Tradeoffs and Limits in Computational Imaging

Oliver Cossairt

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Tradeoffs and Limits in Computational Imaging

Oliver Cossairt

For centuries, cameras were designed to closely mimic the human visual system. With the rapid increase in computer processing power over the last few decades, researchers in the vision, graphics and optics community have begun to focus their attention on new types of imaging systems that utilize computations as an integral part of the imaging process. Computational cameras optically encode information that is later decoded using signal processing. In this thesis, I show three new computational imaging designs that provide new functionality over conventional cameras. Each design has been rigorously analyzed, built and tested for performance. Each system has demonstrated an increase in functionality over tradition camera designs. The first two computational imaging systems, *Diffusion Coding* and *Spectral Focal Sweep*, provide a means to computationally extend the depth of field of an imaging system without sacrificing optical efficiency. These techniques can be used to preserve image detail when photographing scenes that span very large depth ranges. The final example, *Gigapixel Computational Imaging*, uses a computational approach to overcome limitations in spatial resolution that are caused by geometric aberrations in conventional cameras.

While computational techniques can be used to increase optical efficiency, this comes at a cost. The cost incurred is noise amplification caused by the decoding process. Thus, to measure the real utility of a computational approach, we must weigh the benefit of increased optical efficiency against the cost of amplified noise. A complete treatment must take into account an accurate noise model. In some cases, the benefit may not outweigh the cost, and thus a computational approach has no value. This thesis concludes with a discussion on these scenarios.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Research on Spectral Focal Sweep was supported in part by an NSF Graduate Research Fellowship. Research on the Gigapixel Computational Camera was supported in part by DARPA Award No. W911NF-10-1-0214 and an NSF Graduate Research Fellowship. Research on Diffusion Coding Photography was supported in part by the Office of Naval Research through the awards N00014-08-1-0329 and N00014-09-1-0638 and an NSF Graduate Research Fellowship.

This work would not have been possible without the assistance of my fellow graphics and vision researchers, whom I admire greatly on both a professional and personal level. Ravi Ramamoorthi helped advise me during my first few years as Ph.D. student, and he helped set in motion the progress I went on to make in research, writing, and presentation. The numerous discussions about optics and computational imaging with my officemate, Changyin Zhou, were a catalyst for much of my early work, and our collaborations were instrumental in my success as a Ph.D student. The ideas that emerged from our collaborations fueled my imagination for the duration of my Ph.D, and will continue to be a catalyst after I graduate. Daniel Miau contributed his excellent engineering skills to capturing images for the Gigapixel Computational Camera project. Neeraj Kumar contributed his invaluable editing skills on numerous occasions. I credit much of the my progress in both writing and presenting to his exhaustive comments and criticisms. Much of the ideas in the last chapter of this thesis are the result of the numerous and exhaustive conversations with Mohit Gupta on the topic of the limits of computational imaging.

Over the years I have had the pleasure of working side by side with many excellent researchers at Columbia, including Guru Krishnan, Kshitiz Garg, Sujit Kuthirummal, Bo Sun, Dhruv Mahajan, Ryan Overbeck, Kevin Eagan, Alex Berg, Yasunobu Hitomi, Toshihiro Kobayashi, Hajime Nagahara, Dimitri Bitouk, Li Zhang, Francesc Moreno-Noguer and many others. I have also benefited greatly from the company and conversations I have engaged in with peers at vision and graphics conferences and have learned much from the older generation of researchers in computational imaging. I consider myself fortunate to have access to such a talented group of people, and to work within such a vibrant and stimulating community.

My advisors and mentors have been instrumental in both my success as a Ph.D. student, and in the personal growth I have made over the last five years. Shree Nayar has been an continuous source of motivation, inspiration, and admiration. I have been fortunate to have a Ph.D. advisor whose company I take pleasure in, whose advice I can confide in, whose criticisms I can learn from, and whose compliments I can mark as great achievements. Peter Belhumeur has been a great pleasure to get to know on a personal and professional level. I have deeply enjoyed our many discussions about computational photography and consider him a trusted source for good advice. Ravi Athale has shown great enthusiasm for my work, and I am flattered to receive support from someone with such deep roots in computational imaging. As budding professors and researchers still senior to me, I have taken great stock in the discussions and advice from Ashok Veeraraghavan, Jinwei Gu, and Mohit Gupta. Watching them grow as scientists has helped me find direction as I pursue my own career.

Finally, I take great pleasure in acknowledging the contribution from my friends and family. My parents, siblings, and close friends have always been extremely supportive of my work. I am grateful to receive so much encouragement. I have the deepest gratitude for my wife Stephanie, who I sometimes feel has more faith in my abilities than I do. I would be half the person I am today if it weren't for her belief in me and the strength it gives me. Lastly I would like to give appreciation for my daughter Asha, one year old at the time I write this. Asha's gift to me has been greatest of all, for she has taught me how much I can achieve when I pursue my goals with strength, compassion, conviction, and mindfulness.

To Asha. My spirits rise every time I see you smile.

Chapter 1

Introduction

At a fundamental level, all computer vision research is centered around measuring the visual world. We use image sensors to measure the brightness of scenes, and we use these measurements to infer radiometric and geometric properties. As humans, we organize this visual information to build on our understanding of the visual world, and a great deal of computer vision research is focused on extending this capability to machines. In this thesis we focus on the low-level mechanisms underlying the process of image formation, with the goal of developing novel sensing techniques that will better assist in machine driven image understanding. There are two main components of image formation: 1) The optical devices that condition the light as it propagates from the scene towards the optical sensor. 2) The optical sensor that converts the light energy into a measurable signal. Here, we focus primarily on the geometric properties of light, so the means of conditioning are reflection, refraction, transmission and absorption. We are now in the age of the digital camera, and so we focus on the use of digital image sensors such as CMOS and CCD sensors. The choice of optical conditioning and sensing can have a dramatic effect on the information that is captured. A good choice of optical conditioning requires careful consideration about what information content in the scene is most valuable. The digital sensor is a highly complex electrical system, wrought with several sources of uncertainty that corrupt captured images with noise and limit the performance of the imaging system. A thorough treatment of the imaging process jointly considers the optical conditioning and digital sensing together.

For centuries, the human visual system has been a model for conventional cameras.

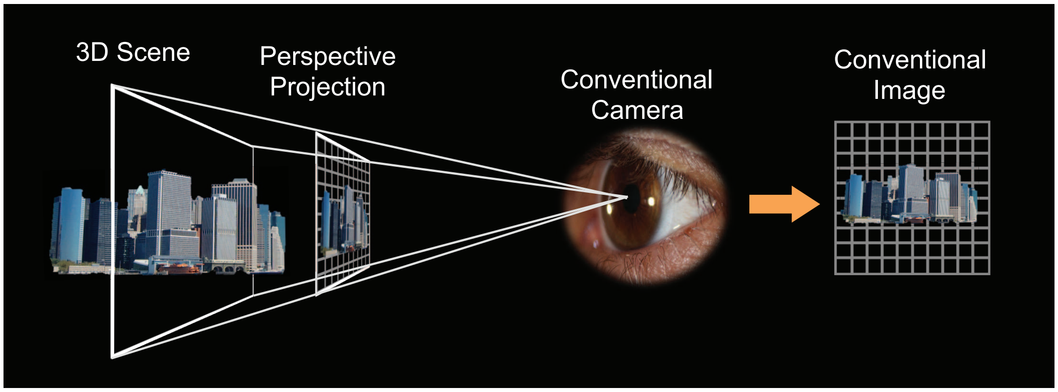


Figure 1.1: Conventional cameras map 3D scene points onto a 2D sensor via perspective projection, mimicking the human eye.

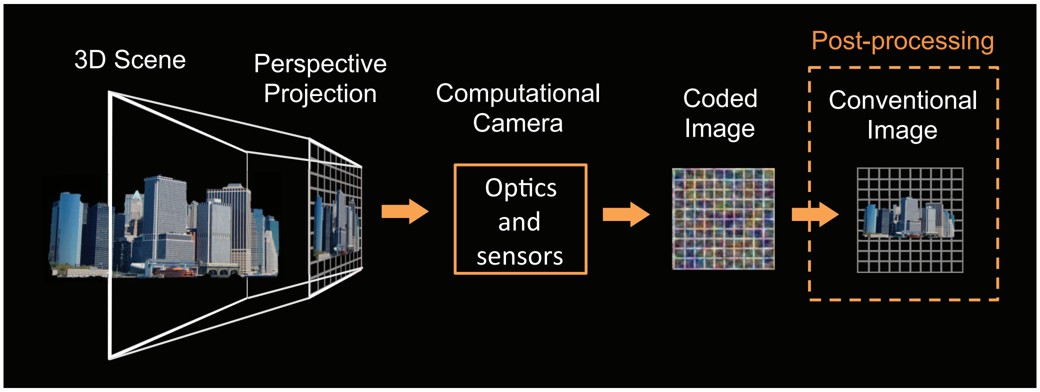


Figure 1.2: Computational cameras include a decoding step as part of the imaging pipeline. A conventional image is recovered offline via signal processing.

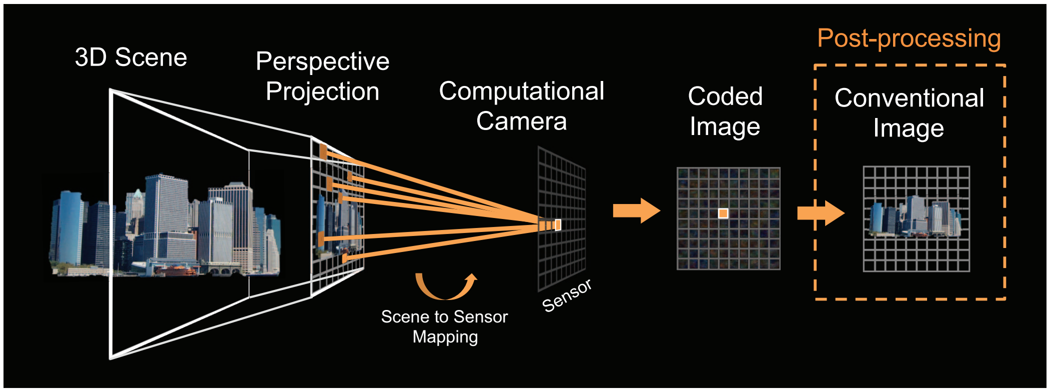


Figure 1.3: In many computational imaging systems, multiple scene points are mapped to the same pixel, which can increase optical efficiency.

Conventional cameras use perspective projection to form a two-dimensional irradiance pattern from the inherently three-dimensional distribution of light intensity (see Figure 1.1). Conventional cameras have the advantage that they produce images that can be *directly interpreted by humans* because, in many cases, they mimic the images produced by our own eyes.

The core idea of a computational imaging system is to utilize a clever combination of optics and sensors to optically encode scene information (see Figure 1.2). What is actually captured by the sensor may not be anything like the images that we are used to seeing. In many cases, there is a conventional image embedded within the captured image that can be recovered computationally. Part of the imaging pipeline is a step where the captured image is decoded offline via signal processing. Computational imaging systems may employ a many-to-one mapping between scene and pixel coordinates (see Figure 1.3), a phenomena known as image blur. These systems can increase optical efficiency because the sampling basis has a much larger support, and much more energy is captured per pixel. This type of computational imaging system is studied extensively in this thesis.

There are two main reasons why we use computational imaging systems. The first is that they offer increased functionality relative to a conventional imaging system. The increase in functionality translates to the ability to capture new types of visual information. There is a whole plethora of functions that computational cameras enable which are not accessible via conventional cameras – including depth estimation, digital refocusing, digital perspective adjustment, multispectral capture, motion blur removal, and defocus blur removal. However, new functionality is not the only reason we use computational cameras. The second reason is that they can offer a performance advantage relative to a conventional imaging system, which translates directly into greater fidelity in measurement and robustness to noise. When computational cameras increase optical efficiency, they increase the strength of captured signals, and often times this can lead to an increase in performance.

In this thesis we look at the design and implementation of a number of different computational imaging systems. We start by looking at the problem of defocus blur. Defocus blur is depth-dependent blur that removes important scene details. For conventional cameras, the only way to remove defocus blur is to stop down the lens aperture. In Chapter 2, we introduce the Diffusion Coding technique for computationally extending Depth Of Field (DOF). This technique can recover details that would otherwise be lost due to defocus blur, without stopping down the aperture. This is done by placing a custom optical diffuser in the aperture of the lens that codes the blur in a manner that is invertible via post-processing.

In Chapter 3 we approach the problem of removing defocus blur from a different angle. We introduce the spectral focal sweep technique, which takes advantage of chromatic aberrations in the camera lens to computationally extend DOF. The lens used in a spectral focal sweep camera is actually simpler than for a conventional camera, and it is this simplicity that we take advantage of to code defocus blur. The chromatic aberrations serve the same purpose as the diffuser in diffusion coding: to code the defocus blur in a way that can be inverted via post-processing.

In Chapter 4 we switch over from talking about defocus blur to talking about the blur caused by geometric aberrations in lenses with imperfect focus. We introduce a gigapixel computational camera that takes advantage of geometric aberrations to create a very high resolution camera with a very compact form factor, and very simple optics. As in diffusion coding and spectral focal sweep, the geometric aberrations code the image blur in a way that is invertible, however, this time the purpose of the coding is to simplify the optical system instead of extending DOF.

To understand how these imaging systems operate and what benefits they afford, let us go back and formalize the notion of a computational camera. For this, we first describe the plenoptic function, a fundamental concept in imaging.

1.1 Plenoptic Function

Digital cameras map visual information to digital numbers that can be processed by computers. But exactly what visual information is measurable? Adelson coined the term “Plenoptic Function” to encompass the set of all measurable visual information [Adelson and Bergen, 1991]. The plenoptic function is a complete description according to the geometric optics model of light. One parameterization of the plenoptic function is

$$P(x, y, \lambda, t, u, v, z), \tag{1.1}$$

where (x, y) are 2D spatial coordinates on the sensor, λ is the wavelength of light, t is time, and (u, v, z) are 3-D spatial coordinates of the aperture of the optical system. Together, these variables are the plenoptic coordinates. We usually refer to the 2D aperture coordinates (u, v) as angular coordinates because they determine the angle that rays are incident on the detector surface.

The plenoptic function essentially measures the radiance per unit wavelength at every 3-D spatial location. With the full plenoptic function, you would be able to watch a multi-spectral movie showing any scene from any location on earth at any point in time. However, we do not measure the plenoptic function directly. Our optical sensors measure optical energy converted to a voltage differential. Sensors average away information because they integrate over space, time, angle, and wavelength.

We measure the plenoptic function indirectly through a plenoptic *sampling basis* [Ihrke *et al.*, 2010]. Formally, the sampling basis are a set of M sampling functions $s_i(x, y, \lambda, t, u, v, z)$. Defining the vector valued plenoptic coordinate $\mathbf{p} = (x, y, \lambda, t, u, v, z)$, the i^{th} plenoptic sample g_i is then given by the inner product

$$\begin{aligned} g_i &= \langle P(\mathbf{p})\,,\, s_i(\mathbf{p})\rangle \qquad (1.2)\\ &= \int_{\Omega_p} P(\mathbf{p}) s_i(\mathbf{p}) d\mathbf{p}, \qquad (1.3) \end{aligned}$$

where Ω_p is the entire plenoptic domain. Typically the plenoptic basis is orthogonal so that $\langle s_i(\mathbf{p})\,,\, s_j(\mathbf{p})\rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. Due to physical constraints, the sampling functions have finite support in each of the plenoptic coordinates, and, in many cases, the sampling bases are separable. Take, for example, a camera system located at depth z_0 with aperture size $(\Delta u, \Delta v)$, with a 1D sensor that has a pixel size and spacing $(\Delta x, \Delta y)$, collecting light uniformly over the wavelength range $\Delta\lambda$, with an exposure time of Δt. The sampling basis for this camera is

$$s_i(\mathbf{p}) = \delta(z_0) \sqcap \left(\frac{x + i\Delta x}{\Delta x}\right) \sqcap \left(\frac{y}{\Delta y}\right) \sqcap \left(\frac{u}{\Delta u}\right) \sqcap \left(\frac{v}{\Delta v}\right) \sqcap \left(\frac{\lambda}{\Delta \lambda}\right) \sqcap \left(\frac{t}{\Delta t}\right), \tag{1.4}$$

where $\sqcap(x)$ is the box function,

$$\sqcap(x) = \begin{cases} 1 & \text{if } |x| < \frac{1}{2} \\ \\ 0 & \text{otherwise} \end{cases} \tag{1.5}$$

The pixel measurements y_i are eventually converted to a digital number with a limited dynamic range. There is a fixed number of bits/second that come out of a sensor that limit the total measurement bandwidth, which essentially limits the information capacity of the imaging system. Capturing the plenoptic function directly would require an enormous amount of computational resources. We therefore typically capture only slices of the plenoptic function, and we have different names for different slices. We call a 2D spatial slice (x, y) an image, a 3D spatio-temporal slice (x, y, t) video, a 3D spatio-spectral slice (x, y, λ) a multispectral volume, a 4D spatio-angular slice (x, y, u, v) a light field, and so on.

The plenoptic function is a useful theoretical tool because it encompasses the space of measurable visual information. We do not discuss it in this thesis, but the concept can be extended to include measurable properties of optical waves (e.g. the mutual coherence function [Brady, 2009]), all possible lighting conditions (e.g. light transport [Kajiya, 1986]), and so on. We do mention, however, that there is a large class of robust methods for estimating geometric and material properties that cannot be analyzed directly using the plenoptic function because they either depend explicitly on lighting conditions or wave properties of light. Examples of these techniques include structured light [Nayar *et al.*, 2006b] [Gupta *et al.*, 2009], BRDF estimation [Sun *et al.*, 2007][Matusik *et al.*, 2003], and optical coherence tomography [Brady, 2009].

1.2 What is Computational Imaging?

Conventional cameras are restricted to have a very specific type of sampling basis: the basis must consist of regularly spaced orthogonal sampling functions. Formally, the sampling basis can be written as $s_i = w(\mathbf{p} - i\mathbf{\Delta}_{x,y})$, where $\mathbf{\Delta}_{x,y}$ is the sample spacing in the x and y coordinates, and w is the sample function. The key property is that this produces a one-to-one, distance preserving map between spatial plenoptic coordinates and pixel coordinates. This allows the spatial information in the plenoptic function to be interpreted directly from pixel measurements once the scale and orientation of the camera are determined. In this way, a conventional camera produces an image that is identical to what what would have been seen by a human observer. Note that by this definition, conventional cameras can only measure spatial information, and a computational camera is the only way to measure spectral, temporal, or angular slices of the plenoptic function.

A computational camera can have much more general sampling basis. In fact, one of the core elements in designing a computational camera is the choice of sampling basis. From the computational imaging perspective, the optics and sensor form a channel that transmits visual information from the scene to the measurement made by an individual pixel. The choice of optics and sensor then determines the sampling basis, which, in turn, also determines the way that visual information is coded in the pixel measurements. From this perspective, we may choose to take advantage of any redundancies in the signals that will be transmitted by choosing our coding strategy appropriately. However, we do not have unlimited flexibility in choosing our coding strategy because we are limited by the space of realizable optical elements and devices. Beyond purely physical constraints, we are further limited by taking into account the complexity, weight, size, and cost of manufacturing optical elements. For instance, we can currently do a good job at creating arbitrary surface profiles out of a single material, but it is quite difficult to arbitrarily control material properties (i.e. index of refraction, absorption, etc.) within a 3D volume. In short, while the computational imaging perspective brings new light to the use of unconventional optics, we are currently restricted to considering the use of optical elements that do not differ too drastically from those that can be realized using current technology.

Computational cameras are allowed to have much more flexible mappings between spatial plenoptic and pixel coordinates. For instance, cameras with radial distortion are simple types of computational cameras – geometric distortions code captured images in a way that is recovered by resampling the image in post-processing. Computational cameras can also capture different slices of the plenoptic function. Another simple type of computational camera is a camera with a Bayer filter. A Bayer filter spatially multiplexes color information onto a single 2D sensor by applying a mapping that reorders plenoptic samples from a 3D wavelength-space volume to 2D spatial locations on a sensor. Recovering the 3D samples is merely a matter of permuting the captured data.

An important point to make is that the plenoptic function is a purely radiometric quantity, and is completely agnostic to the geometric properties of the world. Information about spatial relationships is embedded within different radiometric features such as texture and color. The plenoptic function does not contain any explicit information about 3D spatial relationships. All cameras projectively map 3D scene coordinates to two or fewer dimensions. Projective geometry causes information about the distance of objects from the camera to be lost. As a result, spatial relationships in a conventional image can only be measured accurately in two or fewer dimensions. Three dimensional spatial relationships can only be recovered from the plenoptic function computationally by using triangulation techniques that inherently take advantage of both angular and spatial plenoptic coordinates. For instance, stereo and depth-from-defocus (DFD) methods densely sample angular coordinates together with two or more spatial samples (i.e. translating or changing the size of the lens aperture).

1.2.1 Signal Models and Image Formation

We come back to a discussion about what information content in the scene is most valuable. Formally, we can define a *representation basis* for the class of input signals that we will be imaging. The representation basis is defined by a set of N representation functions $r_j(\mathbf{p})$, each of which is a different slice of the plenoptic function. An input signal $f(\mathbf{p})$ can be represented by a discrete set of N coefficients in this basis:

$$f(\mathbf{p}) = \sum_{j}^{N} f_j r_j(\mathbf{p}). \tag{1.6}$$

We often refer to the N representation coefficients f_j collectively as the signal $\mathbf{f}$, since $f(\mathbf{p})$ can be recovered directly from these coefficients using Equation 1.6, and we note that N

may be countable or infinite. Note Equation 1.6 allows us to write the image formation equation as a linear equation relating the vector of N signal coefficients $\mathbf{f}$ to the vector of M samples $\mathbf{g}$

$$\mathbf{g} = H\mathbf{f}, \tag{1.7}$$

$$H_{ij} = \int_{\Omega_p} r_j(\mathbf{p}) s_i(\mathbf{p}) d\mathbf{p}. \tag{1.8}$$

H is the system transfer matrix, and its conditioning tells us how well we can estimate the unknown signal when using a given sampling method. The uncertainty in the estimation is determined by the assumptions we make about the signal and the algorithm used to invert Equation 1.7.

As an example, consider the representation basis for the set of band-limited signals. Band-limited signals can be represented using a sinc basis $r_j(\mathbf{p}) = \text{sinc}(\mathbf{p} - j\boldsymbol{\Delta}_p)$, where $\boldsymbol{\Delta}_p$ is the sample spacing. When the delta sampling basis $s_i(\mathbf{p}) = \delta(\mathbf{p} - i\boldsymbol{\Delta}_p)$ is used, the sampling and representation basis are orthogonal and $\langle s_i(\mathbf{p})\,,\, r_j(\mathbf{p})\rangle = \delta_{ij}$. Then H is the identity matrix, and Equation 1.7 does not need to be inverted. This is just another way of stating the Nyquist theorem: band-limited signals can be recovered directly from delta sampled measurements.

The representation basis may make more general assumptions about the set of input signals. Whenever possible, we will choose the sampling basis so that it is orthogonal to the representation basis. For instance, we can choose the sampling basis to be the same as the representation basis, which allows us to sample features directly. However, for conventional cameras, we do not have much flexibility in choosing our sampling basis, so we are limited in terms of what features we can measure directly. This is a clear advantage of computational imaging techniques – it allows us to consider a more general representation basis, and choose a sampling basis that is tailored for the capture of specific features.

In some cases, we may have prior information about the statistics of the unknown signal $\mathbf{f}$ that can be used to reduce uncertainty in the measurement process. For instance, the Fourier coefficients of images are known to decay following a $1/\omega$ law when averaged across a large set of natural images [Weiss and Freeman, 2007][Srivastava *et al.*, 2003].

When the Fourier coefficients of a measured signal deviate from this aggregate behavior, we may choose to attribute it to uncertainty in the measurement process. We can modify our estimation algorithm to take into account this prior knowledge and use it to achieve an improved estimate for the unknown signal $\mathbf{f}$. The danger in this approach is that the observed deviation may have been the result of detecting an anomalous signal. Nevertheless, this approach will, on average, reduce the uncertainty over a large set of measurements. We use priors on the Fourier coefficients of natural images to evaluate the performance of the computational imaging techniques introduced in Chapters 2, 3, and 4.

Throughout this thesis, we assume that the number of plenoptic samples M is equal to the number of unknown representation coefficients N. Then image formation for a computational imaging system can be written as a fully determined system of equations. If the conditioning of the system is sufficient, the unknowns can be recovered via linear inversion. Under certain conditions, it is feasible to solve the system of equations when the number of unknown signal coefficients is larger than the number of measurements. Such an imaging system is referred to as compressive because the signal is more compact in the measurement basis than it is when measured directly. This topic will not be treated in this thesis, except for brief discussions in Chapters 5 and 6.

We also mention that in certain cases the captured images may not be intended for human consumption. In this case it may not be necessary to decode images at all, and algorithms can be developed to deal with encoded images directly. It is even possible to design the imaging system so that it is tailored to work efficiently with a specific algorithm. This can be useful if, for instance, the algorithm inherently transforms the data to some embedded lower dimensional space. Then the number of samples used directly by the algorithm may be less than the number of samples captured by the imaging system. In this case, the most efficient sampling scheme will make measurements directly in the lower dimensional space. This strategy will maximize the sampling efficiency, so that all captured information can be used directly by the algorithm. This technique is sometimes called "task-specific" imaging because the imaging system is closely coupled with the computational task at hand. Task-specific imaging systems have been developed for image classification tasks such as face detection and recognition [Nayar *et al.*, 2006a][Nayar *et al.*, 2004][Pal and

Neifeld, 2003][Ashok *et al.*, 2008]. Most of this thesis deals with images that are intended for human consumption, but task-specific imaging is discussed again in Chapter 6.

1.3 Functionality, Resolution, and Blur

We have broadly defined functionality as the ability to flexibly sample the radiometric and geometric properties of a scene. An important aspect of sampling is the resolution that we can sample at. For conventional imaging, choosing the sampling resolution amounts to choosing the size of the support of the sampling basis. We typically want to sample at as high resolution as possible, which would indicate that we want to choose small support. However, the choice of sampling resolution has a large impact on the amount of image blur exhibited by the imaging system.

Image blur is a result of coupling between plenoptic coordinates in the representation basis. For instance, suppose we know that objects are moving at a speed of s in direction θ and we can write our representation basis as

$$r_j(\mathbf{p}) = \sqcap\left(\frac{x - j\Delta x - st}{\Delta x}\right)\sqcap\left(\frac{t}{\Delta t}\right), \tag{1.9}$$

where we consider only a 2D space-time volume for simplicity. If we use the sampling basis from Equation 1.4, the transfer matrix H becomes

$$H_{ij} = \int_{-\infty}^{\infty} \sqcap\left(\frac{x - i\Delta x}{\Delta x}\right)\sqcap\left(\frac{x - j\Delta x}{\Delta x + s\Delta t}\right) dx \tag{1.10}$$

If the exposure duration Δt is small enough so that $s\Delta t < \Delta x$, the transfer matrix H becomes the identity matrix. However, if the exposure duration is larger, the matrix becomes a banded diagonal matrix. This matrix will be ill-conditioned, so that the signal cannot be recovered without the aid of prior information. Even with the aid of prior information, the conditioning may still be poor enough to result in a large amount of uncertainty in the recovered signal. Thus, we are left with two possible ways to ensure a robust signal measurement: either ensure that Δt is small enough, or choose a sampling basis that ensures the system transfer matrix H is well conditioned. This thesis focuses extensively on this problem. In Chapters 2, 3, and 4, we focus on the choice of sampling basis that results in a well-conditioned transfer matrix.

In Equation 1.9, there is a coupling between angular and temporal coordinates that resulted in the transfer matrix H being a blur matrix. The blur is caused by the motion of objects in the scene. We see the same type of coupling between angular and spatial coordinates for defocus blur. Then the blur is the result of objects spanning a range of depths. This type of blur is discussed in Chapters 2 and 3. We also see a coupling between angular and spatial coordinates when lenses exhibit geometric aberrations, which is discussed in Chapters 3 and 4.

1.3.1 Shift Invariant Blur and Convolution

Equation 1.7 is a general expression for the image formation of any computational technique. We have left ourselves open to the possibility that our sampling scheme is arbitrarily complex, and as a result, we must consider a system transfer matrix must have a general form. However, in many cases, the sampling scheme takes a special form that allows us to rewrite the image formation equation in simpler terms.

In the previous section, we discussed scenarios when the system transfer matrix is banded diagonal. This type of blur is unique because it is shift invariant: the amount of blur is identical for each pixel. Shift invariant blur leads to a special relationship between the measured image $\mathbf{g}$ and the input signal $\mathbf{f}$. The vector of measured values $\mathbf{g}$ are samples of an underlying continuous energy distribution that is incident on the sensor $g(x, y)$. When the blur is shift invariant, we can relate the input signal to the blurred signal in the continuous domain

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) h(x' - x, y' - y) dx dy \tag{1.11}$$

Equation 1.11 is a convolution between a shift invariant blur function and the input signal, sometimes written as $g(x, y) = h(x, y) \bigotimes f(x, y)$. The function $h(x, y)$ is referred to as the Point Spread Function (PSF) of the imaging system. If the input image is a single point, the blurred image will be equal to a shifted version of the PSF.

Convolution has the unique property that it can be represented compactly by first performing a transformation on the functions g, h, and f. We define the functions $G(\omega_x, \omega_y)$, $H(\omega_x, \omega_y)$, and $F(\omega_x, \omega_y)$ as the Fourier transform of the functions g, h, and f, respectively. The coordinates (ω_x, ω_y) are spatial frequency coordinates. In the Fourier domain, Equation 1.11 can be written as a multiplication.

$$G(\omega_x, \omega_y) = F(\omega_x, \omega_y) \cdot H(\omega_x, \omega_y) \tag{1.12}$$

The function H is referred to the Optical Transfer Function (OTF), and its modulus is referred to as the Modulation Transfer Function (MTF). The OTF and MTF indicates the amount that different frequencies are suppressed by the imaging system. For imaging systems, H is usually a low-pass filter. Note that Equation 1.12 gives a simple way to solve for the unknown signal $f(x, y)$. The Fourier transform of the signal can be found as

$$F(\omega_x, \omega_y) = \frac{G(\omega_x, \omega_y)}{H(\omega_x, \omega_y)}, \tag{1.13}$$

and then an inverse Fourier transform can be applied to recover the signal. This process is referred to as deblurring the captured image $g(x, y)$. The process of deblurring is complicated by two factors. The first is the possibility of zero values in the OTF that will result in incorrect values calculated in Equation 4.19. The second complication is the presence of noise in the imaging system, which prevents the image $f(x, y)$ from being calculated exactly. In this case, Equation 4.19 will not give the best estimate, and other deblurring techniques should be used instead.

Chapters 2 and 3 analyze the shift invariant blur caused by defocus. Chapter 4 analyzes shift invariant blur caused by geometric aberrations. Different techniques for deblurring images are used throughout this thesis. In some cases, we deblur images directly using Equation 4.19. In other cases, deblurring is done assuming some structure in the Fourier transform of the signal $F(\omega_x, \omega_y)$, as discussed in Section 1.2.1. In other cases, different assumptions are made about the signal to assists in robust estimation of the unknown image. In Chapter 5, we return to the form of generalized multiplexing expressed by Equation 1.7, and we analyze the performance of both general and shift invariant transfer functions within a unified framework.

1.4 Tradeoffs in Imaging

According to the computational imaging paradigm, we jointly consider the optics and sensor as an information channel that transmits information about the plenoptic function. There are physical limitations on this channel that prevent us from achieving an arbitrarily high information capacity. We seek to capture some information about the plenoptic function, be it angular, spatial, wavelength, or temporal information, but we are limited in how we can capture this information. We are forced to make tradeoffs in how we capture data. This thesis discusses five main areas where we are forced to make trade-offs when designing computational imaging systems: plenoptic resolution, efficiency vs. functionality, best case vs. average case performance, resolution vs. scale, and performance vs. complexity.

1.4.1 Plenoptic Resolution

Digital imaging sensors are highly parallel sensing mechanisms. They can sample light energy at millions of different spatial locations within a fraction of a section. Each sample is converted to a digital number with a fixed amount of precision. Ultimately the information capacity of the sensor is determined by the number of bits/second that can be shuffled around and passed on for further digital processing. Because our sensors have a limited bandwidth, we have a fixed number of samples that we can distribute among plenoptic coordinates, and a fixed amount of dynamic range that we can represent each sample with. So there is a trade-off in sampling resolution between space, time, and so on. The same trade-off exists between sampling resolution and dynamic range.

Ultimately we need to map our plenoptic samples to spatio-temporal information captured by a 2D sensor (see Figure 1.4). Methods can be divided into techniques that employ spatial multiplexing to capture all the information in a single frame, and methods that employ temporal multiplexing and therefore require multi-frame capture. Examples of the former include the use of Bayer filters, assorted pixels for High Dynamic Range (HDR)

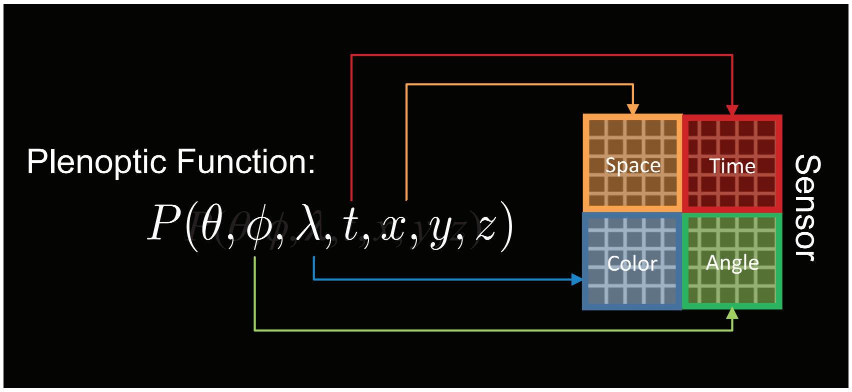


Figure 1.4: One of the main tradeoffs faced in imaging. Because our sensor has a limited bandwidth, we have a fixed number of samples that we can distribute among plenoptic coordinates. So there is a tradeoff in sampling resolution between space, time, angle and wavelength.

[Nayar and Mitsunaga, 2000] and multispectral imaging [Narasimhan and Nayar, 2005], light field capture [Adelson and Wang, 1992] [Ng *et al.*, 2005] [Veeraraghavan *et al.*, 2007] [Lanman *et al.*, 2008], and compressive video capture [Hitomi *et al.*, 2011] [Reddy *et al.*, 2011]. Examples of the latter include sequential HDR capture [Debevec and Malik, 1997] [Hasinoff *et al.*, 2010], panoramic cameras [Wilburn *et al.*, 2005] [Nomura *et al.*, 2007], superresolution [Ben-Ezra *et al.*, 2004] [Ben-Ezra *et al.*, 2005], sequential multispectral capture [Chakrabarti and Zickler, 2011] [Berns *et al.*, 2005], and time-multiplexed light field capture [Liang *et al.*, 2008].

1.4.2 Efficiency vs. Functionality

Conventional cameras typically decrease in functionality as they increase in efficiency. For instance, smaller pixels sample at higher spatial resolution, but collect less light. Narrow bandwidth wavelength filters sample at higher spectral resolution, but are less efficient as a result. Some computational imaging techniques aim to increase resolution without sacrificing efficiency. For instance, superresolution techniques recover images with small pixels from images captured with larger pixels [Ben-Ezra *et al.*, 2004] [Ben-Ezra *et al.*, 2005]. Hadamard spectroscopy recovers narrow band spectral samples from a set of highly

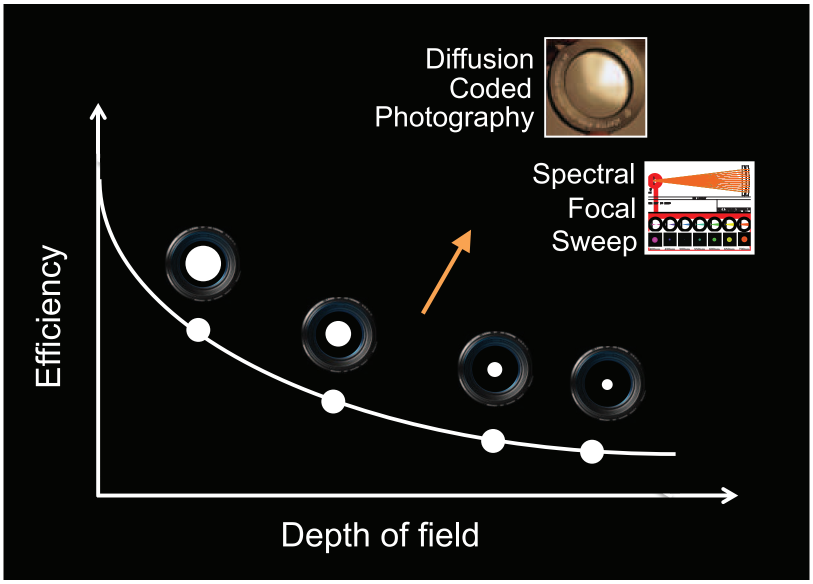


Figure 1.5: The tradeoff between optical efficiency and functionality for large DOF cameras. Conventional cameras decrease in DOF as they increase in efficiency. Computational techniques to extend DOF, such as Spectral Focal Sweep (see Chapter 3) and Diffusion Coding (see Chapter 2) increase efficiency without sacrificing DOF.

efficient spectral filters [Harwit and Sloane, 1979] [Hanley *et al.*, 1999].

We see the same tradeoff between functionality and efficiency when dealing with image blur. For a conventional imaging system, defocus causes blur that is depth dependent. The range of depths that produce defocus blur smaller than a pixel is referred to as the Depth Of Field (DOF) of an imaging system. Defocus blur increases with increasing aperture size, causing a decrease in DOF. In other words, conventional cameras lie on a curve in an Efficiency vs. DOF trade-off space, as seen in Figure 1.5). The Diffusion Coding technique introduced in Chapter 2 and the Spectral Focal Sweep technique introduced in Chapter 3 are examples of Extended DOF (EDOF) techniques. EDOF techniques use computations to simultaneously achieve high efficiency and a large DOF.

We see a similar trade-off between efficiency and resolution in cameras that use lenses with significant geometric aberrations (In fact, we can think of defocus blur as a specific type of geometric aberration). All aberrations produce blur that depends on the size of the

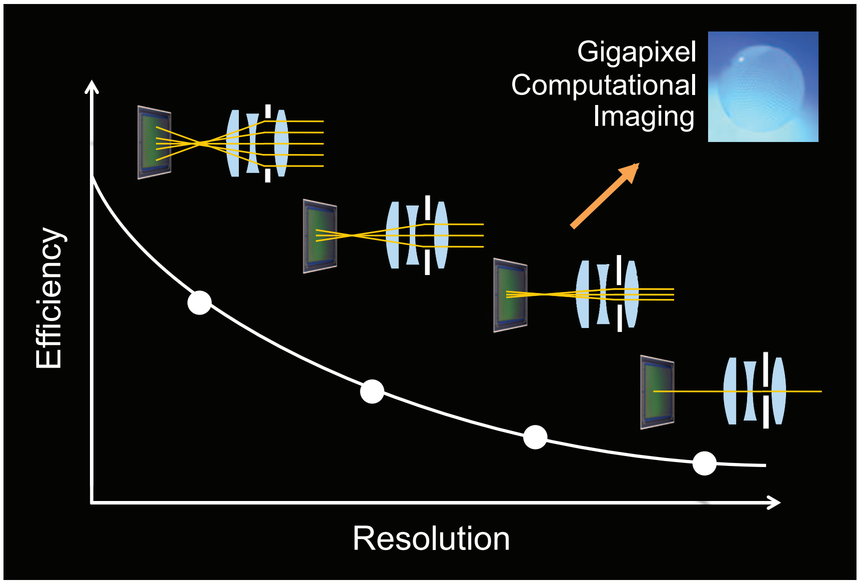


Figure 1.6: The tradeoff between optical efficiency and functionality for high resolution cameras. The resolution of conventional cameras which exhibit geometric aberrations decrease as efficiency increases. The Gigapixel Computational Camera introduced in Chapter 3 increases efficiency without sacrificing resolution.

aperture. The size of the blur limits the resolution of images created by the lens. We can always decrease the size of the blur, and hence increase resolution, by decreasing our aperture size. However, decreasing our aperture size decreases the amount of light collected by the camera. The gigapixel camera introduced in Chapter 4 uses a computational approach to remove blur, and consequently is able to maintain high efficiency at high resolutions (see Figure 1.6).

1.4.3 Best vs. Average Performance

Often we are faced with a dilemma where we want to optimize the performance over a given domain, but there are some constraints that do not allow us to simultaneously maximize average and best case performance. The dilemma is that one one hand, we want performance to be as large as possible, but we want to ensure that performance does not vary significantly over the domain. We are forced to make a tradeoff between best case and average performance. This is the case for the EDOF techniques introduced in Chapters 2

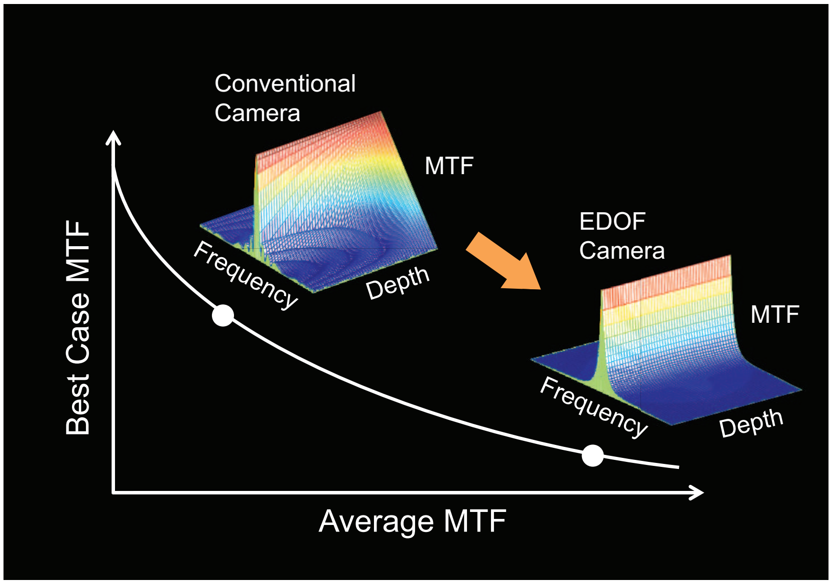


Figure 1.7: EDOF cameras sacrifice best case performance for average case performance. The performance is measured as the MTF of the camera system as a function of depth.

and 3, where the domain of interest is the range of object depths in the scene.

Here, we measure performance in terms of the MTF of the imaging system, which relates directly to the performance of the computational technique. For a conventional camera, the MTF reaches the ideal maximum when objects are located in the focal plane. However, the MTF decreases rapidly when objects are located away from the focal plane. A large variation in the MTF as a function of depth translates to a poor average performance. For an EDOF camera, the MTF does not reach the ideal maximum when objects are located at the focal plane, but the MTF remains constant at other depths, and the average performance is improved (see Figure 1.7).

Ideally we would like the best and average performance to be the same. Then, we could achieve the same performance for an EDOF system as for a camera at best focus. Ultimately, we are forced to make a trade-off due to physical constraints in the imaging system. This means that we cannot create an EDOF camera with the same performance as a conventional camera at best focus – we have to sacrifice best case performance to improve average performance.

This EDOF example demonstrates the trade-off between creating an MTF that is both

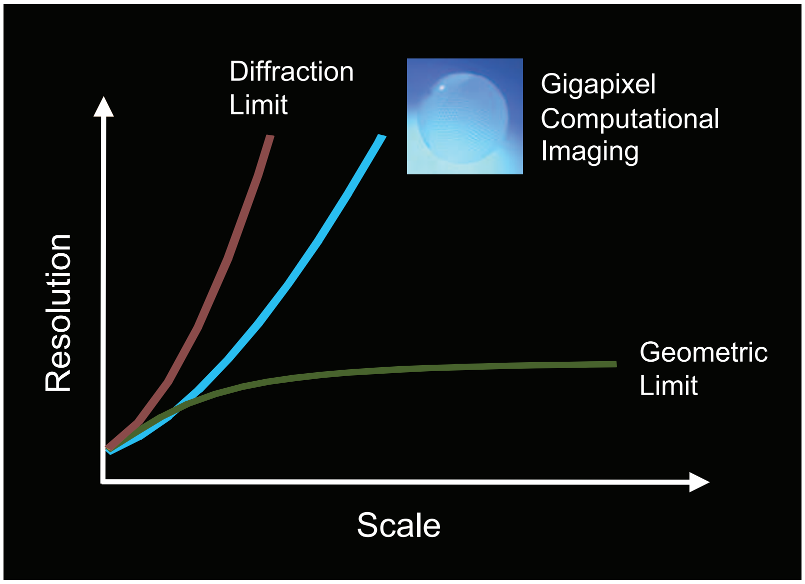


Figure 1.8: Resolution scales rapidly with camera size for ideal diffraction limited lenses. However, in practice, resolution reaches a plateau due to geometric aberrations. The Gigapixel Computational Camera introduced in Chapter 3 breaks the aberration limit so that resolution continues to increase with camera size, despite the presence of geometric aberrations.

maximal and invariant to depth. This example relates to the problem of removing defocus blur, but we see the same trade-off for systems that exhibit motion blur. An EDOF camera is designed to create a depth independent blur that can be removed computationally. Motion invariant cameras create motion invariant blur that can be removed computationally. Chapter 5 discusses performance limits for computational cameras that are invariant to blur.

1.4.4 Resolution vs. Scale

The resolution of a camera system depends on both the amount of blur caused by the optics, and the size of pixels in the sensor. Since the optical resolution is the limiting factor, it usually makes little sense to use pixel sizes greater than the optical blur. The total number of resolvable points of the camera then becomes the optical blur size divided by the size of our sensor. We can usually resolve more points when we uniformly scale up our camera, so

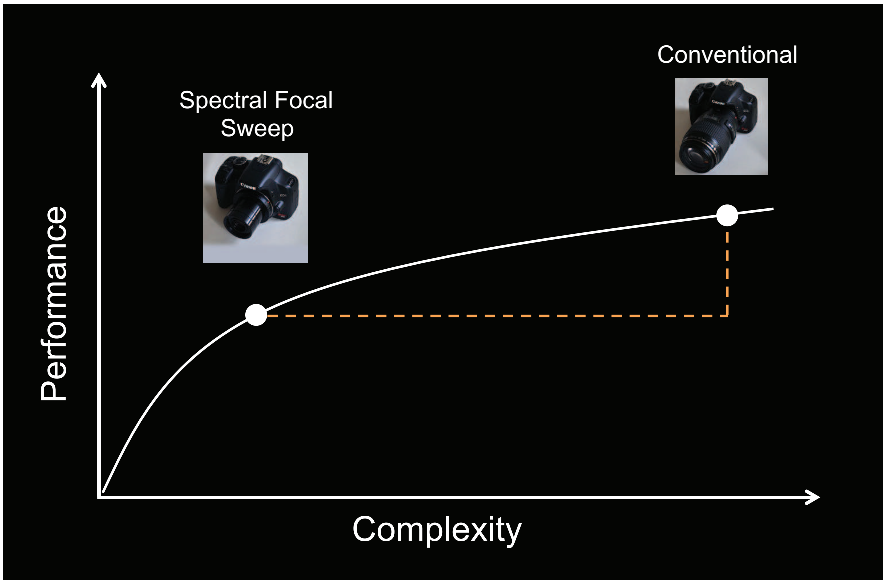


Figure 1.9: Performance vs. Complexity for the Spectral Focal Sweep camera (see Chapter 3). A conventional camera achieves higher performance than the Spectral Focal Sweep camera, but at the cost of a significant increase in complexity.

that the sensor size increases and the FOV and F/# remain fixed.

In the ideal case, blur is only caused by diffraction from the lens aperture, is independent of scale, and resolution scales rapidly with camera size (see Figure 1.8). However, in practice, lenses exhibit geometric aberrations that determine the blur size of the lens. When a lens exhibits geometric aberrations, these aberrations begin to dominate diffraction as the scale increases, causing resolution to reach a plateau. The Gigapixel Computational Camera introduced in Chapter 3 breaks the aberration limit so that resolution continues to increase with camera size, despite the presence of geometric aberrations.

1.4.5 Performance vs. Complexity

From a practical point of view, there are cost factors in building a camera, for instance the size and weight, the power consumption, the number of lenses, and so on. There is a tradeoff between the performance we can achieve and the cost we are willing to accept. For instance, in Chapter 3, we intentionally use a lens which exhibits chromatic aberrations to extend DOF. Because the lens is uncorrected, it is much less complex than a conventional lens (see

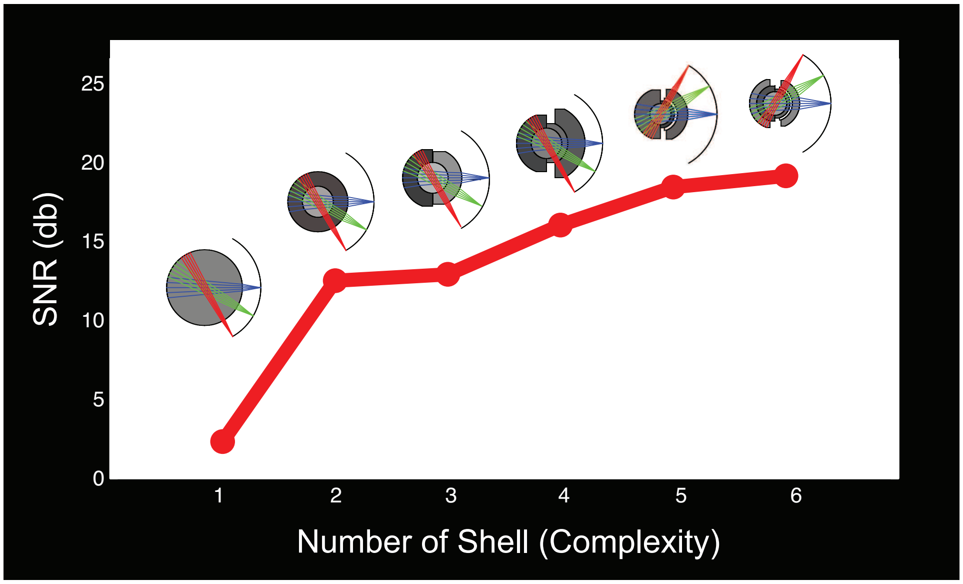


Figure 1.10: The performance of computational cameras with spherical optics as a function of lens complexity. As the complexity increases from left to right, more spherical shells are used in the lens, and the performance increases.

Figure 1.9). The DOF is increased with the uncorrected lens, but the best case performance decreases as a result, as discussed in Section 1.4.3. In this case, we see only a relatively small decrease in performance resulting from a relatively large decrease in complexity. The loss in performance may be acceptable if the cost in manufacturing lenses with increased complexity is significant.

We also see a trade-off between performance and complexity in Chapter 4, where we discuss the performance of computational cameras with spherical optics. Figure 1.10 shows that, as the complexity of spherical lenses increases from left to right, the performance of the computational camera increases. However, the increase in performance is sub-linear, so there is less performance benefit with increasing complexity. Depending on manufacturing, tolerancing and alignment considerations, the small performance advantage offered by lenses with large complexity may not warrant the dramatic increase in cost.

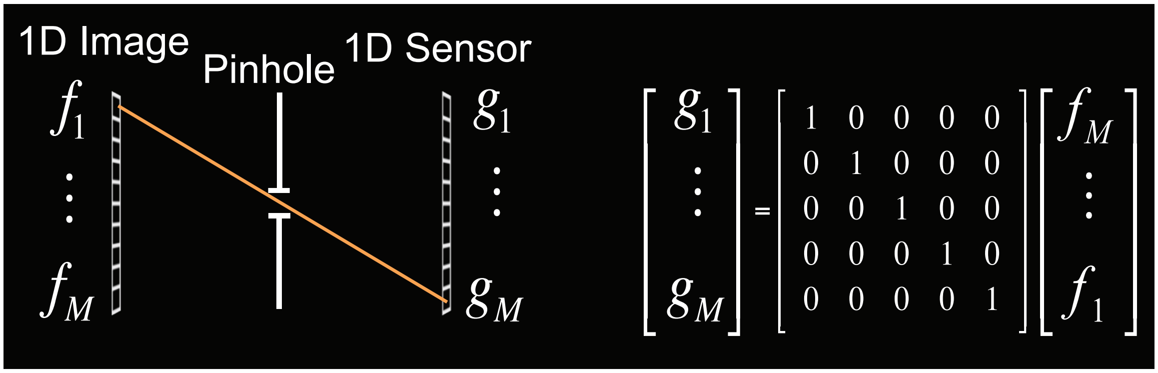


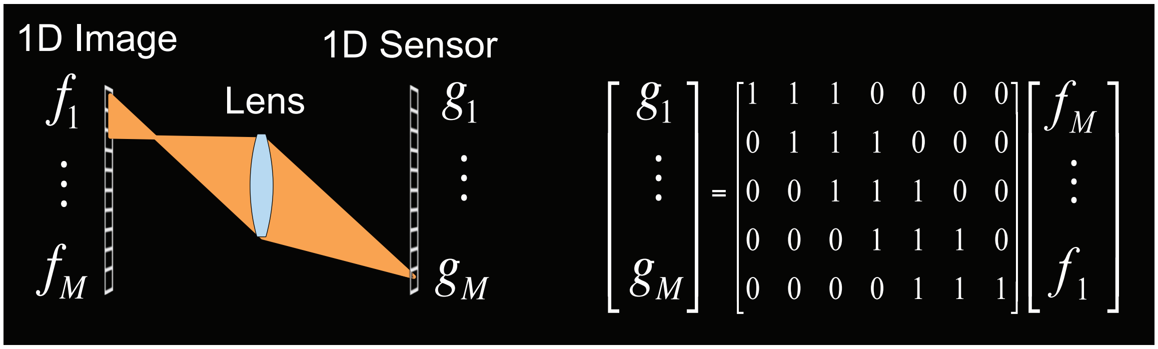
Figure 1.11: A pinhole camera exhibits no defocus blur and produces a system transfer function that is an identity matrix.

1.5 Performance Limits for Computational Imaging

In Section 1.3, we discussed how computational cameras capture images encoded by the system transfer matrix, and how image blur can be removed by inverting Equation 1.7, or in the case of shift-invariant blur, Equation 4.19. Computational cameras allow blur to be removed, and at the same time maintain high optical efficiency. However, we can always remove blur by using a conventional camera that is less optical efficiency (i.e. we can reduce exposure time for motion blur, or reduce aperture size for defocus blur). Therefore, when we evaluating the performance of a computational camera, we need to compare against the performance of a conventional camera.

As an example, consider the problem of defocus blur. A pinhole camera exhibits no defocus blur, and thus the system transfer function is an identity matrix (see Figure 1.11). A pinhole camera is extremely inefficient because it has a very small aperture through which light is allowed to pass before hitting the sensor. The less optically efficient the imaging system, the weaker the signal that is captured by the sensor. Because we want our signal to be as strong as possible, we may consider opening up our aperture to collect more light. However, defocus causes a coupling between spatial and angular coordinates that results in a transfer matrix that is banded diagonal (see Figure 1.12).

The system transfer matrix is no longer an identity matrix, and there is no longer a one-to-one mapping between sample and signal coefficients. We are left with two choices.



$$\begin{bmatrix} g_1 \\ \vdots \\ g_M \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_M \\ \vdots \\ f_1 \end{bmatrix}$$

Figure 1.12: Increasing the aperture size increases the efficiency, but it also produces defocus blur that results in a poorly conditioned system transfer matrix.

We remain in the conventional imaging paradigm by changing to a lower resolution signal representation. Then the mapping becomes one-to-one, but our resolution has decreased. Alternatively, we can stick with the same signal representation and adopt a computational approach. We can estimate the signal by inverting Equation 1.7. However, in this case, the system transfer matrix is ill-conditioned, and therefore the unknown signal $\mathbf{f}$ cannot be estimated from the plenoptic samples $\mathbf{g}$ without the use of prior information.

All is not lost, however, because we have the flexibility of choosing a new sampling strategy. For instance, we can "code" the aperture using a transparency pattern (see Figure 1.13). Depending on the choice of aperture pattern, this sampling strategy can produce a transfer matrix with much better conditioning [Levin *et al.*, 2007][Veeraraghavan *et al.*, 2007][Zhou and Nayar, 2009].

Both the pinhole camera and the coded aperture camera can produce an image that is free of blur, however, the coded aperture camera captures an image with much greater optical efficiency. We have a vague sense that greater optical efficiency is desirable because it increases the signal strength of captured images, but we still haven't determined concretely which technique produces better performance: the pinhole or coded aperture camera. There are two determining factors in evaluating performance: the conditioning of the transfer matrix and the noise model. When we code the the aperture, we increase the conditioning of the transfer matrix so that blur can be removed without sacrificing optical efficiency. However, depending on the noise model, an increase in efficiency may actually increase the

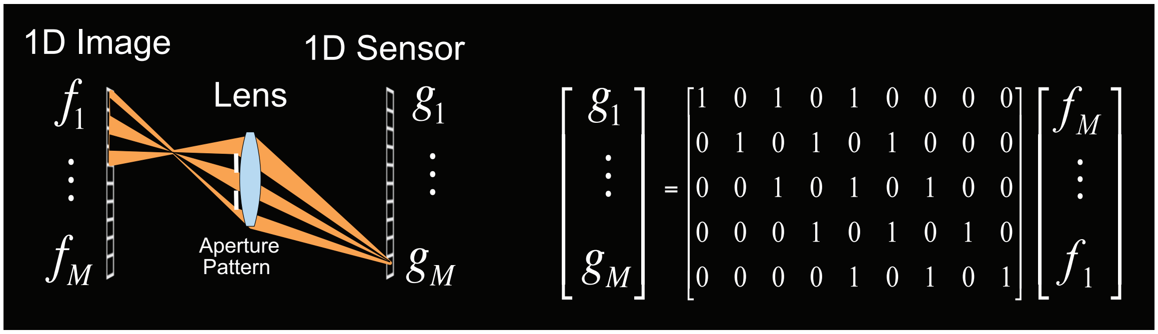


Figure 1.13: By placing a transparency pattern in the aperture of the lens, we can improve the conditioning of the transfer matrix without a significant sacrifice in efficiency.

noise level as well as increasing the signal strength. So we need to be more specific about the the noise model before we can make any concrete statements about the performance of computational cameras. In Chapter 5, we introduce a detailed noise model, and we derive bounds on the maximum performance advantage that a computational camera can have over a conventional camera. The results are somewhat surprising – we will see that an increase in optical efficiency does not always produce the boost in performance that might be expected, and that there are some concrete limits on the performance we can get out of computational cameras.

Part I

Tradeoffs in Computational Imaging

Chapter 2

Diffusion Coding

2.1 Introduction

In Chapter 1 we discussed the trade-off between efficiency and Depth Of Field (DOF). The amount of defocus blur depends on the aperture size and the distance from the focal plane. To decrease defocus blur and increase DOF, the aperture size must be decreased, reducing the signal strength of the recorded image as well. However, stopping down the lens aperture is not always an option, especially in low light conditions, because it it decreases the Signal-to-Noise Ratio (SNR) and corrupts the signal.

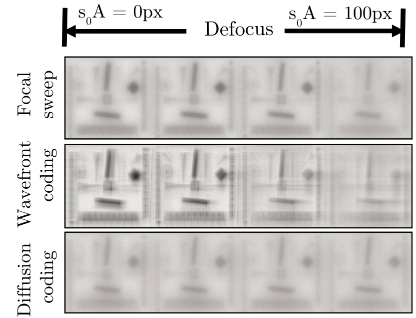
The fundamental problem with increasing the DOF of conventional cameras is that defocus blur is depth dependent. If the depths of objects in the scene are known, it is possible to remove the blur computationally. However, high precision depth estimation is error prone, and difficult (if not impossible) without the aid of additional hardware, such as that used in structured light or laser scanning systems. We are interested in simultaneously maximizing performance averaged over depth, and producing depth-invariant blur, so that we can deblur captured images without knowing depth ahead of time. The cost of maximizing average performance however, is that we must sacrifice best case performance.

Two well-studied techniques that produce a depth-invariant Point Spread Function (PSF) are wavefront coding [E. R. Dowski and Cathey, 1995], which uses a cubic phase plate, and focal sweep [Nagahara *et al.*, 2008] [Häusler, 1972], where either the object, sensor position, or lens focus setting is mechanically varied during exposure. Recently, Baek compared the degree of depth-invariance of these two techniques, and observed that focal sweep gives a near-optimal tradeoff between Modulation Transfer Function (MTF) and depth-invariance at all frequencies [Baek, 2010], while wavefront coding is only guaranteed to be optimal at a single frequency.

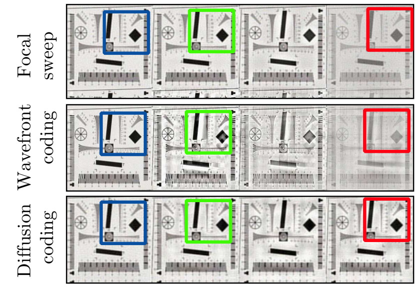
Typically, when deblurring a noisy image, a larger magnitude MTF will result in less deblurring reconstruction error. However, this is only the case if the PSF is completely depth-invariant. This consideration is of utmost importance in the context of Extended Depth Of Field (EDOF) cameras because, in practice, it is only possible to produce a PSF that is *approximately* depth-invariant, and the amount of variation determines the severity of the artifacts that are introduced in the deblurring process.

In this chapter, we introduce a new diffusion coding camera that produces near identical performance to focal sweep, but without the need for moving parts. This is achieved by using optical diffusers placed in the pupil plane, which scatter light in such a way as to produce a depth-invariantblurred image. This image can then be deblurred to create an EDOF image, just like the focal sweep cameras of [Nagahara *et al.*, 2008] [Häusler, 1972], but with without the need for moving parts. Like phase-plates, diffusers have the advantage of being almost completely non-absorptive, and thus do not sacrifice signal intensity. We coin the term *diffusion coding* to mean a camera with a diffuser placed in the pupil plane. We characterize diffusers as kernels that operate on a 4D light field propagating from a camera lens to sensor. As a result, we are able to obtain an analytical solution for the PSF of our diffusion coded camera, which is given in Section 2.4.

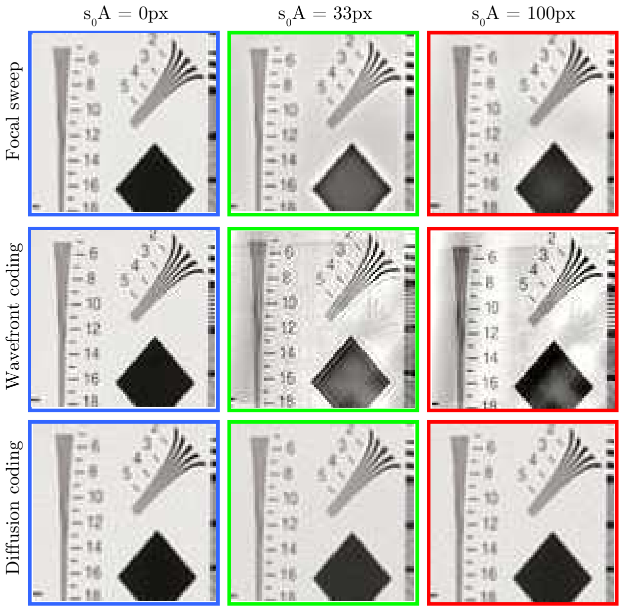
Levin et al. show that wavefront coding produces better results than focal sweep if variation in the PSF is not taken into account [Levin *et al.*, 2009]. As can be seen from Figure 2.1, wavefront coding recovers more detail than other methods for objects at the focal plane when the correct PSF is used for deblurring. However, the method also introduces noticeable artifacts for objects at different depths because the PSF varies significantly with depth. To measure the degree of depth-invariance of a camera, we compute the deblurring reconstruction error for objects at different depths. The result is shown in Figure 2.2, where a flatter curve signifies more similarity between PSFs at different depths. We note that the focal sweep camera produces a PSF that is more depth-invariant than wavefront coding,



(a) Captured images



(b) Recovered EDOF images



(c) Close-ups

Figure 2.1: Simulated image performance for three EDOF cameras. An IEEE resolution chart is placed at different depths. The aperture size A and defocus slope in light field space s_0 are chosen so that the maximum defocus blur diameter is 100 pixels. The center PSF is used for deblurring, producing the images shown in (b). Close-ups in (c) show that the sharpest image is produced by wavefront coding at the center depth ($s_0A = 0$). However, wavefront coding produces significant deblurring artifacts for defocus values as small as $s_0A = 33$ pixels, while diffusion coding produces near identical results for the entire depth range.

and furthermore that our diffusion coded camera produces near identical results to that of focal sweep. The comparison of EDOF Cameras is discussed further in Section 2.5.

We focus our attention on the use of diffusers with predefined scattering properties, and do not address the task of diffuser design. Much work has been done in recent years to develop custom diffusers with tailored scattering profiles. These diffusers are frequently used in lighting and display applications to produce uniform illumination or arbitrary beam shaping. The popularity of these diffusers has also led to much innovation in replication techniques, so that today several companies sell off-the-shelf diffusers reproduced onto plastic sheets up to 36" wide [Luminit, 2011] [RPC, 2011]. In Section 2.6, we introduce our implementation of a diffusion coded camera using a custom diffuser manufactured by RPC Photonics [RPC, 2011]. We conclude with examples of EDOF images taken with our implementation in Section 2.7.

2.2 Related Work

Optical diffusers and other random surfaces have been used to assist in a variety of imaging tasks, including super-resolution [Ashok and Neifeld, 2003][Ashok and Neifeld, 2007], lenseless imaging [Freeman *et al.*, 2006], and extended DOF [García-Guerrero *et al.*, 2007]. In this work, we focus on the task of using diffusers to extend DOF.

Several radially symmetric phase masks have been introduced to extend DOF [Chi and George, 2001] [Ojeda-Castaneda *et al.*, 2005][García-Guerrero *et al.*, 2007]. The work most similar to ours is by Garcia-Guerrero et al., who also use a radially symmetric diffuser. To design their diffuser, the authors take a completely different approach than the technique described in Section 2.6. They derive a random surface that on average produces a PSF whose value at the center is constant over a large depth range, while we derive a diffuser

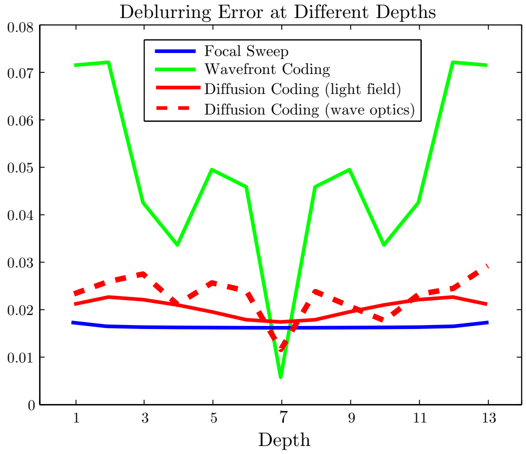


Figure 2.2: The deblurring error (based on simulations in Section 2.5) as a function of depth for three EDOF cameras. A flatter curve denotes less PSF variation. The diffusion coding curves are very similar to that of focal sweep.

whose entire PSF is approximately depth-invariant. The Garcia-Guerrero diffuser consists of annular sections of quadratic surfaces, where the width of the annulus decreases quadratically with distance from the optical axis. This design requires the feature size to decrease from the center to the edge of the diffuser. The minimum feature size is limited by the fabrication technology that is used to make the diffuser. In Section 2.6 we consider the use of laser machining technology that has a minimum spot size on the order of 10μm. The result is that the performance of one instance of the Garcia-Guerrero diffuser varies significantly from the expected performance while the diffuser we introduce in Section 2.6 performs very close to the expected performance (see Figure 2.11). This difference is discussed further in Section 2.6.

Wavefront coding was introduced by Dowski and Cathey [E. R. Dowski and Cathey, 1995], who place a cubic phase plate (CPP) in the pupil plane of a camera system. Dowski et al. show analytically that a camera with a cubic phase plate produces a PSF that is approximately invariant to defocus. Although the CPP does produce a PSF that is approximately depth-invariant, the PSF is not as invariant as the focal sweep camera or our diffusion coded camera (see Figures 2.1 and 2.2).

Focal sweep cameras produce a depth-invariant PSF by sweeping either the object [Häusler, 1972] or sensor [Nagahara *et al.*, 2008] along the optical axis during exposure. The PSFs for these techniques preserves high frequencies because each object is instantaneously in focus at one point during exposure. Focal sweep techniques require the use of moving parts and introduce limitations on the minimum exposure time.

Levin et al. compare the performance of focal sweep and wavefront coding cameras without considering the effect of depth-invariance [Levin *et al.*, 2009]. Hasinoff et al. analyzed the SNR characteristics of both focal sweep and wavefront coding cameras when multiple exposures with different focus settings are used [Hasinoff *et al.*, 2009], and Baek compared the MTF and depth-invariance of focal sweep and wavefront coding cameras [Baek, 2010].

Other works exist in the vision community which recover an EDOF image after first estimating scene depth [Levin *et al.*, 2007] [Levin *et al.*, 2009] [Zhou and Nayar, 2009]. The quality of these techniques, however, is closely coupled to the precision of depth estimation, since each region in the image is deblurred using an estimated defocus PSF.

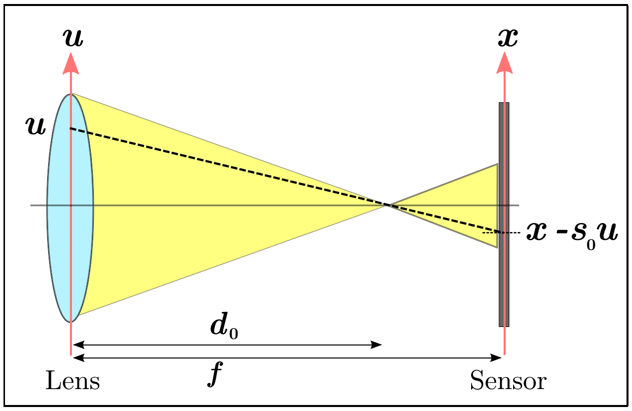


Figure 2.3: The geometry of an image point focused at a distance d_0 from the camera lens aperture. A sensor is located a distance f_l from the aperture. A ray at piercing the aperture at location u intersects the sensor at location $x - s_0 u$. where $s_0 = \frac{d_0 - f_l}{d_0}$.

We use a light field [Levoy and Hanrahan, 1996] parameterization to understand the properties of imaging systems. Several researchers have analyzed the image formation of camera systems as projections of light fields [Ng, 2005][Veeraraghavan *et al.*, 2007][Levin *et al.*, 2009]. In addition, several authors have looked at light fields in the frequency domain, including image formation and interactions between transmissive and reflective objects [Ng, 2005][Durand *et al.*, 2005][Veeraraghavan *et al.*, 2007].

2.3 Light Field Analysis

A light field $l(\mathbf{u}, \mathbf{x})$ can be used to represent the 4D set of rays propagating from an ideal lens with effective focal length (EFL) f_l to a sensor. The vector $\mathbf{u} = (u, v)$ denotes the coordinates on the u-v plane, which is coincident with the exit pupil of the lens. The vector $\mathbf{x} = (x, y)$ denotes the coordinates on the x-y plane that is coincident with the sensor. Note that this is a slightly different convention than used by Levin et al., where the x-y plane is defined in object space [Levin *et al.*, 2009]. The irradiance $g(\mathbf{x})$ observed on the sensor is simply the light field integrated over all ray angles:

$$g(\mathbf{x}) = \int\limits_{\Omega_{\mathbf{u}}} l(\mathbf{u}, \mathbf{x}) d\mathbf{u}, \tag{2.1}$$

where, $\Omega_{\mathbf{u}}$ is the domain of $\mathbf{u}$. For a scene with smooth depth variation, locally, the captured image $g(\mathbf{x})$ can be modeled as a convolution between a depth-dependent PSF kernel $h(\mathbf{x})$ and an all-in-focus image $k(\mathbf{x})$. The EDOF goal is to shape the camera PSF so that the *entire* image $f(\mathbf{x})$ can be recovered from the captured image $g(\mathbf{x})$ by deblurring with a *single* PSF $h(\mathbf{x})$. We analyze the depth-dependence of the camera PSF by considering the image produced by a unit energy point source. Consider a point source whose image comes to focus at a distance d_0 from the aperture of the lens (see Figure 2.3). Assuming a rectangular aperture of width A, the light field produced by this point is

$$l_\delta(\mathbf{u}, \mathbf{x}) = \frac{1}{A^2} \sqcap \left(\frac{\mathbf{u}}{A} \right) \delta(\mathbf{x} - s_0 \mathbf{u}), \tag{2.2}$$

where $s_0 = \frac{d_0 - f_l}{d_0}$ is the defocus slope in light field space, and $\sqcap$ is the multi-dimensional box function

$$\sqcap \left(\frac{\mathbf{x}}{w} \right) = \begin{cases} 1 & \text{if } |\mathbf{x}_i| < \frac{w}{2}, \forall i \\ \\ 0 & \text{otherwise} \end{cases} . \tag{2.3}$$

The image of this point is the camera PSF at the depth d_0, which is the familiar box shaped PSF with defocus blur width $s_0 A$:

$$h(\mathbf{x}) = \frac{1}{s_0^2 A^2} \sqcap \left(\frac{\mathbf{x}}{s_0 A} \right) . \tag{2.4}$$

We now analyze the effect of a general kernel d applied to a light field l, which represents the effect of a diffuser placed in the aperture of a camera lens. The kernel produces a new filtered light field l', from which we can derive the modified PSF h':

$$l'(\mathbf{u},\mathbf{x}) = \int\limits_{\Omega_{\mathbf{u}'}} \int\limits_{\Omega_{\mathbf{x}'}} d(\mathbf{u},\mathbf{u}',\mathbf{x},\mathbf{x}') l(\mathbf{u}',\mathbf{x}') d\mathbf{u}' d\mathbf{x}', \tag{2.5}$$

$$h'(\mathbf{x}) = \int\limits_{\Omega_{\mathbf{u}}} l'(\mathbf{u},\mathbf{x}) d\mathbf{u}, \tag{2.6}$$

where $\Omega_{\mathbf{x}}$ is the domain of $\mathbf{x}$. This approach allows us to express a large class of operations applied to a light field. For instance, consider a kernel of the form

$$d(\mathbf{u},\mathbf{u}',\mathbf{x},\mathbf{x}') = \frac{1}{w^2}\delta(\mathbf{u}-\mathbf{u}') \sqcap \left(\frac{\mathbf{x}-\mathbf{x}'}{w}\right). \tag{2.7}$$

Note that here D takes the form of a separable convolution kernel with finite support in the $\mathbf{x}$ domain. The geometric meaning of this kernel is illustrated in Figure 2.4. Each ray in the light field is blurred so that, instead of piercing the sensor at a single location, it contributes to a square of width w. In order to understand the effect of the diffuser, we compare an image $g(\mathbf{x})$ captured without the diffuser to an image $g'(\mathbf{x})$ captured with it. For this diffuser kernel, substituting Equation 2.7 into Equations 2.5 and 2.6 gives:

$$h'(\mathbf{x}) = \frac{1}{w^2} \sqcap \left(\frac{\mathbf{x}}{w}\right) \otimes h(\mathbf{x}), \tag{2.8}$$

where $\otimes$ denotes convolution. The modified PSF is simply the camera PSF blurred with a box function. Therefore, the effect of the diffuser is to blur the image that would be captured were it not present. Introducing the diffuser given by the kernel in Equation 2.7 is clearly not useful for extending depth of field since it it does not increase depth independence or preserve high frequencies in the camera PSF. We note that, in general, the kernel for any diffuser that is placed in the aperture takes the form

$$d(\mathbf{u},\mathbf{u}',\mathbf{x},\mathbf{x}') = \delta(\mathbf{u}-\mathbf{u}')k(\mathbf{u},\mathbf{x}-\mathbf{x}'), \tag{2.9}$$

where k is called the scatter function. That is, the diffuser has no effect in the $\mathbf{u}$ domain, but has the effect of a convolution in the $\mathbf{x}$ domain. For the diffuser given by Equation 2.7, the scatter function is the 2D box function $k(\mathbf{u},\mathbf{x}) = \frac{1}{w^2} \sqcap \left(\frac{\mathbf{x}}{w}\right)$.

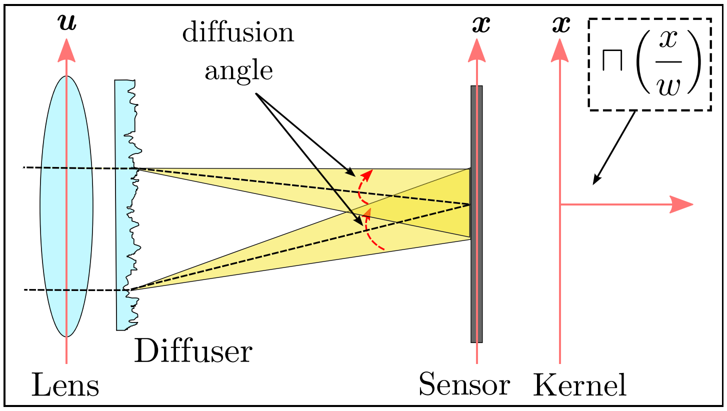


Figure 2.4: For the diffuser defined by the kernel in Equation 2.7, the diffusion angle does not vary across the aperture. Each ray is blurred so that it covers an area on the sensor determined by the the diffuser parameter w.

2.4 Radially Symmetric Light Fields

We now change from rectangular coordinates (u, v, x, y) to polar coordinates (ρ, ϕ, r, θ) using the relations $u = \rho \cos \phi$, $v = \rho \sin \phi$, $x = r \cos \theta$, and $y = r \sin \theta$. We consider a polar system where $\rho, r \in (-\infty, \infty)$ and $\theta, \phi \in (0, \pi)$ and a circular aperture with diameter A. The light field representing a unit-energy point source located at distance d_0 in this new system can be written as

$$l_\delta(\rho, r) = \frac{4}{\pi A^2} \sqcap \left(\frac{\rho}{A}\right) \frac{\delta(r - s_0\rho)}{\pi |r|}, \tag{2.10}$$

which is independent of both θ and ϕ because the source is isotropic. Note that verifying unit-energy can be carried out trivially by integrating $l_\delta(\rho, r)$ in polar coordinates (see Section A.2). Comparing the parameterizations for the light field of a point source in Equations 2.2 and 2.10, we can see that a slice of $l_\delta(\mathbf{x}, \mathbf{u})$ represents a single ray, while a slice $l(\rho, r)$ represents a 2D set of rays. In the radially symmetric parameterization, a slice of the light field represents a conic surface connecting a circle with radius ρ in the aperture plane to a circle of radius r on the sensor (see Figure 2.5).

We now consider the effect of a radially symmetric diffuser on the camera PSF. Somewhat surprisingly, a diffuser that is parameterized in these reduced 2D coordinates produces

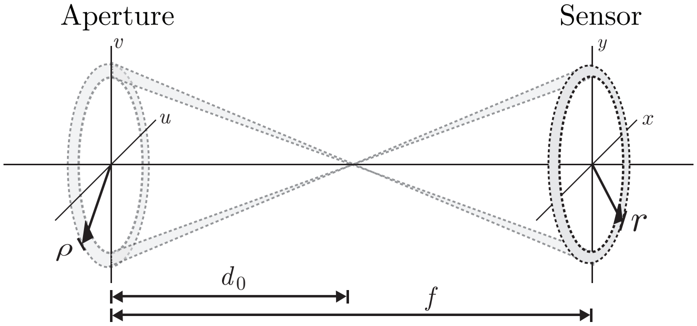


Figure 2.5: The geometry of a radially symmetric light field using reduced coordinates. The light field consists of a point source focused a distance d_0 from the lens aperture. Because the point source is on-axis and isotropic, the light field can be represented as a 2D function $l(\rho, r)$. A 2D slice of the light field $l(\rho, r)$ represents the set of rays traveling from a circle with radius ρ in the aperture plane to a circle with radius r on the sensor. This set of rays forms a conic surface.

a drastically different effect than the diffuser given by Equation 2.7. When a radially symmetric diffuser is introduced, neither the diffuser nor the lens deflects rays tangentially, and therefore we can represent the diffuser kernel and modified light field using the reduced coordinates (ρ, r). Equations 2.5 and 2.6 then become

$$l'(\rho, r) = \pi^2 \int_{\Omega_\rho} \int_{\Omega_r} d(\rho, \rho', r, r') l(\rho', r) |\rho'| d\rho' |r'| dr', \tag{2.11}$$

$$g'(r) = \pi \int_{\Omega_\rho} l'(\rho, r) |\rho| d\rho, \tag{2.12}$$

and the general form of the diffuser kernel becomes

$$d(\rho, \rho', r, r') = \frac{\delta(\rho - \rho')}{\pi|\rho'|} \frac{k(r - r', \rho)}{\pi|r|}. \tag{2.13}$$

We use the same box-shaped scattering function as we did for the diffuser kernel in Equation 2.7:

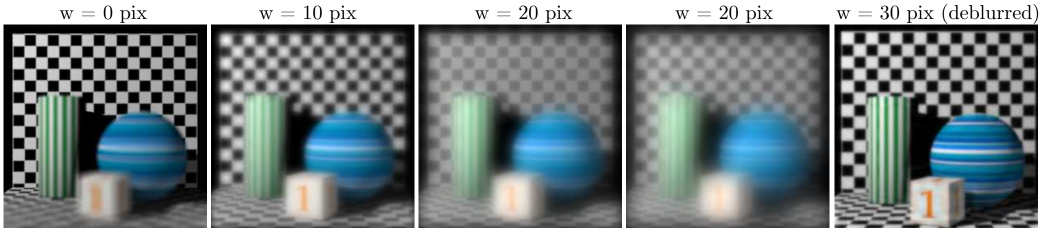


Figure 2.6: Simulated photographs taken of of a light field filtered by the diffuser kernel in Equation 2.14. The parameter w of the diffuser kernel is varied across the columns. The rightmost figure shows a deblurred diffusion coded image with a 10× increase in DOF.

$$k(r, \rho) = \frac{1}{w} \sqcap (\frac{r}{w}). \tag{2.14}$$

However, the physical interpretation of this diffuser is drastically different than for the previous diffuser. For the previous one, each ray in the light field is scattered so that it spreads across a square on the sensor. The effect of the scattering function in Equation 2.14 is illustrated in Figure 2.7. In the absence of the diffuser, light from an annulus of width $d\rho$ and radius ρ in the aperture plane projects to an annulus of width dr and radius r on the sensor. The effect of the scatter function in Equation 2.14 is to spread the light incident on the sensor so that it produces an annulus of width w instead. We can also consider the scattering from the perspective of a single ray, as illustrated by the pink and red volumes in Figure 2.7. In polar coordinates, a ray is a small annular section that travels from the aperture plane to the sensor plane, illustrated by the red volume in Figure 2.7. The pink volume illustrates the effect of the diffuser, which is to scatter a ray along a radial line of width w. We note that a box-shaped scatter function is used here for notational convenience, but we found that a Gaussian scattering function is superior for extended DOF imaging (see Figure 2.10(d)).

The light field of a point source filtered by this diffuser kernel and PSF can be shown to be (see Section A.3 for a complete derivation)

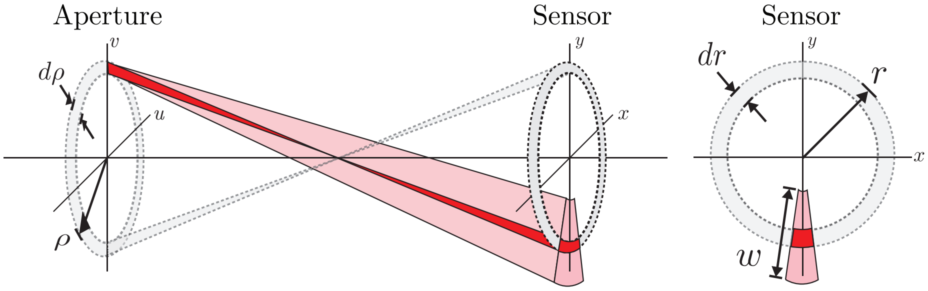


Figure 2.7: The geometry of a radially symmetric diffuser. The diffuser scatters light only in the radial direction, and has no effect in the tangential direction. A thin annulus of light is emitted from the aperture of width $d\rho$ and radius ρ. In the absence of the diffuser, the emitted light projects to an annulus on the sensor of width dr and radius r. When the diffuser is present, the width of the annulus on the sensor becomes w, the diffuser scatter width.

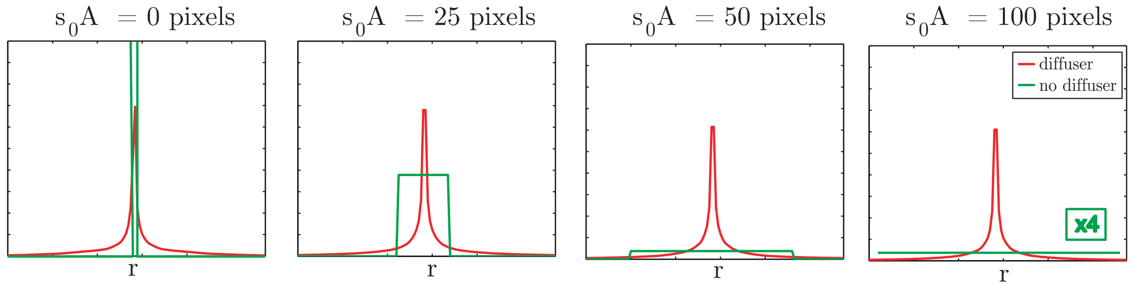
$$l'(\rho, r) = \frac{4}{\pi A^2} \sqcap\left(\frac{\rho}{A}\right) \frac{\sqcap(\frac{r - s_0\rho}{w})}{\pi w |r|}, \tag{2.15}$$

$$h'(r) = \frac{4}{\pi s_0^2 A^2} \frac{1}{w|r|} \left[\sqcap\left(\frac{r}{w}\right) \otimes \left(\sqcap\left(\frac{r}{s_0 A}\right) \cdot |r|\right)\right]. \tag{2.16}$$

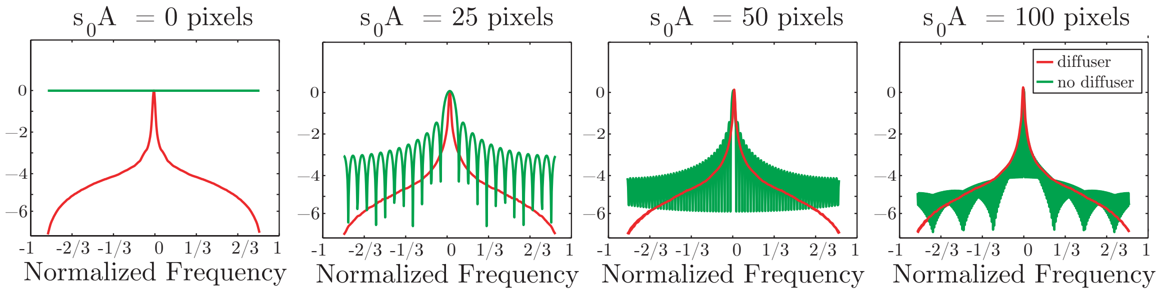
The analytic solution for the PSF is a piecewise function due to the contribution from the term in brackets, which is a convolution between the two rect functions (one weighted by $|r|$). Note that as the scattering width w is reduced to zero, the first rect (combined with $\frac{1}{w}$) approaches a delta function and the result is the familiar pillbox shaped defocus PSF. Also note that if a different scattering function is used, the first rect is simply replaced with the new function. However, the convolution term is far less significant than the $\frac{1}{|r|}$ term, whose effect dominates, resulting in a PSF which is strongly depth-independent while still maintaining a strong peak and preserving high frequencies.

The solution for the PSF may be interpreted in the following way. Please refer to Figure 2.7. Suppose we have a pillbox defocus PSF, and we want to know how a small annular region of width δr and radius r will be affected by the diffuser. Light incident on this region emanates from an annulus in the aperture, and its energy will be proportional to ρ or equivalently r/s_0. This explains the presence of the $|r|$ multiplier within the term in brackets. The term in brackets states that the energy in the PSF annulus is spread uniformly along radial lines of width w, as shown on the right hand side of Figure 2.7. The $\frac{1}{|r|}$ term in Equation 2.16 can be attributed to the fact that the energy density becomes larger for light that is scattered closer to the center of the PSF.

Figure 2.8 shows several PSF/MTF pairs for a camera with and without the diffuser given by Equation 2.16. The defocus blur diameter s_0A varies from 0 to 100 pixels. The scatter function of Equation 2.14 is a Gaussian instead of a box function, and the diffuser parameter w (the variance of the gaussian) is chosen so that $w = 100$ pixels. Note that when the diffuser is present, there is little variation with depth for either the PSF or MTF. Introducing the diffuser also eliminates the zero crossings in the MTF. For smaller defocus values, the diffuser suppresses high frequencies in the MTF. However, because the



(a) The PSF as a function of depth



(b) The MTF as a function of depth (log scale)

Figure 2.8: PSF plots (top) and MTF (bottom) plots for a camera with (red) and without (green) the diffuser kernel defined in Equation 2.14. The defocus blur diameter s_0A is varied across columns from 0 to 100 pixels, and the diffuser parameter $w = 100$ pixels. Both the PSF and MTF exhibit negligible variation when the diffuser is present.

diffuser MTF does not vary significantly with depth, high frequencies can be recovered via deconvolution. Figure 2.6 shows a simulated light field filtered by the radially symmetric diffuser given by Equation 2.14. On the far right of the figure, we show a high contrast, extended depth of field image that is recovered after deconvolution is applied.

2.5 Comparison between EDOF Cameras

All EDOF cameras sacrifice MTF response at high frequencies in order to achieve depth-invariance. High frequencies in captured images are recovered via deconvolution, but this process also amplifies sensor noise which degrades the recovered image. In addition, any variation in the PSF/MTF as a function of depth will result in deblurring artifacts due to a mismatch between the actual PSF and the PSF used for deblurring. The quality of an edof camera can be represented by the deblurring reconstruction error, which takes into account the camera MTF, the degree of depth-invariance of the PSF/MTF, and sensor noise. To calculate the deblurring error we compute the Mean Squared Error (MSE) of deblurred images. The MSE is given by the $L2$ norm on the difference between the ground truth (focused) image and the captured image deblurred by a PSF $h_d(x, y)$. The captured image is the ground truth image $f(x, y)$ blurred by a PSF $h_b(x, y)$ plus noise $\eta(x, y)$.

$$MSE(d) = \left|\left|(f(x,y) \otimes h_b(x,y) + \eta(x,y)) \otimes h_d^{-1}(x,y) - f(x,y)\right|\right|^2. \qquad (2.17)$$

This measure takes into account the camera MTF, since it includes the term $\eta(x, y) \otimes h_d^{-1}(x, y)$, which represents the amplification of sensor noise due to small MTF values. In addition, the measure takes into account the degree of depth-invariance of the camera PSF/MTF because it includes the term $f(x, y) - (f(x, y) \otimes h_b(x, y)) \otimes h_d(x, y)^{-1}$, which is the difference between a ground truth image and the same image blurred by one PSF and then deblurred by another.

To evaluate the performance of an EDOF camera, we calculate the deblurring error over a range of depths. If an EDOF camera performs well, it will have a small deblurring error over all depths. For each camera, we calculated the camera PSF at a variety of discrete depths and used this as the blurring PSF $h_b(x, y)$. For the deblurring PSF $h_d(x, y)$, we used the camera PSF at the center of the depth range. In all simulations, $\eta(x, y)$ was set to be Gaussian white noise with standard deviation $\sigma = .005$. Since the deblurring error can vary with $f(x, y)$, we compute the value over a variety of natural images and take the average. In Figure 2.2, we show the deblurring error for three EDOF methods. Wavefront coding achieves the minimum deblurring error for all cameras when the defocus blur diameter s_0A = 0 pixels. This is because the wavefront coding MTF is greater and therefore preserves more information when deblurred with the correct PSF. However, both diffusion coding and focal sweep produce a flatter curve that results in less deblurring error at all other depth locations.

To demonstrate the performance of our EDOF method, we simulated a scene consisting of an IEEE resolution chart. Simulated defocused images are shown in Figure 2.1(a), where the maximum defocus blur diameter is $s_0A = 100$ pixels. We apply Wiener deconvolution with the PSF at the center depth to obtain the EDOF images shown in (b). Close-ups of the deblurring results are shown in (c). As expected, the sharpest image is produced by wavefront coding for the center depth. However, wavefront coding produces significant deblurring artifacts for defocus values as small as $s_0A = 33$ pixels, while diffusion coding produces near identical results for the entire depth range.

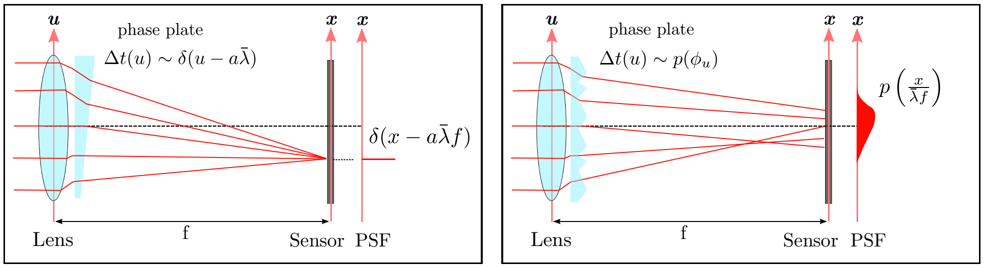
To generate the PSFs for Figures 2.1 and 2.2, we used the analytical solution for the diffusion coding PSF from Equation 2.16. For the focal sweep camera, we numerically integrated a sequence of defocus discs which, for the center PSF, represents a range of defocus blur diameters from 0 to 120 pixels. We performed a numerical search to find the focal sweep range that produces a local minimum in average deblurring error for this simulation. We used the raytracing engine in Zemax to numerically compute the wavefront coding PSFs without the effect of diffraction. To generate the Zemax raytrace, a cubic refractive surface was used such that the light field integration curve takes the form ($x = au^2, y = av^2$). The optimal value for a was chosen to be $a = S/(2A)$ [Levin *et al.*, 2009], where S is the maximum value of the defocus parameter s_0. Furthermore, we performed a numerical search to verify that this a produces a local minimum in average deblurring error for this simulation.

2.6 Implementing the Diffuser

We consider diffusers of the "kinoform" type [Caulfield, 1971], where the scattering effect is caused entirely by roughness variations across a surface. Such a diffuser can be considered a random phase screen, and according to statistical optics, for a camera with effective focal length f_l, and center wavelength λ, the effect of placing this screen in the aperture of the camera results in the following PSF [Goodman, 1985]:

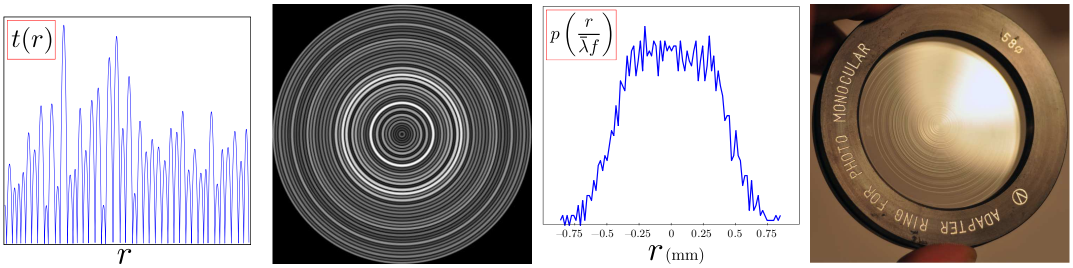
$$h'(x, y) \propto p_{\phi_u,\phi_v}(\frac{x}{\lambda f_l}, \frac{y}{\lambda f_l}), \tag{2.18}$$

where ϕ_u and ϕ_v are the u and v derivatives of the phase shift induced by the surface, and p_{ϕ_x,ϕ_y} is the joint probability of these derivatives. The result of Equation 2.18 is that we can implement a diffuser simply by creating an optical element with thickness $t(u, v)$, where the gradient of this surface $\bigtriangledown t(u, v)$ is sampled from a probability distribution which is also our desired PSF. Intuitively, we can understand this equation as follows: h_{ϕ_u,ϕ_v} denotes the fraction of the surface $t(u, v)$ with slope (ϕ_u, ϕ_v). For small angles, all incoming rays incident on this fraction of the surface will be deflected at the same angle, since the slope is constant over this region. Thus the quantity h_{ϕ_u,ϕ_v} also reflects the portion of light that will be deflected by the slope (ϕ_x, ϕ_y).



(a) A wedge with thickness $t(u) = a\lambda u$ (b) A randomly varying surface

Figure 2.9: A wedge can be thought of as a having a slope drawn from a probability density function which is a delta function. A diffuser can be thought of as a phase plate with a randomly varying thickness with a slope that is drawn from a more general probability density function.



(a) Diffuser profile (b) Diffuser height map (c) Diffuser scatter PDF (d) The diffuser

Figure 2.10: An implementation of the diffuser defined by the kernel in Equation 2.14. (a), (b), and (c) show the radial profile, height-map, and radial scatter function of the diffuser surface, respectively. (d) shows the fabricated diffuser.

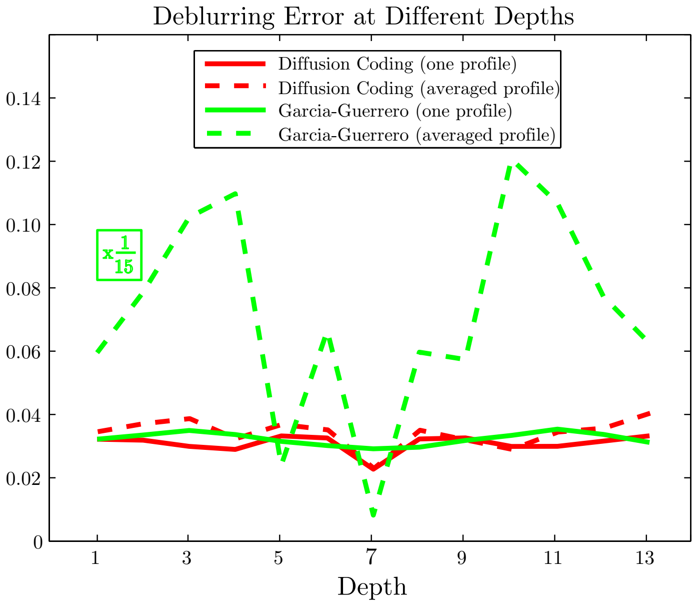


Figure 2.11: The deblurring error as a function of depth for both diffusion coding and the Garcia-Guerrero diffuser. The dotted lines show the deblurring error for a single instance of the diffuser surface. The solid lines show the deblurring error averaged over 100 realizations of the diffuser surfaces. A single instance of the diffusion coding surface performs significantly better than the Garcia-Guerrero diffuser.

In fact, kinoform diffusers can be thought of as generalized phase plates, as shown in Figure 2.9. In Figure 2.9(a), a wedge with thickness $t(u) = a\lambda u$ is placed in the aperture of a lens system. The effect of the wedge is to shift the PSF away from the optical axis. The wedge can be thought of as a having a slope drawn from a probability function $h(\phi_u)$ which is a delta function. The result of placing a wedge in the pupil plane of a camera is to shift the PSF, which can be thought of as convolving $h(\phi_u)$ with the PSF. A kinoform diffuser has a randomly varying surface with a more general probability distribution of slopes (Figure 2.9(b)).

To implement the diffuser defined in Equation 2.14, we follow the procedure in [Sales, 2003], which simply implements a diffuser surface as a sequence of quadratic elements whose diameter and sag is drawn from a random distribution. The scatter function is designed to be roughly Gaussian with 0.5mm variance (corresponding to $w = 1$mm in Equation 2.16) as shown in Figure 2.10(c). To create a radially symmetric diffuser, we create a 1D random profile and then apply a polar transformation to create the final 2D surface (see Figures 2.10(a) and 2.10(b)). The maximum height of the surface is 3μm. The diffuser was fabricated using a laser machining technology which has a minimum spot size of about 10μm. To ensure that each quadratic element was fabricated with high accuracy, the minimum diameter of a single element was chosen to be 200μm, resulting in a diffuser with 42 different annular sections. The diffuser used in all our experiments is shown in Figure 2.10(d), and was fabricated by RPC Photonics [RPC].

To compare the performance of our diffuser surface relative to the analytic PSF from Equation 2.16 derived using light field analysis, we calculated PSFs for the diffuser surface using wave optics, and used them to create a deblurring error curve. The resulting curve is shown as the dotted red line in Figure 2.11, and it is very close to the light field curve shown in solid red.

We also used wave optics to compare the deblurring error for our diffuser and the diffuser proposed by Garcia-Guerrero et al. [García-Guerrero *et al.*, 2007]. For a fair comparison, we also restricted the feature size of the Garcia-Guerrero diffuser to be 200μm. Since this design requires features to reduce in size from the center to the edge of the diffuser, only 21 annular sections could be made to fit within a 22mm aperture. The results are shown in Figure 2.11. The solid red and green lines show the deblurring errors for the diffusion coding and Garcia-Guerrero diffuser, respectively, for PSFs that are averaged over 100 surface realizations. The two curves are very similar, however, a single realization of the diffusion coding surface performs much closer to the average, as seen from the dotted red and green lines. In short, given the imposed fabrication limitations, diffusion coding significantly outperforms the Garcia-Guerrero diffuser.

2.7 Experimental Results

Figure 2.12 shows the PSFs produced when using the diffuser shown in Figure 2.10(d). The PSFs closely resemble the shape predicted by Equation 2.16 as is evident from the depth-invariance shown in the figure. The PSFs are normalized to unit intensity by color channel. The defocus range is chosen so that the normal lens PSF blur diameter ranges between 0 and 1 mm.

Figure 2.14 shows two images taken with a normal lens (Figure 2.14(a) taken with f/4.5 and Figure 2.14(b) taken with f/29) and two images (Figure 2.14(c) before deblurring, and Figure 2.14(d) after deblurring) taken with the diffuser from Section 2.6. All images are taken with a 50ms exposure time and the brightness in the f/29 image is normalized. The example shows that diffusion coding does indeed give far superior results in comparison to stopping down a lens. The deblurred image in Figure 2.14(d) extends depth of field by roughly a factor of six.

Figure 2.13 compares images taken with a normal lens to diffusion coded images taken with the diffuser from Section 2.6. The depth range of each scene is chosen so that the normal lens PSF blur diameter ranges between 0 and 1 mm. Within each figure, all images have the same exposure time and aperture setting. In each figure, three images are taken with the normal lens focusing on the background, middle, and foreground. These three images are then compared to the diffusion coded image(s). In all examples, the deblurred diffusion coded images exhibit a significant increase in DOF.

All images were captured with a Canon 450D sensor. To capture diffusion coded images, the 22mm diameter diffuser from Figure 2.10(d) was inserted into the aperture of a 50mm

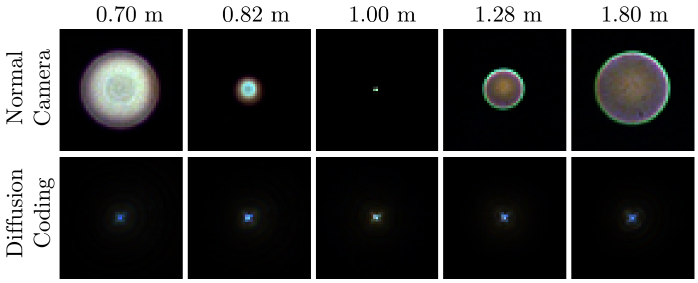


Figure 2.12: Measured PSFs for a 50mm f/1.8 lens without (top) and with diffusion coding (bottom). Almost no variation is visible in the diffusion coding PSF.

f/1.8 Canon lens. Deblurring of all diffusion coded images was performed using the BM3D deblurring algorithm [Dabov *et al.*, 2006]. The BM3D deblurring algorithm enforces a piecewise smoothness prior that suppresses the noise amplified by the deblurring process. Note that, as discussed in Section 2.5, all EDOF cameras amplify noise in the deblurring process, and the amount of amplification can be measured by the deblurring error. The result of using the BM3D algorithm is that while our deblurred images do not look noisy in comparison to images captured without the diffuser, some of the fine details in the deblurred images are not preserved.

2.8 Relating Diffusion Coding and Focal Sweep

Equation 2.16 gives an analytic expression for the PSF produced by a diffuser with the box-shaped scattering function defined by Equation 2.14. In Section 2.7, we experimentally verified that this type of diffusion coding produces very similar results to focal sweep. However, it is possible to show analytically that, for a certain type of scatter function, the diffusion coding produces exactly the same performance as focal sweep.

When we move the sensor to a distance d from the aperture plane, the sensor is no longer located at the (x, y) plane, which is fixed at a distance of f_l. We define the light field slope of the sensor plane $s = (d - f_l)/d$. The PSF for a point that comes to focus at distance d_0 from the aperture plane is then

$$h_s(r) = \frac{4}{\pi(s-s_0)^2A^2} \sqcap \left(\frac{r}{(s-s_0)A}\right). \tag{2.19}$$

For a focal sweep camera that integrates over a range of light field slopes $s \in [-S/2, S/2]$, the PSF is given by

$$h_{fs}(r) = \frac{1}{S}\int_{-S/2}^{S/2} h_s(r)ds \tag{2.20}$$

$$= \frac{1}{S}\int_{-S/2}^{S/2} \frac{4}{\pi(s-s_0)^2A^2} \sqcap \left(\frac{r}{(s-s_0)A}\right). \tag{2.21}$$

In Appendix A, we show that, for a point source focused on the focal plane, this PSF can be written as

$$h_{fs}(r) = \frac{4}{\pi SA}\left(\frac{1}{|r|} - \frac{4}{SA}\right) \sqcap \left(\frac{r}{SA}\right). \tag{2.22}$$

It possible to find an analytic expression for the MTF of the focal sweep camera by taking the Fourier transform of Equation 2.22. In radial symmetric coordinates, the MTF $H_{fs}(\omega_r)$ is found using the Hankel transform

$$H_{fs}(\omega_r) = 2\pi \int_{-\infty}^{\infty} J_0(\pi\omega_r r) h_{fs}(r) r dr \tag{2.23}$$

$$= 2\pi \int_{-\infty}^{\infty} J_0(\pi\omega_r r) \frac{1}{S}\int_{-S/2}^{S/2} h_s(r) ds r dr, \tag{2.24}$$

$$\tag{2.25}$$

where J_k is the k^{th} order Bessel function of the first kind. For point sources located on the focal plane, the focal sweep MTF becomes

$$H_{fs}(\omega_r) = \frac{1}{S}\int_{-S/2}^{S/2} \frac{2J_1(\pi s A\omega_r)}{\pi s A \omega_r} ds. \tag{2.26}$$

$$= {}_1F_2\left(\{1/2\}, \{3/2, 2\}, -\frac{1}{16}\pi^2 S^2 A^2 \omega_r^2\right), \tag{2.27}$$

where ${}_pF_q$ is the Generalized Hypergeometric function [Slater, 1966]. It is also possible to derive a slightly more complicated expression for the focus sweep MTF without the restriction that the point source be located on the focal plane.

We now return to a special form of diffusion coding. We again consider radially symmetric diffusers, but now we consider the case where the scatter profile varies as a function of aperture coordinates

$$k(r,\rho) = \frac{1}{S|\rho|} \sqcap \left(\frac{r}{S|\rho|} \right). \tag{2.28}$$

The physical interpretation of this scatter function is that the amount of diffusion increases with distance away from the optical axis. After passing through the diffuser, the light field of a point source then becomes

$$l'_\delta(\rho, r) = \frac{4}{\pi A^2} \sqcap \left(\frac{\rho}{A} \right) \frac{\sqcap(\frac{r - s_0\rho}{S|\rho|})}{\pi S|\rho||r|}, \tag{2.29}$$

and, as we show in Appendix A, the PSF is also given by the expression in Equation 2.22. Furthermore, it is possible to show that the PSF remains identical even at all depths. This means that a diffuser with the kernel given by Equation 2.28 will also have the same MTF as a focal sweep camera, given by Equation 2.27, and therefore also have exactly the same deblurring performance. Unfortunately, it is not entirely clear how to produce a diffuser with the scatter function given in Equation 2.28, or, for that matter, and scatter function that varies as a function of aperture coordinates.

2.9 Discussion

The diffusion coding technique introduced in this chapter is an attractive method for computationally extending DOF. In Section 2.3, we showed how to model a diffuser as a kernel applied to a light field. We then used this notation to guide the design of a depth-invariant diffuser. The radially symmetric diffuser introduced in Section 2.6 produces a PSF which achieves a similar performance to a focal sweep camera, but without the need for mechanical motion. Since focal sweep cameras achieve a near-optimal tradeoff between MTF and depth-invariance, the introduced diffusion coded camera must also be near optimal.

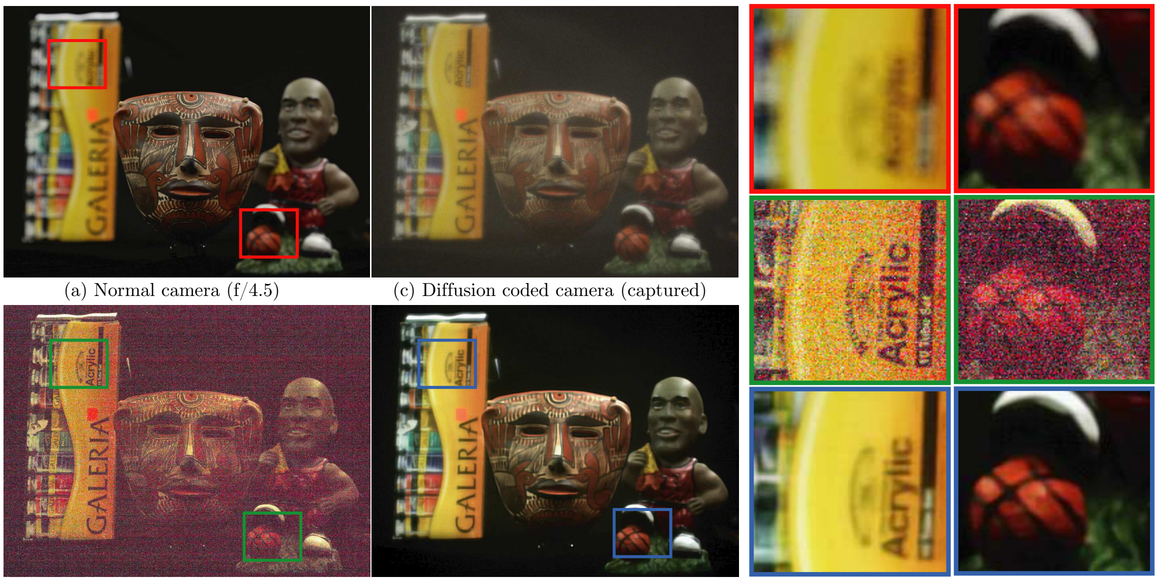
The fabricated diffuser introduced in Section 2.6 functions close to what is predicted by the theoretical analysis of Section 2.3. The example EDOF images captured using the diffusion coded camera demonstrated a significant extension in DOF. However, we have not given a thorough treatment of the noise model in the analysis of this chapter. We compared the performance of different EDOF techniques in Section 2.5, but the only performance comparison between EDOF and conventional cameras (i.e. a stopped down lens) was given in Figure 2.14. In this example, the same camera sensitivity setting was used for EDOF and conventional cameras. This is a fair comparison when the signal is very weak, but for stronger signals, a more fair comparison would be to increase the sensitivity for the less efficient system. This will cause a change in the noise characteristics of the captured image.

This chapter began with the assumption that an increase in efficiency will lead to an increase in performance. This is the case when the noise is signal independent. Then an EDOF technique will have a clear performance advantage over a conventional camera. However, noise is not always signal independent, and therefore the performance advantage of an EDOF technique depends on the noise model used. In Chapter 5, we return to the topic of performance comparison between EDOF and conventional cameras. We introduce a more complete noise model, and ask what conditions, if any, will preclude an EDOF technique from achieving a performance advantage over a conventional camera.

(a) Normal camera at three focus settings

(b) Diffusion coded camera

Figure 2.13: Extending DOF with diffusion coding. All images were taken with a 16ms exposure time. (a) The top, middle, and bottom images were captured using a a 50mm f/1.8 Canon lens focused on the background, middle, and foreground, respectively. The depth of field is too narrow for all objects to be in focus simultaneously. (b) The diffuser from Section 2.6 is inserted into the lens aperture and deblurring is applied to recover the EDOF image in (b). Diffusion coding results in a roughly 10× increase in DOF.



(a) Normal camera (f/4.5)

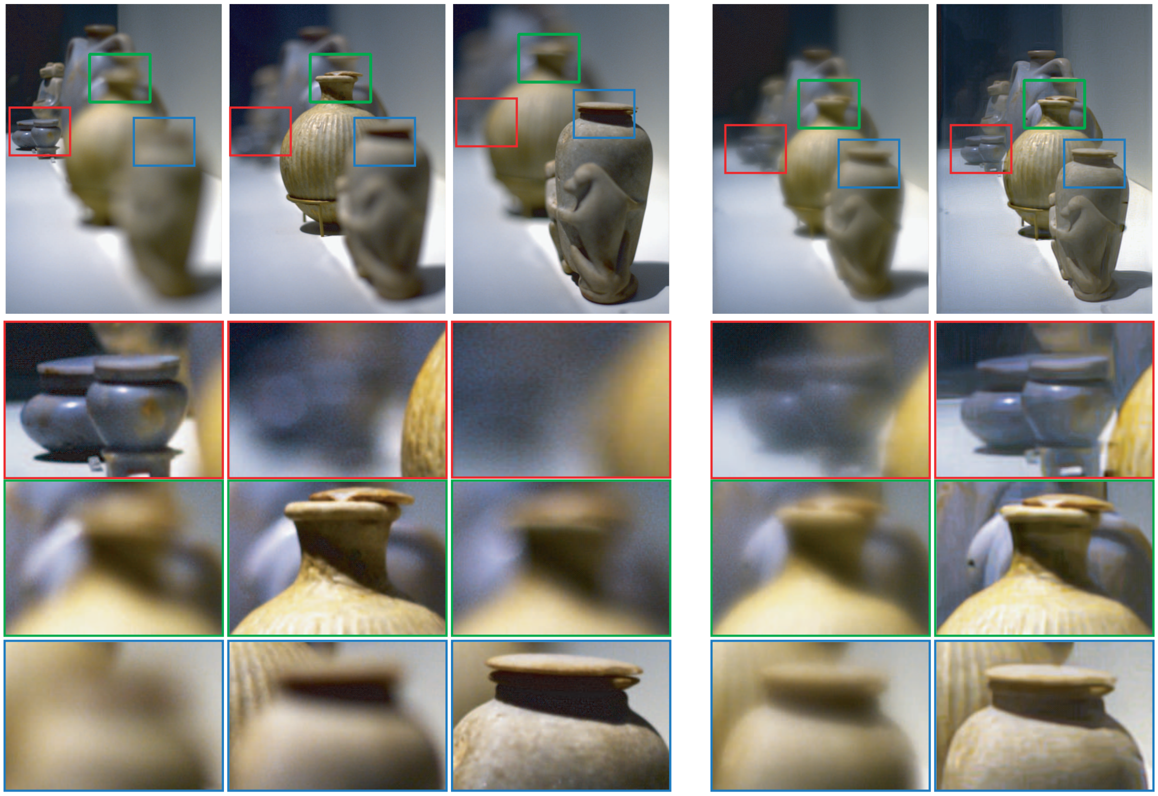
(c) Diffusion coded camera (captured)

(b) Normal camera (f/29)

(d) Diffusion coded camera (deblurred)

(e) Close-ups

Figure 2.14: Noise comparison between a diffusion coded camera and a normal camera. All images were taken with a 20ms exposure time. (a) Image taken with a f/4.5 camera. The DOF is too narrow for all objects to be in focus. (b) Image taken with the lens stopped down to f/29. All the objects are in focus but the noise is significantly increased. (c) Image taken with the same settings as in (a), but with the diffuser from Section 2.6 inserted into the lens aperture. All objects are in focus, but the image exhibits a slight haze. (d) Image obtained by deblurring the one in (c). The image preserves similar detail as in (b), but with significantly less noise. (e) Close-ups of the images in (a),(b), and (d).



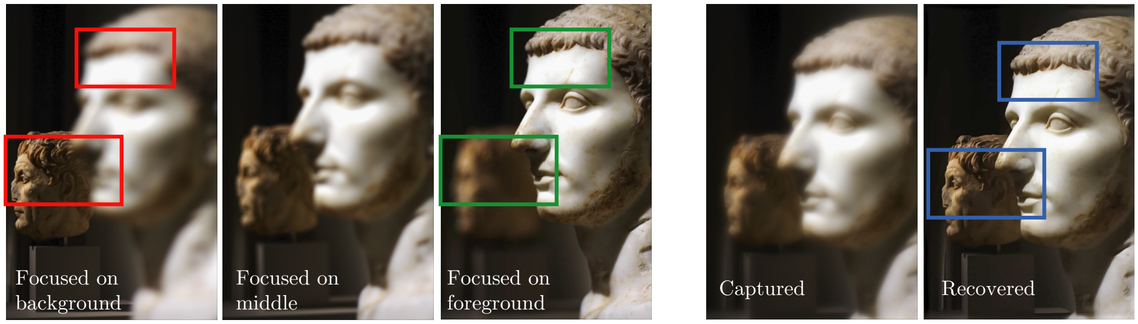
focus at background focus on middle focus on foreground

(a) Normal camera

captured recovered

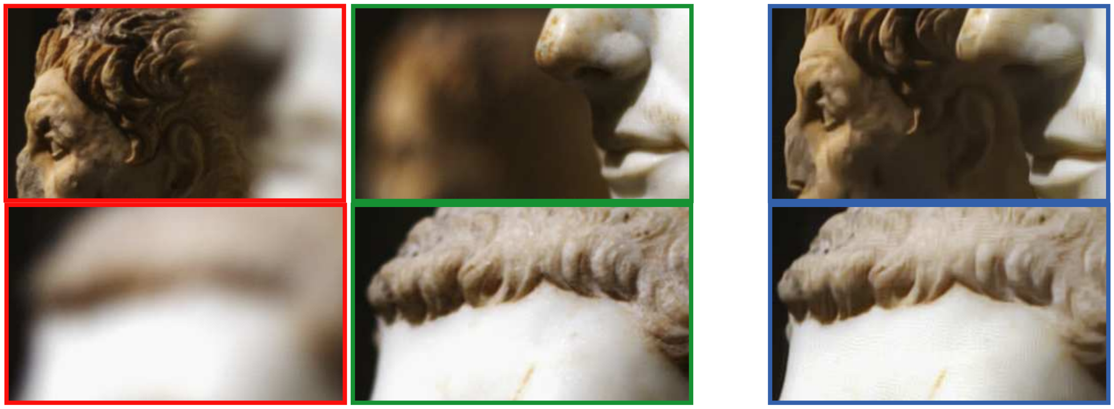
(b) Diffusion coded camera

Figure 2.15: Images of a scene consisting of several vases at different depths shot with a 50mm f/1.8 Canon lens. All images were taken with a 12ms exposure time. (a) Images focused on the background, middle, and foreground from left to right. (b) Images captured using the diffuser from Section 2.6. The right column shows the result after deblurring. Close-ups at the bottom show that the recovered image significantly increases DOF.



(a) Normal camera

(b) Diffusion coded camera



(Focused on background)

(Focused on foreground)

(Recovered)

(c) Close-ups from (a)

(d) Close-ups from (b)

Figure 2.16: Images of a scene consisting of two statues at different depths shot with a 50mm f/1.8 Canon lens. All images were taken with a 10ms exposure time. (a) Images are focused on the background, middle, and foreground from left to right. (b) Images captured using the diffuser from Section 2.6. The right image shows the result after deblurring. Close-ups at the bottom show that the recovered image significantly increases DOF.

Chapter 3

Spectral Focal Sweep

3.1 Introduction

In Chapter 2, we introduced an EDOF technique that computationally increases DOF by placing a diffuser in the aperture of the lens. We have seen a number of other techniques for extending DOF, including the use of coded apertures [Levin *et al.*, 2007][Zhou and Nayar, 2009], phase plates [E. R. Dowski and Cathey, 1995][Levin *et al.*, 2009], or mechanical motion [Nagahara *et al.*, 2008][Häusler, 1972]. All the EDOF techniques discussed thus far *increase* complexity. They require either more optical or mechanical components than a conventional lens. This chapter approaches the problem of extending DOF from another perspective – by *simplifying* the imaging system (see Figure 3.2). The main idea is to take advantage of the dispersive properties of refractive elements to create depth-independent blur. This has the advantage of reducing the number of constraints placed on the camera lens, so that a design with reduced complexity will suffice. The disadvantage is that color imaging performance suffers.

Refractive materials such as glass and plastic bend light rays according to Snell's Law. According to this law, the bending power of a refractive surface is a function of the index of refraction (IOR) of the material. Because the IOR is in turn a function of wavelength, rays incident on a refractive surface are deflected different amounts according to their color. This phenomena is known as chromatic dispersion. In lens design, chromatic dispersion is considered undesirable because it results in lens aberrations which reduce image quality.

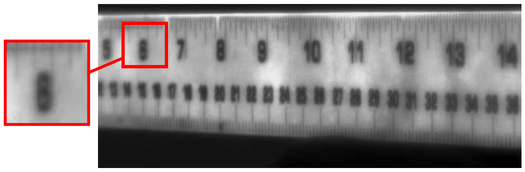
However, chromatic aberrations produce a very useful property that can be exploited; a lens with axial chromatic aberrations has a focal length that varies as a function of wavelength. If such a lens is used with a black and white sensor, the imaging system can be thought of as possessing a continuum of focal lengths simultaneously. We call such a system a “Spectral Focal Sweep” (SFS) camera because it uses chromatic aberrations to create the same effect as existing focal sweep techniques [Nagahara *et al.*, 2008][Häusler, 1972] with one important distinction: it can be used to extend DOF with no moving parts.

To design a SFS lens, we use an optimization that intentionally *maximizes* axial chromatic aberrations while minimizing other aberrations. This approach can greatly simplify lens design, reducing the cost and size of the design relative to a conventional lens design. We use this optimization to engineer a PSF which is not a delta function, but is approximately invariant to depth and preserves image details over a large depth range.

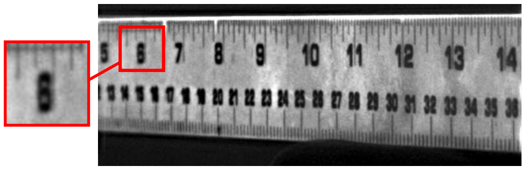
For a SFS camera, the amount of focal sweep depends on the reflectance spectra of objects being imaged. The more broadband an object’s spectrum, the wider the focal sweep. Thus, to function correctly, the camera requires objects being imaged to possess reasonably broad spectral reflectance distributions. Fortunately, the reflectance spectra of most real-world objects is sufficiently broadband [Parkkinen *et al.*, 1989]. We have observed that the SFS camera can effectively increase DOF for a wide variety of scenes (see Section 6, Figures 1, 8-11, and supplementary material). To further verify our claim that a SFS camera works effectively for most real-world spectra, we simulate the performance of our lens using the Munsell color database [of Joensuu Color Group, 2011] in Section 5. The Munsell database consists of 1250 spectrophotometer readings of common reflectance spectra.

It is interesting to note that the SFS camera bears some similarity to NTSC and related video compression techniques. These techniques exploit the fact that the human visual system relies much more heavily on luminance information than color. Before compression is applied, images are first transformed to a different color space such as YUV or NTSC. After transformation, color channels in the image can be compressed more aggressively without significant perceptual degradation. The SFS camera can be thought to apply a similar compression to an image *before* acquisition. For this reason, the SFS camera can be used to capture not only black and white images, but color images as well. To deblur color

(a) An image captured with a corrected lens (8ms exposure)



(b) An image captured with a SFS camera (8ms exposure)



(c) The image from Figure 3.1(b) after deblurring

Figure 3.1: Comparison of the SFS camera with a corrected lens. The image shown in Figure 3.1(a) was taken with a corrected lens. Images shown in Figures 3.1(b) and 3.1(c) were taken with a SFS camera. Figure 3.1(c) demonstrates that after deblurring, more detail is visible over a larger depth range when using the SFS camera.

images, we use an approximate method that produces results which are not exact but *look* good (see Figures 10 and 11, and supplementary material).

3.2 Related Work

There a number of techniques for extending DOF by increasing the complexity of the imaging system. Examples include all-optical techniques such as apodization [Welford, 1960], or the use of zone-plates [Ojeda-Castaneda and Berriel-Valdos, 1990] and computer-generated amplitude holograms [Rosen and Yariv, 1994]. coded aperture techniques. Other examples

Figure 3.2: A comparison showing the relative sizes and complexities of a Cosmicar 75mm F/1.4 lens (left) and our F/4 SFS doublet lens (right). Our lens is significantly lighter and more compact. The corrected lens is stopped down to F/4 in all experiments.

include the use of phase plates which produce PSFs that are approximately depth invariant [Chi and George, 2001][E. R. Dowski and Cathey, 1995]. The focus sweep techniques produce a depth invariant PSF by sweeping either the sensor or object along the optical axis during exposure [Nagahara *et al.*, 2008][Häusler, 1972].

Other works exist in the vision community which recover an extended DOF image after first estimating scene depth [Levin *et al.*, 2007][Levin *et al.*, 2009][Zhou and Nayar, 2009]. These techniques also increase complexity by introducing either phase plates or coded aperture patterns. Furthermore, the quality of these techniques is closely coupled to the precision of depth estimation, since each region in the image is deblurred using an estimated defocus PSF.

The work most similar in spirit to the SFS technique is by DxO Optics [Guichard *et al.*, 2009], which also proposes to extend DOF by exploiting axial chromatic aberrations. This approach finds the color channel which is best focused and then transfers high frequency information from this channel to the remaining color channels. The scene details recovered using this technique are limited by the quality of the best focused channel. We show in the next section that for a system with axial chromatic aberrations, even the best focused color channel is blurred. This is because the spectra of real-world materials and the spectral response of color filters on the image sensor are broadband. Our SFS technique, on the other hand, can be considered analogous to existing focal sweep techniques. SFS imaging creates an approximately depth-invariant PSF. By deconvolving the captured image with the inverse of this PSF, an extended DOF image is recovered with details very close to what can be acquired with a corrected lens. In short, the SFS technique is able to recover more information (and hence DOF) than the frequency transfer method of DxO.

3.3 Theory

In this section, we describe the theoretical foundation for the SFS camera. We first consider the imaging properties of a 'thin' singlet (single element) refractive lens manufactured out of glass with IOR $n(\lambda)$, aperture diameter A, and radii of curvature R_1 and R_2, respectively. The focal length of this thin lens is [Smith, 1966]

$$f_{EFL}(\lambda) = (n(\lambda) - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{3.1}$$

The dependence of focal length on wavelength is a result of the dispersive property of refractive materials, and this dependence, referred to as chromatic focal shift or axial chromatic aberration, is usually considered undesirable (see Figure 3.3). There are several well-established strategies for reducing its effect, e.g., by pairing two or more individual elements made from materials with complementary dispersive properties [Geary, 2002].

A singlet is usually insufficient for imaging onto a sensor because it exhibits strong spherical and field-dependent aberrations. To combat this, more elements are usually introduced to increase the degrees-of-freedom in the lens design optimization. The effective focal length $f_{EFL}(\lambda)$ of a compound lens can be calculated directly using the focal lengths and positions of individual elements. If a compound lens exhibits negligible spherical and field dependent aberrations, the irradiance $E(x, y, \lambda)$ of a point source with distance u from the lens and spectral reflectance $R(\lambda)$ can be written as

$$E(x, y, \lambda) = R(\lambda) \sqcap \left[\frac{r}{d(\lambda)}\right], \tag{3.2}$$

where $r = \sqrt{x^2 + y^2}$, $\sqcap$ is the circ function:

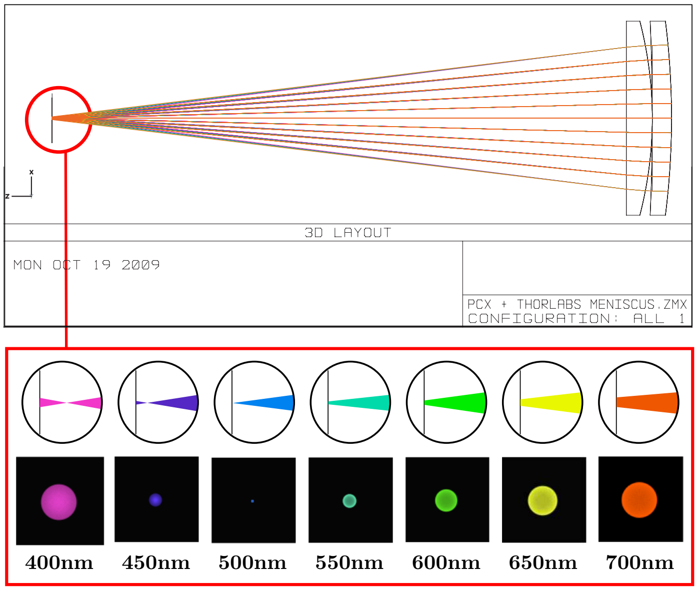


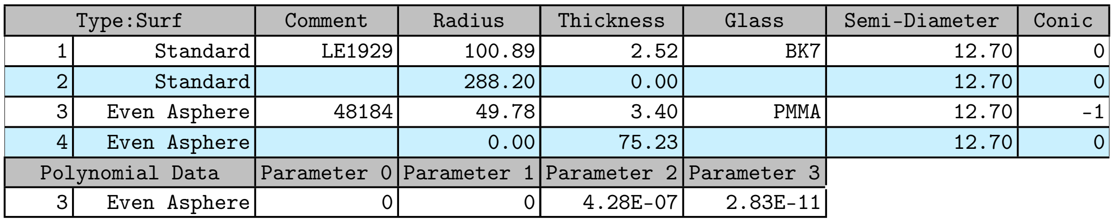
Figure 3.3: A SFS lens design is shown in the top figure. Below, a Zemax raytrace and PSF simulations are shown for various wavelengths. The lens exhibits strong axial chromatic aberration.

$$\sqcap\left(\frac{r}{d}\right) = \begin{cases} \frac{1}{\pi d^2} & \text{if } r < \frac{d}{2} \\ 0 & \text{otherwise} \end{cases}, \tag{3.3}$$

d is the chromatic defocus blur width which is determined from the gaussian lens law as

$$d(\lambda) = Av\left(\frac{1}{f_{EFL}(\lambda)} - \frac{1}{v} - \frac{1}{u}\right). \tag{3.4}$$

Here, v is the sensor-to-lens distance, and A is the lens aperture diameter. A black and white sensor with spectral sensitivity $S(\lambda)$ will then measure a sampled version of the image irradiance $E(x, y)$ averaged over wavelength. If we assume that $S(\lambda)$ is constant with value $\frac{1}{\lambda_2 - \lambda_1}$ between wavelengths λ_1 and λ_2, and zero everywhere else, then we can write our PSF $h(x, y)$ as



Type:Surf		Comment	Radius	Thickness	Glass	Semi-Diameter	Conic
1	Standard	LE1929	100.89	2.52	BK7	12.70	0
2	Standard		288.20	0.00		12.70	0
3	Even Asphere	48184	49.78	3.40	PMMA	12.70	-1
4	Even Asphere		0.00	75.23		12.70	0

Polynomial Data		Parameter 0	Parameter 1	Parameter 2	Parameter 3
3	Even Asphere	0	0	4.28E-07	2.83E-11

Figure 3.4: The lens prescription data for the design shown in Figure 3.3.

$$h(x,y) = \int S(\lambda)E(x,y,\lambda)d\lambda \tag{3.5}$$

$$= \frac{1}{\lambda_2 - \lambda_1}\int_{\lambda_2}^{\lambda_1} R(\lambda)\sqcap\left[\frac{r}{d(\lambda)}\right]d\lambda. \tag{3.6}$$

Thus, the PSF for the SFS camera is a continuous sum of scaled concentric discs. We note that if $f_{EFL}(\lambda)$ varies linearly and the reflectance spectrum happens to be white, then the PSF is identical to the mechanical focal sweep PSF given in [Nagahara *et al.*, 2008]. If, on the other hand, the reflectance spectrum is not white, then the sum is weighted by the magnitude of the spectrum for each wavelength.

3.4 Design and Implementation

The top of Figure 3.3 shows a raytrace of the doublet SFS lens design used in the simulations of Section 3.5 and the experiments of Section 3.6. The lens was designed using Zemax Optical Design software. To optimize our lens, we maximized axial chromatic aberration over the wavelength range 400-700nm, while also minimizing PSF compactness for the center wavelength averaged over all field positions. We ran an optimization to create an F/4 75mm focal length lens consisting of two elements, which images onto a 1/3" sensor with 10μm pixel size. We found that a smaller spot size over a larger field of view can be achieved with a custom lens design. However, we decided to fit a design with off-the-shelf components from stock lens suppliers. The SFS lens design consists of an Edmund Optics plano-convex asphere (part #48184) and a Thorlabs positive meniscus (part #LE1929). The prescription is shown in Figure 3.4.

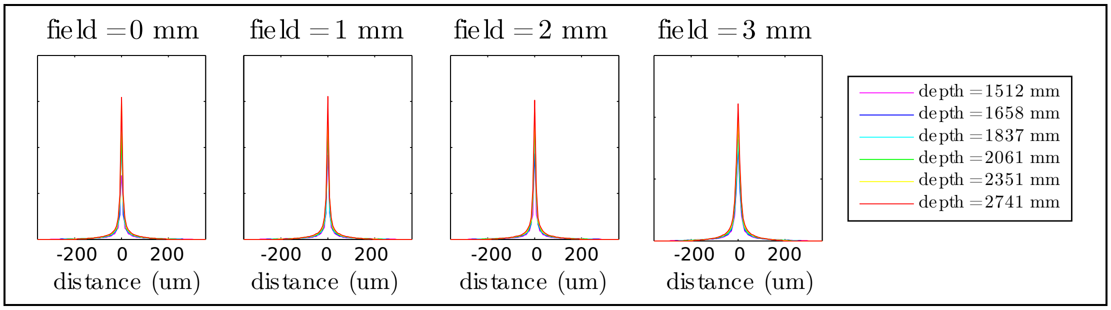


Figure 3.5: The simulated PSF for the lens in Figure 3.3 using a white spectrum. The PSF is shown as a function of depth and field position.

The bottom of Figure 3.3 shows the simulated PSF as a function of wavelength for our lens design. The wavelength-dependent PSF is shown to be the chromatic defocus disc given by Equation A.10, where the disc diameter scales as a function of wavelength. The largest disc diameter, about 100μm, occurs at 400nm and 700nm. Because the focal length is not exactly a linear function of wavelength, the PSF with the smallest spot size is at 500nm, not the center wavelength of 550nm. Figure 3.5 shows the simulated PSF of our lens when using a black and white sensor with a white point source. The depth values were chosen so that the defocus blur size for the center wavelength is 100μm (the same as the maximum chromatic defocus) at the two extreme depths. Note that the PSF does not vary significantly with depth and field positions.

Figure 3.2 shows a side-by-side comparison of our SFS lens with a corrected Cosmicar lens, also designed for use with a 1/3” sensor. The relative complexities of the two designs are obvious from their relative sizes. While the Cosmicar lens is capable of imaging at a smaller F/#, it is significantly larger, heavier, and requires 5-6 elements as opposed to 2. The simplicity of our lens is a direct benefit of the SFS approach. Conventional lens designs minimize chromatic aberrations by adding a constraint to the lens optimization. Optimization with additional constraints requires more degrees of freedom, resulting in designs with the addition of more surfaces, and thus more elements. The SFS lens design does away with this costly constraint, allowing a reduction in complexity of the final design.

3.5 Design Verification

To verify our claim that our camera is useful for a wide variety of real-world scenes, we simulated the PSF for an assortment of reflectance spectra captured with a spectrophotometer. We downloaded the Munsell database of 1250 different recorded spectra and used Zemax to simulate the PSF of these spectra when imaged through our design. In our simulations, we used 50 wavelength samples to simulate the PSF $h_d(x, y)$ at $d = 1, 2, ...12$ depth locations. Again, the depth values were chosen so that the defocus blur size for the center wavelength is the same as the maximum chromatic defocus at the two extreme depths.

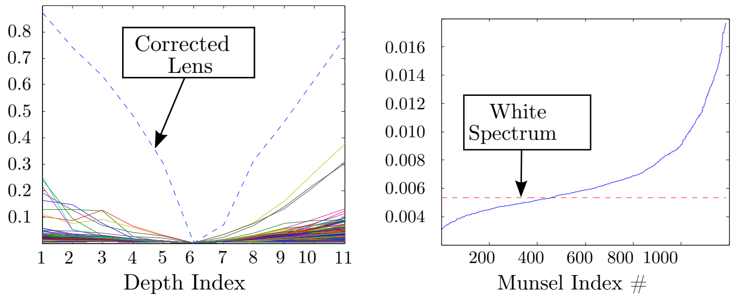
Figure 3.6 shows the results of our simulations. Figure 6(c) shows a cross section of the PSF for a few randomly selected spectra as a function of depth. Note that all of the PSFs have a strong peak, an indication that the PSFs preserve high frequencies. Also note that the PSF for each spectrum is relatively invariant to depth.

To quantitatively evaluate the quality of the PSFs from the Munsell database, we used the PSF distance measure $D(h_1(x, y), h_2(x, y))$ introduced by Zhou et. al [Zhou *et al.*, 2011]. This measure defines the similarity of two PSFs as the L2 norm of the Wiener reconstruction error for an image blurred by one PSF and then deconvolved with the other. For each Munsell color, we calculate the PSF distance for each $h_d(x, y)$ relative to the PSF at the center depth location. A plot of PSF distance is shown in Figure 6(a) for all Munsell colors, along with the PSF distance for a corrected lens (displayed as a dotted line). A flatter profile in this plot indicates less variation of the PSF with depth. The relative PSF distance for all Munsell colors imaged through the SFS lens is always less than for a corrected lens, significantly so for most colors. This indicates that the SFS lens always produces significantly more depth-invariant PSFs relative to a corrected lens.

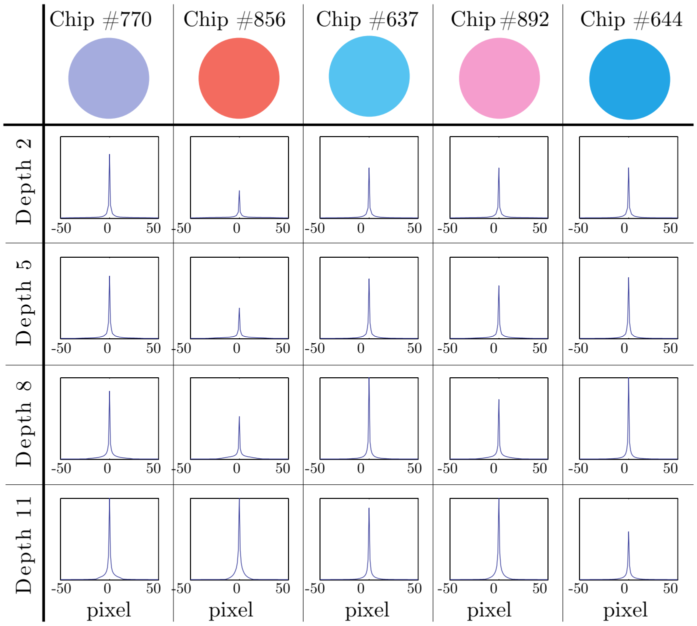
To further evaluate the performance of our camera relative to existing extended-DOF designs, we computed the average PSF distance:

$$A = \frac{1}{12} \sum_{d=1}^{12} D(h_d(x, y), \tilde{h}_6(x, y)), \tag{3.7}$$

where $\tilde{h}_6$ is the PSF of a white point source at the center depth. The quantity A measures the average reconstruction error of a spectrum imaged by our SFS camera when a white



(a) PSF variation (b) Average PSF variation



(c) PSF shape as a function of distance (plot width = 100px)

Figure 3.6: Figure 6(a) shows PSF variation as a function of depth for all Munsell colors when imaged through the SFS lens. The dotted line denotes the PSF variation for all colors using a corrected lens. Note the flatness of all SFS profiles compared to the corrected lens, indicating that the PSF varies little with depth for most real-world colors. Figure 6(b) shows the average PSF variation for 95% of the Munsell dataset when imaged through the SFS camera. The dotted line denotes the average PSF variation for a white spectrum imaged through the SFS camera. Figure 6(c) shows that PSF shape is relatively invariant to depth for randomly selected Munsell colors. PSF height is normalized against the center PSF for each color.

spectrum is used for deblurring. To evaluate the deblurring quality of the Munsell colors, we compare the computed A value to that of a white spectrum. Figure 6(b) shows A for a large number of Munsell colors. As shown in the figure, for a white spectrum, $A \approx .005$. The Munsell colors are sorted in order of ascending A and the bottom 95% percent are shown. Notice that for 95% of the colors, $A \leq .02$. Thus 95% of the Munsell colors have a variation that is within a factor of 4 of a white spectra. This implies that most naturally occurring spectra will not introduce significant deblurring artifacts relative to a black and white scene.

For a corrected lens, $A \approx .5$, which is nearly two orders of magnitude greater than for a white spectrum image through our SFS camera. Figure 3.7 shows that the measured PSF of a white spectrum source imaged through our SFS camera does indeed demonstrate significantly greater depth-invariance relative to a corrected lens.

3.6 Experiments

We now show several examples demonstrating the capabilities of our SFS lens. All black and white SFS images were captured using a Basler A311f VGA 1/3" sensor and the lenses shown in Figure 3.2. Color SFS images were captured using the same doublet SFS lens from Figure 3.2 and a Canon 450D sensor. Corrected lens examples were captured using a Cannon 100mm lens.

Deblurred images were generated using Wiener deconvolution with the PSF measured from a white point source (i.e. the bottom center PSF shown in Figure 3.7).

3.6.1 Black and White Images

Figure 3.8 demonstrates that even for a scene with a variety of colors, image quality is superior to that achieved by stopping down a lens. Figure 3.8(a) shows a scene with plastic toys captured by a F/4 corrected lens. Details in the foreground and background are lost due to defocus blur. Figure 3.8(b) shows an image captured with the same exposure time but stopped down to F/16. The depth of field has been increased, but the SNR is greatly decreased due to weaker signal strength. Figure 3.8(c) shows an image captured with the

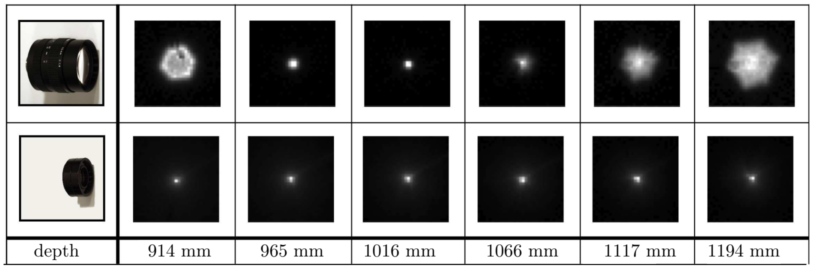


Figure 3.7: The measured PSF using a white point source as a function of distance for both lenses shown in Figure 3.2 (The corrected lens is stopped down to F/4). For the corrected lens, the PSF shape is roughly a disc with diameter proportional to defocus. The SFS lens produces a PSF that is approximately depth invariant.

F/4 SFS lens. Image details are clearly preserved over a larger depth range, but have a light haze due to the soft tail of the PSF. Figure 3.8(d) shows the results of deblurring Figure 3.8(c). The haze has been removed to improve contrast, resulting in crisp details over a larger depth range. The SNR is worse than in Figure 3.8(a), but significantly better than Figure 3.8(b).

3.6.2 Color Images

We have found that it is possible to use our SFS camera to restore color images using a simple and inexact approach that produces good visual results. We capture an RGB image with our SFS lens, then perform a YUV color transformation on the captured image. The resulting luminance channel closely approximates an image that would be captured with a black and white sensor. We deblur the luminance channel only, and transfer the image back to RGB space. The method is inexact because it does not account for color bleeding in the chrominance channels. However, as discussed in the introduction, blurring in these channels is much less perceptible to humans, and we have found that the technique produces satisfactory results for a variety of scenes. Figures 3.10 and 3.11 show details of color reconstructions, demonstrating the fidelity of our inexact deblurring technique.

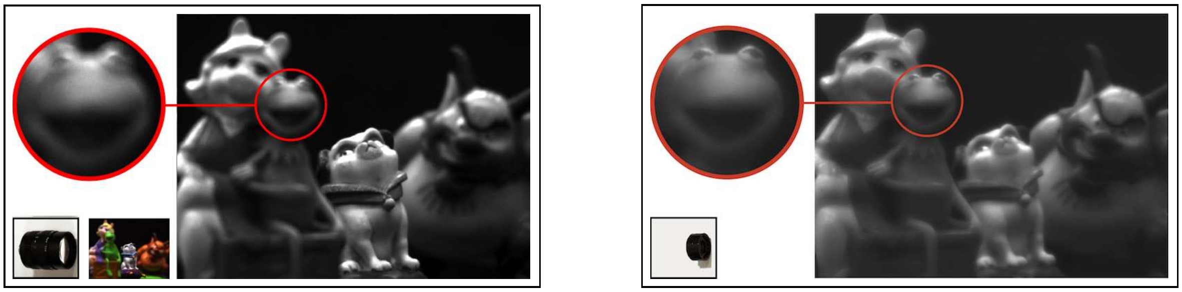
3.7 Limitations

While our technique does work well for a large variety of natural scenes, some naturally occurring spectra are not sufficiently broadband to produce a large spectral focus sweep range, and consequently produce a highly depth dependent PSF. The top 5% of Munsell colors (not shown) in Figure 6(b) have a PSF variation $V \geq .2$, some significantly larger. If our SFS lens is used to photograph a scene that contains narrowband reflectance spectra such as these, a significant amount of artifacts will be introduced after deblurring. Furthermore, while our approximate color deblurring method produces visually pleasing results, it does not correct for blurring in the chrominance channels, and is thus insuitable for many high quality imaging applications.

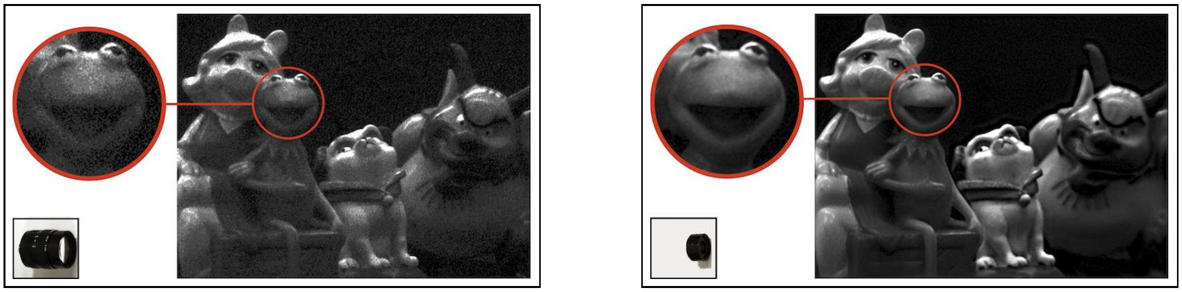
3.8 Discussion

The strategy discussed in this chapter was to increase DOF by reducing complexity. While conventional lenses are designed to *minimize* chromatic aberrations, the lens introduced in Section 3.4 was designed to *maximize* them. These aberrations are exploited for the purpose of extending depth of field. This approach *reduces* lens complexity by relaxing constraints in the lens optimization process. However, it also places restrictions on the scene being imaged. The technique works poorly when imaging narrow band reflectance spectra. However, our experiments with reflectance spectra databases and our prototype camera have indicated that most spectra are sufficiently broadband, and the technique functions well for a wide variety of scenes. The SFS lens introduced in Section 3.4 was built using off-the-shelf components, and produced a number of examples that demonstrate reasonable image quality.

The diffusion coding technique introduced in Chapter 2 and the SFS technique introduced in this chapter represent two different ways of approaching the problem of computationally extending DOF. The diffusion coding technique makes no assumptions about the spectral reflectance of objects being imaged. The SFS technique uses a more restrictive model for the signal. When input signals obey the model, and spectral reflectances are broadband, the performance of the SFS technique is similar to diffusion coding. Then the reduced complexity makes the SFS technique more preferable. However, the added complexity of the diffusion coding technique has a performance advantage associated with it. Diffusion coding will perform better on average over a larger class of input signals. The choice between the two techniques really boils down to the cost of increased complexity relative to the benefit of increased performance. In certain situations, even a small loss in performance may not be acceptable, and diffusion coding is the obvious choice. In other cases, cost may be a limiting factor, and the reduced complexity of the SFS technique may make it a more attractive option.

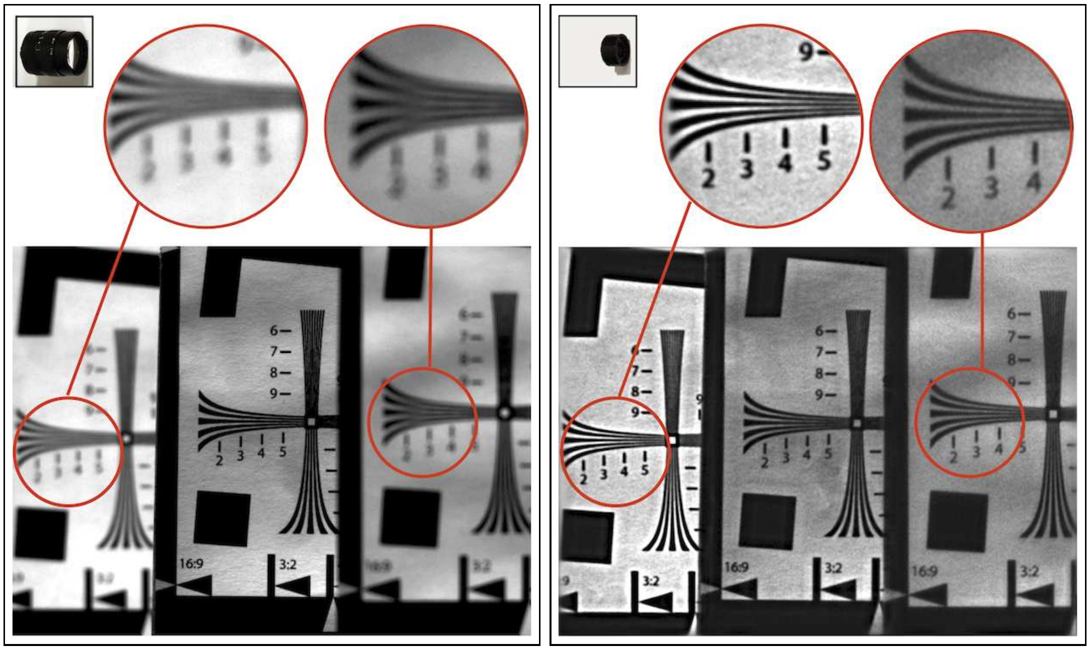


(a) Captured with a F/4 corrected lens (8ms exposure) (b) Captured with our SFS lens (8ms exposure)



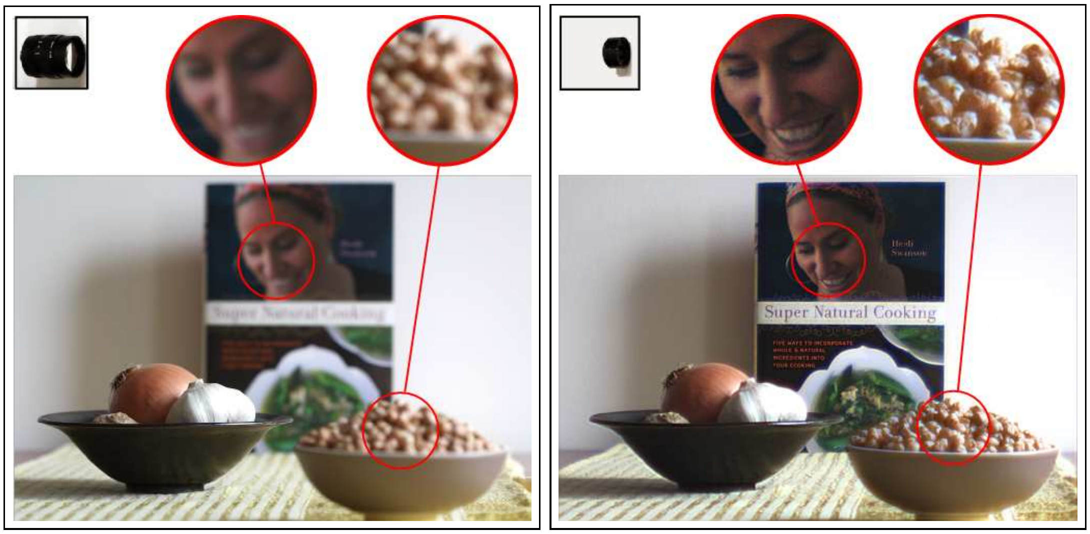
(c) Captured with a F/16 corrected lens (8ms exposure) (d) The image in Figure 8(c) after deblurring

Figure 3.8: Comparison of the SFS camera with a corrected lens. All images are taken with an 8ms exposure time. Images on the left are taken with a corrected lens and images on the right are taken with our SFS camera. As shown in Figure 3.8(a), the DOF using a F/4 corrected lens is too narrow. Figure 3.8(c) shows that if we stop down to F/16 we achieve the desired DOF, but our image is corrupted by noise. When using our SFS camera, we capture the image in Figure 3.8(b), then recover the extended DOF image shown in Figure 3.8(d), which has significantly less noise. A color thumbnail is included in the bottom-left of Figure 3.8(a) to show the colors in the scene.



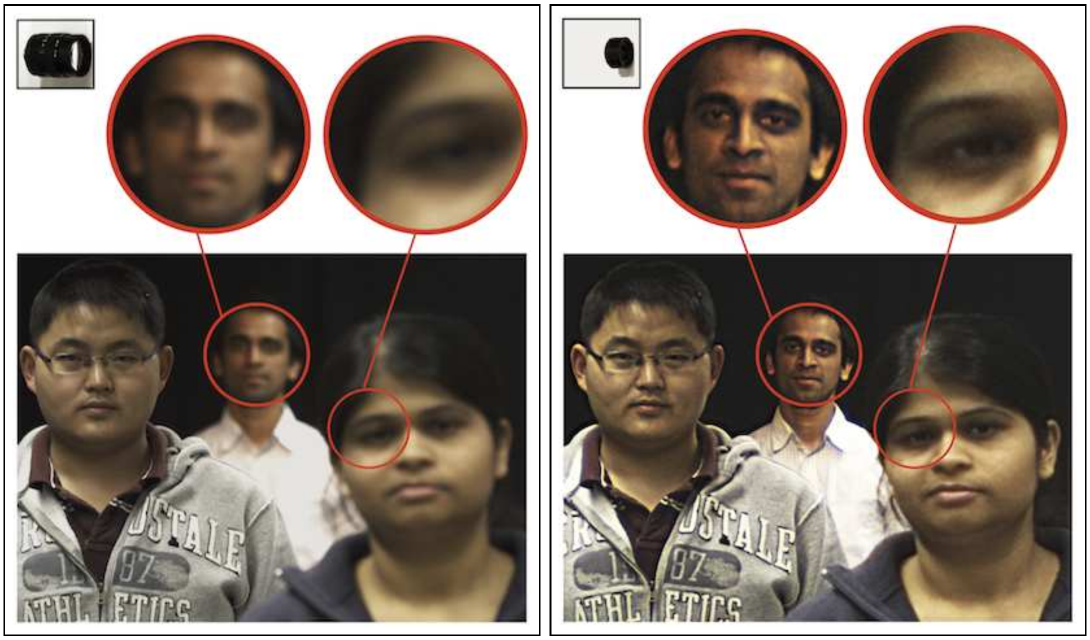
(a) An image captured with a F/4 corrected lens (b) An image captured with our F/4 SFS lens

Figure 3.9: A scene consisting of three identical resolution targets placed at different depth planes. Images were captured with an 8ms exposure time and the corrected lens is stopped down to F/4. The left image was taken with a corrected lens, and the right image was taken with our SFS camera (after deblurring). The insets show that more detail is visible in the front and back planes when using the SFS camera.



(a) An image captured with a F/4 corrected lens (b) An image captured with our F/4 SFS lens

Figure 3.10: A scene consisting of three objects placed at different depths on a table. Both images were taken with a 16ms exposure time and the corrected lens is stopped down to F/4. The image on the left was taken with a corrected lens and on the right is a deblurred version of an image taken with our SFS camera. The insets show that more detail is visible in the front and back objects when using our Spectral Focal Length camera.



(a) An image captured with an F/4 corrected lens (b) An image captured with our F/4 SFS lens

Figure 3.11: A scene consisting of three people located at different depths. Both images were taken with a 16ms exposure time and the corrected lens is stopped down to F/4. The image on the left was taken with a corrected lens and on the right is a deblurred version of an image taken with our SFS camera. The insets show that more detail is visible in the front and back faces when using the SFS camera.

Chapter 4

Gigapixel Computational Imaging

4.1 Introduction

In Chapters 2 and 3, we discussed the problem of computationally extending DOF. In the context of EDOF imaging, we face a tradeoff between best and average case performance. When we try to improve performance over a range of depths, we sacrifice the best possible performance at a single depth. In Chapter 3 we also saw a tradeoff between performance and complexity. We showed that a much simpler lens can be used to extend DOF, but at the price of reduced color performance. In this chapter, we explore this tradeoff further in the context of high resolution cameras. For these cameras, there is a tradeoff between scale and resolution. The scale (overall size) of the camera determines how many pixels we can fit within a given FOV. We can always increase scale to achieve a larger resolution, but there are costs associated with the size, weight, and power consumption of our cameras. Thus, it is attractive to look at the relationship between performance and complexity in order to determine if the cost of increased complexity warrants the resulting gain in performance.

High resolution cameras enable images to be captured with significantly more details than the human eye can detect, revealing information that was completely imperceptible to the photographer at the time of capture. These cameras allow humans to explore minute details of a scene that may have otherwise been overlooked (see Figure 4.2), benefitting a variety of applications including surveillance, inspection, and forensics. Because the performance of low-level automated vision tasks depend highly on the amount of image detail available, greater resolution also helps with computer vision tasks such as object detection, recognition and tracking. For these reasons and more, there is increasing demand for cameras with ever higher resolution. At present, highly specialized gigapixel imaging systems are being developed for aerial surveillance [DARPA, 2010].

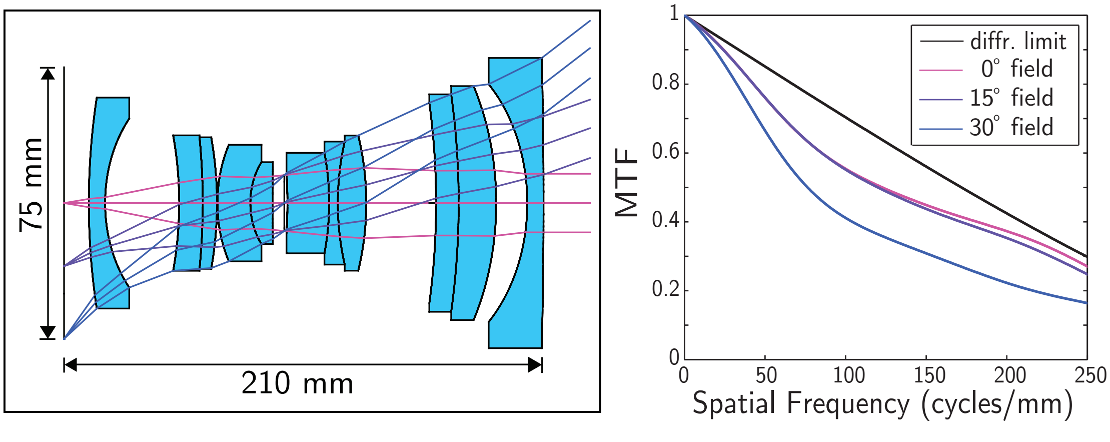
While CMOS and CCD technologies have improved to the point that imaging sensors with pixels in the 1μm range have been demonstrated [Fife *et al.*, 2008], it remains a huge challenge to design and manufacture lenses which have the resolving power to match the resolution of such a sensor. This is because the number of resolvable points for a lens, referred to as the Space-Bandwidth Product (SBP) [Goodman, 2005], is *fundamentally limited by geometrical aberrations.* Ideally, all lenses would be diffraction limited so that increasing the scale of a lens while keeping FOV fixed would increase SBP. Unfortunately, SBP reaches a limit due to geometrical aberrations.

There are two common approaches that are taken to increase SBP in the face of this fundamental limit. The first is to just accept the loss in resolution and increase sensor size. As an example, consider the commercially available F/8 500mm focal length Schneider Apo-Symmar lens. If this lens were diffraction limited, it would be capable of resolving a gigapixel image on a 5” × 5” sensor. However, because of geometrical aberrations, a sensor size of nearly 12” × 12” is necessary to resolve a full gigapixel image.

The second approach taken to increase SBP is to increase complexity as a lens is scaled up. Introducing more optical surfaces increases the degrees of freedom in lens optimization, which can be used to reduce geometric aberrations and achieve diffraction limited performance. Consider the F/4 75mm focal length lens shown in Figure 4.1. The lens is diffraction limited over a 60° FOV so that a gigapixel image can be resolved on a 75mm × 75mm surface, much smaller than for the Apo-Symmar. The increase in performance comes at a great cost, however. The design consists of 11 different elements, ranging from 60-100mm in diameter, resulting in a lens that is both expensive to produce and difficult to align.

We present a new approach to increase SBP - the use of computations to correct for geometrical aberrations. In conventional lens design, resolution is limited by the spot size of the lens. For a lens with aberrations, spot size increases linearly with the scale of the lens. For a computational imaging system, resolution is related to deblurring error. *We observe, however, that for a lens with spherical aberrations, deblurring error does not increase linearly with lens scale.* We use this remarkable fact to derive a scaling law that shows that computational imaging can be used to develop cameras with very high resolution while maintaining low complexity and small size. First, we analytically derive a closed form expression for the Point Spread Function (PSF) and Optical Transfer Function (OTF) of a lens with spherical aberration. We then use this expression to derive a closed form solution for the deblurring error as a function of lens scale. We go on to show how deblurring performance improves when image priors are introduced.

In Section 4.8 we present an imaging architecture that consists of a large ball lens shared by an array of small planar sensors coupled with a deblurring step. Due to our monocentric optical design, field-dependent aberrations are suppressed, and the primary aberrations are spherical and axial chromatic, which are known to code images in a manner that is invertible via post-processing [Robinson *et al.*, 2009] [Robinson and Bhakta, 2009] [Guichard *et al.*, 2009] [Cossairt and Nayar, 2010]. We demonstrate a proof-of-concept gigapixel camera that is implemented by sequentially scanning a single sensor to emulate an array of tiled sensors.



(a) An F/4 75mm focal length lens

(b) The MTF of the lens in (a)

Figure 4.1: (a) An F/4 75mm lens design capable of imaging one gigapixel onto a 75 × 75mm sensor. This lens requires 11 elements to maintain diffraction limited performance over a 60° FOV. (b) The MTF at different field positions on the sensor.

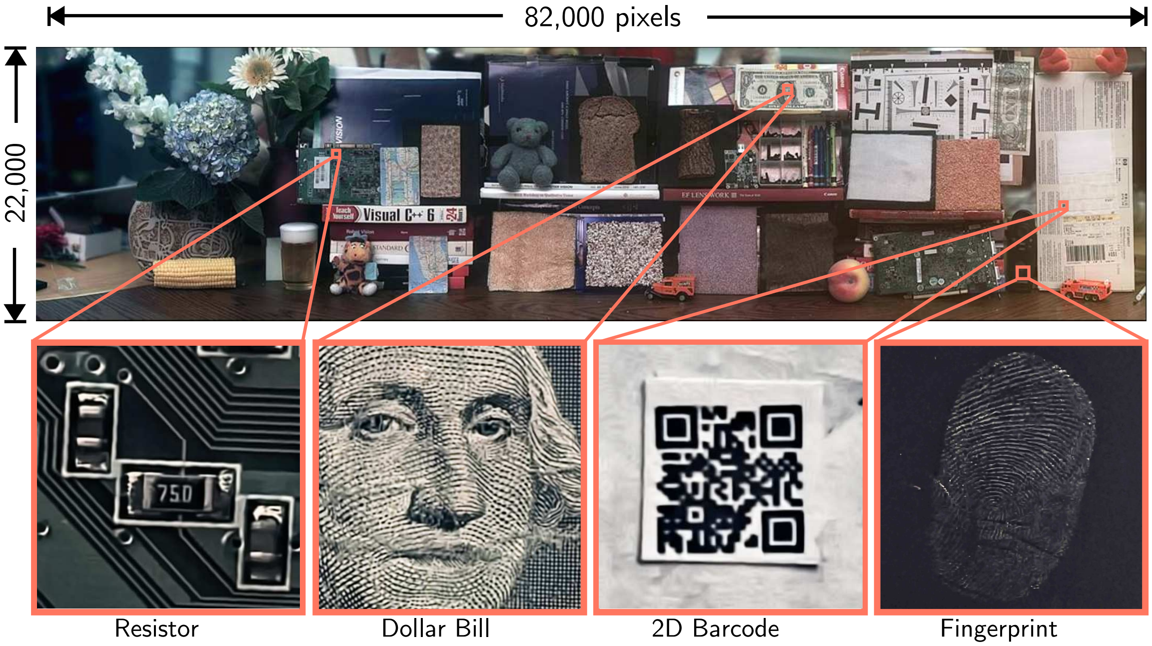


Figure 4.2: **A 1.7 gigapixel image** captured using the implementation shown in Figure 4.13. The image dimensions are 82,000 × 22,000 pixels, and the scene occupies a 126° × 32° FOV. From left to right, insets reveal the label of a resistor on a PCB board, the stippling print pattern on a dollar bill, a miniature 2D barcode pattern, and the fine ridges of a fingerprint on a remote control. The insets are generated by applying a 60×-200× digital zoom to the above gigapixel image.

In addition, we present a single element gigapixel camera design with a contiguous FOV.

In Section 4.9.3 we advocate the use of deblurring to remove the effects of aberrations. However the quality of deblurred images depends on the MTF of the lens, and a diffraction limited lens always has the best possible performance. Unfortunately, achieving diffraction limited performance often requires increasing the complexity of the lens, usually by increasing the number of surfaces. Lenses with greater complexity are typically larger, heavier, more expensive to manufacture, and more difficult to align. We analyze the trade-off between performance and complexity for the special case of spherical optics.

4.2 Related Work

4.2.1 Large Format Imaging Systems

A few custom high resolution imaging systems have been developed using large format lenses. These include systems built with commercial lenses that sequentially scan a large image plane surface [Ben-Ezra, 2010] [Wang and Heidrich, 2004], as well as a system with a custom lens that is photographed on film and later converted to a digital image [Gigapixl, 2007]. These are special purpose cameras that are extremely large (FL $>$ 500mm). In Section 4.8 we show that it is possible to capture images at comparable resolutions with a much smaller form factor.

4.2.2 Camera Arrays and Multiscale Optics

Camera arrays have been used to capture high resolution images by tiling multiple sensors paired with a complex lens [Wilburn *et al.*, 2005] [Nomura *et al.*, 2007]. However, a camera array for gigapixel imaging would be prohibitively large and expensive because it would require tiling an array of long focal length lenses. A related approach taken by Brady and Hagen [Brady and Hagen, 2009] is to use a multiscale optical system consisting of a large single element lens coupled with an array of smaller optical elements, each unique and coupled with a different sensor. The advantage of this approach is that it is a compact design that can correct for geometrical aberrations. The disadvantage is that the system requires a large number of different optical elements, which may be difficult to manufacture and align.

4.2.3 Monocentric Optics and Curved Sensors

Monocentric optical designs are free of field dependent aberrations because they are completely symmetric: the image plane and each lens surface lay on concentric spheres. Monocentric designs date back to the Sutton Panoramic Lens (1859), and later the Baker Ball Lens (1942) [Kingslake, 1989]. Luneburg proposed the use of a monocentric lens with varying index of refraction to correct for aberrations [Luneburg, 1964]. Rim et. al proposed a small diffraction limited camera consisting of a ball lens and curved sensor [Rim *et al.*, 2005]. Krishnan and Nayar proposed the use of a large ball lens and spherical sensor together with deblurring to create a single viewpoint, fully spherical FOV camera [Krishnan and Nayar, 2009]. While several researchers have made progress towards developing curved sensors [Dinyari *et al.*, 2008] [Ko *et al.*, 2008] [Lee and Szema, 2005], the technology is not yet ready for commercialization.

Recently, Marks and Brady proposed a 7-element large format monocentric lens called the Gigagon [Marks and Brady, 2010], which the authors suggest using with a large array of planar sensors. To our knowledge this system has yet to be implemented, but is similar in architecture to some of the designs we propose [1]. Our approach is fundamentally different in that we show how computations can be used to achieve the desired resolution while reducing complexity.

4.2.4 Computational Imaging

In the 90's, Cathey and Dowski proposed a hybrid optical-signal processing system which uses a cubic phase plate to extend depth of field [Dowski and Cathey, 1995]. Later they showed that the same element can be used to reduce the complexity of infrared cameras [Dowski *et al.*, 2000]. Robinson and Stork observed that spherical aberrations are easily invertible via image processing, and proposed the use of simpler lens designs based on this principle [Robinson *et al.*, 2009] [Robinson and Bhakta, 2009] [Robinson and Stork, 2009]. Guichard et. al [Guichard *et al.*, 2009] and Cossairt and Nayar [Cossairt and Nayar, 2010] observed that the effects of axial chromatic aberrations can be inverted using a method that is inexact, but produces images that look good.

4.3 Diffraction Limited Resolution

Lohmann originally observed that lenses obey certain scaling laws that determine how resolution increases as a function of lens size [Lohmann, 1989]. Consider a lens with focal length f, aperture diameter D, and image size Δx by Δy. We introduce a scaling factor

---

[1] Similar camera designs are also being pursued by the DARPA MOSAIC project, led by David J. Brady. *Terrapixel Imaging, ICCP '10 Invited Talk, Mar 2010.*

M, which is defined such that $M = 1$ corresponds to a focal length of $f = 1mm$. If we scale the lens by a factor of M, then f,D, Δx by Δy are all scaled by M, but the F/# and FOV of the lens remain unchanged. If, when we scale the lens, the minimum resolvable spot size has not also increased by a factor of M, then we have increased the total number of points that can be resolved. The number of resolvable points for a lens is referred to as the Space-Bandwidth Product (SBP) [Goodman, 2005]. SBP is a unit-less quantity that tells us the number of distinct points which can be measured over a given FOV.

The minimum spot diameter of a lens due to diffraction is $\delta_d \approx \lambda F/\#$, where λ is the wavelength of light. Since this quantity is independent of lens scale, the SBP for a diffraction-limited lens is

$$R_{diff}(M) = \frac{M^2 \Delta x \Delta y}{(\lambda F/\#)^2}. \tag{4.1}$$

The SBP increases quadratically with the scaling factor M (see the red curve in Figure 4.3).

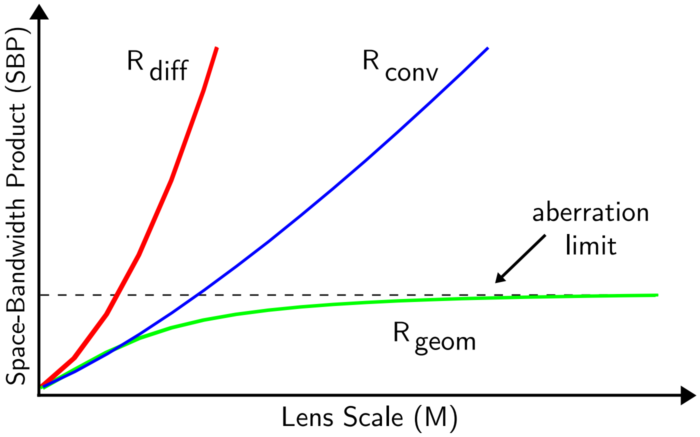


Figure 4.3: A plot showing how Space-Bandwidth Product (SBP) increases as a function of lens size for a perfectly diffraction limited lens (R_{diff}), a lens with geometric aberrations (R_{geom}), and a conventional lens design whose F/# increases with lens size (R_{conv}).

4.4 Aberrations and Image Quality

Ideally, all lenses would be diffraction limited, and resolution would scale quadratically with lens size. Unfortunately, the resolution of most lenses is limited not by diffraction, but by geometrical aberrations. This is because there is no lens shape that can produce a perfect focus for all points on the image plane. The best we can do is to reduce aberrations to the point that their effect is small compared to diffraction.

4.4.1 Aberration Theory

The Optical Path Difference (OPD) generalizes the concept of lens aberrations. The OPD measures the distance between an ideal focusing wavefront and the actual wavefront propagating through the lens as a function of normalized coordinates in the pupil plane (see Figure 4.4). For radially symmetric lenses, the generalized OPD is a function of 2-D polar coordinates $\{\rho \in [-1, 1], \phi \in [0, \pi]\}$ in the aperture plane, and the radial coordinate r on the sensor plane. In optical design, the OPD $W(\rho, \phi, r)$ is typically expressed as a Siedel polynomial, where each term in the polynomial represents a different type of aberration:

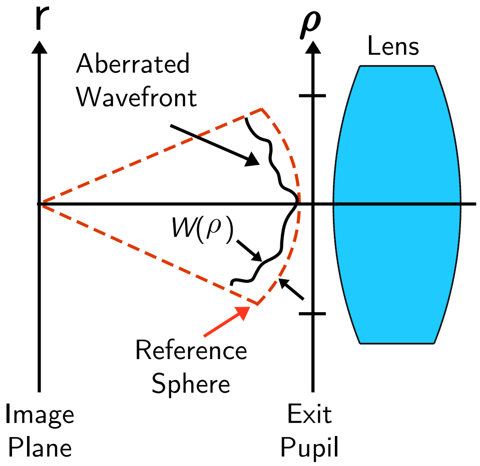


Figure 4.4: The OPD $W(\rho)$ of a lens is the path difference between an ideal spherical wavefront and the aberrated wavefront propagating from the exit pupil of the lens.

$$W(\rho,\phi,r)=\sum_{i,j,k}W_{ijk}r^i\rho^j\cos^k\phi. \tag{4.2}$$

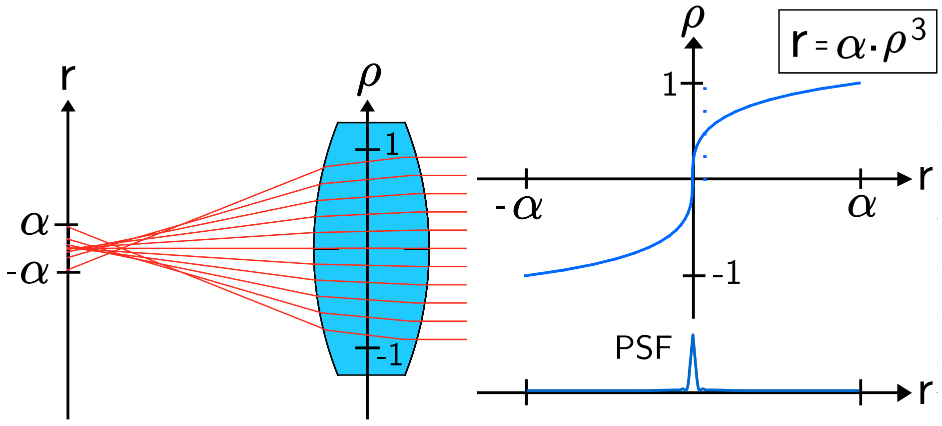
For instance, $W_{020}, W_{040}, W_{131}$ represent the amounts of defocus, spherical aberration, and coma, respectively. For spherical optical systems, the aberrations become independent of position on the sensor due to the symmetry of the system. In this case, the OPD becomes

$$W(\rho)=\sum_{i,j,k}W_{ijk}\rho^j, \tag{4.3}$$

in which case defocus and spherical aberration become the dominant aberrations. For a thin lens, the spherical aberration coefficient W_{040} can be shown to be [Geary, 2002]

$$W_{040}=\frac{\sigma_I}{512}\frac{D}{F/\#^3}, \tag{4.4}$$

where D is again the diameter of the lens aperture, and σ_I is the structural coefficient (a constant that depends only on index of refraction and is usually in the range $\sigma_I = 5-15$).



(a) A singlet with aberrations (b) The rayfan and PSF of (a)

Figure 4.5: (a) A singlet lens with strong spherical aberrations. (b) The rayfan shows ray position on the sensor plane as a function of position in the lens aperture. The PSF has a strong peak because rays are concentrated around the center of the image plane. The PSF's support is enclosed in an area of radius α.

4.4.2 The Aberration Induced PSF

When a lens exhibits aberrations, it can no longer produce a perfect focus. A perfectly focusing lens produces a Point Spread Function (PSF) that is a delta function, which produces the sharpest focus possible. Diffraction and geometric aberrations cause the PSF to deviate from this ideal shape. The OPD can be used to calculate the PSF produced by an optical system with aberrations. If the aberrations are relatively small, then the effect of diffraction needs to be considered and Fourier Optics must be used to derive the correct PSF shape. If the aberrations are large, however, the PSF can be derived using geometric optics. Since rays propagate perpendicular to the aberrated wavefront, we can use the OPD to determine where each ray pierces the sensor plane. The transverse ray-aberration curve $r = T(\rho)$ gives the position of a ray in the sensor coordinates r as a function of coordinates in the pupil plane ρ. For a point source at infinity, this is given by [Geary, 2002]:

$$T(\rho) = 2F/\# \frac{dW}{d\rho}. \tag{4.5}$$

For a lens with spherical aberrations, the transverse aberration curve is given by (see Figure 4.5(b))

$$\begin{aligned} T(\rho) &= \frac{\sigma_I}{64} \frac{D}{F/\#^2} \rho^3 &(4.6)\\ &= \alpha \rho^3, &(4.7) \end{aligned}$$

where α is the spherical aberration coefficient (usually called $SA3$). Because ρ is given in normalized coordinates, the full support of the PSF falls within a circle of radius α (see Figure 4.5(b)). From Equation 4.7 it is clear that if we scale the lens uniformly by a factor of M (such that the F/# remains constant), α increases by the same factor.

We can think of the ray-aberration curve as an integration curve in a radially symmetric light field phase space [Levin *et al.*, 2009] [Levin *et al.*, 2009] [Cossairt *et al.*, 2010]. That is, we can write the light field of a point source propagating through an aberrated lens as

$$l(r, \rho) = \frac{1}{\pi} \sqcap (\rho) \frac{\delta(r - T(\rho))}{\pi |r|}, \tag{4.8}$$

where we use a slightly different definition of the tophat function

$$\sqcap(\rho) = \begin{cases} 1 & \text{if } |\rho| < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

The advantage of the light field representation is that the PSF can be found by integrating over the aperture coordinates. We consider the general monomial OPD $W(\rho) = \alpha/(n+1)\rho^{n+1}$ which leads to the ray-aberration curve $T(\rho) = \alpha\rho^n$. We note that taking the modulus of the radial coordinate inside the ray aberration curve so that $T(\rho) = \alpha|\rho|^n$ does not alter the PSF. The Point Spread Function (PSF) of the lens can then be written as (for a derivation see Appendix B)

$$\begin{aligned} h(r) &= \pi \int_{-\infty}^{\infty} l(r,\rho)|\rho| d\rho && (4.10) \\ &= \frac{1}{\pi n \alpha^{2/n}} \sqcap\left(\frac{r}{\alpha}\right) |r|^{2/n-2}. && (4.11) \end{aligned}$$

The PSF can be shown to be unit normalized so that the integral of the PSF over sensor coordinates is equal to 1 (see Appendix B). The PSF for a lens with spherical aberrations is then written as

$$h(r) = \frac{3}{2\pi\alpha^{2/3}} \sqcap\left(\frac{r}{\alpha}\right) |r|^{-4/3}. \tag{4.12}$$

4.5 Aberrations and Resolution Scaling Laws

4.5.1 The Classical Aberration Limit to Resolution

For a diffraction limited lens, the SBP increases quadratically with the scaling factor M. However, the SBP of a lens also depends on the diameter of the blur circle caused by geometric aberrations. We introduce the variable δ_g, which represents the geometric spot size at lens scale $M = 1$, which we recall corresponds to a focal length of $f_l = 1mm$. Lohmann argues that the combined blur area when diffraction and aberration are taken into account can be expressed as the sum $\delta_d^2 + \delta_g^2$. Since geometric blur increases linearly with the scaling factor M, the SBP becomes [Lohmann, 1989]

$$R_{geom}(M) = \frac{M^2 \Delta x \Delta y}{(\lambda F/\#)^2 + M^2 \delta_g^2}. \tag{4.13}$$

In this case, the SBP plateaus at $\Delta x \Delta y / \delta_g^2$ when the lens is no longer diffraction limited and $M\delta_g >> \lambda F/\#$ (see the green curve in Figure 4.3). For this reason, lens designers typically seek to balance lens aberrations in an effort to minimize the blur circle. For example, defocus can be introduced into a lens with spherical aberrations in order to reduce the geometric blur circle. From a classical perspective, this strategy increases resolution because it decreases the spot size of the lens. As we will show in Section 4.6 however, this strategy is not desirable from a computational imaging perspective because it reduces the conditioning of the PSF, introducing more deblurring error.

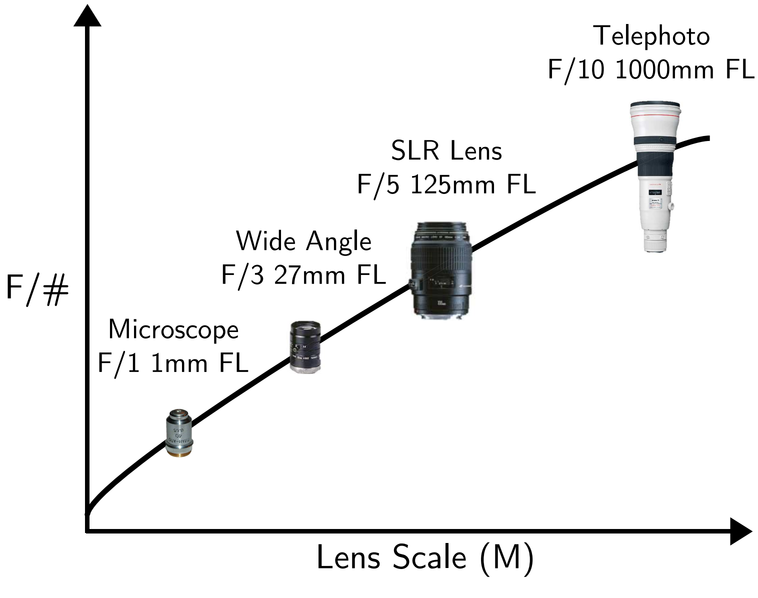


Figure 4.6: For conventional lens designs, the F/# typically scales with the cube root of the focal length in millimeters.

4.5.2 The Scaling Law for Conventional Lens Design

The geometric blur size can always be decreased by stopping down a lens. As a result lens designers typically increase the F/# as a lens is scaled up. A general rule of thumb is that the F/# is increased such that the focal length in mm is approximately equal to $(\mathrm{F}/\#)^3$. Many commercially available lenses follow this general trend (see Figure 4.6). For instance, the $500mm$ focal length Schneider Apo-Symmar operates at F/8, and $8^3 \approx 500$. This heuristic F/# scaling law has a special significance for lenses with spherical aberration. Then the geometric blur size δ_g is proportional to the spherical aberration coefficient α, and from Equation 4.7

$$\alpha = \frac{\sigma_I}{64}\frac{D}{F/\#^2} = \frac{\sigma_I}{64}\frac{f}{F/\#^3}. \tag{4.14}$$

Thus, if the F/# increases with the cube root of the focal length, the geometric blur size δ_g becomes independent of the scaling factor M. However, the diffraction blur size now increases as a function of scale so that $\delta_d = \lambda M^{1/3}$. Then (see the blue curve in Figure 4.3) the SBP becomes [Lohmann, 1989]

$$R_{conv}(M) = \frac{M^2 \Delta x \Delta y}{\lambda^2 M^{2/3} + \delta_g^2}. \tag{4.15}$$

Equation 4.15, derived by Lohmann, is a scaling law that tells us generally how SBP increases with lens size for a conventional lens design. The equation says that when M is large, the diffraction spot size dominates geometric blur. In this regime, the scaling follows the behavior:

$$R_{conv}(M) \propto M^{4/3}, \tag{4.16}$$

which overcomes the resolution threshold set by the aberration limit, but does not attain the ideal M^2 behavior of the diffraction limited scaling law.

4.6 Computational Imaging

We now revisit the imaging equation introduced in Chapter 1. To recap, conventional optical systems are based on the centuries old tradition of modeling optical systems as isomorphic mappings between scene radiance and pixel intensity. In a conventional camera, it is assumed that the brightness measured at a single pixel corresponds directly to the radiance of a single scene point. In the computational imaging paradigm, the imaging system can be written as a system of linear equations that relate the unknown signal coefficients $\mathbf{f}$ to the measurements made at each pixel $\mathbf{g}$. In Chapter 1 we discussed image formation in the absence of noise. Here we alter the image formation equation slightly to include noisy measurements

$$\mathbf{g} = H\mathbf{f} + \eta, \tag{4.17}$$

where $\mathbf{g} \in \mathrm{R}^M$ is a vector consisting of the M measured pixel measurements, H is an $M \times N$ matrix, $\mathbf{f} \in \mathrm{R}^N$ is a vector of N unknown signal coefficients, and $\eta \in \mathrm{R}^M$ is a vector representing the noise measured at each pixel, typically assumed to be gaussian so that $\eta \sim \mathcal{N}(0, \sigma_n^2 I)$. In the context of high resolution imaging, the vector of unknown signal coefficients $\mathbf{f}$ is a discretization of the continuous radiance distribution representing a latent focused image. We assume that the imaging system is non-compressive so that $M = N$.

In the analysis that follows, we assume the optical system is shift invariant, in which case the observation can be modeled as a convolution between the lens PSF and the unknown scene radiance. Convolution can be expressed compactly in the Fourier domain as the product between the Fourier transform of the PSF, referred to as the Optical Transfer Function (OTF), and the Fourier transform of the scene radiance. In our discreet framework, we denote the PSF by the vector $\mathbf{h}$ and the OTF by the vector $\hat{\mathbf{h}} = \mathrm{F}\mathbf{h}$, where F is the Fourier matrix. Under the assumption of periodic boundary conditions, the matrix A becomes a cyclic matrix such that $H_{i,j-i} = h_i$ with the special property that it can be written as $H = \bar{\mathrm{F}}\Lambda\mathrm{F}$, where Λ is a diagonal matrix and $\Lambda_{ii} = \hat{h}_i$, and the $\bar{}$ operator denotes complex conjugate. There is a slight abuse of notation here, because, for a 2D blur kernel, H is actually block-cyclic and diagonalized by the 2D Fourier matrix $\mathrm{F}_{2\mathrm{D}} = \mathrm{F} \bigotimes \mathrm{F}$, where $\bigotimes$ is the Kronecker product.

The image formation equation can be written as a sparse set of linear equations in the Fourier domain:

$$\hat{\mathbf{g}} = \Lambda\hat{\mathbf{f}} + \hat{\eta}, \tag{4.18}$$

where the ˆ operator denotes multiplication with the Fourier matrix F.

4.6.1 Image Deblurring

In the conventional imaging paradigm, pixel measurements correspond directly to scene radiance values. In the computational imaging paradigm, the unknown image $\mathbf{f}$ is blurred by the matrix H. To deblur the captured image $\mathbf{g}$ we must invert Equation 4.17. If the PSF is well conditioned, then the OTF contains no zero crossings and the matrix H is full rank and invertible, and we can estimate the unknown radiance $\mathbf{f}^*$ as

$$\hat{\mathbf{f}}^* = \Lambda^{-1}\hat{\mathbf{g}}. \tag{4.19}$$

Equation 4.19 is a sparse set of linear equations such that the estimate $\mathbf{f}^*$ is found simply by taking the ratio of Fourier coefficients

$$\hat{f}_i^* = \hat{g}_i/\hat{h}_i. \tag{4.20}$$

The final estimate can then be found by simply taking an inverse Fourier Transform. Unfortunately, we cannot recover the unknown image exactly because the original measurements were corrupted by noise. In order to quantify the quality of the deblurred image, we use the mean squared deblurring error σ_d^2 as a metric, which is defined as the expected mean squared difference between the deblurred image $\mathbf{f}^*$ and the ground truth image $\mathbf{f}$. σ_d^2 measures the variance of noise artifacts induced by the deblurring process. In our shift invariant system, this can be written as

$$\sigma_d^2 = \frac{1}{N}E[\|\mathbf{f}^* - \mathbf{f}\|^2] \tag{4.21}$$

$$= \frac{\sigma_n^2}{N}\sum_{i=1}^{N}\frac{1}{\|\hat{h}_i\|^2}, \tag{4.22}$$

where E denotes taking the expectation with respect to the noise η. Equation 4.22 says that, when naive deblurring is applied, the deblurring error is a product between the noise variance and the average squared reciprocal of the OTF.

4.6.2 Spherical Aberrations and Deblurring

In Section 4.4.2 showed that the spherical aberration coefficient α scales linearly with lens size, and we derived the analytic expression for the PSF of a lens with spherical aberrations, given by Equation 4.12. From this expression, we can derive the OTF of the lens. As discussed in Chapter 2, for a radially symmetric PSF $h(r)$, the OTF $\hat{h}(q)$ can be found by applying the zero order Hankel transform:

$$\hat{h}(q) = 2\pi \int_0^\infty J_0(qr)h(r)rdr, \tag{4.23}$$

where $J_0(r)$ is the zero-order Bessel function of the first kind. For the PSF given by Equation 4.12, the OTF becomes

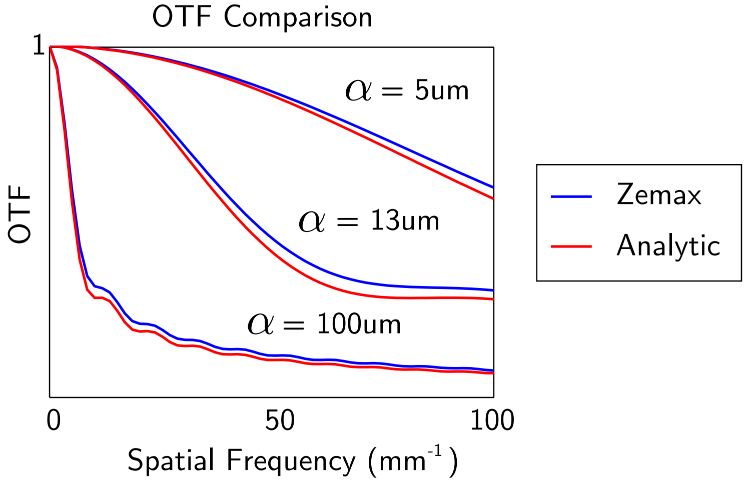


Figure 4.7: A comparison of the OTF for a lens with spherical aberration calculated using Zemax (the blue curves) and using our analytic formula (red curves). The OTF is calculated at various lens scales corresponding to spherical aberration coefficients of $\alpha = \{5\mu m, 13\mu m, 100\mu m\}$

$$\hat{h}(q) = \frac{2}{\alpha^{2/3}} \int_0^{\alpha} J_0(qr) r^{-1/3} dr \tag{4.24}$$

$$= {}_1F_2(\{\frac{1}{3}\}, \{1, \frac{4}{3}\}, -\frac{\alpha^2 q^2}{4}), \tag{4.25}$$

where ${}_1F_2(a; b, c; d)$ is the Generalized Hypergeometric Function [Slater, 1966]. Figure 4.7 shows a comparison between the OTF calculated analytically using Equation 4.24 and the OTF calculated numerically using the Geometric MTF feature in Zemax Optical Design Software [Zemax, 2010]. The OTF is calculated at a variety of lens scales corresponding to spherical aberration coefficients $\alpha = \{5\mu m, 13\mu m, 100\mu m\}$, and the results are highly consistent in all cases.

With an equation for the OTF, it is possible to derive an analytic expression for the deblurring error. In the continuous domain, the deblurring error from Equation 4.22 becomes

$$\sigma_d^2 = \frac{2\sigma_n^2}{\Omega^2} \int_0^{\Omega} \frac{1}{\|\hat{h}(q)\|^2} q dq, \tag{4.26}$$

where the signal is assumed to be bandlimited by the nyquist frequency Ω. Unfortunately, there is no closed form solution for the expression in Equation 4.26 after substituting the Hypergeometric function, so we instead approximate the OTF using the following equation:

$$\hat{h}(q) = \frac{2}{\alpha^{2/3}} \int_0^{\infty} J_0(qr) r^{-1/3} dr \tag{4.27}$$

$$= \frac{2\,\Gamma(7/6)}{\sqrt{\pi}\alpha^{2/3}}, \tag{4.28}$$

where Γ is the gamma function. Equation 4.27 essentially approximates the PSF as having infinite support, which is accurate for large amounts of spherical aberration, but decreases in accuracy as the spherical aberration approaches zero. Figure 4.8 shows a comparison of the OTF calculated using using our analytic formula (red curves) and using the approximation for the OTF given by Equation 4.27. The OTF is calculated at various lens scales corresponding to spherical aberration coefficients of $\alpha = \{20\mu m, 50\mu m, 200\mu m\}$. As the amount of spherical aberrations increase, the approximation increases in accuracy.

Substituting the approximate MTF from Equation 4.27 into the expression in Equation 4.26 gives us an analytic expression for the deblurring error:

$$\sigma_d = \sigma_n \sqrt{\frac{3\pi}{2} \frac{(\Omega\alpha)^{2/3}}{2\,\Gamma(7/6)}}. \tag{4.29}$$

Since we know from Equation 4.7 that scaling a lens by a factor of M also scales α by the same factor, Equation 4.29 gives us the relation

$$\sigma_d = k\sigma_n M^{2/3} \tag{4.30}$$

where k is a constant. Equation 4.30 expresses a remarkable fact: for lenses with spherical aberrations, while the size of the PSF increases linearly with lens scale M, the deblurring error increases *sub-linearly.* While Equation 4.30 is based on an approximation of the geometric OTF, it closely approximates the deblurring error calculated numerically using the OTF from Equation 4.24 (see Figure 4.9).

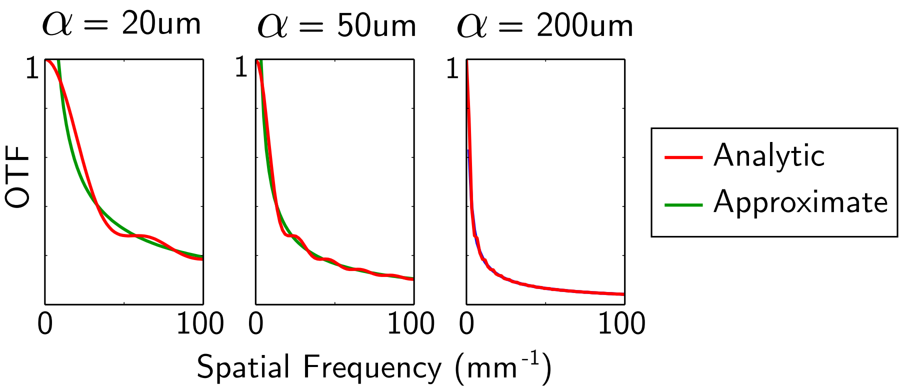


Figure 4.8: A comparison of the OTF for a lens with spherical aberration calculated using using our analytic formula (red curves) and using the approximation for the OTF given by Equation 4.27. The OTF is calculated at various lens scales corresponding to spherical aberration coefficients of $\alpha = \{20\mu m, 50\mu m, 200\mu m\}$. As the amount of spherical aberrations increase, the approximation increases in accuracy.

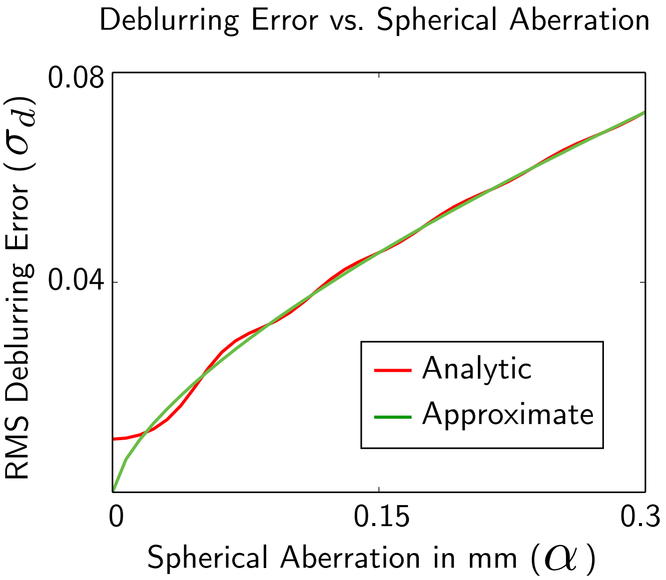


Figure 4.9: A comparison of the RMS deblurring error σ_d as a function of the spherical aberrations coefficient (α) with sensor noise $\sigma_n = .01$ and nyquist frequency $\Omega = 100mm^{-1}$. The red curve shows the error computed numerically using Equations 4.24 and 4.26. The green curve is calculated using the closed form expression for deblurring error given in Equation 4.29. The green curve closely approximates the green curve, with accuracy increasing as α increases.

4.7 A Scaling Law for Computational Imaging

4.7.1 Deblurring Error vs. Resolution

For the scaling laws given in Section 4.5, it is assumed that the minimum resolvable spot size is equal to the blur size due to geometric aberrations, δ_g. For a computational imaging system (i.e., with deblurring), the resolution is given by the pixel size ξ, and SBP does not depend directly on the geometric blur radius δ_g. A more pertinent quantity for measuring image quality is SNR. In the absence of any noise we can theoretically increase SBP by decreasing pixel size until we have reached the diffraction limit. In order to provide a fair comparison between any two computational imaging systems, we must fix the SNR.

By fixing SNR, we establish a relationship between the deblurring error and pixel size.

To show this, we express deblurring error as a function of lens scale M. Assuming the deblurring error is proportional to sensor noise, we can write

$$\sigma_d = \sigma_n s(M), \tag{4.31}$$

where $s(M)$ represents the scale-dependent deblurring factors. In order to force the SNR to remain constant across lens scale, we must adjust the sensor noise appropriately.

We now relate pixel size ξ to sensor noise σ_n. Here we assume that pixels receive sufficient light such that poisson noise dominates. Then the measurement noise can be well approximated by additive gaussian noise with variance proportional to the mean signal intensity [Chakrabarti *et al.*, 2010]. Scaling ξ by a factor of M increases the pixel's area by a factor of M^2. For a fully saturated pixel, assuming a shot noise limited sensor, this will increase the sensor's full well capacity by M^2 and decrease noise by a factor of $1/M$ relative to the signal. The sensor noise is then inversely proportional to pixel size so that

$$\xi(M) \propto \frac{1}{\sigma_n(M)}. \tag{4.32}$$

Equation 4.32 says that in order to make SNR scale independent, the pixel size should be increased as a function of M to exactly cancel out scale-dependent deblurring factors. The number of resolvable points for a computational imaging systems is then

$$R_{comp}(M) = \frac{M^2 \Delta x \Delta y}{(\lambda F/\#)^2 + \xi(M)^2}. \tag{4.33}$$

4.7.2 An Analytic Scaling Law

Using the expression for deblurring error for a lens with spherical aberrations given by Equation 4.30, we see that in order to produce a SNR that is independent of lens scale, the pixel size should be scaled according to the relation $\xi \propto M^{2/3}$. Plugging this into Equation 4.33 gives an analytic scaling law for computational imaging systems:

$$R_{ana}(M) = \frac{M^2 \Delta x \Delta y}{(\lambda F/\#)^2 + k_2^2 M^{4/3}} \tag{4.34}$$

where we have gathered proportionality constants in to k_2. For large M, the scaling law has the behavior

$$R_{ana}(M) \propto M^{2/3}. \tag{4.35}$$

As with conventional lens design curve R_{conv}, Equation 4.34 gives a scaling law that breaks the resolution threshold imposed by the aberrations limit (see the magenta curve in Figure 4.11). However, the analytic scaling law does not behave as close to the ideal diffraction limited scaling law as the R_{conv} curve. At the same time, the R_{conv} curve assumes that F/# reduces and more light is sacrificed as scale increases, while the R_{ana} curve does not make this assumption.

4.7.3 Image Priors for Improved Performance

In the previous section we showed analytically that, when a computational approach is taken, the resolution of a lens with spherical aberrations breaks the classical limit that results when considering geometrical spot size alone. The R_{ana} curve given in Equation 4.34, however, does not increase as rapidly with lens scale as does Lohmann's scaling law for conventional lens designs. We now show that the scaling behavior of computational imaging systems surpasses that of conventional lens designs when image priors are taken into account.

In Section 4.6.1 we used Equation 4.19 to form an estimate of our unknown image. This solution can be seen to be equivalent to the solution found by maximizing the likelihood for the probability distribution [Bertero and Boccacci, 1998]

$$h(\hat{\mathbf{g}}|\hat{\mathbf{f}}) = \exp \|\hat{\mathbf{g}} - \Lambda\hat{\mathbf{f}}\|^2. \tag{4.36}$$

The maximum likelihood solution minimizes the probability of error in the estimate when no information about the the prior distribution $h(\hat{\mathbf{f}})$ is available a priori. In our case however, some information about $h(\hat{\mathbf{f}})$ is known ahead of time since the unknown quantity $\hat{\mathbf{f}}$ belongs to the class of natural images. To make a solution to the estimation problem analytically tractable, we assume a linear distribution on Fourier coefficients of natural images taking the form $h(\hat{\mathbf{f}}) = \exp\|B\hat{\mathbf{f}}\|^2$, where B is a diagonal matrix. We define the vector of Fourier coefficients $\hat{\mathbf{b}}$ such that $B_{ii} = \hat{b}_i$. Given a prior distribution, the maximum a posteriori solution minimizes the probability of error in the estimate. The estimate then becomes

$$\hat{\mathbf{f}}^* = \arg\max_{\mathbf{f}} h(\hat{\mathbf{g}}|\hat{\mathbf{f}})h(\hat{\mathbf{f}}) \tag{4.37}$$

$$= \arg\max_{\mathbf{f}} (\|\hat{\mathbf{g}} - \Lambda\hat{\mathbf{f}}\|^2 + \|B\hat{\mathbf{f}}\|^2) \tag{4.38}$$

$$= (\Lambda^2 + B^2)^{-1}\Lambda^t\hat{\mathbf{g}}, \tag{4.39}$$

which can be written as the set of linear equations

$$\hat{f}_i^* = \frac{\bar{\hat{h}}_i}{\|\hat{h}_i\|^2 + \|\hat{b}_i\|^2}\hat{g}_i, \tag{4.40}$$

We define the average power spectrum $\hat{\mathbf{a}}$ such that $\hat{a}_i = E[\|\hat{f}_i\|^2]$, where the expectation is taken with respect to the set of natural images. Then, as Zhou and Nayar showed, the optimal vector $\hat{\mathbf{b}}$ is such that $\hat{b}_i = \sigma_n^2/\hat{a}_i$, and the squared deblurring error becomes [Zhou and Nayar, 2009]

$$\sigma_d^2 = \sigma_n^2 \sum_{i=1}^{N} \frac{1}{\|\hat{h}_i\|^2 + \sigma_n^2/\hat{a}_i}. \tag{4.41}$$

Figure 4.10 shows the deblurring error σ_d calculated using Equations 4.24 and 4.41. σ_d is shown as a function of spherical aberration α for a variety of sensor noise levels in the range $\sigma_n = [.002, .1]$. A polynomial is fit to each curve, and the best fit is found to be in the range $\sigma_d \propto \alpha^{1/3.4}$ to $\sigma_d \propto \alpha^{1/4.2}$. We approximate the deblurring error as

$$\sigma_d \propto \sigma_n \alpha^{1/3.8}. \tag{4.42}$$

$$\propto \sigma_n M^{1/3.8}. \tag{4.43}$$

In fact, this estimate is slightly pessimistic, as the deblurring error also increases sub-linearly with σ_n as well as α. From Equations 4.43 and 4.33, we conclude that when image priors

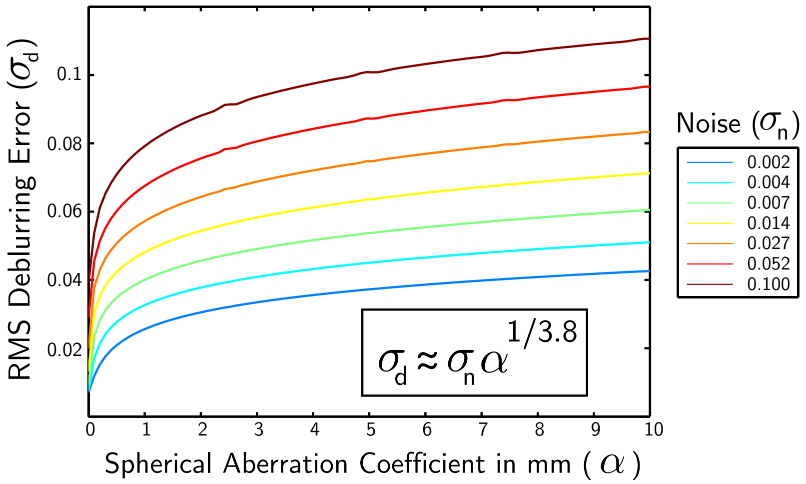


Figure 4.10: RMS deblurring error as a function of spherical aberration (α). As α increases, both the PSF size and the deblurring error increase. While the size of the PSF increases linearly with α, deblurring error increases with $\alpha^{1/3.8}$. In this experiment, the nyquist frequency $\Omega = 250mm^{-1}$.

are used for deblurring, the resolution of a computational imaging system obeys the scaling law given by (see the cyan curve in Figure 4.11)

$$R_{prior}(M) = \frac{M^2 \Delta x \Delta y}{(\lambda F/\#)^2 + k_3^2 M^{2/3.8}}, \tag{4.44}$$

where again we have gathered proportionality constants into k_3. While the analytic scaling law curve R_{ana} does not scale as quickly as the conventional lens design curve R_{conv}, the curve R_{prior} scales more quickly. From this we conclude that in building a camera at a desired resolution, when image priors are taken into account, a computational camera can be built at a smaller scale than a conventional lens design. Again, the R_{conv} curve assumes that F/# reduces and more light is sacrificed as scale increases, while the R_{prior} curve does not make this assumption.

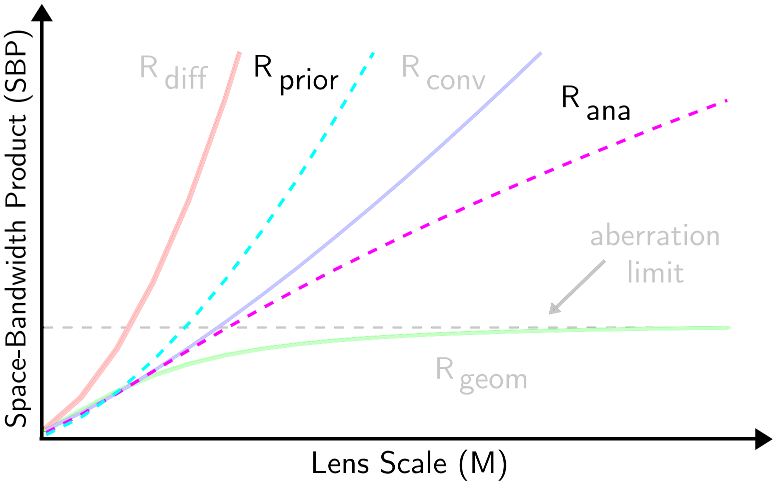


Figure 4.11: Scaling laws for computational imaging systems with spherical aberrations. The R_{ana}, which was analytically derived, shows an improvement upon the aberration limited curve R_{geom}, *without* requiring F/# to increase with M. Performance is further improved when natural image priors are taken into account, as the R_{prior} curve shows. The R_{prior} curve improves upon the conventional lens design curve R_{conv}, also without requiring F/# to increase with M.

4.8 Gigapixel Computational Cameras

According to Equation 4.44, a computational imaging approach can enable a greater resolution to be achieved with a smaller camera size. To demonstrate this principle, we show results from a proof of concept camera that utilize a very simple optical element. By using a large ball lens, an array of planar sensors, and deconvolution as a post processing step, we are able to capture gigapixel images with a very compact camera.

The key to our architecture lies in the size of the sensors relative to the ball lens. Together, a ball lens and spherical image plane produce a camera with perfect radial symmetry. We approximate a spherical image plane with a tessellated regular polyhedron, such as an icosahedron. A planar sensor is placed on each surface of the polyhedron. Note that because sensors are typically rectangular, a different polyhedron, such as a truncated icosahedron, may provide more optimal sensor packing. Relatively small sensors are used so that each sensor occupies a small FOV and the image plane closely approximates the spherical surface. As a result, our camera produces a PSF that is not completely spatially invariant, but comes within a close approximation.

4.8.1 A Proof-of-Concept Gigapixel Camera

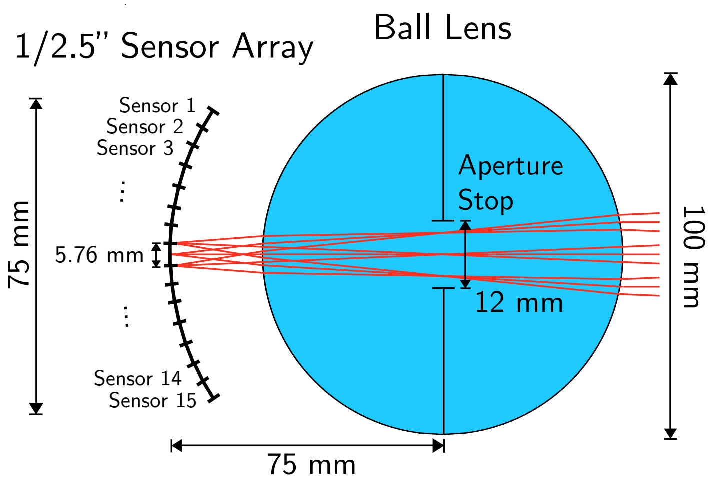
The first system we demonstrate consists solely of a ball lens and an array of planar sensors. We use a 100mm acrylic ball lens and a 5 megapixel 1/2.5" Lu575 sensor from Lumenera [Lumenera, 2010] (see Figure 4.12(a)). We emulate an image captured by multiple sensors by sequentially scanning the image plane using a pan/tilt motor. With this camera, a 1 gigapixel image can be generated over a roughly 60^ox40^o FOV by tiling 14x14 sensors onto a 75mmx50mm image surface. When acquiring images with the pan/tilt unit, we allow a small overlap between adjacent images.

The PSF as a function of field position on each individual sensor is shown in Figure 4.12(b). Note that the PSF shape remains fairly consistent across the FOV of each sensor. The MTF (shown in in Figure 4.12(c)) avoids zero crossings up to the Nyquist frequency of the sensor. The plots were generated using Zemax Optical Design Software [Zemax, 2010].

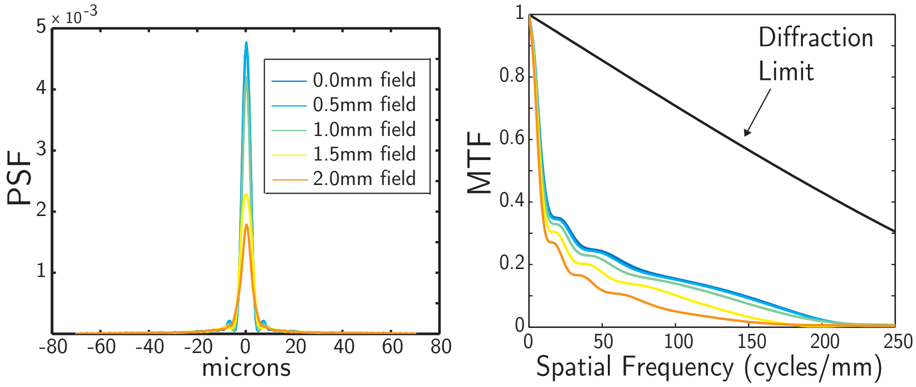
An implementation of this design is shown in Figure 4.13. Figures 4.2 , 4.14, and 4.16 show two gigapixel images captured with this system. Note the remarkable level of detail captured in each of the photographs. Zooming in to Figure 4.2 reveals the label of a resistor on a PCB board, the stippling print pattern on a dollar bill, a miniature 2D barcode pattern, and the extremely fine ridges of a fingerprint. Closeups in Figure 4.14 reveal fine details in a watch, an eye, a resolution chart, and individual strands of hair. Closeups in Figure 4.16 reveal details that are completely invisible in the zoomed out panorama, including a sailboat, a sign advertising apartments for sale, the Empire State Building, and cars and trucks driving on a bridge

4.8.1.1 Color

Because our cameras do not include any color correcting elements, they suffer from axial chromatic aberrations. For our 100mm diameter ball lens that we use, the chromatic



(a) An F/4 75mm focal length ball lens system.



(b) The system PSF of (a)

(c) The system MTF of (a)

Figure 4.12: (a) Our single element gigapixel camera, which consists solely of a ball lens with an aperture stop surrounded by an array of planar sensors. (b) Because each sensor occupies a small FOV, the PSF is nearly invariant to field position on the sensor. (c) The PSF is easily invertible because the MTF avoids zero crossings and preserves high frequencies.

focus shift is about 1.5mm over the visible wavelength range. However, most of the image blur caused by chromatic focus shift is in the chrominance channel of captured images [Guichard *et al.*, 2009] [Cossairt and Nayar, 2010]. Since humans are less sensitive to blur in chrominance channels, axial chromatic aberrations do not cause a significant degradation in perceived image quality. We use the deblurring technique from Cossairt and Nayar [Cossairt and Nayar, 2010], which is inexact but produces images that look good.

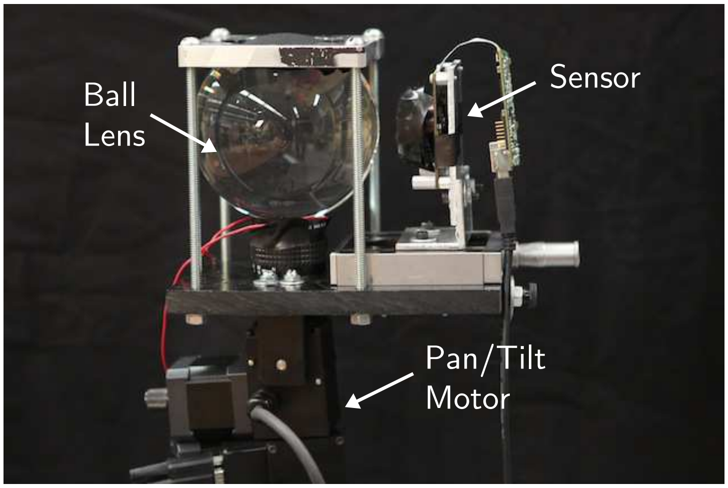


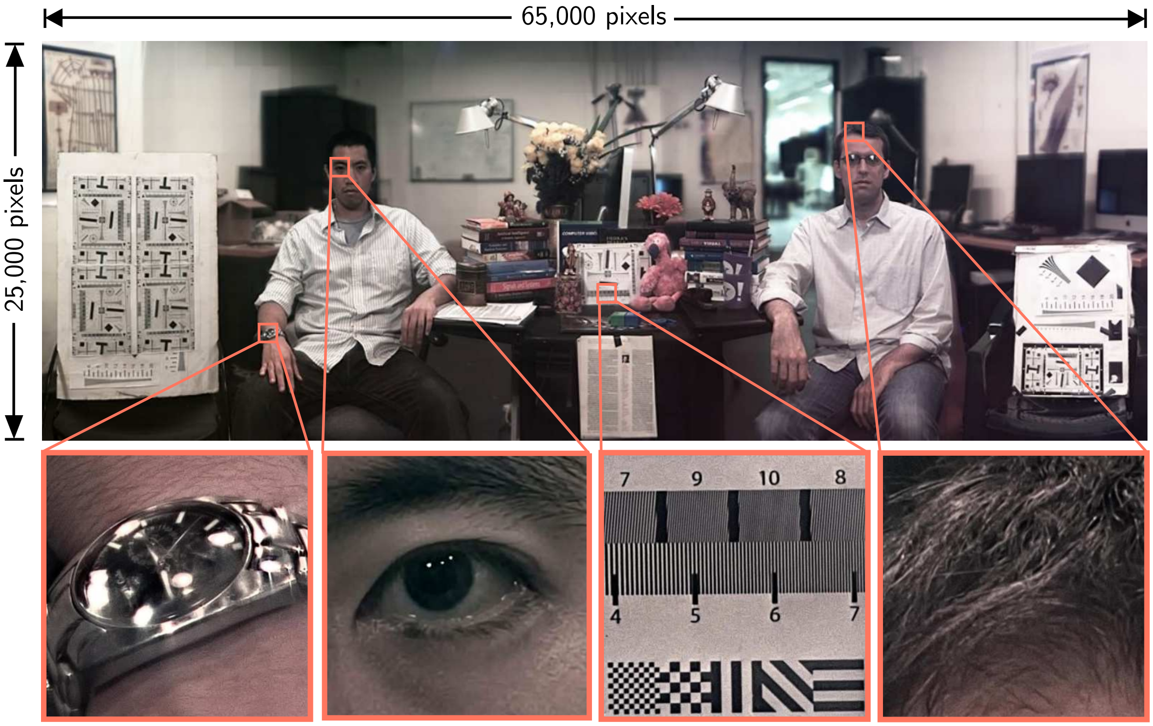
Figure 4.13: A system used to verify the performance of the design shown in Figure 4.12(a). An aperture is placed on the surface of the ball lens. A gigapixel image is captured by sequentially translating a single 1/2.5”, 5 megapixel sensor with a pan/tilt motor. A final implementation would require a large array of sensors with no dead space in between them.

4.8.1.2 Post Processing

The post processing for captured images follows several steps. First, a transformation from RGB to YUV color space is applied. Next, Wiener deconvolution is applied to the luminance channel only, and the image is transformed back to RGB color space. A noise reduction algorithm is then applied to suppress deblurring artifacts. We found the BM3D algorithm [Dabov *et al.*, 2006] to produce the best results. Finally, the set of captured images are stitched to obtain a high resolution image using the Microsoft Image Composite Editor [ICE, 2010].

4.8.2 A Single Element Design

The design in Figure 4.12(a) is extremely compact, but impractical because adjacent sensors must be packed without any dead space in between them. The size of this system is limited by the package size of the sensor relative to the active sensor area. Sensors with a package size that is only 1.5x larger than the active sensor area are currently commercially available. With these sensors, it is possible to build a gigapixel camera that uses only a single optical



Watch Eye Resolution Chart Hair

Figure 4.14: **A 1.6 gigapixel image** captured using the implementation shown in Figure 4.13. The image dimensions are 65,000 × 25,000 pixels, and the scene occupies a 104° × 40° FOV. From left to right, the insets reveal fine details in a watch, an eye, a resolution chart, and individual strands of hair.

element, as shown in Figure 4.15(a). In this design, each sensor is coupled with a smaller acrylic relay lens that decreases the focal length of the larger acrylic ball lens. The relay lenses share a surface with the ball lens, which means that it is possible to combine the entire optical system into a single element that may be manufactured by molding a single material, drastically simplifying the complexity (and hence alignment) of the system.

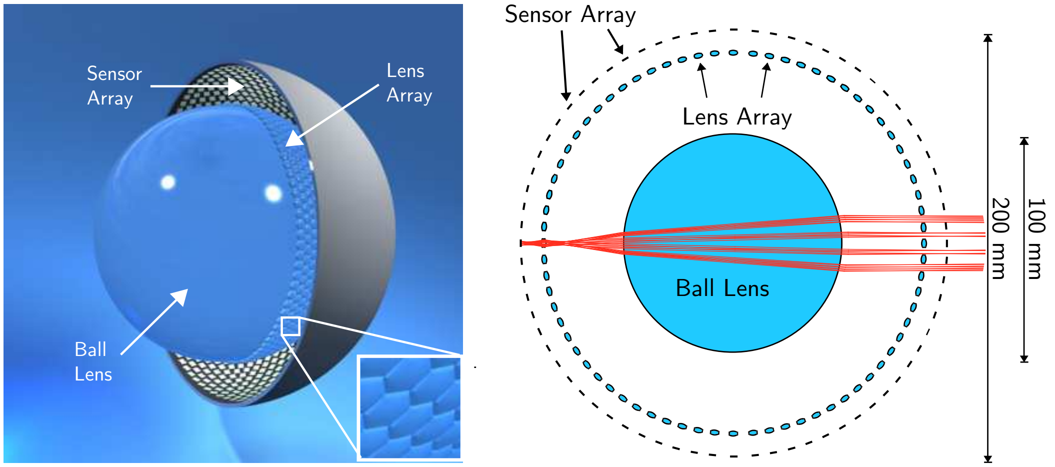
4.8.3 Capturing the Complete Sphere

A great advantage of using a ball lens is that, because it has perfect radial symmetry, a near hemispherical FOV can be captured. In fact, it can even be used to capture the complete sphere, as shown in Figure 4.15(b). This design is similar to the one in Figure 4.15(a) with a large gap between adjacent lens/sensor pairs. Light passes through the gaps on one hemisphere, forming an image on a sensor located on the opposite hemisphere. As a result, the sensors cover the complete 2π FOV at the cost of losing roughly half the incident light.

4.9 Discussion

4.9.1 Limitations of Scaling Laws

In Sections 4.5 and 4.7, we derived scaling laws which express the the general scaling behavior of resolution versus lens scale M, with special attention paid to how the behavior for increasingly large values of M. However, because we have chosen to speak in general terms about the scaling behavior, we have not given attention to how resolution behaves for smaller values of M, which may result in different behavior. For instance, when M is large,



(a) A single element design

(b) A 4π FOV design

Figure 4.15: (a) A single element design for a gigapixel camera. Each sensor is coupled with a lens that decreases focal distance, allowing FOV to overlap between adjacent sensors. (b) A design for a gigapixel camera with a 2π radian FOV. The design is similar to the implementation in Figure 4.15(a) with a large gap between adjacent lens/sensor pairs. Light passes through the gaps on one hemisphere, forming an image on a sensor located on the opposite hemisphere.

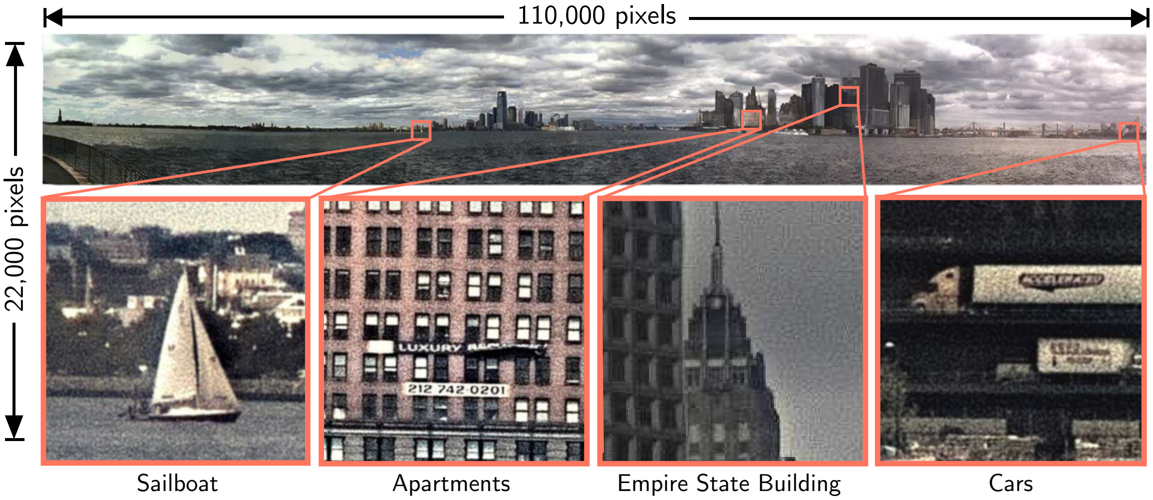


Figure 4.16: **A 1.4 gigapixel image** captured using the implementation shown in Figure 4.13. The image dimensions are 110,000 × 22,000 pixels, and the scene occupies a 170° × 20° FOV. From left to right, insets reveal a sailboat, a sign advertising apartments for sale, the Empire State Building, and cars and trucks driving on a bridge.

conventional lens designs outperform computational imaging without priors, as indicated by the R_{conv} and R_{ana} curves. However, for small M, R_{ana} may actually be greater than R_{conv}, depending on the exact values of the proportionality constant k_1 and the amount of spherical aberration δ_g. These exact values will vary depending on the specific lens design and sensor characteristics, but the aggregate behavior for large values of M will will remain consistent across all scenarios. In this way, the scaling laws encompass the gross behavior of lenses and sensors, but do not always lend themselves to a direct comparison between specific designs.

4.9.2 On Computational Imaging and Scaling Laws

The original scaling laws derived by Lohmann are pleasingly simple in the sense that they keep the problem domain constrained to a single variable: the scale parameter M. In some sense, introducing computational imaging made the problem more complicated because it introduced a new variable in the form of SNR. Looking at the problem in a general way, the resolution scaling behavior of different imaging systems can vary both as a function of lens scale and SNR. While Lohmann made no mention of SNR in his original analysis, there was an implicit relationship between SNR and resolution that was unstated. For example, consider the expression for the scaling behavior of lenses in the presence of geometric aberrations given by Equation 4.13. We recall that, for large M, resolution plateaus at $\Delta x \Delta y/\delta_g$. However, if we choose to match pixel area to blur area, then pixel size increases linearly with M. Thus, according to the arguments in Section 4.7, if we continue to scale a lens beyond the aberration limit, resolution does not increase, while SNR increases linearly with M. On the other hand, for diffraction limited lenses, pixel size, and thus SNR, remains constant while resolution scales quadratically with lens scale. This leads to an interesting observation about the tradeoff between resolution and SNR. In some sense, these two examples are opposite extremes in a two-dimensional design space. When geometric aberrations are present, resolution becomes fixed but SNR can increase, while for diffraction limited lenses, SNR becomes fixed but resolution can increase.

This brings us to the scaling laws for conventional lens design and computational imaging. The conventional lens design curve, R_{conv}, is derived assuming that both F/# and pixel size increase with $M^{1/3}$. In the photon limited noise regime, SNR is proportional to pixel size ξ, and inversely proportional to F/#. Thus, while the R_{conv} curve is derived assuming that more light is sacrificed as lens scale increases, the amount of photons collected per pixel remains fixed, and thus so does SNR. Similarly, in the computational imaging regime, we ask what pixel scaling behavior will produce a deblurring error, and hence SNR, that is independent of lens scale.

The scaling laws for computational imaging and conventional lens design represent the behavior of two competing techniques that are trying to achieve the same goal: maximizing resolution scaling behavior while fixing SNR. Neither technique achieves the ideal scaling performance of diffraction limited lenses. In effect, both techniques are complexity reducing measures, since they aim to maximize performance without introducing the added optical elements required to reduce aberrations below the diffraction limit. This brings us to a third axis in our design space: lens complexity. As we scale a diffraction limited lens, SNR remains fixed and resolution reaches the maximum scaling potential, however lens complexity must also increase in an effort to combat greater amounts of geometrical aberrations. In

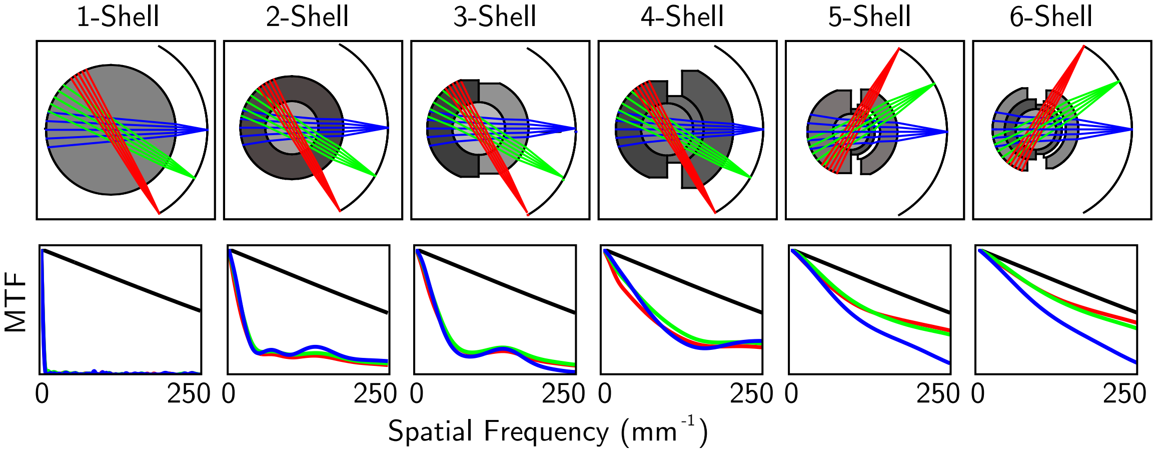


Figure 4.17: The MTF for spherical optical systems with varying amounts of complexity. Complexity is measured as the number of optical surfaces, which increases from left to right as 1 to 6 surfaces. The six surface design is the Gigagon lens designed by Marks and Brady. Each design is a F/2.8 280mm FL lens optimized using Zemax. As the number of surfaces increases, the MTF improves, improving the SNR as well.

contrast, for the computational imaging and conventional lens scaling laws, both SNR and lens complexity remain fixed, but the maximum scaling potential is not achieved.

In an ideal setting, we would like to maximize resolution and SNR while minimizing lens scale and complexity. This cannot be achieved in practice, however, and the best that can be done is to develop a merit function that weighs these measures in terms of their relative importance on an application dependent basis. Lens optimization based on this merit function then gives the design which results in the best performance for this specific application.

4.9.3 The Performance vs. Complexity Trade-off

According to Equation 4.44, with the aid of computations, the resolution of a lens with spherical aberrations will, in general, scale more quickly than for a conventional lens design. However, a lens which requires deblurring will have a smaller SNR than a diffraction limited lens of the same scale. For the designs proposed in Section 4.8, we have chosen designs that favor simplicity, and as a consequence, also result in a lower SNR. Any computational imaging system poses an inherent trade-off between complexity and SNR. In practice, exploring this trade-off requires a carefully designed measure for complexity.

A good complexity measure must take into account many different factors: the number of surfaces, the degree polynomial of each surface, etc. While it is difficult to develop a general measure for complexity that applies to all lens designs, the problem becomes much simpler when we consider only concentric spherical optical elements. In this case, complexity can simply be quantified as the number of surfaces used in the design.

To explore the tradeoff between complexity and SNR for the special case of spherical optics, we created six spherical optics designs, ranging in complexity from 1 shell to 6 shells. The six designs were created in an effort to analyze how the best case performance of a computational imaging system scales as a function of lens complexity. Shells 1-5 were optimized with Zemax using a custom optimization procedure that minimizes the deblurring

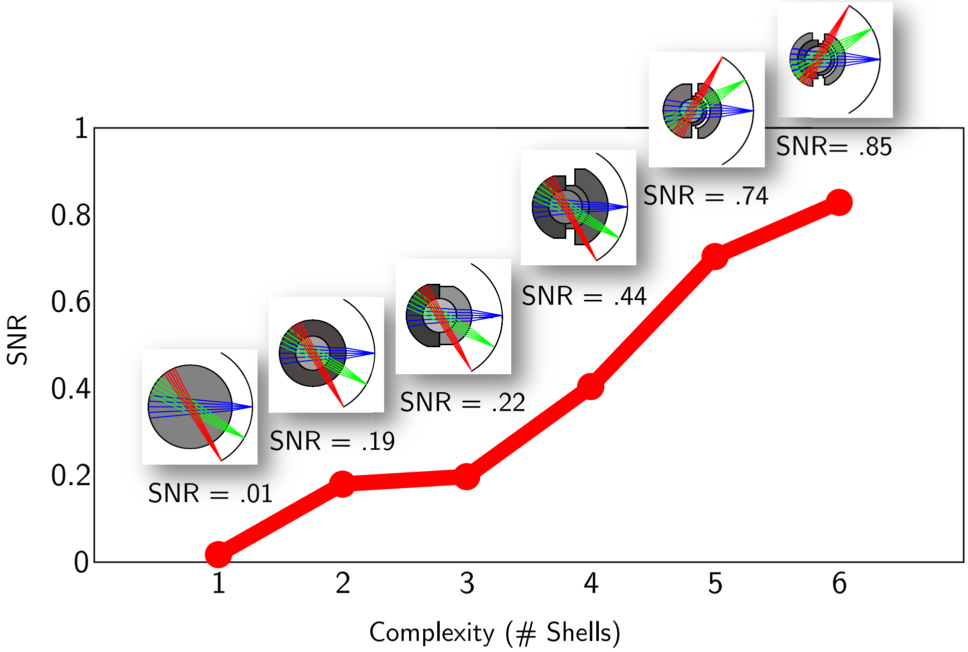


Figure 4.18: SNR vs. complexity for the lens designs shown in Figure 4.18, assuming a computational approach is taken. SNR increases by a factor of 19 when complexity increases from 1 shell to 2 shells, while SNR only increases by a factor of 4 when complexity increases from 2 shells to 6 shells.

error. The six shell design shown is the Gigagon lens designed by Marks and Brady [Marks and Brady, 2010]. The six designs and their relative performance are shown in Figure 4.17. From the MTF plots shown at the bottom of the figure, it can be seen that the six shell design performs near diffraction limited, and the MTF steadily decreases with decreasing complexity.

Figure 4.18 shows how, rather than favoring simplicity, an optimal design may consist of more elements than the designs discussed previously in this chapter. It appears that, for the special case of spherical optics, there is a law of diminishing returns when it comes to improving performance by means of increasing complexity. In particular, we note that SNR increases by a factor of 19 when complexity increases from 1 shell to 2 shells, while SNR only increases by a factor of 4 when complexity increases from 2 shells to 6 shells. Taking this behavior in to account, an optimal design may be found by balancing lens scale and complexity. Such a design would have minimum scale and complexity for a given resolution, and a fixed scale/complexity ratio. Whether or not general lens designs also have similar complexity vs. performance behavior is an open question that is currently under investigation.

4.10 Conclusion

In this chapter, we have given a comprehensive analysis on the tradeoff between the scale and resolution of a camera. We extended Lohmann's analysis to include the resolution scaling behavior of computational imaging systems, with special attention paid to lenses that exhibit spherical aberrations. Closed form expressions for the PSF, OTF, and deblurring error of lenses which exhibit spherical aberrations have been derived. In addition, we have shown that, when image priors are taken into consideration, computational imaging systems exhibit superior scaling performance with respect to conventional lens designs.

Our analysis on the scaling behavior of lenses brought us back to the tradeoff between performance and complexity for computational imaging systems. We explored the design of gigapixel computational imaging systems based on spherical optical elements, which primarily exhibit spherical aberrations. A proof-of-concept system that emulates a spherical lens surrounded by an array of planar sensors was demonstrated. Several examples were shown which evidence promising image quality with a compact camera that consists of only a single optical element. However, the performance of the camera can be improved by increasing lens complexity. We explored the tradeoff between performance and lens complexity, providing a complete exploration of the design space of resolution, SNR, lens scale and complexity for the special case of spherical optics. In the end we see that the tradeoff between scale and resolution is closely related to the tradeoff between performance and complexity.

This chapter concludes our discussion on the tradeoffs of computational imaging systems. In the last three chapters we have seen that there are a number of tradeoffs that exist for cameras that computationally extend DOF and increase resolution. However, these are only a few of the functionalities offered by computational imaging. In the next chapter we return to a general discussion on computational imaging, and focus exclusively on the topic of performance. We return to a comparison between computational and conventional techniques, and ask when computational imaging techniques provide a performance advantage.

Part II

Limits in Computational Imaging

Chapter 5

On the Limits of Computational Imaging

5.1 Introduction

In Chapters 2, 3, and 4, we discussed a number of tradeoffs that for computational imaging systems designed to extend DOF and increase optical resolution. In Chapter 4 we gave a thorough performance comparison between conventional and computational imaging systems. Here the message was clear: a computational approach always reduces performance, but with the benefit of also reducing complexity. We compared performance for the EDOF techniques introduced in Chapters 2 and 3, but the analysis was incomplete. We did not provide any conclusive answers about how the performance of an EDOF camera compares to stopping down a lens.

In this chapter we revisit the topic of performance for EDOF, and more generally, any imaging system that maps a linear combination of signal coefficients to each pixel measurement. We refer to this many-to-one mapping as a *coding* or *multiplexing* of the signal, and the process of inverting the mapping as *decoding* the signal. As we discussed in Chapter 1, coding is a very attractive option in many scenarios because it can be used in intelligent ways to increase the efficiency of the optical system, and increase the Signal-to-Noise Ratio (SNR) of captured images. Unless the coding is perfectly conditioned, however, the process of decoding amplifies noise, and to determine the true benefit of a coding technique we must weigh the benefit of increased efficiency vs. the disadvantage of increased noise. The purpose of this chapter is to explore the relationship between efficiency in the capture process and the conditioning of the coding process.

We have seen a variety of coding applications in the computer vision and optics communities dating back to early work on spectrometry and X-Ray astronomy in the 70's [Skinner, 1988] [Caroli *et al.*, 1987]. Coding techniques have been employed to capture various slices of the plenoptic function, and to measure various properties of light transport. Coded aperture Gamma Ray telescopes use MURA codes to capture images without refractive optics [Gottesman and Fenimore, 1989] [Villela *et al.*, 1995]. Mask-based Hadamard multiplexing has been used for spectrometers that capture a spectrum at a single spatial location [Harwit and Sloane, 1979], as well as for imaging spectrometers that capture irradiance as a function of both wavelength and 2D spatial locations [Hanley *et al.*, 1999]. Color Filter Arrays (CFAs) that multiplex wavelength data have been proposed to capture three color images with more efficiency [Baer *et al.*, 1999] [Ihrke *et al.*, 2010]. Masks placed either in the lens aperture, or near the sensor, have been used to capture light fields [Veeraraghavan *et al.*, 2007] [Lanman *et al.*, 2008] [Liang *et al.*, 2008]. Hadamard codes, as well as newly optimized codes, have been used to capture the appearance of objects under different lighting conditions [Schechner *et al.*, 2003] [Schechner *et al.*, 2007] [Ratner and Schechner, 2007] [Ratner *et al.*, 2007]. Coded aperture masks have been used to deblur images that exhibit defocus blur [Levin *et al.*, 2007] [Veeraraghavan *et al.*, 2007] [Zhou and Nayar, 2009] [Zhou *et al.*, 2011]. Temporal shuttering has been used to remove the effect of motion blur [Raskar *et al.*, 2006]. We refer to this family of coding methods as *multiplexing methods*, and we analyze the performance limits of this family in Section 5.2.

For nearly all examples of coding, there is a corresponding technique that measures the signal directly without the need for any decoding, which we refer to as conventional techniques. For example, a Bayer filter measures RGB colors directly, but less efficiently than a mask with with Cyan, Magenta and Yellow (CMY) filters. Narrowband spectral filters can be used in place of a Hadamard spectrometer. A pinhole array can be used in place of mask-based light field multiplexing techniques. A stopped down aperture can be used in place of a coded aperture, and so on. These conventional techniques are an important baseline for comparing performance. Multiplexing techniques offer an advantage in terms of increased optical efficiency, but whether or not they offer a performance advantage relative to conventional techniques depends on noise characteristics, which in turn depends on lighting conditions. In this chapter, we derive limits on the best possible performance advantage that can be achieved. We relate the performance advantage to different lighting scenarios, and we discover that there is a relatively small set of scenarios where an advantage can be achieved.

Multiplexing methods sacrifice some amount of optical efficiency by masking light in either wavelength, spatial or temporal dimensions. There are a number of coding techniques that do not sacrifice light. Superresolution techniques code high resolution data within multiple images with sub-pixel shifts by using either sensor motion or arrays of parallel imaging systems [Ben-Ezra *et al.*, 2004] [Ben-Ezra *et al.*, 2005] [Tanida *et al.*, 2000] [Tanida *et al.*, 2001]. These codes, however, become very poorly conditioned when trying to increase resolution by more than a factor of 2. We have already discussed a number of well conditioned codes that extend DOF without sacrificing light [Chi and George, 2001] [E. R. Dowski and Cathey, 1995] [Ojeda-Castaneda *et al.*, 2005] [Levin *et al.*, 2009] [Häusler, 1972] [Nagahara *et al.*, 2008] [Cossairt *et al.*, 2010] [Cossairt and Nayar, 2010] [Guichard *et al.*, 2009]. These techniques produce shift invariant blur that is independent of depth, and can be recovered via deconvolution. There are also analogous methods that create motion invariant motion blur that can be removed without sacrificing light [Levin *et al.*, 2008] [Cho *et al.*, 2010]. Both EDOF and motion invariant techniques aim to produce blur that is invariant over some domain (in this case either depth or motion). We discuss methods of *coding for invariance* in Section 5.3.

5.2 Multiplexing Methods

We return to the general form of multiplexing introduced in Chapter 1. As we did in this chapter, we use an image formation equation that relates a vector of signal coefficients to a vector pixel measurements. However, we want to discuss the performance of computational imaging systems that are corrupted by noise. Our image formation equation will be the same as used in Chapter 4

$$\mathbf{g} = H\mathbf{f} + \eta. \tag{5.1}$$

Here we take $\mathbf{g}$ to be the vector of pixel measurements (a vector in $\Re^N$), and $\mathbf{f}$ now represents the vector of unknown signal coefficients, which may represent a number of possible features contained within the plenoptic function. η is a per-pixel additive noise vector, and H is the multiplexing matrix, which may be shift invariant, or it may take a more general form. For now, we assume that the noise is i.i.d and sampled from a zero mean Gaussian distribution $\eta \sim \mathcal{N}(0, \sigma^2 I)$, where I is the identity matrix. We relax this assumption in Section 5.2.1, where we incorporate the effects of signal dependent noise. For noisy images, we measure image quality using the Signal-to-Noise Ratio (SNR) as a metric that quantifies the strength of a signal relative to the amount of noise present. The SNR measures the amount of variation in the signal due to noise (σ) relative to the average signal level (J)

$$SNR = \frac{\sigma}{J}. \tag{5.2}$$

For physical systems, the measurement matrix is usually restricted to be all positive

$$H_{i,j} > 0. \tag{5.3}$$

Assuming the matrix H is invertible, an estimate of the demultiplexed image can be found simply as

$$\begin{aligned} \mathbf{f}^* &= H^{-1}\mathbf{g}, && (5.4) \\ &= \mathbf{f} + H^{-1}\eta. && (5.5) \end{aligned}$$

Note that here we are assuming a linear decoding process. In many cases, the estimate $\mathbf{f}^*$ can be improved by using a non-linear estimation technique, especially when statistical information on the distribution of input signals is accessible ahead of time. We have more to say on this matter in Section 5.4, but for know we stick with the assumption that the estimation is linear to make analysis tractable.

We use the expected Mean Squared Error (MSE) as a metric to quantify the amount of errors in the estimate $\mathbf{f}^*$. This is consistent with the performance evaluations performed in Chapters 2, 3, and 4, where we used the MSE to quantify the amount of deblurring errors after deconvolution. The MSE will provide us with a consistent means of comparing performance for both general multiplexing techniques, as well as techniques that produce shift invariant blur, such as EDOF and motion invariant cameras.

The MSE is essentially a measurement of the noise variance in the estimate $\mathbf{f}^*$ so that we can write the SNR of the demultiplexed image as

$$SNR = \frac{\sqrt{MSE}}{J}. \tag{5.6}$$

The MSE for the estimate $\mathbf{f}^*$ is given by

$$\begin{aligned} MSE &= E\left[\frac{1}{N}\|\mathbf{f}^* - \mathbf{f}\|^2\right] && (5.7)\\ &= E\left[\frac{1}{N}\|H^{-1}\eta\|^2\right]. && (5.8) \end{aligned}$$

Defining the error vector $\mathbf{e} = H^{-1}\eta$, the MSE becomes

$$\begin{aligned} MSE &= E\left[\frac{1}{N}\sum_{i=1}^{N} e_i^2\right] && (5.9)\\ &= \frac{1}{N}\sum_{i=1}^{N} E\left[e_i^2\right]. && (5.10) \end{aligned}$$

Defining the matrix $\tilde{H} = H^{-1}$, the expected squared error for the i^{th} pixel becomes

$$E\left[e_i^2\right] = E\left[\sum_{j=1}^{N}(\tilde{H}_{i,j}\eta_j)^2\right] \tag{5.11}$$

$$= \sum_{j=1}^{N}\tilde{H}_{i,j}^2 E\left[\eta_j^2\right] \tag{5.12}$$

$$= \sigma^2\sum_{j=1}^{N}\tilde{H}_{i,j}^2. \tag{5.13}$$

The MSE can then be written as

$$MSE = \frac{\sigma^2}{N}Tr(H^{-t}H^{-1}). \tag{5.14}$$

Where $Tr()$ is the matrix trace operator. As a point of comparison, we consider the baseline case with no multiplexing. For the baseline case, $H = I$, and the MSE becomes

$$MSE_b = \frac{\sigma_s^2}{N}Tr(I^{-t}I^{-1}). \tag{5.15}$$

$$= \sigma_s^2. \tag{5.16}$$

To measure the performance of a given multiplexing matrix H, we compare the MSE against the baseline case. We refer to the SNR gain Q as the improvement in SNR of a multiplexed measurement, relative to a non-multiplexed measurement. Defining σ_m as the noise variance for the multiplexed measurement, the SNR gain is

$$Q = \sqrt{\frac{MSE_b}{MSE}} \tag{5.17}$$

$$= \sqrt{\frac{\sigma_s^2}{MSE}} \tag{5.18}$$

$$= \sqrt{\frac{N}{Tr(H^{-t}H^{-1})}\frac{\sigma_s^2}{\sigma_m^2}}. \tag{5.19}$$

The SNR gain is maximized when the quantity $Tr(H^{-t}H^{-1})$ is minimized. When the noise is signal independent, $\sigma_s = \sigma_m$, and the matrix that minimizes this quantity is known to be the Hadamard matrix [Harwit and Sloane, 1979], when negative matrix entries are allowed, and the S-matrix when matrix entries are restricted to be non negative. The S-matrix can be derived from the Hadamard matrix, and its entries consist of only ones and zeros. The minimum trace for the S-matrix is $Tr(S^{-t}S^{-1}) = 4(N-1)^2/N$

$$Tr(S^{-t}S^{-1}) = 4\frac{N^2}{(N+1)^2} \tag{5.20}$$

$$\approx 4. \tag{5.21}$$

The SNR gain when the S-matrix is used for multiplexing is then

$$Q \approx \frac{\sqrt{N}}{2}. \tag{5.22}$$

However, when the noise is signal dependent, the S-matrix is no longer optimal, and the maximum SNR gain that can be achieved is lower.

We note that the trace of a matrix H is the sum of the diagonal elements of the matrix, which is also equal to the sum of its eigenvalues λ_i

$$Tr(H) = \sum_{i=1}^{N} a_{i,i} \tag{5.23}$$

$$= \sum_{i=1}^{N} \lambda_i. \tag{5.24}$$

Since the eigenvalues of the inverse of a matrix H^{-1} are just equal to λ_i^{-1}, we can write the MSE from Equation 5.25 as

$$MSE_b = \frac{\sigma_s^2}{N}\sum_{i=1}^{N} \lambda_i^{-2}, \tag{5.25}$$

and the SNR gain can be written as

$$Q = \sqrt{\frac{N}{\sum_{i=1}^{N}\lambda_i^{-2}}\frac{\sigma_s^2}{\sigma_m^2}}.. \tag{5.26}$$

As we discussed in Chapter 4, equations 5.2 and 5.26 are particularly useful if the matrix H is a shift invariant transformation. This is the case when the coding comes in the form of convolution with a blur kernel $\mathbf{h}$. Then H is Toeplitz or block Toeplitz, and its eigenvalues are the Fourier coefficients of the blur kernel $\mathbf{h}$. We revisit the topic of convolution coding in Section 5.3, where we discuss convolution in the continuous domain.

5.2.1 Sources of Noise

Ideally, pixels would would contain no uncertainty in their measurements, and each time we measures the same signal, we would record the same value. There are, however, several sources that cause uncertainty in the measurement of a pixel: photon noise, dark current noise, and read noise.

5.2.1.1 Read Noise

Read noise is signal independent noise that is generated from electrical noise in the camera circuitry. The intensity value of a pixel is generated by first sending the detected electrical signal through an amplifier then an Analaog-to-Digital Converter (ADC). The ADC quantizes the signal into a digital number between 0 and ADU, where ADU represent the maximum number that can be represented after ADC quantization. For instance, for a 12-bit sensor, $ADU = 2^{12} - 1$. Read noise can be divided into two noise sources [Schechner *et al.*, 2007] [Healey and Kondepudy, 1994]. Noise that is generated upstream of the the amplifier is magnified along with the signal, so that the variance in electrons increases with amplification, but the variance in gray levels σ_1^2 remains fixed. For a perfect ADC, σ_1 is simply the variance caused by the uniform quantization of one Gray Level (GL), so that $\sigma_1 = 1/\sqrt{12}$ GL. Typically, however, σ_1 is larger so that the ADC can only effectively quantize to ADU/σ_1 distinct gray levels instead of ADU. Noise that is generated downstream of the amplifier is remains constant with a fixed variance in electrons at σ_0^2. The total read noise is then

$$\sigma_r^2 = g^2\sigma_1^2 + \sigma_0^2, \tag{5.27}$$

where g is the camera gain, measured in units of photo-electrons per gray level (e^-/GL). The camera gain has an inverse relationship with the sensitivity, or ISO. Lower sensitivity settings correspond to higher gain settings, and vice-versa.The read noise is affine in the camera gain so that an increase in the camera gain, corresponding to a decrease in sensitivity, resulting in an increase in read noise.

We typically adjust the camera gain to avoid saturation, however it makes little sense to allow the gain to be set arbitrarily. The minimum gain setting is $1e^-$/GL, since a smaller gain level cannot change the quantization of photoelectrons, and hence will have a null effect on the measured signal. It is common, however, to have a minimum gain setting that is greater than $1e^-$/GL since, according to the affine model, there is no point in allowing the gain to be set such that $g_{min}\sigma_1 \ll \sigma_0$ since the total read noise cannot be reduced below σ_0.

There is also typically a maximum gain level that is set by the maximum amount of photo-electrons that can be stored by a single photodetector element, referred to as the Full Well Capacity (FWC) of the sensor. The largest gain value for a sensor is usually set to be $g_{max} = FWC/ADU$, which ensures that the maximum signal level that can be recorded by the sensor translates to the maximum image intensity. A gain greater than this would map the maximum signal level to a smaller number, effectively reducing the number of quantization levels in the image.

5.2.1.2 Photon Noise

The energy collected when light is absorbed by a pixel is quantized due to the discreet nature of photons. Furthermore, the electrical energy that is generated by the photo-electric conversion process is also quantized into discreet amounts of electrons. The process of converting photons to electrons is a random counting event that can be described by a Poisson process. The variance and mean are equal to each other for a Poisson distribution, so if J_i is the expected number of photo-electrons collected by a pixel, the photon noise variance $\sigma_{p_i}^2 = J_i$.

Poisson distributions have the interesting property that when the variance is larger than around 10, the distribution can be well approximated by a Gaussian. Typically we capture more than 10 photo-electrons per pixel, and we can therefore treat Photon noise as signal-dependent Gaussian noise whose variance depends on the amount of light in the captured signal.

5.2.1.3 Dark Current Noise

Photo-electrons are also randomly generated by thermal excitation in the sensor. Thermally generated electrons also follow a Possion distribution, where the variance is given by $\sigma_d^2 = Dt$. D is referred to the dark current of the camera, and t is the exposure time. Dark current noise becomes significant in applications, such as astronomy, where very long exposure times are prevalent. For most cameras, dark current noise is negligible for exposure times less than one second. From this point on, we focus our attention on cameras that require relatively high speed operation so that dark current noise can be ignored.

5.2.1.4 Total Noise

The noise sources are independent so that the total noise variance for the i$^{\text{th}}$ pixel becomes (ignoring dark current noise)

$$\sigma_i^2 = J_i + g^2\sigma_1^2 + \sigma_0^2. \tag{5.28}$$

If we have an image with N pixels and we define the average signal level as $J = \frac{1}{N}\sum_{i=1}^{N} J_i$, the average noise variance of the image is then

$$\sigma^2 = J + g^2\sigma_1^2 + \sigma_0^2. \tag{5.29}$$

5.2.2 Optimal Multiplexing

Multiplexing with the S-matrix is optimal when measurement noise is dominated by read noise, but when photon noise dominates, it no longer becomes optimal. In fact, the multiplexing gain is no longer greater than one so that a non-multiplexed measurement actually has a greater SNR. In practice, however, photon noise is not infinitely greater than read noise. More realistically, different imaging scenarios will result in different ratios of photon to read noise $\chi = \sigma_p/\sigma_r$. While the S-matrix is no longer optimal when $\chi > 0$, some other matrix may be.

The problem of identifying matrices that are optimal under different values of χ was explored by Ratner et al [Ratner and Schechner, 2007] [Ratner *et al.*, 2007] . We define the optical efficiency C as the number of non-zero elements in a row of the multiplexing matrix H

$$C = \sum_{j=1}^{N} H_{i,j}. \tag{5.30}$$

Assuming the same camera gain setting for multiplexed and non-multiplexed measurements, referring to Equation 5.19, the SNR gain for the matrix H becomes

$$Q(C)^2 = \frac{N}{Tr(H^{-t}H^{-1})} \frac{1+\chi^2}{1+C\chi^2}, \tag{5.31}$$

Ratner et al. used an optimization routine to search for a matrix that maximizes the SNR gain for a given value of χ. In general the trend discovered was that the amount of photon noise decreases as the optimal value of C decreases. This trend can be seen in Figure 5.1, which plots the SNR gain vs. C for a variety of values of χ.

Ratner et al. explored this topic further and found an analytic expression for the minimum trace $T^*(C)$ that can be achieved for any multiplexing matrix

$$\begin{aligned} T^*(C) &= \min\{Tr(H^{-t}H^{-1})\} && (5.32) \\ &= \frac{(N-C)+C(N-1)^2}{(N-C)C^2}. && (5.33) \end{aligned}$$

Equation 5.33 is a very significant result because it allows us to derive the maximum SNR gain that can be achieved for any coding technique that sacrifices light. This means the the same bound holds for all the light masking techniques mentioned in Chapter 5.1, including spectral [Harwit and Sloane, 1979] [Hanley *et al.*, 1999] [Baer *et al.*, 1999], light field [Veeraraghavan *et al.*, 2007] [Lanman *et al.*, 2008] [Liang *et al.*, 2008], illumination multiplexing [Schechner *et al.*, 2003] [Schechner *et al.*, 2007] [Ratner and Schechner, 2007] [Ratner *et al.*, 2007], coded aperture defocus deblurring [Levin *et al.*, 2007] [Veeraraghavan

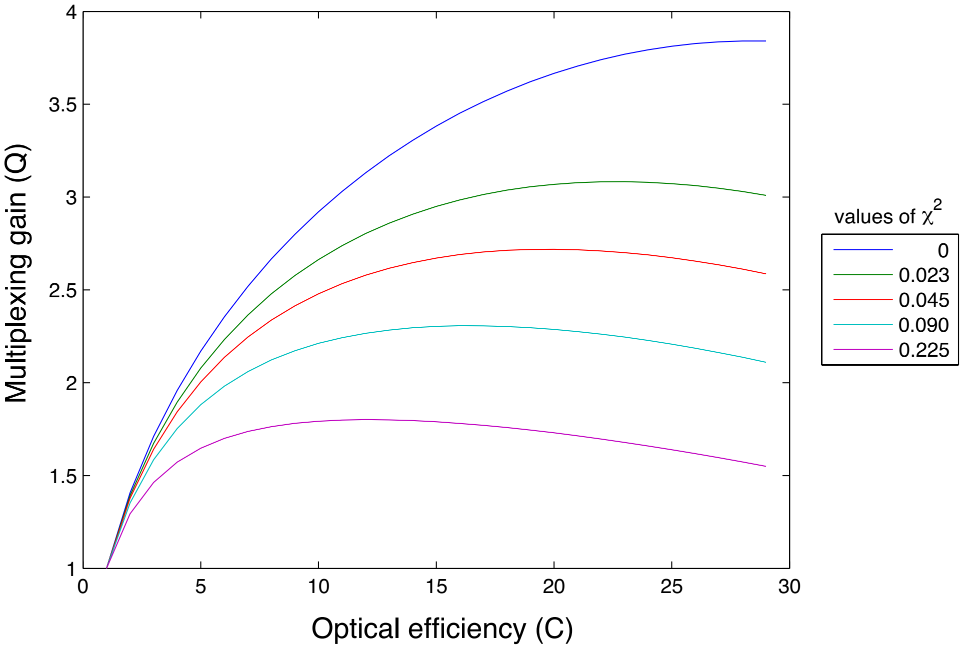


Figure 5.1: Multiplexing gain (Q) vs. optical efficiency (C) for various ratios of photon to read noise variance (χ^2) using a multiplexing matrix with size $N = 57$. The results are calculated using Equation 5.33. When $\chi = 0$, photon noise is absent, the optimal efficiency is $C = 29$, and the optimal multiplexing matrix is the S matrix. As the amount of photon noise increases, both the optimal efficiency, and the maximum SNR gain decrease. When $\chi^2 = .225$, the optimal efficiency is $C = 11$, and the maximum SNR gain is just $Q = 1.75$.

et al., 2007] [Zhou and Nayar, 2009] [Zhou *et al.*, 2011], and shutter-based motion deblurring [Raskar *et al.*, 2006].

While Schechner et al. explored this problem in detail, they only considered the case where the camera gain is identical for multiplexed and non-multiplexed measurements. This scenario only occurs when both the multiplexed and non-multiplexed signals are small enough to remain unsaturated at the lowest gain setting (highest sensitivity). For larger signal levels, the camera gain will need to be set independently for multiplexed and non-multiplexed measurements, and the read noise for the multiplexed and non-multiplexed measurements will no longer be the same.

5.2.3 Multiplexing Noise and Camera Gain

We now relax the restriction that the multiplexed measurement must remain below saturation at the lowest camera gain setting. We derive the noise variance for the multiplexed and non-multiplexed measurements under the assumption that the gain is set appropriately so that each measurement avoids saturation.

The average non-multiplexed image brightness $I_s = J/g_s$ measured in gray levels (GL) depends on the camera gain g_s.

$$I_s = J/g_s. \tag{5.34}$$

The image will have some variation about the average brightness J. As a measure of the dynamic range in the image, we denote D the proportionality constant between the peak signal brightness J_{peak} and the average signal J

$$J_{peak} = D \cdot J. \tag{5.35}$$

To avoid saturation for the non-multiplexed image, we set the gain for the non-multiplexed image so that

$$g_s = \frac{D \cdot J}{ADU}. \tag{5.36}$$

Our multiplexed image is C times brighter than our non-multiplexed image. Denoting the gain for the multiplexed measurement g_m, the brightness in GL for our multiplexed measurement is $I_m = C \cdot J / g_m$. In order to avoid saturation for the multiplexed measurement we set the gain so that

$$g_m = C \frac{D \cdot J}{ADU}, \tag{5.37}$$

$$= C g_s \tag{5.38}$$

where in Equations 5.36 and 5.37 we have assumed that the peak signal level of our multiplexed measurement is less than the FWC of the sensors so that $C \cdot D \cdot J \leq FWC$. The ratio between the average noise for multiplexed and non-multiplexed measurements is then

$$\frac{\sigma_s^2}{\sigma_m^2} = \frac{J + g_s^2 \sigma_1^2 + \sigma_0^2}{CJ + C^2 g_s^2 \sigma_1^2 + \sigma_0^2}. \tag{5.39}$$

We also define an upper bound on the ratio of the two noise levels

$$\frac{\sigma_s^2}{\sigma_m^2} \leq \frac{J + \frac{g_{max}^2}{C^2} \sigma_1^2 + \sigma_0^2}{CJ + \sigma_0^2}. \tag{5.40}$$

The expression in the numerator follows from the fact that since the multiplexed measurement does not saturate at the highest gain setting, the non-multiplexed measurement must have a gain setting that is at least C times smaller. The expression in the denominator follows from the fact that as the camera gain decreases, the total read noise variance approaches σ_0^2.

5.2.4 Multiplexing Limits

We now derive an upper bound on the maximum SNR gain that can be achieved with any multiplexing matrix H. Using the expression for the minimum trace given in Equation 5.33, the optimal SNR gain can be written as

$$Q^2(C,N) = \frac{N}{Tr(H^{-t}H^{-1})}\frac{\sigma_s^2}{\sigma_m^2} \tag{5.41}$$

$$\leq \frac{\sigma_s^2}{\sigma_m^2}\frac{C^2\ (N-C)}{C\,N - 2\,C + 1} \tag{5.42}$$

5.2.4.1 Large Multiplexing Matrices

The upper bound on $Q^2(C,N)$ is a monotonically increasing function. Writing the upper bound as $\tilde{Q}^2(C,N)$, we can be verify this by taking the derivative with respect to N, and observing that the result is strictly positive

$$\frac{d\tilde{Q}^2(C,N)}{dN} = \frac{\sigma_s^2}{\sigma_m^2}\frac{C^2(C-1)^2}{(CN-2C+1)^2} > 0 \tag{5.43}$$

As a result, $\tilde{Q}$ reaches a maximum when $N=\infty$. Denoting $Q_N(C)$ as

$$Q_N^2(C) = \lim_{N\to\infty}\tilde{Q}^2(C,N) \tag{5.44}$$

$$= C\frac{\sigma_s^2}{\sigma_m^2} \tag{5.45}$$

Since the multiplexing gain monotonically increases with N, we expect the upper bound given by Equation 5.45 to decrease for small matrices. The upper bound given holds for all matrices, but the bound is tightest when the multiplexing matrices are large (e.g. $N > 1000$).

We get an upper bound for the SNR gain by substituting the noise bound from Equation 5.40

$$Q_N^2(C) \leq C\frac{J + \frac{g_{max}^2}{C^2}\sigma_1^2 + \sigma_0^2}{CJ + \sigma_0^2} \tag{5.46}$$

We can now find the the value C_{max} that maximizes this upper bound by taking derivatives with respect to C

$$C_{max} = \arg\max_{C} \frac{dQ^2}{dC} \tag{5.47}$$
$$= \frac{Jg_{max}^2\sigma_1^2 + s}{J\sigma_0^2 + \sigma_0^4}, \tag{5.48}$$

where $s = \sqrt{g_{max}^2\sigma_1^2(J^2g_{max}^4\sigma_1^4 + J\sigma_0^2 + \sigma_0^3)}$. Substituting C_{max} into Q_N gives an upper bound that is independent of C

$$Q^2(C, N) \leq Q_N^2(C_{max}) \tag{5.49}$$
$$\leq \frac{2g_{max}^2\sigma_1^2(J + \sigma_0^2)}{Jg_{max}^2\sigma_1^2 + s} \tag{5.50}$$
$$\leq 1 + \frac{\sigma_0^2}{J}, \tag{5.51}$$

where in Equation 5.51 we have used the fact that $s > Jg_{max}^2\sigma_1^2$. We first point out that the upper bound is always greater than one. However, the amount greater than one is inversely proportional to the signal level J. Therefore the multiplexing gain will asymptotically approach one for large signal levels. Thus, while optimal multiplexing can never hurt you, there will be scenarios where the benefit is marginal. We note for instance that if we want to achieve a (squared) multiplexing gain that is ϵ greater than one, this puts an upper bound on the maximum signal level

$$J < \frac{\sigma_0^2}{\epsilon}. \tag{5.52}$$

This bound significantly decreases the space of imaging scenarios that will give the desired multiplexing gain. As an example consider a Canon 1D Mark II sensor. The total read noise variance at the lowest gain setting is $16e^-$, which is typical of a high quality CMOS sensor. For this camera, it is not possible to achieve a multiplexing gain greater than $\sqrt{2}$ if the non-multiplexed signal level is greater than $16e^-$. Note that this is only a minute fraction of the maximum signal level that can be recorded by the system since for this sensor $FWC = 80,000e^-$.

Initially, we assumed that the the sensitivity of the camera was adjusted to avoid saturation and minimize noise for both the multiplexed and non-multiplexed measurements. We note, however, that Equation 5.51 is independent of the gain settings for multiplexed and non-multiplexed measurements. This means that the same bound holds even if the light level is sufficiently small that it is necessary to set the camera gain to the lowest setting for both measurements.

5.2.4.2 Interpretation for Large Multiplexing Matrices

The bound given by Equation 5.51 is particularly useful in evaluating the performance of coded aperture defocus deblurring systems [Levin *et al.*, 2007] [Veeraraghavan *et al.*, 2007] [Zhou and Nayar, 2009] [Zhou *et al.*, 2011] and shutter based motion deblurring systems [Raskar *et al.*, 2006]. In all of these systems, the number of multiplexed measurements N is equal to the number of pixels (i.e $N \approx 10^6$), therefore the bound derived in Equation 5.45 is quite accurate. For these systems it is always possible to find a multiplexing strategy that will give a SNR gain greater than one. However, depending on the signal level, the amount that the SNR gain is greater than one may be negligible. Furthermore, because these systems are inherently shift invariant, there are physical constraints that may prevent implementation of the optimal coding strategy. Also, since the computational technique will always require more resources, even a small amount of SNR gain may not warrant choosing a computational technique over a conventional one. In summary, we expect that the scenarios where a computational technique provides a sufficient performance advantage is extremely limited.

5.2.4.3 Small Multiplexing Matrices

The upper bound on SNR gain given by Equation 5.51 is a very easy to interpret result that holds for all matrices. The SNR gain for any multiplexing matrix is never greater than one plus the ratio of read to photon noise. This bound is tight for large multiplexing matrices, but it becomes weak for smaller matrices. We know derive a tighter bound for smaller matrices.

Let us define the read noise at the highest signal level as $\sigma_{rmax}^2 = g_{max}^2\sigma_1^2 + \sigma_0^2$, and the read noise at the lowest gain setting as $\sigma_{rmin}^2 = g_{min}^2\sigma_1^2 + \sigma_0^2$. Then we can rewrite Equation 5.42 as

$$Q^2(C,N) \leq \frac{J+\sigma_{rmax}^2}{CJ+\sigma_{rmin}^2}\frac{C^2\,(N-C)}{C\,N-2\,C+1} \tag{5.53}$$

$$\leq \frac{J+\sigma_{rmax}^2}{CJ+\sigma_{rmin}^2}\frac{C\,(N-C)}{N}. \tag{5.54}$$

We find the value for C_{max} that maximizes this bound by differentiating

$$C_{max} = \arg\max_C \frac{dQ^2}{dC} \tag{5.55}$$

$$= \frac{s-\sigma_{rmin}^2}{J}. \tag{5.56}$$

where $s = \sqrt{\sigma_{rmin}^2(\sigma_{rmin}^2 + NJ)}$. Substituting the minimizer back into Equation 5.54 gives the expression

$$Q^2(C,N) \leq \frac{\left(J+\sigma_{rmax}^2\right)\left(s-\sigma_{rmin}^2\right)\left(\sigma_{rmin}^2+NJ-s\right)}{sNJ^2}. \tag{5.57}$$

Equation 5.57 implies an upper bound on the maximum signal level that can achieve an SNR gain greater than one

$$J_{max} < \frac{N}{4}\frac{\sigma_{rmax}^4}{\sigma_{rmin}^2}. \tag{5.58}$$

This bound is only meaningful when $J_{max} < FWC$, which requires a relatively small value for N (e.g. $N \ll 1000$) and that the maximum read noise variance is not significantly greater than the minimum read noise variance. For instance, the Canon 1D Mark II sensor has a large difference in the maximum and minimum read noise values. When a multiplexing matrix of size $N = 64$ is used with this sensor, the bound becomes $J_{max} < 1024^2 e^-$, which is greater than the FWC. On the other hand, when a Lumenera Lu570 sensor is used instead, the bound becomes $J_{max} < 1024e^-$. This limits the maximum signal range to within a fraction of the maximum signal level that can be recorded by the sensor.

As a final note, we mention that when both the non-multiplexed and the multiplexed signal are below saturation at the lowest gain level, the total read noise is the same for both measurements, and then the largest signal that will allow an SNR gain greater than one becomes

$$J_{max} < \frac{N}{4}\sigma_{rmin}^2. \tag{5.59}$$

For instance if a Canon 1D Mark II is used with a multiplexing matrix of size $N = 64$, the maximum signal level that will result in a multiplexing gain is $J_{max} < 64e^-$, which is a very weak signal.

5.2.4.4 Interpretation for Small Multiplexing Matrices

We saw that for large multiplexing matrices, there is always a coding strategy that will result in a multiplexing gain greater than one. However, for small matrices, we see that this is not the case, and there will be a maximum signal level above which it is not possible to achieve a SNR gain. The bound given by Equation 5.58 is particularly useful in evaluating the performance of color [Baer *et al.*, 1999] [Ihrke *et al.*, 2010], light field [Veeraraghavan *et al.*, 2007] [Lanman *et al.*, 2008] [Liang *et al.*, 2008] and illumination multiplexing systems [Schechner *et al.*, 2003] [Schechner *et al.*, 2007] [Ratner and Schechner, 2007] [Ratner *et al.*, 2007]. In all of these systems, the number of multiplexed measurements N is small (i.e $N \leq 16$). Therefore the maximum signal level that will result in a multiplexing gain greater than one is extremely small (i.e. $J < 64e^-$). When the signal level is greater than this, conventional techniques will exhibit superior performance. We discuss how to map lighting conditions into photon counts in Section 5.4.1. For now we simply state that, as is the case when N is large, the scenarios where these techniques achieve a performance advantage over conventional techniques is extremely limited.

5.3 Coding For Invariance

Blur is a common problem in imaging systems. For a perfectly focusing imager, the Point Spread Function (PSF) is a perfect delta function. Then, the brightness at a point on an object maps directly to a point on the image. When an imaging system exhibits blur, object points are blurred so that they map to an area on the sensor instead of a point. This blurring in the image typically suppresses the energy at high frequencies in the Modulation Transfer Function (MTF). When the PSF is known, we can apply deconvolution to remove blur. Then the MTF tells us directly what the SNR of recovered images will be.

The deconvolution problem is analogous to the coding problem discussed in the previous sections. In the discrete case, blurring a focused image with a PSF corresponds to multiplication by a coding matrix H that is block Toeplitz, and the MTF corresponds to the singular values of this matrix. In the equations that follow, we use the continuous form of the PSF and MTF to derive the MSE and SNR gain after deconvolution.

In a variety of applications, a conventional camera introduces an undesirable variation in MTF over some domain, for instance object depth or motion. This variation is particularly undesirable when all states in the domain are equally likely. Then, the best strategy is to introduce coding in the imaging system that equally preserves energy in the MTF over all states. Parseval's theorem, however, dictates that there is a limit to how much we can simultaneously maximize MTF and achieve MTF invariance [Levin *et al.*, 2008] [Levin *et al.*, 2009] [Cho *et al.*, 2010] . As a consequence, there is an upper bound on the maximum domain invariant MTF that can be achieved.

We need to be careful when we talk about domain invariant blur. There are two types of domain invariant blur. For the first type, both the MTF and PSF are invariant. This is the most desirable type of invariance because it allows deconvolution to be applied without first estimating the PSF. Unfortunately it is very difficult to derive upper bounds on the MTF for this type of invariance. It is however, possible to derive upper bounds for the second type of invariance, where the MTF is invariant, but the PSF may not be. Of course if the PSF is not invariant, then we need a way to estimate the PSF before we can apply deconvolution. However, we can think of these bounds as being somewhat conservative in the sense that they tell us, assuming the best case where we can estimate the PSF perfectly, what is the minimum MSE that can be achieved. Furthermore, the upper bounds derived for the second type of invariance must also apply for the first type of invariance (perhaps weakly) since the first type is more constrained than the first.

Here we focus on two application areas: motion and defocus deblurring. In both these scenarios, blur suppresses important scene details, and the amount of blur can vary significantly within an image. The amount of motion blur varies depending on object speed, while the amount of defocus blur depends on object depth. In both cases, their is a tension between the amount of light captured, and the amount of blur introduced. Increasing exposure time increases the amount of light captured, but also increases motion blur. Analogously, increasing aperture size increases the amount of light captured but increases defocus blur. Consequently, it is always possible to remove blur by sacrificing light: either by reducing exposure time, or stopping down the lens aperture.

A common strategy that has been used to combat both types of blur without sacrificing light is to remove the object dependent variation in MTF. For example, parabolic camera motion can be utilized to create a motion independent MTF, and translating the sensor along the optical axis during exposure can produce blur that is independent of depth. The MTFs for these techniques have been derived in [Levin *et al.*, 2008] [Levin *et al.*, 2009]. Using the MTF, a lower bound on the MSE can be derived, and using the machinery from the previous sections, the maximum SNR gain can be calculated for these techniques.

The problem of engineering a depth-invariant MTF is closely related to the problem of engineering a motion invariant MTF. Both problems can be well understood by using a phase space representation to analyze the problem. For the case of 2D motion blur, we use a 3D space-time parameterization, for defocus blur, we use a 4D light field parameterization. In both cases, optimal performance is achieved when energy in Fourier space is maximized over a restricted domain, subject to conservation of energy constraints. For motion blur, energy in Fourier space is restricted to lie within an double cone whose apex angle is determined by the range of object speeds. For defocus blur, energy is restricted to a 3D manifold within 4D Fourier space. Levin et al. and Cho et al. provide an excellent overview of these concepts [Levin *et al.*, 2008] [Levin *et al.*, 2009] [Cho *et al.*, 2010] .

5.3.1 MSE in Continuous Form

At this point, we transition from the discrete to continuous domain, which will provide a convenient means of calculating the MSE for each of these techniques. We return to the analysis of shift-invariant systems introduced in Chapter 1. A blurry image $g(x, y)$ is then the convolution of a latent focused image $f(x, y)$ and the system Point Spread Function

$h(x, y)$,

$$g(x, y) = f(x, y) \otimes h(x, y). \tag{5.60}$$

This equation can be expressed compactly in the Fourier domain, where convolution becomes a multiplication

$$G(\omega_x, \omega_y) = F(\omega_x, \omega_y) \cdot H(\omega_x, \omega_y), \tag{5.61}$$

where ω_x and ω_y are spatial frequencies in the x and y directions, respectively. The Fourier transform of the blur kernel $H(\omega_x, \omega_y)$ is the Optical Transfer Function (OTF) of the camera. The modulus of the OTF is the MTF, which is an indicator of how much different spatial frequencies are attenuated by the system. With this notation, we can then write the average image intensity as

$$\begin{aligned} J &= \frac{1}{D_x \cdot D_y} \int_{-W/2}^{W/2} \int_{-H/2}^{H/2} g(x, y) dx dy \qquad (5.62) \\ &= \frac{F(0,0)H(0,0)}{D_x \cdot D_y}, \qquad (5.63) \end{aligned}$$

where D_x and D_y are the height and width of the image, respectively. Note that if the optical system allows C times more light through so that $H(0, 0)$ becomes $C \cdot H(0, 0)$, the average signal intensity becomes $C \cdot J$. When we capture an image, the blurred image is perturbed by additive noise, represented by the random function $\psi(x, y)$ and its Fourier spectrum $\Psi(\omega_x, \omega_y)$. The captured image spectrum Q_0 is then

$$G_0(\omega_x, \omega_y) = F(\omega_x, \omega_y) \cdot H(\omega_x, \omega_y) + \Psi(\omega_x, \omega_y). \tag{5.64}$$

We form an estimate of our focused image spectrum F^* by dividing the captured image spectrum G_0 by the OTF. Assuming that there are no zero crossings in the OTF, the estimate becomes

$$F^*(\omega_x, \omega_y) = \frac{G_0(\omega_x, \omega_y)}{H(\omega_x, \omega_y)} \tag{5.65}$$

$$= F(\omega_x, \omega_y) + \frac{\Psi(\omega_x, \omega_y)}{H(\omega_x, \omega_y)}. \tag{5.66}$$

The MSE is the expected L2 norm on the difference between the estimated image and the focused image

$$MSE = E[\|F(\omega_x, \omega_y) - F^*(\omega_x, \omega_y)\|^2] \tag{5.67}$$

$$= E\left[\left\|\frac{\Psi(\omega_x, \omega_y)}{H(\omega_x, \omega_y)}\right\|^2\right] \tag{5.68}$$

When the noise is Gaussian, the MSE becomes

$$MSE = \sigma^2 \left\|\frac{1}{H(\omega_x, \omega_y)}\right\|^2 \tag{5.69}$$

Although the assumption that the noise is Gaussian is somewhat restrictive, the assumption works quite well for Poisson distributed noise with variance greater than about about 10.

We assume that the focused image is bandlimited by the Nyquist frequency $\Omega = 1/2\Delta$, where Δ is the pixel size. Then the frequencies outside the range $|\omega_x|, |\omega_y| < \Omega$ do not contribute to the MSE. Defining the normalized frequency coordinates $\omega'_x = \omega_x\Delta$ and $\omega'_y = \omega_y\Delta$, we can write the MSE as

$$MSE = \sigma^2 \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \frac{1}{\|H(\omega'_x, \omega'_y)\|^2} d\omega'_x d\omega'_y. \tag{5.70}$$

Note that, if the captured image intensity is measured in terms of photons, according to this definition, the MSE is properly normalized so that if the PSF is a delta function, $h(x, y) = \delta(x, y)$, the OTF is constant, $H(\omega_x, \omega_y) = 1$, and the MSE is equal to $\sigma_s^2 = J + g_s^2\sigma_1^2 + \sigma_0^2$. This is consistent with the discrete form of the MSE, which also reduces to $MSE = \sigma_s^2$ when the multiplexing matrix is equal to the identity. Also note that when the optical system becomes C times more efficient so that $H(0,0)$ becomes $C \cdot H(0,0)$, and the camera gain is adjusted to avoid saturation at the increased light level, the noise variance becomes $\sigma^2 = \sigma_m^2 = C \cdot J + C^2 g_m^2 \sigma_1^2 + \sigma_0^2$. Thus, the notation for the noise variances is consistent with the definitions in Section 5.2.3.

5.3.2 Performance Limits for Motion Blur

Cameras exhibit motion blur when objects move during exposure so that points in the image are blurred along the direction of motion. A scene may consist of multiple objects moving at different speeds and directions. Levin et al. showed that, for 1D motion blur, parabolic camera motion creates a motion invariant PSF that nearly achieves the upper bound on the MTF [Levin *et al.*, 2008]. The technique can therefore be used to be remove blur without first estimating motion, while at the same time nearly achieving the maximum possible SNR. However, the technique is not applicable for general 2D motion.

Here we consider the general case where objects motion is a 2D vector in image space coordinates. Cho et al. developed a technique that comes close to achieving the MTF upper bound, but the technique requires multiple exposures and motion estimation. No techniques currently exist that achieve a motion invariant MTF without also requiring motion estimation.

5.3.2.1 The Baseline for Motion Blur

For an exposure time T, an object moving at constant velocity 2D $\mathbf{s}_m = (s_x, s_y)$ in image space, is blurred along the direction of motion by a box shaped kernel with width equal to $b_m = |\mathbf{s}_m|T$. Denoting the maximum object speed as S_m so that $-S_m \leq |\mathbf{s}_m| \leq S_m$, the maximum blur size for the scene then becomes $b_{max} = TS_m$. Motion blur becomes apparent when the maximum blur size is larger than a pixel so that $b_{max} > \Delta$. Thus, motion blur can be removed by setting the exposure time T so that the maximum blur size is equal to one pixel. We define the baseline exposure time T_b as

$$T_b = \frac{\Delta}{S_m}. \tag{5.71}$$

When the exposure time is equal to T_b, the motion blur becomes so small that the PSF effectively becomes a delta function for all motions. We use this case as a baseline for comparing the performance of any motion deblurring technique. Here we assume that the image intensity $f(x)$ is measure in units of power, and the PSF $h_{mi}(x)$ is measured in units of time, which ensures that $J = F(0,0)H(0,0)/(H \cdot W)$ is in units of photons. Then the D.C. component of the OTF becomes $H_{mi}(0,0) = T_b$. Defining the average power $\bar{P} = F(0,0)/(D_x \cdot D_y)$, the average signal becomes $J = \bar{P}T_b$. The MTF for the motion blur baseline is constant

$$H_{mb}(\omega_x, \omega_y) = T_b, \tag{5.72}$$

and, using the same notation for the noise in the absence of multiplexing, σ_s, as used in the previous section, the MSE is

$$MSE_{mb} = \frac{\sigma_s^2}{T_b^2}. \tag{5.73}$$

5.3.2.2 Motion Invariant Blur

When the exposure time is increased by a factor of C to $T = C \cdot T_b$, the amount of light captured increases by a factor of C, but the maximum blur size also increases by the same factor, causing motion blur to become larger than one pixel. When motion blur is larger than one pixel, the blur can be engineered to become motion invariant, allowing it to be be removed via deconvolution. The question becomes whether or not the combination of increased light and deconvolution results in a net gain in MSE.

5.3.2.3 MSE for 1D Motion Invariant Blur

Levin et al., analyzed techniques that produce an MTF that is motion invariant over a desired range of 1D velocities [Levin *et al.*, 2008]. They derived an upper bound on the best possible MTF that can be achieved, and introduced a technique that produces an MTF that is close to the upper bound, while at the same time producing a motion invariant PSF. In this technique, constant acceleration is intentionally applied to the camera during exposure so that the camera speed matches the speed of all objects in the scene at least once during exposure. In this setting, the D.C. component of the OTF is $H_{mi}(0) = C \cdot T_b$. The upper bound on the squared motion invariant MTF is

$$\|H_{1D}(\omega_x)\|^2 \leq \frac{CT_b}{2S_m|\omega_x|},, \tag{5.74}$$

where ω_x is assumed to be the direction of object motion. In this setting, the range of camera speeds that are spanned during exposure is $2S_m$, and the MTF is motion invariant for moving objects that would otherwise produce a blur kernel with widths less than or equal to the maximum blur size $b_{max} = CT_bS_m$. Using the same notation for noise in presence of multiplexing as used in the previous section, the lower bound on the motion invariant MSE then becomes

$$MSE_{1D} \geq \sigma_m^2 \int_{-1/2}^{1/2} \frac{1}{\|H(\omega_x')\|^2} d\omega_x'. \tag{5.75}$$

$$= \frac{2\sigma_m^2 S_m}{CT_b\Delta} \int_{-1/2}^{1/2} |\omega_x'| d\omega_x'. \tag{5.76}$$

$$= \frac{\sigma_m^2}{2CT_b^2} \tag{5.77}$$

5.3.2.4 SNR Bound for 1D Motion Invariant Blur

The maximum gain in SNR Q_{mi} that can be achieved when using a 1D motion invariant technique, relative to using a shorter exposure time, then becomes

$$Q_{1D}^2 \leq \frac{MSE_{mb}}{MSE_{1D}}. \tag{5.78}$$

$$= 2C\frac{\sigma_s^2}{\sigma_m^2}. \tag{5.79}$$

Interestingly, the upper bound on the SNR gain for a motion invariant camera is equal to $\sqrt{2}$ times the SNR gain for optimal multiplexing, and hence the optimal increase in exposure time is equal to the optimal row sum C_{max} for multiplexing given by Equation 5.48.

Substituting for C gives an upper bound for the motion invariant SNR gain in terms of only signal strength and sensor noise

$$Q_{1D}^2 \leq 2 + \frac{2\sigma_0^2}{J}. \tag{5.80}$$

In this case, the SNR gain is always greater than $\sqrt{2}$. However, the amount greater than $\sqrt{2}$ decreases asymptotically with the signal level, and again we have an upper bound on the maximum signal level that will result in a (squared) SNR gain that is ϵ greater than one

$$J < \frac{2\sigma_0^2}{\epsilon - 1}. \tag{5.81}$$

Again, the bound significantly reduces the space of imaging scenarios that will give the desired multiplexing gain. In this case, when using the Canon Mark II, it is not possible to achieve a multiplexing gain greater than $\sqrt{3}$ if the uncoded signal level is greater than $32e^-$.

5.3.2.5 MSE for 2D Motion Invariant Blur

Cho et al., analyzed the problem of producing a motion invariant MTF for the case of general 2D motion [Cho *et al.*, 2010] . They derived an upper bound on the best possible MTF that can be achieved, and introduced a technique that produces an MTF that is close to the upper bound. However, the technique does not produce a motion invariant PSF and therefore requires motion estimation. In this technique, two images are captured with constant acceleration in orthogonal directions.

The upper bound on the motion invariant MTF for 2D motion is

$$\|H_{2D}(\omega_x, \omega_y)\|^2 \leq \frac{CT_b}{2S_m|\omega_r|}, \tag{5.82}$$

where $|\omega_r| = \sqrt{\omega_x^2 + \omega_y^2}$. the lower bound on the motion invariant MSE for general 2D motion then becomes

$$
\begin{aligned}
MSE_{2D} &\geq \sigma_m^2 \int_{-1/2}^{1/2} \frac{1}{\|H(\omega'_x, \omega'_y)\|^2} d\omega'_x d\omega'_y. && (5.83)\\
&= \frac{2\sigma_m^2 S_m}{CT_b\Delta} \int_{-1/2}^{1/2} \sqrt{\omega'^2_x + \omega'^2_y} d\omega'_x d\omega'_y. && (5.84)\\
&= \frac{\sigma_m^2 S_d}{CT_b\Delta} \frac{\sqrt{2} + \sinh^{-1}(1)}{6} && (5.85)\\
&> \frac{3}{4} \frac{\sigma_m^2}{CT_b^2} && (5.86)
\end{aligned}
$$

5.3.2.6 SNR Bound for 2D Motion Invariant Blur

The maximum gain in SNR Q_{mi} that can be achieved when using a 2D motion invariant technique, relative to using a shorter exposure time, is

$$
\begin{aligned}
Q_{2D}^2 &\leq \frac{MSE_{mb}}{MSE_{2D}}. && (5.87)\\
&= \frac{4}{3} C \frac{\sigma_s^2}{\sigma_m^2}. && (5.88)
\end{aligned}
$$

This upper bound on the SNR gain for 2D motion is only slightly better than the multiplexing gain given in Equation 5.45, and slightly worse than the SNR gain for 1D motion blur. As in the case of 1D motion, the optimal increase in exposure time is equal to the optimal row sum for multiplexing given by Equation 5.48. The upper bound for the SNR gain is

$$
Q_{2D}^2 \leq \frac{4}{3} + \frac{4}{3} \frac{\sigma_0^2}{J}. \tag{5.89}
$$

In this case, when using the Canon Mark II, it is not possible to achieve a multiplexing gain greater than $\sqrt{4/3}$ if the uncoded signal level is greater than $32e^-$.

5.3.2.7 Motion Extension for Motion Invariant Blur

There is another interesting way to interpret Equations 5.80 and 5.89. Suppose that instead of asking what the maximum SNR gain I can get from motion invariant photography, we ask instead what is the most we can extend the range of speeds using motion invariant photography without sacrificing SNR. We derive the result for 1D motion, but the results are the same for 2D motion as well. Then the range of speeds for the motion invariant exposure increases to $M_{1D} \cdot S_m$, and the new MTF H'_{1D} becomes

$$\|H'_{1D}(\omega_x)\|^2 \leq \frac{CT_b}{2M_{1D}S_m|\omega_x|}, \tag{5.90}$$

which makes the upper bound for the SNR gain

$$\begin{aligned} Q'^2_{1D} &\leq 2\frac{C}{M_{1D}}\frac{\sigma_s^2}{\sigma_m^2} && (5.91) \\ &= \frac{1}{M_{1D}}Q^2_{1D}. && (5.92) \end{aligned}$$

Thus, if we want to increase the range of speeds by a factor of M_{1D} while also ensuring the SNR gain Q'^2_{1D} remains greater than 1, we find that

$$M_{1D} \leq Q^2_{1D}, \tag{5.93}$$

so the increase in motion speeds has the exact same bound as the SNR gain given by Equations 5.80 and 5.89. These equations summarize the best possible performance that can be achieved using a motion invariant technique. We can think of these upper bounds as either the maximum SNR gain that can be achieved for a fixed range of speeds, or as the maximum increase in the range of speeds that can be achieved for a fixed SNR.

For 1D motion, the upper bound is always greater than one for all signal levels. However, it approaches 2 for large signal levels. Furthermore, the upper bound is only met when the optimal motion invariant MTF achieved. The parabolic exposure technique developed by Levin et al. comes closest, but it only achieves the bound for infinitely large exposures. In practice, the actual SNR gain will be less than $\sqrt{2}$, and the maximum motion extension will be less than 2, even for large signal levels. We therefore expect that for 1D object motion, we can achieve at least some performance advantage at all light levels when using a motion invariant technique.

For the case of general 2D motion, the upper bound is always greater than one for all signal levels, but, is only slightly greater than one for large signal levels. The orthogonal parabolic exposure technique developed by Cho et al. [Cho *et al.*, 2010] comes closest, but it does not achieve the bound. Furthermore, the technique requires motion estimation which may propagate errors in PSF estimation that reduce the SNR gain even further. In summary, the performance advantage in using either a 1D or 2D motion invariant technique is little to none for large signal levels, even when the optimal MTF is achieved. There is only hope in achieving a performance advantage when the light levels are very low, as dictated by Equations 5.52 and 5.81.

Finally, we note that the upper bound was derived under the assumption that the range of object speeds are distributed uniformly. This is of course the most general case, but one can imagine scenarios where this distribution is not entirely uniform, and for instance, only two object motions exist that differ significantly in magnitude. When prior information on the distribution of motions is known, the upper bounds on the MTF in Equations 5.74 and 5.82 no longer hold so that the bounds derived here are no longer valid.

5.3.3 Performance Limits for Defocus Blur

The discussion on defocus blur will largely parallel the discussion on motion blur. Cameras exhibit defocus blur when objects are located at depths other than the focal plane of the camera. The further objects are from the focal plane, the greater the amount of defocus blur will be. The problem of defocus blur arises due to the finite size of the camera aperture. Pinhole cameras exhibit no defocus blur, while larger aperture sizes introduce greater amounts of defocus blur. We can thus always remove defocus blur by stopping down the aperture. We do so however, at the cost of reducing the amount of light captured by the sensor. An alternative to stopping down the aperture is to use an EDOF camera to produce a depth-invariant blur. Because the blur is depth invariant, it can be removed via deconvolution. However, deconvolution amplifies the noise in captured EDOF images. In the following sections we compare the performance of EDOF techniques relative to a baseline imaging system with a stopped down aperture.

5.3.3.1 The Baseline for Defocus Blur

For a square aperture of width A, an object located at a depth d will produce a square defocus kernel with width equal to $b_m = |s_d|A$, where $s_d = (d - d_0)/d$, and d_0 is the distance from the lens to the object side focal plane of the camera. When the light field is parameterized by the 2D lens aperture coordinates (u, v) and the 2D lens focal plane coordinates (x, y), image points map to hyperplanes in light field space with slope s_d, according to the two relations $x = s_d u$, $y = s_d v$. As in [Levin *et al.*, 2009], for objects in a depth range $d \in [d_{min}, d_{max}]$, we choose the focal distance to be $d_0 = \frac{2d_{min}d_{max}}{d_{max}-d_{min}}$. Then the range of light field slopes becomes $S_d/2 \leq s_d \leq S_d/2$ and the maximum blur size becomes $b_{max} = AS_d/2$, where $S_d/2 = \frac{d_{max}-d_{min}}{d_{max}+d_{min}}$.

Defocus blur becomes apparent when the maximum blur size is larger than a pixel so that $b_{max} > \Delta$. In the same way that motion blur can be removed by setting the exposure time T appropriately, so can defocus blur be removed by setting the aperture width A so that the maximum blur size is equal to one pixel. We define the baseline aperture width A_b as

$$A_b = \frac{2\Delta}{S_m}. \tag{5.94}$$

When the aperture width is equal to A_b, defocus blur becomes so small that the PSF effectively becomes a delta function at all depths. We use this case as a baseline for comparing the performance of any defocus deblurring technique. Here we assume that the image intensity $f(x)$ is measure in units of energy per unit area (irradiance multiplied by time), and the PSF $h_(x, y)$ is measured in units of area (in the aperture plane), which ensures that $J = F(0, 0)H(0, 0)/(D_x \cdot D_y)$ is in units of photons. Under the assumption that the MTF is essentially depth invariant, the MTF for the baseline imaging system is constant

$$H_{db}(\omega_x, \omega_y) = A_b^2, \tag{5.95}$$

and the MSE is

$$MSE_{db} = \frac{\sigma_s^2}{A_b^4}. \tag{5.96}$$

5.3.3.2 Defocus Blur and Diffractive Blur

Defocus blur is a purely geometrical phenomenon that depends only on object depth, aperture size, and the focal length of the lens. However, lenses also exhibit some amount blur due to the diffraction of light from the aperture. Furthermore, while defocus blur is directly proportional to aperture size, diffractive blur is inversely proportional to the aperture size. Therefore, the two types of blur compete, and any attempt to remove one type of blur will increase the other.

To make a fair comparison between the baseline case with the stopped down aperture and an EDOF system, we need to account for the diffractive blur in both cameras. We have two choices, we can either assume that the diffractive blur is removed from both cameras, or that it is not removed from either camera. To make the problem easier to analyze, we take the former approach and assume that diffractive blur is removed from the stopped down camera.

Under the assumption that the MTF is essentially depth invariant for the baseline case, the MTF for a lens with a square aperture of width A_b is

$$H_{db}(\omega_x, \omega_y) = A_b^2 \Lambda\left(\frac{\omega_x}{\rho}\right) \Lambda\left(\frac{\omega_y}{\rho}\right), \tag{5.97}$$

where $\Lambda(x) = (1 - |x|)$ is the tophat function, $\rho = A_b/(\lambda d_0)$, and λ is the average wavelength of light. The MSE for the baseline imaging system taking diffraction into account becomes

$$\begin{aligned} MSE_{db} &= \sigma_m^2 A_b^2 \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \frac{1}{(1 - |\frac{\omega_x}{\rho\Delta}|)^2 (1 - |\frac{\omega_y}{\rho\Delta}|)^2} d\omega'_y d\omega'_x, && (5.98) \\ &= \sigma_m^2 A_b^2 \left(\frac{2\rho\Delta}{2\rho\Delta - 1}\right)^2, && (5.99) \\ &= \sigma_m^2 A_b^2 \left(\frac{\rho}{\rho - \Omega}\right)^2. && (5.100) \end{aligned}$$

The MSE for the baseline case then depends on the relationship between the Nyquist frequency Ω and the diffraction cutoff frequency ρ. The expression for the MSE in Equation 5.100 is only correct when $\rho < \Omega$. For any larger values of ρ, the MSE becomes infinite because the MTF at frequencies in the range $\omega_x, \omega_x > \rho$ have zero magnitude. Note that implicitly puts a constraint on the maximum F/# for the baseline imaging system

$$F\# = \frac{d_0}{A_b} \leq \frac{2\Delta}{\lambda}. \tag{5.101}$$

For instance, for a pixel size of $\Delta = 2\mu m$, and assuming an average wavelength of $\lambda = .5\mu m$, the maximum allowable F/# is F/8. This brings us to an important point about extending DOF. It may be impossible to reduce the aperture to a small enough size without reducing the diffraction cutoff frequency below the Nyquist frequency. When this is the case, so long as it is necessary to maintain some energy at frequencies larger than the diffraction cutoff, the only option available for doing so while at the same time maintaining the desired DOF is to use an EDOF technique. In this case, the benefit of using an EDOF technique does not come in the form of SNR gain relative to stopping down the aperture, it is simply the only way to achieve the desired DOF extension and at the same time fend off the negative effects of diffraction.

5.3.3.3 Defocus Invariant Blur

The problem of producing depth invariant blur is analogous to the problem of producing motion invariant blur. When the aperture width is increased by a factor of $\sqrt{C}$ to $A = \sqrt{C} \cdot A_b$, the amount of light captured increases by a factor of C, but the maximum blur size also increases by a factor of $\sqrt{C}$, so that the defocus blur is now larger than one pixel. When defocus blur is larger than one pixel, the blur can be engineered to become defocus invariant, allowing it to be be removed via deconvolution without the need to first estimate depth. As in the case of motion blur, the question becomes whether or not the combination of increased light and deconvolution results in a net gain in MSE.

While several techniques have been developed to extend DOF, there is no way to completely remove depth-dependent blur from a camera with a finite aperture size. The best that can be done is to produce depth-independent blur over some depth range. Baek analyzed the tradeoff between MTF and depth-invariance [Baek, 2010]. He derived a bound on the optimal tradeoff that can be achieved, and showed that that the focal sweep technique is nearly optimal in simultaneously maximizing MTF and depth-invariance. Levin et al. derived an approximate expression for the focal sweep MTF, as well as an upper bound on the maximum MTF that can be achieved regardless of the degree of depth-indpendence. The approximate MTF for focal sweep (H_{fs}), and the upper bound for any technique (H_{ub}) are

$$\|H_{fs}(\omega_x,\omega_y)\|^2 \approx \frac{CA_b^2\alpha^2(\omega_x,\omega_y)}{S_d^2|\omega_r|^2}\Lambda\left(\frac{\omega_x}{C\rho}\right)\Lambda\left(\frac{\omega_y}{C\rho}\right), \tag{5.102}$$

$$\|H_{ub}(\omega_x,\omega_y)\|^2 \leq \frac{C^{3/2}A_b^3}{S_d|\omega_r|}\Lambda\left(\frac{\omega_x}{C\rho}\right)\Lambda\left(\frac{\omega_y}{C\rho}\right), \tag{5.103}$$

where $\alpha(\omega_x,\omega_y) = \frac{|\omega_r|}{\max(|\omega_x|,|\omega_y|)}$, which has values in the range $[1,\sqrt{2}]$. In this setting, the range of depths is set by S_m, and the blur is depth invariant for depths that would otherwise produce blur widths less than or equal to the maximum blur size $b_{max} = \sqrt{C}A_bS_m/2$.

The upper bound in Equation 5.103 is significantly larger than the focal sweep MTF given by Equation 5.103. Levin et al. introduced the lattice focal lens, which comes closest to achieving the upper bound in Equation 5.103. However, the technique does not produce a depth invariant PSF, and therefore requires depth estimation in order to apply deconvolution to captured images.

5.3.3.4 SNR Gain for Focal Sweep

The focal sweep technique creates a depth-independent blur by utilizing sensor motion during exposure. The sensor is translated along the optical axis during exposure so that the each object is in focus during at least one instant. In the setting analyzed here, the depth range swept by the sensor is equal to the range of object depth in the scene.

We first simplify the expression for the focal sweep MTF by ignoring the effect of diffraction on the opened up aperture

$$\|H_{fs}(\omega_x, \omega_y)\|^2 \leq \frac{CA_b^2}{S_d^2 \max^2(|\omega_x|, |\omega_y|)}. \tag{5.104}$$

This upper bound will be tight when we open up the aperture much larger than the baseline case. Then C is large and the new diffraction cutoff frequency becomes much larger than the Nyquist frequency (i.e. $C\rho \gg \Omega$ and $\Lambda(\frac{\omega_y}{C\rho}) \approx 1$). Equation 5.104 is a convenient expression to work with analytically. However, it is important to keep in mind that the upper bound becomes weak when C is small.

We observe that the max function can be broken into four quadrants which integrate to the same amount, and that for the quadrant $\omega_x > |\omega_y|$, $\max(\omega_x, \omega_y) = \omega_x$. We can then change MSE so that the x integrand becomes $[0, 1/2]$, and the y integrand becomes $[-\omega_x, \omega_x]$, and the MSE becomes

$$\begin{aligned}
MSE_{fs} &\geq 4\sigma_m^2 \int_0^{1/2} \int_{-\omega'_x}^{\omega'_x} \frac{1}{\|H_{fs}(\omega'_x, \omega'_y)\|^2} d\omega'_y d\omega'_x, && (5.105)\\
&= \frac{4\sigma_m^2 S_d^2}{CA_b^2 \Delta^2} \int_0^{1/2} \int_{-\omega_x}^{\omega_x} {\omega'}_x^2 d\omega'_y d\omega'_x, && (5.106)\\
&= \frac{32\sigma_m^2}{CA_b^4} \int_0^{1/2} {\omega'}_x^3 d\omega'_x, && (5.107)\\
&= \frac{\sigma_m^2}{2CA_b^4} && (5.108)\\
& && (5.109)
\end{aligned}$$

The multiplexing gain for focal sweep is then

$$Q_{fs}^2 \leq 2C \left(\frac{\rho}{\rho - \Omega}\right)^2 \frac{\sigma_s^2}{\sigma_m^2} \tag{5.110}$$

The SNR gain for focal sweep equal to the SNR gain for 1D motion invariant photography, weighted by a diffraction dependent term. When diffraction can be ignored, the two expressions are identical. As a result, the optimal increase in aperture area, the optimal increase in exposure time for motion invariant photography, and the optimal row sum for multiplexing are all equal to the same value C_{max} given in Equation 5.48. The upper bound on the SNR gain for focal sweep is

$$Q_{fs}^2 \leq \left(2 + \frac{2\sigma_0^2}{J}\right)\left(\frac{\rho}{\rho - \Omega}\right)^2. \tag{5.111}$$

For focal sweep, the SNR gain is always greater than $\sqrt{2}$. However, the amount greater than $\sqrt{2}$ decreases when either the amount of diffraction or the signal level increases. The bound significantly reduces the space of imaging scenarios that will give the desired multiplexing gain. In this case, when using the Canon Mark II, it is not possible to achieve a multiplexing gain greater than 3 if the non-multiplexed signal level is greater than $32e^-$.

5.3.3.5 DOF Extension for Focal Sweep

In the previous section we asked what the SNR gain will be for an EDOF system, relative to a stopped down camera, when considering a fixed range of depths. Now we ask instead what is the maximum extension in DOF we can achieve for an EDOF system, while at the same time ensuring that SNR is not less than a stopped down camera. We increase the depth range $d_{max} - d_{min}$ by a factor of M_{fs} so that the parameter S_d also increases by a factor of M_{fs}. Then the bound on the focal sweep MTF becomes

$$\|H'_{fs}(\omega_x, \omega_y)\|^2 \leq \frac{CA_b^2\alpha^2(\omega_x, \omega_y)}{M_{fs}^2 S_d^2 |\omega_r|^2}, \tag{5.112}$$

and the multiplexing gain becomes

$$Q'^2_{fs} \leq 2\frac{C}{M_{fs}^2}\frac{\sigma_s^2}{\sigma_m^2}, \tag{5.113}$$

and enforcing the constraint that the SNR gain Q'_{fs} is greater than one gives the upper bound on the DOF extension

$$M_{fs} \leq Q_{fs}. \tag{5.114}$$

This result is slightly different than the result for motion invariant photography. Here the upper bound on the increase in DOF M_{fs} for focal sweep is equal to the bound on the

SNR gain (not the square of the SNR gain). The difference here is due to the fact that the amount of light for the focal sweep system is proportional to the aperture area, which is proportional to the square of the DOF. In this case, Equation 5.111 gives a bound on both the maximum SNR gain that can be achieved for a fixed range of depths, and the maximum increase in the range of depths that can be achieved for a fixed SNR. The upper bound is always greater than one for all signal levels. When diffraction effects are negligible, the bound approaches $\sqrt{2}$ for large signal levels so that both the maximum SNR gain and DOF extensions become equal to $\sqrt{2}$. The bound increases when the amount of diffraction decreases, which happens when the F/# for the stopped down camera becomes much larger than the ratio between the pixel size and the wavelength of light.

5.3.3.6 Focal Sweep Summary

We summarize the results for focal sweep by considering two different scenarios that demonstrate the effect of diffraction on the SNR gain. In the first scenario, we assume that the diffraction cutoff frequency for the stopped down camera is twice the cutoff frequency ($\rho \geq 2\Omega$). This implies that for the stopped down camera, $F/\# = \Delta/\lambda$. For instance, assuming a pixel size of $4\mu m$ and a center wavelength of $\lambda = .5\mu m$, the F/# is then F/8. When this is the case, for large signal levels, the DOF extension for the focal sweep technique can be no greater than a factor of $2\sqrt{2}$. Thus, when the signal is large, the absolute best we can hope to do with focal sweep without decreasing the SNR below the stopped down camera is to open up the aperture to F/2.8. This is a reasonably large extension in DOF. However, we point out that in order to derive this bound, we accounted for the effects of diffraction in the uncoded measurement, but not the coded one. Thus, the actual amount we will be able to open up the aperture will be less than F/2.8. A numerical simulation taking into account diffraction in the multiplexed measurement found that the optimal aperture setting to be F/3.5 instead of F/2.8.

When the diffractive blur is large for the stopped down camera, the benefit of focal sweep is large enough to be attractive. However, the benefit decreases significantly when the diffractive blur decreases. For instance, consider the case when the diffraction cutoff frequency is at least four times the cutoff frequency ($\rho \geq 8\Omega$). This implies that for the stopped down camera, F/# $> \Delta/(2\lambda)$. Then, assuming we have larger $16\mu m$ pixels, the F/# remains at F/8. The DOF extension for the focal sweep technique can then be no greater than a factor of $8/7 \cdot \sqrt{2} \approx \sqrt{2}$. In this scenario, when the signal is large, the absolute best we can hope to do with focal sweep without decreasing the SNR below the stopped down camera is to open up the aperture to about F/5.6. This is a relatively marginal improvement that most likely will not warrant the effort required in implementing the EDOF technique.

Since the focal sweep technique is nearly optimal in simultaneously maximizing MTF and minimizing PSF invariance, we expect that this is the best possible DOF extension without requiring depth estimation. Thus, we expect Equation 5.111 to summarize the best possible performance of any EDOF technique that does not require depth estimation, including the diffusion coding technique discussed in Chapter 2 and spectral focal sweep technique discussed in Chapter 3. However, note that the previous two examples illustrate the best possible performance when signal levels are large. The focal sweep technique will have maximum performance advantage for small signal levels. In this case, the DOF extension can increase significantly beyond the performance described in the previous two paragraphs. We also reiterate a point made in Section 5.3.3.2: when the diffraction cutoff frequency is larger than the Nyquist frequency ($\rho > \Omega$), the only available option for maintaining energy at all frequencies, while at the same time maintaining the desired DOF, is to use an EDOF technique. This is a particularly important problem when considering cameras with small pixel sizes that have large Nyquist frequencies.

As in the case for motion invariant photography, the upper bound here has been derived under the assumption that the range of object depths are distributed uniformly. It is of course possible to imagine scenarios where prior information on the distribution of object depths is known, in which case a depth invariant MTF is no longer optimal, and the upper bound in Equation 5.110 is no longer valid.

5.3.3.7 SNR Gain for Optimal Defocus Coding

We use the 4D light field parameterization to analyze light propagating through lenses. However, neglecting the effect of occlusions and highly specular reflectances, the light field produced by a set of image points focused at different depths is parameterized by only 3D spatial coordinates. Interestingly, this 3D subspace corresponds to a 3D manifold in 4D Fourier Space [Ng, 2005] [Levin *et al.*, 2009]. When a camera maximally preserves energy along this 3D manifold, it also maximizes the MTF that can be achieved across a range of depth values, giving an upper bound on the MTF H_{ub}. We emphasize that the upper bound H_{ub} is the maximum possible MTF that can be achieved while maximally distributing energy along the focal manifold, regardless of the degree of depth-invariace. For instance, for the lattice focal lens, which comes closest to the upper bound, produces a highly depth-dependent blur [Levin *et al.*, 2009]. We may therefore consider H_{ub} as a weak upper bound for depth-invariant cameras. On the other hand, H_{ub} is a tight upper bound on the maximum MTF that can be achieved over a range of depths when depth is known *a-priori*.

The lower bound on the MSE for any camera imaging over a range of depths is then

$$MSE_{ub} \leq \frac{\sigma^2 S_d}{C^{3/2} A_b^3 \Delta} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \sqrt{\omega_x'^2 + \omega_y'^2} d\omega_x' d\omega_y' \tag{5.115}$$

$$= \frac{\sigma^2 S_d}{C^{3/2} A_b^3 \Delta} \frac{\sqrt{2} + \sinh^{-1}(1)}{6} \tag{5.116}$$

$$< \frac{\sigma^2 S_d}{2C^{3/2} \cdot A_b^3 \Delta^2}. \tag{5.117}$$

The best possible SNR gain for any camera optimized over a range of depths is then

$$Q_{ub}^2 \leq \frac{MSE_{db}}{MSE_{ub}}. \tag{5.118}$$

$$= \frac{2C^{3/2}\Delta}{A_b S_d} \frac{\sigma_s^2}{\sigma_m^2} \left(\frac{\rho}{\rho - \Omega} \right) \tag{5.119}$$

Under the assumption that the aperture is sufficiently large for the stopped down camera that $\rho > 2\Omega$, after substituting the expression for the noise ratio from Equation 5.39, the bound on the SNR gain becomes

$$Q_{ub}^2 < 4C^{3/2} \frac{J + g_s^2 \sigma_1^2 + \sigma_0^2}{CJ + C^2 g_s^2 \sigma_1^2 + \sigma_0^2} \tag{5.120}$$

This upper bound on SNR gain increases more rapidly with aperture size than focal sweep. We can solve for the optimal increase in aperture size C_{ub} by taking derivatives

$$C_{ub} = \frac{J + \sqrt{12\sigma_1^2\sigma_0^2 g_s^2 + J^2}}{2g_s^2\sigma_1^2} \tag{5.121}$$

$$\approx \frac{J}{g_s^2\sigma_1^2}, \tag{5.122}$$

where we have made the approximation based on the assumption that the signal is sufficiently large so that $J^2 >> 12g_s^2\sigma_1^2\sigma_0^2$. Substituting C_{ub} into the equation for the upper bound gives

$$Q_{ub}^2 \leq 2\frac{J + g_s^2\sigma_1^2 + \sigma_0^2}{\sqrt{J}g_s\sigma_1}. \tag{5.123}$$

Again, assuming the signal is sufficiently large, $J \gg g_s^2\sigma_1^2 + \sigma_0^2$, and the SNR gain becomes

$$Q_{ub}^2 \leq 2\frac{\sqrt{J}}{g_s\sigma_1}. \tag{5.124}$$

Since we need to increase the gain to avoid saturation as the signal level increases, the expression in the numerator increases also, until the maximum gain level is reached. Substituting the expression for camera gain from Equation 5.36, the SNR gain becomes

$$Q_{ub}^2 \leq 2\frac{ADU}{\sigma_1 D\sqrt{J}}. \tag{5.125}$$

So, once the once the signal level is large enough so that photon noise dominates read noise, the bound on the SNR gain actually decreases monotonically with increasing signal level until the signal saturates. Unless the dynamic range is exceptionally large, the bound will always be larger than one. For instance, we consider the case when $D = 8$, which is a typical value for natural images. Then, if a Canon 1D Mark II is used, the signal will saturate when $J = 10,000e^-$. The read noise at the largest camera gain setting is $g_s\sigma_1 \approx 30.6e^-$, so the SNR gain would be $Q_{ub} < 2.5$ for a saturated signal.

5.3.3.8 Upper Bound on DOF Extension

Following the same procedure as in Section 5.3.3.5, we find that for a DOF extension of M_{ub} that does not sacrifice SNR,

$$M_{ub} \leq Q_{ub}^2. \tag{5.126}$$

Thus, Equation 5.124 gives an upper bound on both the maximum squared SNR gain, and the maximum DOF extension that can be achieved with any technique. This upper bound applies to all EDOF techniques, including both techniques that do and do not require depth estimation. The MTF for the lattice focal technique from Levin et al. comes close to the upper bound in Equation 5.124, but it requires depth estimation, which propagates errors in the deconvolution process that increase the MSE and reduce the SNR gain.

5.3.3.9 EDOF Upper Bound Summary

In summary, The upper bound on the SNR gain for any EDOF technique given by Equation 5.124 is clearly greater than the upper bound for focal sweep given by Equation 5.110. Therefore, an EDOF technique that comes close to the upper bound, such as the lattice focal lens, will maintain a performance advantage even at larger signal levels. However, the focal sweep is near optimal at maximizing MTF and depth-invariance. We therefore expect that an EDOF technique which comes closer to the bound given by Equation 5.124 will require depth estimation, and errors in the estimation will effectively reduce the SNR below this bound.

5.4 Conclusion

In Section 5.2 we found a bound on the maximum SNR gain that any multiplexing technique can achieve. In Section 5.3 we found bounds for motion invariant and EDOF techniques. The performance of all these techniques depends on the signal level. At small signal levels, all techniques do well. At large signal levels, the SNR gains approach a maximum. We summarize the results in Table 5.1, where we show the upper bound on the performance at large signal levels for all techniques.

	SNR Gain	Extension
Multiplexing	$G \leq 1$	-
1D Motion Invariant	$Q_{1D} < \sqrt{2}$	$M_{1D} < \sqrt{2}$
2D Motion Invariant	$Q_{2D} < \sqrt{4/3}$	$M_{2D} < \sqrt{4/3}$
Focal Sweep	$Q_{fs} < \sqrt{2}\frac{\rho}{\rho-\Omega}$	$M_{fs} < \sqrt{2}\frac{\rho}{\rho-\Omega}$
EDOF Upper Bound	$Q_{ub} < \sqrt{2}\frac{J^{1/4}}{\sqrt{g_s\sigma_1}}\frac{\rho}{\rho-\Omega}$	$M_{ub} < 2\frac{\sqrt{J}}{g_s\sigma_1}\left(\frac{\rho}{\rho-\Omega}\right)^2$

Table 5.1: The SNR gain for several techniques at large signal level. From top to bottom, the techniques are multiplexing, 1D motion invariant photography, 2D motion invariant photography, focal sweep, generalized EDOF. The middle column shows the SNR gain, and the right column shows the motion extension for motion invariant photography, and the defocus extension for focal sweep and generalize EDOF.

Summarizing Table 5.1, there is no SNR gain for multiplexing techniques. The maximum SNR gain for 2D motion invariant photography is only $\sqrt{4/3}$. This means that any extension of motion for 2D motion invariant photography will be negligible if the SNR gain is at least one. It is possible to get a rather negligible SNR gain or motion extension of $\sqrt{2}$ for 1D motion invariant photography, but this requires motion to be restricted along a line.

Analyzing the performance of the EDOF techniques is slightly more complicated. When the effects of diffraction can be ignored, the SNR gain and DOF extension for focal sweep is $\sqrt{2}$, which is negligible. This happens when the diffraction cutoff frequency is much larger than the Nyquist frequency, which only occurs when the F/# is much smaller than the ratio of the pixel size to the wavelength of light. Since focal sweep is near optimal at maximizing MTF and depth invariance, we expect that the performance of focal sweep is an upper bound on any depth-invariant technique. The performance can be significantly increased if the EDOF technique is allowed to exhibit depth-dependence, as indicated by the EDOF upper bound, which outperforms focal sweep at all signal levels.

5.4.1 Signal Levels and Lighting Conditions

While the performance of most coding techniques reaches a threshold at large signal levels, the performance is always greater at smaller signal levels. The reason is that the amount of photon noise decreases at smaller signal levels so that read noise dominates in the non-coded measurement. Assuming a read noise value of $\sigma_0 = 4e^-$, which is a typical value for today's high quality CMOS sensors, the upper bounds in Table 5.1 are correct to within four percent when the signal is greater than $100e^-$. To give an idea of what lighting conditions will result in a signal of $J = 100e^-$, we derive the relationship between illumination irradiance and the number of photons collected by a pixel.

For convenience, we will work with photometric units, which are equivalent to radiometric units weighted by the spectral response of the standard human observer $S(\lambda)$. This will allow us to derive an expressions for the number of photons collected by a pixel in terms of standard illumination conditions. Let us assume that the scene being photographed is lambertian with reflectance R, and is lit by a source with illuminance E_{src}. Then the luminance of the object is

$$L = \frac{R \cdot E_{src}}{\pi}. \tag{5.127}$$

The illuminance falling on the detector is [Horn, 1986]

$$E_{det} = \frac{1}{4}\frac{1}{F/\#^2} E_{src} R. \tag{5.128}$$

We assume that the detector spectral response $V(\lambda)$ is matched to the spectral response of the standard human observer $S(\lambda)$. For most color cameras, this will be the case for the luminance channel of captured images. Letting the quantum efficiency of the detector be η, the irradiance collected by the detector in Joules/m^2/s is

$$I_{det} = \frac{1}{680 \text{ lumens/watt}} \frac{1}{4}\frac{1}{F/\#^2} E_{src} R \eta. \tag{5.129}$$

The energy in Joules of a single photon is given by $\hbar c/\lambda$, where $\hbar$ is Planck's constant, and c is the speed of light. Assuming a mean wavelength of $\lambda = .55\mu m$, the energy per unit photon collected by a pixel can be approximated as

$$\frac{\hbar c}{\lambda} \approx 3.6 * 10^{19}. \tag{5.130}$$

The irradiance in photons/μm^2/s is then

$$I_{photon} \approx 10^4 \frac{1}{F/\#^2} E_{src} R\,\eta. \tag{5.131}$$

So that for a pixel with area Δ^2 in μm^2 and exposure time t seconds, the total number of photons collected is

$$J \approx 10^4 \frac{1}{F/\#^2} E_{src} R\,\eta\,\Delta^2\,t. \tag{5.132}$$

As the F/# is increased, and the aperture is stopped down, the number of photons collected by a pixel decreases. However we typically don't allow the F/# to increase so that the diffraction cutoff frequency increases beyond the Nyquist frequency. The minimum number of photons collected then occurs when the diffraction cutoff frequency and the Nyquist frequency are equal, which means that $F/\# = 2\Delta/\lambda$. The minimum photon count is then

$$J_{min} \approx 10^4 \frac{\lambda^2}{4} E_{src} R\,\eta\,t. \tag{5.133}$$

Typical illumination conditions and the corresponding photon counts are shown in Table 5.2. In the last column, we calculate the photon count assuming a reflectivity of $R = .5$, and a quantum efficiency of $\eta = .5$ (typical of a high quality sensor). The number of photons increases with exposure time, but we assume that the camera should be able to operate at video rates, which limits the maximum exposure time that can be used. Here we use the maximum exposure time that will achieve video rates, which we assume to be $t = 1/50$

seconds. From the table, we can see that even for very weak illumination conditions (living room lighting), the largest possible F/#, and the longest possible exposure time, enough photons are collected so that the bounds in Table 5.1 are correct to within nearly four tenths of one percent.

	lumens/m^2	photons/μm^2/s	photons
Starlight	10^{-4}	5	< 1
Full Moon	1	$5 * 10^4$	19
Living Room	50	$2.5 * 10^6$	965
Office Lighting	$4 * 10^2$	$2 * 10^7$	$7.7 * 10^3$
Overcast Day	10^3	$5 * 10^7$	$1.9 * 10^4$
Daylight	10^4	$5 * 10^8$	$1.9 * 10^5$
Direct Sun	10^5	$5 * 10^9$	$1.9 * 10^6$

Table 5.2: Lighting conditions and their corresponding illuminance in terms of photon counts. The left-most table shows typical illuminance values in lumens/m^2 for different lighting conditions. The center column shows the same values in terms of photons/μm^2/s. The right column shows the photon counts calculated using Equation 5.133 assuming a reflectivity of $R = .5$, quantum efficiency of $\eta = .5$, and exposure time of $t = 1/50$ seconds. Even for living room lighting conditions, enough photons are collected so that the bounds in Table 5.1 are correct to within four tenths of one percent.

5.4.2 Discussion

In Chapter 1 we discussed that computational imaging systems offer advantages in terms of increased functionality and performance. In chapters 2 and 3, we introduced new techniques to provide increased functionality in the form of computationally extending DOF. However, this work left many questions open about the performance advantages of computational imaging systems. In this chapter, we have introduced a comprehensive framework for analyzing the performance of a large class of computational imaging systems. We used this framework to derive performance limits for a variety of computational imaging techniques.

The analysis in this chapter provides insight into the scenarios where a computational technique provides a performance advantage. However, the analysis itself is somewhat restrictive, and it should be understood that there are some limitations to the conclusions that can be made. We have used the MSE as a metric to quantify errors in the decoding process of computational imaging systems. We have also assumed a linear decoding process. A linear decoding process makes the MSE a convenient measure because analytic calculations, such as the bounds derived in Sections 5.2 and 5.3, become tractable. However, a linear decoding process is not always optimal, and MSE is not always the best measure of performance.

5.4.2.1 Image Priors

In many cases, we have access to useful information prior to performing the decoding process that is useful in reducing decoding errors. For instance, in Chapters 2 and 4, we took advantage of the fact that the distribution of Fourier spectra for natural images decays rapidly with increasing spatial frequency. This property is also closely related to the fact that the distribution of gradient magnitudes of natural images also decays rapidly. Both of these properties have been exploited to increase the accuracy of the decoding process [Levin *et al.*, 2007] [Zhou and Nayar, 2009] [Cho *et al.*, 2010]. In some cases, the decoding process is still linear, in some cases it is not. However, in either case, it no longer becomes analytically tractable to calculate the MSE.

Prior information on the distribution of the class of input signals provides information on how compactly signals can be represented. The concept of an image prior is generalized by the concept of sparse modeling and compressed sensing. In compressed sensing, we operate under the assumption that measured signals can be represented as a sparse set of coefficients in some prior basis. This assumption works well when the signal coefficients in the prior basis decay rapidly. Then the signal can be reconstructed faithfully from only a small set of the coefficients. For instance, natural images are typically sparse in the Haar Wavelet basis. When the assumption of sparsity is correct, a non-linear reconstruction technique can be used to decode captured images with better performance than a linear recovery process. Compressive imaging systems have been developed for a variety of imaging modalities, including superresolution [Marcia and Willett, 2008] [Duarte *et al.*, 2008], imaging spectrometry [Wagadarikar *et al.*, 2009] [Wagadarikar *et al.*, 2008], space-time volumes [Reddy *et al.*, 2011] [Hitomi *et al.*, 2011], 3D measurements of transparent objects [Gu *et al.*, 2008], and light transport matrices [Peers *et al.*, 2009]. However, the non-linearity in the recovery process makes it difficult to compare the performance of a compressive imaging technique to a non-compressive one.

While the bounds derived in Sections 5.2 and 5.3 are only strictly applicable to the case when a linear decoding process is used, we do expect that they will still be strongly informative in all cases. The reason is that, in our analysis, we have compared the performance of coded and uncoded imaging systems. While the multiplexed performance will certainly improve when image priors are taken into account, the non-multiplexed performance will improve also. The question becomes: "what will the ratio of performance improvement be?". This is a very deep question that depends strongly on the amount of sparsity, the prior basis, and the multiplexing matrix, and is therefore difficult to make generalizations about. One approach may be to use the coherence between the prior basis and the multiplexing matrix as a performance metric, since the quantity bounds the MSE. Then the performance could be quantified in terms of the sparsity of the signal. The bounds derived in Sections 5.2 and 5.3 would hold for non-sparse signals, and the bounds would increase with increasing sparsity (i.e stronger image priors).

5.4.2.2 Limitations of the MSE Metric

To quantify the quality of a decoded image we need a single representative value that takes into account all the errors in a recovered image. In reality each pixel will have a different error and we will have a distribution of errors in our recovered image. When we use the MSE to calculate the performance, we choose the mean error across all pixels as the representative value. This may give us a biased result, particularly if there is a large variation in the error at each pixel. For instance, if there just a few pixels with very large error, the average error calculated by the MSE may be quite large, when in fact most of the image has been decoded faithfully. In this case, the MSE will not be a desirable metric for evaluating performance. Other metrics, such as the L1 norm, would be more suitable in this case. The L1 norm would choose the median error as the representative value instead of the mean. Since the median is more robust to outliers, this performance metric would eliminate bias created by small groups of pixels with large errors.

Part III

Conclusions

Chapter 6

Conclusions on the Computational Imaging Advantage

In this thesis we have given a thorough analysis of computational imaging systems. In Chapter 1, we established that the general goal of any imaging system is to capture a slice of the plenoptic function. Conventional cameras are extremely limited in the way they can sample the plenoptic function. In contrast, computational cameras open up a world of possibilities by allowing for more flexible sampling strategies. We can use computational cameras to estimate depth, or correct for geometric distortions in the lens. We can also use a computational approach to measure more of the plenoptic function at once – we can capture hyperspectral volumes, space-time volumes, and light fields with just a 2D sensor. We can also create a many to one mapping between plenoptic samples and pixels, such as the blur produced by defocus blur and geometric aberrations, which were studied extensively in Chapters 2, 3, and 4.

In Chapters 2, we introduced the diffusion coding technique for capturing Extended Depth Of Field (EDOF) images. The technique works by placing a radially symmetric diffuser inside the aperture of a lens. The diffuser codes the light propagating to the sensor in such a way that blur becomes depth-independent. The result is that an EDOF image can be recovered by deblurring without the need for depth estimation. Diffusion coding produces nearly identical performance to the most optimal EDOF technique known today:

focal sweep. However, focal sweep requires the use of moving parts, while diffusion coding does not.

In Chapters 3, we introduced the spectral focal sweep technique for capturing EDOF images. This technique uses a lens that has intentionally not been corrected for chromatic aberrations. These aberrations produce a focal length that is dependent on wavelength, essentially imaging onto multiple focal planes simultaneously. When objects with broadband reflectance spectra are imaged through a spectral focal sweep lens onto a black and white sensor, the captured images exhibit depth-independent blur in a manner that is identical to a focal sweep camera. Spectral focal sweep enables EDOF images to be captured without the use of moving parts, and at the same time reduces lens complexity. While the technique works best when reflectance spectra are perfectly broadband, it works well for more general scenes that consist of naturally occurring reflectance spectra.

In Chapter 4, we introduced the gigapixel computational imaging technique for computationally increasing optical resolution. Increasing the optical resolution of a camera is a challenging problem because resolution is fundamentally limited by geometric aberrations. Conventional cameras stop down the lens aperture to reduce the amount of aberrations, but this increases the amount of diffractive blur, which forces camera size to increase. On the other hand, some aberrations can be removed via deblurring without significantly reducing image quality. We showed that, by taking a computational approach, high resolution cameras can be built with a very compact size, and a very simple lens.

Chapters 2, 3, and 4 introduced three new coding techniques. For each technique, we compared the performance of our computational imaging implementations to convention cameras wherever applicable. However, the comparisons were limited to a few imaging conditions, and did not give conclusive answers about the performance over all possible imaging conditions. This led us to pursue a deeper analysis on the limits of computational imaging systems in Chapter 5. Here we considered computational imaging systems that use coding to increase optical efficiency. From a naive perspective, an increase in optical efficiency translates directly to an increase in performance, since it increases the signal strength of captured images. However, an increase in optical efficiency does not always translate directly to an increase in performance. To properly analyze performance, we must take into account a detailed noise model that takes into account signal dependent noise. We also must take into account the conditioning of the imaging system. In Chapter 5, we analyzed a wide variety of computational imaging techniques using the general imaging model introduced in Chapter 1, including general multiplexing techniques that code information by masking light, as well as fully efficient, shift invariant techniques that code for invariance.

6.1 Tradeoffs in Computational Imaging

Tradeoff	Where Discussed
Efficiency vs. Functionality	Sections 2.1, 3.1, and 4.5
Best vs. Average Performance	Sections 2.1 and 3.1
Resolution vs. Scale	Section 4.7
Performance vs. Complexity	Sections 3.1 and 4.9.3

Table 6.1: Tradeoffs in computational imaging. Each tradeoff is listed along with the corresponding sections in this thesis where the tradeoff is discussed.

Looking closer at the coding techniques in Chapters 2, 3, and 4 led us to a number of tradeoffs that reappeared throughout this thesis (see Table 6.1). Perhaps the most significant tradeoff we saw was between optical efficiency and functionality. In Chapters 2 and 3, functionality came in the form of increased DOF. In Chapter 4, functionality came in the form of increased optical resolution. We saw a tradeoff between best case and average case performance for EDOF systems. Our EDOF systems attempt to maximize the performance averaged over a range of depths, which prevents them from achieving the best possible performance at a single depth. In Chapter 4, we saw that functionality, performance, and complexity are closely interlinked. We explored the tradeoff between resolution and camera scale, and we said that simpler optics could be used to increase resolution, but at the cost of decreased performance. This led us to further investigate the relationship between performance and complexity. We said that complexity has a cost associated with it, and the cost of increased complexity is not always warranted by the achieved performance gain. In Chapter 3, we introduced the spectral focal sweep camera, which extends DOF by reducing the complexity of the lens. We used a simple lens to increase DOF, but the cost was reduced color imaging performance. In Chapter 4, we explored the tradeoff between complexity and resolution in the context of spherical optics. We saw that there is a law of diminishing return – an increase in complexity does not always result in an equal increase in performance.

6.2 The Limits of Computational Imaging

In Chapter 5, we analyzed the performance of computational imaging systems using a detailed noise model, and the conclusions were somewhat surprising. The lighting conditions must be extremely dim in order to achieve a performance advantage for a computational technique. For instance, if the lighting conditions are similar to a typical living room, the best possible SNR gain for a multiplexing technique is only a fraction of a percent. When the lighting conditions are similar to the light cast by a full moon, the best possible SNR gain is only forty percent. This has significant implications for computational techniques that capture hyperspectral volumes, color images, light fields, coded aperture defocus deblurring, and shuttered motion deblurring. In general, we see that these techniques are only useful when the average signal level of captured images is about the same as the read noise. This is only the case when the lighting conditions are extremely dim. In all other cases, a conventional imaging technique will perform at least as well, regardless of the efficiency of the computational imaging system.

We also saw some surprising results for the performance limits of techniques that code for invariance. We saw that the performance limit for 1D and 2D motion blur is only marginally better than for multiplexing techniques. We therefore draw the same conclusion that any motion invariant technique is only useful when the average signal level of captured images is about the same as the read noise. The performance of EDOF systems is not quite as straightforward to analyze. When we ignore diffraction, the performance of techniques that produce a depth-invariant PSF is the same as for 1D motion blur. We see a performance advantage of about forty percent at large signal levels, and the advantage increases when the average signal level is about the same as the read noise. However, when we take diffraction into account, the performance advantage can become significant, depending on the combination of F/# and pixel size. In some cases, the only way to extend DOF and also preserve information at all spatial frequencies is to use a computational technique. When we remove the restriction that the PSF must be depth-invariant, the performance limit significantly increases, and it becomes possible, in theory, to achieve a significant increase in performance even at very large signal levels. However, when the PSF is depth-dependent, it is necessary to perform depth estimation before deblurring, and errors in the estimation will reduce performance.

6.2.1 Measuring Performance

The conclusions drawn from the analysis on performance limits from Chapter 5 may seem to be largely negative, but we should keep in mind the assumptions made in the analysis before we develop strong convictions about the benefit (or lack thereof) of computational imaging as a whole. At first glance, it may be tempting to conclude that computational imaging rarely has any performance benefit. However, we should keep in mind that we have used MSE as a performance measure. The MSE metric is biased towards error measurements that are evenly distributed. When error measurements are widely distributed, large errors in a few measurements will strongly penalize performance.

In essence, the MSE is implicitly making assumptions about the imaging model. When images have a large dynamic range, photon noise is large for a small number of pixels. The MSE for these images will be large even though the errors are concentrated at a few measurement locations. The MSE may indicate that a computational technique will not have a performance advantage, when in fact, this will not be not be true for most measurements in the image – the error might be greater at a few pixels but significantly less at most pixels. The difficulty in using MSE as an error metric is that it gives large importance to a few errors that may have little importance.

At the end of the day, if we want a metric to compare performance across different measurement strategies, we need a way to represent the error at each pixel using a single number. The MSE is a convenient metric because, in many cases, it simplifies analysis.

Other metrics may be less biased, and consequently may result in different conclusions about the performance limits of computational imaging techniques.

The other important drawback of the performance analysis in Chapter 5 is that it assumed only linear inversion techniques. This excludes the use of non-linear reconstruction techniques that use image priors to improve performance. However, we do need to keep in mind that priors apply equally to images captured by both conventional and computational cameras, so we do not expect that priors will have a dramatic impact when comparing performance. More precise conclusions will require specific information about the type of prior and sampling method used. This is certainly a topic that deserves further attention. There are a number of image prior models worth exploring further, including the sparsity models discussed in 5, that may be lead to superior performance for computational imaging methods.

6.2.2 Computationally Increasing Efficiency

In this thesis, we have focused on techniques that produce a many-to-one mapping between the input signal and pixel measurements. This was central to the performance analysis of Chapter 5. It was also a critical part of the coding schemes introduced in Chapters 2 3, and 4. One general conclusion of this thesis is that we need to maintain a healthy skepticism when considering the efficiency benefit that results from a many-to-one mapping. As we've seen in this thesis, the decoding process can cause a surprisingly large decrease in performance.

We have given a thorough treatment of sampling the plentopic function, paying special attention to techniques that encode signals using multiplexing and shift invariant blur. However, there are many applications of computational imaging that do not require a many-to-one mappings such as this. In Chapter 1, we discussed a general image model, where signals are represented as coefficients in a representation basis. We can think of this basis as a type of image prior. The representation tells us what features embedded within the plenoptic function are most useful to measure. When using a computational imaging technique, we have the flexibility to adapt our sampling to measure these features directly, without any explicit coding. This may be the greatest benefit of computational imaging systems. Not that we have the flexibility in sampling to make explicitly coded measurements that can be decoded offline, for, as we have seen, the decoding process can be problematic. Rather, the crucial difference may be that we have the flexibility in sampling needed to directly record implicitly coded measurements. In Chapter 1, we referred to this strategy as “task-specific” imaging. From the task-specific perspective, we begin to ask the question “what is the most efficient sampling strategy for measuring the specific features we are interested in?”. This is clearly related to compressive imaging techniques discussed in Chapter 1, where we attempt to recover a large number of signal coefficients from a small number of pixel measurements. Both task-specific and compressive techniques focus on recovering the desired signal from as few measurements as possible. This is a natural extension of the computational imaging perspective, where we consider the imaging system as a channel that transmits visual information. This discussion on efficient measurement techniques is an interesting starting point for continued research on the topics covered in this thesis.

Part IV

Appendix

Appendix A

Diffusion Coding Derivations

In this appendix, we give derivations for several equations that were provided in Chapter 2, which analyzed the diffusion coding technique.

A.1 Derivation for Diffuser with constant 2D Scatter Function

The first derivation we give is for the PSF of a Diffusion Coded camera with a constant 2D scatter function, as described in Chapter 2. From Equation A.1, the kernel for this diffuser is

$$d(\mathbf{u}, \mathbf{u}', \mathbf{x}, \mathbf{x}') = \frac{1}{w^2}\delta(\mathbf{u} - \mathbf{u}') \sqcap \left(\frac{\mathbf{x} - \mathbf{x}'}{w}\right). \tag{A.1}$$

$$l'_\delta(\mathbf{u},\mathbf{x}) = \int\limits_{\Omega_\mathbf{u}}\int\limits_{\Omega_\mathbf{x}} \frac{1}{w^2}\delta(\mathbf{u}-\mathbf{u}')\sqcap\left(\frac{\mathbf{x}-\mathbf{x}'}{w}\right) l'_\delta(\mathbf{u}',\mathbf{x}')d\mathbf{u}'d\mathbf{x}' \tag{A.2}$$

$$= \frac{1}{w^2}\int\limits_{\Omega_\mathbf{x}} \sqcap\left(\frac{\mathbf{x}-\mathbf{x}'}{w}\right) l_\delta(\mathbf{u},\mathbf{x}')d\mathbf{x}' \tag{A.3}$$

$$h'(\mathbf{x}) = \frac{1}{w^2}\int\limits_{\Omega_\mathbf{u}}\int\limits_{\Omega_\mathbf{x}} \sqcap\left(\frac{\mathbf{x}-\mathbf{x}'}{w}\right) l_\delta(\mathbf{u},\mathbf{x}')d\mathbf{x}'d\mathbf{u} \tag{A.4}$$

$$= \frac{1}{w^2}\int\limits_{\Omega_\mathbf{x}} \sqcap\left(\frac{\mathbf{x}-\mathbf{x}'}{w}\right)\left[\int\limits_{\Omega_\mathbf{u}} l_\delta(\mathbf{u},\mathbf{x}')d\mathbf{u}\right]d\mathbf{x}' \tag{A.5}$$

$$= \frac{1}{w^2}\sqcap\left(\frac{\mathbf{x}}{w}\right)\otimes h(\mathbf{x}) \tag{A.6}$$

Which is the same result as Equation A.6, the result being that the effect of the diffuser is to blur the image E that would be captured were it not present.

A.2 Radially-Symmetric Light Field Derivation

In this section we verify mathematically that Equation 2.10 represents the light field of a unit energy point source. The equation for the light field is

$$l_\delta(\rho, r) = \frac{4}{\pi A^2}\sqcap\left(\frac{\rho}{A}\right)\frac{\delta(r - s_0\rho)}{\pi|r|}. \tag{A.7}$$

In polar coordinates, the energy e of a light field is calculated by integrating over all variables

$$e = \pi^2\int\limits_{\Omega_\rho}\int\limits_{\Omega_r} l_\delta(\rho, r)|\rho|d\rho|r|dr \tag{A.8}$$

or equivalently

$$e = \pi\int\limits_{\Omega_r} h(r)|r|dr, \tag{A.9}$$

where $h(r)$ is the PSF resulting from the image of the point source l_δ. The PSF for the point source is

$$h(r) = \pi \int\limits_{\Omega_\rho} \frac{4}{\pi A^2} \sqcap\left(\frac{\rho}{A}\right) \frac{\delta(r - s_0\rho)}{\pi|r|} |\rho| d\rho \tag{A.10}$$

$$= \frac{4}{\pi s_0^2 A^2} \frac{1}{|r|} \int\limits_{\Omega_\rho} \delta(r - \rho) \sqcap\left(\frac{\rho}{s_0 A}\right) |\rho| d\rho. \tag{A.11}$$

The integral in Equation A.11 is just a convolution between $\sqcap\left(\frac{r}{s_0 A}\right)|r|$ and a delta function. Thus, the resulting PSF takes the familiar shape of a pillbox with diameter $s_0 A$

$$h(r) = \frac{4}{\pi s_0^2 A^2} \sqcap\left(\frac{r}{s_0 A}\right). \tag{A.12}$$

The energy for the point source light field is then

$$e = \pi \int\limits_{\Omega_r} \frac{4}{\pi s_0^2 A^2} \sqcap\left(\frac{r}{s_0 A}\right) |r| dr \tag{A.13}$$

$$= \frac{4}{s_0^2 A^2} \int\limits_{s_0 A/2}^{-s_0 A/2} |r| dr \tag{A.14}$$

$$= 1, \tag{A.15}$$

which verifies that the point source has unit energy.

A.3 Radially-Symmetric Diffuser Derivation

We now give a derivation for the PSF of a Diffusion Coded camera using the diffuser kernel from Equation 2.14. The light field of a point source filtered by the radially symmetric kernel is

$$l'_\delta(\rho, r) = \pi^2 \int\limits_{\Omega_{\rho'}} \int\limits_{\Omega_{r'}} d(\rho, \rho', r, r') l_\delta(\rho', r) |\rho'| d\rho' |r'| dr' \tag{A.16}$$

$$= \frac{4\pi}{A^2} \int\limits_{\Omega_{\rho'}} \int\limits_{\Omega_{r'}} \frac{\delta(\rho - \rho')}{\pi|\rho'|} \frac{\sqcap(\frac{r-r'}{w})}{\pi w |r|} \sqcap\left(\frac{\rho'}{A}\right) \frac{\delta(r' - s_0\rho')}{\pi|r'|} |\rho'| d\rho' |r'| dr' \tag{A.17}$$

$$= \frac{4}{\pi A^2} \sqcap\left(\frac{\rho}{A}\right) \frac{1}{w|r|} \int\limits_{\Omega_r} \delta(r' - s_0\rho) \sqcap\left(\frac{r - r'}{w}\right) dr' \tag{A.18}$$

$$= \frac{4}{\pi A^2} \sqcap\left(\frac{\rho}{A}\right) \frac{\sqcap(\frac{r - s_0\rho}{w})}{\pi w |r|}. \tag{A.19}$$

The PSF for the light field filtered by this diffuser is

$$h'(r) = \pi \int\limits_{\Omega_\rho} l'_\delta(\rho, r) |\rho| d\rho \tag{A.20}$$

$$= \pi \int\limits_{\Omega_\rho} \frac{4}{\pi A^2} \frac{\sqcap(\frac{r - s_0\rho}{w})}{\pi w |r|} \sqcap\left(\frac{\rho}{A}\right) |\rho| d\rho \tag{A.21}$$

$$= \frac{4}{\pi A^2 w |r|} \int\limits_{\Omega_\rho} \sqcap(\frac{r - s_0\rho}{w}) \sqcap\left(\frac{\rho}{A}\right) |\rho| d\rho \tag{A.22}$$

$$= \frac{4}{\pi s_0^2 A^2 w |r|} \int\limits_{\Omega_\rho} \sqcap\left(\frac{r - \rho}{w}\right) \sqcap\left(\frac{\rho}{s_0 A}\right) |\rho| d\rho \tag{A.23}$$

$$= \frac{4}{\pi s_0^2 A^2} \frac{1}{w|r|} \left[\sqcap\left(\frac{r}{w}\right) \otimes \left(\sqcap\left(\frac{r}{s_0 A}\right) \cdot |r|\right)\right], \tag{A.24}$$

which is the same result as the PSF given by Equation 2.16.

A.4 Focal Sweep Comparison

In this section, we derive the expressions used to compare the focal sweep and a special form of diffusion coding discussed in Section 2.8. First we derive the expression for the focal sweep PSF given by Equation 2.22. For a point source located on the focal plane, $s = 0$, and the focus sweep PSF can be written as

$$h_{fs}(r) = \frac{1}{S} \int_{-S/2}^{S/2} \frac{4}{\pi s^2 A^2} \sqcap\left(\frac{r}{sA}\right) ds. \tag{A.25}$$

Using the change of variables $s' = sA$, the PSF becomes

$$h_{fs}(r) = \frac{4}{\pi SA} \int_{-SA/2}^{SA/2} \frac{1}{\pi s'^2} \sqcap\left(\frac{r}{s'}\right) ds'. \tag{A.26}$$

$$= \frac{8}{\pi SA} \int_{0}^{SA/2} \frac{1}{\pi s'^2} \sqcap\left(\frac{r}{s'}\right) ds'. \tag{A.27}$$

From the definition of the box function, $\sqcap\left(\frac{r}{s'}\right) = 1$ when $s' \geq 2|r| \geq SA$, which allows us to rewrite the PSF as

$$h_{fs}(r) = \frac{8}{\pi SA} \sqcap\left(\frac{r}{SA}\right) \int_{2|r|}^{SA/2} \frac{1}{s'^2} ds' \tag{A.28}$$

$$= \frac{4}{\pi SA} \left(\frac{1}{|r|} - \frac{4}{SA}\right) \sqcap\left(\frac{r}{SA}\right), \tag{A.29}$$

which is the same expression given by Equation 2.22. We now show that the scatter function given by Equation 2.28 results in the same PSF as focal sweep. We derive the diffusion coding PSF from the light field given by Equation 2.29. For a point source located on the focal plane, the PSF is

$$h'(r) = \pi \int\limits_{\Omega_\rho} l'_\delta(\rho, r)|\rho| d\rho \tag{A.30}$$

$$= \frac{4}{A^2} \int\limits_{\Omega_\rho} \sqcap\left(\frac{\rho}{A}\right) \frac{\sqcap(\frac{r}{S|\rho|})}{\pi S|\rho||r|} |\rho| d\rho. \tag{A.31}$$

Using the change of variable $\rho' = S\rho$, the PSF becomes

$$h'(r) = \frac{4}{\pi S^2 A^2 |r|} \int_{-SA/2}^{SA/2} \sqcap\left(\frac{r}{|\rho'|}\right) d\rho'. \tag{A.32}$$

$$= \frac{8}{\pi S^2 A^2 |r|} \int_{0}^{SA/2} \sqcap\left(\frac{r}{|\rho'|}\right) d\rho'. \tag{A.33}$$

By definition, $\sqcap\left(\frac{\rho}{A}\right) \sqcap(\frac{r}{S|\rho'|}) = 1$ when $2|r| \geq |\rho'| \geq SA$. This allows us to rewrite the PSF as

$$h'(r) = \frac{8}{\pi S^2 A^2 |r|} \sqcap \left(\frac{r}{SA} \right) \int_{2|r|}^{SA/2} d\rho'. \tag{A.34}$$

$$= \frac{4}{\pi SA} \left(\frac{1}{|r|} - \frac{4}{SA} \right) \sqcap \left(\frac{r}{SA} \right), \tag{A.35}$$

which proves that, for point sources located at the focal plane, the diffusion coding and focal sweep PSFs are the same. Although the analysis becomes a bit tricky, the same approach can be taken to prove that the PSFs are the same at all depths.

Appendix B

Gigapixel Computational Imaging Derivations

In this appendix, we give derivations for several equations that were provided in Chapter 4, which analyzed the performance of computational imaging systems that exhibit spherical aberrations.

B.1 Appendix A: PSF Derivation

From Equations 4.8 and 4.11, the PSF produced by a lens with a monomial OPD is given by

$$h(r) = \int_{-\infty}^{\infty} \sqcap(\rho) \frac{\delta(r - \alpha\rho^n)}{\pi|r|} |\rho| d\rho \tag{B.1}$$

$$\tag{B.2}$$

We introduce a change of variables $z = \alpha\rho^n$, giving the relations

$$\rho = \left(\frac{|z|}{\alpha}\right)^{1/n} \tag{B.3}$$

$$d\rho = \frac{1}{n\alpha}\left(\frac{|z|}{\alpha}\right)^{1/n-1} dz. \tag{B.4}$$

After substitution, the PSF becomes

$$h(r) = \frac{1}{\pi|r|}\int_{-\infty}^{\infty} \sqcap\left(\frac{z}{\alpha}\right)\delta(r-z)\frac{1}{n\alpha}\left(\frac{|z|}{\alpha}\right)^{1/n}\left(\frac{|z|}{\alpha}\right)^{1/n-1} dz \tag{B.5}$$

$$= \frac{1}{\pi|r|n\alpha}\int_{-\infty}^{\infty} \sqcap\left(\frac{z}{\alpha}\right)\delta(r-z)\left(\frac{|z|}{\alpha}\right)^{2/n-1} dz \tag{B.6}$$

$$= \frac{1}{\pi|r|n\alpha}\sqcap\left(\frac{r}{\alpha}\right)\left(\frac{|r|}{\alpha}\right)^{2/n-1} \tag{B.7}$$

$$= \frac{1}{\pi n\alpha^{2/n}}\sqcap\left(\frac{r}{\alpha}\right)|r|^{2/n-2}. \tag{B.8}$$

B.2 Appendix B: PSF Normalization

The energy for the PSF given in Equation 4.11 is

$$e = \pi\int_{-\infty}^{\infty} P_r(r)|r|dr \tag{B.9}$$

$$= \pi\int_{-\infty}^{\infty}\frac{1}{\pi n\alpha^{2/n}}\sqcap\left(\frac{r}{\alpha}\right)|r|^{2/n-2}|r|dr \tag{B.10}$$

$$= \frac{1}{n\alpha^{2/n}}\int_{-\alpha}^{\alpha}|r|^{2/n-1}dr \tag{B.11}$$

$$= \frac{1}{n\alpha^{2/n}}\left[\frac{n}{2}\frac{r}{|r|}|r|^{2/n}\right]_{\alpha}^{-\alpha} \tag{B.12}$$

$$= \frac{1}{n\alpha^{2/n}}\left(n\alpha^{2/n}\right) = 1. \tag{B.13}$$

which verifies that the PSF is properly normalized,

Bibliography

[Adelson and Bergen, 1991] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[Adelson and Wang, 1992] E.H. Adelson and J.Y.A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992.

[Ashok and Neifeld, 2003] A. Ashok and M. Neifeld. Information-based analysis of simple incoherent imaging systems. *Opt. Express*, 11(18):2153–2162, 2003.

[Ashok and Neifeld, 2007] A. Ashok and M. A. Neifeld. Pseudorandom phase masks for superresolution imaging from subpixel shifting. *Appl. Opt.*, 46(12):2256–2268, 2007.

[Ashok *et al.*, 2008] A. Ashok, P. K. Baheti, and M. A. Neifeld. Compressive imaging system design using task-specific information. *Appl. Opt.*, 47(25):4457–4471, Sep 2008.

[Baek, 2010] J. Baek. Transfer Efficiency and Depth Invariance in Computational Cameras. In *to appear in Proc. ICCP*, 2010.

[Baer *et al.*, 1999] R.L. Baer, W.D. Holland, J. Holm, and P. Vora. A comparison of primary and complementary color filters for ccd-based digital photography. In *Proc. SPIE Electronic Imaging Conference*, page 16. Citeseer, 1999.

[Ben-Ezra *et al.*, 2004] M. Ben-Ezra, A. Zomet, and S.K. Nayar. Jitter camera: high resolution video from a low resolution detector. In *Proceedings of the 2004 IEEE computer* *society conference on Computer vision and pattern recognition*, pages 135–142. IEEE Computer Society, 2004.

[Ben-Ezra *et al.*, 2005] M. Ben-Ezra, A. Zomet, and S.K. Nayar. Video super-resolution using controlled subpixel detector shifts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 977–987, 2005.

[Ben-Ezra, 2010] M. Ben-Ezra. High Resolution Large Format Tile-Scan - Camera Design, Calibration, and Extended Depth of Field. In *ICCP*, Mar 2010.

[Berns *et al.*, 2005] R.S. Berns, L.A. Taplin, M. Nezamabadi, M. Mohammadi, and Y. Zhao. Spectral imaging using a commercial color-filter array digital camera. In *Proc. of The 14th Triennal ICOM-CC meeting*, 2005.

[Bertero and Boccacci, 1998] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. Taylor & Francis, 1998.

[Brady and Hagen, 2009] D. J. Brady and N. Hagen. Multiscale lens design. *Opt. Express*, 17(13):10659–10674, 2009.

[Brady, 2009] D.J. Brady. *Optical imaging and spectroscopy*. Wiley, 2009.

[Caroli *et al.*, 1987] E. Caroli, JB Stephen, G. Cocco, L. Natalucci, and A. Spizzichino. Coded aperture imaging in x-and gamma-ray astronomy. *Space Science Reviews*, 45(3):349–403, 1987.

[Caulfield, 1971] H. J. Caulfield. Kinoform Diffusers. In *SPIE Conference Series*, volume 25, page 111, 1971.

[Chakrabarti and Zickler, 2011] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *CVPR '11*, pages 193–200, 2011.

[Chakrabarti *et al.*, 2010] A. Chakrabarti, K. Hirakawa, and T. Zickler. Computational color constancy with spatial correlations. *Harvard Technical Report TR-09-10*, 2010.

[Chi and George, 2001] W. Chi and N. George. Electronic imaging using a logarithmic asphere. *Opt. Lett.*, 2001.

[Cho *et al.*, 2010] T.S. Cho, A. Levin, F. Durand, and W.T. Freeman. Motion blur removal with orthogonal parabolic exposures. In *Computational Photography (ICCP), 2010 IEEE International Conference on*, pages 1–8. IEEE, 2010.

[Cossairt and Nayar, 2010] O. Cossairt and S. K. Nayar. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *ICCP*, Mar 2010.

[Cossairt *et al.*, 2010] O. Cossairt, C. Zhou, and S. K. Nayar. Diffusion Coding Photography for Extended Depth of Field. In *SIGGRAPH '10)*, Aug 2010.

[Dabov *et al.*, 2006] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising with block-matching and 3 D filtering. In *Proc. SPIE*, volume 6064, pages 354–365. Citeseer, 2006.

[DARPA, 2010] Darpa at 50. `"www.darpa.mil/Docs/1-25013846_Eprint_200811141152151.pdf"`, 2010.

[Debevec and Malik, 1997] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97*, pages 369–378, 1997.

[Dinyari *et al.*, 2008] R. Dinyari, S.B. Rim, K. Huang, P.B. Catrysse, and P. Peumans. Curving monolithic silicon for nonplanar focal plane array applications. *Applied Physics Letters*, 92:091114, 2008.

[Dowski and Cathey, 1995] E.R. Dowski and J.W.T. Cathey. Extended depth of field through wave-front coding. *Appl. Opt*, 34:1859–1866, 1995.

[Dowski *et al.*, 2000] E.R. Dowski, R.H. Cormack, and S.D. Sarama. Wavefront coding: jointly optimized optical and digital imaging systems. In *Proc. SPIE*, volume 4041, pages 114–120. Citeseer, 2000.

[Duarte *et al.*, 2008] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk. Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):83–91, 2008.

[Durand *et al.*, 2005] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. X. Sillion. A frequency analysis of light transport. In *SIGGRAPH '05*, 2005.

[E. R. Dowski and Cathey, 1995] Jr. E. R. Dowski and W. T. Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 1995.

[Fife *et al.*, 2008] K. Fife, A. El Gamal, and HSP Wong. A 3MPixel Multi-Aperture Image Sensor with 0.7 um Pixels in 0.11 um CMOS. In *IEEE ISSCC Conference*, 2008.

[Freeman *et al.*, 2006] W. Freeman, R. Fergus, A. Torralba, and W.T. Freeman. Random Lens Imaging. *MIT CSAIL*, 2006.

[García-Guerrero *et al.*, 2007] E. E. García-Guerrero, E. R. Méndez, H.r M. Escamilla, T. A. Leskova, and A. A. Maradudin. Design and fabrication of random phase diffusers for extending the depth of focus. *Opt. Express*, 15(3):910–923, 2007.

[Geary, 2002] J. M. Geary. *Introduction to Lens Design: With Practical Zemax Examples (Hardcover)*. Willmann-Bell, 2002.

[Gigapixl, 2007] The Gigapixl Project website. `"http://www.gigapixl.org/"`, 2007.

[Goodman, 1985] J. W. Goodman. *Statistical optics*. Wiley series in pure and applied optics, 1985.

[Goodman, 2005] J.W. Goodman. *Introduction to Fourier optics*. Roberts & Company Publishers, 2005.

[Gottesman and Fenimore, 1989] S.R. Gottesman and EE Fenimore. New family of binary arrays for coded aperture imaging. *Applied Optics*, 28(20):4344–4352, 1989.

[Gu *et al.*, 2008] J. Gu, S. Nayar, E. Grinspun, P. Belhumeur, and R. Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. *Computer Vision–ECCV 2008*, pages 845–858, 2008.

[Guichard *et al.*, 2009] F. Guichard, H. Nguyen, R. Tessières, M. Pyanet, I. Tarchouna, and F. Cao. Extended depth-of-field using sharpness transport across color channels. In *Digital Photography V*, volume 7250. Proc. SPIE, 2009.

[Gupta *et al.*, 2009] M. Gupta, Yuandong Tian, S.G. Narasimhan, and Li Zhang. (de) focusing on global light transport for active scene recovery. In *CVPR '09*, june 2009.

[Hanley *et al.*, 1999] Q.S. Hanley, P.J. Verveer, and T.M. Jovin. Spectral imaging in a programmable array microscope by hadamard transform fluorescence spectroscopy. *Applied spectroscopy*, 53(1):1–10, 1999.

[Harwit and Sloane, 1979] M. Harwit and N.J. Sloane. Hadamard transform optics. *New York: Academic Press, 1979*, 1, 1979.

[Hasinoff *et al.*, 2009] S.W. Hasinoff, K.N. Kutulakos, F. Durand, and W.T. Freeman. Time-constrained photography. In *Proc. ICCV*, pages 1–8, 2009.

[Hasinoff *et al.*, 2010] S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR '10*, pages 553–560, 2010.

[Häusler, 1972] G. Häusler. A method to increase the depth of focus by two step image processing. *Optics Comm.*, 1972.

[Healey and Kondepudy, 1994] G. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 267–276, 1994.

[Hitomi *et al.*, 2011] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K. Nayar. Video from a Single Coded Exposure Photograph using a Learned Over-Complete Dictionary. In *IEEE International Conference on Computer Vision (ICCV)*, Nov 2011.

[Horn, 1986] B. Horn. *Robot vision.* The MIT Press, 1986.

[ICE, 2010] Microsoft Image Composite Editor (ICE) website. `"http://research.microsoft.com/en-us/um/redmond/groups/ivm/ICE/"`, 2010.

[Ihrke *et al.*, 2010] I. Ihrke, G. Wetzstein, and W. Heidrich. A Theory of Plenoptic Multiplexing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2010. oral.

[Kajiya, 1986] James T. Kajiya. The rendering equation. In *SIGGRAPH '86*, pages 143–150, 1986.

[Kingslake, 1989] R. Kingslake. *A history of the photographic lens.* Academic Press, 1989.

[Ko *et al.*, 2008] H.C. Ko, M.P. Stoykovich, J. Song, V. Malyarchuk, W.M. Choi, C.J. Yu, J.B. Geddes Iii, J. Xiao, S. Wang, Y. Huang, et al. A hemispherical electronic eye camera based on compressible silicon optoelectronics. *Nature*, 454(7205):748–753, 2008.

[Krishnan and Nayar, 2009] G. Krishnan and S.K. Nayar. Towards A True Spherical Camera. In *SPIE Human Vision and Electronic Imaging*, Jan 2009.

[Lanman *et al.*, 2008] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin. Shield fields: modeling and capturing 3d occluders. *ACM Transactions on Graphics (TOG)*, 27(5):131, 2008.

[Lee and Szema, 2005] L.P. Lee and R. Szema. Inspirations from biological optics for advanced photonic systems. *Science*, 310(5751):1148, 2005.

[Levin *et al.*, 2007] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH '07*, 2007.

[Levin *et al.*, 2008] A. Levin, P. Sand, T.S. Cho, F. Durand, and W.T. Freeman. Motion-invariant photography. In *ACM SIGGRAPH 2008 papers*, pages 1–9. ACM, 2008.

[Levin *et al.*, 2009] A. Levin, S. Hasinoff, P. Green, F. Durand, and W. T. Freeman. 4d frequency analysis of computational cameras for depth of field extension. In *SIGGRAPH '09*, 2009.

[Levoy and Hanrahan, 1996] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH '96*, 1996.

[Liang *et al.*, 2008] C.K. Liang, T.H. Lin, B.Y. Wong, C. Liu, and H.H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM SIGGRAPH 2008 papers*, pages 1–10. ACM, 2008.

[Lohmann, 1989] A. W. Lohmann. Scaling laws for lens systems. *Appl. Opt.*, 28(23):4996–4998, 1989.

[Lumenera, 2010] Lumenera Corporation company website. `"http://www.lumenera.com/"`, 2010.

[Luminit, 2011] Luminit company website. `http://www.luminit.com`, 2011.

[Luneburg, 1964] R.K. Luneburg. *Mathematical theory of optics.* University of California Press, 1964.

[Marcia and Willett, 2008] R.F. Marcia and R.M. Willett. Compressive coded aperture superresolution image reconstruction. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 833–836. IEEE, 2008.

[Marks and Brady, 2010] D.L. Marks and D.J. Brady. Gigagon: A Monocentric Lens Design Imaging 40 Gigapixels. In *Imaging Systems.* OSA, 2010.

[Matusik *et al.*, 2003] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. In *SIGGRAPH '03*, pages 759–769, 2003.

[Nagahara *et al.*, 2008] H. Nagahara, S. Kuthirummal, C. Zhou, and S.K. Nayar. Flexible Depth of Field Photography. In *ECCV '08*, 2008.

[Narasimhan and Nayar, 2005] S.G. Narasimhan and S.K. Nayar. Enhancing Resolution along Multiple Imaging Dimensions using Assorted Pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, Apr 2005.

[Nayar and Mitsunaga, 2000] S.K. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *CVPR '00*, volume 1, pages 472–479, Jun 2000.

[Nayar *et al.*, 2004] S.K. Nayar, V. Branzoi, and T. Boult. Programmable Imaging using a Digital Micromirror Array. In *CVPR '04*, volume I, pages 436–443, Jun 2004.

[Nayar *et al.*, 2006a] S. K. Nayar, V. Branzoi, and T. E. Boult. Programmable Imaging: Towards a Flexible Camera. *International Journal on Computer Vision*, Oct 2006.

[Nayar *et al.*, 2006b] S.K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination. *Siggraph '06*, Jul 2006.

[Ng *et al.*, 2005] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2, 2005.

[Ng, 2005] R. Ng. Fourier slice photography. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, 2005.

[Nomura *et al.*, 2007] Y. Nomura, L. Zhang, and S. Nayar. Scene collages and flexible camera arrays. In *Proc. EGSR*, 2007.

[of Joensuu Color Group, 2011] University of Joensuu Color Group. Spectral database. http://spectral.joensuu.fi/, 2011.

[Ojeda-Castaneda and Berriel-Valdos, 1990] J. Ojeda-Castaneda and L. R. Berriel-Valdos. Zone plate for arbitrarily high focal depth. *Appl. Opt.*, 1990.

[Ojeda-Castaneda *et al.*, 2005] J. Ojeda-Castaneda, J. E. A. Landgrave, and H. M. Escamilla. Annular phase-only mask for high focal depth. *Opt. Lett.*, 2005.

[Pal and Neifeld, 2003] H. Pal and M. Neifeld. Multispectral principal component imaging. *Opt. Express*, 11(18):2118–2125, Sep 2003.

[Parkkinen *et al.*, 1989] J. Parkkinen, J. Hallikainen, and T. Jaaskelainen. Characteristic spectra of munsell colors. *J. Opt. Soc. Am.*, 1989.

[Peers *et al.*, 2009] P. Peers, D.K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi, and P. Debevec. Compressive light transport sensing. *ACM Transactions on Graphics (TOG)*, 28(1):1–18, 2009.

[Raskar *et al.*, 2006] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Transactions on Graphics (TOG)*, 25(3):795–804, 2006.

[Ratner and Schechner, 2007] N. Ratner and Y.Y. Schechner. Illumination multiplexing within fundamental limits. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[Ratner *et al.*, 2007] N. Ratner, Y.Y. Schechner, and F. Goldberg. Optimal multiplexed sensing: bounds, conditions and a graph theory link. *Opt. Express*, 15:17072–17092, 2007.

[Reddy *et al.*, 2011] D. Reddy, A. Veeraraghavan, and R. Chellappa. P2c2: Programmable pixel compressive camera for high speed imaging. *CVPR '11*, pages 329–336, 2011.

[Rim *et al.*, 2005] S.B. Rim, P.B. Catrysse, R. Dinyari, K. Huang, and P. Peumans. The optical advantages of curved focal plane arrays. In *Proc. SPIE*, volume 5678, pages 48–58, 2005.

[Robinson and Bhakta, 2009] M.D. Robinson and V. Bhakta. Experimental validation of extended depth-of-field imaging via spherical coding. In *Computational Optical Sensing and Imaging*, page CThB4. OSA, 2009.

[Robinson and Stork, 2009] M.D. Robinson and D.G. Stork. Extending depth-of-field: Spherical coding versus asymmetric wavefront coding. In *Computational Optical Sensing and Imaging*, page CThB3. OSA, 2009.

[Robinson *et al.*, 2009] M.D. Robinson, G. Feng, and D.G Stork. Spherical coded imagers. In *Proc. SPIE*, volume 7429, page 20, 2009.

[Rosen and Yariv, 1994] J. Rosen and A. Yariv. Synthesis of an arbitrary axial field by computer-generated holograms. *Opt. Lett.*, 1994.

[RPC, 2011] Rpc photonics company website. `http://www.rpcphotonics.com`, 2011.

[Sales, 2003] T. R. M. Sales. Structured microlens arrays for beam shaping. *Optical Engineering*, 42(11):3084–3085, 2003.

[Schechner *et al.*, 2003] Y.Y. Schechner, S.K. Nayar, and P.N. Belhumeur. A theory of multiplexed illumination. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 808–815. IEEE, 2003.

[Schechner *et al.*, 2007] Y.Y. Schechner, S.K. Nayar, and P.N. Belhumeur. Multiplexing for optimal lighting. *IEEE Transactions on pattern analysis and machine intelligence*, pages 1339–1354, 2007.

[Skinner, 1988] GK Skinner. X-ray imaging with coded masks. *Scientific American;(USA)*, 259(2), 1988.

[Slater, 1966] L. J. Slater. *Generalized Hypergeometric Functions*. Cambridge University Press, 1966.

[Smith, 1966] Warren J. Smith. *Modern optical engineering; the design of optical systems [by] Warren J. Smith*. McGraw-Hill, New York,, 1966.

[Srivastava *et al.*, 2003] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S c. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18:17–33, 2003.

[Sun *et al.*, 2007] B. Sun, K. Sunkavalli, R. Ramamoorthi, P. Belhumeur, and S. Nayar. Time-Varying BRDFs. *IEEE Transactions on Visualization and Computer Graphics*, Mar 2007.

[Tanida *et al.*, 2000] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (tombo): an optoelectronic image capturing system. In *Proceedings of SPIE*, volume 4089, page 1030, 2000.

[Tanida *et al.*, 2001] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (tombo): concept and experimental verification. *Applied Optics*, 40(11):1806–1813, 2001.

[Veeraraghavan *et al.*, 2007] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *SIGGRAPH '07*, 2007.

[Villela *et al.*, 1995] T. Villela, J. Braga, F. D'Amico, and UB Jayanthi. A mura-based coded mask telescope. *Advances in Space Research*, 15(5):95–98, 1995.

[Wagadarikar *et al.*, 2008] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008.

[Wagadarikar *et al.*, 2009] A.A. Wagadarikar, N.P. Pitsianis, X. Sun, and D.J. Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Opt. Express*, 17(8):6368–6388, 2009.

[Wang and Heidrich, 2004] S. Wang and W. Heidrich. The design of an inexpensive very high resolution scan camera system. In *Computer Graphics Forum*, volume 23, pages 441–450. Citeseer, 2004.

[Weiss and Freeman, 2007] Y. Weiss and W. T. Freeman. What makes a good model of natural images. In *CVPR '07*, pages 1–8, 2007.

[Welford, 1960] W. T. Welford. Use of annular apertures to increase focal depth. *J. Opt. Soc. Am.*, 1960.

[Wilburn *et al.*, 2005] B. Wilburn, N. Joshi, V. Vaish, E.V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Transactions on Graphics (TOG)*, 24(3):776, 2005.

[Zemax, 2010] Zemax Optical Design Software. `"http://www.zemax.com/"`, 2010.

[Zhou and Nayar, 2009] C. Zhou and S. Nayar. What are Good Apertures for Defocus Deblurring? In *ICCP '09*, 2009.

[Zhou *et al.*, 2011] Changyin Zhou, Stephen Lin, and Shree Nayar. Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring. *International Journal on Computer Vision*, 93(1):53, May 2011.