, "The frequency response of a defocused optical system" *Proceeding Royal Society of London*, A 231, 1955, pp 91-103.
- [21] B. K. P. Horn, "Focusing", Artificial Intelligence Memo No. 160, Massachusetts Institute of Technology, Cambridge, MA, USA, May 1968.

- [22] B. K. P. Horn, *Robot Vision*, McGraw-Hill Book Company, 1986.
- [23] R.A. Jarvis, "Focus Optimization Criteria for Computer Image Processing", *Microscope* **24** (2), pp. 163-180 (1976).
- [24] R. A. Jarvis, " A perspective on range finding techniques for computer vision", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 5(2): pp.122-139, March 1983.
- [25] S. M. Johnson, "Best Exploration for Maximum is Fibonacci", *Technical Report* on P-856, RAND Corporation, Santa Monica CA, 1956.
- [26] K. Kaneda, Y. Wakasu, E. Nakamae and E. Tazawa, "A Method of Pan-Focused and Stereoscopic Display Using a Series of Optical Microscopic Images", *Proceedings of Fourth Symposium on Image Sensing Technologies in Industry*, pp. 189-194, June, 1988.
- [27] T. Kanade, A. Gruss and L. R. Carley, "A very fast VLSI rangefinder", *Proceedings on International Conference on Robotics and Automation*, pp. 1322-1329, April 1991.
- [28] A. Krishnan and N. Ahuja, "Range estimation from focus using a non-frontal imaging camera", *Proceedings of AAAI Conference*, pp. 830-835, July 1993.
- [29] E.P. Krotkov, "Exploratory Visual Sensing for Determining Spatial Layout with an Agile Stereo Camera System", *Ph.D Dissertation MS-CIS-87-29*, University of Pennsylvania, May 1987.

- [30] E.P. Krotkov, "Focusing", *International Journal of Computer Vision*, 1, 223-237, 1987.
- [31] E.P. Krotkov and R. Kories, "Cooperative focus and stereo ranging", *Proceedings The Fourth Conference on Artificial Intelligence Applications*, Pages 76-81, March 1988.
- [32] S. Lai and C. Fu, "A Generalized Depth Estimation Algorithm with a Single Image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, NO. 4, April 1992, pp 405-411.
- [33] H. C. Lee, "A review of image blur model in a photographic system using the principles of optics", *Optical Engineering*, May, 1990.
- [34] J. J. Leu, C. J. Tsai, Y. P. Hung, and C. H. Chen, "Depth Recovery by Integrating Depth-from-defocus with Stereo", *International Conf. on Automation, Robotics, and Computer Vision*, Singapore, pp. CV-6.7.1 - CV-6.7.5, Sep. 1992.
- [35] G. Lighthart and F.C.A. Groen, "A Comparison of Different Autofocus Algorithms", *Proceedings on International Conference on Pattern Recognition*, pp. 597-600, 1982.
- [36] P. Meer and I. Weiss, *Smoothed differentiation filters for images*, Tech. Report No. CS-TR-2194, Center for Automation Research, University of Maryland, College Park, MD 20742-3411.
- [37] R. A. Muller and A. Buffington, "Real-time Correction of Atmospherically degraded Telescope Images through Image Sharpening",

- J. Opt. Soc. America*, 64(9), pp. 1200-1210 (Sep. 1974).
- [38] H. N. Nair and C. V. Stewart “Robust Focus Ranging” *IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 309-314, June 1992.
  - [39] S. K. Nayar, “Shape from Focus System” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 302-308 (June 1992).
  - [40] S. K. Nayar and Y. Nakagawa. “Shape from focus: an effective approach for rough surfaces”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-16, No. 8, pp. 824-831, August 1994.
  - [41] S. K. Nayar, M. Watanabe and M. Noguchi, “Real-time focus range sensor”, *Proc. of Intl. Conference on Computer Vision*, pp. 995-1001, June 1995.
  - [42] M. Noguchi and S. K. Nayar, “Microscopic shape from focus using active illumination”, *Proceedings on International Conference on Pattern Recognition*, October 1994.
  - [43] M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo”, *IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition*, (1991).
  - [44] A. Papoulis, “*Probability, Random Variables, and Stochastic Processes*”, McGraw-Hill Book Company, the 3rd, 1991.
  - [45] A. P. Pentland, “Depth of Scene from Depth of Field”, *Proc. Image Understanding Workshop*, pp. 253-259 (April 1982).

- [46] 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.
- [47] A. Pentland, T. Darell, M. Turk and W. Huang, "A Simple Real-time Range Camera", *IEEE Comp. Soc. Conf. Comp. Vision and Pattern Recognition*, pp. 256-261 1989.
- [48] A. Pentland, S. Scherrock, T. Darell and B. Girod, "Simple range cameras based on focal error", *Journal of Optical Society of America*, 11(11):2925-2935, November 1994.
- [49] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C: the art of scientific computing*, 2nd ed., Cambridge University Press, 1992.
- [50] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Vol 1, Academic Press, New York (1982).
- [51] B. Ross, "A Practical Stereo System", *IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition*, pp. 148–153 (1993).
- [52] 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", *Tech. Report CMU-RI-TR-83-14*, Robotics Institute, Carnegie-Mellon University (August 1983).
- [53] C.V. Stewart and C.R. Dyer, "Local Constraint Integration in a Connectionist Model of Stereo Vision", *IEEE Comp. Soc. Conf. Computer*

*Vision and Pattern Recognition*, pp. 165-170 (1988).

- [54] M. Subbarao, "Direct Recovery of Depth Map I: Differential Methods", *Proc. IEEE Computer Society Workshop on Computer Vision*, pp. 58-65, December 1987. (Also in Tech. Report 87-02, Computer Vision Lab, Dept. of Electrical Engineering, SUNY at Stony Brook.)
- [55] M. Subbarao, "Direct Recovery of Depth Map II: A new Robust Approach", *Tech. Report 87-03*, Computer Vision Lab, Dept. of Electrical Engineering, SUNY at Stony Brook, April 1987.
- [56] M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters", *Second International Conference on Computer Vision*, Florida, USA, pp. 149-155, December 1988.
- [57] M. Subbarao, "Efficient Depth Recovery through Inverse Optics", Editor: H. Freeman, *Machine Vision for Inspection and Measurement*, Academic press, Boston, pp. 101-126, 1989.
- [58] 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.
- [59] 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.



- [60] M. Subbarao, T. Choi, and A. Nikzad, "Focusing Techniques", *Proceedings of SPIE Conference, OE/TECHNOLOGY '92*, Vol. 1823, pp 163 - 174, Boston MA, November 1992. (Also appear in *the Journal of Optical Engineering*, Vol. 32, No. 11, pp. 2824-2836, November 1993.)
- [61] M. Subbarao and T. Choi, "A New Method for Shape from Focus", *Proceedings of SPIE Conference*, Vol. 2064, September 1993.
- [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 Y.F. Liu, "Accurate Reconstruction of 3D Shape and Focused Image from a Large Sequence of Noisy Defocused Images", *Proceedings of SPIE Conference on Three-Dimensional Imaging and Laser-Based Systems for Metrology and Inspection II*, Vol. 2909, pp. 178-191, Boston MA, November 1996.
- [64] M. Subbarao, and M.C. Lu, "Computer Modeling and Simulation of Camera Defocus", *Machine Vision and Applications*, (1994) 7, pp. 277-289.
- [65] M. Subbarao, and A. Nikzad, "Model for image sensing and digitization in computer vision", *Proceedings of SPIE conference*, Vol 1385, pp70-84, Boston MA, October 1990.

- [66] M. Subbarao and G. Surya, "Application of Spatial-Domain Convolution/Deconvolution Transform for Determining Distance from Image Defocus", *Proceedings of SPIE conference*, OE/TECHNOLOGY '92, Vol. 1822, pp. 159-167, Boston MA, November 1992. (Also appear in *International Journal of Computer Vision*, 13, 3, pp. 271-294, 1994.)
- [67] M. Subbarao and J.K. Tyan, "The Optimal Focus Measure for Passive Autofocusing and Depth-from-Focus", *Proceedings of SPIE Conference on Videometrics IV*, Vol. 2598, pp. 89-99, Philadelphia PA, October 1995.
- [68] M. Subbarao, and J.K. Tyan, "Selecting the Optimal Focus Measure for Passive Autofocusing and Depth-from-Focus", *Tech. Report No. 96.8.03*, Computer Vision Laboratory, Dept. of Electrical Engineering, SUNY, Stony Brook, NY 11794-2350. (Also a revised version to appear in *IEEE Transactions on Pattern Analysis and Machine Intelligence*.)
- [69] M. Subbarao and J.K. Tyan, "Root-Mean Square Error in Passive Autofocusing and 3D Shape Recovery", *Proceedings of SPIE Conference on Three-Dimensional Imaging and Laser-Based Systems for Metrology and Inspection II*, Vol. 2909, pp. 162-177, Boston MA, November 1996.
- [70] M. Subbarao, and J.K. Tyan, "Noise Sensitivity Analysis of Depth-from-Defocus by a Spatial-Domain Approach", *Proceedings of SPIE*

- Conference on Videometrics V*, Vol 3174, pp 174-187, San Diego CA, July 1997.
- [71] M. Subbarao, T. Yuan and J.K. Tyan, "Integration of Defocus and Focus Analysis with Stereo for 3D Shape Recovery", *Proceedings of SPIE Conference on Three-Dimensional Imaging and Laser-Based Systems for Metrology and Inspection III*, Vol 3204, Pittsburgh PA, October 1997.
- [72] M. Subbarao, and J.K. Tyan, "Noise Sensitivity Analysis of Depth-from-Defocus by a Spatial-Domain Approach", *Tech. Report No. 97.12.09*, Computer Vision Laboratory, Dept. of Electrical Engineering, SUNY, Stony Brook, NY 11794-2350. (Also a revised version is submitted to *the Journal of SPIE - the intel. society for optical engineering*)
- [73] M. Subbarao and T. Wei, "Depth from Defocus and Rapid Autofocusing : A practical Approach", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 773-776, June 1992.
- [74] G. Surya, "Three-dimensional scene recovery from image defocus", Ph.D. Thesis, Dept. of Electrical Engg., SUNY at Stony Brook, 1994.
- [75] G. Surya, and M. Subbarao, "Depth from Defocus by Changing Camera Aperture: A Spatial Domain Approach", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, pp. 61-67, June 1993.

- [76] J. M. Tenenbaum, "Accommodation in Computer Vision", Ph.D. Dissertation, Stanford University, November 1970.
- [77] John B. Thomas "An introduction to statistical communication theory" , John Wiley & Sons, Inc, 1969.
- [78] J.K. Tyan, "Noise Sensitivity Analysis of Depth-from-Focus Methods", *Proceedings of AAAI conference on the student abstract and poster program*, Providence RI, July 1997.
- [79] M. Watanabe and S. K. Nayar, "Telecentric Optics for Constant Magnification Imaging", *Technical Report CUCS-026-95*, Dept. of Computer Science, Columbia University, New York NY, September 1995.
- [80] M. Watanabe, S. K. Nayar and M. Noguchi, "Real-time computation of depth from defocus", *Proceedings of SPIE: Three-Dimensional and Unconventional Imaging for Industrial Inspection and Metrology*, Vol 2599:A-03, November 1995.
- [81] T. Wei, "Three-dimensional machine vision using image defocus", Ph.D. Thesis, Dept. of Electrical Engg., SUNY at Stony Brook, 1994.
- [82] R. G. Willson and S. A. Shafer, "Dynamic Lens Compensation for Active Color Imaging and Constant Magnification Focusing", *Technical Report CMU-RI-TR-91*, The Robotics Institute, Carnegie Mellon University, Pittsburgh PA, October 1991.
- [83] R. G. Willson and S. A. Shafer, "Modeling and Calibration of Automated Zoom Lenses", *Technical Report CMU-RI-TR-94-03*, The

Robotics Institute, Carnegie Mellon University, Pittsburgh PA, January 1994.

- [84] G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos CA , 1990.
- [85] Y. Xiong and S. A. Shafer, “Depth from focusing and defocusing”, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 68-73, June 1993. Also, Technical Report CMU-RI-TR-93-07, Pittsburgh, PA, USA, March, 1993.
- [86] Y. Xiong and S. A. Shafer, “Variable window gabor filters and their use in focus and correspondence”, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 668-671, June 1994.
- [87] Y. Xiong and S. A. Shafer, “Moment filters for high precision computation of focus and stereo”, *Proc. of Intl. Conf. on Robotics and Automation*, pp. 108-113, August 1995. Also, Technical Report CMU-RI-TR-94-28, Pittsburgh, PA, USA, September, 1994.
- [88] T. Yeo *et al.* “Autofocus for Tissue Microscopy”, *Image and Vision Computing*, Vol. 11, no. 10, pp. 629-639, 1993.