A Prototype Imaging System with a Biologically Inspired Imager for an Extended Range Walking Stick

by

Albert James M. Licup

Advisors:

Paolo Del Giudice

Complex Systems Unit, Department of Technologies and Health

Istituto Superiore di Sanità

Silvio Franz

Condensed Matter Group

the Abdus Salam International Centre for Theoretical Physics

A Thesis presented for the Master's degree

Modeling and Simulation of Complex Realities

The Abdus Salam International Centre for Theoretical Physics
Trieste, Italy

September 2004

To my family

A Prototype Imaging System with a Biologically Inspired Imager for an Extended Range Walking Stick

Albert James M. Licup

Modeling and Simulation of Complex Realities
September 2004

Abstract

This thesis describes the first attempt at simulating a prototype imaging system involving a biologically inspired imager for the purpose of estimating depth information in ambient scenes. This is part of the series of tests intended for the neuromorphic device developed at the ISS Complex Systems Unit under the HAPTIC project, a feasibility study on the integration of neuromorphic imagers with tactile transducers for the real-time transduction of dynamic visual stimuli. It has long been recognized that the focus information contained in an image is an important depth cue for both human and machine vision. This work implements distance estimation by optimizing the amount of image focus on a blurred image in the space of the imaging system parameters where the blurring is due to the imaging system's point spread function. The defocused image is first restored via blind deconvolution and then the amount of focus from both the original and restored images are calculated using a suitable focus measure. Using a simple golden section line search, the object distance is calculated from the optimal parameters that minimize the square error between the amount of focus from the original image and the reblurred restored image obtained at each optimization step. Results from simulated defocused images are acceptably good in the sense that the algorithm was able to provide good best distance estimates within a narrow range of object distances. However the uncertainties associated to those estimates are quite large, mainly due to blind deconvolution not being able to provide unique ideal image restorations.

Declaration

The work presented in this thesis is based on research carried out at the Complex Systems Unit, the Department of Technologies and Health, Istituto Superiore di Sanità in Rome, Italy. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all the author's own work unless referenced to the contrary in the text.

Copyright © 2004 by Albert James M. Licup.

"The copyright of this thesis rests with the author. No quotations from it should be published without the author's prior written consent and information derived from it should be acknowledged".

Acknowledgments

I give thanks and praise to the Almighty Father for without Him all these things would not have been possible.

I would like to extend my sincerest gratitude to my supervisors, Prof. Silvio Franz and Prof. Paolo Del Giudice for providing me the opportunity to work on this interesting and practical topic. Their valuable support in terms of work and future directions is very much highly appreciated.

I would also like to acknowledge Dr. Vittorio Dante for the space and time he has kindly afforded me at the physics lab, and for helpful insights regarding the neuromorphic retina and the PCI-AER board. Many thanks to Dr. Maurizio Mattia for enlightening discussions with Paolo. Thanks to Massimiliano Giulioni for explaining more on the details of the retina and also among other things like running which has been a much needed break from the monotony during the last few months of staying in Rome.

Thanks to all my classmates and friends from ICTP. It has been a pleasure having met such amiable friends from around the world.

Lastly, I acknowledge the generous support of the Abdus Salam ICTP which has given me this very great opportunity to participate in the Joint ICTP/SISSA Masters Program in Modeling and Simulation of Complex Realities. It has been such a fulfilling and wonderful experience having worked in an environment as ICTP. It has opened new doors for me and for my future.

Contents

List of Figures

Chapter 1

Introduction

For almost ten years now, there has been an ongoing research on development and testing issues for a neuromorphic system at the Complex Systems Unit of the Istituto Superiore di Sanità (ISS) together with the physics departments of Rome universities "La Sapienza" and "Tor Vergata". The term "neuromorphic" is used to describe those electronic devices that try to mimic the operation of some neural subsystem [1]. In a nutshell, the infrastructure of the system developed at the laboratory consists of: (i) an analog very large scale integrated (VLSI) neural chip consisting of a network of spiking neurons and plastic synapses, that autonomously develops spike representations of the input stimuli and extracts the relevant part to be forwarded to the observer; and (ii) a PCI-AER board that interfaces the standard peripheral component interconnect (PCI) bus of a PC with an asynchronous *address event representation* (AER) bus that handles communication among the neuromorphic chips. Currently the said neuromorphic system is subjected to a number of tests via systems relying on the PCI-AER board for real-time event-driven communication. One of the offshoots of those tests is the HAPTIC project, which is a feasibility study on the transduction of real-time dynamic visual signals into tactile information. The first phase of the project was to come up with two prototypes for *la mano che vede* (literally "the hand that sees") in order to design a mouse-type device that would serve as a supplement to the Braille system. It involves a small neuromorphic imager, commonly known in literature as the *neuromorphic retina*, that gathers dynamic visual stimuli by moving the mouse over printed objects and

outputs a real-time dynamical signal representation. It does such tasks in such a way that its sensing elements or *pixels* activate (i.e., light up) in the presence of a high luminosity gradient and quickly adapt (i.e., get dimmer) if the luminosity pattern does not change. Spike represention of the imager output are then passed to the PCI-AER interface board after which the output signals are sent to a tactile transducer [2]. This system, albeit primitive, led to interesting results from experiments where real tactile representations of simple input shape images were actually observed.

The second phase of the HAPTIC project is quite a big step compared to the first phase. It is a feasibility study for *il bastone esteso* (literally "the extended walking stick"). The idea is to implement a similar tactile transducer of visual signals as an electronic travel aid (ETA) device, i.e., a system that aids in the independent locomotion of the visually-impaired. The system will have to be tested with highly complex ambient scenes composed of cluttered objects at varying distances. However, attempts to resolve images of real scenes into tactile form appear to be unattainable and also useless. Instead, a sensible alternative is to provide real-time information on the distances to small image patches from the scene. In effect, the system acts as a range finder, thus consequently *extending the range of the walking stick*. The main aim of this work is to simulate and test the possibility of having such a system and propose a suitable distance estimation algorithm. The idea is that the system does the following tasks: (i) exploration of the scene and the instantaneous acquisition of images at regular time intervals (in general the images would be out of focus); (ii) partitioning each image into small patches or subimages, the size of the patch is such that it contains obstacles and other conspicuous objects; (iii) estimate the distance to such patches by a numerical optimization of the amount of focus in that patch. The ultimate goal would be a kind of a coarse depth map representation of the scene.

Distance or depth estimation from an image (or usually a set of images) is a well-known machine vision problem. The information obtained by imaging systems may be categorized into two types: *photometric* or image irradiance information and

geometric or depth information. There are two approaches that address the problem of recovering either of the two. Those that project energy onto the object and detect the reflected energy (as in the case of sonar and laser ranging) belong to the *active* approach while *passive* approaches are those that solely use ambient illumination [3]. Perhaps the most popular among passive approaches is *stereopsis*, or otherwise known as binocular, stereoscopic or stereo vision. It takes advantage of the slight disparities between an image pair obtained simulataneously from two different perspectives. Depth information is then recovered from a so-called *disparity map*. Although the technique is very accurate and quite attractive in the sense that it is inspired by biological visual systems, stereopsis inherently suffers from the problem of determining which among the points on the image pair correspond to the same point on the scene when the image disparity is large. This is known as the *correspondence problem*. Stereopsis also suffers from the problem of *occlusion*. It might occur that a point visible on one image is hidden from view on the other image. Almost as quite as popular as stereopsis is the *depth from defocus* (DFD) approach. It is also a passive approach that measures a *focal gradient* or an error in focus as a depth cue. The focal gradient is computed from (at least) two images of the scene at one perspective but obtained with different parameter settings (for instance, using small and large apertures). In this method [4], two images of an arbitrary scene are used, one formed by a pinhole camera while the other is formed by a wide-aperture camera. The images are thus defocused by different amounts. Depth information is then recovered by calculating the differences in focus for each corresponding image points. Another passive approach is *depth from focus* (DFF). Here, one directly relates the amount of image focus (or equivalently the image blur) with depth ([5], [6]). Such an approach may be carried out by *active focusing*, which involves the acquisition of a sequence of several images in better and better focus by continuously adjusting the system parameters [7]. This is used mostly in camera systems that have autofocusing features. *Passive focusing* on the other hand performs a "numerical autofocusing" of the image by performing a search of the optimal parameters that yield the best focused image. Both DFD and DFF methods have the common advantage over stereopsis in that both do not suffer from

correspondence or occlusion problems.

This work is limited to the simulation of an imaging system that, in addition to a suitable image defocusing model, would include a model for an image detector that has the general characteristics of a typical neuromorphic imager. One does not have to restrict oneself too much in having a model of an imager that closely matches the actual neuromorphic imager's response since one has enough flexibility in choosing from a wide range of available neuromorphic imagers. Having decided on a suitable model for the imaging system, a depth estimation algorithm will be proposed. The main constraint is to use only a single image and that focusing must be done by numerical optimization of the amount of image focus. Stereopsis and DFD are definitely out of the choices since they both require more than one image. The only choice left is to use DFF with passive focusing because for active focusing, one also needs a sequence of several images.

The development of this thesis is as follows. Chapter 2 presents a model of image defocusing. It begins with a short review of the basic concepts from Fourier optics on the propagation of light in free space as well as transmission through thin lenses. The main ideas are used together with concepts from linear systems theory to arrive at a model for image formation in a diffraction-limited system under monochromatic illumination. The model is generalized to include aberrations caused by defocusing under incoherent illumination. While chapter 2 discusses image formation and image defocusing models given an arbitrary image detector, chapter 3 focuses on the image detector itself. It focuses on modeling the overall behavior of typical neuromorphic imagers. Starting with an overview of the mechanisms of vision and signal processing on vertebrate retinas, it explains the fundamentals on the operation and on the general characteristics of a neuromorphic imager. The main results are then wrapped up in the image detector model. The depth estimation algorithm is proposed in the fourth chapter. It starts with the imaging system model by combining image defocusing with the neuromorphic imager. After choosing an appropriate measure for the image focus, a cost function is defined and the distance

estimation algorithm is outlined. In chapter 5, simulation results using artificially defocused images are presented. Finally the whole work is summarized and conclusions drawn at the end as well as provisions on some recommendations that might be needed for future work.

Chapter 2

Image Defocusing Model

In order to understand the process of image formation, one has to have a firm understanding from first principles the propagation of an optical field through imaging systems. Since the two most basic components of imaging systems are lenses and free space, this chapter begins with the Huygens-Fresnel principle of light propagating in free space and its mathematical interpretation expressed by Kirchhoff's diffraction integral as a solution to the scalar wave equation. The discussion proceeds with two practical approximations that lead to comparatively simple diffraction calculations. Next, the phase transforming property of thin lenses is introduced. These results are then used to analyze the properties of a simple diffraction-limited imaging system in the case of objects illuminated with monochromatic light. Finally the system is generalized under incoherent illumination and includes the case of focusing aberrations.

2.1 Optical Field Propagation

The wave nature description of light tells us that being an electromagnetic field, its behavior is wrapped up neatly by Maxwell's equations. In a linear, isotropic, homogeneous, nonmagnetic, nondispersive dielectric medium having a refractive index n, the behavior of each scalar field component u as light propagates through this medium can be described by the scalar wave equation

$$\Delta u(r,t) - \frac{n^2}{c^2}\partial_t^2 u(r,t) = 0. \tag{2.1.1}$$

Any solution to (2.1.1) is a representation of an optical field. If the field is monochromatic having temporal frequency ν, the solution expressed in complex form is given by

$$u(r,t) = \psi(r)e^{-j2\pi\nu t}$$

where the complex amplitude function ψ (sometimes called a phasor) represents the spatial variation of the field amplitude and field phase. Substituting this into the wave equation and eliminating the time dependence in the process yields the Helmholtz equation

$$\Delta\psi(r) + k^2\psi(r) = 0 \tag{2.1.2}$$

where the so-called wave number k is defined by

$$k \equiv \frac{2\pi}{\lambda}$$

and the wavelength of the field within the medium defined as $\lambda \equiv c/n\nu$.

The mathematical description of the optical field as it propagates through a given medium is therefore provided by the solution to equation (2.1.2), which can be obtained by using tools like Green's theorem. Such techniques can be too involved, in practice however, an alternative solution is given by Huygens-Fresnel principle:

> *... that every point on a wavefront* (optical field) *serves as a source of secondary wavelets such that the wavefront* (optical field) *at any later instant is given by the envelope of these wavelets...*

Kirchhoff provided a sound mathematical interpretation of the Huygens-Fresnel principle in the form of a superposition integral [8]. Since the Helmholtz equation is linear, then the linear superposition of solutions is allowed; hence the optical field at any later point is the sum of the corresponding *spherical wavelets* emanating from a point source:

$$\psi(r) = \frac{e^{\pm jkr}}{r}$$

where r is the radial distance from the source to the point of observation and the sign in the exponential term indicating a diverging $(+)$ or converging $(-)$ spherical wave. From figure 2.1.1, if $\psi(x', y')$ describes the field on the source plane at (x', y'),

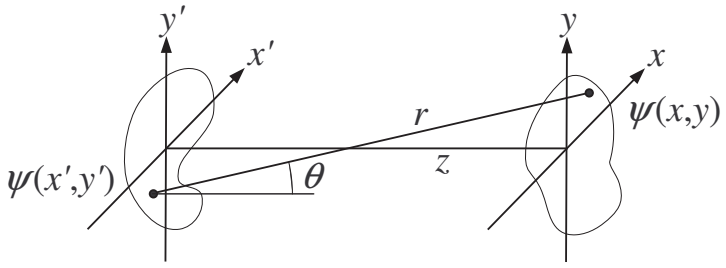

Figure 2.1.1: Diffraction geometry to illustrate Kirchhoff's diffraction integral.

then Kirchhoff's diffraction integral expresses the propagated field $\psi(x, y)$ on the observation plane at (x, y) in terms of $\psi(x', y')$ as:

$$\psi(x,y) = \iint_{\mathbb{R}^2} \psi(x',y') \frac{e^{jkr}}{j\lambda r} \cos\theta \mathrm{d}x'\mathrm{d}y' \tag{2.1.3}$$

where $\cos\theta$ is an obliquity factor [9]. Hence, an arbitrary optical field can be seen as a collection of spherical point sources and that the field at any point is simply given by the linear superposition of propagating spherical waves.

2.2 Fresnel and Fraunhofer Approximations

In most cases, the variation of the optical field is much slower compared to the wavelength and that the field is appreciable only within a small region around the optic axis. The dimension of this region is usually much smaller than the separation distance between source and observation planes. In this *paraxial approximation* case, the radial distance may be well approximated by:

$$\begin{aligned} r &= \sqrt{(x-x')^2 + (y-y')^2 + z^2} \\ &\approx z\Big\{1 + \tfrac{1}{2z^2}\big[(x-x')^2 + (y-y')^2\big]\Big\}. \end{aligned} \tag{2.2.1}$$

The obliquity factor also approaches the value unity. Substituting to equation (2.1.3), the resulting expression in rectangular coordinates for the field at (x, y) is given by the *Fresnel diffraction integral*

$$\psi(x,y) = \frac{e^{jkz}}{j\lambda z} \iint_{\mathbb{R}^2} \psi(x',y') e^{j\frac{k}{2z}\left[(x-x')^2+(y-y')^2\right]} \mathrm{d}x'\mathrm{d}y' \tag{2.2.2}$$

and the observation plane is said to be in the Fresnel or *near-field* diffraction region. This approximation is only valid when the third and higher order terms in the Binomial expansion for the radial distance r are very small such that

$$z^3 \gg \frac{k}{8}\left[(x-x')^2+(y-y')^2\right]^2_{\max}. \tag{2.2.3}$$

A further simplification is obtained in the case when the plane separation distance is even much larger than the maximum lateral dimension of the source plane, i.e.,

$$z \gg \frac{k}{2}\left(x'^2+y'^2\right)_{\max}. \tag{2.2.4}$$

In this case, the Fresnel diffraction integral reduces to

$$\psi(x,y) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}(x^2+y^2)} \iint_{\mathbb{R}^2} \psi(x',y') e^{-j\frac{k}{z}(xx'+yy')} \mathrm{d}x'\mathrm{d}y' \tag{2.2.5}$$

which is the *Fraunhofer diffraction integral*. The observation plane is then said to be in the Fraunhofer or *far-field* diffraction region. The interesting thing with this integral is its resemblance to a Fourier transform. Recasting equation (2.2.5) in the form

$$\psi(x,y) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}(x^2+y^2)} \iint_{\mathbb{R}^2} \psi(x',y') e^{-j2\pi\left(\frac{x}{\lambda z}x'+\frac{y}{\lambda z}y'\right)} \mathrm{d}x'\mathrm{d}y'$$

one sees that the integral term is simply the Fourier transform of the input field $\psi(x',y')$ evaluated at the spatial frequencies $\xi = \frac{x}{\lambda z}$ and $\eta = \frac{y}{\lambda z}$. Neglecting a constant phase term, one can express the Fraunhofer diffraction integral as

$$\psi(x,y) = e^{j\frac{k}{2z}(x^2+y^2)} \Psi(\xi,\eta)$$

where $\Psi(\xi,\eta)$ and $\psi(x',y')$ are a Fourier transform pair. One can see that up to a multiplicative phase factor, *the observed optical field in the far-field region is simply the scaled Fourier transform of the field itself at the source plane.*

2.3 Phase Transforms by Thin Lenses

Besides free space, the most important optical imaging component are lenses, what with their ability to form images. It is known from geometric optics that light is refracted as it traverses through media of different refractive indices as in the case of

free space and lens combinations. When there is a negligible coordinate translation (i.e., refraction) of the optical field as it propagates through the lens, then such lens is said to be a *thin lens* and it simply introduces *phase delays* to the incident field by an amount proportional to the lens thickness. Under the paraxial approximation, the field $\psi'(x_\ell, y_\ell)$ immediately exiting the lens with *focal length* F and refractive index n_ℓ is related to the incident field $\psi(x_\ell, y_\ell)$ by a multiplicative quadratic phase factor:

$$\psi'(x_\ell, y_\ell) = \psi(x_\ell, y_\ell) e^{jkn_\ell \delta_m} e^{-j\frac{k}{2F}(x_\ell^2 + y_\ell^2)} \tag{2.3.1}$$

where δ_m is the maximum lens thickness. Note that equation (2.3.1) holds for any lens shape if one provides the correct sign of F, i.e. $F > 0$ for convex and $F < 0$ for concave. This phase transformation can introduce one of two effects: it *collimates* a spherical wave diverging from a point source at a distance F from the lens into a plane wave, or it focuses a collimated beam to a focused spot at a distance F from the lens.

Real lenses have finite size. To take into account the finite extent of the lens aperture, one simply associates a *pupil function* $p(x, y)$ which could be any function of the form

$$p(x, y) = \begin{cases} 1 & \text{inside the lens aperture} \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the field immediately exiting the finite aperture lens becomes

$$\psi'(x_\ell, y_\ell) = \psi(x_\ell, y_\ell) p(x_\ell, y_\ell) e^{jkn_\ell \delta_m} e^{-j\frac{k}{2F}(x_\ell^2 + y_\ell^2)}. \tag{2.3.2}$$

2.4 Image Formation with Monochromatic Light

Having determined how the properties of an optical field changes as it propagates through free space and through a finite-aperture thin lens, it is now possible to construct a simple imaging system. From figure 2.4.1, suppose that a planar object illuminated by monochromatic light is placed at a distance z_o in front of a positive lens having a focal length F. The complex optical field emanating from the object is represented by $\psi_o(x', y')$. At a distance z_i behind the lens, there appears a field

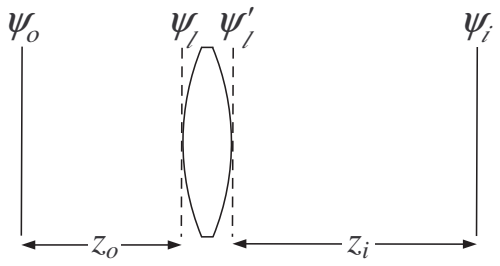

Figure 2.4.1: Image formation geometry.

$\psi_i(x, y)$. As already seen that wave propagation is a linear phenomenon, in any case the field $\psi_i(x, y)$ can be expressed by a superposition integral

$$\psi_i(x, y) = \iint_{\mathbb{R}^2} \psi_\delta(x, y; x', y')\psi_o(x', y')\mathrm{d}x'\mathrm{d}y'$$

where the function $\psi_\delta(x, y; x', y')$ expresses the field at (x, y) due to a unit-amplitude point source object at (x', y'). This function is known in linear systems theory as the *impulse response* which completely characterizes the system. In optics applications this is termed as the *point spread function* (PSF) [10].

In order to determine the PSF when light emanating from the object is monochromatic, let the point source object be a unit-amplitude impulse represented by a Dirac delta δ at (x', y'). The field incident on the lens is a spherical wave diverging from that point source. Its paraxial approximation is obtained by direct application of the Fresnel diffraction integral:

$$\psi_\ell(x_\ell, y_\ell) = \frac{e^{jkz_o}}{j\lambda z_o} e^{j\frac{k}{2z_o}\left[(x_\ell - x')^2 + (y_\ell - y')^2\right]}.$$

The field after passage through the lens undergoes a phase delay and becomes

$$\psi'_\ell(x_\ell, y_\ell) = \frac{e^{jkz_o}}{j\lambda z_o} e^{j\frac{k}{2z_o}\left[(x_\ell - x')^2 + (y_\ell - y')^2\right]} p(x_\ell, y_\ell) e^{jkn_\ell\delta_m} e^{-j\frac{k}{2F}(x_\ell^2 + y_\ell^2)}.$$

Provided that the thin lens equation

$$\frac{1}{F} = \frac{1}{z_o} + \frac{1}{z_i} \tag{2.4.1}$$

is satisfied (which is a requirement in order for imaging to occur), a second application of the Fresnel diffraction integral to account for the propagation through the distance z_i finally yields the result

$$\psi_\delta(x,y;x',y') = \frac{A}{\lambda z_i}\iint_{\mathbb{R}^2} p(x_\ell,y_\ell)e^{-j\frac{k}{z_i}[(x-Mx')x_\ell+(y-My')y_\ell]}\mathrm{d}x_\ell\mathrm{d}y_\ell$$

where the image *magnification* is defined as $M \equiv -\frac{z_i}{z_o}$, the minus sign accounting for possible image inversion. The constant A accounts for all the constant phase factors which may be safely dropped by an appropriate normalization. To remove such magnification/inversion effects, one can redefine the coordinates in the object plane as

$$\begin{aligned} x' &= Mx' \\ y' &= My' \end{aligned}$$

such that the PSF becomes

$$\psi_\delta(x-x',y-y') = \frac{A}{\lambda z_i}\iint_{\mathbb{R}^2} p(x_\ell,y_\ell)e^{-j\frac{k}{z_i}[(x-x')x_\ell+(y-y')y_\ell]}\mathrm{d}x_\ell\mathrm{d}y_\ell.$$

At this point it is useful to define the ideal image as *predicted by geometric optics*. If the imaging system is perfect, then *the image is just an inverted and magnified replica of the object*, i.e.,

$$\psi_g(x',y') \equiv \frac{1}{|M|}\psi_o\left(\frac{x'}{M},\frac{y'}{M}\right).$$

All these results wrap up neatly in the following convolution relation:

$$\psi_i(x,y) = \iint_{\mathbb{R}^2} \psi_\delta(x-x',y-y')\psi_g(x',y')\mathrm{d}x'\mathrm{d}y' \tag{2.4.2}$$

where

$$\psi_\delta(x,y) = \frac{A}{\lambda z_i}\iint_{\mathbb{R}^2} p(x',y')e^{-j\frac{k}{z_i}(xx'+yy')}\mathrm{d}x'\mathrm{d}y'. \tag{2.4.3}$$

The convolution (2.4.2) which may be rewritten in shorthand as

$$\psi_i = \psi_\delta \otimes \psi_g \tag{2.4.4}$$

indicates that the system is linear and shift-invariant (LSI). Also note that the PSF ψ_δ according to equation (2.4.3) is a scaled Fraunhofer diffraction pattern of the

pupil function p. Hence under monochromatic illumination, the imaging system is *linear and shift-invariant in complex optical field amplitude*. One can obtain the field amplitude at the image plane by *convolving the ideal geometric optics image with a point spread function which is the scaled Fourier transform of the pupil function*.

2.5 Generalized Imaging Systems

When the imaging system has a combination of several positive and/or negative lenses at various displacements relative to each other but still producing a real image in space, one may treat the imaging elements as lumped altogether in a "black box" (figure 2.5.1) such that the system properties are completely described by specifying only the "terminal" properties, i.e., the entrance and exit pupils. Geometric optics is adequate in describing the passage of light from the entrance to the exit pupil so diffraction effects will only play a role *either* during the passage of light from the object plane to the entrance pupil or from the exit pupil to the image plane. The imaging system is said to be *diffraction-limited* if a diverging spherical wave incident on the entrance pupil comes out as a converging spherical wave at the exit pupil. Any deviation from the ideal spherical shape indicates an *aberration* of the imaging system. The diffraction-limited imaging system under monochromatic illumination was already shown in the previous section to be a LSI system. In cases where the system contains a combination of lenses, the same situation still applies and the PSF is now the scaled Fourier transform of either the entrance or the exit pupil function. These two points of view are entirely equivalent since one pupil is just an image of the other [9]. In the following discussions we will take the latter view to be consistent with the distance z_i being the distance from the exit pupil to the image plane.

When the object illumination is spatially coherent, the resulting amplitude point spread functions on the image plane vary with time in a perfectly correlated manner. Hence *linear superposition may be done on a complex amplitude basis*. The results from our monochromatic analysis in the preceding section still applies di-

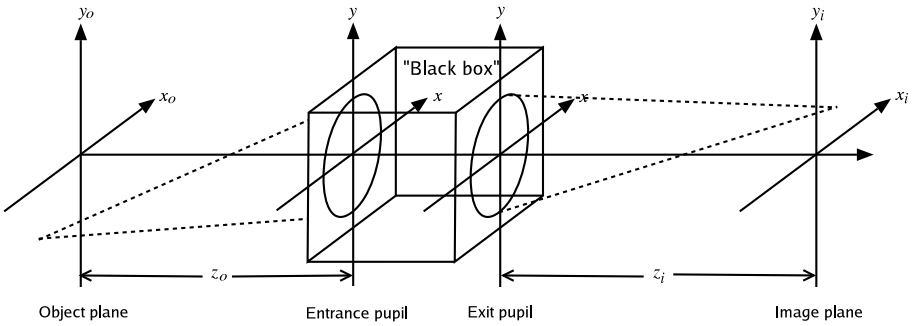

Figure 2.5.1: Generalized imaging system.

rectly to such systems. However in the general case when the light reflected from the object is incoherent such as those coming from diffuse or extended light sources like gas discharge lamps or sunlight, the temporal variations of the point spread functions are uncorrelated. For this reason they must be summed up on a *power or intensity basis*. Hence the imaging system under incoherent illumination is *linear in intensity*. If one has the object and image intensities f and g as well as the *intensity point spread function* h, then the imaging system is modeled as [11]

$$g = h \otimes f. \tag{2.5.1}$$

Up to this point we have been talking about diffraction-limited imaging systems. We will now characterize the imaging system in the general case when there exists a focusing aberration. Consider an imaging system consisting of a positive lens with focal length F, aperture (i.e., the exit pupil) diameter D and an image detector placed at a distance z_d behind the aperture plane as shown in figure 2.5.2. For now assume that the detector is simply a photosensitive plate which responds to light intensity in the same way as an ordinary photographic film or an array of photodetectors would. From geometric optics, a point light source on the *object plane* at a distance z_o in front of the *aperture plane* will have its corresponding sharp focused image on the *image plane* at a distance z_i behind the aperture plane according to the thin lens equation (2.4.1). We will call the plane that cuts perpendicularly across the *optic axis* at F as the *focal plane*. For notational convenience, the *camera parameters* will be defined collectively as the vector

$$\boldsymbol{e} \equiv (z_d, F, D). \tag{2.5.2}$$

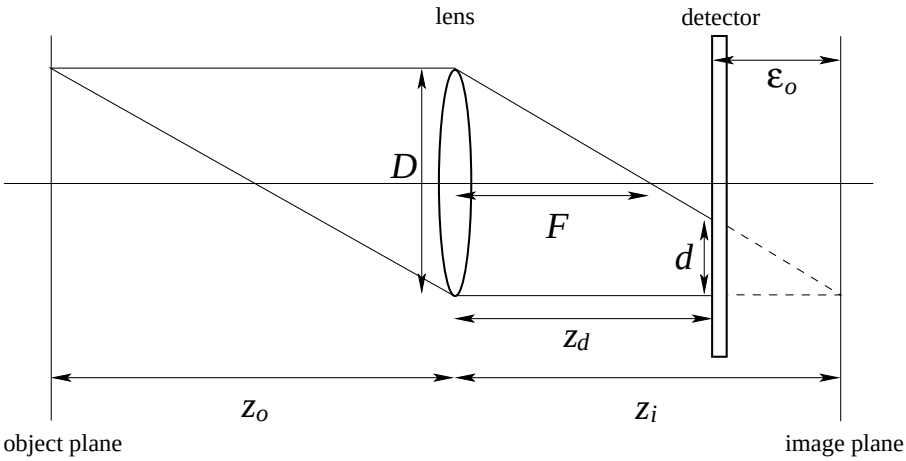

Figure 2.5.2: Image formation in a positive lens.

Obviously from the one-to-one correspondence between z_o and z_i dictated by equation (2.4.1), one can obtain a sharp image of the scene at the *detector plane* only when $z_d = z_i$. Otherwise a defocused image results and that the amount of defocus is directly related to the deviation $\varepsilon_o = z_i - z_d$. The subsequent discussions on image defocusing models will consider only a circular aperture so that the imaging system is taken to be circularly symmetric around the optic axis. The origin is fixed at the center of the aperture and all distance measurements are made with respect to this reference point. For convenience we also focus the imaging system at infinity, i.e., the detector plane is held fixed at $z_d = F$.

The *geometric optics model* for image defocus lies on the basic idea that the defocused image of a point source object has the same shape as the aperture but only scaled by a factor. Thus for a circular aperture, the image of a point source object will also be a circle having uniform intensity inside and zero elsewhere [12]. This so-called *blur circle* will have a diameter d as shown in the geometry of figure 2.5.2. The *blur circle radius* can be easily derived using similar triangles and is given by $r = D\varepsilon_o/2z_i$ or

$$r(\boldsymbol{e}, z_o) = \frac{Dz_d}{2}\left(\frac{1}{z_d} - \frac{1}{F} + \frac{1}{z_o}\right). \tag{2.5.3}$$

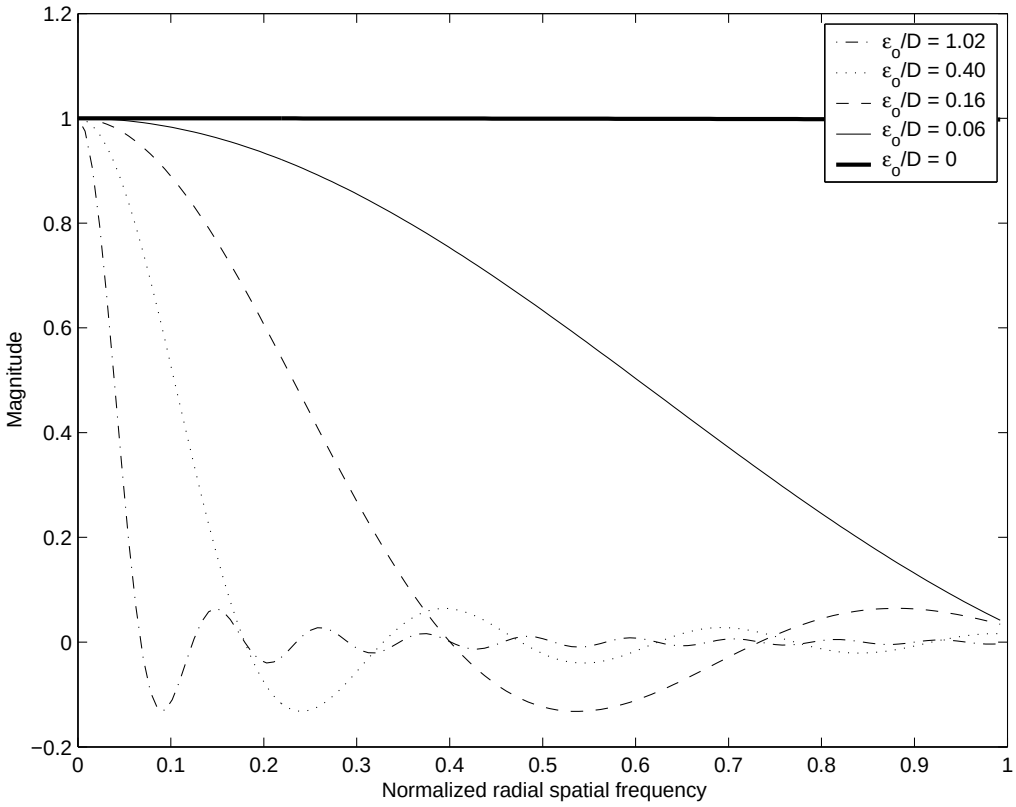

Figure 2.5.3: Geometric optics OTF.

For a lossless imaging system (i.e., no light is absorbed), the *geometric optics PSF* is defined by a circular pillbox function

$$h_c(x, y; \boldsymbol{e}, z_o) = \begin{cases} \frac{1}{\pi r^2} & \text{if } x^2 + y^2 \leq r^2 \\ 0 & \text{otherwise} \end{cases} . \tag{2.5.4}$$

where its dependence on $\boldsymbol{e}$ and z_o comes directly from r. At this point we introduce the concept of the imaging system's *optical transfer function* (OTF). It is simply the Fourier transform of the PSF which allows ease in characterizing the system according to its *spatial frequency response*. Thus we have the *geometric optics OTF*

$$\mathcal{H}_c(\rho; \boldsymbol{e}, z_o) = 2\frac{J_1(r\rho)}{r\rho} \tag{2.5.5}$$

where ρ is the radial spatial frequency and J_1 is a first order Bessel function of the first kind. Figure 2.5.3 shows the cross-section of $\mathcal{H}_c$ at varying amounts of focus.

As an alternative to the geometric optics model, a *Gaussian PSF model* is suggested

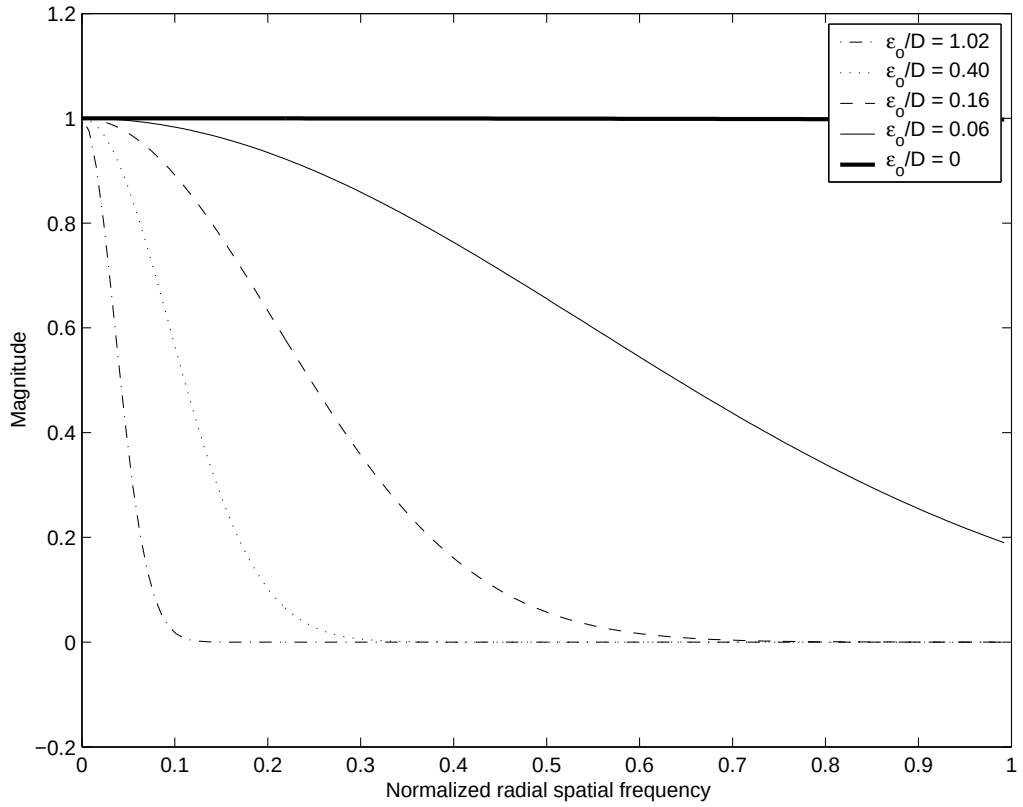

Figure 2.5.4: Gaussian OTF.

([4], [12]) in order to take into account diffraction, polychromatic illumination, and other lens aberration effects among others. Hence the image of a point source object will be roughly a circular region with intensity falling off gradually at the edges instead of a crisp circular patch of uniform intensity as suggested by the circular pillbox of the geometric optics PSF. The *Gaussian PSF* is given by

$$h_g(x, y; \boldsymbol{e}, z_o) = \frac{1}{\pi r^2} e^{-\frac{x^2+y^2}{r^2}} \tag{2.5.6}$$

and the corresponding *Gaussian OTF*

$$\mathcal{H}_g(\rho; \boldsymbol{e}, z_o) = e^{-\frac{1}{4}\rho^2 r^2}. \tag{2.5.7}$$

Figure 2.5.4 shows the cross-section of the circularly symmetric Gaussian OTF. Note that both PSFs/OTFs depend explicitly on the vector of camera parameters $\boldsymbol{e}$ and the object distance z_o.

In both cases, the image defocusing system will have the form

$$g = h_{\text{defocus}}(\boldsymbol{e}, z_o) \otimes f \tag{2.5.8}$$

where h_{defocus} can be the geometric optics PSF, a Gaussian PSF, or the actual PSF of the system which may be determined from calibration. Let us first examine the image defocusing process in the spatial domain by looking at the behavior of the PSF. In both models presented, when the PSF is very small (corresponding to a very small blur circle radius) such that it is essentially a Dirac delta, the detected image g will be at its best focus since it will be exactly the same as the ideal image f. As the blur circle radius (and hence the PSF) gets larger, the amount of focus decreases due to the smearing out effect of the convolution operation. In order to see this clearly, an alternative way of looking at things is to look into the system's spatial frequency characteristics by examining its OTF. Intuition and basic photography experience tell us that the presence of high detail on an image signifies that it is in a much better focus compared to one which has less. These high details correspond to the high spatial frequency components of the image spectrum. The more high frequency content the image has, the higher detail content it will have and the sharper and better its focus will be. The blurring effect brought about by the PSF is because of an attenuation of those high frequencies by a multiplicative weighting factor applied on the image spectrum in the spatial frequency domain. This weighting factor which has the overall characteristics of a *low-pass filter* is precisely the OTF, which is the Fourier transform of the PSF (or the Hankel transform for circularly-symmetric PSFs). When the PSF is so small, the OTF is essentially an all-pass filter and all frequencies are preserved resulting in a sharp image. As the PSF gets larger, the OTF bandwidth gets narrower hence more high spatial fequencies are attenuated leading to a decrease in the amount of image focus.

We have seen that as a consequence of the linearity property of light propagation according to the Huygens-Fresnel principle, the entire process of light propagation from source plane to detector plane through an imaging system whether under coherent or incoherent illumination is modeled as a linear system. Furthermore, this property holds irrespective of the position of the contributing point sources at the

source plane. Hence the linear imaging system is also shift-invariant. Finally image defocusing can be understood simply by considering the imaging system's PSF (or OTF) as a low-pass filter that attenuates the image's high frequency components more as the defocusing aberration increases.

Chapter 3

Image Detector Model

The previous chapter dealt with a model for image formation and defocusing, and the discussion did not mention anything regarding the nature of the image detector. This chapter is dedicated to the detector itself. In this work a biologically inspired neuromorphic imager (sometimes known in literature as the neuromorphic retina) will be used. The latter term is nowadays not preferred since such devices mimic only a small portion of the much more complex real vertebrate retina. For this reason, it is more appropriate to use the term neuromorphic imager. The discussion begins with a short review of the mechanisms of vertebrate vision and especially on the signal processing aspect of the retina followed by a brief description on the basic architecture of a typical neuromorphic imager.

3.1 Signal Processing in the Vertebrate Retina

The retina is a thin, around half a millimeter, sheet of tissue that partially lines the inside of the eyeball. This small central nervous system outpost is responsible for gathering all visual information that reaches the brain's visual cortex. It consists of five discrete layers of alternating cells and processes (figure 3.1.1) that transduce light into a neural signal. The *photoreceptor cells* made up of *rods* and *cones* actually lie in a layer against the back of the eyeball contrary to what one might intuitively expect that they would be lining the retina's surface. In the *inner nuclear layer* next to the photoreceptors lie the *horizontal cells*, *bipolar cells*, and

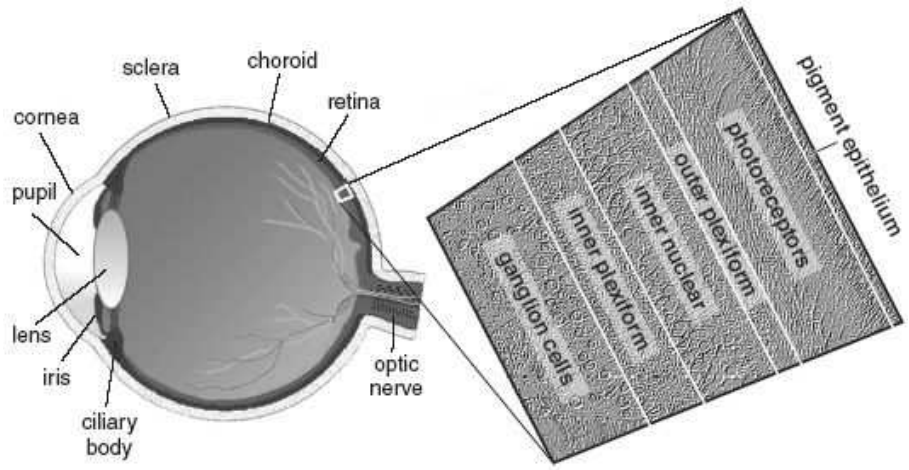

Figure 3.1.1: Cross-section of the eye showing the retina and its structure (from [13]).

amacrine cells, the numbers of each depending on species. The surface of the retina contains the *ganglion cells* where impulses travel to the brain via the optic nerve fibers. The spaces separating these three cell layers are also distinct, and are essentially synaptic links. The region containing the synapses linking the photoreceptors with bipolar and horizontal cells is the *outer plexiform layer*. The area where bipolar and amacrine cells connect to the ganglion cells is the *inner plexiform layer* (figure 3.1.2).

When light enters the retina, it enters from the ganglion cell layer first and must penetrate all cells until it reaches the photoreceptors. Both rods and cones respond to light directly over them, thus their collection areas or *receptive fields* are very narrow. Counterintuitively, for as long as light continues to shine on the rods and cones, they remain in the *hyperpolarized* state (i.e., their membrane potentials become very negative) and do not emit a neurotransmitter. Although both rods and cones respond to light with a slow hyperpolarizing response, they report quite different image properties. Rods, detecting dim light, usually respond to relatively slow changes. Cones on the other hand, dealing with bright light, can detect rapid light fluctuations. In both cases, photoreceptors begin the process of decomposing images into separate parts.

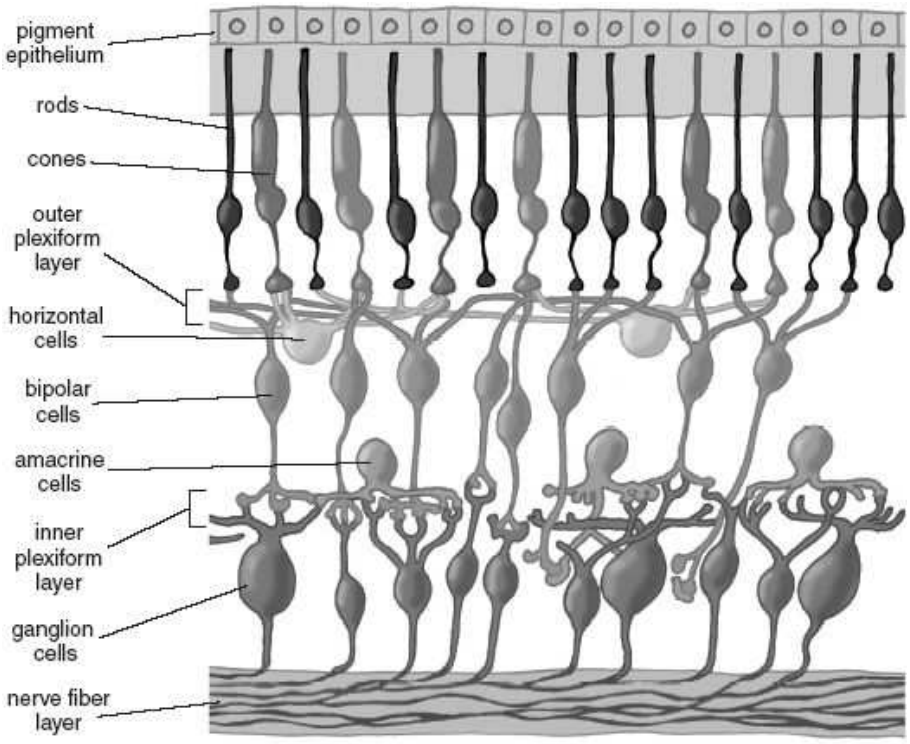

Figure 3.1.2: Arrangement of the cells in the retina (from [13]).

An image continues to be broken down into component elements in the outer plexiform layer, where axons of photoreceptors contact the bipolar and horizontal cells. A single bipolar cell receives input from a handful of photoreceptors and thus has a *medium-sized receptive field*. For the same stimulus, different bipolar cells respond in different ways. The ON bipolar cells are inhibited (depolarized) in the presence of the neurotransmitter thus activating the ON pathway for detecting bright images on a dark background. The OFF bipolar cells become hyperpolarized and this activates the OFF pathway for detecting dark images on a bright background [13]. These parallel sets of visual channels for ON and OFF image qualities are fundamental to vertebrate vision, which depends on perceiving the *contrast* between images and backgrounds. Further processing in the retina defines precise edges and fine detail. The sharpening of the image starts where horizontal cells receive input from photoreceptors, mainly cones. Each horizontal cell receives input from many cones, so it has a *large receptive field*. Moreover, their receptive fields become even broader because they also link with those of neighboring horizontal cells. Whereas a

single bipolar cell with its OFF or ON response would carry a fairly blurry response to its ganglion cell, horizontal cells add an opposing signal that is spatially constrictive, giving the bipolar cell what is known as a *center-surround* organization. This means that the center must be mainly light and the surrounding region mainly dark, or vice versa. To understand why, suppose that diffuse light hyperpolarizes a photoreceptor cell thereby exciting the bipolar cell directly underneath. However, the neighboring photoreceptor cells also excite horizontal cells which send processes laterally and inhibit the center bipolar cell. The result is that in addition to shaping the receptive field of the bipolar cells, horizontal cells modulate the photoreceptor signal under different lighting conditions, allowing signaling to become less sensitive in bright light and more sensitive in dim light. Through this, the visual system is able to adapt to light over several orders of magnitude.

Finally, the amacrine cell circuitry in the inner plexiform layer conveys additional information to the ganglion cell, possibly sharpening the boundary between center and surround even further than the horizontal cell input already does. In human retinas, two basic types of ganglion cells (ON center and OFF center) form the majority of the retina's output to the visual centers in the brain. ON center ganglion cells are activated when a spot of light falls in the center of their receptive field and are deactivated when light falls on the periphery. OFF-center ganglion cells react in the opposite way: their activity increases when the periphery of their receptive field is lit and decreases when light falls on the center of the field. Hence the receptive fields of ganglion cells are modeled as a Mexican-hat shape reflecting their integration of opposing information about centers and surrounds. This kind of processing which is akin to spatial bandpass filtering helps sharpen the boundaries of images. Amacrine cells also play an important role in the perception of very dim light since they collect messages from many rod-connected bipolar cells. They transmit information from rod photoreceptors and feed information directly to OFF center ganglion cells ganglion cells.

Having reviewed the basic mechanisms of signal processing in the retina, the impor-

tant point is that the entire visual system exists to perceive *borders* and *contours*. It sees the world as a pattern of lines even for complex objects. Colors and brightness are discriminated by comparison of *contrast* and not by any absolute scale. The system of lateral inhibition in the retina is the first step towards sharpening contours much like Laplacian spatial filtering does in image processing. It is also geared towards picking up on borders between light and dark, adaptation to a wide range of viewing conditions, and producing an output which is independent of the absolute illumination level.

3.2 The Neuromorphic Imager

All the cameras commercially available are designed for the aim in obtaining good photographs; pictures of the world that people see. What some people have started calling "neuromorphic retina" is something very much different. People who designed such devices are interested in things like getting more information behind the images, to identify the objects in the visual field, whether they are moving and how fast they are moving, etc., in other words, they are interested in getting a list of properties that can be used to interact with the environment, for instance, to walk around.

In the past, digital cameras were made using CCD technology while the retina uses CMOS technology. The CCD allows one to pack together a lot of pixels in a small area to obtain high resolution and high quality images. In CMOS technology, one cannot design a good photopixel: there are always secondary effects that degenerate image quality. It was easy then to distinguish a camera from a retina by simply identifying which technology is used. Nowadays things have changed a bit. The CMOS technology has improved and allows the use of *active pixels*. Active pixels are made up not only of photodiodes and phototransistors but also has other small circuits like an amplifier or a simple source follower. People have studied a lot on these things ([14], [15]) and they find that active pixels perform well and have better signal-to-noise ratios (SNR). Moreover, they are less expensive than CCD

pixels. With this, digital camera technology is now beginning to move from CCD to CMOS, thereby making it difficult to distinguish them from the retinas. Other digital cameras are now even capable of some features like edge and motion detection. So now there is a fine line demarcating cameras and retinas. Despite all this, there are still some fundamental differences between the two. First, the retina is an asynchronous analog system, i.e., no clock is required. Digital cameras have a clock which makes them require a lot of power because of these clock signals. Second, while cameras are still made with the "photographic" aim, retinas or neuromorphic imagers still try to mimic the vertebrate visual system.

Whereas conventional image sensors including those of digital cameras usually have linear response to brightness detected at each pixel, neuromorphic imagers typically involve a nonlinear mapping function, adaptation, and local spatiotemporal filtering [16]. These are just some of the important features observed on vertebrate retinas. We describe briefly the most basic model of the "silicon retina" as presented in [1]. The model essentially has the following three elements: (i) a photoreceptor circuit that takes the logarithm of light intensity; (ii) a resisitive network that performs spatiotemporal filtering of the photoreceptor output; and (iii) a circuitry for computing the output proportional to the difference between photoreceptor signals and horizontal cell signals.

The primary function of the photoreceptor is to transduce light into an output voltage, which for intermediate illumination levels, is proportional to the logarithm of the light intensity. The purpose of this logarithmic mapping is twofold. First it compresses the intensity range of incoming light of many orders of magnitude into a manageable scale of output voltages. Second it encodes the contrast ratio, which is proportional to the output voltage differences, thus removing the dependence on background illumination. Local spatiotemporal filtering is achieved by interconnecting the individual sensing elements with resistive networks that mediate excitatory and inhibitory interactions. The combination of these interactions results in a spatial filtering that enhances edges and contours. One can see that the output of the

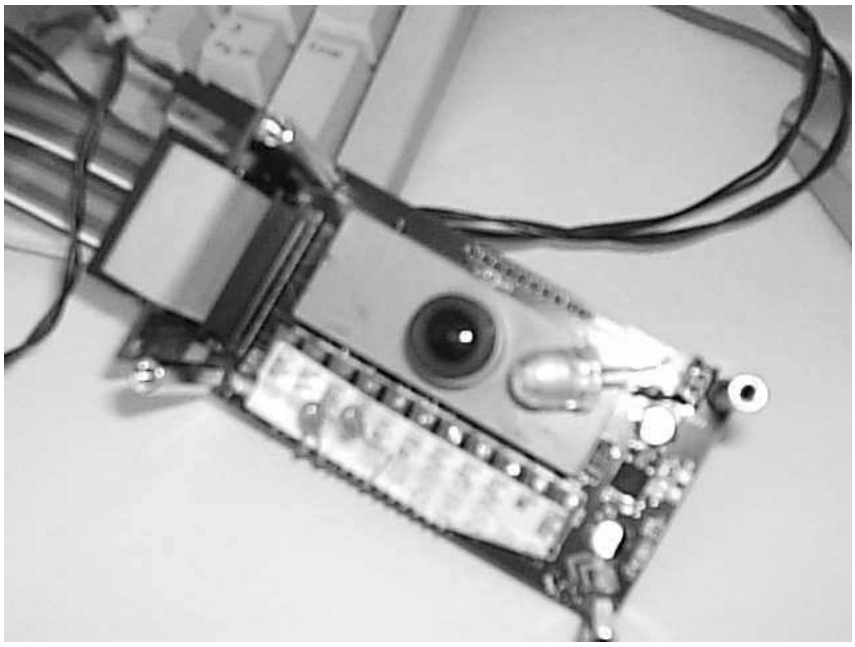

Figure 3.2.1: The neuromorphic imager.

neuromorphic retina is analogous to the center-surround response of the ganglion cell which is computed from the interaction between the photoreceptors, the horizontal cells, and the bipolar cells.

The neuromorphic retina at the laboratory (figure 3.2.1) is a slightly modified version from the basic one discussed above but still has the common features of contrast adaptation and detection of edges and contours. Figure 3.2.2 shows a schema of the structure of an optical matrix embedded in the neuromorphic imager. It is an array of photopixels, each made up of a combination of an adaptive photoreceptor and a rectifying and thresholding differentiating element. The photoreceptor encodes positive (dark-to-bright) or ON transitions and negative (bright-to-dark) or OFF transitions on separate output nodes. Recall that separated ON and OFF channels are found in the retina. To a first approximation, these transient responses encode the temporal contrast, and are insensitive to absolute irradiance. Still in the pixel is another subcircuit that is responsible for the conversion of these analog signals into spikes and for interfacing with asynchronous readout circuitry given a two-dimensional array of such pixels. Two integrate-and-fire (IF) neurons translate the

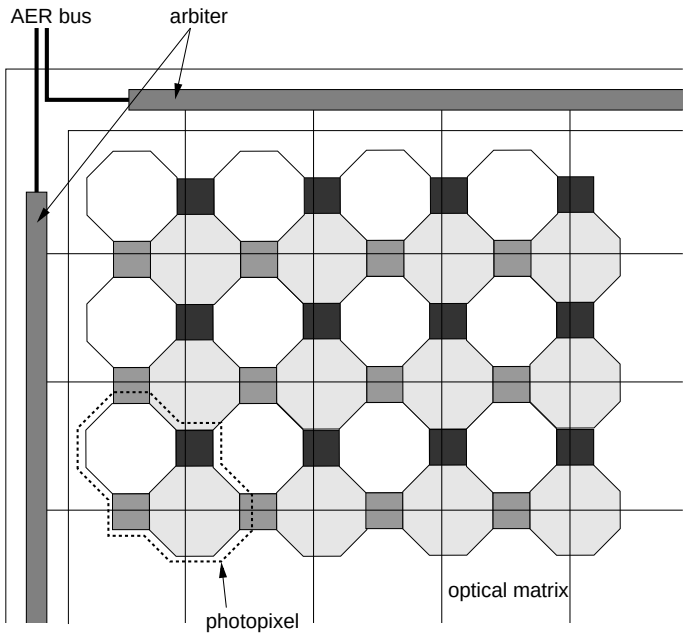

Figure 3.2.2: Schematic diagram of the structure of an optical matrix embedded on a neuromorphic imager. The white octagonal regions represent the photosensitive elements while the shaded areas are a combination of analog and digital circuitry.

analog currents coming from the adaptive photoreceptor's ON and OFF channels into spike trains, digital impulses separated by analog time intervals. This initiates a *handshaking* cycle between the pixel array and an arbiter, which is responsible for managing the selection of spiking neurons according to priority competing to access the external AER bus. In the first chapter we have briefly mentioned that the AER bus is the one that takes care of the communication among neuromorphic chips. A more detailed explanation may be found in ([18], [15], [16]). The overall effect, what the imager perceives are edges emphasized by the contrast transitions. Coding the temporal contrast also implies the need for a dynamic visual stimulus. Otherwise when the stimulus is stationary, the photoreceptor signals rapidly decay to very small values and hence the edges quickly fade away. Figure 3.2.3 shows an eye image obtained by a neuromorphic imager as compared to one by a conventional digital camera.

Figure 3.2.3: Image of an eye as seen by a digital camera (left) and by a neuromorphic imager (right). Photo courtesy of Giacomo Indiveri.

3.3 Imager Model

Different neuromorphic imagers produce different images but they all have one thing in common, that of *edge detection*. For this reason, we model the imager as an edge-detection operator. This might seem to be a naïve model but could be a good starting point nevertheless for this work. Basically, edges are curves in the image where a rapid transition in brightness or in the spatial derivatives of brightness occur [12]. There are several ways to detect edges but the majority of those methods may be grouped into two categories, gradient and Laplacian. Common practice for gradient-based approaches is to use the image gradient magnitude, which for a given image $f(x, y)$, is given by

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}. \tag{3.3.1}$$

On the other hand, the image Laplacian is expressed by

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \tag{3.3.2}$$

It can be shown easily that taking the image gradient or image Laplacian is equivalent to a high-pass filtering. Edge detection in digital image processing, whether gradient or Laplacian-based, may be done using frequency domain methods where one has to use high-pass (or in some cases bandpass) filtering of the image spectrum. However, it is usually done via *local spatial filtering* methods ([12], [19]). The general approach to spatial filtering is to let the image f within a predefined

neighborhood of a pixel at (x, y) determine the value of the output image g at the same location (x, y). The local neighborhood of pixels is defined by the size of a *mask* or *linear filter*. Basically the filter is a small (much smaller than the size of the image) 2D array of coefficients, the values of which determine the kind of filtering it performs. These coefficients are usually obtained from the approximation with finite differences. The procedure is to sum products between the filter coefficients and the values of f in the neighborhood within the filter. So for a filter h having a size $[-K, K] \times [-L, L]$, the response would be such that at any location (x, y) on f, one has

$$g(x, y) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} f(x - k, y - l)h(k, l).$$

Observe that this is a discrete convolution between f and h. Thus our imager model will have the following general form:

$$g = h_{\text{imager}} \otimes f \tag{3.3.3}$$

where h_{imager} is a filter having high-pass (or bandpass) frequency characteristics, i.e., an edge-detector filter. Earlier in the first section of this chapter, we have mentioned that the system of lateral inhibition in the vertebrate retina is the first step towards sharpening contours similar to Laplacian spatial filtering. For this reason, we choose h_{imager} in 3.3.3 as the Laplacian filter. From 3.3.2, one can implement numerically in several ways the Laplacian spatial filtering. The most basic requirement in defining the Laplacian filter is such that the center coefficient must be positive and that the outer coefficients negative. Because the Laplacian is a derivative, the sum of the filter coefficients has to be zero. Laplacian filters can be constructed using:

$$h(\alpha) = \frac{4}{\alpha + 1} \begin{pmatrix} \frac{-\alpha}{4} & \frac{\alpha-1}{4} & \frac{-\alpha}{4} \\ \frac{\alpha-1}{4} & 1 & \frac{\alpha-1}{4} \\ \frac{-\alpha}{4} & \frac{\alpha-1}{4} & \frac{-\alpha}{4} \end{pmatrix} \tag{3.3.4}$$

where the parameter α controls the shape of the Laplacian and must be in the range $(0, 1)$.

Chapter 4

Distance Estimation

The present chapter shows the imaging system model that simulates image blurring caused by an aberration in focusing where the detector is modeled by the general response of typical neuromorphic imagers. It also outlines the proposed distance estimation algorithm to be applied on an image obtained from such a system. The algorithm involves a numerical optimization of a cost function which has been defined according to a suitable measure for the amount of image focus. The optimal system parameters are then used to calculate distance.

4.1 Imaging System Model

This section sets up the prototype imaging system to be used in the simulations. We take the imaging system as before from figure 2.5.2 but now the detector is replaced by a neuromorphic imager. A schematic diagram is shown in figure 4.1.1. Since both image defocus and imager response were modeled as LSI systems (i.e., a

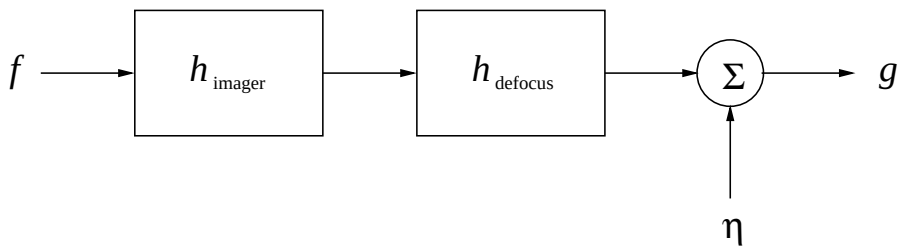

Figure 4.1.1: Prototype imaging system.

convolution between an input signal and a linear filter) as expressed by equations (2.5.8) and (3.3.3), we can combine both filters h_{imager} and h_{defocus} into the cascade of LSI systems as shown in figure 4.1.1. We choose the Laplacian edge detector of (3.3.4) as the h_{imager} filter. The image defocus is modeled by the Gaussian PSF (2.5.6). One can obtain a finite impulse response (FIR) representation of the filter h_{defocus} by sampling the Gaussian OTF with a two-dimensional windowing method. Setting the vector of camera parameters $\boldsymbol{e}$ constant, the only remaining parameter that controls the shape of the blurring PSF will be the object distance z_o (i.e., the distance from the lens aperture to the image patch f) so we can write $h_{\text{defocus}}(\boldsymbol{e}, z_o) = h_{\text{defocus}}(z_o)$. One might also take into account an additive noise term η. The noise term is important because in practical imaging situations, additive noise is not negligible. Common types of noise may come from (i) electronic noise resulting from thermal motion of electrons in the electronic components of the imaging system, (ii) photoelectric noise due to the statistical nature of light in the photoelectric conversion process in the photopixels, and (iii) quantization noise which occurs during image digitization. Therefore given a small image patch f of the scene we have the image g at the detector expressed as:

$$\begin{aligned} g &= h_{\text{defocus}}(z_o) \otimes \{h_{\text{imager}} \otimes f\} + \eta \\ &= h_{\text{defocus}}(z_o) \otimes f_\Delta + \eta \end{aligned} \tag{4.1.1}$$

where the ideal image has been denoted by $f_\Delta = h_{\text{imager}} \otimes f$.

In our simulations, we will be using the 256×256 image at the left of figure 4.1.2 as the input f and assume that it is only a small patch of the entire scene that we want to explore. At the right of the same figure is the ideal image f_Δ as obtained from the neuromorphic imager using the Laplacian filter with $\alpha = 0.5$. Figure 4.1.3 shows the defocused image where η is an additive white Gaussian noise (WGN) with variance 0.1% of the image gray values. Here, the object is placed at $z_o = 4$ m and the camera parameters are set to $F = 100$ mm, $z_d = 101$ mm, and $D = 6$ mm.

Figure 4.1.2: An image patch of a scene that one obtains from a conventional digital camera (left). The ideal image obtained from a neuromorphic imager as modeled by Laplacian spatial filtering (right).

Figure 4.1.3: The defocused image.

4.2 Distance Estimation Algorithm

To lay the groundwork for the proposed distance estimation algorithm, some assumptions regarding the imaging system have to be raised. These will eventually set some practical constraints on our optimization algorithm. First of all, the distance estimation algorithm is only for one subimage of the whole scene that has to be explored. We define a large but finite *maximum working distance* z_{max} and fix the detector plane behind the lens exit pupil at a distance z_d such that it is at the image plane for objects at z_{max}, i.e.,

$$\frac{1}{z_d} = \frac{1}{F} - \frac{1}{z_{max}}.$$

This assumption is not too restrictive. Of course one can always fix the detector at $z_d = F$ which means that the system is focused at infinity. However, it seems impractical to measure distances to objects at such extremely large distances for our purposes. Thus we fix z_d close to but greater than F so that we will have a reasonably large z_{max} which is not too far from the imaging system. For the same reason, we also set a not too close *minimum working distance* z_{min} since it would be trivial to estimate the distance to objects which are very near to the imaging system. For notational convenience, one defines the set of working distances as

$$\Omega_z \equiv \{z \in \mathbb{R}^+ | z_{max} \geq z \geq z_{min}\} \tag{4.2.1}$$

and that any $z \notin \Omega_z$ will be ignored. That is, any object at $z > z_{max}$ is said to be very far (i.e., "at infinity") and that objects nearer than z_{min} are just too near to the system that it is a trivial thing to go through all the optimization for the distance estimation. Consequently from (4.2.1), a set of values for the distance difference ε between the detector plane and the image plane may be defined as

$$\Omega_\varepsilon \equiv \{\varepsilon \in \mathbb{R}^+ | 0 \leq \varepsilon \leq \varepsilon_m\}. \tag{4.2.2}$$

Obviously from the thin lens equation, there is a one-to-one correspondence between the sets Ω_z and Ω_ε. That is, for any $z_o \in \Omega_z$, there corresponds only one $\varepsilon_o \in \Omega_\varepsilon$ such that

$$\frac{1}{z_o} = \frac{1}{F} - \frac{1}{z_d + \varepsilon_o}.$$

The system is now set up such that the detector (i.e., the neuromorphic imager) is fixed at z_d or equivalently at $\varepsilon = 0$. For the moment we assume that the system does not introduce any noise. In general, since the object plane is at $z \neq z_{max}$, one obtains a blurred image brought about by the defocusing PSF. Consider the case of an object at $z_o \in \Omega_z$, $z_o \neq z_{max}$. This leads to a blurred image $g(z_o)$ due to low-pass filtering by $h_{\text{defocus}}(z_o)$. Our main problem is to extract the distance z_o solely from a single measurement of $g(z_o)$. This will have to be done by a passive focusing approach where one "virtually" moves the imager plane through a distance ε. The object distance is estimated by searching in the Ω_ε space for the optimal image focus. This implies that at each optimization step, one calculates the new image $g_\varepsilon(z_o)$ such that

$$g_\varepsilon(z_o) = h_{\text{defocus}}(\varepsilon, z_o) \otimes f_\Delta \tag{4.2.3}$$

for $\varepsilon \in \Omega_\varepsilon$. Observe that it is necessary to have prior information about the ideal image f_Δ which obviously is unknown. Hence we have to find a way of recovering f_Δ. We point out that this problem of recovering the ideal image from an image which is degraded by an unknown or partially known degradation model is an *ill-posed* problem. For instance in our case, there is only one measurement $g(z_o)$ but two unknowns, $h_{\text{defocus}}(\varepsilon, z_o)$ and f_Δ. Even if we have partial knowledge about the PSF since its analytical form is known, we never know its exact shape. Otherwise it would mean that z_o is already known in advance which is of course a trivial situation. The major obstacle is how to retrieve f_Δ from the $g(z_o)$. The usual way of recovering the undegraded image is by *inverse filtering* techniques or some of its variants like Weiner filtering or regularized inverse filtering techniques. However, these doesn't help in any way since all these methods require the explicit knowledge of the blurring filter, which as already mentioned, must not be known a priori. Hence we are faced with two problems. The first and most pressing is to recover the ideal image f_Δ solely from $g(z_o)$ (i.e., independently from the PSF). The second is to determine the shape of $h_{\text{defocus}}(\varepsilon, z_o)$ that causes the blurring. Solving the latter eventually leads to the estimation of z_o.

We first address the second problem. Suppose that we have a way of recover-

ing f_Δ from $g(z_o)$. The problem of using the approach in equation (4.2.3) is that $h_{\text{defocus}}(\varepsilon, z_o)$ is unknown since it depends on the unknown z_o. An alternative to (4.2.3) is to exploit the one-to-one correspondence between the sets Ω_z and Ω_ε so that instead of searching in the Ω_ε space, one can instead use Ω_z as the search space. That is, one can equivalently search z instead of ε. Therefore as an alternative to (4.2.3), we take

$$g(z') = h_{\text{defocus}}(z') \otimes f_\Delta. \tag{4.2.4}$$

In other words, one has to *reblur* the image by convolving the recovered f_Δ with h_{defocus} for different values of $z' \in \Omega_z$. Now we turn back to the first problem. One possibility is to use *blind deconvolution* techniques since they do not require any knowledge about the blurring PSF. Blind deconvolution can only yield an approximate restoration $\hat{f_\Delta}$ to f_Δ. In that case, one has the reblurred image (4.2.4) at $z = z'$:

$$\hat{g}(z') = h_{\text{defocus}}(z') \otimes \hat{f_\Delta} \tag{4.2.5}$$

where $\hat{f_\Delta}$ denotes the restoration of f_Δ using blind deconvolution. Note that having blind deconvolution recover the ideal image is quite tricky. We must consider the fact that for the real-time walking stick application, we must have a suitably fast and reliably accurate distance estimation algorithm. However, most blind deconvolution algorithms converge slowly and in some cases, may never even arrive to a unique restoration.

The flow of the distance estimation algorithm as described in equation (4.2.5) is shown in figure 4.2.1. The following subsections will discussed the details of each step.

Recovering the Ideal Image

The general image blind deconvolution problem refers to the task of separating two convolved signals when they are either unknown or partially known [20]. An important consideration is that the restored signals must be unique and should not lead to any ambiguity. However, in general, the image restoration problem using blind

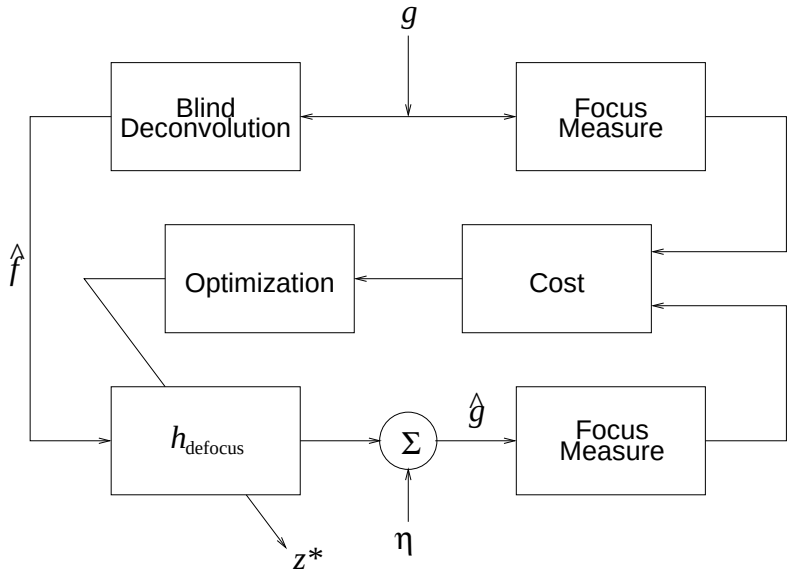

Figure 4.2.1: Distance estimation algorithm.

deconvolution is *ill-conditioned* and can only give an approximate restoration which may not be unique.

There are several existing techniques for image blind deconvolution. These include among others iterative techniques as well as heuristic approaches like simulated annealing. These methods tend to be very slow for our purposes. However there exist algorithms that have been implemented for the acceleration of some iterative blind deconvolution methods. In this work, we will use an accelerated Richardson-Lucy (R-L) blind deconvolution algorithm [21]. Given the image blurring model (4.1.1), The iterative R-L algorithm can be expressed as

$$\hat{f_\Delta}^{(k+1)} = \hat{f_\Delta}^{(k)} \operatorname{corr}\left(h^{(k)}, \frac{g}{h^{(k)} \otimes \hat{f_\Delta}^{(k)}}\right) \tag{4.2.6}$$

where $\hat{f_\Delta}^{(k)}$ and $h^{(k)}$ are the estimate of the ideal image and the PSF after k iterations, respectively. The PSF estimate at each iteration can be obtained by using any variant of inverse or Weiner filtering. Additional constraints like nonnegativity have to be satisfied for both image and PSF estimates. The PSF also has to satisfy normalization constraints such that its sum always equals unity. What the iterative R-L does is that it attempts to do a maximum likelihood of the restored image and

has been used heavily for the restoration of noisy astronomical images. The accelerated version has been reported to yield a considerable reduction in restoration times given a small image distortion [21].

Cost Function Definition

In defining the cost function, we need some Φ that measures the amount of image focus. The main requirements are that Φ must be monotonic and unimodal in the search space. We define the cost function as

$$C(z') = \|\Phi\{\hat{g}(z')\} - \Phi\{g(z_o)\}\|^2. \tag{4.2.7}$$

The optimization aim is to minimize C to get the optimal object distance

$$z^* = \underset{z' \in \Omega_z}{\operatorname{argmin}} C(z'). \tag{4.2.8}$$

Recall that in chapter 2 we have observed that the defocusing PSF acts as a low-pass filter and that the OTF increasingly attenuates the high frequency components of an image due to the decrease of the OTF bandwidth resulting from the increase of the size of the blur circle. Recall also that the size of the blur circle is related to the object distance and that we have set up the system in such a way that the OTF bandwidth decreases as the object distance decreases. We will define Φ from these observations.

Since convolution in the spatial domain corresponds to multiplication in the spatial frequency domain, the blurring model (2.5.8) may be written as

$$\hat{\mathcal{G}}_{z'}(\omega, \nu) = \mathcal{H}^{(z')}_{\text{defocus}}(\omega, \nu)\hat{\mathcal{F}}_{\Delta}(\omega, \nu) \tag{4.2.9}$$

where (ω, ν) are the spatial frequency domain coordinates, $\hat{\mathcal{G}}_{z'}$ and $\hat{\mathcal{F}}_{\Delta}$ are the Fourier spectra of $\hat{g}(z')$ and $\hat{f_{\Delta}}$ respectively, and $\mathcal{H}^{(z')}_{\text{defocus}}$ is the OTF that corresponds to $h_{\text{defocus}}(z')$. We follow [6] in defining the focus measure in terms of the aforementioned observations by taking the integral

$$\Phi\{\hat{g}(z')\} = \iint \alpha(|\hat{\mathcal{G}}_{z'}(\omega, \nu)|)\mathrm{d}\omega\mathrm{d}\nu$$

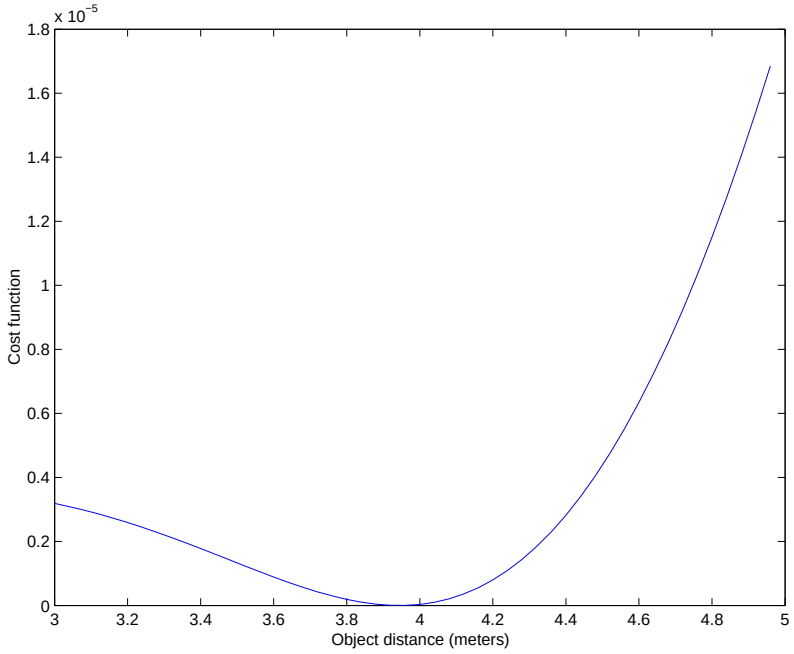

Figure 4.2.2: Cost function at different object distances when the object is at 4 m.

where $\alpha(\cdot)$ is a monotonically increasing function of its argument. Based upon their results [6], we take $\alpha(\cdot)$ as a bandpass filtering operation such that it is equivalent to taking the energy of the low-pass filtered Laplacian of the image. The good thing about this choice for Φ is that it coincides perfectly to our model. That is, since we already have $\hat{g}(z')$ which is a low-pass filtered image Laplacian, applying $\Phi\{\cdot\}$ to $\hat{g}(z')$ is just equivalent to taking

$$\Phi\{\hat{g}(z')\} = \iint |\hat{\mathcal{G}}_{z'}(\omega, \nu)|^2 \mathrm{d}\omega \mathrm{d}\nu. \tag{4.2.10}$$

In other words, $\Phi\{\hat{g}(z')\}$ is just the total power of the image spectrum $\hat{\mathcal{G}}_{z'}$. It can be easily shown that Φ is monotonic and unimodal such that the cost function (4.2.7) will have a unique minimum when $z' = z_o$. Figure 4.2.2 shows a plot of the cost function for different values of z' when the object is at $z_o = 4$ m.

Minimizing the Cost Function

Having seen from the previous subsection that the cost function is unimodal in the one-dimensional search space Ω_z, we now decide which function minimization technique we have to use. Again we have to consider that what we will eventually need is

a simple and straighforward technique that is not too computationally expensive but still able to yield accurate results. Usually the computational effort is dominated by the number of function evaluations such that what is essentially desired is to have a method that performs as few function evaluations as possible. For this reason it is decided to use a simple line search technique, in particular a golden section search.

From the plot of $C(z)$ in figure 4.2.2, we see that if we normalize z in the interval $[0, 1]$, there is a value $0 < c < 1$ such that C is strictly decreasing in $[0, c)$ and strictly increasing in $[c, 1]$. The only operation that we have to do is to choose a value x, $0 \leq x \leq 1$, and probe the value of $C(x)$. In this work we will use the algorithm from [22]. The algorithm goes as follows. First, one sets an error tolerance $\epsilon(0 < \epsilon \leq 1)$. Next, one initializes the quadruple (a, b, c, d) with $a = 0$, $b = 1$, $c = \phi^2$, and $d = \phi$, where $\phi = \frac{\sqrt{5}-1}{2}$. The function is evaluated at each point in the quadruple. At each step of the search, if the current search interval is $[a, b]$, two probes are created, one at $c = a + (b - a)\phi^2$ and the other at $d = a + (b - a)\phi$ and the function is evaluated for c and d. There are two possibilities: (i) if $C(c) < C(d)$, the search continues in $[a, d]$, otherwise (ii) the search continues in $[c, b]$. Note that in both cases, the next probe d is the same as c from the previous iteration, so only one function evaluation is needed each time. It is shown [22] that the total number of function evaluations is only $1 + \log_{\phi}(\frac{1}{\epsilon})$. The algorithm is outlined below.

Algorithm 1 Golden Section Search

Initialize ϵ;

$a \Leftarrow 0$; $b \Leftarrow 0$; $c \Leftarrow \phi^2$; $d \Leftarrow \phi$;

$fa \Leftarrow C(a)$; $fb \Leftarrow C(b)$; $fc \Leftarrow C(c)$; $fd \Leftarrow C(d)$;

while $|a - b| > \epsilon$ **do**

  if $fc < fd$ **then**

    $b \Leftarrow d$; $fb \Leftarrow fd$;

    $d \Leftarrow c$; $fd \Leftarrow fc$;

    $c \Leftarrow a + (b - a) * \phi^2$; $fc \Leftarrow C(c)$;

  else

    $a \Leftarrow c$; $fa \Leftarrow fc$;

    $c \Leftarrow d$; $fc \Leftarrow fd$;

    $d \Leftarrow a + (b - a) * \phi$; $fd \Leftarrow C(d)$;

  end if

end while

Return a;

Chapter 5

Results

5.1 Simulation

The imaging system is set up according to the model in figure 4.1.1 and equation (4.1.1). The camera parameters are chosen such that $z_d = 101$ mm, $F = 100$ mm, and $D = 6$ mm. The set of working distances are such that $z_{max} = 4.5$ m and $z_{min} = 3.2$ m. We take the Gaussian PSF model for h_{defocus} and construct a 5×5 FIR filter representation out of it by sampling its OTF with a Gaussian windowing method. The imager is modeled by the 3×3 Laplacian spatial filter (3.3.4) where $\alpha = 0.2$. Figure 5.1.1 shows an image of both filters, the defocusing filter is due to an object placed at a distance $z_o = 4$ m from the lens aperture. In all simulations we introduce an additive white Gaussian noise (WGN) having zero mean and variance

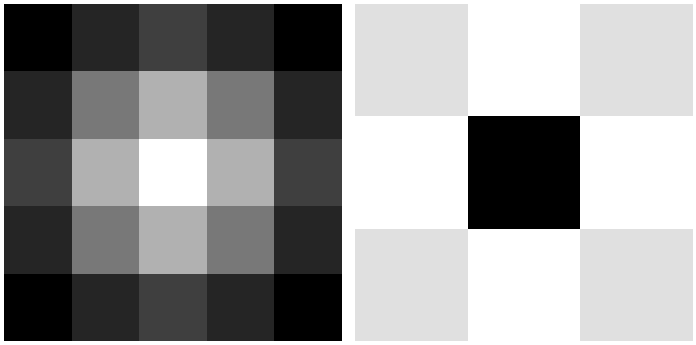

Figure 5.1.1: The blurring PSF when the object is at $z_o = 4$ m (left) and the Laplacian filter with $\alpha = 0.2$.

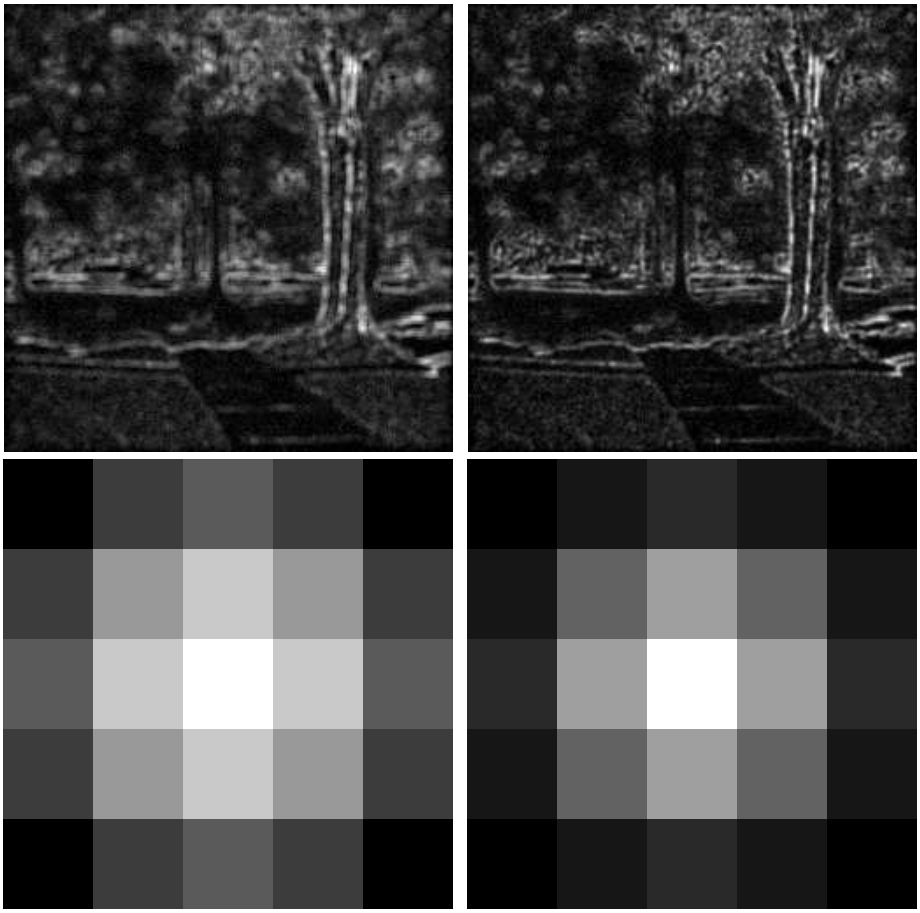

Figure 5.1.2: The images obtained from the imaging system simulation when the object is at z_{min} (left) and at z_{max} (right). The bottom figures show the corresponding blurring PSFs. Note their difference with respect to distance.

0.1% of the image gray levels. The gray levels are also normalized in the range $[0, 1]$ such that 0 corresponds to black and 1 corresponds to white. Figure 5.1.2 shows the images from the simulation of the imaging system when the object is at z_{min} and at z_{max}, together with their corresponding defocusing PSFs.

5.2 Distance Estimation

Now that we have the images at the detector, we proceed with the estimation of distance. Following the flow described in figure 4.2.1, the ideal image has to be recovered first by blind deconvolution. It should be pointed out that this algorithm

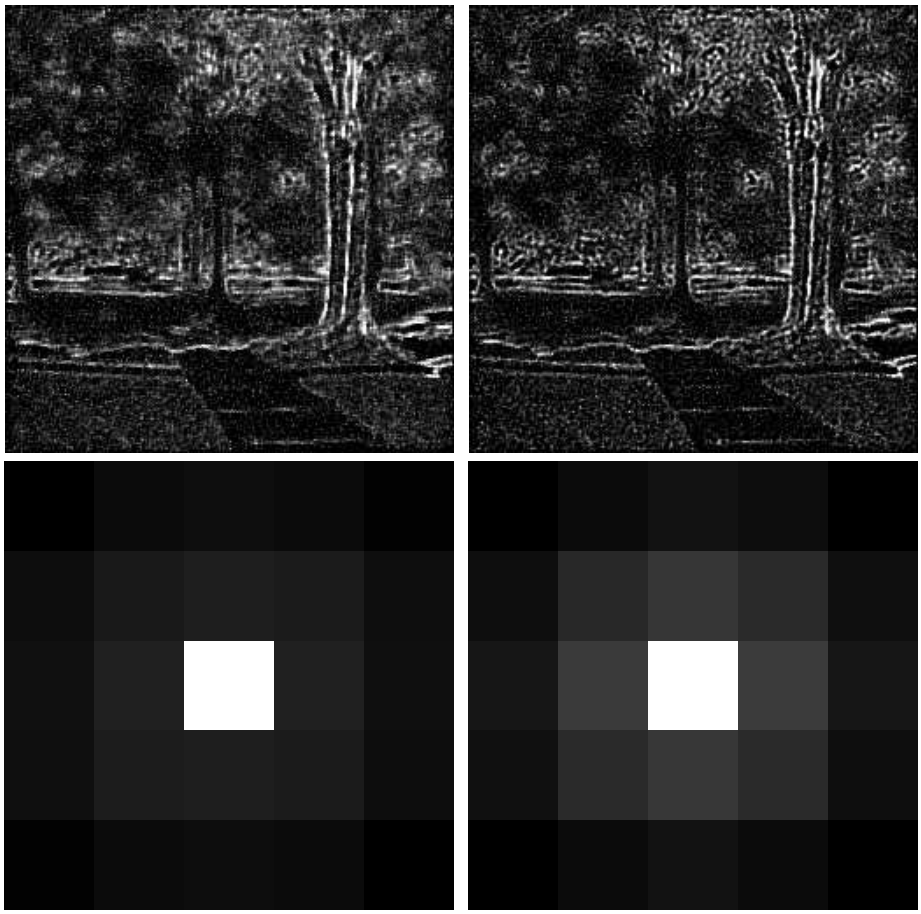

Figure 5.2.1: The restored images when the object is at z_{min} (left) and at z_{max} (right). The bottom figures show the corresponding restored PSFs.

is only applied to one small subimage of the entire scene. The MATLAB routine `deconvblind.m` was used to do the deconvolution task. It implements an accelerated R-L algorithm by iteratively restoring both the image and the PSF. An array of ones having the same size as the blurring filters serves to initialize the PSF and as the deconvolution terminates, it returns the recovered PSF as well as the image. Figure 5.2.1 shows the restoration results when the object is at z_{min} and at z_{max}. Each restoration takes 20 iterations. The quality of the restoration can be inferred by comparing the restored images with the ideal image shown at the right of figure 4.1.2. One can also compare the restored PSFs with the actual blurring PSFs of figure 5.1.2.

The next step is to calculate the amount of focus on the image obtained from the detector. In order not to introduce any unwanted high frequency contributions caused by the discontinuity at the image borders, a smooth window function (e.g., a Gaussian) is applied on the image which smoothly tapers the gray values at the borders. What follows next is the distance optimization. Essentially we minimize the cost function (4.2.7) by using the golden section search algorithm described in the previous chapter. The error tolerance was fixed at $\epsilon = 0.001$ for all runs. Figure 5.2.2 shows the results from the optimization algorithm. Each point corresponds to the best distance estimate for 40 runs. For accurate distance estimates, the points should all lie along the straight line. However we see that there is only a very narrow region where the algorithm has acceptable performance. Figure 5.2.3 shows the relative error that compare the estimated distance with the actual distance. Outside the range where the relative error is small, one gets at most 6% difference corresponding to an error of about ± 6 cm. Within the range of best distance estimates, an error of about ± 0.5 cm. However, looking at figure 5.2.4 which is a plot of the standard deviation of each run of the algorithm for each distance, we can see that the algorithm does not give reliable distance estimates. It yields uncertainties as high as 40% or around ± 40 cm.

The major contributor to these high uncertainties comes from the blind decon-

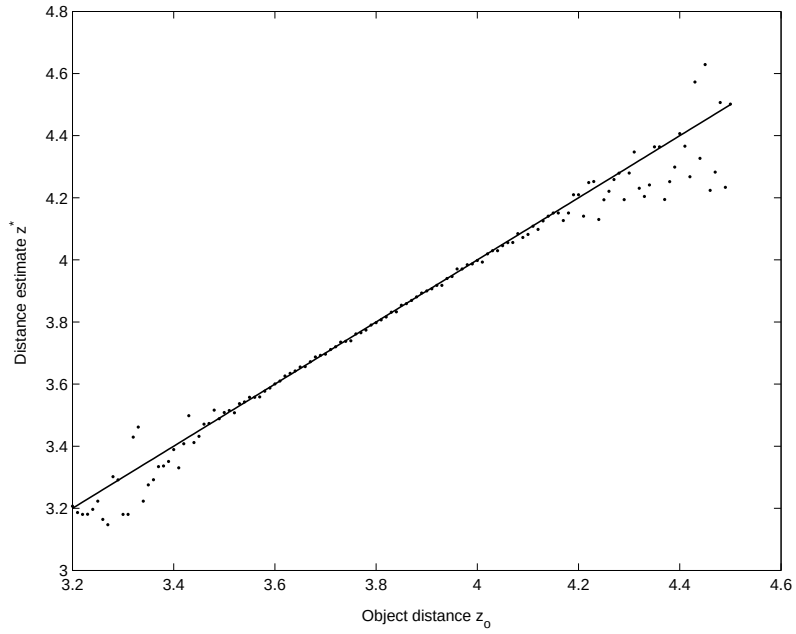

Figure 5.2.2: The estimated distances resulting from the optimization algorithm plotted against the actual object distances.

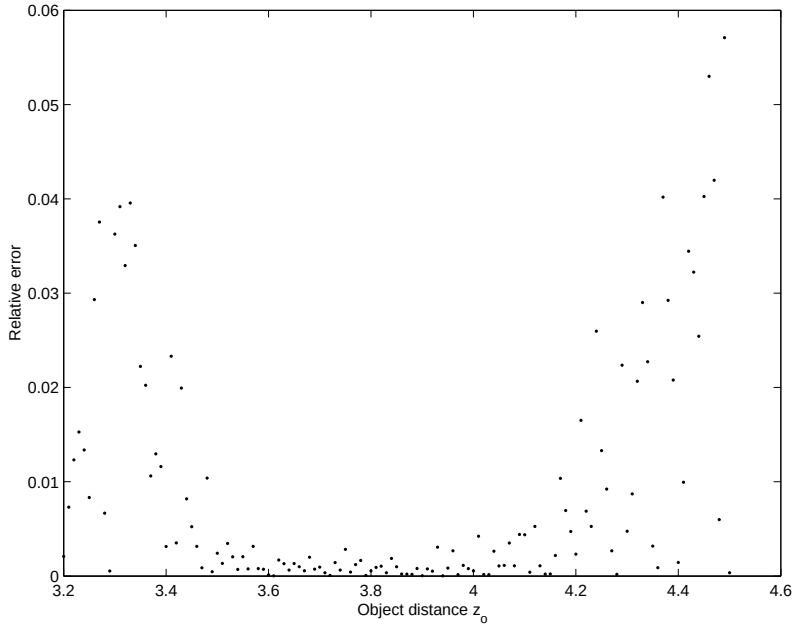

Figure 5.2.3: Relative error.

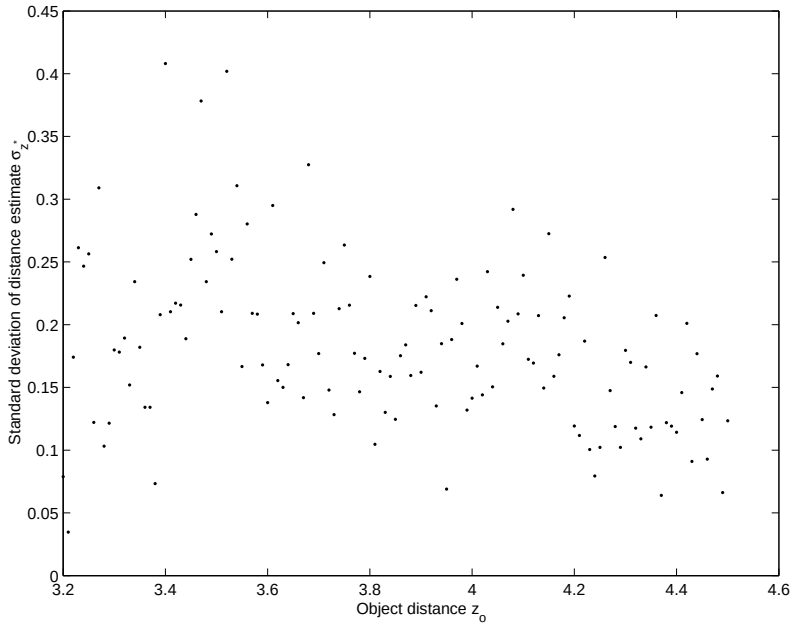

Figure 5.2.4: The variability of the output of the optimization algorithm run 40 times for each object distance.

volution itself. One can see this by plotting results similar to figures 5.2.2, 5.2.3, and 5.2.4 but without using blind deconvolution. That is, we assume that we have knowledge about the ideal image from a second imaging system which is able to capture the ideal image (e.g., an imaging system that has a very small aperture). Here we can see that the uncertainties as shown in figure 5.2.7 are much less compared to those in figure 5.2.4.

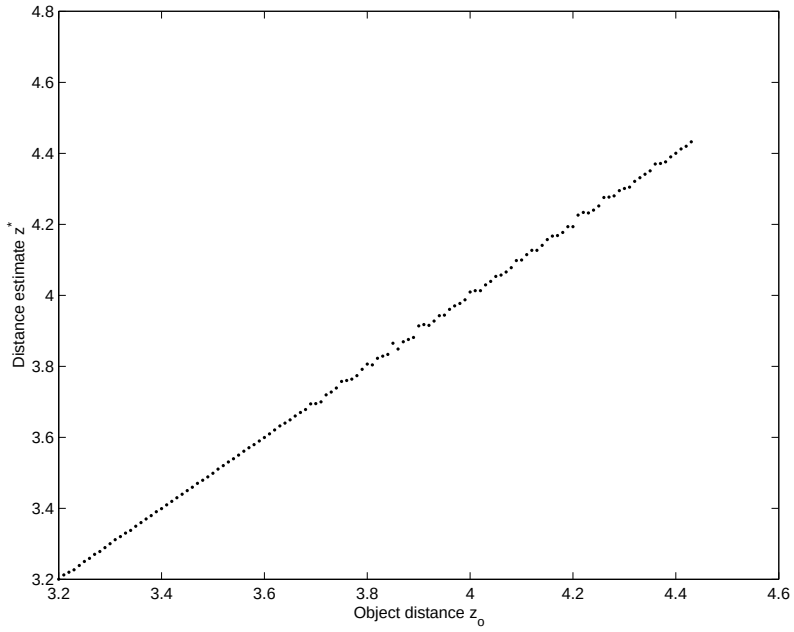

Figure 5.2.5: The estimated distances resulting from the optimization algorithm plotted against the actual object distances when the ideal image is known.

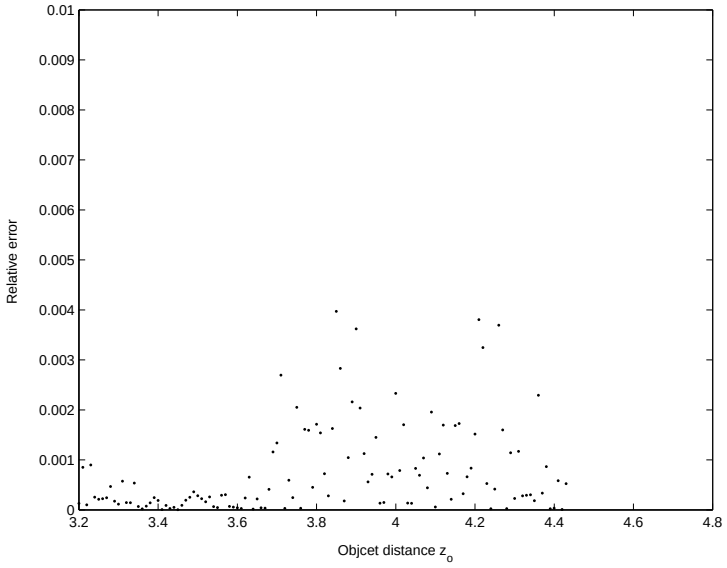

Figure 5.2.6: Relative error when the ideal image is known.

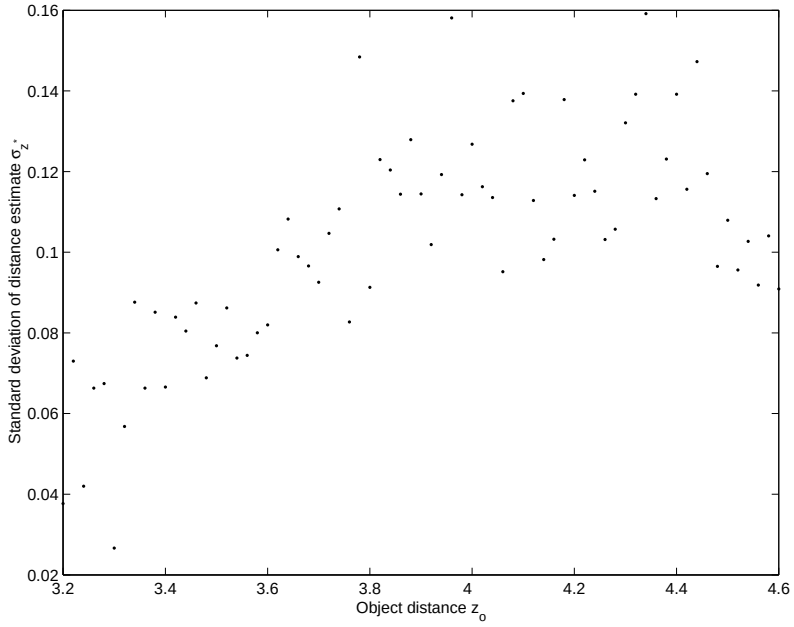

Figure 5.2.7: The variability of the output of the optimization algorithm run 40 times for each object distance when the ideal image is known.

Chapter 6

Conclusions

In this work we have proposed a distance estimation algorithm which will be used for an imaging system that has a neuromorphic imager as image detector. We have simulated such an imaging system that includes the optics that causes image defocusing and a model for the imager's response. The model is essentially that of a two-dimensional LSI system that convolves the ideal image (i.e., the image observed at the detector when the imaging system is diffraction-limited) with the system's point spread function. Although we have been a bit loose in the sense that we have modeled the imager as a Laplacian edge detector, this is sufficient as a first step in this feasibility study. Even if there is a wide variety of neuromorphic imagers available, they all have one common feature, that of edge-detection. This is due to the fact that these imagers try to mimic the system of lateral inhibition in the vertebrate retina which is the first step towards sharpening contours similar to that of Laplacian spatial filtering.

The difficult part of the work was in choosing the distance estimation algorithm. We are basically faced with an ill-posed and ill-conditioned inverse problem of trying to separate two unknown convolved signals that yield the defocused image. That is, we are given a single observation (i.e., the image at the detector) which is modeled as a convolution of an ideal image and the system's PSF, plus noise. We have seen how the PSF determines the amount of blur incurred on the observed image and how the PSF size relates with object distance. As the object gets nearer, the PSF

size gets bigger and the amount of defocus increases. The size of the PSF is related to the imaging system parameters and that includes object distance. Since we want to estimate the distance, we must find a way of extracting the distance information from the PSF by somehow optimizing the amount of focus on the image. The crucial part is in knowing the ideal image. In this work we have opted to use blind deconvolution to recover the ideal image since it is the only known method that restores defocused images without any prior knowledge of the PSF. Of course, knowing the exact shape of the PSF beforehand would make the problem trivial since it implies that the object distance is already known.

As the results show, although our algorithm was able to provide a fairly accurate distance estimate within a narrow range of object distances, the uncertainties associated to those estimates are quite large. The main contributing factor to these high uncertainties comes from the use of blind deconvolution itself since it doesn't yield a unique restoration of the ideal image. Although it is tempting to recommend trying to look for better blind deconvolution algorithms that yield more or less unique solutions in order to minimize the uncertainties of the distance estimates, we have to consider also the computational expenses that surely will lead to big tradeoffs. What is essentially needed is a fast, computationally cheap, and reliable algorithm in order to implement it real-time on the extended walking stick. Usually these blind deconvolution algorithms are very slow for such practical applications and when one wants to explore an entire scene by using this approach for every subimage, obviously the computational expenses will increase by a factor equal to the number of subimages found on the scene.

All these complications stem from the fact that we have been trying to solve a hopelessly impossible to solve inverse problem. However there is hope if one has a way to obtain the ideal image by other means besides blind deconvolution. Perhaps one can use a second image using another imaging system having different parameter settings. For instance, the second imaging system may have a very small aperture diameter such that its PSF size is very small and deos not change so much with ob-

ject distance. In this case one can opt for other depth estimation techniques which have been proven to yield accurate results and are less computationally expensive such as DFD methods. Another alternative would be to construct a kind of a lookup table of the different PSFs for known object distances and use them to initialize the blind deconvolution algorithm. Also it is possible to do away entirely with blind deconvolution. Since one already has a reference of PSFs, then one might be able to use ordinary inverse filtering techniques to recover the ideal images.

Bibliography

[1] C. Mead (1989), **Analog VLSI and neural systems**, Addison-Wesley.

[2] V. Dante, P. Del Giudice and M. Mattia (2001), *Implementation of Neuromorphic Systems: from discrete components to analog VLSI chips (testing and communication issues)*, Ann. Ist. Super. Sanita, **37** 2, pp 231-239.

[3] Y. F. Liu (1998), *A unified approach to image focus and defocus analysis*, Ph.D. Thesis, Dept. of Electrical Engg., SUNY at Stony Brook.

[4] A. Pentland (1987), *A new sense of depth of field*, IEEE Trans. Patt. Anal. Machine Intell., **PAMI-9**, pp 523-531.

[5] P. Grossman (1987), *Depth from focus*, Pattern Recognition Letters, **5**, pp 63-69.

[6] M. Subbarao, T. Choi, and A. Nikzad (1993), *Focusing techniques*, Optical Engineering, **32** 11, pp 2824-2836.

[7] Y. Xiong and S. Shafer (1993), *Depth from focusing and defocusing*, Tech. report CMU-RI-TR-93-07, Robotics Institute, Carnegie Mellon University.

[8] M. Born and E. Wolf (1999), **Principles of Optics: electromagnetic theory of propagation, interference, and diffraction of light, Seventh (Expanded) Edition**, Cambridge.

[9] J. Goodman (1996), **Introduction to Fourier Optics, Second Edition**, McGraw Hill.

[10] J. Gaskill (1978), **Linear systems, Fourier transforms, and Optics**, Wiley.

[11] A. Papoulis (1968), **Systems and Transforms with Applications in Optics**, McGraw Hill.

[12] B. Horn (1986), **Robot Vision**, MIT Press.

[13] H. Kolb (2003), *How the retina works*, American Scientist, **91** 1, pp 28-35.

[14] A. Moini (1997), *Vision chips or seeing silicon*, Tech. report, Centre for High Performance Integrated Technologies and Systems, The University of Adelaide, March 1997.

[15] M. Giulioni (2003), *Analisi e progetto di sensori ottici neuromorfi*, Tesi di Laurea (Università degli Studi di Roma La Sapienza).

[16] J. Kramer (2002), *An On/Off transient imager with event-driven, asynchronous read-out*, 2002 IEEE International Symposium on Circuits and Systems.

[17] J. Kramer, *An integrated optical transient sensor*, unpublished.

[18] V. Dante, P. D. Giudice (2001), *The PCI-AER interface board*, Proc. Neuromorphic Engineering Workshop.

[19] R. Gonzalez and R. Woods (1992), **Digital Image Processing**, Addison-Wesley.

[20] D. Kundur and D. Hatzinakos (1996), *Blind image deconvolution*, IEEE Signal Processing Magazine, **13**, 3 pp 43-64.

[21] D. Biggs and M. Andrews (1997), *Acceleration of iterative image restoration algorithms*, Applied Optics, **36**, 8 pp 1766-1775.

[22] B. Ravikumar and G. Cooper (1997), *Fibonacci search algorithm for finding extremum using unary predicates*, submitted for publication.

[23] M. Subbarao and M. Lu (1992), *Computer modeling and simulation of camera defocus*, Tech. report 92.01.16, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350.

[24] O. Ghita and P. Whelan (2000), *Real time 3D estimation using depth from defocus*, Vision, MVA (SME), **16**, 3 pp 1-6 (Invited paper).

[25] J. Kramer and G. Indiveri (1998), *Neuromorphic vision sensors and preprocessors in system applications*, Advanced Focal Plane Arrays and Electronic Cameras, Zurich, (May 1998).

[26] T. Delbrück (2000), *Silicon retina for autofocus*, Institute for Neuroinformatics (INI), Zurich.

[27] P. Lichtsteiner and T. Delbrück (2004), *WP1.5 AER retina characterization*, INI Workshop: Integration of Neuromorphic Systems to Sensory-Motor processing, Zurich (March 2004).

[28] L. Lucy (1974), *An iterative technique for the rectification of observed distributions*, The Astronomical Journal, **79**, 6 pp 745-754.

[29] L. Lucy (1994), *Optimum strategies for inverse problems in statistical astronomy*, Astron. Astrophys., **289**, pp 983-994.

[30] G. Ricort et al (1993), *Application of the Richardson-Lucy algorithm to the deconvolution of two-fold probability density functions*, Pure Appl. Opt., **2**, pp 125-143.

[31] D. Kundur and D. Hatzinakos (1998), *A novel blind deconvolution scheme for image restoration using recursive filtering*, IEEE Trans. Sig. Proc., **46**, 2 pp 375-390.

[32] R. Neelamani, H. Choi, R. Baraniuk (2004), *ForWaRD: Fourier-Wavelet regularized deconvolution for ill-conditioned systems*, IEEE Trans. Sig. Proc., **52**, 2 pp 418-433.

[33] M. Subbarao, and J.K. Tyan (1995), *The optimal focus measure for passive autofocusing and depth from focus*, Proceedings of SPIE conference on Videometrics IV, Philadelphia, Oct 1995.

[34] N. Nill and B. Bouzas (1992), *Objective image quality measure derived from digital image power spectra*, Optical Engineering, **31**, pp 813-825.

[35] J. Kautsky, et al (2002), *A new wavelet-based measure of image focus*, Pattern Recognition Letters, **23**, pp 17851794.

[36] V. Aslanta (2003), *Estimation of depth from defocusing using a neural network*, IJCI Proc. of Int. Conf. on Sig. Proc., **1**, 2.

[37] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery (1992), **Numerical Recipes in C, Second Edition**, Cambridge University Press.

[38] H. Merklinger (2002), **The Ins and Outs of Focus, Internet Edition**, http://www.trenholm.org/hmmerk/download.html.