Theses and Dissertations

Student Graduate Works

3-11-2011

Vision Aided Inertial Navigation System Augmented with a Coded Aperture

Jamie R. Morrison

Follow this and additional works at: https://scholar.afit.edu/etd

Part of the Navigation, Guidance, Control and Dynamics Commons

VISION AIDED INERTIAL NAVIGATION SYSTEM AUGMENTED WITH A CODED APERTURE

DISSERTATION

Jamie R. Morrison, Major, USAF

AFIT/DCE/ENG/10-14

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/DCE/ENG/10-14

VISION AIDED INERTIAL NAVIGATION SYSTEM AUGMENTED WITH A CODED APERTURE

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Jamie R. Morrison, BS, MS

Major, USAF

March 2011

AFIT/DCE/ENG/10-14

VISION AIDED INERTIAL NAVIGATION SYSTEM AUGMENTED WITH A CODED APERTURE
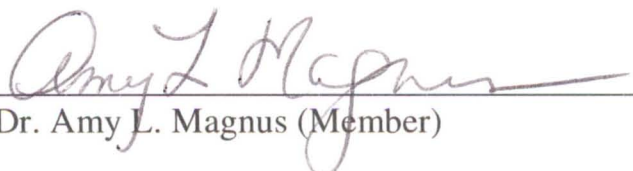
Jamie R. Morrison, BS, MS

Major, USAF

Approved:

_____
Dr. John F. Raquet (Chairman)

14 MAR 11
Date

_____
Lt Col Michael J. Veth, PhD (Member)

8 MAR 2011
Date

_____
Dr. Amy L. Magnus (Member)

14 Mar 11
Date

Accepted:

_____
M. U. Thomas
Dean, Graduate School of Engineering and Management

16 Mar 11
Date

Abstract

Navigation through a three-dimensional indoor environment is a formidable challenge for an autonomous micro air vehicle. A main obstacle to indoor navigation is maintaining a robust navigation solution (i.e. air vehicle position and attitude estimates) given the inadequate access to satellite positioning information. A MEMS (micro-electro-mechanical system) based inertial navigation system provides a small, power efficient means of maintaining a vehicle navigation solution; however, unmitigated error propagation from relatively noisy MEMS sensors results in the loss of a usable navigation solution over a short period of time. Several navigation systems use camera imagery to diminish error propagation by measuring the direction to features in the environment. Changes in feature direction provide information regarding direction for vehicle movement, but not the scale of movement. Movement scale information is contained in the depth to the features.

Depth-from-defocus is a classic technique proposed to derive depth from a single image that involves analysis of the blur inherent in a scene with a narrow depth of field. A challenge to this method is distinguishing blurriness caused by the focal blur from blurriness inherent to the observed scene. In 2007, MIT's Computer Science and Artificial Intelligence Laboratory demonstrated replacing the traditional rounded aperture with a coded aperture to produce a complex blur pattern that is more easily distinguished from the scene. A key to measuring depth using a coded aperture then is to correctly match the blur pattern in a region of the scene with a previously determined set of blur patterns for known depths.

As the depth increases from the focal plane of the camera, the observable change in the blur pattern for small changes in depth is generally reduced. Consequently, as the depth of a feature to be measured using a depth-from-defocus technique increases, the measurement performance decreases. However, a Fresnel zone plate aperture produces diffraction patterns that change the shape of the focal blur pattern. When used as an aperture, the Fresnel zone plate produces multiple focal planes in the scene. The interference between the multiple focal planes produce changes in the aperture that can be observed both between the focal planes and beyond the most distant focal plane. The Fresnel zone plate aperture and lens may be designed to change in the focal blur pattern at greater depths, thereby improving measurement performance of the coded aperture system.

This research provides an in-depth study of the Fresnel zone plate used as a coded aperture, and the performance improvement obtained by augmenting a single camera vision aided inertial navigation system with a Fresnel zone plate coded aperture. Design and analysis of a generalized coded aperture is presented and demonstrated, and special considerations for the Fresnel zone plate are given. Also techniques to determine a continuous depth measurement from a coded image are presented and evaluated through measurement. Finally the measurement results from different aperture configurations are statistically modeled and compared with a simulated vision aided navigation environment to predict the change in performance of a vision aided inertial navigation system when augmented with a coded aperture.

Acknowledgments

I first wish to thank my Lord and Savior Jesus Christ for providing me the privilege of working with so many wonderful and talented people. I am especially grateful for my wife for keeping my spirits up during many setbacks and supporting me through all of my "all-nighters" in the AFIT Penthouse. I am thankful for my children for always being ready to spend time with their father no matter how late I came home.

I am grateful for Dr. John Raquet's leadership and foresight, Lt Col Michael Veth's purposefulness and encouragement, and Dr. Amy Magnus's thoroughness and counsel. I am also thankful for Dr. Paul Havig and Alan Pinkus of the Human Effectiveness Directorate for showing me the Zoom Lane and allowing me open access to it. I also thank Dave Sivert of the Human Effectiveness Directorate for his eternally positive attitude and for patiently showing me how to care for the optics.

I am grateful for fellow classmate Ben Crossley for being the sounding board to most of my bad ideas, helping me construct the apertures, and being a friend when the stress levels rose.

I wish to thank the students and faculty of the Air Force Institute of Technology for providing cherished camaraderie and a rich environment in which to learn, explore, make mistakes, discover and grow.

I also thank the members of Hope Church for helping me keep my life and decisions in perspective and for helping me continually mature personally and spiritually.

Jamie R. Morrison

vi

Table of Contents

List of Figures

List of Tables

List of Symbols

Symbol	Description	Units
x^b	x in body frame	n/a
x^i	x in inertial frame	n/a
x^c	x in camera frame	n/a
x^n	x in navigation frame	n/a
$\lfloor x \rfloor_3$	Third component of vector x	n/a
$\tilde{x}$	Approximation of x	n/a
$\bar{x}$	Mean value of x	n/a
$\hat{x}$	Estimate of x	n/a
δx	Small value approximation of x	n/a
λ	Wavelength	m
ρ	Radial distance	m
Ψ	Small angle matrix	rad
Ω	Angular rate	$\dfrac{\text{rad}}{\text{s}}$
ω	Angular displacement	rad
A_d	Diameter of traditional round aperture	m
A_{psf}	Amplitude point spread function	n/a
a_x	Radius of aperture opening number x	m
C_x^y	Direction cosine matrix from x to y	n/a
C_a	Approximate diameter of focal blur for clear aperture	m

xii

Symbol	Description	Units
C_n	Approximate diameter of focal blur for clear aperture number n	m
$C(x,y)$	Laplacian of Gaussian for image over x and y	n/a
$F(\lambda_\rho, \theta)$	Fourier transform of image in polar coordinates λ_ρ and θ	n/a
$\mathcal{F}(x)$	Fourier transform of x	n/a
f_l	Focal length of lens	m
$f_i(\rho, \theta)$	Image in polar coordinates ρ and θ	m
g	Gravitational force	$\frac{m}{s^2}$
$G_\sigma(x,y)$	Gaussian kernel over x and y with standard of deviation σ	n/a
$G(\rho, \sigma)$	Gaussian kernel over ρ with standard of deviation σ	n/
H_x^z	Observation model of x in measurement z	n/a
$\mathcal{H}$	Optical transfer function	n/a
$h(x,y)$	Convolution matrix over x and y	n/a
$I(x,y)$	Image over coordinates x and y	n/a
I_{psd}	Power spectral density	$\frac{V^2}{Hz}$
I_{psf}	Intensity Point Spread Function	n/a
L	Likelihood	n/a
m_s	Magnification of optical system	n/a
$N(0, \psi)$	Zero mean normal distribution with covariance ψ	n/a
P_e	East position error	m
P_n	North position error	m
P_u	Up position error	m

Symbol	Description	Units
$P_{x,x}$	Covariance of x	n/a
$P(\eta,\xi)$	Pupil function along Cartesian coordinates η and ξ	n/a
$\mathcal{P}(\eta,\xi,s_a)$	Generalized pupil function along η and ξ for feature at distance s_a	n/a
p^a	Position of feature in scene in navigation frame	m
R_0	Local Earth radius	m
R_x	Measurement noise of x	n/a
$\mathbb{R}[x]$	Real portion of x	n/a
s_a	Distance from lens to feature in scene	m
s_d	Distance from lens to focal plane in scene	m
s^f	Vector from camera to feature in camera frame	m
$\underline{s}^c$	Pointing vector from camera to feature in camera frame	m
T_c^{pix}	Translation matrix from pointing vector to pixel plane coordinates	n/a
T_{pix}^c	Translation matrix from pixel plane coordinates to pointing vector	n/a
$W(\eta,\xi,s_a)$	Aberration function along η and ξ for feature at distance s_a	n/a
$Y(\nu,\omega)$	Coded image over coordinates ν and ω	n/a
z_a	Distance from lens to focused image near pixel plane	m
z_d	Distance from lens to pixel plane	m
z_i	Distance from lens to image plane	m

List of Abbreviations

Abbreviation	Description
2D	Two Dimensional
3D	Three Dimensional
4D	Four Dimensional
ASA	Angle-Side-Angle
DARPA	Defense Advanced Research Agency
GENMAV	Generic Micro Aerial Vehicle
GPS	Global Positioning System
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
LIDAR	Light Detection and Ranging
MAV	Micro Aerial Vehicle
MITE	Micro Tactical Expendable
OTF	Optical Transfer Function
PDF	Probability Density Function
PSD	Power Spectral Density
PSF	Point Spread Function
PVA	Position Velocity Attitude
SIFT	Scale-invariant Feature Transform
SLAM	Simultaneous Location and Mapping

VISION AIDED INERTIAL NAVIGATION SYSTEM AUGMENTED WITH A CODED APERTURE

1. Introduction

Loss of access to the Global Positioning System (GPS) through intentional or unintentional interference from or obstruction by urban structures is a significant challenge for a modern vehicle navigation system to overcome [40]. When this challenge is anticipated, a favored solution is reliance on a self-contained inertial navigation system (INS). When an INS is set or initialized to correct vehicle location and pose estimates, navigation errors are often small such that they may be neglected for short periods of time; however errors grow over time such that an uncorrected INS becomes largely unusable [6, 16]. An INS is often supplemented by complimentary sensor systems to correct navigation errors such that they remain relatively small. Camera-based sensing systems have been used to provide correction to an INS by exploiting the persistence of image features that are detectable from different perspectives [6, 39]. These systems use multiple camera images to estimate constraints to either the navigation solution or the slant range to the features [6, 39].

This research proposes correcting an INS solution using a depth from defocus camera system that captures both heading and slant range to each feature in a single image, thereby enabling estimation of the location of the feature relative to the vehicle with just one camera. Also, this research proposes using a Fresnel zone plate as an aperture coding to improve the performance for a depth from defocus vision system. To

the best of this author's knowledge, this research is the first to propose using depth from defocus for navigation, using a coded aperture and depth from defocus for navigation, and using a Fresnel zone plate as an aperture for a depth from defocus system. The resulting first of its kind vision aiding system will provide robust INS correction, allowing greater non-GPS navigation performance. This chapter presents a top level summary of material that is more thoroughly covered in Chapters Two through Five.

1.1 Conventional Vision Aided INS.

1.1.1 Navigation With a Map. Traditional vision systems used to aid an INS employ cameras with an infinite depth of field to minimize focal errors when detecting significant image features. When the slant range to a feature is large relative to the movement of the vehicle, such as a high altitude air vehicle capturing images of land for navigation, the features may be treated as static landmarks of the local topography. In some instances, a map of the locations of uniquely identifiable landmarks may be acquired a priori with a significant degree of precision. A single camera image may then be used to provide heading information to detected landmarks relative to the frame of the air vehicle, but the distance from the camera to the landmark must be inferred. If the locations of multiple uniquely identified landmarks are known (for example, a set of roadway intersections or lakes) then the distance between the landmarks are also known. From distance and heading information to landmarks with known locations, the location and orientation of the air vehicle may be determined [16, 39].

1.1.2 Navigation Without a Map. Often a map of uniquely identified landmarks with known locations does not exist. The locations of observed landmarks may be estimated relative to the location of the aircraft, with the noteworthy limitation that the location of the camera is itself an estimate. This problem of estimating a map of the location of landmarks and the location at which images of the landmarks are captured is commonly known as simultaneous location and mapping (SLAM) [31]. A single image from a traditional camera provides heading information from the camera to a landmark, constraining the possible locations that the landmark may occupy. Without slant range, there is not sufficient information to fully define a location estimate of a landmark relative to the camera, and the location of the landmark is only defined up to scale [11]. Common approaches to fully define a landmark location estimates involve use of multiple images from one or more standard cameras, where each image further constrains possible landmark locations [11].

1.1.2.1 Methods with Multiple Cameras. If feature locations are unknown but multiple cameras are available, then the vector separating the locations of each camera may be known precisely. Knowledge of this vector allows stereopsis techniques to be employed. With stereopsis, the angle from the focal point of each camera to a given feature is known, and the distance between the focal points is known [11]. Triangular congruency then emerges as an Angle-Side-Angle (ASA) problem, where the "side" is defined by the vector separating the two camera locations and the "angles" are the heading measurements from each camera to a given landmark. Because

the vector separating the two cameras is known a priori, the distance between any focal point and the feature may be fully determined.

1.1.2.2 Methods with a Single Camera.

If a single camera is available, then multiple images used to constrain the landmark location estimate may be obtained from the single camera at different times when the camera is in motion. Triangular congruency again emerges as an ASA problem; however the vector separating the two camera locations is not known a priori. Instead, the two camera locations are determined by solutions to the navigation system. This method of correcting the INS is ill-posed however as the triangular congruency is derived from the INS solution that the measurements are intended to correct. Using the previous estimated location of the landmark and an INS prediction of vehicle movement, a prediction may be made as to where the landmark will appear in the subsequent image [39]. Similar systems offer improved performance by incorporating depth as an inverse rather than directly [13, 21, 27].

1.1.2.3 Light Detection and Ranging.

Another method of determining range is through use of light detection and ranging (LIDAR) [3]. Laser range scanners emit light in a given direction and measure properties of the light backscatter [3]. For example, if a reflection of a pulse of light is detected, and the transmission velocity of the light pulse through the air is assumed, then the total distance traveled by the reflected light is estimated from the total time of flight [3]. As an active sensor, LIDAR requires significantly more power than does a passive camera. Also,

whereas a camera may use a global shutter to capture all pixel data simultaneously, non-flash LIDAR normally scans each point sequentially [2]. The sequential scanning of each point introduces distortion and bias error due to movement of the laser resulting from vehicle movement and vibration [2].

1.2 Depth Measurement from Focal Blur.

1.2.1 Depth from Defocus.
Rather than comparing images from multiple from cameras with an infinite depth of field, an alternative technique to estimate depth is to compare multiple images from a single camera with a varying depth of field. Depth to a feature is then determined from the amount of defocus, which is used to estimate the distance to a feature by observing the change in blurring of that feature as the focal length is varied [30]. Using a camera with a conventional rounded aperture, the technique requires multiple images from the same camera perspective in order to distinguish blurring due to defocus from the appearance of blurring due to naturally occurring smooth gradients of the captured image [22]. To meet the requirement that the images be derived from the same perspective, either the camera should remain effectively stationary or an optical arrangement devised such that multiple pixel planes capture an image simultaneously with multiple focal lengths [41]. To the best of this author's knowledge, depth from defocus techniques have not previously been proposed for use in navigation systems.

1.2.2 Augmenting with a Coded Aperture.
The replacement of the rounded aperture of the conventional camera with a coded aperture, as shown in

Figure 1-1, replaces the normally smooth defocus blur with a structured defocus blur. This structure enables separability of defocus blur from naturally occurring smooth image gradients within a single image. Because of this separability, depth from defocus using a camera with coded aperture has been shown to provide a means of estimating the angle and distance from a camera to observed features in a single image [22]. The ability to capture range in a single image removes the requirement that multiple camera images be captured from the same perspective. The challenge then is to establish a correspondence between the a priori focal blur patterns for various distances and the focal blur observed in the image. Chapter Three provides a method to predict the focal blur for a given aperture coding and lens system, to measure the focal blur, and to design the camera system to select the focal blur for a given range. Chapter Four describes a method to measure range from the coded image and characterizes the measurement noise for various apertures and scenarios.
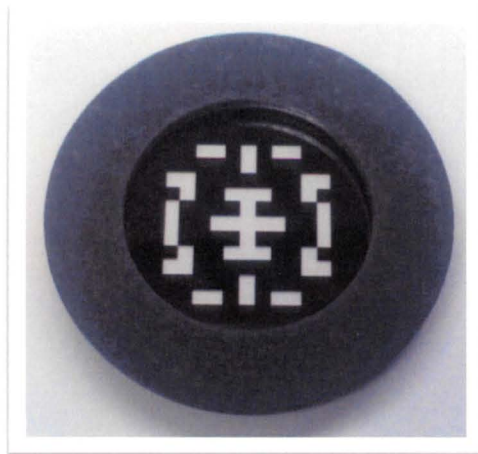
Figure 1-1 Coded Aperture to Replace Traditional Rounded Aperture

1.2.3   Fresnel Zone Plate for Depth from Defocus Aperture

Coding.   The Fresnel zone plate is a set of clear and opaque annular rings, as shown in Figure 1-2, spaced in such a manner as to create a diffractive lens with multiple focal points [12]. When paired with a refractive lens, the focal blur from a Fresnel zone plate aperture changes in a unique manner with distance, and the multiple focal points each contribute to the total focal blur [19]. This research is the first to propose and demonstrate the Fresnel zone plate aperture improving the performance of a coded aperture depth from defocus system by enhancing correspondence determination between the a priori focal blur patterns for various distances and the focal blur observed in the image. Chapter Three describes and analyzes the focal blur from a defocused lens and a Fresnel zone plate aperture, and Chapter Four describes a method to measure range from the coded image and characterizes the measurement noise for various apertures and scenarios.

Figure 1-2 Fresnel Zone Plate Aperture as Coded Aperture

1.3 Proposed Fresnel Zone Plate Aperture Vision Aided INS.

1.3.1 Coded Aperture for Slant Range Measurement.   This research proposes using depth from defocus with a coded aperture to measure slant range in a single image for aiding inertial navigation systems. No navigation system to date uses any depth from defocus techniques for navigation. Using the slant range information from a coded aperture depth from defocus system enables estimating landmark location relative to the vehicle from a single image. Because the relative landmark location measurements are independent of the navigation solution, observed changes in the landmark location relative to the vehicle provide well-posed corrections to the inertial navigation system. This research describes the depth observation model and the covariance of the observation noise to augment the extended Kalman filter for inertial navigation system correction.

1.3.2 Related Efforts.   Depth from defocus using a coded aperture camera is related to other recent efforts to recover depth information from a single camera that involve capturing the four dimensional (4D) light field of the two dimensional (2D) image [14, 22, and 38]. A 4D light field describes all light passing through a given point in space, and a conventional image is the 2D projection of the 4D light field that is incident to the camera pixel plane [38]. From the 4D light field, the depth from defocus technique is more easily solved as the change in blur at different distances from the pixel plane can be inferred. Cameras with a micro lens array (called plenoptic cameras), capture the 4D light field by redirecting sections of the optical path to separate locations of the pixel plane, resulting in multiple lower resolution images from

slightly different perspectives [14, 43]. Dappled photography is a similar to the plenoptic camera approach except that a cosine mask captures a Fourier transform of each image at various angles rather than low resolution images [38]. Multiple coded images have also been used, with each image identical except using a different aperture coding, to improve the depth measurement performance [7, 45]. None of these techniques have been proposed for use in navigation systems.

1.4 Application of Proposed System for MAVs.

An example application of a Fresnel zone plate aperture aided INS is autonomous indoor navigation of micro aerial vehicles (MAVs). DARPA's Grand Challenge highlighted and prompted progress in overall autonomous navigation by demonstrating several automobiles autonomously navigating across a 142 mile course [36]. Like many automobiles participating in DARPA's Grand Challenge, Stanford's winning vehicle used a combination of the Global Positioning System (GPS), an Inertial Measurement Unit (IMU), and a suite of several cameras and laser range finders for navigation [36]. Following was the similar success of DARPA's Urban Grand Challenge in which automobiles navigated autonomously through a city environment with traffic [24, 36]. Along with this growth of ground based systems, a number of MAVs have also recently been developed, such as Air Force Research Laboratory's GENMAV [34], Honeywell's Micro Air Vehicle [26], U.S. Naval Research Laboratory's MITE [20], and University of California Berkeley's micromechanical flying insect [10]. While the Grand Challenge vehicles demonstrate the rising maturity of autonomous navigation technology for automobiles, indoor autonomous navigation of MAVs face additional challenges of more

complex processing of fewer sensors and a need to minimize sensor size and weight requirements. Also, GPS is often not available or performs poorly in indoor environments [25].

The significantly smaller size and weight of a MAV necessitates judicious use of both limited payload of the sensors and sensor power requirements. The generally higher power requirements of active sensors, such as LIDAR, render such sensors less desirable than passive sensors. While INS performance determines in part the amount of time that the INS location estimates are sufficient, smaller systems necessary for a MAV are accurate for significantly shorter time periods than those more commonly used for larger systems [9]. Aiding an INS using stereopsis between multiple cameras provides significant correction to the INS solution [39]; however a single camera vision system is also preferential over a multi-camera system for the smaller vehicle. Several vision aided INS proposals have been shown to perform well if a topographic map is available [5], the image data constrains the navigation solution [6], or the slant range to the location points is estimated [39]. For an indoor environment, reliance upon a map is impractical due to moveable occlusions such as furniture and doorways as well as shadowing due to changes in illumination. Using visual information to constrain the navigation solution during indoor navigation has been shown to be susceptible to significant attitude errors over a short period of travel (e.g. one minute) [5]. Also, when estimating slant range to a feature, an incorrect range estimate has been shown to significantly degrade the corrective performance of the vision aided system [39]. The proposed coded aperture vision aiding of the INS requires no map, estimates slant range rather than attempting to

constrain the navigation solution, and does not translate errors in the uncorrected INS solution to the slant range estimate. The proposed navigation system offers superior performance for indoor MAV navigation.

The raw images from a camera with a defocused lens with or without a coded aperture would include blurring that may not be acceptable for other uses of the images such as surveillance. Some of the image blur may be removed through deconvolution after the depths in the scene are determined, however artifacts of the defocus may remain. A solution would be to forgo the depth from defocus system and instead use two-camera stereoscopy to correct the INS. However, if only one of these two cameras were defocused for navigation, then the other camera could capture focused images and be used for non-navigation purposes as desired. This would free the focused camera to tilt, pan, or zoom to capture images of interest in the scene without coordination with the defocused camera and without losing the navigation solution.

1.5 Organization of Document

Chapter One provides a high-level overview of the entire work and describes the motivation for the effort. Chapter Two provides the background material for the research presented. First, inertial navigation is discussed, followed by a description of methods to correct an inertial navigation solution using a vision system. Next, standard techniques for determining depth in an image are presented. Finally, defocus of an image for a general aperture is explained with special consideration for a Fresnel zone plate aperture.

Chapter Three presents analysis and design of the coded aperture system. A method of modeling an arbitrary aperture at an arbitrary depth is presented first. The model is then applied to the previously generated Levin coded aperture and then to a newly proposed Fresnel zone plate aperture, followed by comparisons of the model to real measurements. The changes in the focal blur relative to distance are analyzed, followed by a discussion of behaviors specific to the Fresnel zone plate.

Chapter Four discusses depth measurement techniques for a coded image. Several different measurement methods are evaluated using two different aperture codings and a traditional rounded aperture. The measurement noise from the various methods are evaluated and characterized for the different apertures.

In Chapter Five, the noise characterized in Chapter Four is further analyzed and modeled for integration into the Kalman filter of a navigation system. The method for integrating the depth measurements into the navigation is given, and the results of simulating the vision aided navigation with the popular aperture and Fresnel zone plate are compared with each other and with other vision aided navigation methods. Chapter Six then presents the conclusions of the overall effort.

2. Background

Navigation through indoor environments is a key challenge associated with autonomous micro aerial vehicles (MAVs). While current GPS technology is sufficient for systems in areas with clear access to GPS signals, it is often not available within indoor environments. An alternative navigation method is a visually aided inertial navigation system. One approach to implementing a visually aided system is to use sensors to measure the navigation environment to aid in determining the system position, velocity, and attitude at any given time.

This chapter will provide an overview of inertial navigation and the effect of various sources of error on the navigation solution. Integration of a vision system using a conventional camera is then described to correct the inertial navigation system. Next, the optics of focal error with an aperture coding is presented, followed by a summary of some unique optical properties of the Fresnel zone plate as both a lens and an aperture. Finally, an overview of depth determination from imagery is given that includes stereopsis, depth from defocus, and coded aperture techniques.

2.1 Inertial Navigation System

2.1.1 Inertial Navigation System Overview.

An INS includes a collection of accelerometers and gyros. Figure 2-1 shows a basic description of a two-dimensional inertial navigation system with two accelerometers and one gyroscope, whereas the three-dimensional system would include three accelerometers and three gyroscopes [16]. The accelerometers and

gyroscope are attached to a rigid body frame, and the gyroscope detects rotations about the perpendicular plane by being oriented orthogonally to a pair of accelerometers.
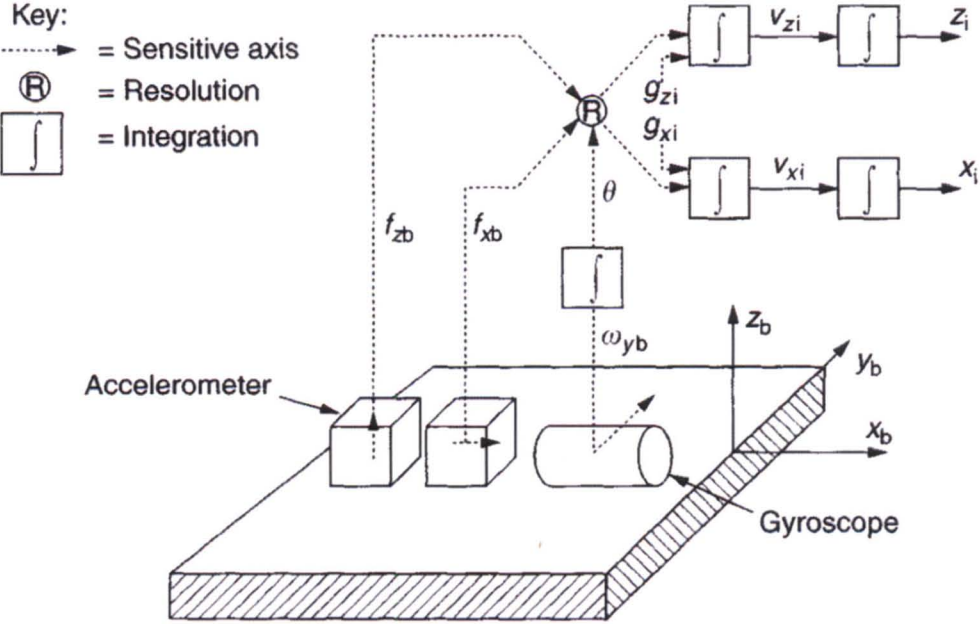
Figure 2-1 Two-Dimensional INS Description [37]

Integration of accelerometer outputs f_{zb} and f_{xb} of Figure 2-1 results in velocity estimates V_{zi} and V_{xi}, and integration of the velocity estimates results in position estimates z_i and x_i [37]. Integration of the accelerometer output ω_{yb} produces the angular displacement θ of the body frame from the inertial navigation system's reference frame, as shown in Figure 2-2.

The system of linear equations that describe two-dimensional inertial navigation system, with gravitational force g, local Earth radius R_0, and distance to center of local Earth radius sphere z, are shown in Equations (2-1) through (2-7) [37]:
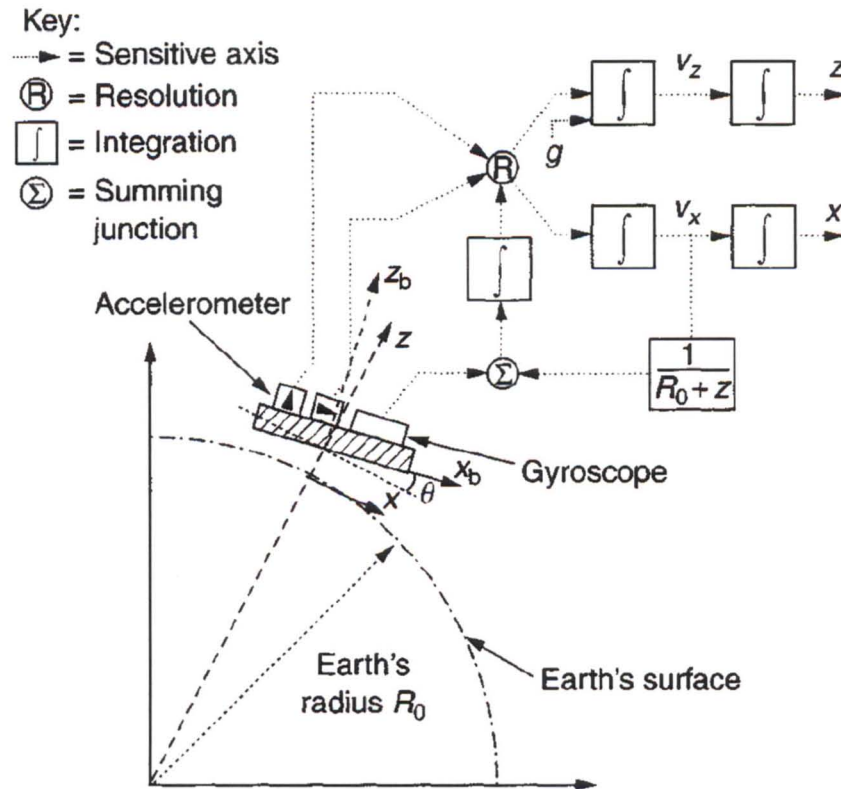
Figure 2-2 Two-Dimensional INS Block Diagram [37]

$$\dot{\theta} = \omega_y^b - v_x / (R_0 + z) \tag{2-1}$$

$$f_x^i = f_x^b \cos\theta + f_z^b \sin\theta \tag{2-2}$$

15

$$f_z^i = -f_x^b \sin\theta + f_z^b \cos\theta \tag{2-3}$$

$$\dot{v}_x^i = f_x^i + v_x^i v_z^i / (R_0 + z) \tag{2-4}$$

$$\dot{v}_z^i = f_z^i + g - \left(v_x^i\right)^2 / (R_0 + z) \tag{2-5}$$

$$\dot{x} = v_x^i \tag{2-6}$$

$$\dot{z} = v_z^i \tag{2-7}$$

The value θ is the body attitude, ω_y^b is the angular rate measured by the gyroscope aligned to the y-axis of the body, f_k^i is the specific force in the k axis of the inertial frame, f_k^b is the specific force in the k axis of the body frame, and v_k^i is the velocity in the k axis of the inertial frame.

2.1.2 Inertial Navigation System Error Model. This section will describe the INS error model for this work, which includes error in altitude, angular rate, position and velocity. For INS error notation, Ψ represents attitude error, Ω represents angular rate error, p represents position error, and v represents velocity error. For superscript and subscript notation, i indicates representation in the inertial reference frame, b indicates representation in the body reference frame, and n indicates representation in the navigation reference frame.

Beginning with attitude error modeling, the direction cosine matrix C_b^n describes the rotation from the body frame to the navigational reference frame. Misalignments of sensors within the INS can be characterized by $\tilde{C}_b^n = BC_b^n$, where $\tilde{C}_b^n$ is an estimation of C_b^n and B includes the misalignments [37]. For small misalignments, B can be defined as $B = [I - \Psi]$ where a Ψ is given by Equation (2-8).

$$\Psi = \begin{bmatrix} 0 & -\delta\gamma & \delta\beta \\ \delta\gamma & 0 & -\delta\alpha \\ -\delta\beta & \delta\alpha & 0 \end{bmatrix} \tag{2-8}$$

The parameters $\delta\alpha$ and $\delta\beta$ are tilt errors and $\delta\gamma$ is the azimuth error [37]. The relations for Equations (2-9) through (2-12) hold [37].

$$\tilde{C}_b^n = BC_b^n \tag{2-9}$$

$$\tilde{C}_b^n = [I - \Psi]C_b^n, \text{ by substitution} \tag{2-10}$$

$$\Psi = I - \tilde{C}_b^n \left(C_b^n\right)^T, \text{ rearrangement} \tag{2-11}$$

$$\dot{\Psi} = -\dot{\tilde{C}}_b^n \left(C_b^n\right)^T - \tilde{C}_b^n \left(\dot{C}_b^n\right)^T, \text{ by differentiation} \tag{2-12}$$

Continuing with the attitude error modeling, the body angular rate, Ω_{ib}^b, is given by Equation (2-13) for small values with respect to the inertial frame as a function of gyroscope measurements ω_z, ω_y, and ω_x [37].

17

$$\Omega_{ib}^b = \begin{bmatrix} 0 & -\omega_z & \omega_x \\ \omega_z & 0 & -\omega_y \\ -\omega_x & \omega_y & 0 \end{bmatrix} \tag{2-13}$$

The navigation frame angular rate, Ω_{in}^n, is similarly defined for small values with respect to the inertial frame [37]. Note that Ω_{ib}^b, Ω_{in}^n, and C_b^n are used to determine $\dot{C}_b^n$ by the relation $\dot{C}_b^n = C_b^n \Omega_{ib}^b - \Omega_{in}^n C_b^n$ [37]. Given estimates of Ω_{ib}^b and Ω_{in}^n that are $\tilde{\Omega}_{ib}^b$ and $\tilde{\Omega}_{in}^n$ respectively, the estimate $\dot{\tilde{C}}_b^n$ is then $\dot{\tilde{C}}_b^n = \tilde{C}_b^n \tilde{\Omega}_{ib}^b - \tilde{\Omega}_{in}^n \tilde{C}_b^n$. Equations (2-14) through (2.21) show application of this relation in determining $\dot{\Psi}$, where $\delta f(x)$ is defined as $\tilde{f}(x) - f(x)$ [37].

$$\dot{\Psi} = -\tilde{C}_b^n \tilde{\Omega}_{ib}^b \left(C_b^n \right)^T + \tilde{\Omega}_{in}^n \tilde{C}_b^n \left(C_b^n \right)^T + \tilde{C}_b^n \Omega_{ib}^b \left(C_b^n \right)^T - \tilde{C}_b^n \left(C_b^n \right)^T \Omega_{in}^n \tag{2-14}$$

$$\dot{\Psi} = -\tilde{C}_b^n \left[\tilde{\Omega}_{ib}^b - \Omega_{ib}^b \right] \left(C_b^n \right)^T + \tilde{\Omega}_{in}^n \tilde{C}_b^n \left(C_b^n \right)^T - \tilde{C}_b^n \left(C_b^n \right)^T \Omega_{in}^n \tag{2-15}$$

$$\dot{\Psi} = -\left[I - \Psi \right] C_b^n \left[\tilde{\Omega}_{ib}^b - \Omega_{ib}^b \right] \left(C_b^n \right)^T + \tilde{\Omega}_{in}^n \left[I - \Psi \right] C_b^n \left(C_b^n \right)^T - \left[I - \Psi \right] C_b^n \left(C_b^n \right)^T \Omega_{in}^n \tag{2-16}$$

$$\dot{\Psi} = -\left[I - \Psi \right] C_b^n \left[\delta \Omega_{ib}^b \right] \left(C_b^n \right)^T + \left[\tilde{\Omega}_{in}^n - \tilde{\Omega}_{in}^n \Psi \right] - \left[\Omega_{in}^n - \Psi \Omega_{in}^n \right] \tag{2.17}$$

$$\dot{\Psi} = -\left[I - \Psi \right] C_b^n \delta \Omega_{ib}^b \left(C_b^n \right)^T + \delta \Omega_{in}^n + \Psi \Omega_{in}^n - \tilde{\Omega}_{in}^n \Psi \tag{2.18}$$

$$\dot{\Psi} = -\left[I - \Psi \right] C_b^n \delta \Omega_{ib}^b \left(C_b^n \right)^T + \delta \Omega_{in}^n + \Psi \Omega_{in}^n - \left[\delta \Omega_{in}^n + \Omega_{in}^n \right] \Psi \tag{2.19}$$

$$\dot{\Psi} = -\left[I - \Psi \right] C_b^n \delta \Omega_{ib}^b \left(C_b^n \right)^T + \delta \Omega_{in}^n - \delta \Omega_{in}^n \Psi + \Psi \Omega_{in}^n - \Omega_{in}^n \Psi \tag{2.20}$$

18

$$\dot{\Psi} = -[I - \Psi]C_b^n \delta\Omega_{ib}^b (C_b^n)^T + \delta\Omega_{in}^n [I - \Psi] + \Psi\Omega_{in}^n - \Omega_{in}^n \Psi \qquad (2.21)$$

For position and velocity error modeling, the velocity rate may be expressed as $\dot{v} = C_b^n f^b - (2\omega_{ie}^n + \omega_{en}^n) \times v + g_l$, where v is the velocity, f^b is specific force coordinatized in the body axes, g_l is the portion of the specific force due to gravity, ω_{ie}^n is the turn rate of the Earth with respect to the inertial, and ω_{en}^n is the turn rate of the Earth with respect to the navigation frame [37]. Equation (2-22) is then an estimate for velocity rate , and the estimate error is given in Equation (2-23) [37].

$$\dot{\tilde{v}} = \tilde{C}_b^n \tilde{f}^b - (2\tilde{\omega}_{ie}^n + \tilde{\omega}_{en}^n) \times \tilde{v} + \tilde{g}_l \qquad (2-22)$$

$$\delta\dot{v} = \dot{\tilde{v}} - \dot{v} = \tilde{C}_b^n \tilde{f}^b - C_b^n f^b - (2\tilde{\omega}_{ie}^n + \tilde{\omega}_{en}^n) \times \tilde{v} + (2\omega_{ie}^n + \omega_{en}^n) \times v + \tilde{g}_l - g_l \qquad (2-23)$$

Equation (2-24) incorporates the relations $\delta f(x) = \tilde{f}(x) - f(x)$ and $\tilde{C}_b^n = [I - \Psi]C_b^n$ [37].

$$\delta\dot{v} = -\Psi C_b^n + C_b^n \delta f^b - (2\omega_{ie}^n + \omega_{en}^n) \times \delta v - (2\delta\omega_{ie}^n + \delta\omega_{en}^n) \times v - \delta g \qquad (2-24)$$

The position error can be expressed as Equation (2-25) [37].

$$\delta\dot{p} = \delta v \qquad (2-25)$$

Error rates due misalignments have been shown, as well as position and velocity error rates due to error in estimates of specific force, gravity, and the turn rate of the Earth. Sensor

19

information may be used to correct attitude, position and velocity errors. The next section describes such a system using a multiple cameras to correct the inertial navigation system.

2.2 Vision Aided Inertial Navigation System

This section provides an overview of the foundation vision aided inertial navigation system, proposed by Veth in [39], to be augmented in this work. Veth's system uses two cameras and stereoscopy to aid the navigation system [39]. Chapter 5 will present the performance of Veth's proposed navigation system using both cameras, and the substantial loss of performance that occurs when one camera is removed. Chapter 5 will also show the significant recovery of much of the lost performance when the remaining camera is augmented with a coded aperture.

Veth's system uses observations of features in images to provide correction to the navigation solution produced by an INS. In Veth's vision system, two-camera stereoscopy is used to solve for depth in the initial identification of features from which corrections will be obtained. From the two images, the location of each feature in the navigation frame is estimated. As the vehicle moves, the location of each feature is predicted in subsequent images and the predictions are compared to the feature locations observed in images captured from the cameras. The differences in observed and predicted feature locations, as measured in pixels, are then used to correct both the vehicle's position, velocity and attitude (PVA) estimate and the location of each feature in the navigation frame. Depth estimates are assumed to be Gaussian distributed and zero mean. Inaccurate depth estimates produce linearization errors that bias INS and feature location correction. [39]

As illustrated in Figure 2-3, an INS and a vision system provide inputs to a Kalman filter in Veth's proposed system. The Kalman filter uses the vision system measurements to estimate the errors in the inertial navigation system. The INS can then be corrected by these error estimates. [39]
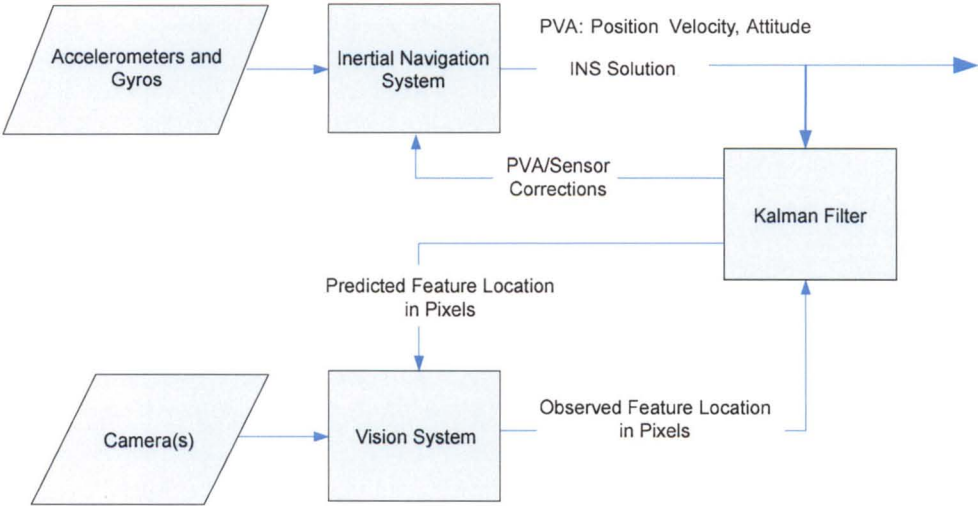
Figure 2-3 System Design of Vision Aided INS [39]

For each feature currently tracked, Veth's navigation system maintains a location estimate and descriptor to aid in identifying the feature in subsequent images. When selecting new features to track, the features with the greatest feature quality are chosen. The reader is referred to [39] for a description of the feature quality metric and descriptor.

Veth's vision system receives images from the camera and a list of tracked feature location estimates in the camera frame from the Kalman filter. The vision system uses the feature descriptor and location estimate to establish correspondence from the tracked features to the features in the observed image. For each feature in the camera frame for which

21

correspondence to a tracked feature is established, a two dimensional homogenous pointing vector is found. The homogenous pointing vectors are submitted to the Kalman filter as new measurements of the direction to their corresponding tracked features. [39]

The Kalman filter incorporates the INS solution and the feature location measurements from the vision system to create an optimal estimate of the PVA and feature location in the navigation frame in Veth's system. The feature location estimates and covariance propagate from the Kalman filter to the vision system. [39]

In the image aided INS system of Veth's proposed system, the influence matrix of the Kalman filter is created using the homogenous pointing vector to give the direction from the camera to the identified feature because the vector is known only up to scale. As an example, linearization of the homogenous pointing vector with respect to the position of the vehicle in the navigation frame is described below; however the reader is referred to [39] for a full description of the Kalman filter.

The vector from the camera to the feature is s^c, where c indicates that the vector is represented in the camera frame. Equation (2-26) presents the linearization of s^c with respect to the position of the vehicle in the navigation frame, p^n. The rotation matrix from the navigation frame to the camera frame is C_n^c. [39]

$$\frac{\delta s^c}{\delta p^n} = -C_n^c$$

(2-26)

22

Because the true value of s^c is not observable from a single image measurement, linearization about the homogenous pointing vector, $\underline{s}^c$, is determined by Equation (2-27). [39]

$$\frac{\delta \underline{s}^c}{\delta p^n} = \lfloor s^c \rfloor_3 \left(\frac{\delta s^c}{\delta p^n} \right) - \frac{s^c \left\lfloor \dfrac{\delta s^c}{\delta p^n} \right\rfloor_3}{\left(\lfloor s^c \rfloor_3 \right)^2}$$

(2-27)

The notation $\lfloor \cdot \rfloor_3$ indicates that only the third component of the vector, or the third column of a matrix, is used. Given matrix T_c^{pix} which transforms $\underline{s}^c$ to pixel plane coordinates, Equation (2-28) shows the influence of the feature location on the vehicle position estimate. [39]

$$H_{p^n}^z = T_c^{pix} \frac{\delta \underline{s}^c}{\delta p^n}$$

(2-28)

Linearization of additional Kalman filter states uses a similar method as shown in Equations (2-27) and (2-28) by using the partial derivative of s^c with respect to the state [39]. The reader is referred to [39] for a more complete description of applying partial derivatives of s^c in the Kalman filter.

It was shown by [27] that significant improvements may be made to the systems such as the one proposed by [39] by representing depth as an inverse rather than directly. Systems proposed by [13] and [21] demonstrate similar monocular configurations similar to [39] that represent depth as an inverse.

2.3 Image Depth Determination.

Depth is a measure of the distance between a camera lens and features or scene segments observed in the image. Determining depth is a classic challenge in the field of computer vision, and several methods have been developed to solve the depth determination problem. This section discusses a set of methods of determining depth from one or more images. Of the methods presented, only stereopsis has been previously proposed for use in a navigation system.

2.3.1 Stereopsis.

Stereopsis is the practice of extracting depth information from a scene given images taken from different perspectives. The perspective of the images must be such that points of interest in the one image correspond to points in another image. Figure 2-4 illustrates an example of a pair of parallel image planes from which scene depth may be inferred. The focal point of the camera for each image is separated by a baseline of length B, the face edges of the box in the right image are shown in red, and the face edges of the box in the left image are shown in green. The displacement of an edge between the left and right image is the disparity (d), and the depth to the edge from the baseline (h) is found by $\frac{-B}{d}$.

The key to finding depth through stereopsis is determining which feature in one image corresponds to the same feature in another image. This is commonly referred to as the correspondence problem. One technique to solve the correspondence problem is the SIFT algorithm [23]. The first part of the SIFT algorithm is to approximate a Laplacian pyramid using the difference of Gaussian method. Rather than lowering the resolution of the image at each

level of the pyramid, the resolution is kept constant and the bandwidth parameter σ is scaled instead. The next step is to find keypoints in the pyramid. Points that are locally maximum or locally minimum with respect to the 26 neighboring pixels in the volume of the pyramid and exceed a given threshold are marked as candidate keypoints. Keypoints are also pruned if they are determined to be edge points rather than corner points [23].

From the remaining keypoints, rotation invariant descriptor orientations must be found. Descriptor orientation is found by accumulating the gradient magnitudes of pixels in the pyramidal neighborhood of the keypoints into a histogram of gradient orientation. A threshold is applied and each orientation peak of the remaining histogram is given a unique descriptor. A curve is then fit to histogram values surrounding each peak to estimate the orientation of maximum magnitude [23]. A more detailed description of the SIFT algorithm can be found in [23].

Once the keypoints and orientations are found, the SIFT descriptors can be formed. The neighborhood gradient orientation is rotated with respect to the found descriptor orientation to achieve rotation invariance. The neighborhood gradient magnitudes are scaled by a Gaussian window centered on the keypoint to mitigate edge effects. The neighborhood is then segmented into multiple regions and an orientation histogram is made of the gradient magnitudes for each region. Each region's histogram is then normalized to a unit vector, and then saturated to a maximum value of β where $\beta < 1$, and finally renormalized to a unit vector [23]. With the correspondence problem solved by the SIFT algorithm, the disparity in a set of stereoscopic

images can be determined. If the baseline between the images can also be discerned, then an estimate of the range to feature points in the scene may also be established.

Height (h)

Baseline (B)

Disparity (d)
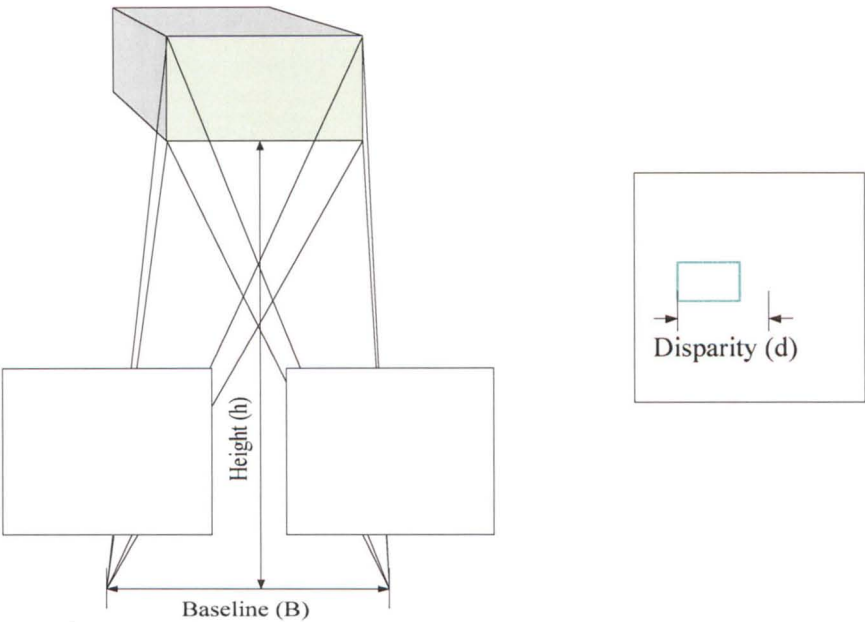
Figure 2-4 Stereopsis Example

2.3.2 Spatial Phase Imaging.

The spatial phase sensor was developed under Nichols Research Corporation and is currently part of the Spatial Phase Video Camera produced by a company called Photon-X [1]. The spatial phase sensor places an arrangement of polarizers and wave retarders behind a lenslet array, allowing both intensity and the phase of light incident to the pixel plane to be determined [1]. Figure 2-5 shows the arrangement of the polarizers on the pixel plane, with a dark box grouping a set of polarizers in the upper left corner that form a "super" pixel, from which phase and intensity determination is made [1]. From the polarization data, a Stokes vector is

determined, where a Stokes vector is a four element real valued vector describing the polarization state of light [1]. According to Fresnel's Laws, the polarization of light reflected from a surface is a function of the angle of the light that is incident upon that surface [1]. This phenomenon is similar to changes in shading that would also be observable in light reflected from the surface [1]. When observing surfaces, the gradients in the polarization of the reflected light are a function of gradients in the orientation of the surfaces relative to the pixel plane. The super pixels allow observation of changes in light polarization along the three-dimensional contours of such surfaces, provided the surfaces are both continuous and non-occluded. Changes in depth from the camera to various points along the observed surface may then be determined from the changes in the polarization of the reflected light; however, the absolute range to any given point cannot be determined by this method alone.
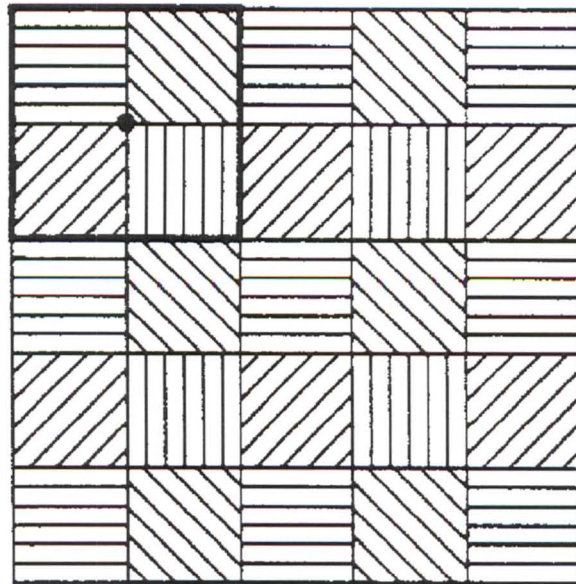
Figure 2-5 Illustration of Polarization Arrangement and Super Pixel [1]

2.3.3 Depth from Multiple Defocused Images.

Another method to determine depth in a scene is through analysis of defocused images. The original approach presented by Pentland [30] is to compare two images of the same scene taken at two focal lengths, then evaluate the focal gradient for each point in the image to determine depth. Pentland's original proposal compared a defocused image to a perfectly focused image as would be obtained using a pinhole aperture, then using spatial filtering to estimate depth [30, 8]. Two means of spatial filtering were proposed that include using feature points in a single image and comparing multiple images of the same view of the scene using different apertures. Both of these methods are ill-posed in that small errors induced into the data produce large fluctuations in the answer [8].

When determining depth using feature points, it is assumed that the location in the image of the feature point is known and that the point spread function is the primary cause of information surrounding the feature point. Let $C(x, y)$ be the Laplacian of Gaussian for image data $I(x, y)$. Then Equation (2-29) describes evaluation in the x direction [30].

$$C(x, y) = \delta \frac{dG_\sigma(x, y)}{dx} = \left(\frac{-\delta x}{\sqrt{2\pi\sigma^3}} \right) e^{\left(\frac{-x^2}{2\sigma^2} \right)} \qquad (2\text{-}29)$$

To find $\hat{\sigma}$, a maximum-likelihood estimate of σ, a linear regression in x^2 of the form $Ax^2 + B = C$ is found using the natural log $\ln \left| \frac{C(x, y)}{x} \right| = \ln \left(\frac{\delta x}{\sqrt{2\pi\sigma^3}} \right) - \left(\frac{x^2}{2\sigma^2} \right)$. Then $A = -\frac{1}{2\sigma^2}$, $B = \ln \left(\frac{\delta}{\sqrt{2\pi\sigma^3}} \right)$ and $C = \ln \left| \frac{C(x, y)}{x} \right|$. Equations (2-30) through (2-32) solve for $\hat{\sigma}$ letting $\bar{x}$ be the

28

mean value of x_i, and $\bar{C}$ be the average value of C [30]. The two resulting solutions for $\hat{\sigma}$ correspond to solutions for the two possible distances $s'_a < s_a$ and $s'_a > s_a$ that could have produced $G_{\hat{\sigma}}(x,y)$.

$$A = \frac{\sum_i \left(x_i^2 - \bar{x}^2\right) C_i}{\sum_i \left(x_i^2 - \bar{x}^2\right)^2} \tag{2-30}$$

$$B = \bar{C} - \bar{x}^2 A \tag{2-31}$$

$$\hat{\sigma} = (-2A)^{-\frac{1}{2}} \tag{2-32}$$

The second form of spatial filtering involves the comparison of multiple images of the same view of a scene with different apertures. In this method, $f_i(\rho,\theta)$ is defined as a polar coordinate description of an image region i, with (ρ,θ) as a coordinate frame with (x_0,y_0) as the center. The intensity due to a given an image point source that is maximum at $\rho=0$ and trails from the center point by a scaled two-dimensional Gaussian $G(\rho,\sigma)$. The value of $f_i(\rho,\theta)$ for a point source is found by the relation $f_i(\rho,\theta)=I_i(x_0+\rho\cos\theta, y_0+\rho\sin\theta)G(\rho,\sigma)$. Letting $f_0(\rho,\theta)$ be a perfectly focused image, then the relation between any pair of images of different aperture size is found by Equation (2-33) [30].

$$\frac{f_1(\rho,\theta)}{f_2(\rho,\theta)} = \frac{f_0(\rho,\theta) \otimes G(\rho,\sigma_1)}{f_0(\rho,\theta) \otimes G(\rho,\sigma_2)} \tag{2-33}$$

Letting $F_i(\lambda_p, \theta)$ be the two-dimensional Fourier transform of $f_i(\rho, \theta)$, the two-dimensional Fourier transform of $f_i(\alpha_s \rho, \theta)$, with α_s as a scaling factor, is of the form

$\dfrac{1}{|\alpha_s|} F\!\left(\dfrac{\lambda_p}{\alpha_s}, \theta\right)$. The relation between $F_i(\lambda_p, \theta)$ and $F_0(\lambda_p, \theta)$ may be approximated by the equation

$F_i(\lambda_p, \theta) = F_0(\lambda_p, \theta) G\!\left(\lambda_p, \dfrac{1}{\sqrt{2\pi}\sigma_i}\right)$ [30]. An integration of the Fourier transform over θ produces

$F_i(\lambda_p) = \int_{-\pi}^{\pi} F_i(\lambda_p, \theta)\, d\theta$, allowing the convolution of $\dfrac{f_1(\rho,\theta)}{f_2(\rho,\theta)}$ to be solved in the Fourier domain by

Equations (2-34) through (2-36) [30].

$$\frac{f_1(\rho,\theta)}{f_2(\rho,\theta)} = \frac{f_0(\rho,\theta) \otimes G(\rho,\sigma_1)}{f_0(\rho,\theta) \otimes G(\rho,\sigma_2)} \tag{2-34}$$

$$\frac{F_1(\lambda_p)}{F_2(\lambda_p)} = \frac{G\!\left(\lambda_p, 1/\sqrt{2\pi}\sigma_1\right)}{G\!\left(\lambda_p, 1/\sqrt{2\pi}\sigma_2\right)} \tag{2-35}$$

$$\frac{G\!\left(\lambda_p, 1/\sqrt{2\pi}\sigma_1\right)}{G\!\left(\lambda_p, 1/\sqrt{2\pi}\sigma_2\right)} = e^{\lambda_p^2 2\pi^2 \left(\sigma_2^2 - \sigma_1^2\right)} \tag{2-36}$$

To find maximum likelihood estimates $\hat{\sigma}_1$ and $\hat{\sigma}_2$, a linear regression in λ_p^2 of the form

$A\lambda_r^2 = C$ is found using the natural log $\lambda_p^2 2\pi^2 \left(\sigma_2^2 - \sigma_1^2\right) = \ln F_1(\lambda_p) - \ln F_2(\lambda_p)$ [30]. Then

$A = 2\pi^2 \left(\sigma_2^2 - \sigma_1^2\right)$ and $B = n F_1(\lambda_p) - \ln F_2(\lambda_p)$.

Like stereoscopy, a spatial filtering approach to depth from defocus is fundamentally a triangulation problem [32]. Whereas stereoscopy introduces distance to the triangulation problem through the baseline distance between image pairs and range is measured by the disparity of pixel locations between the images, spatial filtering introduce distance to the triangulation problem through the radius of the lens and range is measured by the disparity of blur kernel sizes between images [32]. Because stereoscopy involves matching points in one image to corresponding points in another image, and depth from defocus involves matching blur kernels in one image to blur kernels in another image, both techniques also involve solving a correspondence problem [32].

A more accurate approach to determining depth from defocus is through matrix-based deconvolution, in which the problem is characterized as a system of linear equations [8]. Two images of the same view are captured at different levels of defocus. The point spread function for each defocused image is represented as convolution matrices $h_1(x, y)$ such that a perfect focus image, $I_0(x, y)$, when convolved with a matrix $h_1(x, y)$ or $h_2(x, y)$, produces observed image $I_1(x, y)$ and $I_2(x, y)$, respectively [8]. A convolution ratio matrix $h_3(x, y)$, when convolved with observed image $I_1(x, y)$, also produces $I_2(x, y)$ [8]. The set of linear equations is then presented in Equations (2-37) and (2-38) [8].

$$I_0(x,y) \otimes h_1(x,y) \otimes h_3(x,y) = I_0(x,y) \otimes h_2(x,y) \qquad (2\text{-}37)$$

$$h_1(x,y) \otimes h_3(x,y) = h_2(x,y) \qquad (2\text{-}38)$$

31

Matrices $h_1(x, y)$ and $h_2(x, y)$ may be determined experimentally and used to solve for a table of $h_3(x, y)$ patterns [8]. When depth is to be determined, the $h_3(x, y)$ pattern for which convolution with $I_1(x, y)$ produces $I_2(x, y)$ is selected to indicate the depth.

Traditional depth from defocus assumes a rounded aperture shape. A Gaussian distribution is often used to approximate the focal blur from a rounded aperture, and both high frequency content and edge strength decrease as the standard deviation of the blur's Gaussian shape increases [11, 15].

2.3.4 Depth a from Single Image using Coded Aperture.

A key limitation of the traditional depth from defocus technique is that multiple images of a scene must be captured that are almost identical except for a change in the amount of defocus. Although traditional depth from defocus is well suited for scenes that remain static as the camera optics change between image captures, this limitation may be prohibitive when applied to the dynamic scenes encountered by a camera aiding navigation. This section describes Levin's proposal for a novel depth from defocus technique using a coded aperture that estimates depth through evaluation of a single image rather than by comparison of multiple images [22]. The key limitation is thus removed, allowing determination of both direction and depth to features in the scene from a single camera on a moving platform such as a micro air vehicle.

The replacement of the rounded aperture on the conventional camera with a coded aperture replaces the normally smooth defocus blur with a structured defocus blur. Figure 2-6 shows examples of a rounded defocus blur from a conventional camera and the structured

defocus blur from a camera with a coded aperture. This structure enables separability of defocus blur from naturally occurring smooth image gradients within a single image. Because of this separability, depth from defocus using a coded aperture provides a means of estimating the angle and distance from the camera to observed features in a single image. Depth is discernible from the image because, for a portion of the scene observed at a given range, the coded aperture should introduce zeros in the spatial frequency domain. The location of the zeros is a function of the given range. [22]
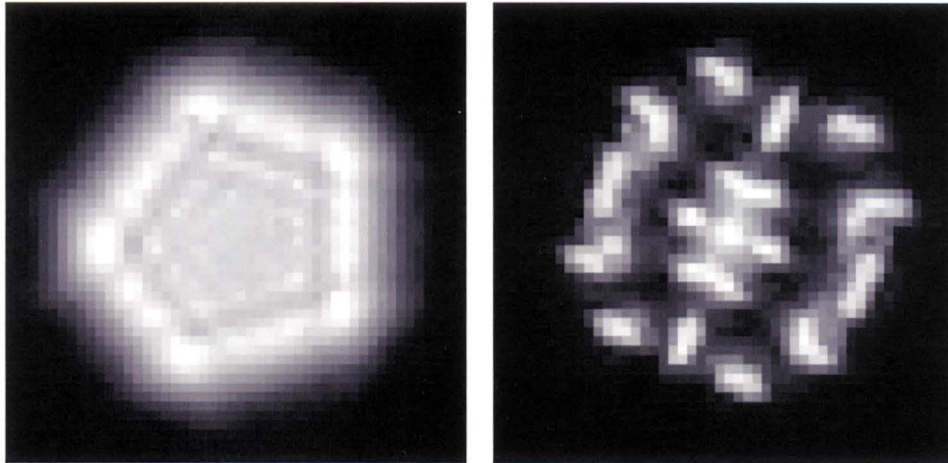
Figure 2-6 Typical (left) and Coded (right) Aperture Point Spread Functions [22]

The technique proposed by considers the image as a distribution of derivatives that can be characterized as $N(0, \psi)$. By also noting that, at a given range, the coded aperture will convolve a specific point spread function with the all focus image, the resulting image will be of a Gaussian distribution of point spread functions. The method Levin suggests for determining range is to evaluate Equation (2.39), where $Y(v, \omega)$ is the Fourier transform of the coded

33

aperture image, $\sigma_r(v,\omega)$ is the Fourier domain response of the coded aperture vision system to a point light source at range r, and L is a likelihood. The value of r that maximizes L is selected as the measured range. [22]

$$L = \exp\left(-\frac{1}{2}\sum_{v,\omega}\left(\left.|Y(v,\omega)|^2\middle/\sigma_r(v,\omega)\right)\right)\right)$$
(2.39)

2.4. Apertures and Defocus.

This section describes focal blur as a function of the depth in the scene and the aperture of the camera. In Chapters Three and Four of this work, the relationships presented here are further analyzed in the modeling and designing of the optical portion of a coded aperture augmented vision system.

Focal blur in a camera with a narrow depth of field and a coded aperture may be characterized by the point spread function, $I_{psf}(s_a)$, and optical transfer function, $\mathcal{H}(s_a)$, for a point source at a given distance s_a. First, the general effect of the coded aperture on the I_{psf} and $\mathcal{H}$ on an optical system will be described, then $I_{psf}(s_a)$ and $\mathcal{H}(s_a)$ will be considered with focal error added.

The shape of an aperture is observed in the pupil function, $P(\eta,\xi)$, which describes complex amplitude transmittance of the lens and aperture [15]. Here η and ξ are Cartesian coordinates on the surface of the aperture facing the pixel plane [15]. For an aperture with only opaque and transparent openings, $P(\eta,\xi)$ is one for transparent aperture locations and zero for opaque aperture locations. The amplitude point spread function, A_{psf}, describes the complex spreading of the light's amplitude from the scene's point sources to the pixel plane. The function A_{psf} is found by the Fourier transform of $P(\eta,\xi)$, as given by Equation (2.40) [15].

$$A_{psf} = \mathcal{F}\left(P(\eta,\xi)\right)$$
(2.40)

The intensity point spread function, I_{psf}, is the intensity of the light, and is therefore real and the normalized squared magnitude of A_{psf} as given by Equation (2.41), when normalized such that I_{psf} sums to one [15]. The function I_{psf} may be measured directly by capturing images of point sources of light in an otherwise darkened environment

$$I_{psf} = \left|A_{psf}\right|^2$$
(2.41)

When observing a natural scene, the resultant image is a convolution of the light incident upon the lens from each point of the illuminated surfaces of the environment, and I_{psf} for the location of the points in the scene relative to the camera [15].

The optical transfer function, $\mathcal{H}$, is the normalized autocorrelation of the transfer function for A_{psf}, and is related to I_{psf} through Equation (2.42), with the final result normalized to a maximum value of one [15].

$$\mathcal{H} = \mathbb{R}\left[\mathcal{F}\left(I_{psf} \right) \right]$$

(2.42)

Geometrically, $\mathcal{H}$ is the normalized area of overlap, $\mathcal{A}$, between pupil functions $P(\eta, \xi)$ and $P(-\eta, -\xi)$ centered at $(\lambda z_i f_\eta / 2, \lambda z_i f_\xi / 2)$ and $(-\lambda z_i f_\eta / 2, -\lambda z_i f_\xi / 2)$ respectively [15]. This relation is expressed in Equation (2.43), where λ is the light wavelength and z_i is the distance from the lens to the image plane. $\mathcal{H}$ may be measured by capturing the I_{psf}, and changes in $\mathcal{H}$ with respect to changes in s_a may be observed by changing the distance to the light source [15].

$$\mathcal{H}(f_\eta, f_\xi) = \frac{\iint P\left(\eta + \frac{\lambda z_i f_\eta}{2}, \xi + \frac{\lambda z_i f_\xi}{2} \right) P\left(\eta - \frac{\lambda z_i f_\eta}{2}, \xi - \frac{\lambda z_i f_\xi}{2} \right) d\eta d\xi}{\iint P(\eta, \xi) d\eta d\xi}$$

(2.43)

For a camera with a narrow depth of field focused at distance s_d, the I_{psf} becomes a function of s_a, and $I_{psf}(s_a)$ for a given value of s_a may be measured by capturing an image of a point source at a distance s_a from the lens.

Focal error can be modeled as an aberration across the surface of the aperture. This aberration is described in Equation (2.44), where $W(\eta, \xi, z_a)$ is the aberration error, z_a is the

distance behind the lens at which the image is formed, and z_i is the distance from the lens to the image plane [15].

$$W(\eta,\xi,z_a)=\left(-\frac{1}{2}\right)\left(\frac{1}{z_a}-\frac{1}{z_i}\right)(\eta^2+\xi^2)$$

(2.44)

Note that z_a can be found for a given s_a for a lens with focal length f_l by Equation (2.45).

$$z_a=\frac{s_a f_l}{s_a-f_l}$$

(2.45)

Substituting Equation (2.45) into Equation (2.44) gives the focal error with respect to s_a in Equation (2.46).

$$W(\eta,\xi,s_a)=\left(-\frac{1}{2}\right)\left(\frac{1}{s_d}-\frac{1}{s_a}\right)(\eta^2+\xi^2)$$

(2.46)

Aberrations may be modeled as phase distortions that combined with $P(\eta,\xi)$ to form the complex pupil function, $\mathcal{P}(\eta,\xi,s_a)$, given in Equation (2.47) [15].

$$\mathcal{P}(\eta,\xi,s_a)=P(\eta,\xi)e^{jkW(\eta,\xi,s_a)}$$

(2.47)

The focal error propagates to $\mathcal{H}(s_a)$ over the area of overlap by Equation (2.48) [15].

$$\mathcal{H}\left(f_{\eta}, f_{\xi}, s_a\right) = \frac{\iint\limits_{A(f_\eta, f_\xi)} e^{j\frac{2\pi}{\lambda}\left[W\left(\eta + \frac{\lambda z_i f_\eta}{2}, \xi + \frac{\lambda z_i f_\xi}{2}\right) - W\left(\eta - \frac{\lambda z_i f_\eta}{2}, \xi - \frac{\lambda z_i f_\xi}{2}\right)\right]} d\eta d\xi}{\iint\limits_{A(0,0)} d\eta d\xi} \qquad (2.48)$$

2.5 The Fresnel Zone Plate.

Chapters Three through Five of this work will show that optical properties of the Fresnel zone plate provide performance advantages when used as an aperture for a coded aperture augmented vision system. An overview of the optical properties of the zone plate are described in this section when the zone plate is used as a diffractive lens or when used as an aperture with a refractive lens.

The Fresnel zone method was introduced by Fresnel's classic work in 1826 [12]. The first zone plate was constructed by Lord Rayleigh in 1871, and extensive visible light experiments started with R. W. Wood in 1898 [28, 42]. The Fresnel zone plate is a set of annular rings, as shown in Figure 2-7 and described by Equation (2.49), where k is a scale factor and ρ is the distance from the center [15].

$$P(\rho) = \frac{1 + \text{sgn}\left(\cos\left(k\rho^2\right)\right)}{2} \qquad (2.49)$$

38
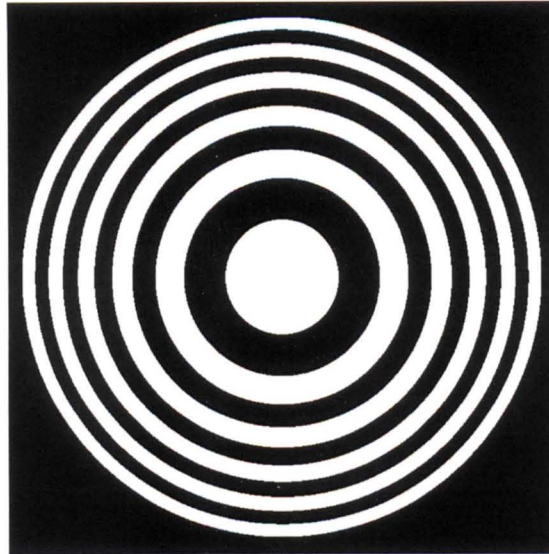
Figure 2-7 Fresnel Zone Plate with Eleven Zones

2.5.1 The Fresnel Zone Plate Lens.

The Fresnel zone plate acts as a diffractive lens with multiple focal points spaced at a_1^3/λ, $a_1^3/3\lambda$, $a_1^3/5\lambda$, etc., where a_1 is the radius of the central circle, and λ is the wavelength of the light [42]. The Fresnel zone plate shares several optical properties with the well known refractive lens, but uses diffraction of waves rather than refraction [17].

Table 2-1 compares several optical properties of the refractive lens with the zone plate, however the zone plate also acts as a lens for x-rays, sound waves, and electromagnetic frequencies for which an equivalent refractive lens may not exist [28, 44].

The Fresnel zone plate also possesses multiple focal lengths, with the lengths described by Equation (2.50) [28]. Here n refers to the higher order of the lens, for which there are no equivalent higher orders for a refractive lens, and a_n refers to the diameter to the edge of a zone.

$$\frac{1}{f_l} = \frac{1}{a_n - n\lambda} - \frac{1}{a_n + n\lambda} \tag{2.50}$$

Only a portion an incident wave may then be focused at any given focal length, and the relative amount of a wave focused at each focal length is determined as A_n in Equation (2.51), where A_n is equal to zero for even values of n [19].

$$A_n = \frac{\sin\left(\frac{n\pi}{2}\right)}{n\pi} \tag{2.51}$$

Table 2-1 Comparison of Refractive and Diffractive Lenses [28].

Property	Refractive Lens	Fresnel Zone Plate
Chromatism	Present	Pronounced
Image Forming	$\frac{1}{f_l} = \frac{1}{s_d} + \frac{1}{z_d}$	$\frac{1}{f_l} = \frac{1}{s_d} + \frac{1}{z_d}$
Higher Orders	None	$1, (2), 3, (4), 5, \ldots$

2.5.2 The Fresnel Zone Plate Aperture.

Using a Fresnel zone plate with a refractive lens was previously explored by [19] with the goal of using the Fresnel zone plate's multiple focal points to extend the depth of field in a camera. Although [19] does not consider, as this research effort proposes, using the Fresnel zone plate to induce measureable modulations in the focal blur as a means to improve depth measurement performance, this section summarizes the analysis from [19] of the change in the optical transfer function as a function of scene depth.

To analyze the function $\mathcal{H}(s_a)$ of an optical system having a Fresnel zone plate pupil, consider the generalized pupil function $\mathcal{P}$. Function $\mathcal{P}$ can be written as Equation (2.52), where N is the number of zones, ρ is the radial spatial frequency, and z is measured in units of depth of field. By changing to Cartesian coordinates (η, ξ) and considering the trace $(\eta, \xi = 0)$, the resultant $\mathcal{H}$ may be expressed by Equation (2.53), where function $\phi(\eta', \eta)$ is given by Equation (2.54). Using Equation (2.52) and assuming $|\mathcal{P}| \leq 1$, the product in Equation (2.54) is given by Equation (2.55) with amplitudes A_n and A_m found by Equation (2.56). [19]

$$\mathcal{P}(\rho, z) = \sum_n \exp\left[i4\pi(z+nN)\rho^2\right]\mathrm{circ}(\rho) \tag{2.52}$$

$$\mathcal{H}(\eta, 0, z) = \int \phi(\eta', \eta)\exp\left[-i8\pi z\eta'\eta\right]d\eta' \tag{2.53}$$

$$\phi(\eta', \eta) = \int \mathcal{P}\left(\eta' + \frac{\eta}{2}, \xi'\right)\mathcal{P}^*\left(\eta' - \frac{\eta}{2}, \xi'\right)d\xi' \tag{2.54}$$

$$\mathcal{P}\left(\eta'+\frac{\eta}{2},\xi'\right)\mathcal{P}^*\left(\eta'-\frac{\eta}{2},\xi'\right)=$$

$$\sum_{m,n}A_mA_n\exp\left\{i4\pi N\left[(m-n)\left(\eta'^2+\xi'^2+\frac{\eta^2}{4}\right)\right]+(m+n)\eta'\eta\right\}\mathrm{rect}\left(\xi'\right)\mathrm{rect}\left(\frac{\eta'}{1-\eta}\right)\quad(2.55)$$

$$A_n=\frac{\sin\left(\frac{n\pi}{2}\right)}{n\pi}\qquad(2.56)$$

The ξ'^2 term in Equation (2.55) only contributes to $\phi(\eta',\eta)$, and consequently $\mathcal{H}$, when $m=n$ [19]. Recall that a geometrical interpretation of $\mathcal{H}$ is as the normalized area of overlap between pupil functions $P(\eta,\xi)$ and $P(-\eta,-\xi)$ centered at $\left(\lambda z_i f_\eta/2,\lambda z_i f_\xi/2\right)$ and $\left(-\lambda z_i f_\eta/2,-\lambda z_i f_\xi/2\right)$ respectively [15]. The overlap of identical Fresnel zone plates displaced by a distance r_d produces straight Schuster fringes perpendicular to the direction of the displacement and with a fundamental wavelength of $\left|a_1^2/r_d\right|$, where a_1 is the diameter of the innermost zone plate opening [18]. In Equations (2.53) through (2.55), these identical Fresnel zone plates and the associated straight Schuster fringes are present for $m=n$ when using a zone plate of four or more zones [19]. Figure 2-8 illustrates the straight Schuster fringes observable in overlapping identical zone plates. In the formation of $\mathcal{H}$, the summation of the contribution from values for $m\neq n$ tend to average out, while the Schuster fringes from values for $m=n$ produce large values in $\mathcal{H}$ for certain amounts of defocus [19, 33]. Zone plates with fewer than four zones are circular or annular pupils that lack such Moiré features [19].

The function $\mathcal{H}$ may be approximated using only the $m = n$ terms of Equation (2.55) by the much simpler form of Equation (2.57), where z is a unit of defocus, n is the zone number, N is the total number of zones in the aperture, ρ is radial spatial frequency, and amplitude A_n is found by Equation (2.56) [19].
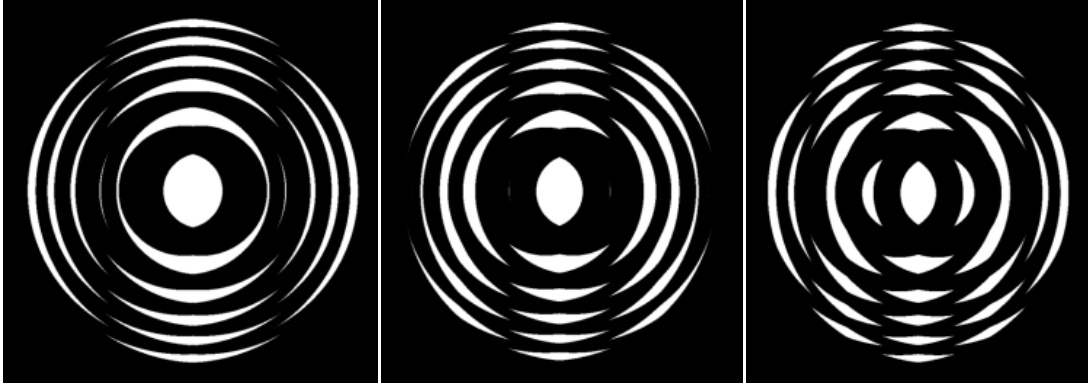
Figure 2-8 Schuster Fringes from Overlapping Fresnel Zone Plates

$$\mathcal{H} \sim \sum A_n^2 \delta(z - nN) \otimes \frac{\sin[4\pi z \rho (1-\rho)]}{4\pi z \rho}$$

(2.57)

Although Equation (2.57) provides insight into the overall behavior of $\mathcal{H}(s_a)$ for a Fresnel zone plate, Equation (2.48) is a valid and more exact calculation of $\mathcal{H}(s_a)$ that accounts for interactions beyond the Schuster fringes [19, 35].

2.6 Overview of Background.

Because an inertial navigation system can be made small, light weight and low power using microelectromechanical accelerometers and gyroscopes, it is well suited for indoor

navigation of a micro air vehicle. However, the error in the navigation solution of an INS grows quickly without correction, and an example of a stereoscopic vision system is described that provides this correction. The correction is provided by comparing the relative locations of features in the scene with predictions of where those features should be if the INS solution were correct. To do this comparison using two dimensional images requires solving the ambiguity of scale, which is done in the described system using two-camera stereoscopy.

For the proposed single-camera design that uses focal blur to solve scale ambiguity, the relationships between focal blur, the camera's aperture, and the depth to points in the scene are described. The special case of the Fresnel zone plate aperture is included because the focal blur modulations can be exploited to infer depth. Previous efforts at determining depth from an image are presented that include stereoscopy, traditional depth from defocus, and the coded aperture approach. As of the date of this publication, the Fresnel zone plate aperture has not been proposed for use with a depth from defocus system, and no depth from defocus system has been proposed for use in navigation. These two proposals, and their combination of the Fresnel zone plate aperture for navigation, are original contributions of this work.

3. Coded Aperture Analysis and Design

This chapter presents an analysis of use of the coded aperture in depth-from-defocus techniques.

First a model for defocus with a generalized coded aperture is presented. The modeling includes both the point-spread function and the power spectral density of a point source observed at any given distance. The intensity point spread functions and power spectral densities of both the Levin and Fresnel zone plate apertures are measured to verify the models.

After modeling, a means of determining the overall size of the intensity point spread function is given. Because depth determination using a coded aperture occurs using a region of the image, the focal blur size determines the size of the region that must be evaluated. Also, a method of selecting a point-spread function for a given coded aperture and distance is presented, which may be used to select the maximum point-spread function size.

Finally, the characteristics of the point-spread function and power spectral density unique to the Fresnel zone plate aperture are presented. The changes in the point-spread function and power spectral density are described with respect to various focal distances and the point source depth.

3.1 Aperture Modeling

Modeling allows prediction of the performance of a given aperture when presented with a variety of scenarios. Fresnel diffraction approximation presented in Chapter Two allows

modeling of the imaging system when equipped with a general aperture shape and a point source

of light at an arbitrary distance from the lens. A variety of image noise sources are not

accounted for with this technique, such as partial occlusion of points in the scene, pixel noise,

lens distortion, lens placement misalignment, and the shape and fill factor of pixels in the pixel

plane. The model does provide insight into the overall change in the point spread function and

optical transfer function of the lens and aperture system. These insights may aid in determining

optimal design decision for a given scenario in which the system is to be employed.

3.1.1 Point Spread Function. The point spread function, $I_{psf}(s_a)$, for a light source at

distance s_a can be found for given aperture shape by first finding the general pupil function, $\mathcal{P}$.

The general pupil function may be found using Equations (2.46) and (2.47) from [15], which for

convenience are presented again here as Equations (3.1) and (3.2) accordingly. For Equation

(3.1), recall that η and ξ are coordinates along the surface of the aperture plane, s_d is the

distance to the focal plane in the scene, and s_a is the distance to the point source of interest. The

resulting $W(\eta,\xi,s_a)$ describes the complex defocus aberration across the (η,ξ) aperture surface.

$$W(\eta,\xi,s_a) = \left(-\frac{1}{2}\right)\left(\frac{1}{s_d} - \frac{1}{s_a}\right)\left(\eta^2 + \xi^2\right) \tag{3.1}$$

For Equation (3.2), recall that $P(\eta,\xi)$ is the pupil function of the aperture, such that

$P(\eta,\xi)=1$ for all values of (η,ξ) at which the pupil is transparent, and $P(\eta,\xi)=0$ for all values of

(η,ξ) at which the pupil is opaque. Pupil function values may be between zero and one if semi-

transparent materials are used, and may contain an imaginary component if the materials

introduce a phase shift in the light. Without loss of generality, this document will assume the aperture pupil may be described as fully transparent or opaque, and does not introduce an appreciable phase shift. Equation (3.2) does represent the focal error as a phase shift that, combined with the $P(\eta,\xi)$ provides complex values for $\mathcal{P}(\eta,\xi,s_a)$ across the (η,ξ) aperture surface.

$$\mathcal{P}(\eta,\xi,s_a) = P(\eta,\xi)e^{jkW(\eta,\xi,s_a)} \tag{3.2}$$

Once $\mathcal{P}$ is determined, Equations (2.40) and (2.41) from [15] may be used to find I_{psf}. For convenience, Equations (2.40) and (2.41) are presented again here as Equations (3.3) and (3.4) accordingly, however the pupil function $P(\eta,\xi)$ of (2.40) is replaced with the more general $\mathcal{P}(\eta,\xi,s_a)$.

$$A_{psf} = \mathcal{F}\left(\mathcal{P}(\eta,\xi,s_a)\right) \tag{3.3}$$

Recall that A_{psf} is the complex amplitude point spread function. From the A_{psf}, the intensity point spread function, I_{psf}, is found using Equation (3.4).

$$I_{psf} = \left|A_{psf}\right|^2 \tag{3.4}$$

The I_{psf} for a given $P(\eta,\xi)$, s_d and s_a can also be measured if the optical system is constructed. To validate a modeled I_{psf}, optical systems were created and the I_{psf} measured using point sources of light. The camera used in the validation is a Prosilica GE4900 with a 16 megapixel monochrome CCD. The pixel plane of this camera is the 35mm Kodak KAI-16000 CCD image sensor. The pixel pitch is 7.4 μm, and the sensor has 4872×3248 pixels. The lens system consisted of a Zeiss 50mm Distagon lens. A green P01 filter (filter factor four, exposure

factor 2) was attached to the lens to allow only green light to pass through the lens system, thereby reducing any chromatic aberration. The apertures consist of chrome deposited on a plate of soda lime. Chrome plated locations on the plate are considered opaque, whereas locations on the plate that are not covered in chrome are considered transparent. Two such apertures were constructed. The first aperture is similar to the aperture proposed by Levin in [22]. The image from which this aperture is created is shown as Figure 3-1. Black regions in Figure 3-1 correspond to areas of the aperture covered with chrome and for which $P(\eta,\xi)=0$. White regions in Figure 3-1 correspond to areas of the aperture not covered with chrome and for which $P(\eta,\xi)=1$.
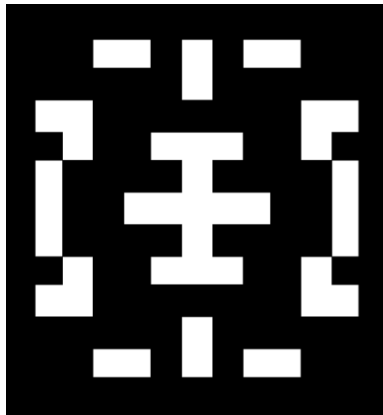
Figure 3-1 The Levin Aperture used to Validate the I_{psf} Model

The soda lime plate was attached to a washer, and the edges of the glass were light-sealed with black paint. The assembly, shown in Figure 3-2, was attached to the back of the lens with black bookbinding tape. The bookbinding tape was also used to provide a light-seal for the interface between the washer and the lens.
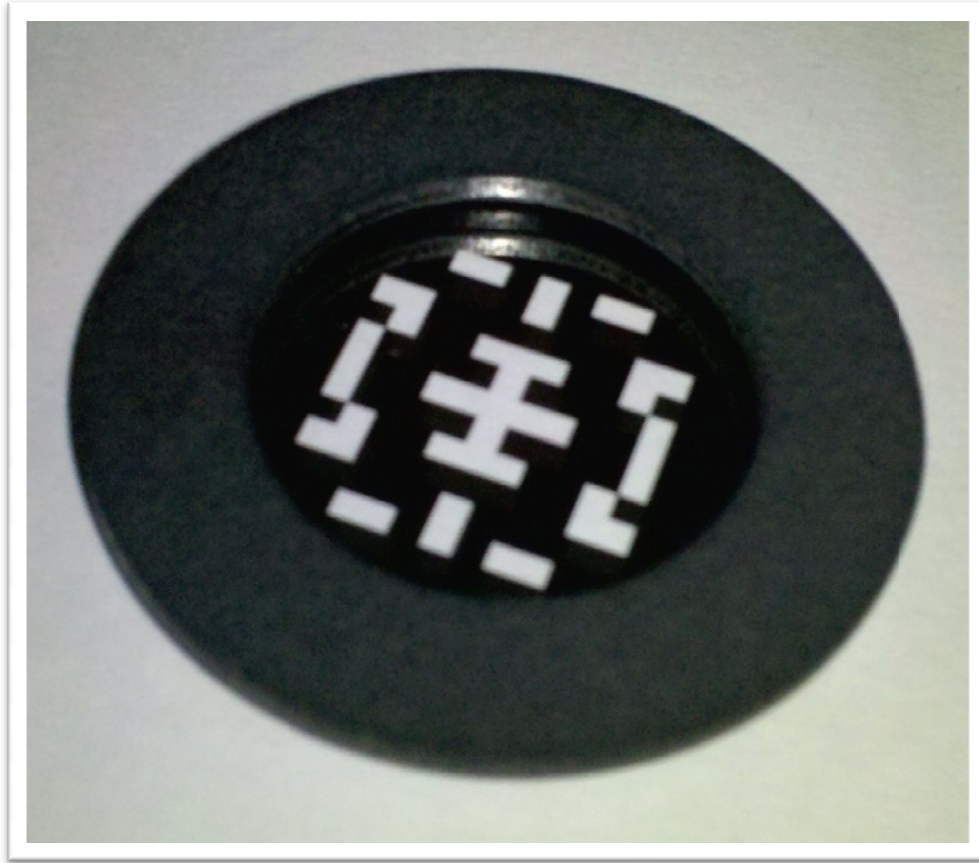
Figure 3-2 Aperture Applied to 50mm lens

After installation in the camera, it was observed that the aperture was rotated. The model of the aperture was altered to account for this rotation, and the resultant rotated image used to represent the Levin aperture in the model is shown as Figure 3-3. The orientation of the aperture does not affect depth measurement performance provided the point spread functions are correctly aligned with the physical aperture.
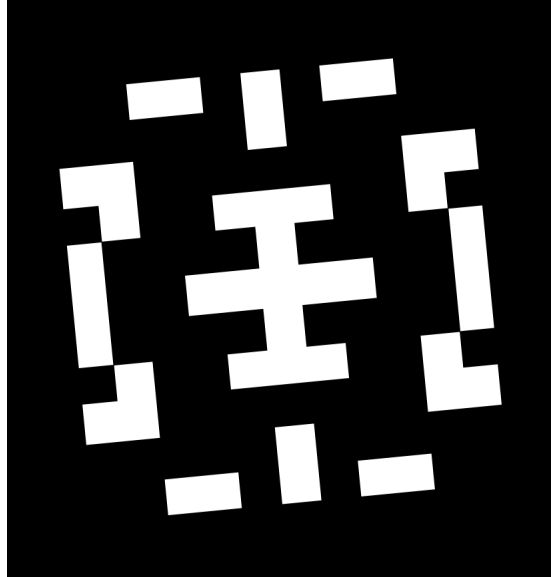
Figure 3-3 The Levin Aperture Model Rotated to Align with Physical Installation

The second aperture constructed is a Fresnel Zone plate with eleven zones. Six of the zones are transparent, while the other five are blocked by chrome. The image from which this aperture is created is shown as Figure 3-4. Black regions in Figure 3-4 correspond to areas of the aperture covered with chrome and for which $P(\eta,\xi)=0$. White regions in Figure 3-4 correspond to areas of the aperture not covered with chrome and for which $P(\eta,\xi)=1$.

As with the Levin aperture, the soda lime plate is attached to a washer and light-sealed with black paint. The Fresnel zone plate attached to the washer is shown in Figure 3-5. The washer is then attached to the back of the lens assembly, and the interface between the washer and the lens is light sealed with black bookbinding tape. Because the Fresnel zone plate is rotationally symmetric, it was not necessary to adjust the model to account for the rotation of the aperture after installation.
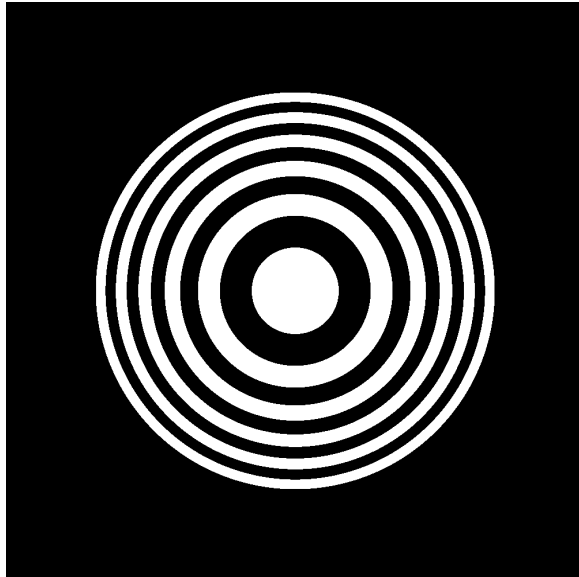
Figure 3-4 Fresnel Zone Plate Aperture

Once the lens system is constructed, for both the Levin and Fresnel zone plate aperture, the camera is focused at 1.7 m, and the built-in aperture of the 50mm lens is fully opened to an f-stop of 1.4.

A point light source is then placed in front of the lens in an otherwise darkened environment. Each point light source is created by puncturing a small hole into a sheet of aluminum foil, and then wrapping that foil around a flashlight. The interface between the foil and the flashlight is light-sealed using black bookbinding tape. The size of the hole is determined by the angular resolution of the camera such that an in-focus projection of the illuminated hole on the pixel plane is no larger than a single pixel.
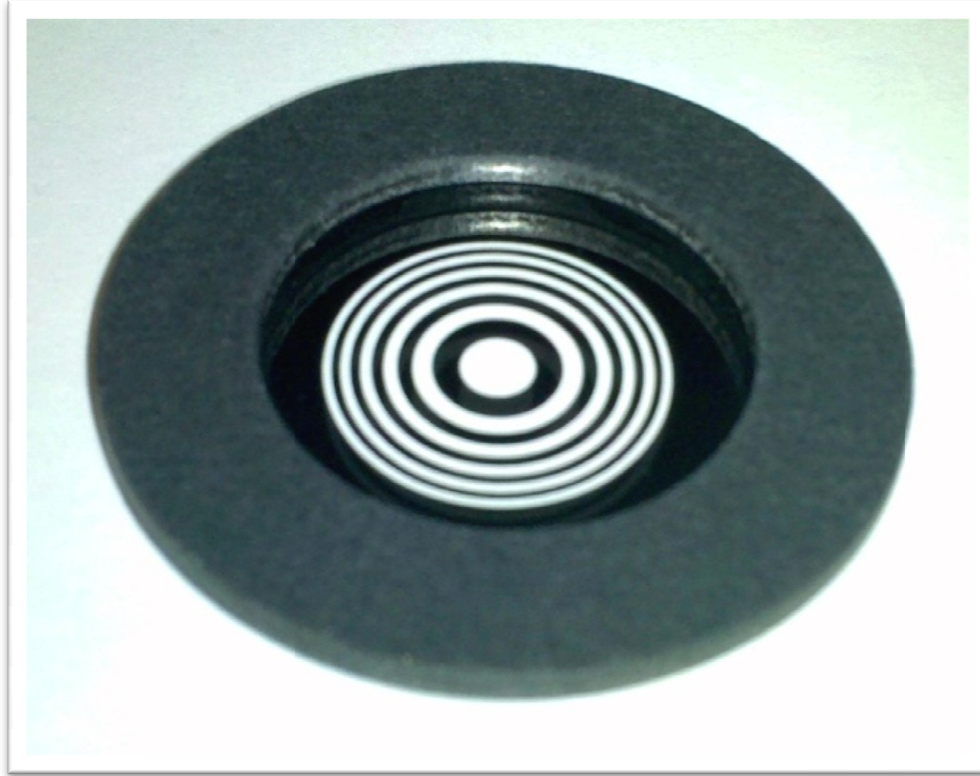
Figure 3-5 Fresnel Zone Plate Aperture to be Attached to 50mm Lens

When the point source is placed at a given distance s_a from the camera, the image produced is the I_{psf}. The recorded I_{psf} and the modeled I_{psf} may then be compared. For the Levin aperture, Table 3-1 shows a selection of measured I_{psf} distributions for various values of s_a and modeled I_{psf} distribution for the same values of s_a. Appendix A provides a complete table of measured and modeled I_{psf} distributions for a larger set of s_a values.
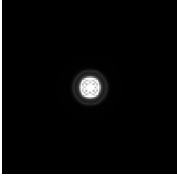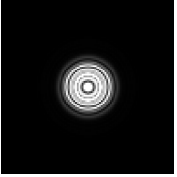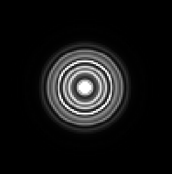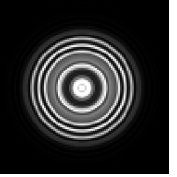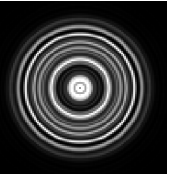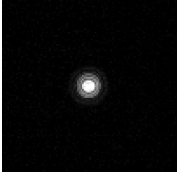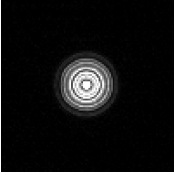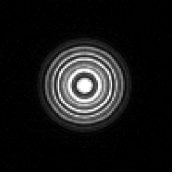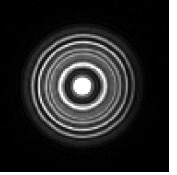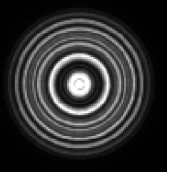
Table 3-1 Modeled and Measured Levin Aperture I_{psf} for Various Values of s_a

I_{psf} Type	Distance from lens to point source of light				
	2.022 m	2.654 m	3.859 m	7.074 m	42.449 m
Modeled					
Measured					

Although small differences may be observed upon close inspection of the modeled and measured I_{psf} distributions for given values of s_a, Table 3-1 shows that the modeled I_{psf} is a reasonable spatial domain approximation of the true I_{psf} for the Levin aperture. A portion of the differences between the modeled and measured I_{psf} may be attributed to pixel noise, lens distortion, lens placement misalignment, and the shape and fill factor of pixels in the pixel plane.

For the Fresnel zone plate aperture, Table 3-2 shows a selection of measured I_{psf} distributions for various values of s_a and modeled I_{psf} distribution for the same values of s_a. Appendix B provides a complete table of measured and modeled I_{psf} distributions for a larger set of s_a values.

Table 3-2 Modeled and Measured Zone Plate Aperture I$_{psf}$ for Various Values of s$_a$

I$_{psf}$ Type	Distance from lens to point source of light				
	2.023 m	2.656 m	3.863 m	7.084 m	42.498 m
Modeled					
Measured					

As with the Levin aperture, small differences are observable between the modeled and measured I_{psf} distributions for given values of s_a, and Table 3-2 shows that the modeled I_{psf} is a reasonable spatial domain approximation of the true I_{psf} for this Fresnel zone plate aperture. Again, a portion of the differences between the modeled and measured I_{psf} may be attributed to pixel noise, lens distortion, lens placement misalignment, and the shape and fill factor of pixels in the pixel plane.

A Fresnel zone plate aperture with several hundred zones was attempted early in this research effort. The features in this aperture were too fine in that they introduced aliasing errors that prevented matching the I_{psf} measured for a point source with the I_{psf} observed in a natural scene at the same distance. Future work should determine an optimal number of zones that can be used to maximize performance of the aperture.

3.1.2 Optical Transfer Function. The optical transfer function, $\mathcal{H}(s_a)$, for a light source at distance s_a can be found for a given aperture shape by first finding $I_{psf}(s_a)$. The optical transfer function is found using Equation (2.42) from [15], which for convenience are presented here as a function of s_a in Equation (3.5).

$$\mathcal{H}(s_a) = \mathbb{R}\left[\mathcal{F}\left(I_{psf}(s_a)\right)\right] \tag{3.5}$$

Since the $I_{psf}(s_a)$ distribution may be found by model or measurement, $\mathcal{H}(s_a)$ may likewise be found by model or measurement. To validate the $\mathcal{H}(s_a)$ model, Equation (3.5) was applied to the modeled and measured $I_{psf}(s_a)$ distributions found for the Levin and Fresnel zone plate apertures in Section 3.1.1.

Table 3-3 shows the modeled and measured $\mathcal{H}(s_a)$ distributions with the Levin aperture and a set of chosen values of s_a. For clarity, the images presented in Table 3-3 are adjusted so that the maximum one percent and minimum one percent of the image values are saturated to fully white or black. If the images were not so adjusted, the central region of each image contains such large values that the rest of each image is rendered too dark to observe. Appendix A provides a complete table of measured and modeled $\mathcal{H}(s_a)$ distributions for a larger set of s_a values.

The differences in Table 3-3 between the modeled and measured distributions of $\mathcal{H}$ for the Levin aperture are more easily observed than the differences between the modeled and measured distributions of the I_{psf} in Table 3-1. These differences result from differences in the

55

modeled and measured I_{psf} distributions, of which a portion may be attributed to pixel noise, lens distortion, lens placement misalignment, and the shape and fill factor of pixels in the pixel plane. Table 3-3 shows that the modeled $\mathcal{H}$ is a reasonable frequency domain approximation of the true $\mathcal{H}$ for the Levin aperture; however various types of image noise become more pronounced in the frequency domain.

Table 3-3 Modeled and Measured Levin Aperture $\mathcal{H}$ for Various Values of s_a
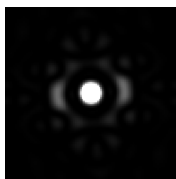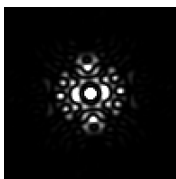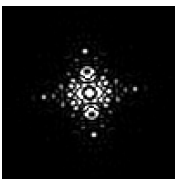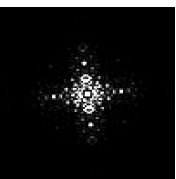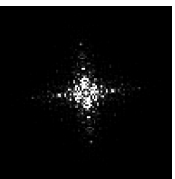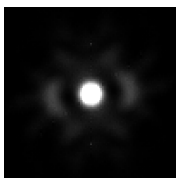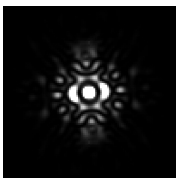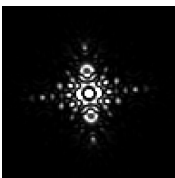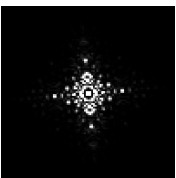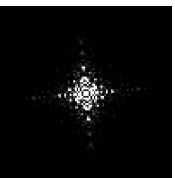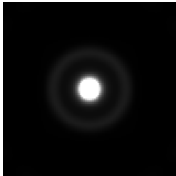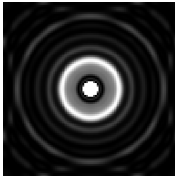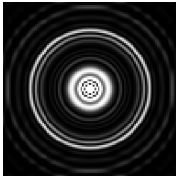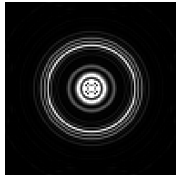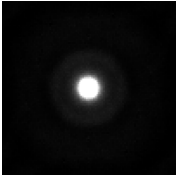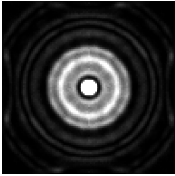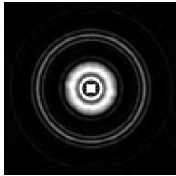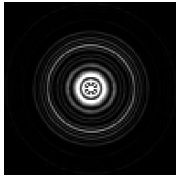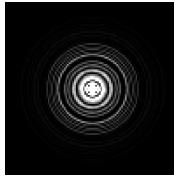
I_{psd} Type	Distance from lens to point source of light				
	2.022 m	2.654 m	3.859 m	7.074 m	42.449 m
Modeled					
Measured					

Table 3-4 shows the modeled and measured $\mathcal{H}(s_a)$ distributions with the Fresnel zone plate aperture and a set of chosen values of s_a. For clarity, the images presented in Table 3-4 are also adjusted so that the maximum one percent and minimum one percent of the image values are saturated to fully white or black. Appendix B provides a complete table of measured and modeled $\mathcal{H}(s_a)$ distributions for a larger set of s_a values.

Table 3-4 Modeled and Measured Zone Plate Aperture $\mathcal{H}$ for Various Values of s_a

I_{psd} Type	Distance from lens to point source of light				
	2.023 m	2.656 m	3.863 m	7.084 m	42.498 m
Modeled					
Measured					

As with the Levin aperture, the differences in Table 3-4 between the modeled and measured distributions of $\mathcal{H}$ for the Fresnel zone plate aperture are more easily observed than the differences between the modeled and measured distributions of the I_{psf} in Table 3-2. These differences result from differences in the modeled and measured I_{psf} distributions, of which a portion may be attributed to pixel noise, lens distortion, lens placement misalignment, and the shape and fill factor of pixels in the pixel plane. Table 3-4 shows that the modeled $\mathcal{H}$ is a reasonable frequency domain approximation of the true $\mathcal{H}$ for this Fresnel zone plate aperture. Like the Levin aperture, various types of image noise become more pronounced in the frequency domain.

3.2 Aperture Design Considerations

3.2.1 Size of Focal Blur. An increase in focal error generally increases the overall size of the observed I_{psf} distribution. For a traditional clear aperture, the I_{psf} blurred by defocus is often approximated as either a two-dimensional Gaussian or disk smoothing function. The maximum diameter of the I_{psf} for an image blurred by defocus and using a coded aperture is analogous to the diameter of the I_{psf} for a similarly sized clear aperture. The approximate diameter of the focal blur for a clear aperture, c_a, is given by Equation (3.6) [11]. In Equation (3.6), A_d is the diameter of the aperture, and m_s is the magnification of the optical system and is a function of s_d. For a point source closer to the camera than the focal plane, c_a increases without bound as $s_a \rightarrow 0$. For a point source beyond the focal plane, $c_a \rightarrow A_d m_s$ as $s_a \rightarrow \infty$.

$$c_a = A_d m_s \left(\frac{|s_a - s_d|}{s_a} \right) \qquad (3.6)$$

The value of $s_a \in (0,\infty)$ can be found from c_a by Equation (3.7). There are two values of s_a for every $c_a \in (0, A_d m_s)$, and one value of s_a for $c_a \in [A_d m_s, \infty)$. For $c_a \in (0, A_d m_s)$, there is one value of s_a smaller than s_d to satisfy Equation (3.6), and another value of s_a larger than s_d to satisfy Equation (3.6). However, neglecting other aberrations, the I_{psf} for the two values of s_a differ in that $I_{psf}(x,y)$ for a given c_a when $s_a \geq s_d$ is equivalent to $I_{psf}(-x,-y)$ for the same value

of c_a when $s_a < s_d$. If the aperture has horizontal and vertical reflection symmetry, then the I_{psf}

for the two values of s_a are equivalent.

$$s_a = \begin{cases} \dfrac{s_d A_d m_s}{A_d m_s - c_a}, & s_a \in [s_d, \infty), \ c_a \in [0, A_d m_s) \\[4mm] \dfrac{s_d A_d m_s}{A_d m_s + c_a}, & s_a \in (0, s_d), \ c_a \in (0, \infty) \end{cases} \tag{3.7}$$

For every value of s_a greater than s_d, there exists a value of s_a less than s_d with the same,

though vertically and horizontally reflected, resultant I_{psf}. For an aperture with reflective

symmetry, the I_{psf} distributions are equivalent. For an aperture without reflective symmetry,

the reflection is not observable in $\mathcal{H}$; therefore, the two values of s_a will produce the same $\mathcal{H}$.

3.2.2 Choosing Infinite Focal Blur.   Recall in Equation (2.46) from [15],

presented again here for convenience as Equation (3.8), that the aberration due to focal blur is

captured through the relation between s_a and s_d. Using Equation (3.9) to replace $\dfrac{1}{s_a}$ and $\dfrac{1}{s_d}$ with

the value k produces Equation (3.10).

$$W(\eta, \xi, z_a) = \left(-\frac{1}{2} \right) \left(\frac{1}{s_d} - \frac{1}{s_a} \right) (\eta^2 + \xi^2) \tag{3.8}$$

$$\frac{1}{s_d} - \frac{1}{s_a} = k \tag{3.9}$$

$$W(\eta, \xi, z_a) = \left(-\frac{k}{2} \right) (\eta^2 + \xi^2) \tag{3.10}$$

59

Equation (3.10) highlights that the focal blur will be the same for any values of s_d and s_a that produce the same value for k. If an I_{psf} occurs at s_{a_1} with s_{d_1} and it is preferable that the same I_{psf} occur instead at s_{a_2} with s_{d_2}, then the value of k should be found using Equation (3.9) with $s_a = s_{a_1}$ and $s_d = s_{d_1}$. Once k is found, a solution to s_{d_2} may be found by solving Equation (3.9) for s_d with $s_a = s_{a_2}$ and $s_d = s_{d_2}$. Equation (3.11) provides a value of s_{d_2} to reproduce the desired I_{psf} at the s_{a_2}, subject to the physical limitations of the lens.

$$ s_{d_2} = \frac{s_{a_2}}{k s_{a_2} + 1} \tag{3.11} $$

Using the I_{psf} modeling technique presented in Section 3.1 of this document, an I_{psf} can be modeled for a given aperture using arbitrary values of s_a and s_d. It was shown in Section 3.2.1 that the size of the focal blur when $s_a = s_d$ is zero, and that the focal blur will strictly increase in size as s_a increases from s_d, and will reach maximum when $s_a = \infty$. Then a value for s_d may be found to place a modeled I_{psf} at $s_a = \infty$ using Equations (3.9) and (3.11). Then, any $I_{psf}(s_a)$ for $s_{d_2} \leq s_a < \infty$ will be smaller in size than $I_{psf}(\infty)$ for the selected value of s_d. By selecting the $I_{psf}(\infty)$, the maximum size of the I_{psf} is then constrained, provided no point source is closer to the camera than the distance to the focal plane.

3.3 Fresnel Zone Plate Design

The Fresnel zone plate offers advantages when employed as a coded aperture for depth determination. Section 2.5.2 shows that the I_{psf} from the Fresnel zone plate may be approximated as a sum of contributions from clear apertures with differing amounts of defocus. The differing amounts of defocus are due to the multiple focal points, which then produce multiple focal planes in the scene at multiple distances of s_d. This section will show that the greatest rate of change in the I_{psf} from a Fresnel zone plate is due to the focal point closest to the lens, while focal points further from the lens aid in preserving contrast.

To analyze changes in the I_{psf} with respect to depth, consider the approximation of the contributions of the Schuster fringes given a Fresnel zone plate as an aperture in Section 2.5.2. Equation (3.6) from Section 3.2.1 describes the size of the disk or Gaussian shaped focal blur for a traditional clear aperture, c_a, with respect to the distance to the focal plane, s_d, and the distance to a given point source of light, s_a. Without loss of generality, the focal blur produced from Fresnel zone plate aperture, c_a, will be approximated for this section as a superposition of disk shaped focal blurs, c_{a_n}, that result from the different focal planes, s_{d_n} of the optical system.

Recall that A_d is the diameter of the aperture and m_s is the magnification of the optical system and is a function of s_{d_n}. Because the interest is in changes to the I_{psf} respect to distance, consider the rate of change in c_{a_n} with respect to s_a which is given by Equation (3.12), when s_a is greater than s_{d_n}.

$$\frac{dc_{a_n}}{ds_a} = \frac{A_d m_s s_{d_n}}{s_a^2}, \ s_a \in \left(s_{d_n}, \infty\right) \tag{3.12}$$

Figure 3-6 illustrates the relationship between values of c_{a_n} to s_a for various values of s_{d_n}, showing that the nearest focal plane results in both the largest change in focal blur size as s_a changes, and a larger focal blur as $s_a \to \infty$. The plot shows that the rate of change in the blur size increases as the focal plane is moved towards the camera lens. The plot also shows that the blur size as $s_a \to \infty$ increases as the focal plane is moved towards the camera lens. The focal blur may be approximated by disk or two-dimensional Gaussian, and an increase in the diameter of the focal blur is analogous to an increase in the standard deviation of the equivalent Gaussian. For an image convolved with a Gaussian, increasing the standard deviation of that Gaussian reduces edge strength and attenuates higher frequencies in the resultant image [11, 15]. A trade-off then exists between selecting a focal plane near the lens to increase the observability in the rate of change in the focal blur as s_a increases, and selecting a focal plane further from the camera to reduce attenuation of edge strength and high frequency content as $s_a \to \infty$.

The value for m_s is related to s_d by Equation (3.13), where z_d is the distance from the lens to the pixel plane [15].

$$m_s = \frac{z_d}{s_d} \tag{3.13}$$
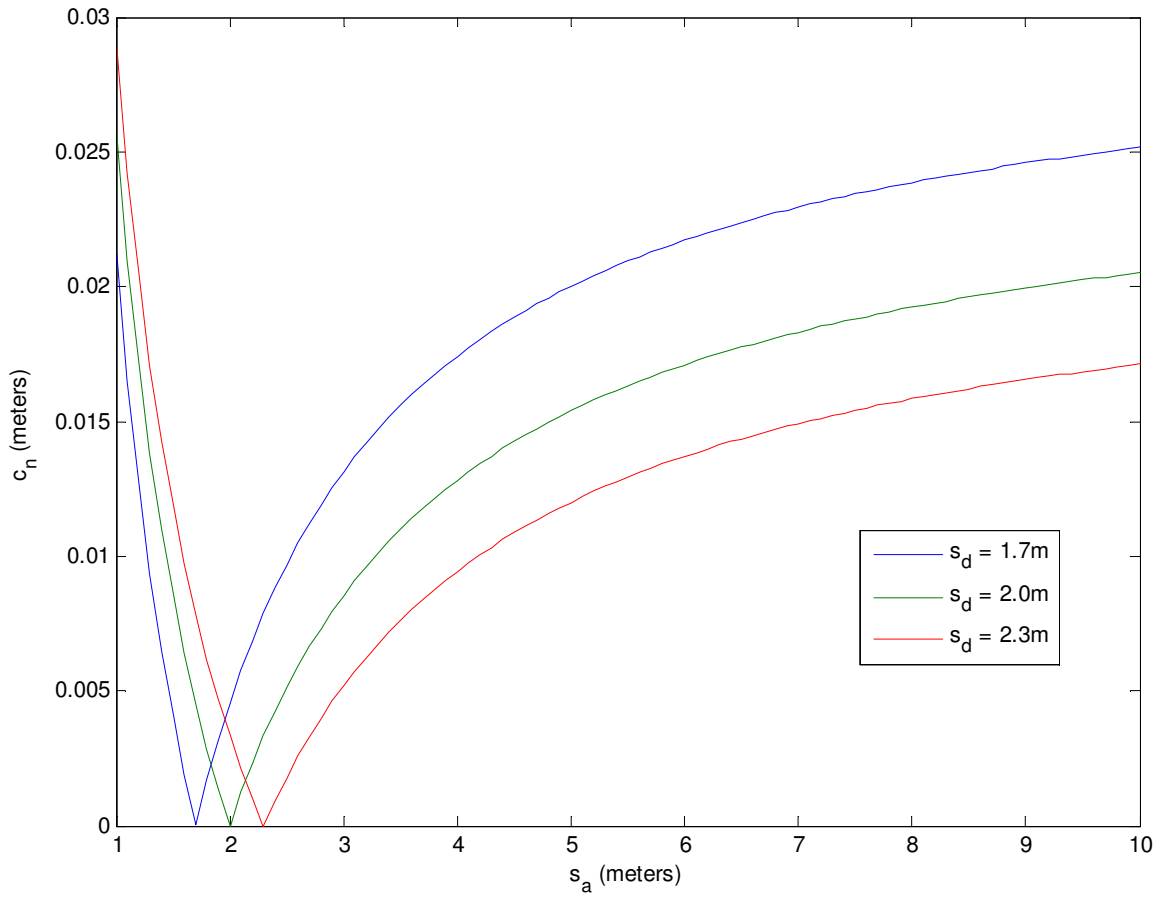
Figure 3-6 Clear Aperture Blur Diameter versus Distance for Three Focal Planes

Combining the thin lens Equation (3.14) from [15] with Equation (3.13) can be simplified to Equation (3.15) to define m_s in terms of s_d and the focal length of the lens, f_l [15].

$$\frac{1}{z_d} + \frac{1}{s_d} = \frac{1}{f_l} \tag{3.14}$$

$$m_s = \frac{f_l}{s_d - f_l} \tag{3.15}$$

Recalling that Equation (3.6) shows that $c_a \to A_d m_s$ as $s_a \to \infty$, Equation (3.15) shows that, as $s_a \to \infty$, the portion of c_a produced by each value of n decreases in diameter as s_{d_n} increases. Replacing m_s in Equation (3.12) with the definition for m_s found in Equation (3.15) yields Equation (3.16), which may be simplified to Equation (3.17).

$$\frac{dc_{a_n}}{ds_a} = \frac{A_d \left(\dfrac{f_l}{s_{d_n} - f_l} \right) s_{d_n}}{s_a^2}, \; s_a \in \left(s_{d_n}, \infty \right) \tag{3.16}$$

$$\frac{dc_{a_n}}{ds_a} = \frac{A_d}{s_a^2} \left(\frac{1}{f_l} - \frac{1}{s_{d_n}} \right)^{-1}, \; s_a \in \left(s_{d_n}, \infty \right) \tag{3.17}$$

Equation (3.17) shows that as $s_a \to \infty$, the portion of dc_{a_n}/ds_a produced by each value of n also decreases as s_{d_n} increases. Assuming a clear aperture, an approximation of the focal blur from an optical system with a Fresnel zone plate aperture is given by Equation (3.18).

$$c_{a_n} = k_n A_d \frac{f_l \left(s_a - s_{d_n} \right)}{s_a \left(f_l - s_{d_n} \right)}, \; s_a \in \left(\max \left(s_{d_n} \right), \infty \right) \tag{3.18}$$

It is important to note that this approximation only accounts for the contributions of the Schuster fringes in formation of the I_{psf}, whereas the approximation method described in Section 3.1 is more accurate because it is not limited to Schuster fringes. Allowing c_{a_n} to be the value of c_a for a given value of n, the change in c_{a_n} with respect to s_a for a given s_{d_n} is approximated by Equation (3.19).

64

$$\frac{dc_{a_n}}{ds_a} = A_d \left(\frac{f_l s_{d_n}}{s_a^2 \left(f_l - s_{d_n} \right)} \right), \ s_a \in \left(s_{d_N}, \infty \right) \tag{3.19}$$

Without loss of generality, now consider $c_{d_{1,2}}$ as the difference in size of c_a for two different values of n. Equation (3.20) gives the value for $c_{d_{1,2}}$, and Equation (3.21) gives the change in $c_{d_{1,2}}$ with respect to $s_a \in \left(s_{d_N}, \infty \right)$.

$$c_{d_{1,2}} = \left(\frac{A_{d_1} f_l s_{d_1}}{s_a^2 \left(f_l - s_{d_1} \right)} \right) - \left(\frac{A_{d_2} f_l s_{d_2}}{s_a^2 \left(f_l - s_{d_2} \right)} \right), \ s_a \in \left(s_{d_N}, \infty \right) \tag{3.20}$$

$$\frac{dc_{d_{1,2}}}{ds_a} = f_l \left(\frac{f_l A_{d_1} s_{d_1} - f_l A_{d_2} s_{d_2} + s_{d_1} s_{d_2} \left(A_{d_2} - A_{d_1} \right)}{s_a^2 \left(f_l - s_{d_1} \right) \left(f_l - s_{d_2} \right)} \right), \ s_a \in \left(s_{d_N}, \infty \right) \tag{3.21}$$

For $s_a \in \left(s_{d_{n-1}}, s_{d_n} \right)$, the relations are similar to those given in Equation (3.18), however $\left| s_a - s_{d_n} \right| = -\left(s_a - s_{d_n} \right)$, $s_a \in \left(0, s_{d_n} \right)$ and the change in c_{a_n} with respect to s_a becomes negative instead of positive. The difference between two values of c_a with one value of s_d closer than the point source and the second further from the point source changes Equations (3.20) and (3.21) to Equations (3.22) and (3.23) respectively.

$$c_{d_{1,2}} = \left(\frac{A_{d_1} f_l s_{d_1}}{s_a^2 \left(f_l - s_{d_1} \right)} \right) - \left(\frac{A_{d_2} f_l s_{d_2}}{s_a^2 \left(f_l - s_{d_2} \right)} \right), \ s_a \in \left(s_{d_1}, s_{d_2} \right) \tag{3.22}$$

$$\frac{dc_{d_{1,2}}}{ds_a} = f_l \left(\frac{f_l A_{d_1} s_{d_1} + f_l A_{d_2} s_{d_2} - s_{d_1} s_{d_2} \left(A_{d_2} + A_{d_1} \right)}{s_a^2 \left(f_l - s_{d_1} \right) \left(f_l - s_{d_2} \right)} \right), \ s_a \in \left(s_{d_1}, s_{d_2} \right) \tag{3.23}$$

This section shows the relationship between s_{d_n} and the change in the focal blur as s_a changes. Smaller values of s_{d_n} result in greater changes in the focal blur, whereas larger values

of s_{d_n} produce a generally smaller focal blur, which increases the contrast of the coded image. Recall from Section 3.2.2 that, except for various aberrations in the lens system, values of s_a and s_d that satisfy Equation (3.9) for a given value of k will result in an equivalent focal blur. If desired, a focal blur may then be selected such that $s_a \in \left(s_{d_{N-1}}, s_{d_N} \right)$, and an equivalent design may be produced using Equation (3.9) such that $s_a \in \left(s_{d_{N-1}}, s_{d_N} \right)$ for any given value of s_a.

Recalling that the focal blur from a clear aperture may be approximated as a disk or a two-dimensional Gaussian, then c_{a_n} is analogous to the standard deviation of the equivalent Gaussian focal blur approximation. While the focal planes closer to the camera lens, those with smaller values for s_{d_n}, produce greater values of $\dfrac{dc_{a_n}}{ds_a}$, they also reduce edge strength and higher frequency content in the coded image, as shown in Section 2.3.3 [11, 15]. The focal planes further from the camera lens, those with larger values for s_{d_n}, produce smaller values of $\dfrac{dc_{a_n}}{ds_a}$, but also do not reduce edge strength or higher frequency content as significantly as the nearer focal planes [11, 15]. The combination of multiple focal planes in a single image allows exploitation of larger values of $\dfrac{dc_{a_n}}{ds_a}$ from smaller s_{d_n} to aid in depth discrimination during measurement, while also retaining edge strength and high frequency content from larger values of s_{d_n}.

3.4 Overview of Coded Aperture Analysis and Design

This chapter presented analysis of both the coded aperture in general and the Fresnel zone plate aperture in particular. The Fresnel zone plate presents unique changes in the point-spread function with respect to scene depth. By characterizing these changes, design decisions may be made to improve the performance of the coded aperture system for a given application. Also discussed was a method of choosing a specific depth at which a given point-spread function is to occur. To improve the characterization of the behavior of the zone plate and compare the behavior to a variety of other coded aperture designs, Section 3.1 provides a more accurate method of modeling the point-spread function of the coded aperture system.

4. Coded Image Depth Measurement

This chapter presents a method for measuring depth from a coded image. To measure depth, a set of intensity point spread functions for various depths are measured a priori. A region of interest is selected within the coded image to measure, and a set of the intensity point spread functions are then compared to that region. A real valued fitness is found for each comparison to aid in determining which intensity point spread function best matches the observed image coding. Because each point spread function corresponds to a depth, an interpolation of depth values versus fitness values is used to determine the depth for which the fitness is expected to be greatest. The depth expected to provide the greatest fitness value then becomes the measured depth for the coded image region.

The method of determining depth from an interpolated fitness maximum is discussed first, along with the distribution of depths at which to measure a priori intensity point spread functions to support this interpolation. Then the three different fitness determination methods are tested in a scenario for three different apertures and the results are compared. The three apertures include the traditional round aperture, Levin's proposed aperture, and a Fresnel zone plate. The three fitness determination methods include power spectral density analysis, textural contrast analysis of a deconvolved image, and textural entropy analysis of a deconvolved image. Because textural entropy analysis generally provides the better depth determination performance of the fitness methods evaluated, this method is then applied using each of the three apertures to three additional scenes.

4.1 Fitness Interpolation

Levin's localized method of determining depth to points in a coded image is to select a collection of candidate depths and to use the likelihood of each candidate depth as a fitness value [22]. For the study Levin presented, the candidate depths chosen are from 2m to 3m in 10cm increments [22]. For a window around each pixel, the depth with the highest fitness is then classified as the measured depth [22]. The result of any depth measurement is then limited to one of ten values between 2m and 3m. Levin also introduces a global method which shows significant improvement in depth determination while segmenting the image into regions of constant depth [22]. For each region, the fitness is determined by the minimum textural entropy of the region when deconvolved with the I_{psf} for each candidate depth [22]. As with the localized method, the depth with the highest fitness is then classified as the measured depth [22]. The section seeks to expand the range of allowable depth measurements from a set of discrete values to a continuous measurement within an interval of depth values. It is assumed that the true depth is approximately constant over a region for which depth is to be determined. Also it is also assumed that the true depth is within the interval of measureable depths, and that the function used to establish I_{psf} fitness values is both continuous and differentiable throughout that interval.

Section 2.4 of this document explains that an image captured using a coded aperture and a narrow depth of field is the convolution of a focused image of the scene and the intensity point spread function given the depth of each point, $I_{psf}(s_a)$. The focused image is not known a priori, however the $I_{psf}(s_a)$ and corresponding optical transfer function, $\mathcal{H}(s_a)$, for a discrete number of

69

pre-determined values of s_a can be determined a priori and compared to the image. For a given region of the image, a fitness value may be determined for each $I_{psf}(s_a)$ that establishes a likelihood that the observed image of the scene is coded by $I_{psf}(s_a)$.

The $I_{psf}(s_a)$ must then be established for various values of s_a. Equation (3.8) of Section 3.2.2 shows that the I_{psf} of a general aperture is a function of the distance between the focal plane in the scene and the distance to a point light source in the scene. To select values of s_a at which to sample, consider the special case of the traditional round aperture. The diameter of the I_{psf} for a traditional round aperture, c_a, as a function of s_a is shown through Equation (3.6) which is presented again here as Equation (4.1).

$$c_a = A_d m_s \left(\frac{|s_a - s_d|}{s_a} \right)$$ (4.1)

Conversely, values for s_a may be determined from values of c_a through Equation (3.7), presented here again as Equation (4.2).

$$s_a = \begin{cases} \dfrac{s_d A_d m_s}{A_d m_s - c_a}, & s_a \in [s_d, \infty), \ c_a \in [0, A_d m_s) \\[4mm] \dfrac{s_d A_d m_s}{A_d m_s + c_a}, & s_a \in (0, s_d), \ c_a \in (0, \infty) \end{cases}$$ (4.2)

The values of s_a in which to capture samples of the I_{psf} may be selected through equidistant values of c_a. For a sample lens system, Figure 4-1 shows c_a versus s_a as a black

line, and the red dots indicate locations on the line for a set of equidistant values of c_a. For

$s_a \in [s_d, \infty)$, the values of s_a in which the I_{psf} is to be captured are then closer together near the

focal plane.
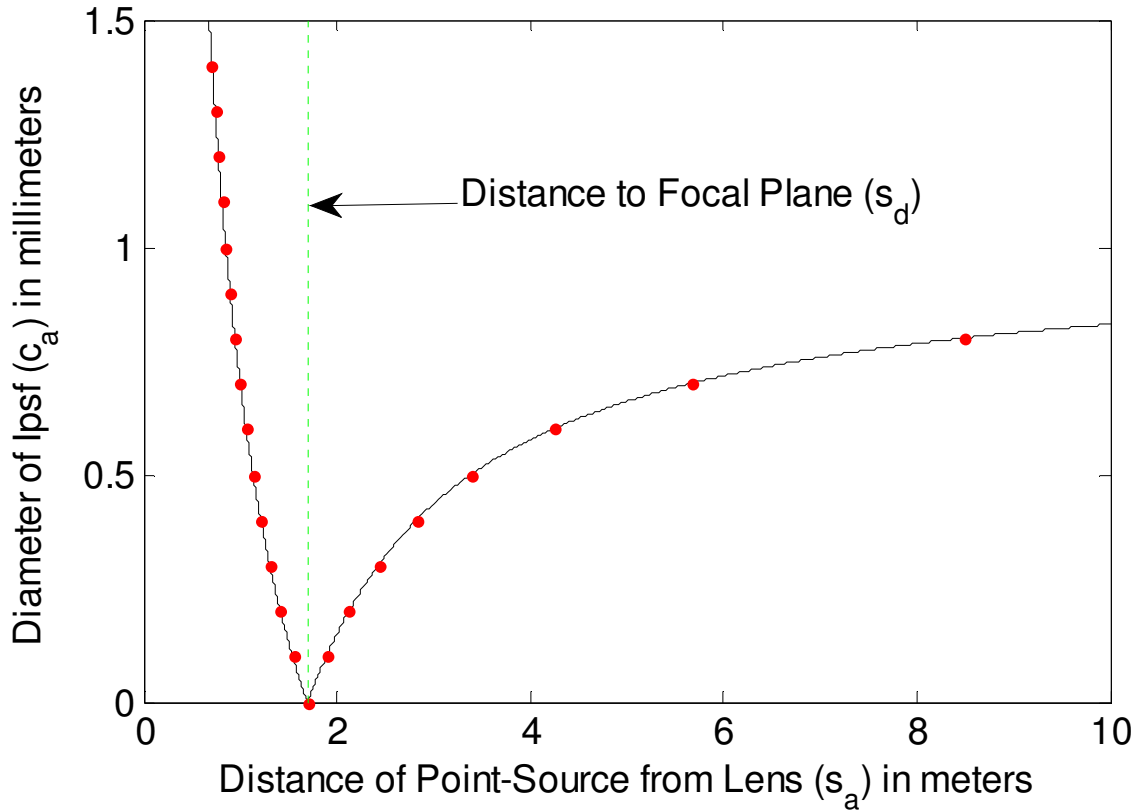
Figure 4-1 Values for c_a Versus s_a When Selecting Equidistant Values of c_a

The fitness method used to determine the likelihood for each $I_{psf}(s_a)$ and corresponding

$\mathcal{H}(s_a)$ may be modeled using the methods described in Section 3.1 to establish that the

differences in the values of c_a are sufficiently small to avoid aliasing over the interval of tested

depths. Also, the model should ensure that the $I_{psf}(s_a)$ is both continuous and differentiable over the interval. Section 3.1 also describes the method for measuring the $I_{psf}(s_a)$ and $\mathcal{H}(s_a)$ of the built coded aperture system.

Upon measurement of an actual image of a scene captured with a coded aperture, the fitness values may then be determined for each $I_{psf}(s_a)$ distribution captured. The results may be linearized over c_a and the value of c_a, for which the fitness is maximal may be estimated. For this research, the value of c_a for which the fitness value is expected to be maximal is selected by first detecting all locally maximal measured fitness values. A parabola is fitted to the locally maximal fitness point and its upper and lower neighbor points. The three points, defined by c_a and fitness value, describe a parabola with an apex between the upper and lower neighbor points. The location of the apex is found, and the fitness value and c_a value for this apex is accepted as the estimated maximal fitness over that interval. If the fitness distribution contains several locally maximal fitness points, then estimated maximal fitness points are compared and the maximum kept. If c_a is maximal only at the upper or lower bound of the interval of depths over which c_a is determined, then the maximum value of c_a is assumed the be outside the interval.

Once a continuous value of c_a is found for which the fitness is estimated to be maximum, then Equation (4.2) may be used to find the equivalent s_a, and thereby the depth to the point.

4.2 Fitness Methods

Section 4.1 describes interpolating the values of a fitness determination to estimate a continuous depth to a region of interest within a coded image. This section describes and evaluates several methods for determining the fitness of each $I_{psf}(s_a)$ distribution.

To analyze the depth measurement characteristics associated with each fitness method, a traditional round aperture was compared with two apertures coded by chrome deposited on a plate glass. One coded aperture is based on Levin's proposed design [22] with a maximum diameter of 14 mm. The second aperture is an eleven zone Fresnel zone plate. Six of the zones are transparent, including the center, and the remaining five zones are opaque. The outer diameter of the eleventh zone is also 14 mm. A Prosilica GE4900 camera was paired with a 50 mm lens focused at 1.7 meters for testing. When installed in the camera, the apertures are affixed to the back of the camera lens and result in a maximum opening that is approximately equivalent to an f-stop, N, of 2.4. The f-stop is defined as the ratio of the lens focal length and the diameter of the aperture opening, therefore the aperture opening is f_ℓ/N. For a 50mm focal length and 2.4 f-stop, the maximum opening of each aperture then is approximately equivalent to 21 mm opening for a thin-lens system. The built-in aperture of the lens set to f-stop 2.4 was also evaluated for comparison to the coded aperture configurations. Also, a green P01 filter is added to the front of the lens to prevent chromatic aberration.
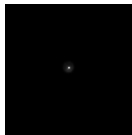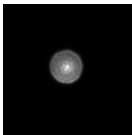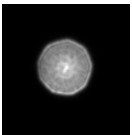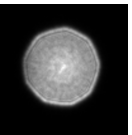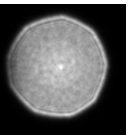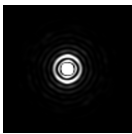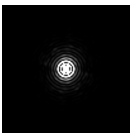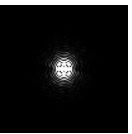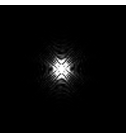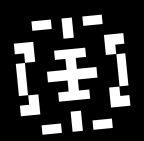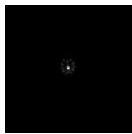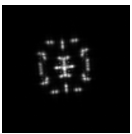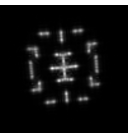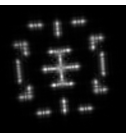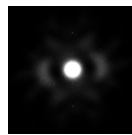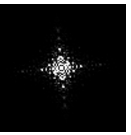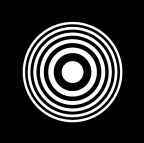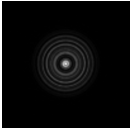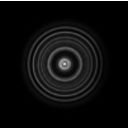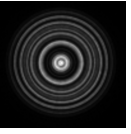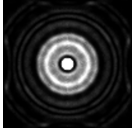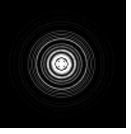
To measure the $I_{psf}(s_a)$ of the resultant system, a point source of light was placed in a darkened corridor at twenty four distances from the lens, with the distances chosen by using

Equation (4.2). The distances chosen are all 1.7 meters or greater since the distance to the focal plane is 1.7 meters. Samples of the measured intensity point spread functions for the traditional aperture and two coded apertures are presented in Table 4-1. Appendix A presents the modeled and measured point spread functions for the Levin aperture at each value of s_a, and Appendix B presents the modeled and measured point spread functions for the Fresnel zone plate aperture at each value of s_a. Each region of the image for which depth is assumed approximately constant is a 100×100 pixel square. Depth measurements interpolated to be outside the interval from two to forty meters are discarded.

To evaluate the fitness method for depth determination in a scene, four different scene types are presented. The four scenes include a black and white acquity chart, a poster at the end of a hallway, a sign with raised lettering, and a bookshelf containing several books. This section will only consider measurements made of the poster at the end of the hallway; however Section 4.3 will show the results with the other scenes to highlight the effect of the scene on depth measurement. Figure 4-2 shows an in-focus image of the poster at the end of the hallway. The posters are white with black lettering, contain pictures, diagrams and graphs, and are posted on a black board with an aluminum border. The images of the poster were captured approximately one meter apart. In each picture a planar region normal to the face of the camera is selected, and the region is separated into 16 segments arranged in a 4×4 grid. The strongest corner in each segment is selected as the center about which a 101×101 pixel window is measured for range. The corners selected must be at least 50 pixels from the edge of the segment to ensure that no single pixel in the image contributes to the depth measurement of more than one pixel window. The truth depth for each segment is derived from the measured boresight distance from the face

74

of the camera lens to the normal surface in the scene using a Ryobi RP4010 laser range finder

with a stated accuracy of $\frac{1}{16}$ inches. Assuming the surface is normal to the camera face, the

true depth to each corner triangulated from the angular offset of the selected pixel and the

boresight depth.

Table 4-1 Measured Point Spread Functions for Three Aperture Configurations

Aperture Shape		Approximate distance from lens to point source of light				
		2.02 m	2.66 m	3.86 m	7.08 m	42.4 m
Built-In	I_{psf}					
	I_{psd}					
Levin	I_{psf}					
	I_{psd}					
FZP	I_{psf}					
	I_{psd}					
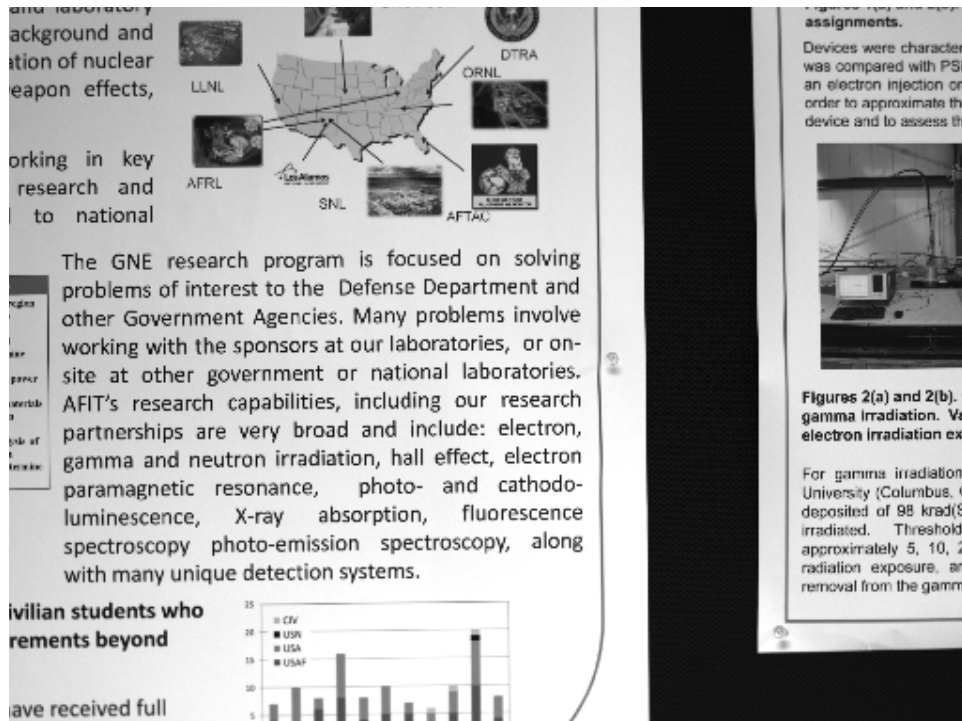
Figure 4-2 Posters at the End of a Hallway

4.2.1 Power Spectral Density of Coded Image. The first method to consider is the one initially proposed by Levin for localized depth measurement [22]. This method, described in Section 2.3.4, assumes that the Fourier transform of regions of constant depth s_a in the coded image are the product of the Fourier transform of the fully focused image and $\mathcal{H}(s_a)$ [22]. Except for noise, any locations of the $\mathcal{H}(s_a)$ for which a zero is present should likewise produce a zero in the same locations in the Fourier transform of the coded image [22]. The fitness is then given by Equation (2.39).

It was observed during testing that the fitness value is consistently maximal near the focal plane. As $s_a \rightarrow s_d$, the total number of zeros in $\mathcal{H}(s_a)$ reduce until there are no zeros present at

$\mathcal{H}(s_a = s_d)$. With no zeros in the denominator of Equation (2.39), the likelihood grows higher. To account for this phenomena, the fitness of the three other scenes (acquity chart, large sign and bookshelf) were measured at a variety of depths. For the fitness at each s_a, the top and bottom 10% of the fitness measurements were discarded and the remaining values were averaged. The result is an average fitness value for each $\mathcal{H}(s_a)$ when used to measure a variety of scenes and depths. The fitness value used for depth evaluation is then the difference between the fitness observed for a particular measurement and the average observed fitness. This adjustment differs from Levin's originally proposed method; however, the depth interval for this study covers two to forty meters whereas Levin's depth interval covers only two to three meters [22].

Figure 4-3 shows the depth measurements found using this method for each of the tested apertures and Table 4-2 provides a statistical summary of the results. The high rate at which the measurements were discarded for each aperture and the overall shape of the plots suggest that although the Fresnel zone plate may be useful for measuring depths up to approximately seven meters, little useful information is gained outside this operating region. The primary challenge with this technique is overcoming tendency towards higher fitness values near the focal plane. This method may provide better performance over an interval of depths in which the amount of $\mathcal{H}(s_a)$ near remains approximately constant even while the locations of near zeros regions change.

Overall the depth measurement performance appears worse than the results observed in [22] measuring power spectral density with the Levin aperture. Three primary causes for the

decrease in performance include the differences in the depth intervals over which the measurements are to occur, the distribution of depths at which the I_{psf} distributions were measured, and the means by which depth measurements are rejected as invalid. In [22], the depth interval used was from two and three meters, and eleven I_{psf} distributions were captured that were ten centimeters apart [22]. For this test, the depth interval used was from two to forty meters, and although twenty-one I_{psf} distributions were captured over the entire interval, only eight I_{psf} distributions occur over the two to three meter interval. In addition, measurements for which a single maxima appears at the lower depth interval bound of two meters are discarded for this test due to significant pixel noise error when measuring larger values of s_a, but all depth measurements are retained as valid in [22].
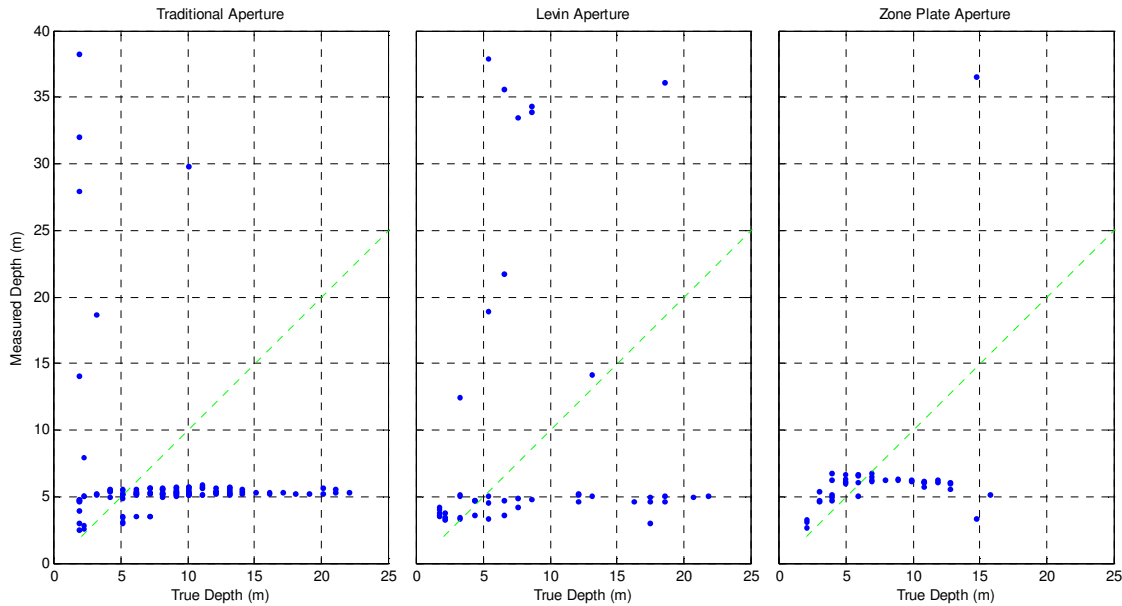
Figure 4-3 Depth Measurement using Power Spectral Density

Table 4-2 Statistics for Depth Measurement using Power Spectral Density

Aperture	Discard Rate (%)	Mean Square Error (m^2)	Median Square Error (m^2)
Traditional	71.9	65.7	15.6
Levin	86.0	144.8	14.3
Zone Plate	88.5	25.4	2.90

4.2.2 Contrast of the Deconvolved Image.   The second method to consider is the contrast of the deconvolved image. The resultant image is an estimate of the scene that, convolved with the $I_{psf}(s_a)$, would produce the coded image. An estimated scene was made assuming each of the $I_{psf}(s_a)$ distributions, and the contrast of the final image was measured. The assumption with this technique is that deconvolution with an incorrect $I_{psf}(s_a)$ is a form of aberration in the final image, and that aberrations cannot increase and often decrease the contrast of any component of an image [15].

As with the power spectral density method, it was observed during testing that the fitness value is consistently maximal near the focal plane. The cause in this case is white pixel noise superimposed with the coded image. The pixel noise is equivalent to an in-focus aberration. When the image is deconvolved with an $I_{psf}(s_a \neq s_d)$, the pixel noise becomes blurred and averaged over many other pixels in the contribution to the final image. Because the noise is approximately white, the contrast of the pixel noise in the final image decreases as the blur

increases. The contrast of the pixel noise is then maximal when $I_{psf}(s_a = s_d)$, and as a result the fitness value of the contrast measurement is also maximal. As with the power spectral density fitness measurement, the fitness of the three other scenes (acquity chart, large sign and bookshelf) were measured at a variety of depths. For the fitness at each s_a, the top and bottom 10% of the fitness measurements were discarded and the remaining values are averaged. The result is an average fitness value for each $I_{psf}(s_a)$ when used to measure a variety of scenes and depths. The fitness value used for depth evaluation is then the difference between the fitness observed for a particular measurement and the average observed fitness.

Figure 4-4 shows the depth measurements found using this method for each of the tested apertures and Table 4-3 provides a statistical summary of the results. Using the contrast for fitness determination discarded the depth measurements at a much lower rate than with the power spectral density technique, and the shape of the overall plots suggest the traditional rounded aperture provides more depth information beyond five meters.

Table 4-3 Statistics for Depth Measurement using Contrast

Aperture	Discard Rate (%)	Mean Square Error (m^2)	Median Square Error (m^2)
Traditional	31.9	154.2	63.0
Levin	39.9	124.9	94.7
Zone Plate	35.9	124.1	62.4
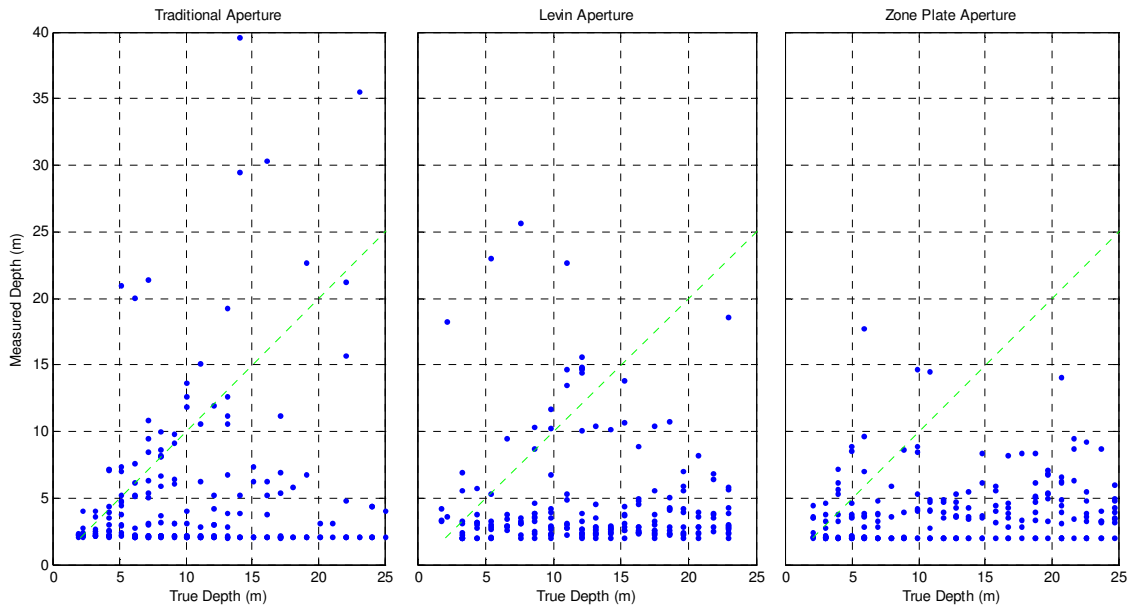
Figure 4-4 Depth Measurement using Contrast
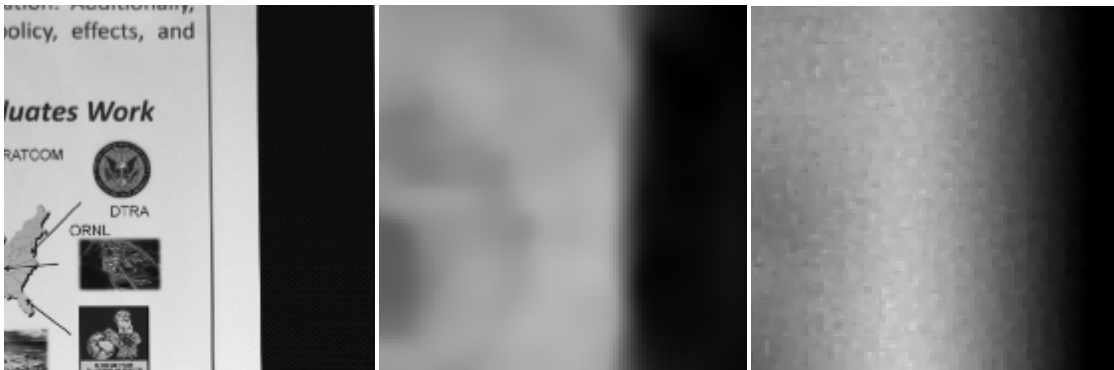
Figure 4-5 True Image Deconvolved at 8.5 and 2.0 Meters

Recall that the scene contains large image components, such as white posters on a black

board, and small image components, such as text. Figure 4-5 shows an example image, truth

observed on the left, at approximately 9.1 meters and deconvolved using $I_{psf}(8.49\,\text{meters})$ in the center and $I_{psf}(2.02\,\text{meters})$ in the right. At greater distances, the text is obscured and is removed from the deconvolved image, however the edge of the poster is reproduced as a sharp line when deconvolved at the correct depth. When deconvolved near the focal plane, the sharp edge is blurred and the contribution of pixel noise becomes more prominent.

4.2.3 Entropy of the Deconvolved Image.   The third method to consider is the entropy of the deconvolved image. The method is used locally, but was proposed by Levin for a global depth measurement [22]. A deconvolution is made using the sparse image priors. The resultant image is an estimate of the scene that, convolved with the $I_{psf}(s_a)$, would produce the coded image. An estimated scene was made assuming each of the $I_{psf}(s_a)$ distributions, and the entropy of the final image is measured. The assumption with this technique is that a natural image is generally smooth with sparse edges, and that the mostly smooth regions of a natural image are of lower entropy than random noise. Deconvolution then with an incorrect $I_{psf}(s_a)$ is a form of random noise in the final image, and that the additional noise will increase the measured entropy of the deconvolved image. The best fit $I_{psf}(s_a)$ produces the lowest textural entropy measurement [22]. The fitness is taken as the inverse of the textural entropy of the deconvolved scene.

As with the power spectral density and contrast methods, it was observed during testing that the fitness value is consistently maximal near the focal plane. As with the contrast method, the cause is white pixel noise superimposed with the coded image. The pixel noise is equivalent

to an in-focus aberration. When the image is deconvolved with an $I_{psf}(s_a \neq s_d)$, the pixel noise becomes blurred and averaged over many other pixels in the contribution to the final image. Because the noise is approximately white, the contrast of the pixel noise in the final image decreases as the blur increases. The contrast of the pixel noise is then maximal when $I_{psf}(s_a = s_d)$, and as a result the fitness value of the contrast measurement is also maximal. As with the power spectral density fitness measurement, the fitness of the three other scenes (acquity chart, large sign and bookshelf) were measured at a variety of depths. For the fitness at each s_a, the top and bottom 10% of the fitness measurements were discarded and the remaining values are averaged. The result is an average fitness value for each $I_{psf}(s_a)$ when used to measure a variety of scenes and depths. The fitness value used for depth evaluation is then the difference between the fitness observed for a particular measurement and the average observed fitness.

Figure 4-6 shows the depth measurements found using this method for each of the tested apertures and Table 4-4 provides a statistical summary of the results. Using the entropy for fitness determination resulted in a lower rate at which depth measurements were chosen to be discarded as compared to both the power spectral density method and the contrast method. The overall shape of the plots suggests depth information may reasonably be obtained from all three apertures for a depth interval less than approximately fifteen meters, but that the traditional aperture is not useful for greater ranges. The Levin and Fresnel zone plate apertures provide depth measurements that appear bimodal, with the modes separating for true depth values

beyond ten meters. Using the Fresnel zone plate with an entropy fitness evaluation method provides both the lowest median square error of the three apertures considered.
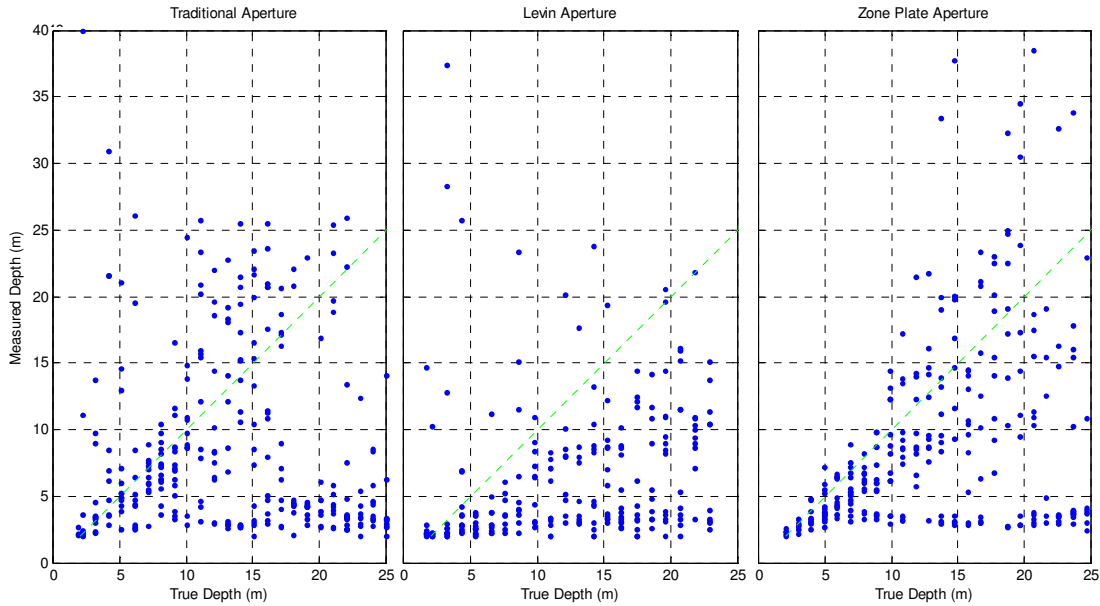
Figure 4-6 Depth Measurement using Entropy

Table 4-4 Statistics for Depth Measurement using Entropy

Aperture	Discard Rate (%)	Mean Square Error (m^2)	Median Square Error (m^2)
Traditional	25.0	158.2	55.5
Levin	29.5	85.8	32.2
Zone Plate	29.2	71.7	11.0

4.3. Measuring Various Scenes.

The previous section discussed the depth measurement for the scenario of set of posters on a board at the end of a hallway. Several other scenes were also evaluated, and the measurement error changes significantly with each. This next section compares the changes in the measurement error when measuring depth in three other scenarios in addition to the hallway. To account for the maximal fitness values near the focal plane of each scene, the fitness values for all measurements made of the three other scenes were measured. The top and bottom 10% of the fitness measurements from the three other scenes were discarded and the remaining values were averaged. The resulting average fitness distribution was applied to the scene under consideration so that the measured fitness is the difference between the fitness observed for a particular measurement and the average observed fitness for all measurements of the three other scenes.

The first is the idealized scenario of measuring depth to an acquity chart. The chart, shown as Figure 4-7, consists of black and white regions, with some shading due to variation in illumination. The scene has very high contrast. True depth was then taken as the intersection between a projection in the scene of a feature's pixel coordinates and the plane described by the acquity chart. Approximately half of the measurement were made with values very near s_a, and the other half are approximately equidistant from each other.
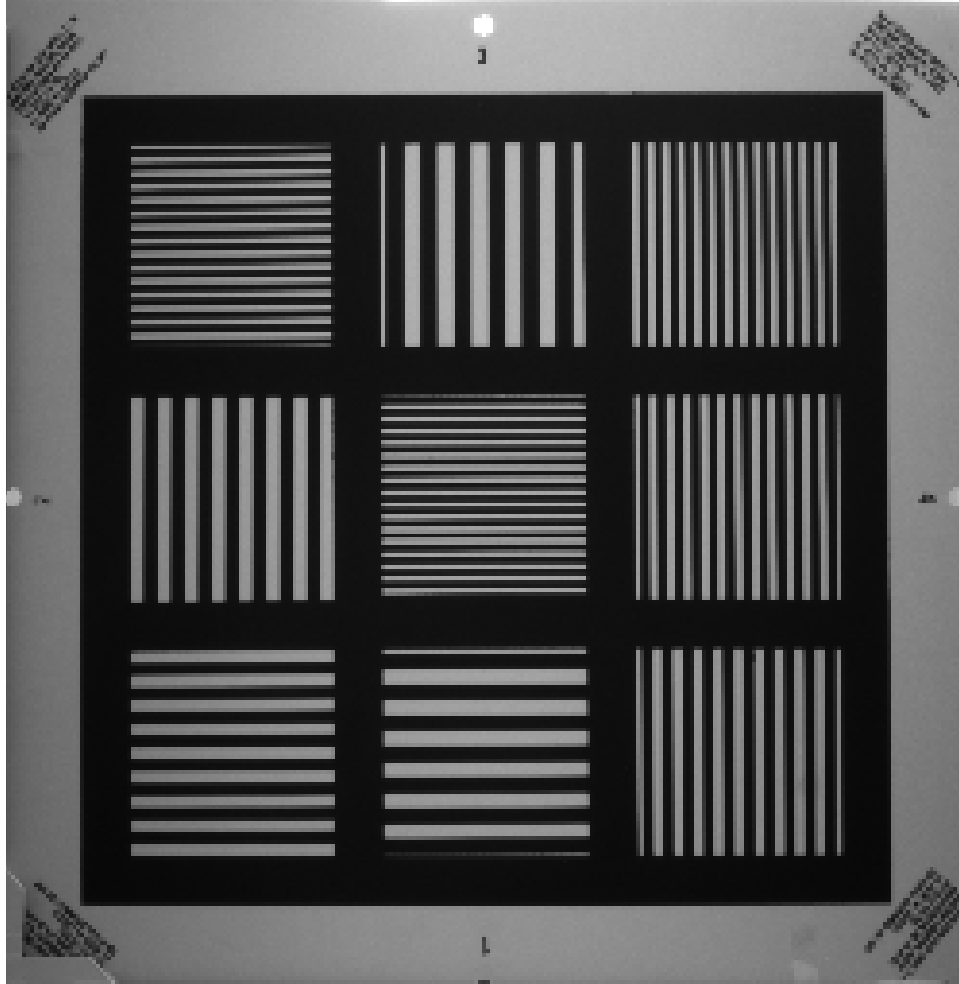
Figure 4-7 Acquity Chart Scenario

The second additional scenario is a metallic sign as shown in Figure 4-8. The sign made of is simulated aluminum lettering raised 1.5 centimeters from a wall made of wooden panels. The sign is illuminated from the side producing strong shadowing around the letters. The shadows are higher contrast, and the contrast in the wood and aluminum are lower contrast. True depth was then taken as the intersection between a projection in the scene of a feature's pixel

coordinates and the plane described by the wooden panels. The measurements are approximately one meter apart.

Figure 4-8 Metallic Sign Scenario

The third scene is of a bookshelf populated with books. The bookshelf is shown in Figure 4-9. This scene introduces variations in depth, because the spines of the books are up to 17 centimeters behind the front of the shelf. This scene offers some areas of high contrast, and other areas of low contrast. The front of the bookcase shelves can be used to define a plane that is normal to boresight of the camera. Because all books are behind the front of the shelves, this plane also describes a minimum distance from the camera to a feature observed at a given set of pixel coordinates. True depth was then taken as the intersection between a projection in the

scene of a feature's pixel coordinates and the plane described by the shelves. The measurements were taken with the camera moved approximately one meter between image captures.

Figure 4-10 and Table 4-2 shows the measurements for all four scenarios using a traditional rounded aperture and the entropy fitness method. Figure 4-11 and Table 4-6 shows the measurements for all four scenarios using a Levin aperture and the entropy fitness method. And Figure 4-12 and Table 4-7 shows the measurements for all four scenarios using a Fresnel zone plate aperture and the entropy fitness method.

Figure 4-9 Bookshelf Scenario
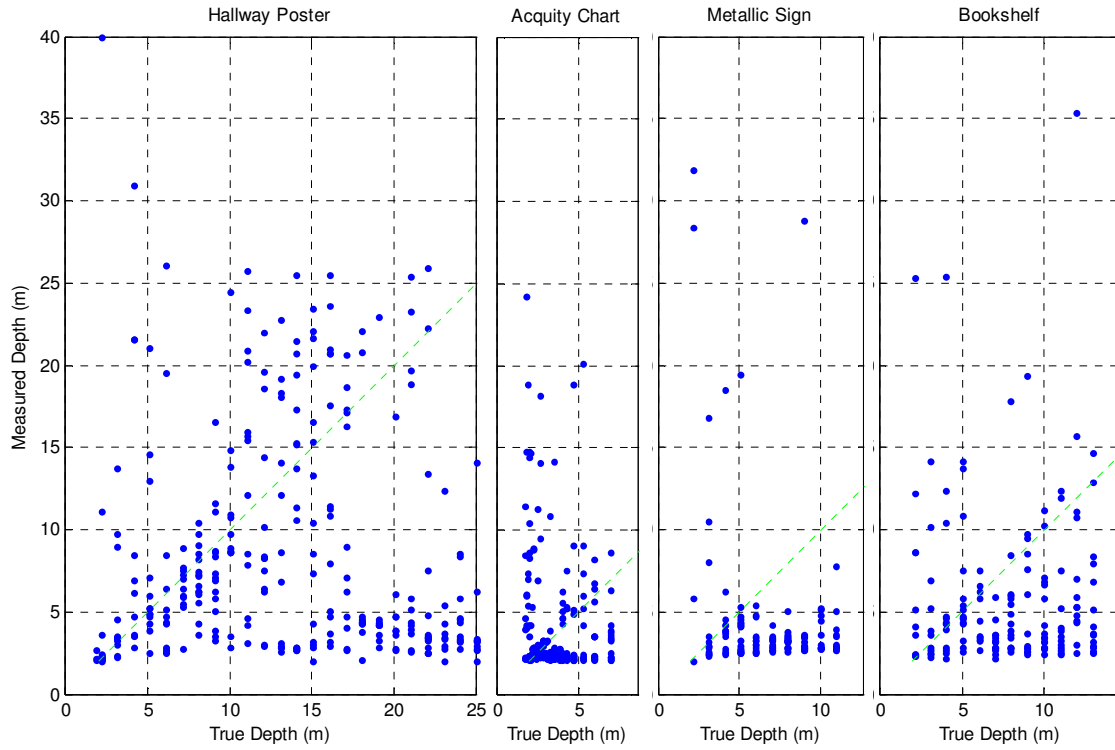
Figure 4-10 Traditional Aperture and Entropy

Table 4-5 Statistics for Traditional Aperture Measuring Various Scenes

Scenario	Discard Rate (%)	Mean Square Error (m^2)	Median Square Error (m^2)
Hallway Poster	25.0	158.2	55.5
Acquity Chart	20.9	13.2	0.823
Metallic Sign	13.1	39.4	17.7
Bookshelf	13.5	36.5	14.5
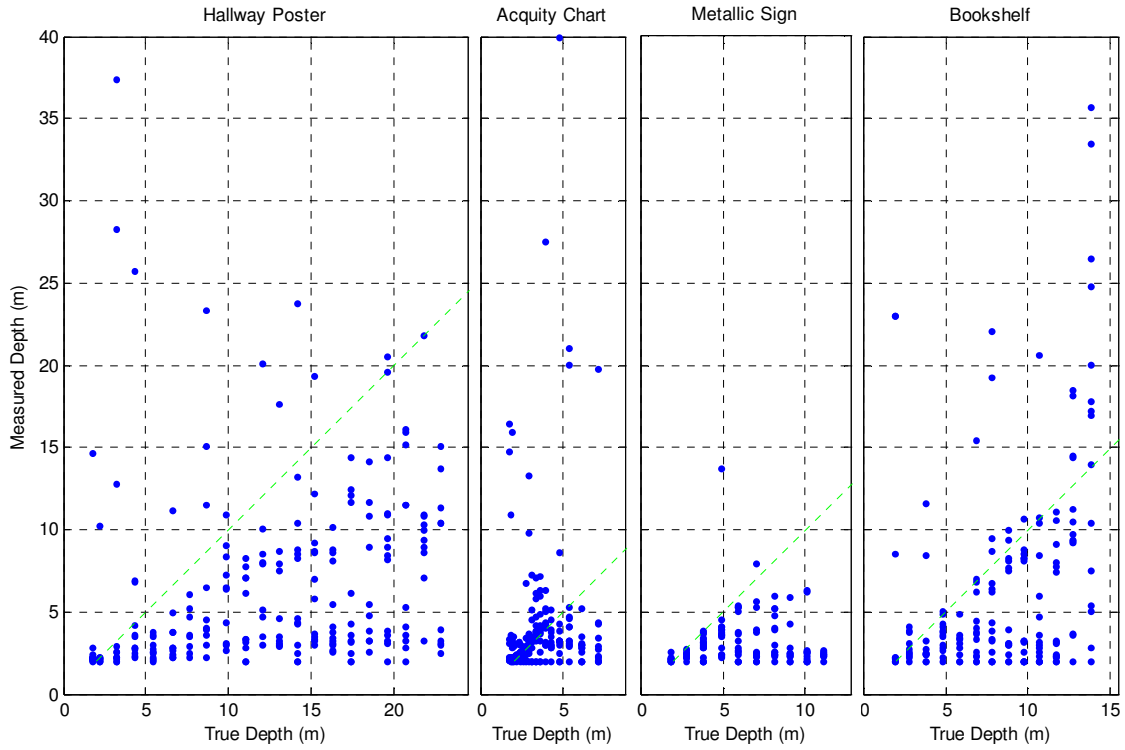
Figure 4-11 Levin Aperture and Entropy

Table 4-6 Statistics for Levin Aperture Measuring Various Scenes

Scenario	Discard Rate (%)	Mean Square Error (m^2)	Median Square Error (m^2)
Hallway Poster	29.5	85.8	32.2
Acquity Chart	13.1	13.9	0.141
Metallic Sign	15.6	18.6	6.26
Bookshelf	16.4	36.6	11.6
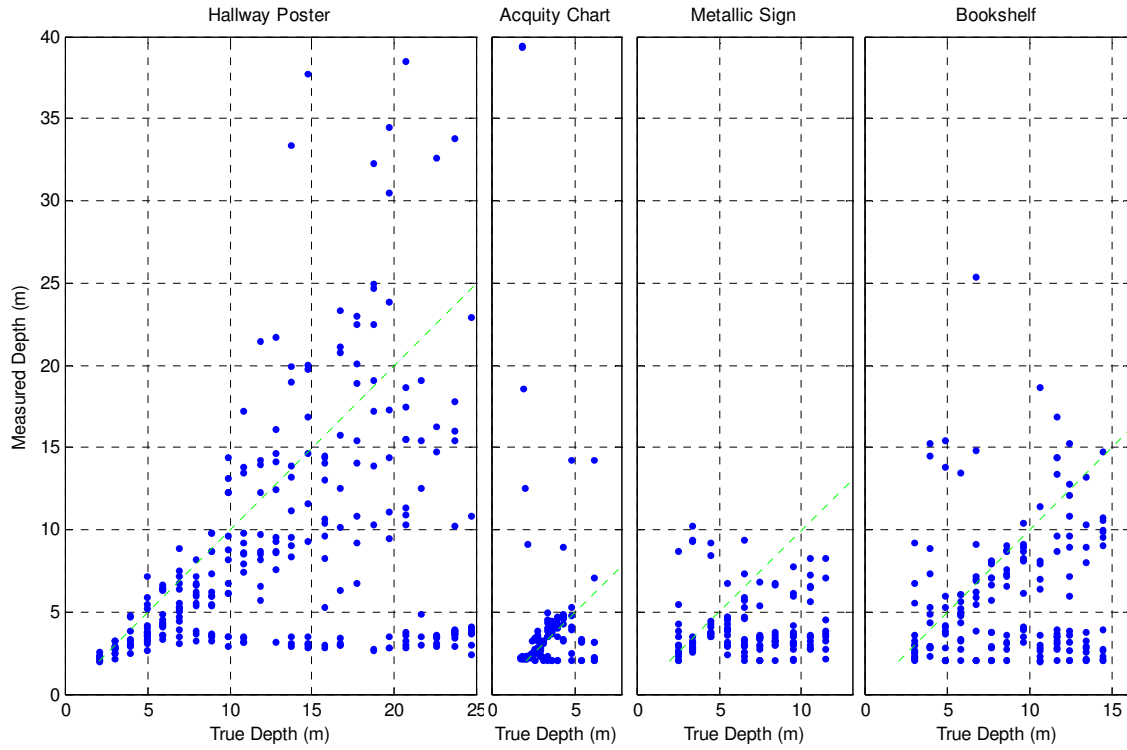
Figure 4-12 Fresnel Zone Plate Aperture and Entropy

Table 4-7 Statistics for Fresnel Zone Plate Aperture Measuring Various Scenes

Scenario	Discard Rate (%)	Mean Square Error (m^2)	Median Square Error (m^2)
Hallway Poster	30.2	71.7	10.7
Acquity Chart	14.5	14.1	0.0978
Metallic Sign	15.6	17.6	10.2
Bookshelf	13.9	33.1	9.23

Figure 4-10 through Figure 4-12 show that the measurement error increases with the true depth to be evaluated. As the depth increases, the resulting measurements are bimodal with the center of one mode increasing in depth in relation to the true depth of the region, and the other mode near the focal plane. It can also be observed that, for the mode that increases with the true depth, the variance increases with depth.

Because the error varies with the true depth, and each scenario does not contain approximately equal distributions of true depths, the mean and median error values for a given scenario in Table 4-2 through Table 4-7 cannot be compared with error values of another scenario. However comparisons can be made between scenarios over common depths measured in Figure 4-10 through Figure 4-12.

Figure 4-10 suggests that, of the two measurement modes of the traditional aperture, the mode following the correct depth is strongest in the hallway poster and acquity chart scenarios. For the multi-depth metallic sign and bookshelf scenarios, erroneous measurements near the focal plane increased. Figure 4-11 suggests that, of the two measurement modes of the Levin aperture, the mode following the correct depth has a greater bias than that exhibited by the same mode of the Fresnel zone plate in Figure 4-12. And, although the hallway poster scenario for the Fresnel zone plate in Figure 4-12 doesn't show the occasional outlier for true depth values under 10 meters that the traditional and Levin apertures show, these outliers do occur with the Fresnel zone plate aperture with the three other scenarios.

Table 4-2 through Table 4-7 show that the median square error for the Levin aperture is less than the median square error of the Traditional aperture for all scenarios, and the median

square error of the Fresnel zone plate is less than the median square error of the Levin aperture for all but one scenario. The Levin aperture showed a lower median square error for the metallic sign scenario.

The collective results of the measurements for all three apertures, and all four scenarios, suggest that a given measurement result may be modeled as a binomial distribution with a weighted combination of two Gaussian distributions and a uniform distribution. The binomial distribution corresponds to the rate at which measurements are identified as unusable and are discarded. For measurements that are not discarded, the two Gaussian distributions and uniform distribution characterize a measurement with colored noise. The uniform portion of the combined distribution characterizes the measurements that are produced by pixel noise rather than the true depth to any point within the scene. Of the two Gaussian distributions, one characterizes the near zero mean error about the true depth and has a standard of deviation that increases with the true depth. The second Gaussian distribution characterizes the spurious measurement noise that consistently produces measurement values near the focal plane. The mean and standard deviation of this second Gaussian distribution changes little with the true depth. For the combined distribution, the means and covariance of the Gaussian distributions as well as the weightings of the uniform and Gaussian distributions vary with both the true depth and the type of scene to be observed.

4.4. Overview of Coded Image Depth Measurement.

The depth measurement performance of three different apertures was compared against three different methods of determining fitness for four different scenarios. To perform the measurements, intensity point spread functions for the three apertures were measured a priori. The distances selected at which to measure the intensity point spread functions were chosen to facilitate interpolation of a maximum solution to a fitness function.

Overall, the fitness method based on entropy analysis of the deconvolved image provides better depth measurement performance, for the aperture configurations and depth intervals selected, than either contrast analysis or power spectral density analysis. Using the entropy analysis method, the Fresnel zone plate provides a lower median square error than the Levin or traditional round aperture for all but one of the scenarios explored.

A reasonable sensor model for coded aperture depth measurement is a pair of Gaussian distributions, a uniform distribution and a binomial distribution. The binomial distribution models the probability that a given depth measurement is discarded as outside an allowable interval of depth measurements. One Gaussian distribution models the error of the measurement around the true depth. A second Gaussian distribution models spurious outlier depth measurements, and the uniform distribution models the error induced by pixel noise.

5. Performance and Analysis of Augmented Navigation System

This chapter presents application of coded aperture techniques to single-camera vision aided inertial navigation. A design to augment a vision aided inertial navigation system is proposed, including a method of selecting features to track and a method of augmenting a Kalman filter to incorporate the additional depth measurements. A statistical depth measurement model is created and used in simulation to evaluate the performance of the proposed navigation system. The resultant system demonstrates vision aided inertial navigation using a single camera.

For smaller vehicles, multi-camera systems may be undesirable because of the added size, weight and power consumed by additional cameras. Also, a small overall vehicle size limits the maximum allowable distance between any two installed cameras, which then constrains stereographic depth measurement capability. For vehicles with multiple cameras, a single-camera method of aiding navigation frees additional cameras to perform non-navigation related tasks. The additional cameras would not be constrained to observing the same region of the scene as the camera used for navigation, nor would there be a requirement that a location estimate be established for any features of the captured images. The additional cameras may then tilt, pan and zoom while performing non-navigational tasks.

The stereoscopic navigation system proposed by Veth in [39] provides the foundation for the coded aperture image aided inertial navigation system used in this work. Veth's system is equipped with two synchronized cameras providing overlapping images of the scene. Stereoscopy is used to concurrently apply two simultaneously captured images in solving scale

ambiguity for the initial creation of feature location estimates for the navigation solution. Subsequent pairs of simultaneously captured images of the features are applied sequentially, rather than concurrently, to update the navigation solution using only the headings observed to the features relative to each camera. Because the images are captured simultaneously and the distances between the cameras are known, stereoscopy implicitly resolves scale ambiguity. [39]

The image aided inertial navigation system proposed by this work uses only one camera with a narrow depth of field and a coded aperture to observe the scene. For both the initial creation of feature location estimates and to update the navigation solution, depth from the defocus of a single captured image is used to solve ambiguity of scale. Because depth measurements from a coded aperture vision system present greater noise than measurements from a stereoscopic system, the proposed system is not expected to perform as well as a two camera system. In an unaided MEMS inertial navigation system, the navigation solution may become unstable after several seconds due to error growth in the navigation solution. Assuming noise in a coded aperture system's range measurements are Gaussian distributed and zero mean, the single camera coded aperture system should maintain a navigation solution that does not diverge and is within several meters of the true position.

5.1 Overview of Augmented System

As illustrated in Figure 3-6, an INS and a vision system provide inputs to a Kalman filter. The Kalman filter uses the vision system measurements to estimate the errors in the inertial navigation system. These INS can then be corrected by these error estimates.
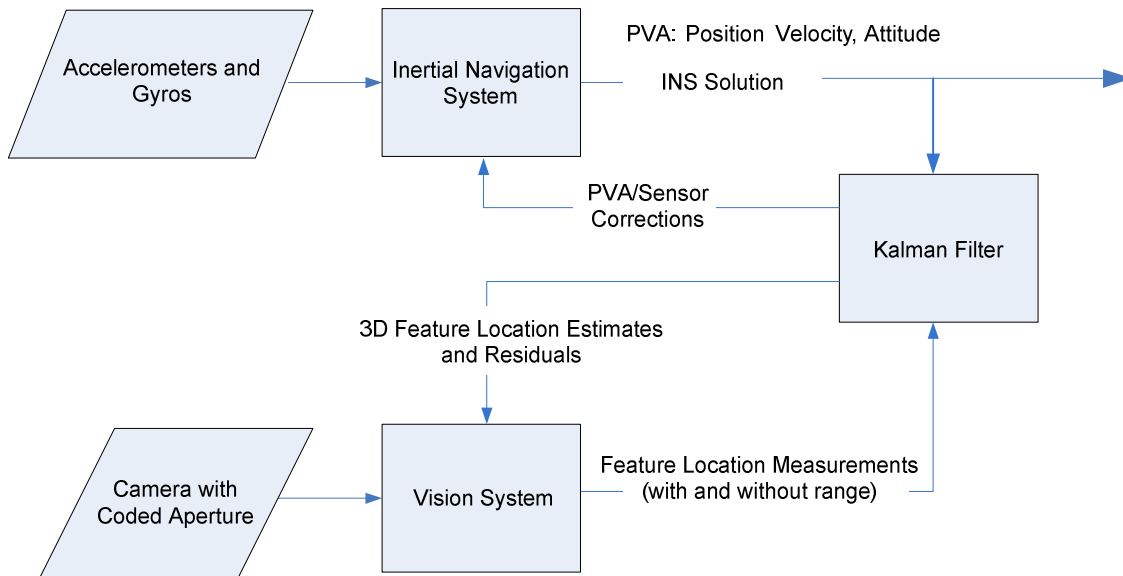
Figure 5-1 System Design of Vision Aided INS Augmented with Coded Aperture

The vision system receives images from the camera and a list of tracked feature location estimates in the camera frame from the Kalman filter. The vision system establishes correspondence from the tracked features to the features in the observed image. For each feature for which correspondence is established, a unit length two dimensional homogenous pointing vector is found. Each pointing vector describes the direction to a given feature and is submitted to the Kalman filter as a measurement. The vision system will use coded aperture information to measure the depth of the feature from the camera. If the measured depth is identified as unusable and selected to be discarded, then the measured depth is not submitted to the Kalman filter. However, when the vision system has determined a measured depth that is usable, the depth measurement and an estimation of the standard deviation of the noise in the depth measurement are submitted to the Kalman filter as an additional measurement.

5.1.1 Kalman Filter Design.   The Kalman filter incorporates the INS solution and the feature location measurements from the vision system, with and without range, to create an optimal estimate of the PVA and three dimensional feature locations in the navigation frame. The feature location estimates and covariance are propagated from the Kalman filter and provided to the vision system.

This research proposes using the depth measurement provided by the coded aperture augmented vision system to reduce linearization errors of the image aided navigation system. Because the feature location s_c is not fully observable with a single two-dimensional camera, Equations (2-26) through (2-28) show the Kalman filter linearization about the unit length homogenous pointing vector $\underline{s}^c$. The estimated relative location of a feature s_c is found by multiplying the unit length homogenous pointing vector $\underline{s}^c$ by the depth estimate. From the determined s_c, the values for $\dfrac{\delta \underline{s}^c}{\delta p^n}$ and $H_{p^n}^z$ may subsequently be found by Equations (2-27) and (2-28). The covariance of the observed pixel noise is not changed by the depth estimate.

This research also proposes using depth itself as a measurement for the Kalman filter. To include the depth, the change in the depth with respect to the change in s_c must be determined and can be found by Equation (5.1) where r (for "range") is the depth.

$$\frac{dr}{ds_c} = \frac{s_c}{|s_c|} \tag{5.1}$$

With depth information included, $H_{p^n}^z$ for a pixel measurement which includes both a pixel measurement and depth is replaced by $H_{p^n}^s$ given by Equation (5.2).

$$H^s_{p^n} = \begin{bmatrix} T_c^{pix} \dfrac{\delta \underline{s}^c}{\delta p^n} \\[2ex] \dfrac{dr}{ds_c} \end{bmatrix} \qquad (5.2)$$

The observation noise of the depth measurement, σ_r^2, is described in Section 5.2. However linearization of the range error due to camera misalignment is given in Equations (5.3) and (5.4). Equation (5.3) shows the small angle change in depth with attitude misalignment of the camera where α^{cam} is the camera attitude error.

$$\frac{dr}{d\alpha^{cam}} = skew\left(\alpha^{cam}\right)\frac{s_c}{|s_c|} \qquad (5.3)$$

Equation (5.4) shows the change in depth with position misalignment of the camera where p^{cam} is the attitude error.

$$\frac{dr}{dp^{cam}} = -C_{c_0}^c \frac{s_c}{|s_c|} \qquad (5.4)$$

The total observation error for the depth is then described by the Equation (5.5) as the sum of the direct depth measurement error and the depth error contributions from camera misalignment.

$$R_r = \sigma_r^2 + \left(\frac{dr}{d\alpha^{cam}}\right)\left(diag\left(\sigma_{\alpha^{cam}}^2\right)\right)\left(\frac{dr}{d\alpha^{cam}}\right)^T + \left(\frac{dr}{dp^{cam}}\right)\left(diag\left(\sigma_{p^{cam}}^2\right)\right)\left(\frac{dr}{dp^{cam}}\right)^T \qquad (5.5)$$

Because the pixel measurement observation noise is assumed independent of the depth measurement observation noise, the covariance of the noise observation for the pixels and depth are as shown by Equation (5.6), where R_z is the $[2\times2]$ pixel covariance and R_r is the $[1\times1]$ depth covariance.

99

$$R_s = \begin{bmatrix} R_z & \vdots & 0 \\ & \vdots & 0 \\ \hline 0 & 0 & \vdots & R_r \end{bmatrix} \tag{5.6}$$

The covariance of the s_c estimate is then given by Equation (5.7).

$$P_{s_c, s_c} = \left(\left(H^s_{p^n} \right)^T (R_s)^{-1} H^s_{p^n} \right)^{-1} \tag{5.7}$$

By making the proposed changes to the observation model and noise observation covariance, depth information from the coded aperture camera is an independent measurement that can be used to ameliorate linearization errors from pixel measurement. However, the covariance of the depth measurement is required to determine the covariance of the overall noise of a given pixel observation and for the covariance of the estimate of the feature location. The next section describes obtaining the covariance of the depth measurement.

5.1.2 Selecting Features to Track.   Section 2.2 explains that in a given frame the features with the highest feature imaging quality, as described by [39], are selected to be tracked by the navigation system. For the augmented system, Section 5.2 will show the process for determining the standard deviation of the true depth estimate which varies with the noisy depth measurement. When selecting features to track in the augmented system, the estimated quality of the feature's depth measurement may be considered in addition to the feature's imaging quality to mitigate initial linearization errors from noisy depth measurements.

For this research, a depth quality metric d_q is derived from the standard deviation of the current true depth estimate and a prediction of the standard deviation of the subsequent true depth estimate for a given feature. Given the current depth measurement r_t, the standard

deviation of the current true depth estimate is denoted as $\sigma_{\hat{s}_t}^2(r_t)$ and a prediction of the standard deviation of the next true depth estimate is $\sigma_{\hat{s}_{t+1}}^2(r_t)$. Section 5.2 will show that depth measurements that are the result of pixel noise corruption have a relatively large $\sigma_{\hat{s}_t}^2(r_t)$. Selecting features to track with smaller values of $\sigma_{\hat{s}_t}^2(r_t)$ provides greater precision in the current depth measurement and avoids erroneous depth measurements that are the result of pixel noise corruption. Also, selecting smaller values of $\sigma_{\hat{s}_{t+1}}^2(r_t)$ increases the likelihood of greater precision in a subsequent depth measurement, as well as avoiding true depths for which the likelihood of pixel noise corruption of subsequent depth measurements is higher.

Section 5.2 will describe the process for determining the mean and standard deviation of the true depth estimate given a noisy depth measurement. The value of $\sigma_{\hat{s}_{t+1}}^2(r_t)$ is found by Equation (5.8), where $P(s_t \mid r_t)$ is the probability of a true feature depth at time t given a depth measurement at time t, $P(s_{t+1} \mid s_t)$ is the probability of a true feature depth at time $t+1$ given a true feature depth at time t, $P(r_{t+1} \mid s_{t+1})$ is the probability of a measured feature depth at time $t+1$ given a true feature depth at time $t+1$.

$$\sigma_{\hat{s}_{t+1}}^2(r_t) = \sum_{s_t}\sum_{s_{t+1}}\sum_{r_{t+1}} \sigma_{\hat{s}_{t+1}}^2(r_{t+1}) P(r_{t+1} \mid s_{t+1}) P(s_{t+1} \mid s_t) P(s_t \mid r_t) \qquad (5.8)$$

For computational simplicity, it is assumed for this system that the standard deviation of the true depth estimate given a depth measurement does not change over time, so that $\sigma_{\hat{s}_{t+1}}^2(r_{t+1}) = \sigma_{\hat{s}_t}^2(r_t)$. For determining $\sigma_{\hat{s}_{t+1}}^2(r_t)$ only, it is assumed that the estimated true depth and the actual true depth are the same, so that $P(s_t = \hat{s}_t(r_t) \mid r_t) = 1$. Also for determining $\sigma_{\hat{s}_{t+1}}^2(r_t)$ only,

it is assumed that the true depth in the next image frame is the same as the true depth for the current image frame, so that $P(s_{t+1} = s_t \mid s_t) = 1$. These assumptions reduce Equation (5.8) to Equation (5.9).

$$\sigma_{\hat{s}_{t+1}}^2 (r_t) = \sum_{r_{t+1}} \sigma_{\hat{s}_t}^2 (r_t) P(r_{t+1} \mid s_{t+1} = \hat{s}_t (r_t)) \tag{5.9}$$

Depth quality for a feature given a depth measurement is then found by Equation (5.10).

$$d_q = \frac{1}{\sqrt{\sigma_{\hat{s}_t}^2 (r_t) \sigma_{\hat{s}_{t+1}}^2 (r_t)}} \tag{5.10}$$

For the augmented system, the features selected to track in a given frame are those for which the geometric mean of the feature quality and depth quality are highest.

5.2 Depth Measurement Statistics

This section describes the process by which a noisy depth measurement is translated into a mean and variance for a true depth estimate. The method of simulating depth measurement noise given an aperture and true depth is also discussed. The simulated navigation environment is a hallway which most resembles the poster scenario of Section 4.3. The scenarios used in this section assume the measurement distributions resemble those of the "Poster" scenario when evaluated for fitness using the entropy method described in Section 4.2.3.

To establish a mean and standard of deviation for an estimate of the true depth given an aperture and a depth measurement, a joint probability density function (pdf) of the true depth and measured depth is estimated for each aperture. Each of the three joint pdf estimates is derived from the true depth versus measured depth observations from Section 4.2.3 and presented in

Figure 4-6. The joint pdf is estimated using Expectation Maximization (EM) of a mixture of ten Gaussian distributions. EM is performed ten times, and each EM is initialized with ten randomly selected Gaussian distributions with equal weighting and diagonal covariance. The distribution produced by each of the ten EM results is normalized, and then all ten distributions are point-wise averaged with the final result again normalized to form the estimated joint pdf. Using this estimate of the joint pdf for true depth and measured depth, the conditional pdf of the true depth can be found for a given depth measurement. It is then assumed that the conditional pdf of the true depth given a depth measurement can be adequately approximated by a single fitted Gaussian distribution. Because the depth measurement characteristics of the systems presented in Figure 4-6 are evaluated for only true depths within twenty-five meters, measured depths beyond twenty-five meters will be used to estimate the joint pdf but are otherwise not used to estimate the mean and standard deviation for the true depth.

For all three tested apertures, Figure 5-2 through Figure 5-7 present the true depth versus measured depth observations along with the corresponding fitted Gaussian approximation of the conditional pdf for the true depth given a measured depth. Figure 5-2, Figure 5-4 and Figure 5-6 show the true versus measured depth observations from the poster scenario using entropy fitness determination for the traditional aperture, Levin aperture, and zone plate aperture respectively. Figure 5-3, Figure 5-5 and Figure 5-7 the mean and standard deviation of the fitted Gaussian approximation of the true depth given a measured depth, again for the traditional aperture, Levin aperture and zone plate aperture respectively. For Figure 5-3, Figure 5-5 and Figure 5-7, the mean of each plot is shown as a solid green line and the standard deviation about the mean is shown in red dotted lines.
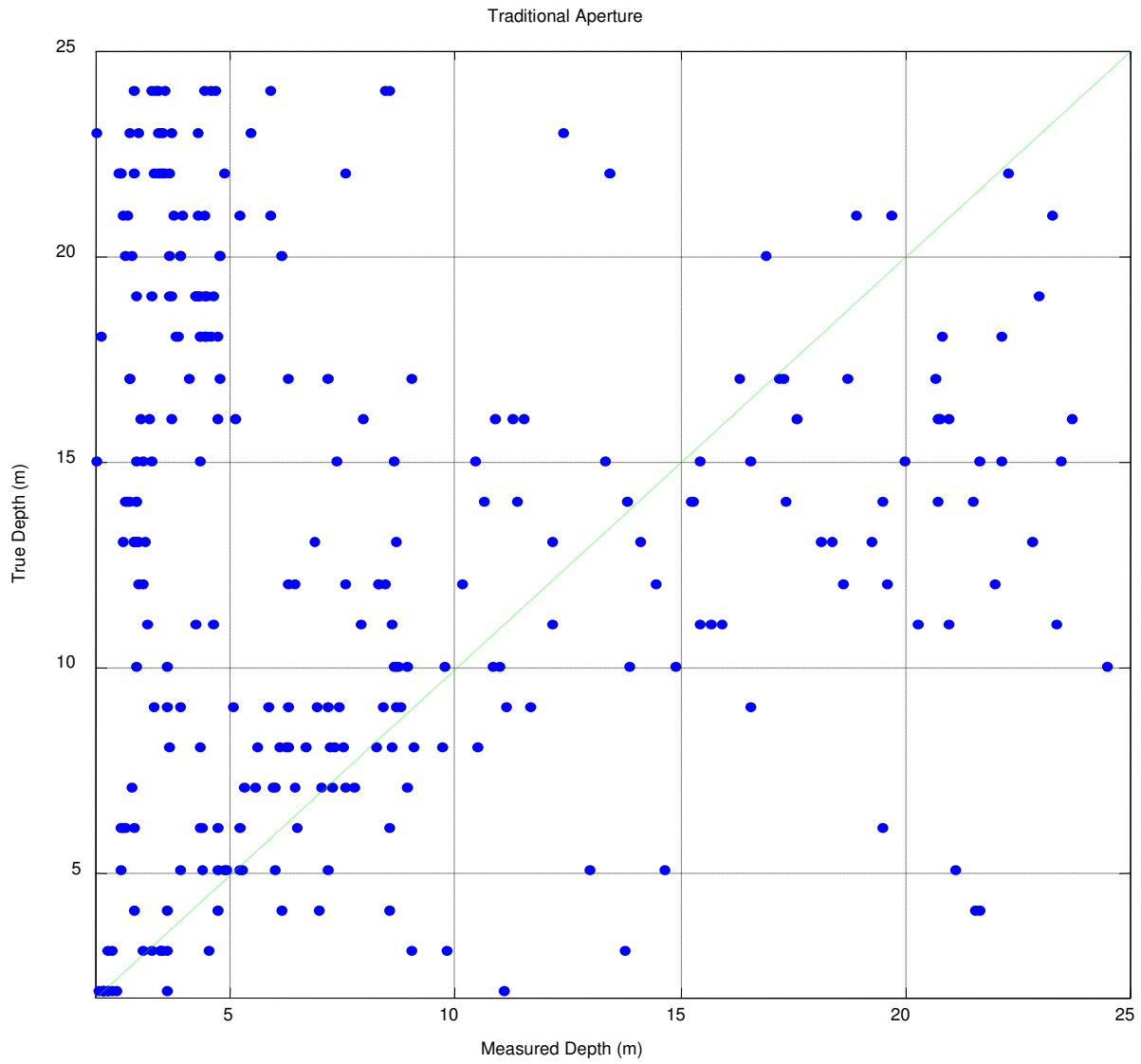
Figure 5-2 True versus Measured Depth using Traditional Aperture

104
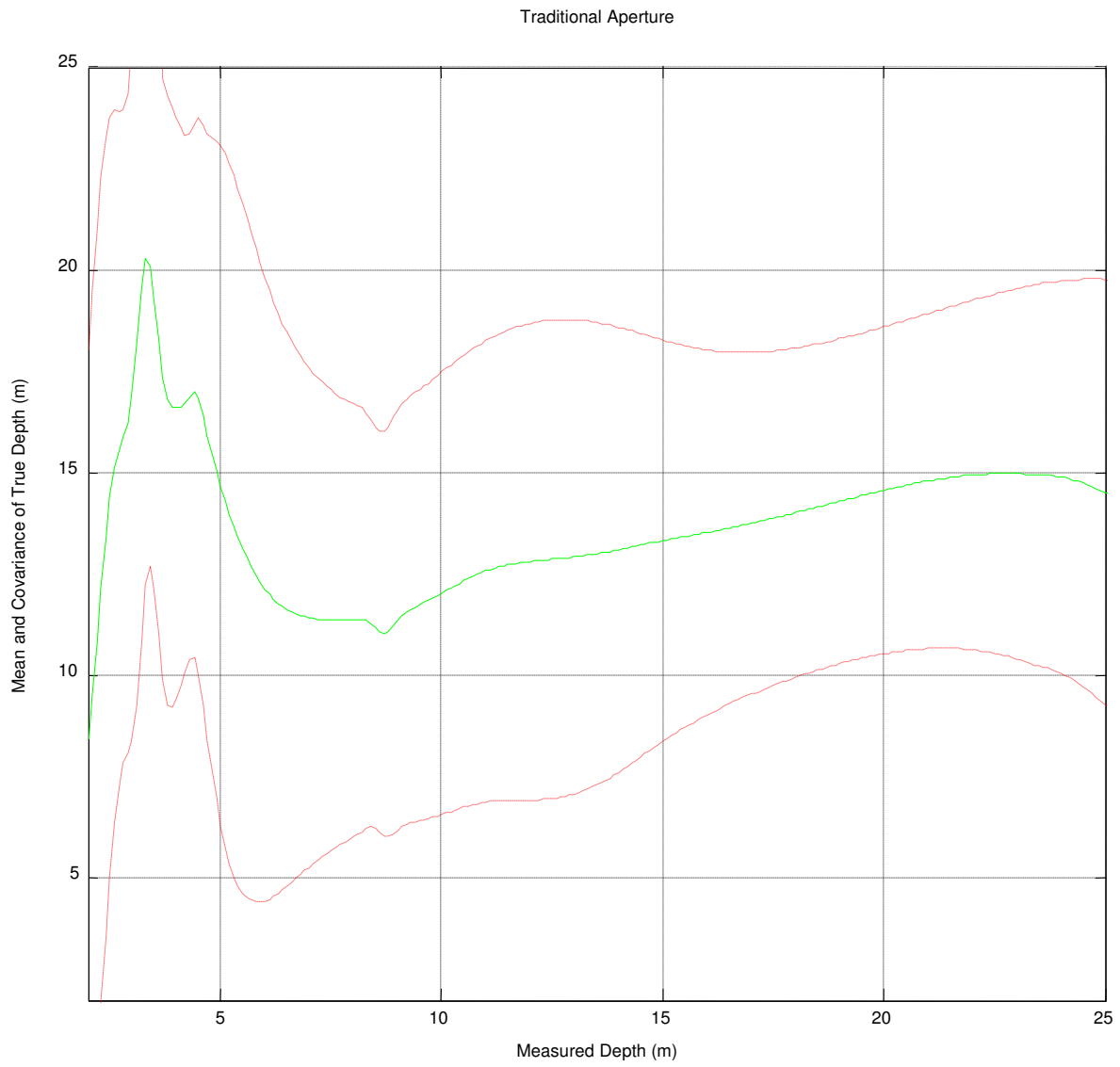
Figure 5-3 Statistics of Depth Given Measured Depth for Traditional Aperture
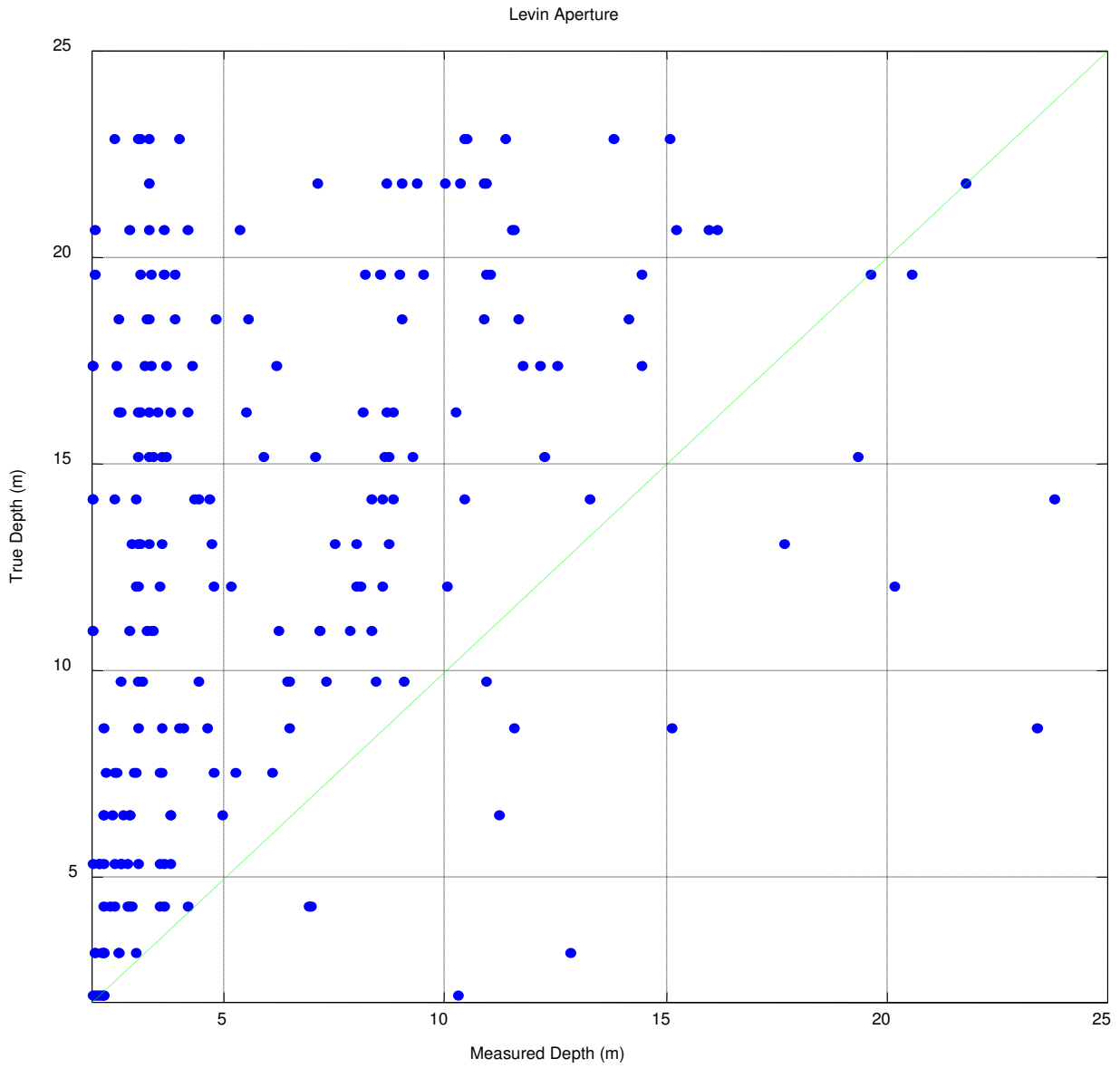
Figure 5-4 True versus Measured Depth using Levin Aperture
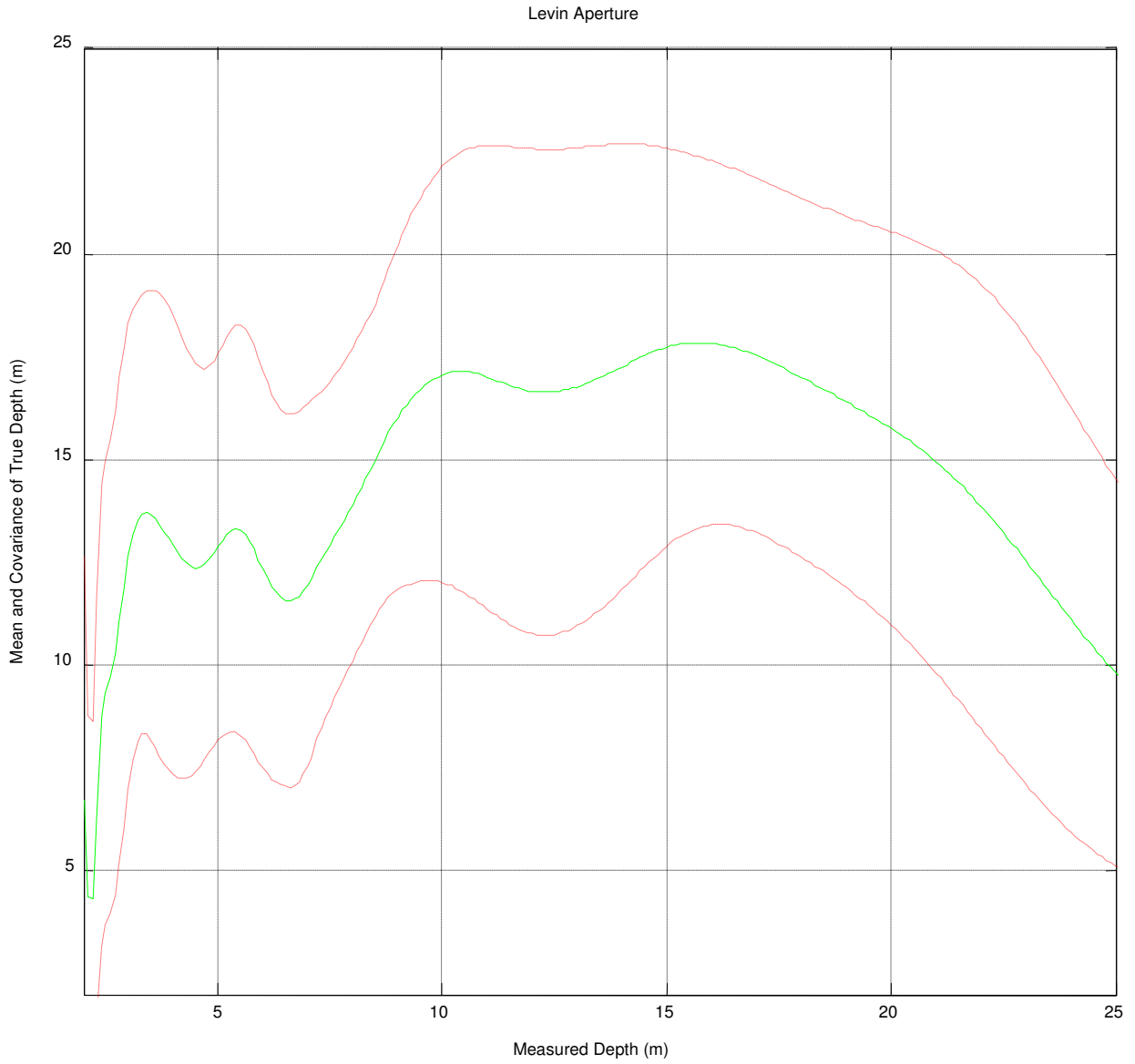
Figure 5-5 Statistics of Depth Given Measured Depth for Levin Aperture
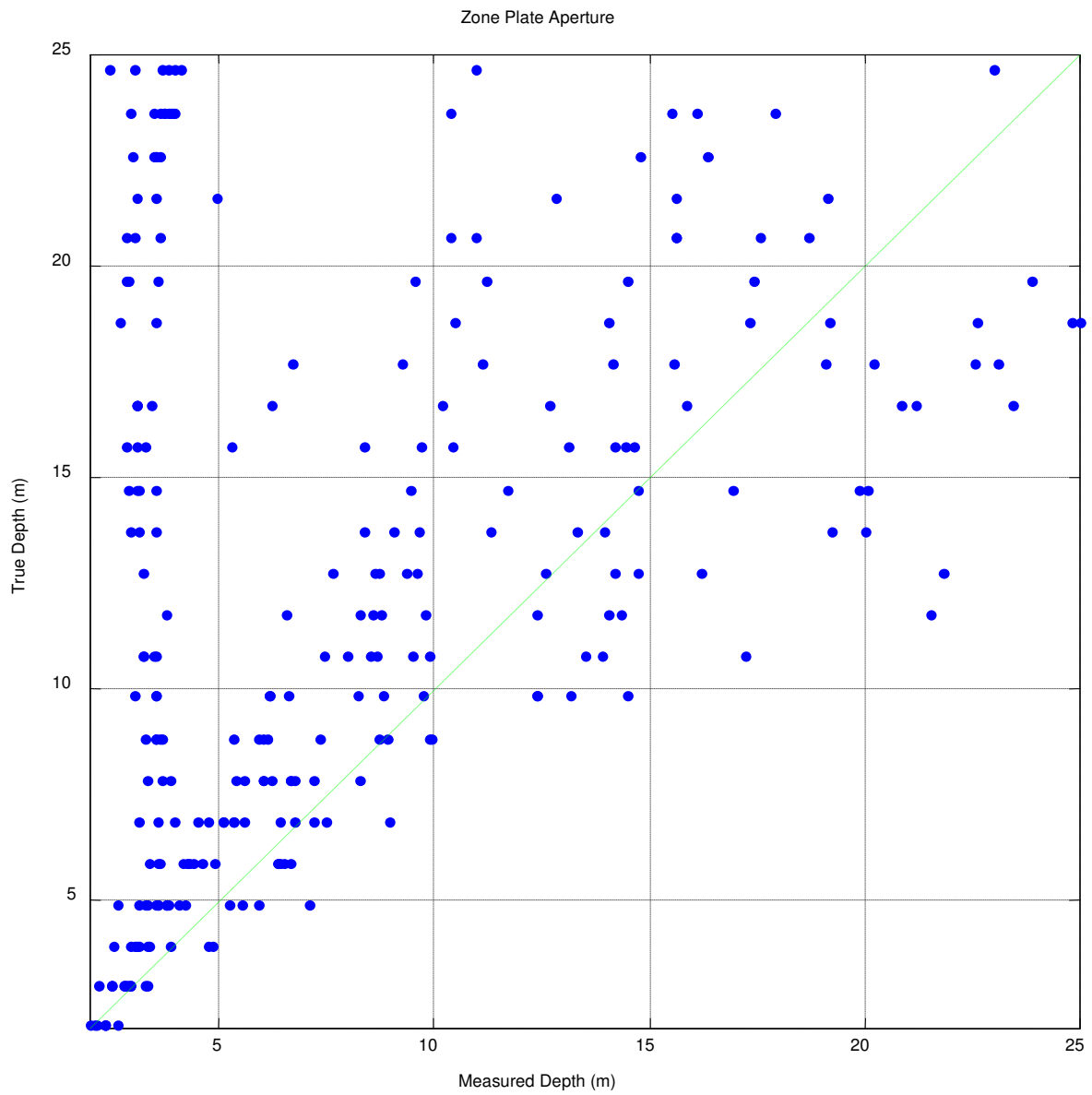
Figure 5-6 True versus Measured Depth using Zone Plate Aperture
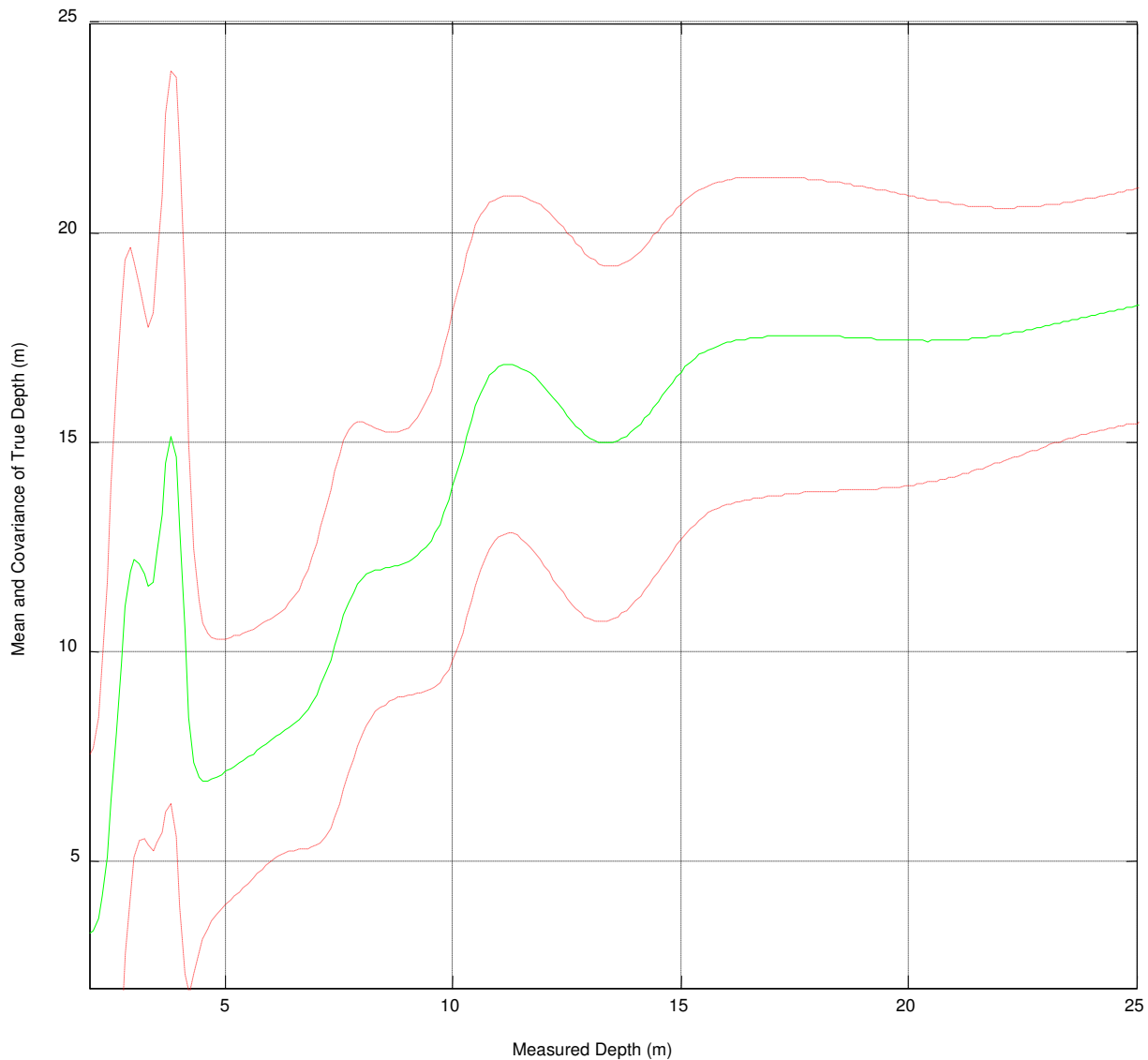
Figure 5-7 Statistics of Depth Given Measured Depth for Zone Plate Aperture

When true depth versus measured depth observations are not available for a given aperture, Section 4.3 suggests a depth measurement noise model that is a weighted mixture of two Gaussian distributions and a uniform distribution. The uniform portion models depth measurements that are the result of pixel noise, and one Gaussian distribution models spurious measurements near the focal plane. The second Gaussian distribution models near zero mean noise about the true depth and has a standard of deviation that increases with the true depth.

Because the true depth versus measured depth observations are available for each aperture of this research, the depth measurement noise model is derived from the previously estimated joint pdf. The conditional pdf for a depth measurement is found from the joint pdf given a true depth. The value for a noise corrupted depth measurement is then chosen by generating a random value from the found conditional pdf.

To validate the measurement noise model for each aperture, the noise models are applied to the same true depth values used to create each joint pdf and compared to the observed depth measurements shown in Figure 4-6. Figure 5-8, Figure 5-9 and Figure 5-10 show the observed depth measurement noise in a green box for the traditional aperture, Levin aperture, and zone plate aperture respectively. The other three plots in each figure are simulated noisy depth measurements using the same true depth values as the plot in the green box. The simulated depth measurement noise significantly resembles the observed depth measurement noise for each modeled aperture.
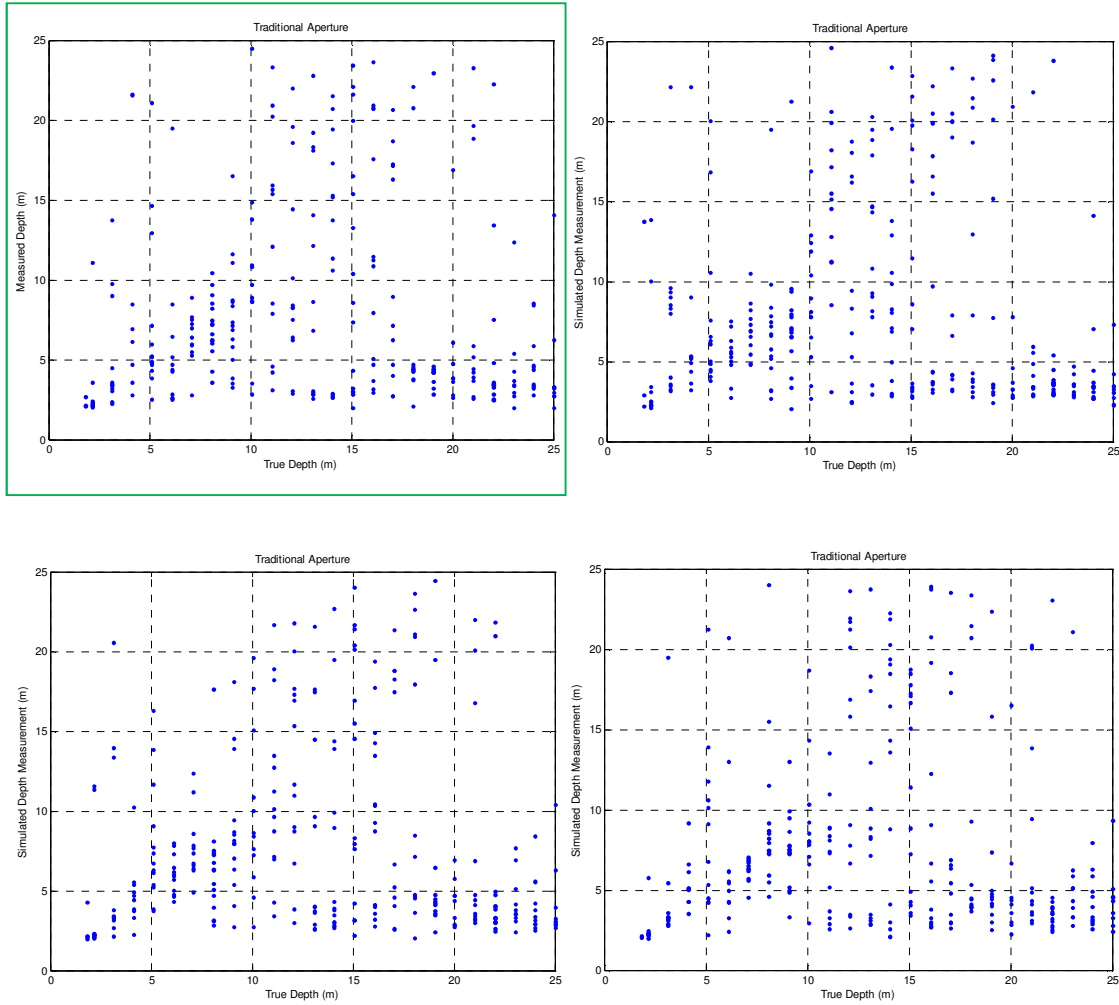
Figure 5-8 Simulation of Measurement Noise using Traditional Aperture

111
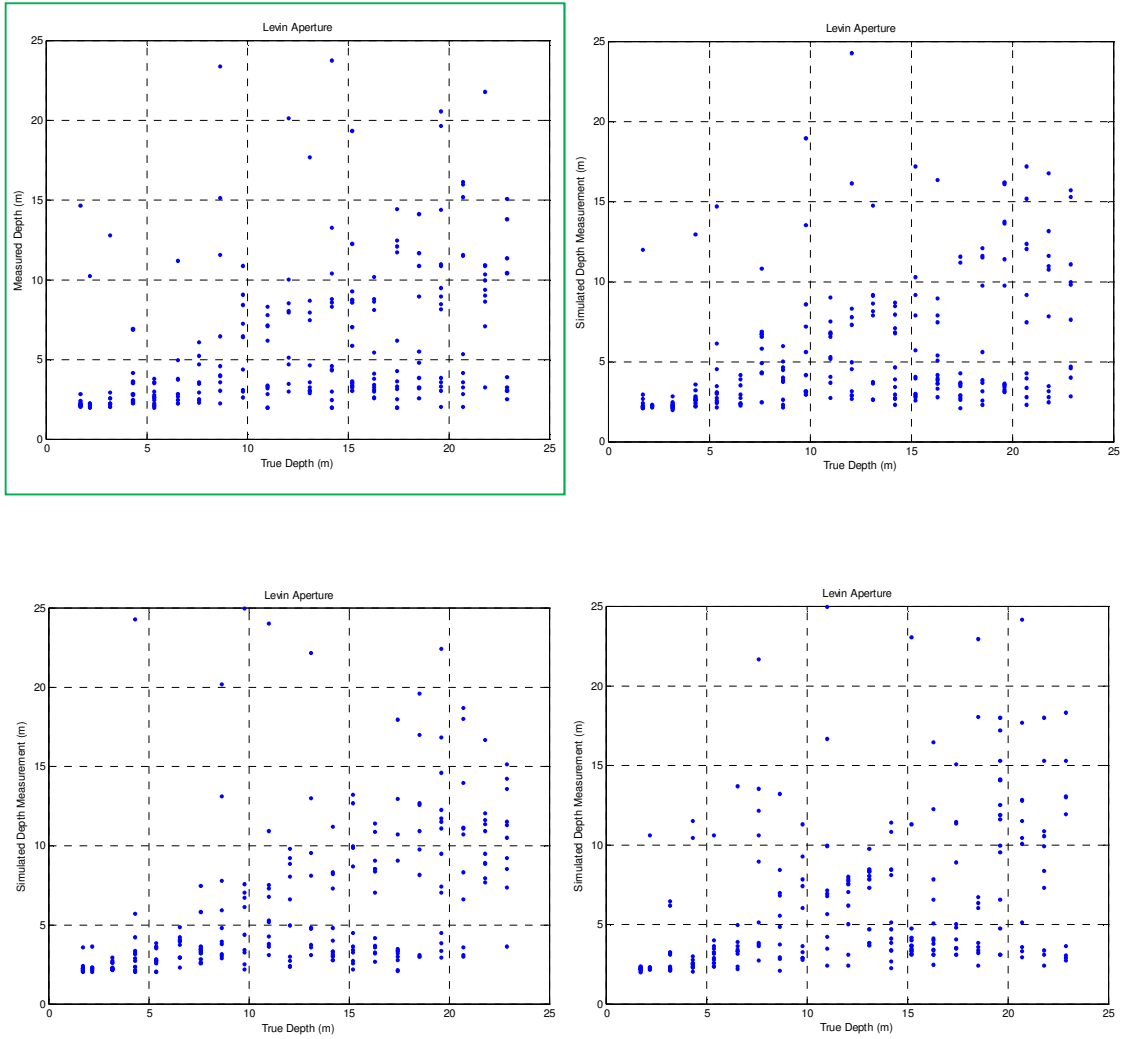
Figure 5-9 Simulation of Measurement Noise using Levin Aperture
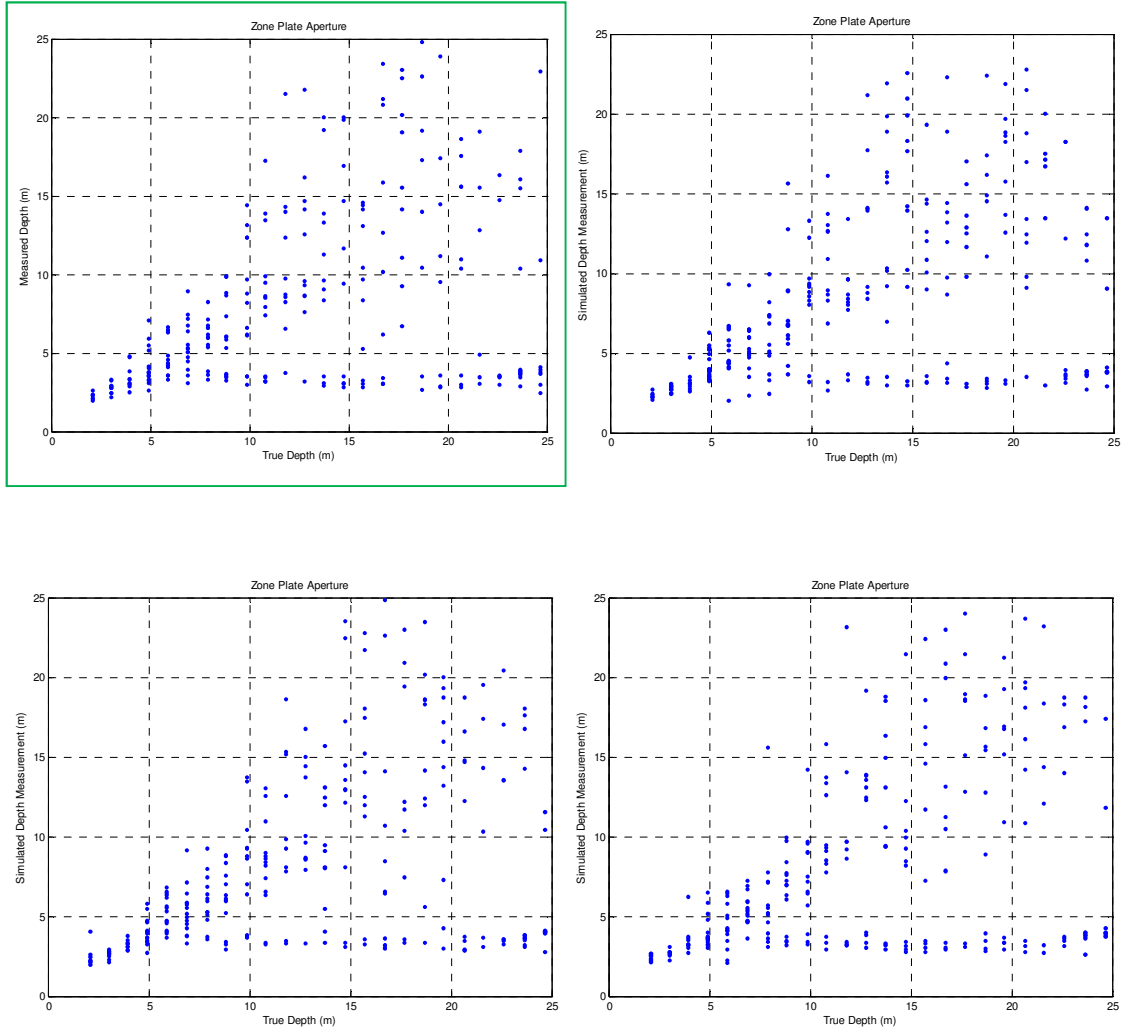
Figure 5-10 Simulation of Measurement Noise using Zone Plate Aperture

5.3 Performance Evaluation of Augmented Navigation System

For this research, a simulation of the coded aperture vision system generates image measurements such that heading and range may be determined independently for a set of feature points in each image. The pixel location and range of each feature point are then combined as a measurement of the vector from the camera to the feature, s_0^c. Full s_0^c measurements are then

compared to propagated estimates in the next image capture, and, as in [39], a search for correspondence is made within the new image. Providing a measurement of s_0^c allows not only the inclusion of the additional range measurement information in determining a navigation solution, but also mitigates errors due to linearization assumptions when updating $\underline{s}_0^c$ estimates with new measurements. Because range is determined for each image without use of previous images, a Markov process results such that each new state estimate depends only on the previous estimate and the current measurement.

To simulate the coded aperture vision aided inertial navigation system, the simulator produced by [39] for optimal image aided inertial navigation using standard images was altered. Two hundred test scenarios were created in which the vehicle is simulated as remaining stationary for 60 seconds, then accelerating until traveling north at approximately 0.5 meters per second until 500 seconds. This scenario simulates moving along a straight hallway. The INS is simulated as a commercial grade Crista MEMS INS operating at 100 Hz, and the camera is assumed to have a single coded aperture from which depth is measured using depth-from-defocus. The depth measurement statistics are assumed to be the same as those discussed in Section 5.2 and the camera operates at two frames per second. The cameras are positioned on the vehicle north and slightly east of the INS. Feature correspondence between image frames is determined using expected pixel location and SIFT as described in [39], however feature depth is not considered when establishing correspondences.

For each true depth to a corresponding feature, noise is added to the measurement using the non-Gaussian noise models given in Figure 5-8, Figure 5-9 or Figure 5-10 as appropriate. From the noisy measurement, an estimate of the depth is made using the Gaussian noise models

given in Figure 5-3, Figure 5-5 or Figure 5-7 as appropriate. Because the noise models only cover a true depth interval of two to twenty-five meters, the features detectable by the navigation system are limited to this interval.

5.3.1 Traditional Aperture System. Figure 5-11 shows the ensemble position error of two hundred runs of using the traditional aperture for depth determination. When the vehicle first begins to move, there is a position error in the direction of travel in which the vehicle first overestimates, then underestimates the movement. This error is due to bias in depth measurements that are uncorrected by monocular information while the vehicle is stationary.
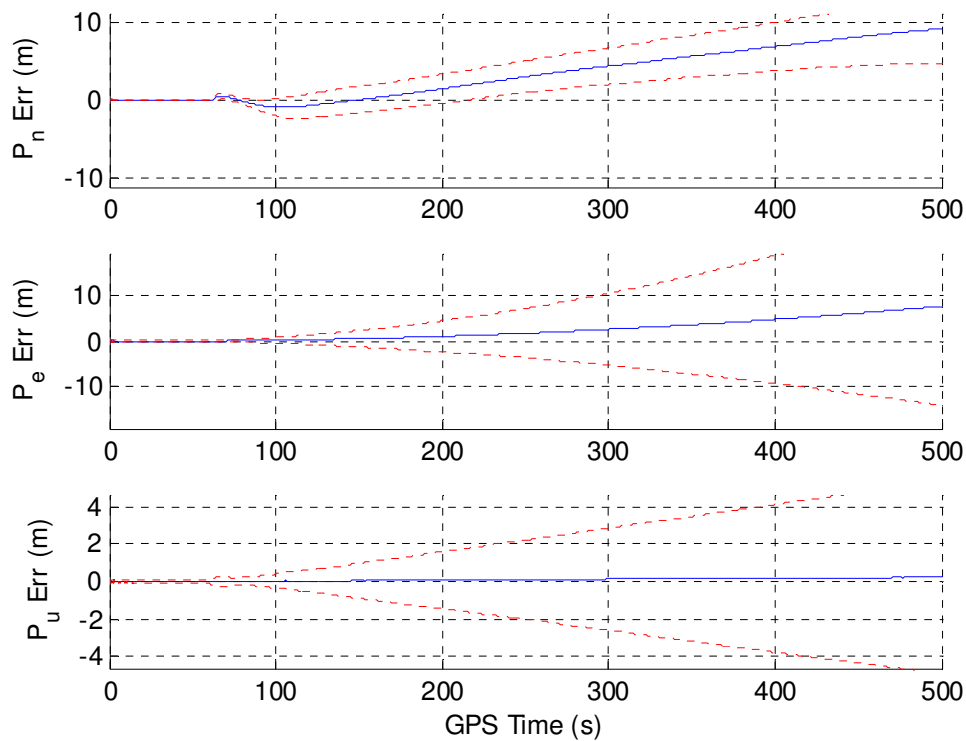
Figure 5-11 Traditional Aperture System Position Error

(Ensemble results, blue line is mean and red is mean $\pm$ standard deviation)

115

As the vehicle begins to move, new features are tracked to aid in navigation. Figure 5-8 shows that range error due to pixel noise is greater at more distant features. Also, Figure 5-3 shows a true depth estimate mean of 18 meters and standard deviation of only 6.8 meters for the peak uncorrected pixel noise depth measurements of 3.5 meters. For features with a true depth near 3.5 meters or features with an uncorrected measurement near 3.5 meters due to pixel noise, the 18 meter depth estimate is a significant bias. This bias results in a navigation solution overestimation of movement in the direction of vehicle movement.

5.3.2 Levin Aperture System. Figure 5-12 shows the ensemble position error of two hundred runs of using the Levin aperture for depth determination. The navigation performance using depth from defocus and a Levin aperture significantly improves the consistent error in the direction of movement present in the Traditional aperture.

When the vehicle first begins to move, there is an error in the direction of travel in which the vehicle first overestimates, then underestimates the movement. This error is due to bias in depth measurement that is uncorrected by monocular information while the vehicle is stationary. Because the Levin aperture does not distinguish larger depths as well as the traditional aperture, the initial movement error in the direction of vehicle movement is generally greater than the similar error for the traditional aperture.

As the vehicle begins to move, new features are tracked to aid in navigation. Figure 5-9 shows that range error due to pixel noise is not significantly greater at more distant features. Also, Figure 5-5 shows a true depth estimate mean of 14 meters and standard deviation of 5.2

meters for the peak uncorrected pixel noise depth measurements of 3.3 meters. For features with

a true depth near 3.3 meters or features with an uncorrected measurement near 3.3 meters due to

pixel noise, the 14 meter depth estimate is not as significant of a bias as was present with the

traditional aperture.
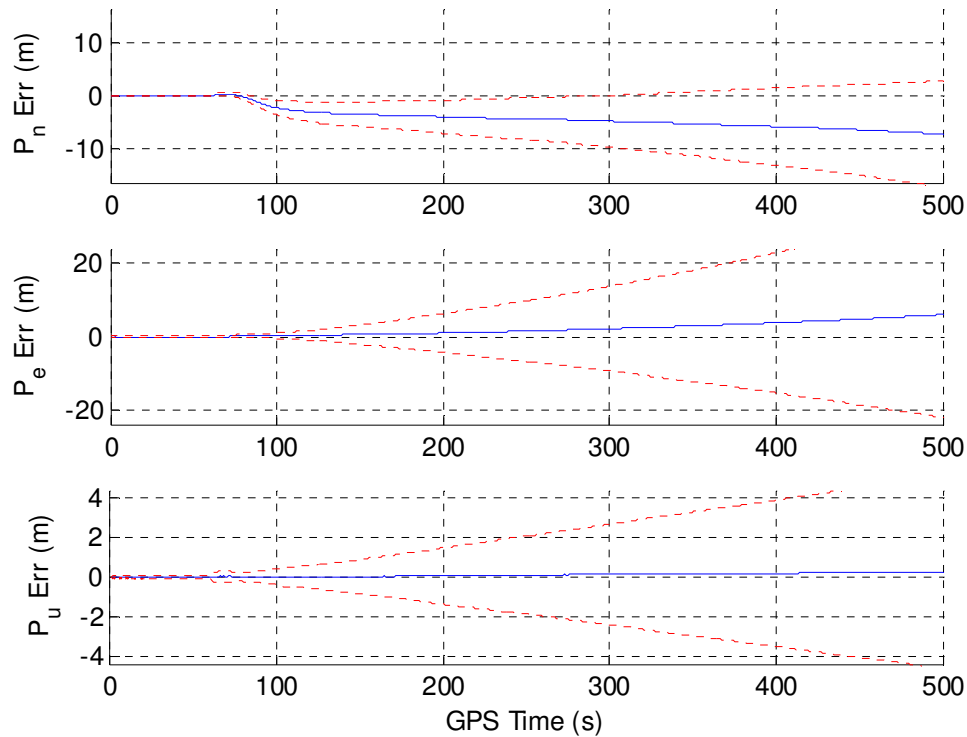
Figure 5-12 Levin Aperture System Position Error

(Ensemble results, blue line is mean and red is mean $\pm$ standard deviation)

5.3.3 Zone Plate Aperture System. Figure 5-13 shows the ensemble position

error of two hundred runs of using the Fresnel zone plate aperture for depth determination. The

standard of deviation in the navigation solution using depth from defocus and a Fresnel zone

plate aperture is improved from the Levin aperture. Also the Fresnel zone plate aperture system does not show the consistent error during movement observed with the traditional aperture.

When the vehicle first begins to move, there is a position error in the direction of travel in which the vehicle first overestimates, then underestimates the movement. This error is due to bias in depth measurements that are uncorrected by monocular information while the vehicle is stationary.

As the vehicle begins to move, new features are tracked to aid in navigation. However, Figure 5-10 shows that range error due to pixel noise is not significantly greater at more distant features. Also, Figure 5-7 shows a true depth estimate mean of 14 meters and standard deviation of 8.5 meters for the peak uncorrected pixel noise depth measurements of 3.8 meters. For features with a true depth near 3.8 meters or features with an uncorrected measurement near 3.8 meters due to pixel noise, the 14 meter depth estimate is not as significant a bias as was with the traditional aperture. As a result, the mean of the north position error at 500 seconds is similar to the mean north position error at 100 seconds.
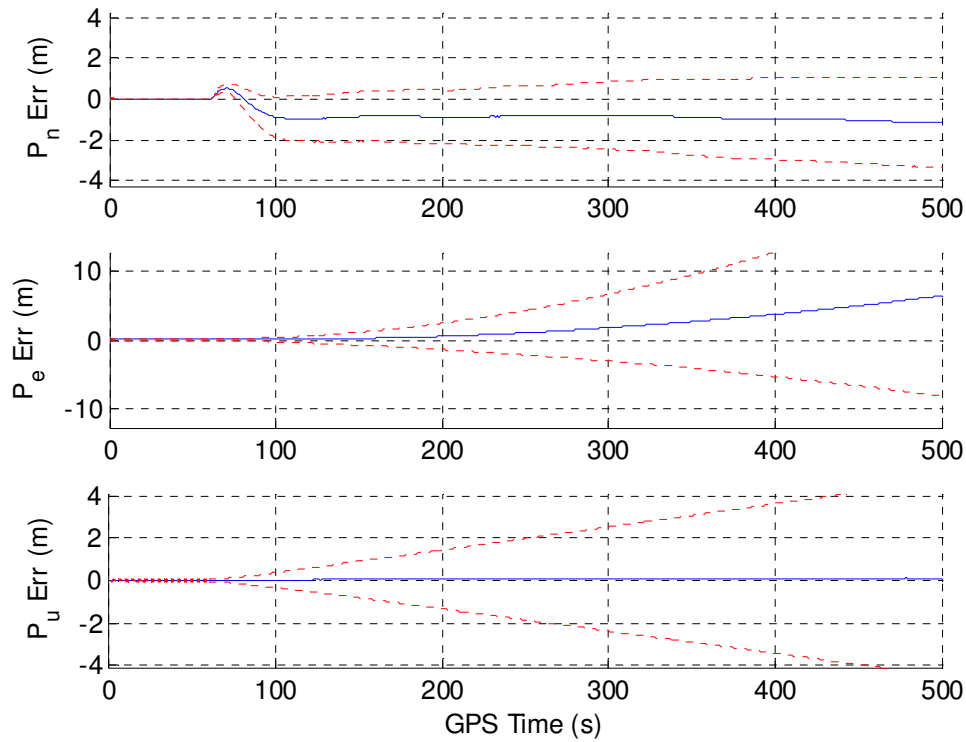
Figure 5-13 Zone Plate Aperture System Position Error

(Ensemble results, blue line is mean and red is mean $\pm$ standard deviation)

5.4 Comparison to Similar Systems

As a comparison, a stereoscopic version of the system was tested using the simulator presented in [39] using the same Crista MEMS INS operating at 100 Hz and two cameras each operating at two frames per second. For consistency with the depth-from-defocus cameras presented in Section 5.3, cameras in the stereoscopic system are limited to observations within 25 meters in depth.

The existing simulator allows initialization of a feature location through stereoscopic measurement. Then, an initial estimate of the feature is established, and each camera measures the homogenous vector from its own camera frame to the feature as an independent update to the navigation solution. The range to the feature is incorporated in the measurement indirectly by the differences in the perspectives of the two cameras. Depth is used directly rather than using the inverse of depth as suggested by [27]. Figure 5-14 shows the ensemble position error of the two hundred runs of the simulated stereoscopic system. P_n is the north position error, P_e is the east position error, and P_u is the up position error, were all position errors are in meters.

Figure 5-14 suggests the position error of the stereoscopic system relative to the Fresnel zone plate system is decreased by an order of magnitude in the north direction, decreased by half in the east direction, and similar in the up direction. The cameras are positioned on either side of the INS so as to be east and west of each other at the start of the simulation. The east position error is reduced as a result of the east-west separation of the cameras.

A monocular version of the system depicted in Figure 5-14 was also tested by removing one of the cameras and allowing initial depth estimates to be randomly selected. Only the initial depth estimate was randomly selected; subsequent measurements rely on heading measurements to features only. The depths were selected with a mean distance of 20 meters, but only features with a distance of less than 25 meters were selected for tracking to resemble the systems presented in Section 5.3. The relatively long mean distance of 20 meters was chosen to favor low parallax angles, thereby reducing linearization errors [27]. The simulation also used a 100 MHz Crista INS and a single two frame per second camera. Two hundred runs were tested and

120

at the end of 500 seconds, the navigations solution consistently diverged such that the standard deviation of the position error in all directions was on the order of 10^5 meters.
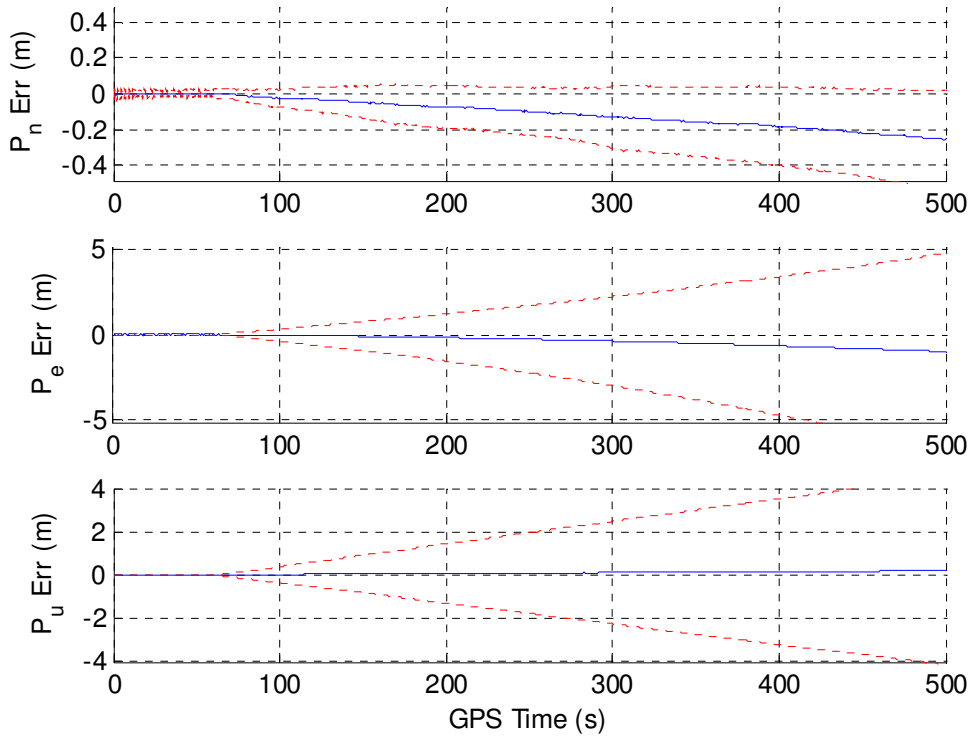
Figure 5-14 Stereoscopic System Position Error

(Ensemble results, blue line is mean and red is mean $\pm$ standard deviation)

The navigation solution diverges dramatically in the monocular system because the vision system is no longer able to establish correspondence between observed features and tracked features. Using the vehicle location estimate and feature location estimates, a mean and covariance of the predicted feature locations relative to the new vehicle position is determined for each frame. When establishing correspondence between a tracked feature and an observed

feature in a new frame, only observed features of the frame that are near the location in which the tracked feature is predicted to appear are searched. The overall search space for observed features in the new image is equal to twice the standard deviation of the predicted relative location of a tracked feature. When the actual relative feature location in the scene is greater than twice the standard deviation of the predicted relative feature location, the actual location of the feature in the image frame is not within the search space of observed features. Correspondence between a tracked feature and an observation of that feature then does not occur. When the navigation error becomes sufficiently large, all actual relative feature locations in the scene become greater than twice the standard deviation of the predicted relative feature locations. Correspondence between observed features and tracked features cannot occur, resulting in navigation solution error growth equivalent to an unaided inertial navigation system.

A monocular system has been demonstrated by [21] using an inverse depth representation. The INS in [21] operates at 256 Hz, whereas the INS used in this study operates at 100 Hz. Also, the camera used in [21] provides 28 frames per second, which is significantly greater than the two frames per second of the cameras used in this study. Although [21] does not provide position error, the navigation solution does not diverge. The greater frame rates of the cameras in [21] may be sufficient to maintain correspondence between tracked and observed features, thus preventing divergence. The authors note an approximate 10% scale drift, or increase in estimated feature depths towards the end of test. The system appears to then overestimate movement in the direction of vehicle movement.

5.5 Conclusions From Results

This chapter presents the performance of a simulated navigation system with one camera using depth from defocus, or two cameras using stereoscopy. With the two cameras, the direction and depth to a feature are measured, and the location of the feature relative to the movement of the vehicle is used to correct the navigation solution. Removing one of these cameras and not employing depth from defocus resulted in a single-camera system with no means of estimating the depth upon the first observance of a feature. The incorrect initial depth estimates introduce linearization errors into the navigation solution. Depth may be triangulated using subsequent observances of the feature as the camera moves; however movement of the camera is part of the navigation solution that the camera is meant to solve. As a result, the navigation solution error was shown to be several orders of magnitude greater for a single camera system without an initial depth measurement method than with either the single camera depth from defocus or two camera stereoscopy systems.

Of the systems using one camera and depth from defocus presented, the Fresnel zone plate provides the greatest navigation performance. Because the pixel noise does not become prominent for features observed at greater depths, the depth estimates corrupted by pixel noise are not as strongly biased in the direction of movement as with the traditional aperture. Because the differences in depth to more distant features are observable, the initial error in the depth estimate of a newly tracked feature is lower with the zone plate than with the Levin aperture. With the notable exception of initial movement error, the variance in the position error increases for all three apertures more slowly in the direction of travel than in other directions.

As described earlier, feature location estimates are maintained by the Kalman filter to aid in navigation; however the system design tested in this research does not include this estimate when measuring the depth to a feature. Also described earlier are errors that are introduced into the depth measurements because of pixel noise that can be approximated by an aberration located at the focal plane. The performance of all three depth measurement systems may be improved by constraining the depth estimates using the feature location covariance of the navigation solution. The resultant system may significantly improve the performance of depth from defocus by reducing error when measuring more distant features, which would in turn improve the performance of the navigation system overall.

6. Conclusion

The original contributions of this research are to develop a method for using depth from defocus to measure feature direction and depth for navigation, to improve the proposed depth from defocus navigation performance by using a coded aperture, and to further improve navigation performance by using an aperture coding that is a Fresnel zone plate. This research is also the first to propose the Fresnel zone plate aperture for any depth from defocus system, and to characterize the behavior of the Fresnel zone plate aperture when used for depth measurement from defocus.

6.1 Navigation using Depth from Defocus.

This document evaluates the performance of simulated inertial navigation systems aided by a vision system that employs either two camera stereoscopy or single camera depth from defocus. With stereoscopy or depth from defocus, the direction and depth to a feature are measured, and the location of the feature relative to the movement of the vehicle is used to correct the navigation solution. Using a single-camera vision system with no means of estimating the depth upon the first observance resulted in navigation error five orders of magnitude greater than the stereoscopic system and overall divergence of the navigation solution.

Equipping the single camera system with a narrow depth of field and measuring depth to features using depth from defocus was shown to significantly restore much of the lost navigation performance. The navigation solution error of this system was shown to be only one order of

magnitude greater than the two camera system, and the solution has a positive bias in the direction of travel of the camera. This bias results from a combination of pixel noise and the smoothing effect of the focal blur from the rounded aperture. The increasing focal blur from a rounded aperture decreases edge strength and the contrast of high frequency content in the image. Pixel noise may be approximated as elements of the scene that are in focus, hence located at the focal plane. Because the pixel noise is not affected by the focal blur, apparent edge strength and high frequency content of the pixel noise is unattenuated. For distant features the focal blur may become sufficiently large, and edges and high frequency content of the scene sufficiently reduced such that the pixel noise produces depth measurements at the focal plane instead of the true depth.

As the camera moves during navigation, features that exit the field of view must be replaced with new features to track. Often the new features that are not already being tracked are more distant features, which have a greater likelihood of depth measurements that are affected by pixel noise. Linearization errors are then introduced similar to those of the single-camera system for which range is randomly selected. However, because depth measurements do not depend on the navigation solution and pixel noise induced error decreases as the true depth to the feature decreases, the navigation solution errors are mitigated.

6.2 Coded Aperture Navigation.

Augmenting the depth from defocus system with a coded aperture improves the navigation performance of the single camera system, but with a bias as the vehicle begins to move. Defocus with a coded aperture is modeled and it is shown that the high frequency content

126

from the scene does not consistently decrease as the focal blur increases. Although depth measurements from a defocus system with a coded aperture are still corrupted by the same pixel noise problem as the traditional round aperture, the pixel noise does not dominate the depth measurements beyond a given true feature depth. The coded aperture is also shown to provide an overall decrease in depth measurement noise, although with a bias at greater true depths. This bias produces feature depth estimate errors when an approximately stationary camera observes more distant features. As the camera moves, the bias in the depth measurements is abated and the overall navigation solution is unbiased.

6.3 Improvement With A Zone Plate Aperture.

The Fresnel zone plate was also modeled and analyzed as an aperture. The navigation solution with the Fresnel zone plate is, like the system with the clear aperture and first coded aperture, comparable to the two camera system. Like the first aperture coding, the high frequency content from the scene does not consistently decreases as the focal blur increases. Also like the first aperture, the pixel noise does not dominate the depth measurements beyond a given true feature depth. Unique to the zone plate, the Schuster fringes producing multiple focal planes in the scene produce multiple focal blurs in the coded image. The focal plane closer to the camera has the greatest change in the focal blur as the feature depth changes, and the focal plane furthest from the camera reduces pixel noise depth measurement obfuscation with greater edge strength and high frequency content from the scene. The bias in the depth estimates to features in the scene is reduced as compared to the traditional and Leven apertures, thereby reducing the bias in the navigation solution when the vehicle initiates movement. The overall navigation solution is also unbiased for the moving camera.

6.4 Proposed Future Work

It is assumed in this research that if the depth to a feature is known, then the focal blur is also known and can be deconvolved from the coded image to reveal an estimate of the uncoded image. From the estimated uncoded image, the pixel location of the corner provides the direction information used to correct the navigation solution. However the relationship between the coded image and corner detection and location error should be further explored.

The Kalman filter of the proposed system includes a depth estimate for each tracked feature. This depth estimate may be used to mitigate pixel error in the depth from defocus system. Pixel error may be approximated as an aberration located at the focal plane in the scene, and the feature location estimate from the Kalman filter may be used to exclude the focal plane from the interval over which the depth to the feature is to be measured. The depth measurement performance of the depth from defocus system would improve, and the performance of the overall navigation system would improve as well.

In the proposed system, when a feature is selected to be tracked, a depth measurement to the feature is made and it is immediately included in the Kalman filter. An alternative method may selectively delay the inclusion of one or more candidate features until multiple observances are made. Also, an optimal means of selecting a feature to be tracked may be determined.

In the proposed system, depth is used directly rather than inversely as suggested by [27]. Several systems have shown significant improvement by using the inverse of depth in the

Kalman filter [13, 21, 27]. The change in performance of this system should be determined when using inverse depth rather than using depth directly.

A color camera was tested during this research; however the color filter introduced errors in the measurement of the point spread function. Establishing a reliable method for capturing the point spread function in the presence of a color filter, such as the Bayer color filter array [3], would allow multiple wavelengths in the scene to be measured concurrently. The performance of a depth from defocus system that uses each color to produce multiple independent depth measurements may also be explored.

A Fresnel zone plate with several hundred zones was attempted for depth measurement; however the diffraction introduced aliasing errors in the coded images. The relationship between diffraction from a zone plate aperture and aliasing can be explored to allow determination of an optimal number of zones to use in the zone plate for a given optical system.

The modeling method given provides estimates of the effect of scaling the optical system with respect to the wavelength of light, however measurements of an optical system scaled to approximately the size of a cellular phone camera should be captured and compared for validation. Micro optical system may also be evaluated to determine if there exists a lower bound on the scale of an optical system for which depth from defocus with a given aperture may be used. This would also aid in determining the navigation performance of very small MAVs equipped with a depth from defocus system.

Bibliography

1. Barbour, B. A., Chenault, D. B., United States: Patent 5890095, 1997.

2. Bates, D. P., *Navigation Using Optical Tracking of Objects at Unknown Locations*. Master's Thesis, Russ College of Engineering and Technology of Ohio University, 2007.

3. Bayer, B. E., United States: Patent 3971065, 1976.

4. Brenner, C., *Tutorial "Registration and 3D modeling of aerial laser scanner data".* International Summer School Digital Recording and 3D Modeling, 24-29 April, Aghios Nikolaos, Crete, Greece, 2006.

5. Brown, A., Silva R., "Video-aided GPS/INS Positioning and Attitude Determination," *Proceedings on the 12th International Technical Meeting of the Satellite Division of the Institute of Navigation ION GPS*, September, 1999.

6. Diel, David D, P. DeBitetto, S. Teller, "Epipolar Constraints for Vision-Aided Inertial Navigation," *Proceedings of the IEEE Workshop on Motion and Video Computing*, vol. 2, p. 221-228, 2005.

7. Dou, Q., Favaro, P., "Off-Axis Aperture Camera: 3D Shape Reconstruction and Image Restoration," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 1 – 7, August, 2008.

8. Ens, J., Lawrence, P., "An Investigation of Methods for Determining Depth from Focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-15, no. 2, p. 97-108, February, 1993.

9. Ettinger, S.M.; Nechyba, M.C.; Ifju, P.G.; Waszak, M., "Vision-guided flight stability and control for micro air vehicles," *Intelligent Robots and System, 2002. IEEE/RSJ International Conference on* , vol.3, p. 2134-2140, 2002

10. Fearing, R.S.; Chiang, K.H.; Dickinson, M.H.; Pick, D.L.; Sitti, M.; Yan, J., "Wing transmission for a micromechanical flying insect," *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol.2, p.1509-1516, 2000

11. Forsyth, David A. and Jean Ponce, *Computer Vision: A Modern Approach.* Prentice Hall, Upper Saddle River, New Jersey, 2003.
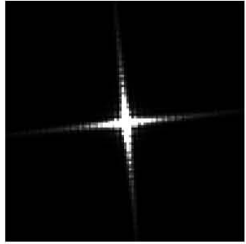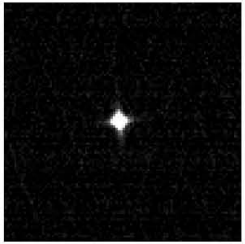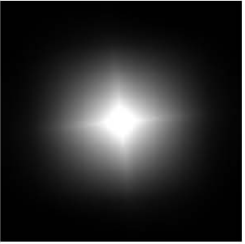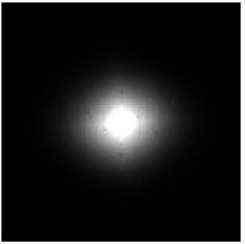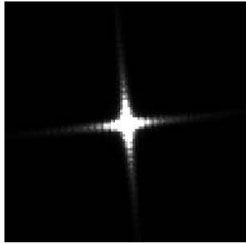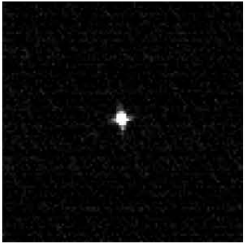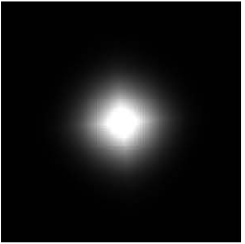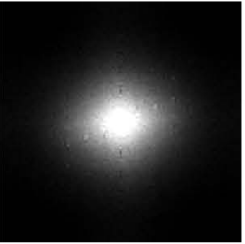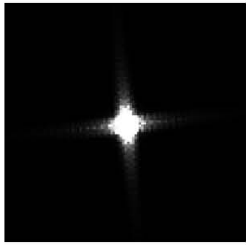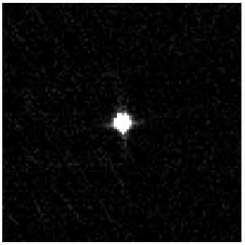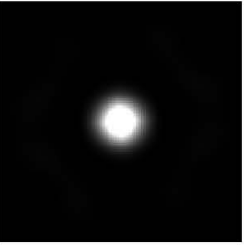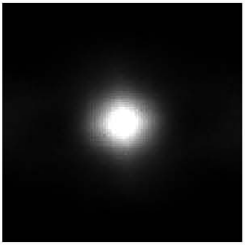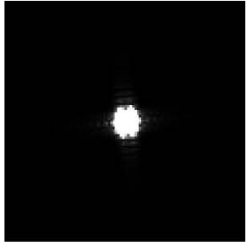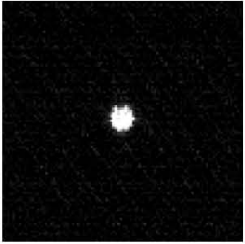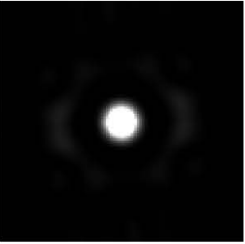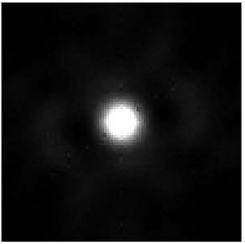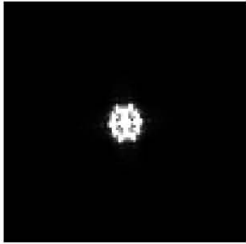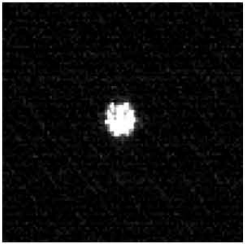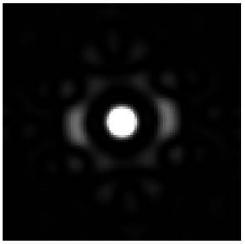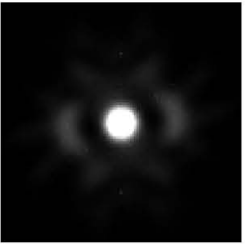
12. Fresnel, Augustin-Jean, "Calcul de l'intensité de la Lumière au centre de l'ombre d'un Écran et d'une Ouverture circulaires éclaires par un point radieux," *Mémoires de l'Académie des sciences de l'Institut de France*, Vol. 1, Note 1, L'Imprimerie Royale, France, 1826.

13. George, Michael and Salah Sukkarieh, "Inertial Navigation Aided by Monocular Camera Observations of Unknown Features," *IEEE International Conf. on Robotics and Automation*, April, 2007.

14. Georgiev, T., Zheng, K. C., Curless, B., Salesin, D., et al., "Spatio-Angular Resolution Tradeoff in Integral Photography," *ACM Transactions on Graphics*. vol. 26, July, 2007.

15. Goodman, J. W., *Introduction to Fourier Optics, 3rd Ed.*, Roberts and Company, Englewood, CO, 2005.

16. Groves, Paul D., *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. Artech House, Boston, MA, 2008.

17. Guenther, Robert, *Modern Optics*, John Wiley & Sons, New York, 1990.

18. Harburn, G., T. R. Welberry and R. P. Williams, "A Fourier-series approach to moiré patterns with special reference to those produced by overlapping zone plates," *Optica Acta*, vol. 22, Taylor & Francis Ltd., London, 1975.

19. Indebetouw, Guy and B. Hanxian, "Imaging with Fresnel zone pupil masks: extended depth of field," *Applied Optics*, vol. 23, p. 4299-4302, Optical Society of America, 1984.

20. Kellogg, J., C. Bovais, and R. Foch, et al, "The NRL micro tactical expendable (MITE) air vehicle," *Aeronautical Journal*, 106(1062), p. 431-441, August, 2002.

21. Kleinert, Markus and S. Schleith, "Inertial Aided Monocular SLAM for GPS-Denied Navigation," *2010 IEEE International Conf. on Multisensor Fusion and Integration for Intelligent Systems*, p. 20-25, September 5-7, 2010.

22. Levin, A., Fergus, R., Durand, F., Freeman, W., "Image and Depth from a Conventional Camera with a Coded Aperture," *ACM Trans. Graph.* 26, 3, Article 70, July, 2007.

23. Lowe, David G. "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, 2004

24. Markoff, J., "Crashes and Traffic Jams in Military Test of Robotic Vehicles," *The New York Times*, November 5, 2007.
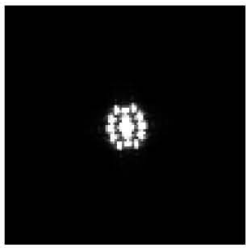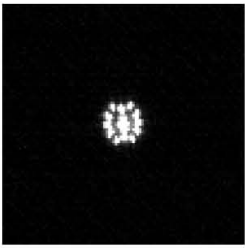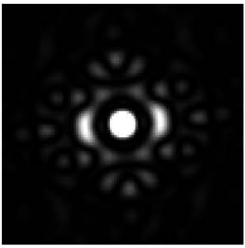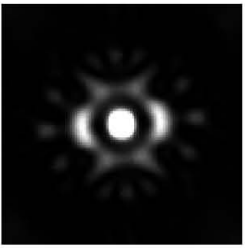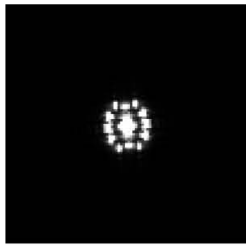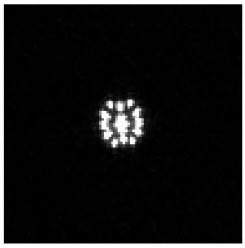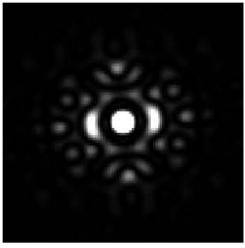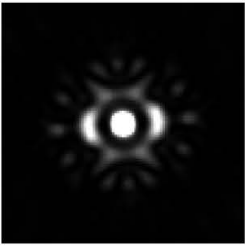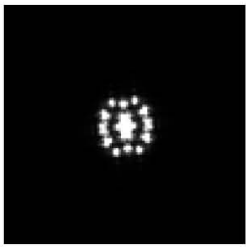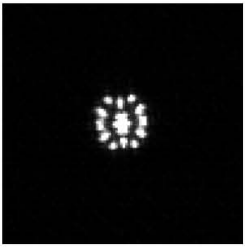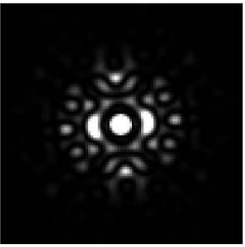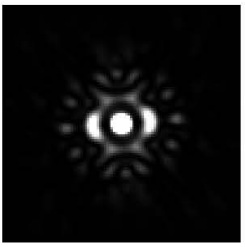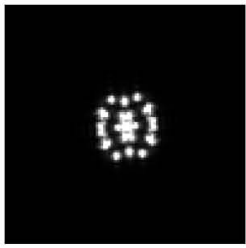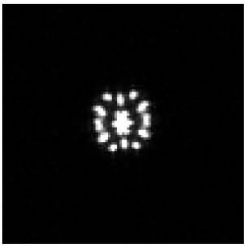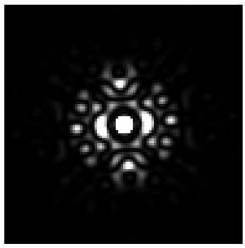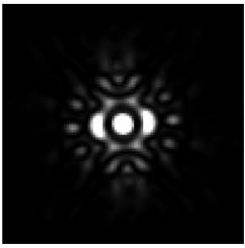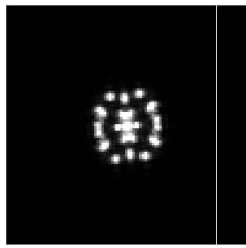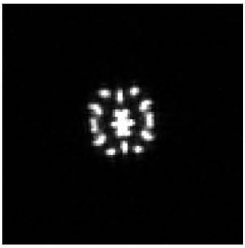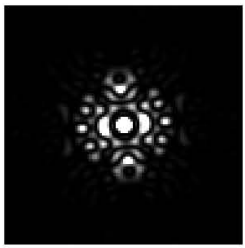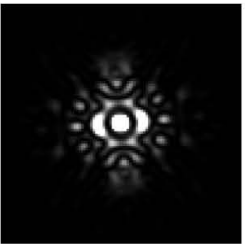
25. Merico, D.; Bisiani, R., "Indoor Navigation with Minimal Infrastructure," *Positioning, Navigation and Communication*, 2007. WPNC '07. 4th Workshop on, p.141-144, March 22, 2007.

26. Mohr, B.B.; Fitzpatrick, D.L., "Micro air vehicle navigation system," *Aerospace and Electronic Systems Magazine, IEEE* , vol. 23, no.4, p.19-24, April, 2008.

27. Montiel, J. M. M., Civera, J., Davison, A. J., "Unified Inverse Depth Parameterization for Monocular SLAM," *Robotics Science and Systems*, 2006.

28. Myers, Ora E., "Studies of Transmission Zone Plates," *American Journal of Physics*, vol. 19, p. 359-365, American Association of Physics Teachers, 1951.

29. Nielsen, M. B., "Development and Flight Test of a Robust Optical-Inertial Navigation System using Low-Cost Sensors," M. S. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, 2008.

30. Pentland, A. P., "A New Sense for Depth of Field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, p. 523-531, July, 1987.

31. Russel, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey, 2003.

32. Schechner, Y. Y., Kiryati, N., "Depth from Defocus vs. Stereo: How Different Really are They?" *Fourteenth International Conference on Pattern Recognition*, vol. 2, p. 1784-1786, August 16-20, 1998.

33. Schuster, Arthur, "On Passage of Light Through Transparent Plates," *Philosophical Magazine*, series 6, vol. 48, p. 609-619, October 1924.

34. Stewart, K, Wagener, J., Abate, G., Salichon, M., "Design of the Air Force Research Laboratory Micro Aerial Vehicle Research Configuration," *AIAA Aerospace Sciences Meeting*, Reno, NV, January 8-11, 2007.

35. Sun, Jin-Ao and Cai, A., "Archaic focusing properties of Fresnel zone plates," *Journal of Optical Society of America A*, vol. 8, p. 33-35, January 1991.

36. Thrun, S., Montemerlo, M., Dahlkamp, D., et.al, "Stanley: The Robot That Won The DARPA Grand Challenge," *Journal of Field Robotics*, vol. 23, no. 9, p. 661-692, June 9, 2006.

37. Titterton, D. H. and J. L. Weston, *Strapdown Inertial Navigation Technology*. Peter Peregrinus Ltd., Lavenham, United Kingdom, 1997.

38. Veeraraghaven, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J., "Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing," *ACM Transactions on Graphics*. vol. 26, July, 2007.

39. Veth, Michael J. and J. F. Raquet, *Fusion of Imaging and Inertial Sensors for Navigation*. PhD dissertation, Air Force Institute of Technology, 2006.

40. Volpe, J. A., *Vulnerability Assessment of the Transportation Infrastructure Relying on the Global Positioning System.* National Transportation Systems Center, Office of the Assistant Secretary for Transportation Policy, U. S. Department of Transportation, August, 2001.

41. Watanabe, M., Shree, N. K., "Telecentric Optics for Focus Analysis,", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, December, 1997.

42. Wood, R. W., "Phase-Reversal Zone-Plates, and Diffraction-Telescopes," *Philosophical Magazine*, Series 5, vol. 45, p. 511-523, June 1898.

43. Yang, H., J. Xia, B. Wang, "Computational Reconstruction Three-Dimensional Object Using Integral Imaging," *1st International Conference on Information Science and Engineering*, p. 512 – 515, April, 2010.

44. Young, M., "Zone Plates and Their Aberrations," *Journal of the Optical Society of America*, vol. 62, p. 972-976, August 1972.

45. Zhou, C., S. Lin, S. Nayar, "Coded Aperture Pairs for Depth from Defocus," *2009 IEEE 12th International Conference on Computer Vision*, p. 325 - 332, October, 2009.
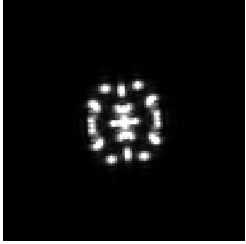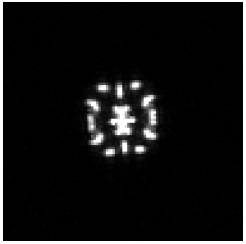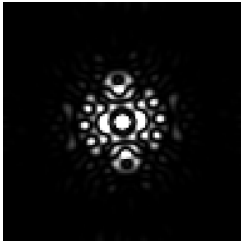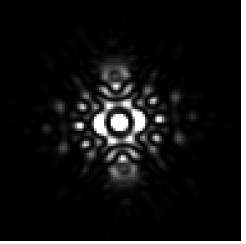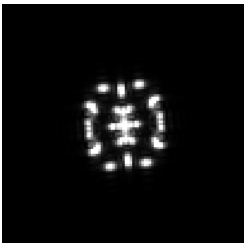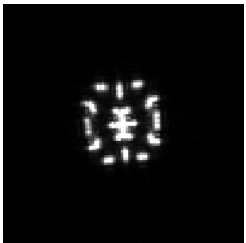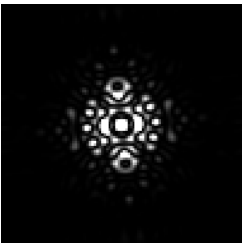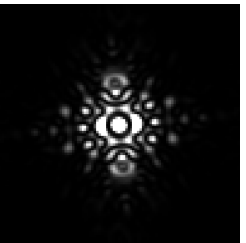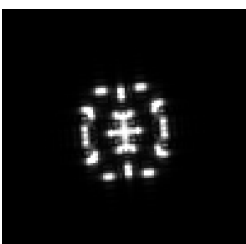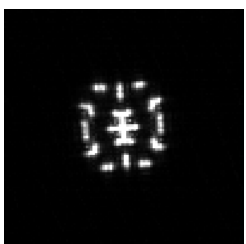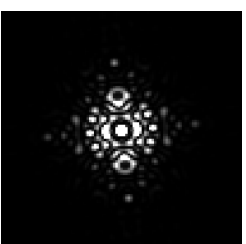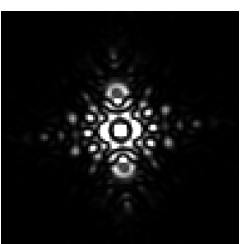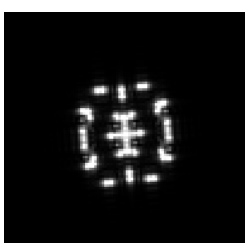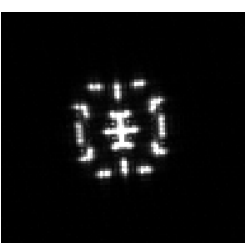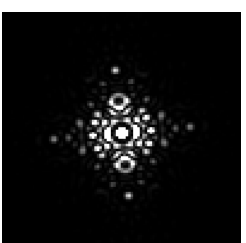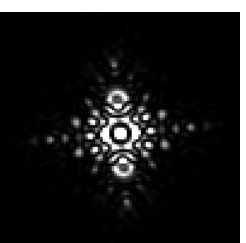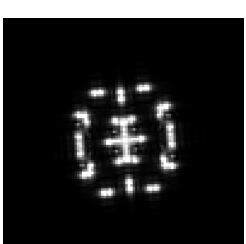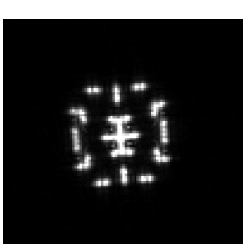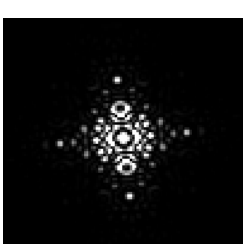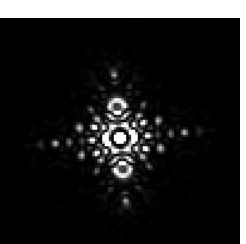
Appendix A
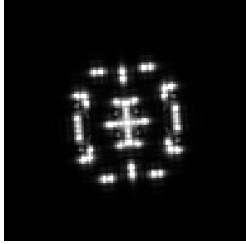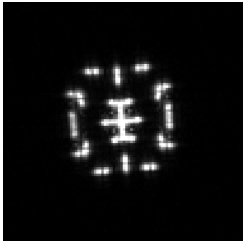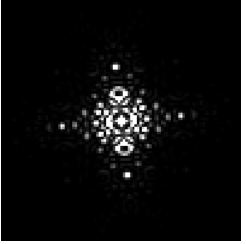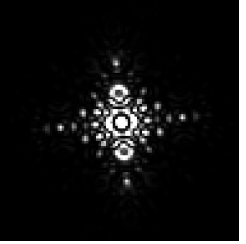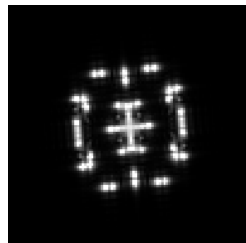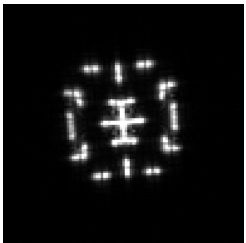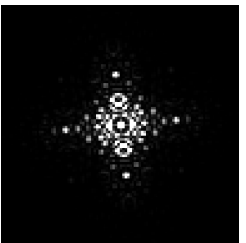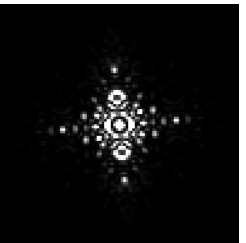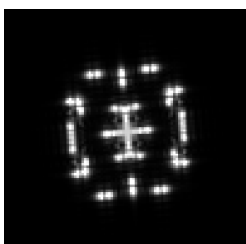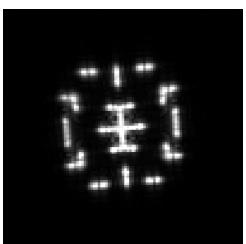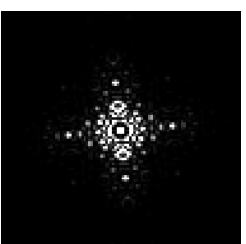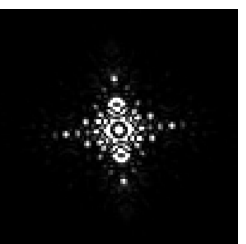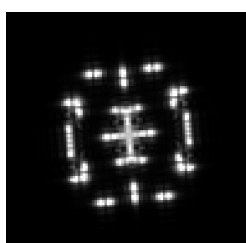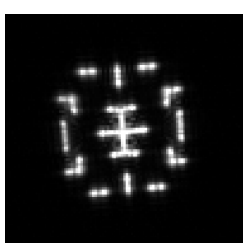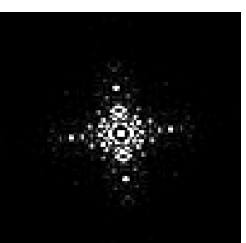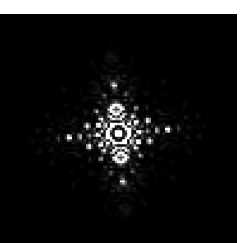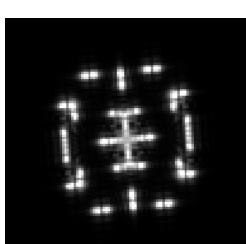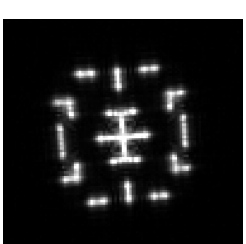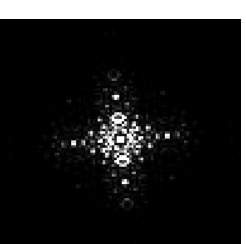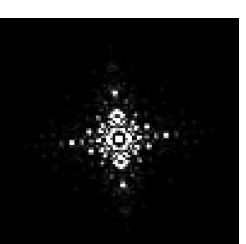
Measured and Modeled Levin Aperture I_{psf} and I_{psd}
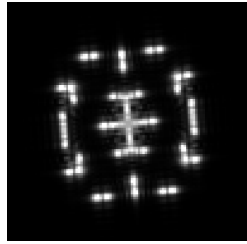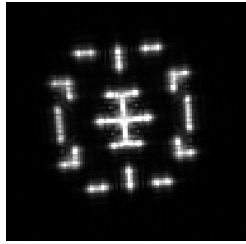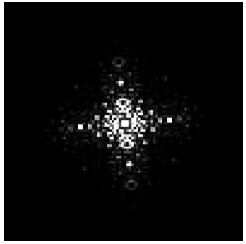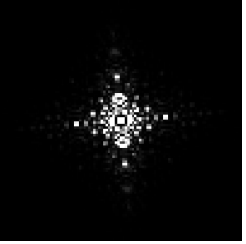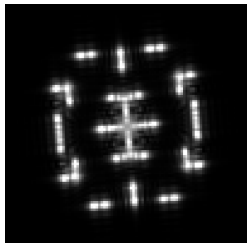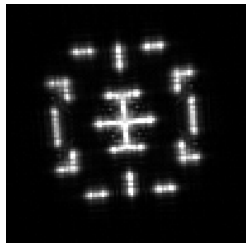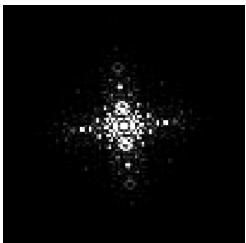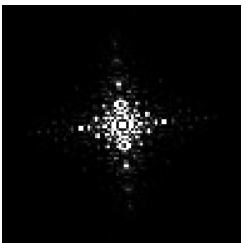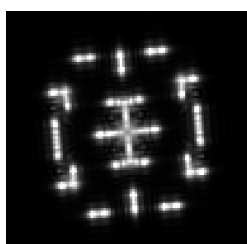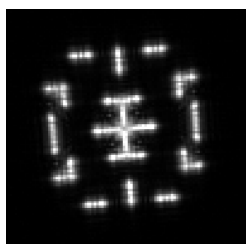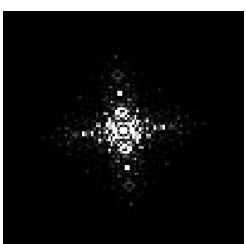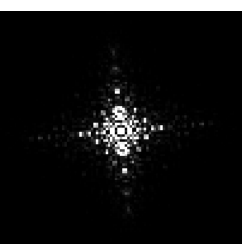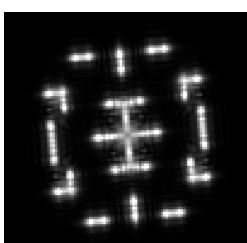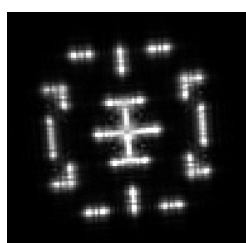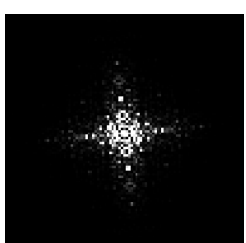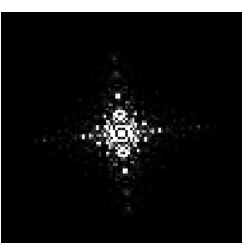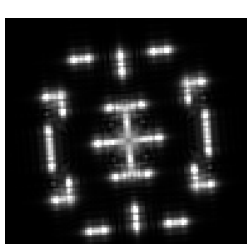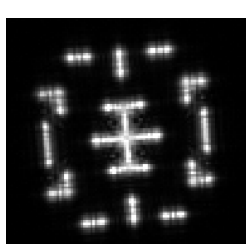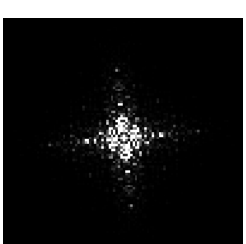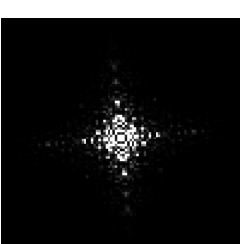
Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
1.695				
1.768				
1.846				
1.930				
2.022				

Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
2.123				
2.235				
2.359				
2.497				
2.654				
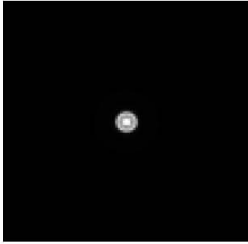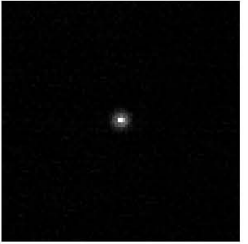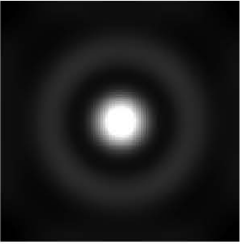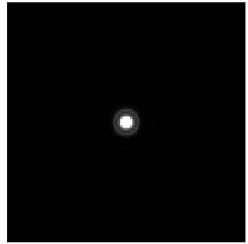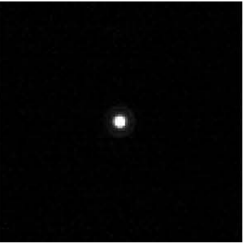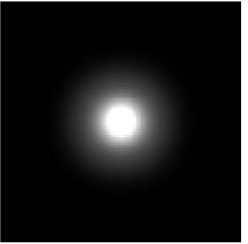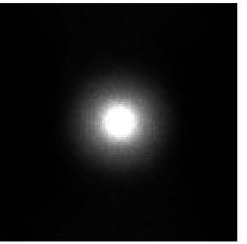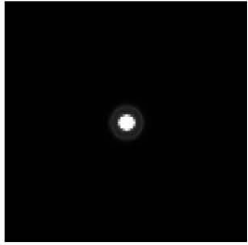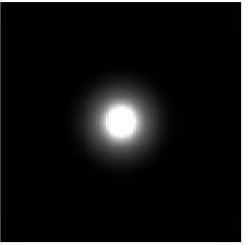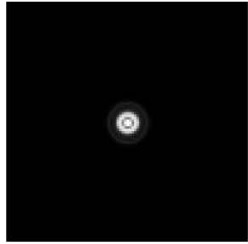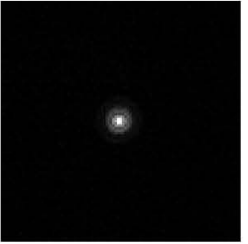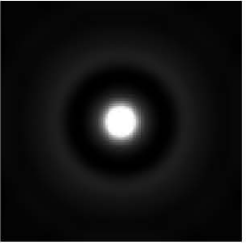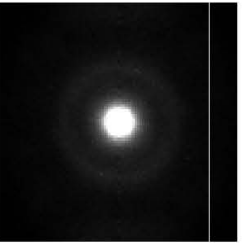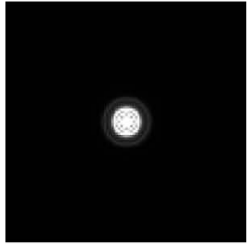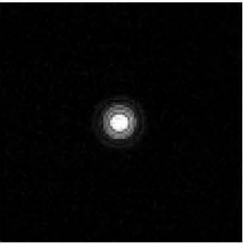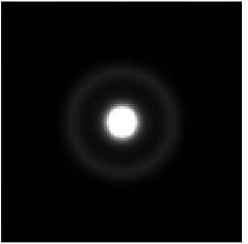
Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
2.830				
3.033				
3.265				
3.538				
3.859				

Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
4.244				
4..717				
5.307				
6.064				
7.074				
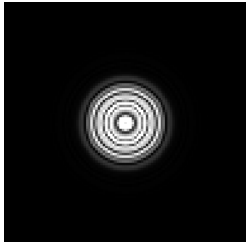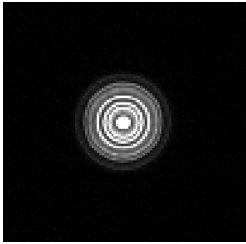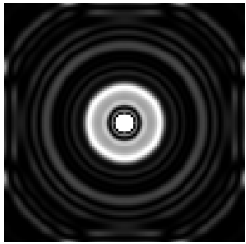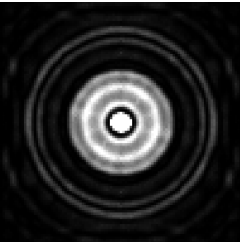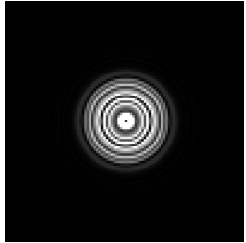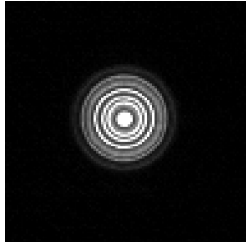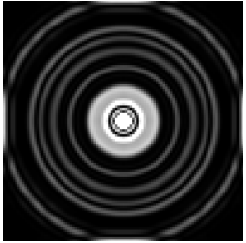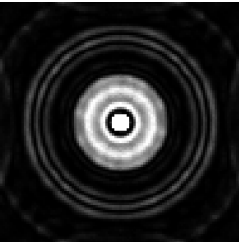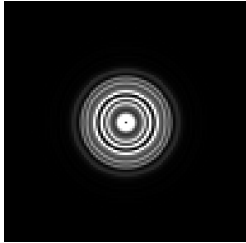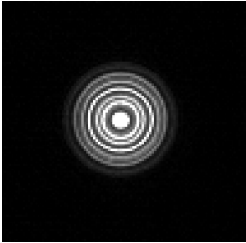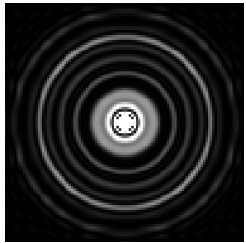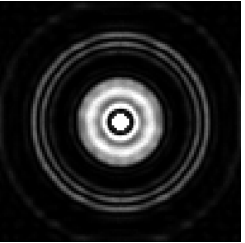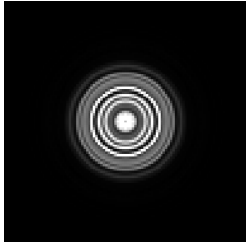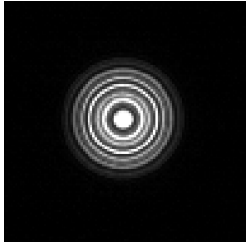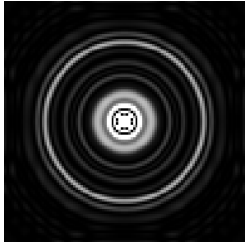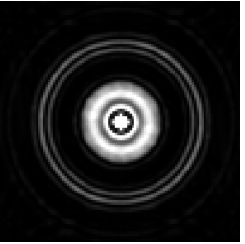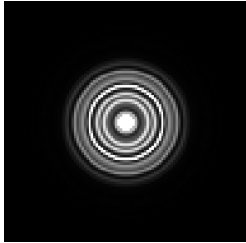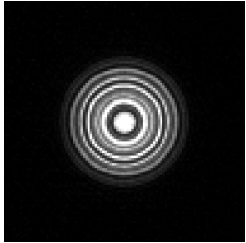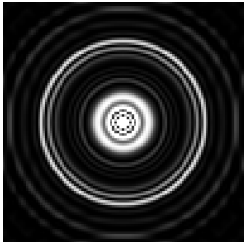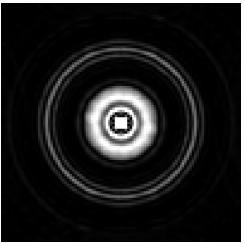
Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
8.489				
10.612				
14.149				
21.224				
42.449				

138
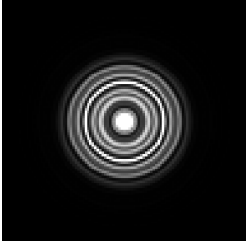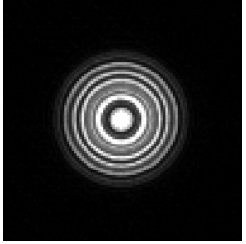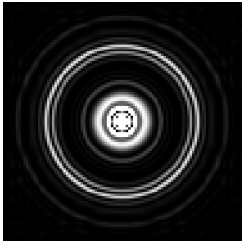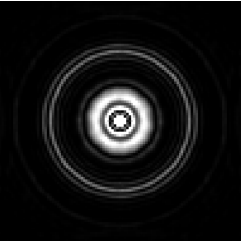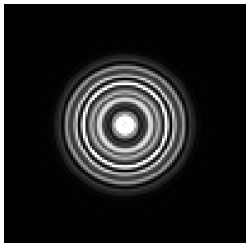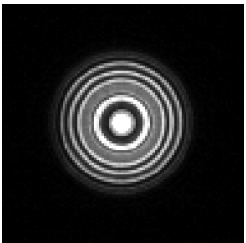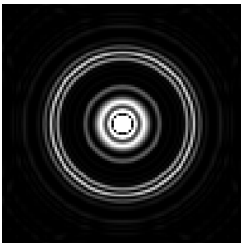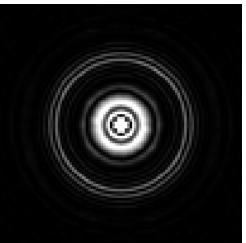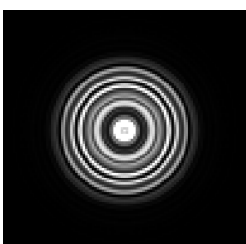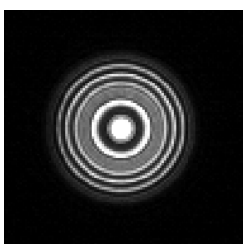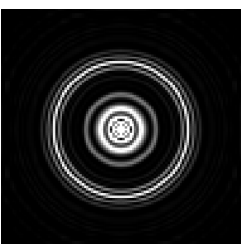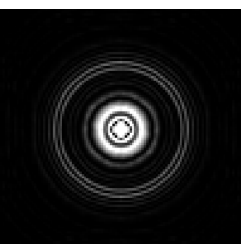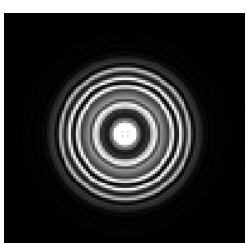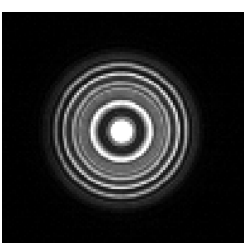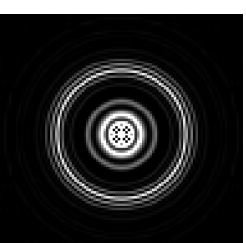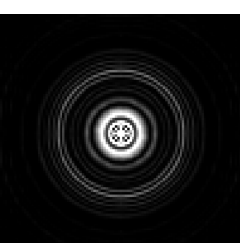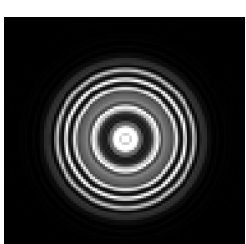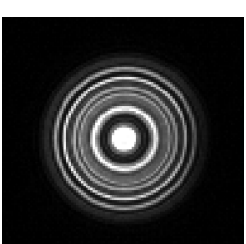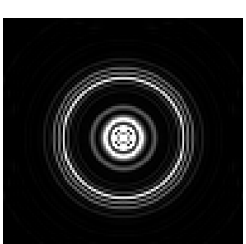
Appendix B
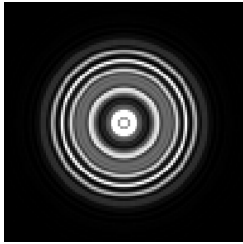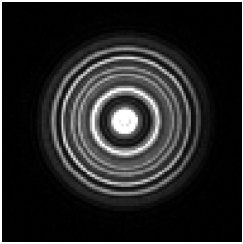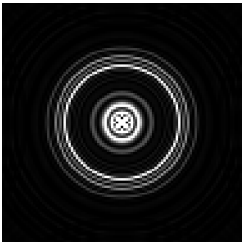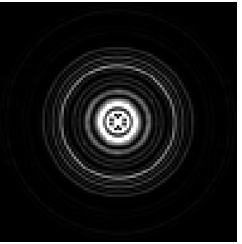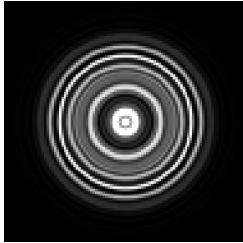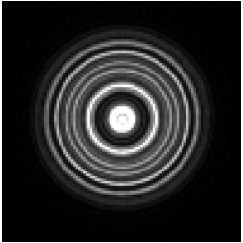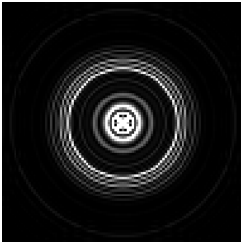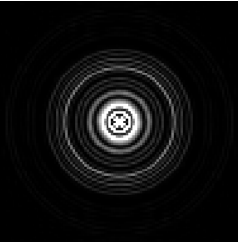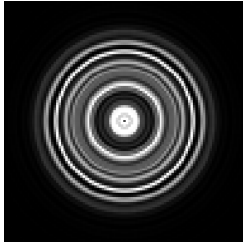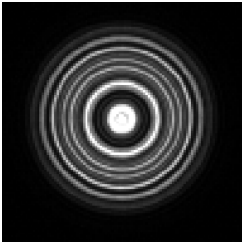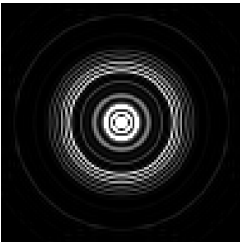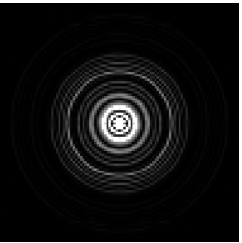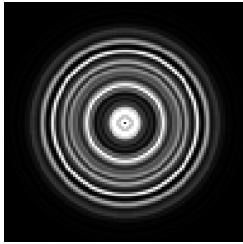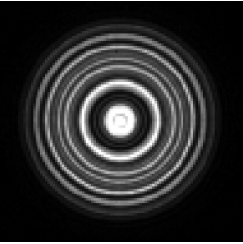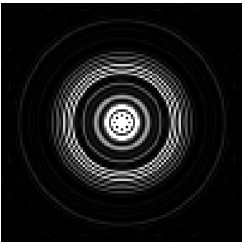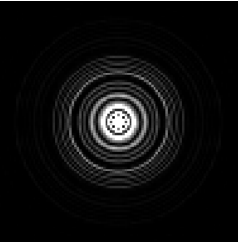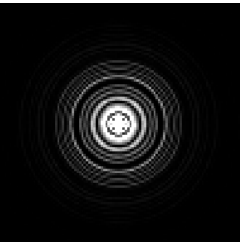
Measured and Modeled Zone Plate I_{psf} and I_{psd}

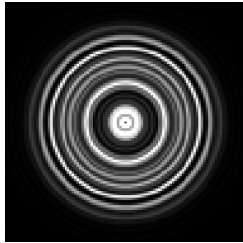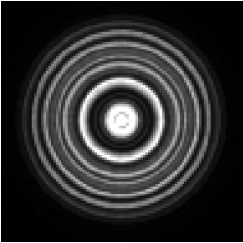| Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
1.700				
1.770				
1.847				
1.931				
2.023				

Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
2.125				
2.236				
2.361				
2.500				
2.656				

Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
2.833				
3.035				
3.2690				
3.541				
3.863				

Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
4.250				
4.721				
5.313				
6.071				
7.084				

Range	Modeled I_{psf}	Measured I_{psf}	Modeled I_{psd}	Measured I_{psd}
8.499				
10.625				
14.166				
21.249				
42.498				

Vita

       Major Jamie R. Morrison graduated from Chamberlain High School in Tampa, FL in 1989. He received the BS degree in electrical engineering from the University of Wisconsin (UW), Madison in 2000 and the MS degree in computer engineering from the Air Force Institute of Technology (AFIT) in 2005. Prior to commissioning as an officer in the United States Air Force, he served as an enlisted electronics and computer switching systems specialist. His first assignment after commissioning in 2000 was as a munitions test engineer at Eglin Air Force Base, FL. He has also served as an embedded information systems research and development engineer with Information Directorate of the Air Force Research Laboratory. His main research interests are computer vision, high performance computing, embedded systems and vision aiding of inertial navigation systems.

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
24-03-2011	Doctoral Dissertation	May 2007 – Mar 2011

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
VISION AIDED INERTIAL NAVIGATION SYSTEM AUGMENTED WITH A CODED APERTURE	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Jamie R. Morrison, Major, USAF	10-117
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765	AFIT/DCE/ENG/10-14

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
Dr. Stewart L. DeVilbiss (937-255-6127 ext. 4274) Air Force Research Laboratory 2241 Avionics Cir Wright-Patterson AFB, 45433-7301	AFRL/RYMN
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Navigation through an indoor environment is a formidable challenge for an autonomous micro air vehicle. One solution is a vision aided inertial navigation system using depth-from-defocus to determine heading and depth to features in the scene. Depth-from-defocus uses a focal blur pattern to estimate depth. As depth increases, the observable change in the focal blur is generally reduced. Consequently, as the depth of a feature to be measured increases, the measurement performance decreases. The Fresnel zone plate, used as an aperture, introduces multiple focal planes. Interference between the multiple focal planes produce changes in the aperture that extend the depth at which changes in the focal blur are observable. This improved depth measurement performance results in improved performance of the vision aided navigation system as well. This research provides an in-depth study of the Fresnel zone plate used as a coded aperture and the performance improvement obtained by augmenting a single camera vision aided inertial navigation system.

15. SUBJECT TERMS

Coded Aperture, INS, Image Aided Navigation, Fresnel Zone Plate, Depth-From-Defocus

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr John F. Raquet (ENG)
U	U	U	UU	165	19b. TELEPHONE NUMBER *(Include area code)* (937) 785-3636, ext 4580 (John.Raquet@afit.edu)