Multiplexed Photography: Single-Exposure Capture of Multiple Camera Settings

by

Paul Elijah Green

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author ....
Department of Electrical Engineering and Computer Science
~~August 17~~, 2009

Certified by ....
Frédo Durand
Associate Professor
Thesis Supervisor

Accepted by ....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Multiplexed Photography: Single-Exposure Capture of Multiple Camera Settings

by

Paul Elijah Green

Submitted to the Department of Electrical Engineering and Computer Science
on August 17, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science

Abstract

The space of camera settings is large and individual settings can vary dramatically from scene to scene. This thesis explores methods for capturing and manipulating multiple camera settings in a single exposure. Multiplexing multiple camera settings in a single exposure can allow post-exposure control and improve the quality of photographs taken in challenging lighting environments (e.g. low light or high motion).

We first describe the design and implementation of a prototype optical system and associated algorithms to capture four images of a scene in a single exposure, each taken with a different aperture setting. Our system can be used with commercially available DSLR cameras and photographic lenses without modification to either. We demonstrate several applications of our multi-aperture camera, such as post-exposure depth of field control, synthetic refocusing, and depth-guided deconvolution.

Next we describe multiplexed flash illumination to recover both flash and ambient light information as well as extract depth information in a single exposure. Traditional photographic flashes illuminate the scene with a spatially-constant light beam. By adding a mask and optics to a flash, we can project a spatially varying illumination onto the scene which allows us to spatially multiplex the flash and ambient illuminations onto the imager. We apply flash multiplexing to enable single exposure flash/no-flash image fusion, in particular, performing flash/no-flash relighting on dynamic scenes with moving objects.

Finally, we propose spatio-temporal multiplexing, a novel image sensor feature that enables simultaneous capture of flash and ambient illumination. We describe two possible applications of spatio-temporal multiplexing: single-image flash/no-flash relighting and white balancing scenes containing two distinct illuminants (e.g. flash and fluorescent lighting).

Thesis Supervisor: Frédo Durand
Title: Associate Professor

Acknowledgments

I would like to thank the following people for their help, advice, support and friendship during my graduate studies: my advisor Frédo Durand, thesis committee members William Freeman and Ramesh Raskar, machine shop gurus John Barnwell and Jonathan Westhues, mentors Jan Kautz, Wojciech Matusik and Sylvain Paris, my friends and colleagues in the MIT Computer Graphics Group, Daniel Vlasic, Marco da Silva, Eugene Hsu, Tom Mertens, Sara Su, Jonathan Reagan-Kelley, Jaakko Lehtinen, Robert Sumner, Eric Chan, Jiawen Chen, Robert Wang, Soonmin Bae, Yeuhi Abe, Emily Whiting, Vladimir Bychkovsky, Ilya Baran, Anat Levin, Sam Hasinoff, Matthias Zwicker, SeBaek Oh, Addy Ngan, Bryt Bradley, Jovan Popović, Tom Buehler, Adel Hanna, Thouis Jones, Jernej Barbič, the friendly members of the ARTIS group at INRIA Grenoble, Xavier Décoret, Elmar Eisemann, Stephane Grabli, Emmanuel Turquin, and the greatest distraction of all, MIT noonball. I would also like to acknowledge the financial support I received from the Ford Foundation graduate fellowship and the Norman Leventhal fellowship.

Finally, I would like to add a special dedication to my family — Vicky, Galen, Gelando, Klaus, Al, Mary Jane, Paul, Angel, Elisa, Elizabeth, and Jane. I am eternally grateful to each of you for your unwavering support and patience.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "The very definition of a new medium is that it alters the relationship of people to the world around them. It is only after that alteration has occurred that people can sense the change. Obvious examples of this phenomenon are the effects of the telephone or television on our culture as a whole. And so, in a much smaller but yet significant way, we find that our kind of photography can change the interaction of people with the world around them" - Edwin H. Land, 1972 [35]

Edwin Land, photographic visionary, esteemed vision scientist and founder of the Polaroid Corporation, saw the potential of instant photography to radically transform our society and "the interaction of people with the world around them". He envisioned a photographic system that was fast, seamless, and simple to use. His Land instant camera revolutionized the landscape of photography, transforming a process that previously took hours to mere seconds, allowing almost instantaneous photographs. His inventions helped popularize personal photography, making it simple for people to recompose and retake a photo if they were unsatisfied with the original result. Moreover, instant photography found other less traditional applications, such as for taking ID and passport photos, and aiding the police in documenting crime scenes.

The advent of digital photography has now almost entirely usurped Land's original film based camera systems, providing even faster, cheaper and more flexible photography. In a sense, digital photography has taken Land's dream to the next level, providing truly instant and effectively free photos. Having such abundant and easily created photographs has opened up many new applications and uses for digital photography, and made a significant impact on society today. The number and variety of cameras has also seen an explosion of growth, from high-end large format and Digital SLR cameras to inexpensive and portable point-and-shoot cameras and camera-phones, with many people owning several types. This rapid increase in the number of cameras and amateur photographers has produced hundreds of millions of photographs available online at photo-sharing and social networking websites like flickr.com and facebook.com.

Spurred on by digital photography's recent propulsion into the mainstream, the new field of computational photography is beginning to fundamentally re-think the way we capture and process images. In traditional photography, standard optics directly form the final image. In contrast, computational approaches replace traditional lenses with generalized optics that form a *coded* image onto the sensor which cannot be visualized directly but, through *computation*, can yield higher quality or enable flexible post-processing and the extraction of extra information. Some designs use multiple sensors to extract more information about the visual environment. Other approaches exploit the rapid growth in sensor spatial resolution, which is usually superfluous for most users, and instead use it to capture more information about a scene. These new techniques open exciting possibilities, and in particular give us the ability to modify image formation parameters *after* the image has been taken.

This thesis describes multiplexed photography, a collection of computational photography methods for simultaneously capturing multiple camera settings in a single exposure. By capturing multiple camera settings at one time, we can extend the capabilities of digital photography, and help photographers manage the large space of camera settings by enabling post-exposure editing and control. There are many different camera settings (e.g. focus, aperture, shutter speed, and flash) and erroneously setting any one can ruin an otherwise good photograph. In this thesis we focus on methods that allow post-exposure editing and control of physical camera settings as well as higher-level controls such as depth of field. One common thread in all of the projects described in this thesis is that we trade image resolution to capture more information, such as multiple aperture or flash settings. This thesis is comprised of three projects: multi-aperture photography, multiplexed illumination, and spatio-temporal multiplexing. While each project takes a different approach to the goal of capturing multiple camera settings, exploring several areas and approaches of computational photography, they all rely on spatial multiplexing and the emerging abundance of image resolution now available on modern image sensors.

1.1 Trends in Computational Photography

Marc Levoy broadly defined computational photography as "computational methods that enhance or extend the capabilities of digital photography"[46]. To give some context for this thesis we can construct a taxonomy of recent computational photography research by examining the various ways researchers have modified or augmented traditional cameras.

digital image processing The first modification, which in essence created digital photography, was to replace film with a discrete digital sensor. The new digital sensor essentially emulated film, but instead produced digital files, instead of negatives and prints. The first category in the taxonomy is digital image processing, loosely defined as the processing of captured images to create new images. This is very broad, but includes methods like tone mapping and dynamic range compression[19, 23, 7, 55], multi-exposure high dynamic range images[16, 29], and digital image compression[27].

Computational optics and cameras The next change, and a fairly radical departure from traditional photography, was to replace the standard optics in the camera, that produced nice focused images, with new coding optics, that no longer directly produced images fit for human consumption. Instead computation is required to decode the captured data into normal images. Some examples of new computational optics and camera designs include wavefront coding[9, 12, 18, 24] which extends depth of field, coded aperture[43, 82] which can provide depth information, and lightfield photography[48, 66, 65, 44] which allows refocusing of the image after it has been taken. Our multi-aperture camera (discussed in chapter 3) falls under the area of computational optics and cameras.

Computational Sensors Next, people again replaced the digital sensor, which as mentioned previously, was essentially emulating film, with a new sensor of smart pixels that combine sensing and processing. These smart pixels have been used for applications like adaptively adjusting their individual exposures to avoid saturation[1] and capturing a high dynamic range image[62]. Spatio-temporal multiplexing, discussed in chapter 5 is an example of a new computational sensor.

Computational Illumination Finally, we can replace the passive lighting, with structured or controlled illumination. Some examples include fast separation of direct and global illumination[61], spatially adaptive flash [2], Shader Lamps [73, 30] to change the appearance of the scene, tabletop lighting for digital photography[57], and using flash shadow edges to estimate depth[72]. Our multiplexed flash illumination method (presented in chapter 4) falls under the category of computational illumination.

1.2 Organization of Thesis

This thesis explores methods for capturing and manipulating multiple camera settings in a single exposure. Multiplexing multiple camera settings in a single exposure can allow post-exposure control and improve the quality of photographs taken in challenging lighting environments (e.g. low light or high motion). To this end, we introduce three new computational photography methods for improved digital photography.

Chapter 3 describes the design and implementation of a prototype optical system and associated algorithms to capture four images of a scene in a single exposure, each taken with a different aperture setting. Our system can be used with commercially available DSLR cameras and photographic lenses without modification to either. We demonstrate several applications of our multi-aperture camera, such as post-exposure depth of field control, synthetic refocusing, and depth-guided deconvolution.

Chapter 4 introduces a coded illumination method we call multiplexed flash illumination to recover both flash and ambient light information as well as extract depth information in a single exposure. Traditional photographic flashes illuminate the scene with a spatially-constant light beam. By adding a mask and optics to a flash, we can project a spatially varying illumination onto the scene which allows us to spatially multiplex the flash and ambient illuminations onto the imager. We apply flash multiplexing to enable single exposure flash/no-flash image fusion, in particular, performing flash/no-flash relighting on dynamic scenes with moving objects.

Finally, in chapter 5 we propose spatio-temporal multiplexing, a novel image sensor integration strategy that enables simultaneous capture of flash and ambient illumination. We describe two possible applications of spatio-temporal multiplexing: single-image flash/no-flash relighting and white balancing scenes containing two distinct illuminants (e.g. flash and fluorescent lighting).

In chapter 2 we review basic optics, camera settings and other preliminaries useful for reading the thesis. We end with concluding remarks and future directions in Chapter 6.

Chapter 2

Background

In this chapter, we present background and related work useful for the rest of this thesis. We begin with a review of geometric optics and image formation. Next, we discuss camera settings and their effects on image formation. We conclude the chapter with a description of the bilateral filter[81] and related flash / no-flash image fusion methods[21, 70] which we reference frequently in chapters 4 and 5. In this chapter we present a broad overview of related work, while each subsequent chapter contains a more detailed and focused related work section.

2.1 Geometrical Optics

In this section we present a brief review of geometrical optics that forms the basis for our discussion of camera lens settings and controls as well as provides a foundation in optics useful for reading the following chapters of the thesis. We make two main approximations in our discussion of optics in this section. The first approximation, and the classic definition of geometrical optics, is that the wavelength of light is small enough such that the wave nature light can be ignored, and that light propagation can instead be approximated by rays that travel in straight lines. By neglecting the wave nature of light, we ignore phase and interference effects such as diffraction. Our second approximation is that we only consider paraxial rays — rays that make a small angle with the optical axis of the system. The paraxial approximation allows us to

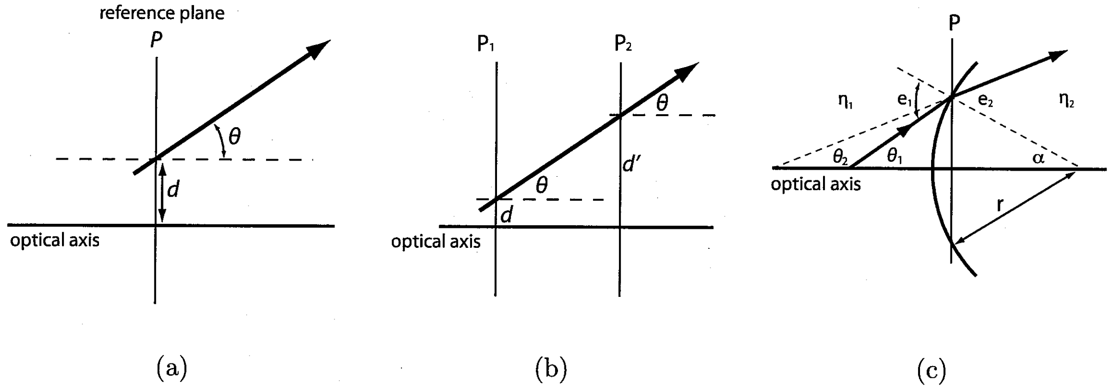


Figure 2-1: *Ray parameterization. (a) A ray is parameterized at a reference plane* P *by its distance* d *and angle* θ *from the optical axis. (b) Rays travel along a straight path in a homogeneous medium. The angle* θ *remains constant, while the distance from the optical axis can vary from reference plane* P_1 *to* P_2*. (c) Refraction of a ray as it passes through a spherical interface between two media with different indices of refraction.*

use simple approximations of $\sin\theta$, $\cos\theta$ and $\tan\theta$ for small values of θ (e.g. less than 10°):

$$\cos\theta = 1 - \tfrac{\theta^2}{2!} + \tfrac{\theta^4}{4!} - \cdots \approx 1 \tag{2.1}$$

$$\sin\theta = \theta - \tfrac{\theta^3}{3!} + \tfrac{\theta^4}{4!} - \cdots \approx \theta \tag{2.2}$$

$$\tan\theta = \theta + \tfrac{\theta^3}{3} + \tfrac{2\theta^5}{15} + \cdots \approx \theta \tag{2.3}$$

This assumption is commonly referred to as paraxial, Gaussian (after Carl Friedrich Gauss), or first-order optics, because we are using a first-order approximation of the Taylor expansion of the trigonometric functions. Paraxial optics provides an idealized version of the optical characteristics of a system and does not capture third-order (or higher) aberrations such as spherical or comatic aberration.

Ray parameterization For simplicity, we limit our discussion to 2D rays, and as such, a ray traveling through our optical system can be described by its distance d and angle θ from the optical axis, as measured at a sequence of reference planes (see figure 2-1(a)). In order to simplify later steps in our analysis, we do not describe rays using the angle θ directly, we instead use the so called *ray direction-cosine*, $\eta \sin \theta$ (which can be simplified to $\eta\theta$ using the paraxial approximation), where η is the index of refraction of the medium the ray is traveling in. Thus a ray r, in a medium of index η, can be describe by a two dimensional column vector $r = [d, \eta\theta]^T$. The advantage of using ray direction-cosines instead of angles directly is that, by Snell's law[34, 74], the direction-cosine $V = \eta \sin \theta \approx \eta\theta$ (assuming paraxial rays) remains constant for a ray as it crosses a planar boundary between two media.

Matrix Methods It is common to describe the behavior of paraxial optical systems using a matrix formulation that relates the state of rays as they travel from one reference plane to another reference plane. This allows complex optical systems to be described by combining multiple matrices. In particular, in order to describe compound optical systems consisting of multiple spherical glass lenses in air, we only need to derive matrix forms for two cases: the *translation* (or propagation) of rays between two reference planes in the same medium, and the *refraction* of rays at the boundary between media of two different refractive indices.

Translation Let us assume we have a ray r described by $r = [d, \eta\theta]^T$ at reference plane P_1, and the ray travels a distance t through a homogeneous medium with index of refraction η. We would like to describe the translated ray $r' = [d', \eta\theta']^T$ at a second reference plane P_2 (see figure 2-1(b) for a diagram).

$$\begin{aligned} d' &= d + t \tan\theta && (2.4) \\ &= d + t\theta && (2.5) \\ &= d + (\frac{t}{\eta})(\eta\theta) && (2.6) \end{aligned}$$

Step 2.5 follows from the paraxial ray assumption. The quantity $T = \frac{t}{\eta}$ is called the "reduced thickness," and adjusts the distance relative to a vacuum, accounting for the index of refraction of the medium. Light rays travel in straight lines through a homogeneous medium, thus implying $\theta' = \theta$. These two equations can be captured in matrix form as:

$$\begin{bmatrix} d' \\ \eta\theta' \end{bmatrix} = \begin{bmatrix} 1 & t/\eta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d \\ \eta\theta \end{bmatrix} \tag{2.7}$$

From equation 2.7 we see that the translation matrix Δ that describes the translation of ray r a distance t through a medium with index of refraction η is:

$$\Delta = \begin{bmatrix} 1 & t/\eta \\ 0 & 1 \end{bmatrix} \tag{2.8}$$

Refraction We would like to compute the refraction matrix $\Re$, given a spherical refraction boundary, with radius of curvature r, and refractive indices η_1 and η_2 on either side of the boundary (see figure 2-1(c) for a diagram). In the case of refraction, we only examine a single reference plane P, and describe the ray just before, and just after P (we can treat this as the limit of two reference planes as the distance between them goes to zero). In a sense refraction is the dual of translation in that the ray height stays the same, i.e. $d' = d$, and we must calculate a new ray direction-cosine $\eta_2\theta'$.

By Snell's law we have:

$$\eta_1 \sin e_1 = \eta_2 \sin e_2 \tag{2.9}$$

which can be simplified using the paraxial approximation as

$$\eta_1 e_1 = \eta_2 e_2. \tag{2.10}$$

Using the exterior angle theorem and the paraxial approximation (we assume α is small like θ_1 and θ_2) we can solve for e_1 and e_2 in terms of θ_1,θ_2, r and d.

$$e_1 = \theta_1 + \alpha \approx \theta_1 + \sin\alpha = \theta_1 + \frac{d}{r} \tag{2.11}$$

$$e_2 = \theta_2 + \alpha \approx \theta_2 + \sin\alpha = \theta_2 + \frac{d}{r} \tag{2.12}$$

Substituting for e_1 and e_2 into equation 2.10 and rearranging terms we get:

$$\eta_2\theta_2 = \eta_1\theta_1 - (\eta_2 - \eta_1)d/r \tag{2.13}$$

which can be written in matrix form as:

$$\begin{bmatrix} d' \\ \eta_2\theta' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -(\eta_2 - \eta_1)/r & 1 \end{bmatrix} \begin{bmatrix} d \\ \eta_1\theta \end{bmatrix} \tag{2.14}$$

It is common to make the substitution $P = (\eta_2 - \eta_1)/r$, where P is called the refractive power of the surface. Then, the refraction matrix $\Re$ is simply:

$$\Re = \begin{bmatrix} 1 & 0 \\ -P & 1 \end{bmatrix} \tag{2.15}$$

2.1.1 Modeling Lenses

The power of the matrix representation is that complex optical systems can be constructed by composing simple combinations of Δ and $\Re$ matrices. For example, a singlet lens L (a lens consisting of a single element) can be constructed by composing two refractive matrices with a translation matrix between them:

$$L = \Re_2\Delta\Re_1, \tag{2.16}$$

where $\Re_1$ and $\Re_2$ are chosen to match the index of refraction and radii of curvature of the lens, and Δ matches the thickness. Note that the convention is that radius values, r, are positive if the center of curvature is to the right of the surface, and negative if to the left.

Thin Lens Approximation The thin lens approximation models a lens as having zero center thickness, and thus it can be described using only two refraction matrices.

$$L = \Re_2\Re_1 \tag{2.17}$$

Assuming the glass has refractive index η, and is surrounded by air, the compound matrix is:

$$L = \begin{bmatrix} 1 & 0 \\ -P_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -P_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -(P_1 + P_2) & 1 \end{bmatrix}, \tag{2.18}$$

with

$$P_1 + P_2 = (\eta - 1)/r_1 + (1 - \eta)/r_2 = (\eta - 1)(1/r_1 + 1/r_2) = 1/f \tag{2.19}$$

where f is the focal length of the thin lens.

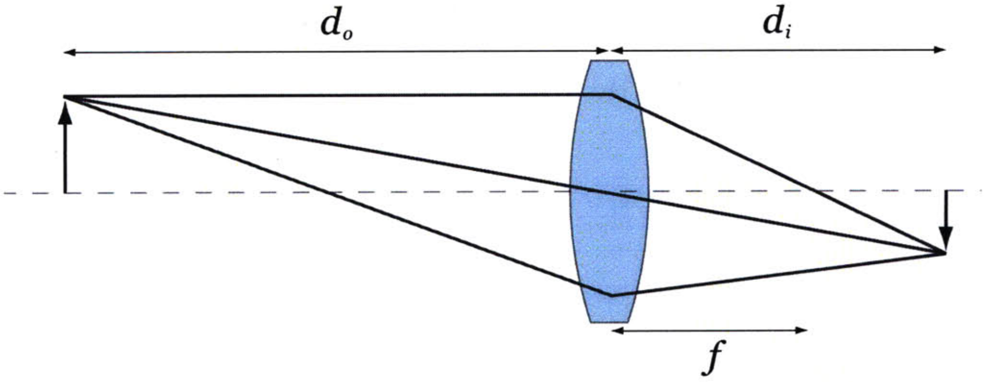


Figure 2-2: *An ideal thin lens with focal length f focuses light leaving an object a distance d_o in front of the lens into a point d_i behind the lens. Notice that the lens creates an inverted image of the object.*

Thin lens equation We can use our matrix representation to derive the well known thin lens equation that describes the relationship of object and image conjugate points and the focal length of the lens:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \tag{2.20}$$

where f is the lens focal length, d_o is the distance to the object reference plane, and d_i is the distance to the image reference plane (see figure 2-2). The thin lens equation describes the focusing behavior of an idealized (paraxial) aberration-free lens, and thus is one of the most useful equations for modeling camera lenses.

In order to derive equation 2.20 we will construct a matrix C that models the scene and then investigate some of its properties. The scene can be modeled by composing three matrices, a translation matrix that models the ray propagation from the object to the lens, a refraction matrix that models the lens using the thin lens approximation (equation 2.18), and finally another translation matrix that propagates the ray from the lens to the image plane.

$$C = \begin{bmatrix} 1 & d_i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} 1 & d_o \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - d_i/f & (1 - d_i/f)d_o + d_i \\ -1/f & 1 - d_o/f \end{bmatrix} \tag{2.21}$$

Now suppose that d_o and d_i are chosen such that C is of the form

$$C = \begin{bmatrix} \alpha & 0 \\ -1/f & \beta \end{bmatrix}, \tag{2.22}$$

where the upper right entry vanishes. Then, if we transform a ray $r = [d, \theta]^T$ by C we see that $d' = \alpha d$ for any value of θ. In other words, C "focuses" all object reference plane rays leaving from d, regardless of their initial angle θ, to point d' on the image reference plane. It is trivial to see that C takes the form in equation 2.22 when the thin lens equation (equation 2.20) is satisfied.

2.2 Camera Settings

Is this section we move away from our discussion of optics and now focus on the standard camera controls available to photographers when using commercial camera systems. Our goal is to highlight the basic controls available on most cameras and lenses, their effect on the captured image, and any work related to multiplexing and post-exposure editing.

2.2.1 Focus

Focus is arguably the most important camera control in terms of its effect on the final image. Focus controls are used to select a scene (or object) distance that will be sharp in the final image. Although commercial lenses are complex, multi-element optical systems (often containing between four and as many as twenty or more lenses), they can be modeled as a single ideal thin lens with focal length f, that follows the thin lens equation (equation 2.20). Under this model, focus controls are directly adjusting the lens-sensor distance d_i in order to focus at the desired object distance d_o.

Handheld lightfield cameras and camera arrays [3, 66, 65, 47] have been used to capture enough information about a scene to enable post-capture refocusing. Typically, refocusing is accomplished by recording the lightfield, or the 4D set of light rays that enters the camera, and digitally performing the lens integration normally performed with optics. An alternative approach was proposed by McGuire et. al. [54] where they constructed a network of N copies of the incoming light, using a tree of beam-splitters, and each copy can be imaged with different camera settings. In particular, each of the N copies could be captured with a different focus setting, allowing some post-exposure control of focus. Wavefront coding methods[12, 9, 18] attempt to avoid the focusing problem altogether by capturing an image that has a depth-invariant blur, which can be removed using deconvolution and deblurring.

2.2.2 Exposure

Exposure is a measure of the total amount of light that reaches the image sensor. There are two main controls that effect the overall exposure of an image: shutter speed and aperture size. One useful concept when discussing exposure, shutter speed, and aperture size is exposure value (EV). A particular EV describes all combinations of exposure time and aperture size that produce the same total image exposure. EV can be calculated as:

$$EV = \log_2 \frac{N^2}{t} \tag{2.23}$$

where t is the exposure time and N is the aperture size (or $f/\#$), both of which will be described below.

2.2.3 Exposure Time

The exposure time, also commonly called the shutter speed, determines the length of time the shutter remains open during image capture. As noted above, a particular EV can be attained by many combinations of exposure time and aperture size. However, although each combination will have the same exposure, the images can appear dramatically different for different combinations. Shutter time is primarily relevant for controlling the motion blur of scenes with moving objects. For example, it is common to use a fast shutter speed (e.g. $1/250^{th}$ of a second or shorter) when photographing sporting events and other action scenes in order to "freeze" the motion of an instant in time. Alternatively, a long shutter speed can be used to emphasize the motion and dynamism of the scene. This technique is commonly used when photographing ocean tides, waterfalls, and the movement of the stars in the night sky over an extended period of time (often employing exposure times of minutes or even hours).

Most photo-editing software tools (e.g. Adobe Photoshop [4]) contain exposure adjustment controls that can emulate the effect of changing the shutter speed after the picture has been taken (assuming the scene is static, and thus no motion blur). Shutter speed, which has essentially a linear effect on the final exposure, can be digitally adjusted by linearly scaling the pixel values[1] by the ratio of the desired to the captured shutter speeds.

It has become fairly common for photographers to capture a sequence of images, each taken with a different shutter speed, and then merge the "exposure stack" into a single high dynamic range (HDR) image [16]. There has been extensive research on methods to directly capture HDR images [5, 62, 60] in a single image without needing to construct the entire "exposure stack". The multiplexing methods presented in this thesis build on many of the ideas used to capture HDR images.

[1] We assume pixels are in a linear colorspace, or gamma correction can be inverted.

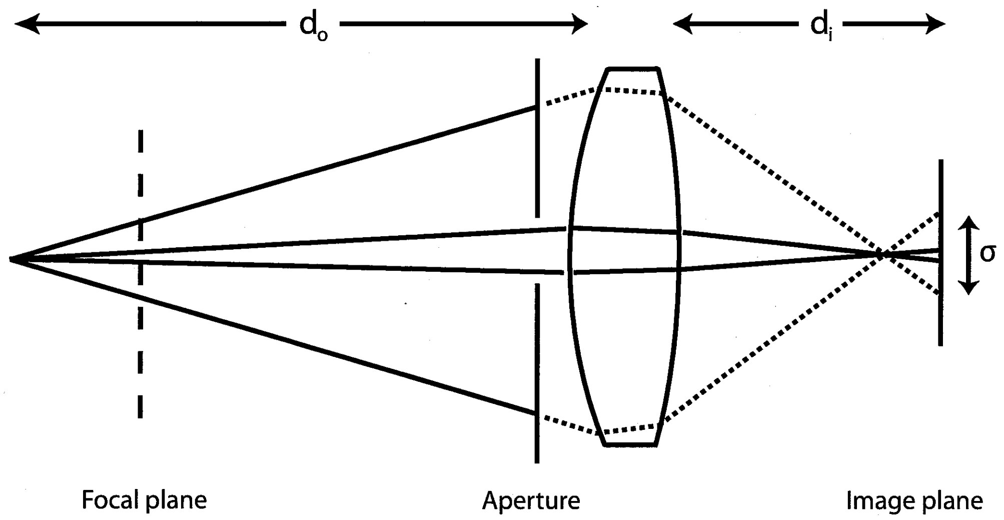


Figure 2-3: *The relationship between aperture size and depth of field. The diameter σ of the defocus blur is dependent on the aperture size. Consequently, the depth of field is also dependent on the aperture size.*

2.2.4 Aperture

The aperture setting on a lens controls the physical size of the opening through which light rays can pass. Recall that aperture size is the other parameter that determines exposure value, and hence adjusting the aperture size primarily effects the overall exposure of the captured image. However, aperture size is a critical imaging parameter for another reason: aperture size directly controls the depth of field of the captured image.

Depth of Field Depth of field (DOF) is the term photographers use to describe the range of distances in a scene that appear sharp in the final image. Assuming aberration free paraxial optics, exact focus is only possible for a single depth plane (equation 2.20). When located at the focus distance, a point object source will produce a point image. However, when located any any other depth, a point object source will image to a blurry spot, the size of the spot being directly related to the distance from the plane of focus. The size of the blurry spot on the image sensor, σ, is commonly called the *circle of confusion*, or defocus blur. We can define DOF as the

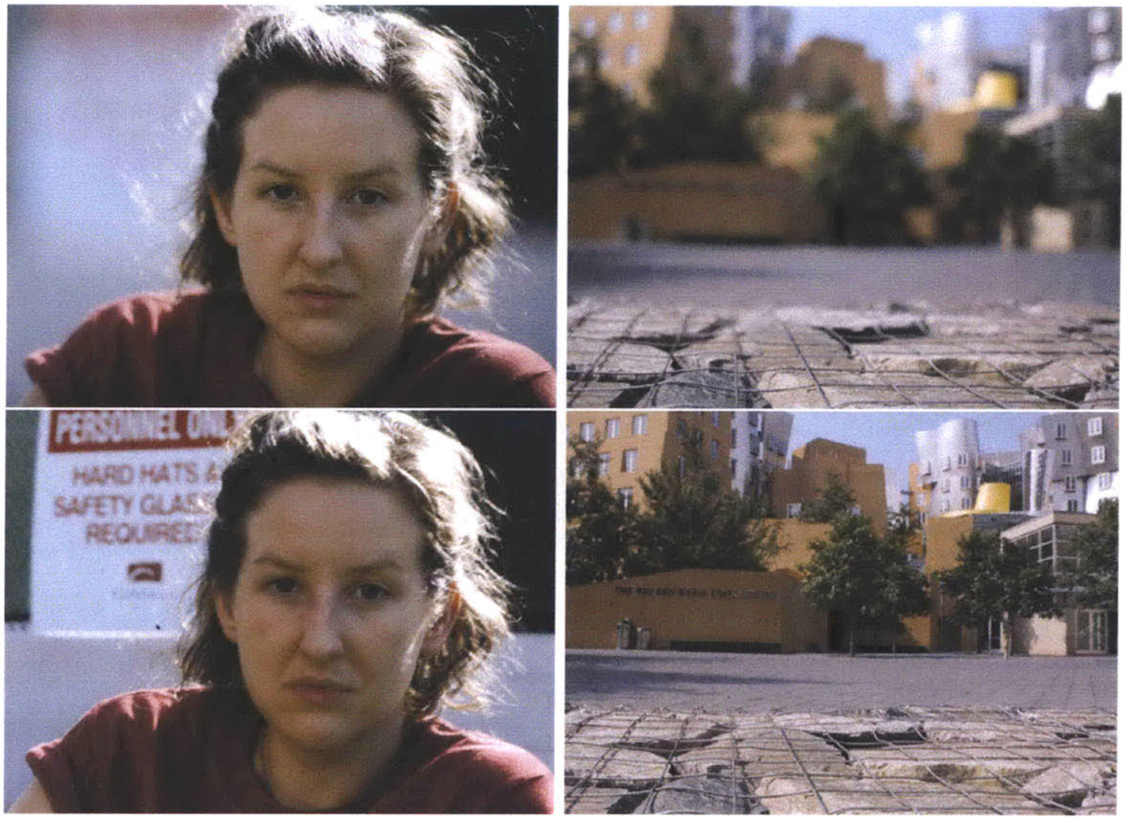


Figure 2-4: *Effect of aperture on depth of field. The top row shows two scenes photographed with a large aperture. The bottom row shows the same scenes photographed using a small aperture. The shallow depth of field obtained when using a large aperture helps remove the distracting background of the portrait scene (left column) and focus the attention on the subject. While large depth of field is necessary to capture sharply the full range of depth in the landscape scene (right column).*

set of scene depths that produce an acceptably small circle of confusion such that the image still appears sharp. The definition of DOF is somewhat vague, partly because the perceived sharpness or blurriness of the final image will depend on a number of factors, including the size of the sensor pixels, the physical size and resolution of the display media (e.g. physical prints vs. viewed on a monitor), viewing distance, etc. and thus the size of an acceptably small circle of confusion can vary.

While any point not located at the plane of focus will produce a nonzero circle of confusion, the exact size of the blur circle will depend on the distance from the focal plane *and* the aperture setting of the lens. Figure 2-3 shows that as the size of the aperture decreases the circle of confusion also decreases. It is in this way that aperture can be used to control DOF. This coupled behavior of the aperture, influencing both exposure and DOF, makes controlling the aperture settings challenging for many amateur photographers.

Aperture settings are indicated using f-numbers, denoted as $f/\#$ or N, which measures the ratio of the lens focal length f to the diameter D of the aperture opening:

$$N = f/\# = \frac{f}{D} \tag{2.24}$$

Typical consumer lenses have f-numbers that range from f/2.0 to f/22, (larger f-numbers correspond to smaller aperture sizes) often in $\sqrt{2}$ increments. Each $\sqrt{2}$ increase in the f-number halves the amount of light that enters the lens. For example a f/2.0 aperture captures twice as much light as a f/2.8 aperture. Image quality and sharpness is also dependent on aperture size. At large aperture settings (e.g. f/2.0), rays strike the periphery of the lens (no longer satisfying the paraxial assumption) and can suffer from increased aberrations. At the other extreme, small aperture settings (e.g. f/16) are reaching the size where diffraction can limit the ability of the system to resolve small details. Often the best image quality is obtained for settings between f/5.6 and f/8.0.

Figure 2-4 shows the qualitative effect of the aperture and DOF for both portrait and landscape photography. The top row of images shows two scenes photographed using a large aperture, producing shallow DOF. The bottom row shows the scenes photographed with a small aperture, and thus a large DOF. In the portrait, shallow DOF is used to blur the distracting background and bring attention to the subject. A large DOF is necessary to capture the entire landscape image sharply.

2.2.5 Flash

A flash is an invaluable accessory in low-light situations, unfortunately many amateur photographers have difficulty properly using the flash, and their photos often suffer from flash artifacts. Red eye, unflattering highlights, strong shadows, glare and under-exposed backgrounds are some of the most common flash artifacts. Photographers have developed methods and guidelines for improving the quality of flash images[31]. For example, flashes are often bounced off large reflectors or walls to create an area light source that produces softer lighting and can reduce highlights and strong shadows. Slow-sync is a method that combines a flash with a long exposure to help increase the background exposure level. Researchers have also developed computational methods to improve flash photography. Several methods have been developed that combine flash and no-flash images of a scene in order to create better images with pleasant lighting and remove artifacts like shadows, glare and reflections[21, 70, 6, 51]. Flash and no-flash images have also been used to extract background and foreground mattes[79, 78].

2.2.6 Focal Length

The last camera setting we will discuss is the lens focal length f. Given a fixed sensor size, the focal length f of the lens determines the field of view of the image. There are two types of lenses, prime and zoom. Prime lenses have fixed focal lengths, while the focal length of a zoom lens can be adjusted between a range of focal lengths. In general, prime lenses are optically less complex than zoom lenses, requiring fewer lens elements, and achieving higher optical quality and with larger apertures.

Cropping the field of view of an image to a smaller size can emulate the effect of using a larger focal length lens. In this way, "digital zoom" can be applied after the photo has been taken to adjust the focal length and field of view. Of course, the obvious limitations are that you can only decrease the field of view (i.e. increase the focal length) and the resolution of the final image is decreased.

2.3 Bilateral Filtering

In this final section of the chapter we review the bilateral filter[81] and its applications. The bilateral filter is a nonlinear edge-preserving filter that has recently begun being used extensively in the computer graphics, computer vision and computational photography communities. In particular, the bilateral filter, and its extensions have

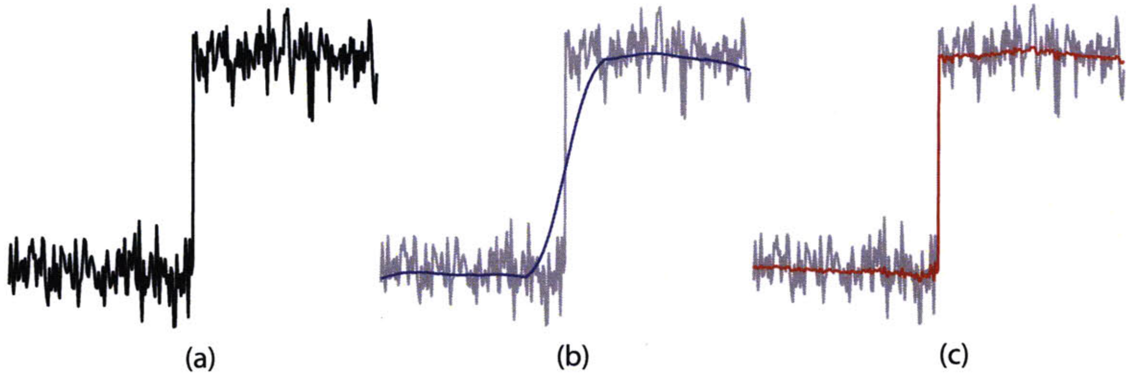


Figure 2-5: *Comparison of Gaussian and Bilateral filtering. A noisy step function (a) is denoised using Gaussian filtering (b) and bilateral filtering (c). The noisy input function is shown in light gray for comparison. The Gaussian filter overly smoothes the step edge, while the non-linear bilateral filter is able to preserve the sharp step edge.*

been used to combine two images of a scene, one taken with a flash, one without, into a single image with the best properties of each[21, 70]. In chapters 4 and 5 we investigate methods to capture both the flash and no-flash images at the same time.

Given an input image I, the output of the bilateral filter at a pixel p, J_p, is defined as:

$$J_p = \frac{1}{W_p} \sum_{s \in \Omega} N_{\sigma_s}(s-p) N_{\sigma_r}(I_s - I_p) I_s, \tag{2.25}$$

with the normalization factor W_p defined as:

$$W_p = \sum_{s \in \Omega} N_{\sigma_s}(s-p) N_{\sigma_r}(I_s - I_p), \tag{2.26}$$

and N_{σ_s} and N_{σ_r} are zero mean Gaussian functions with standard deviations σ_s and σ_r respectively. As can be seen from equation 2.25, the bilateral filter combines Gaussian weighting on both the spatial distance *and* on the intensity difference between neighboring pixels. The term N_{σ_r}, often called the range Gaussian, provides the weighting based on the intensity differences, and is what gives the bilateral filter its edge-preserving properties. Essentially, while standard Gaussian filtering gives the most weight to spatially close pixels, regardless if they are across an intensity edge, the N_{σ_r} term down-weights pixels on different sides of an intensity edge, even if they are close spatially. The effect of the range Gaussian can be seen in figure 2-5, where we compare standard linear Gaussian filtering with nonlinear bilateral filtering. The Gaussian filtered version smoothes across the step edge, while the bilateral filtered version keeps the edge intact.

The range Gaussian term N_{σ_r} makes bilateral filtering nonlinear, as it is dependent on the image intensities and not just spatial positions. This nonlinearity makes evaluating the bilateral filter directly computationally intensive, and also prevents using standard acceleration methods for linear filters such as the Fast Fourier Transform (FFT). Fortunately, several methods have been developed that significantly accelerate bilateral filtering[19, 14, 68, 85].

Chapter 3

Multi-Aperture Photography

3.1 Introduction

In this chapter we focus on one of the central aspects of optical imaging: the effects of a finite aperture. Compared to pinhole optics, lenses achieve much higher light efficiency at the cost of integrating over a finite aperture. The choice of the size of the aperture (or $f/\#$) is a critical parameter of image capture, in particular because it controls the depth of field or range of distances that are sharp in the final image. Depending on the type of photography, more or less depth of field can be desirable. In portraits, for example, shallow depth of field is desirable and requires a wide physical aperture. Unfortunately, many users do not have access to wide aperture cameras because of cost, in the case of SLRs, and limited physical sensor size, in the case of compact cameras. Photographs of multiple subjects are even more challenging because the aperture diameter should be large enough to blur the background but small enough to keep all subjects in focus. In summary, aperture size is a critical imaging parameter, and the ability to change it during post-processing and to extend it beyond the physical capabilities of a lens is highly desirable.

Design goals We have designed an imaging architecture that simultaneously captures multiple images with different aperture sizes using an unmodified single-sensor camera. We have developed a prototype optical system that can be placed between

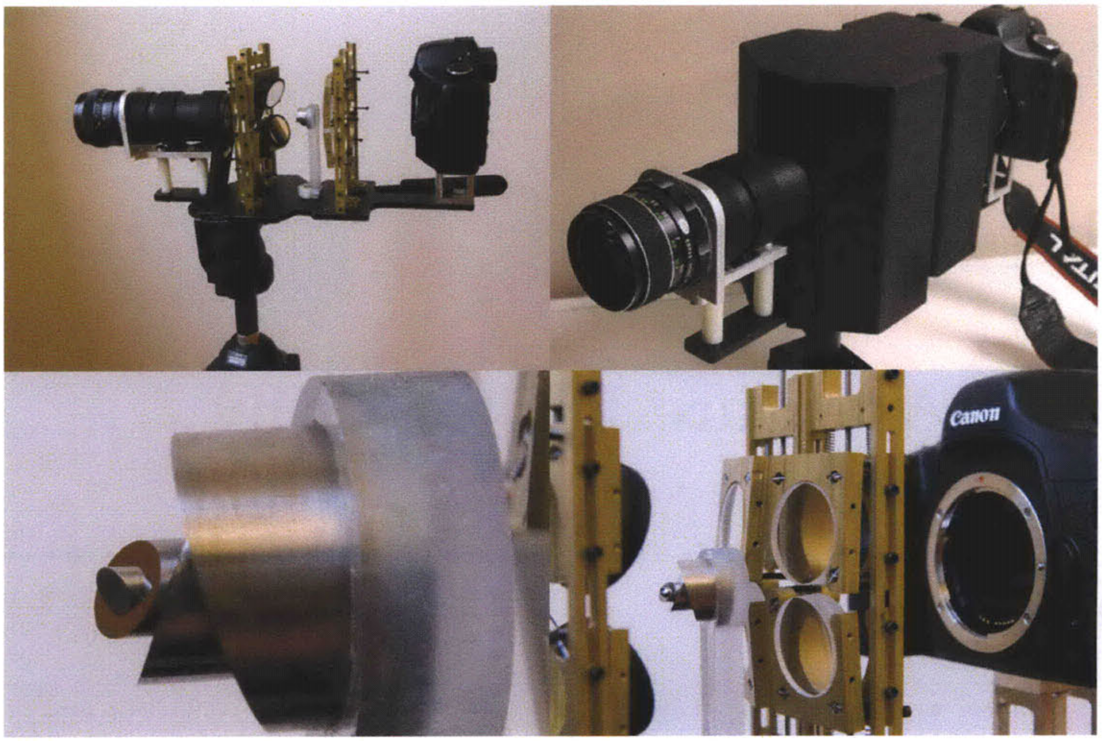


Figure 3-1: *Photographs of our prototype optical system to capture multi-aperture images. The system is designed as an extension to a standard DSLR camera. The lower right image shows a close-up of the central mirror used to split the aperture into multiple paths.*

the camera and an unmodified lens to split the aperture into four concentric rings and form four images of half resolution onto the camera sensor.

We designed our optical system to meet four goals:

- Sample the 1D parameter space of aperture size and avoid higher-dimensional data such as full light fields.
- Limit the loss of image resolution, in practice to a factor of 2 × 2.
- Design modular optics that can be easily removed in order to capture standard photographs.
- Avoid using beam splitters that cause excessive light loss (e.g., [53]).

One advantage of our design is that the captured images can be added directly to form new images which correspond to various aperture settings, without requiring nonlinear processing and image analysis. More advanced post-processes can be performed using depth from defocus.

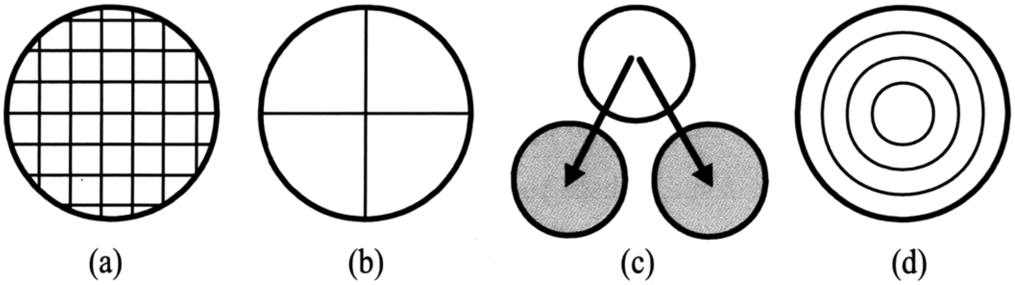


Figure 3-2: *(a-c) A pictorial representation of previous methods of splitting the aperture. (a) The standard Light field camera design [Adelson and Wang 1992; Ng 2005; Georgiev et al. 2006]. (b) The splitting used by Aggarwal and Ahuja [2004] for high dynamic range imaging. (c) Beam splitters [Mcguire et al. 2007], and (d) our decomposition.*

3.2 Related Work

We focus on work related to modifying the aperture of a camera, which includes both camera systems that permit the capture of richer data streams and image processing algorithms that take advantage of this captured information.

Plenoptic cameras instantaneously capture the full light field entering the optical system. Various designs have been investigated and implemented [3, 67, 59, 64, 26]. These designs vary in size and optical components, but, in principle, plenoptic cameras trade spatial resolution to capture directional information about the rays entering the optical system. This also can be seen to split the main aperture into a number of rectangular areas and form a separate image from each of these sub-apertures (Fig. 3-2(a)). A typical drawback of these approaches is a severely reduced spatial resolution, where the grid subdivision of the aperture results in a reduction that is quadratic in the number of samples along one axis. An advantage of these approaches is that the final image can be a simple linear combination of the recorded data [64]. Non-linear reconstruction can afford better resolution trade-offs, but is more prone to artifacts.

Another interesting way of splitting the light entering an optical system is to use a pyramid mirror placed behind the main lens [5]. This effectively subdivides the aperture into "pie slices" and each of these sub-apertures is captured using a separate sensor (Fig. 3-2(b)).

Perhaps the most common way of splitting the light entering an optical system is to use beam splitters to replicate the optical path (Fig. 3-2(c)). Prisms and half-silvered mirrors are typical elements used to perform this task. In this context, 3-CCD cameras use a dichroic prism to split the light and create three copies of the image, each with a different spectral band. Many other designs have been investigated. In particular, McGuire et al. use different aperture and focus settings to perform matting [54]. Watanabe et al. have demonstrated a real-time depth from defocus system that uses beam splitters and active illumination [83]. We have considered designs with beam splitters to decompose the aperture, but they usually require multiple sensors and lose light because they need to rely on occlusion by a mask to select a sub-region of the aperture.

Hasinoff and Kutulakos use a brute force approach by capturing all possible combinations of aperture and focus settings for use in a depth from focus method [33]. This method produces very high quality depth maps but requires several hundred exposures.

Applications of splitting the aperture include: extending dynamic range [5, 60], computing depth [3, 22, 36], alpha matting [54], multi-spectral imaging [60], high-speed imaging [32], changing viewpoint [67, 59, 64], digital refocusing [39, 64, 26], synthetically changing depth of field [26], and extending depth of field [64, 53].

3.3 Optical Design

3.3.1 General Principle

The optical system must accomplish two tasks simultaneously: 1) split the circular aperture of the main photographic lens into a central "pinhole" image and several

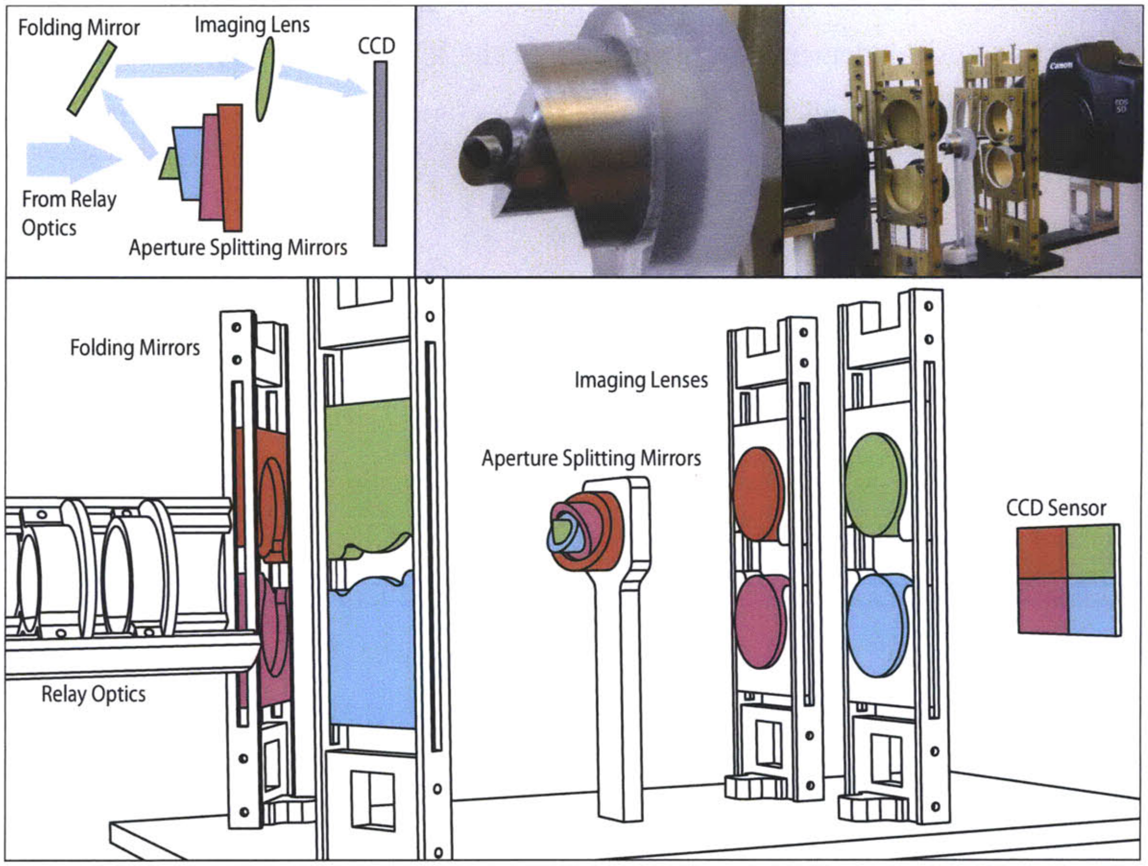


Figure 3-3: *Schematic diagrams and photographs of our optical system and a sample image taken with the camera. Our design consists of a main photographic lens imaged through relay optics and split using a set of tilted mirrors. The relay optics produce an image of the main lens' aperture onto the aperture-splitting mirrors. The diagram is color coded to display the four separate optical paths. The right image shows data acquired from our camera. Each quadrant of the sensor captures an image from a different aperture ring. Colors are used to denote the separate optical paths of each aperture ring. The unit is enclosed in a custom cover during normal operation (see Fig. 3-1).*

concentric rings and 2) re-sort and image the light rays from the "pinhole" and rings onto the imaging sensor of a digital camera.

We use a relay system to image the physical aperture diaphragm of the photographic lens to a plane outside of the lens, called the exit pupil [34]. The exit pupil is then divided into a central disc region and a number of concentric rings. Refractive/reflective optical elements are used to steer the light rays passing through different regions. Finally, additional lenses are used to form images on a single imaging sensor.

3.3.2 Our Design

Our optical design for splitting the aperture into a central disc and a set of concentric rings is conceptually similar to a Cassegrain lens. A schematic is shown in Fig. 3-3. The self-contained optical assembly is placed between a regular photographic lens and the camera body. The entire optical package includes the relay optics tube, a 4-way aperture-splitting mirror to divide the lens aperture, and four sets of image forming mirrors and lenses. We chose to divide the full photographic lens aperture into *N=4* sub-aperture areas because this division achieves a good trade-off between the loss of sensor resolution and the ability to perform our proposed post-exposure edits. We use a 12.8MP *Canon EOS-5D* digital SLR camera, and achieve around 3MP spatial resolution for each of the four images. From four images we are able to acquire depth maps, interpolate and extrapolate depth of field, and synthetically refocus.

Relay optics are necessary for two reasons. First, to relay the intermediate image formed by the photographic lens to the camera's sensor. More importantly, relay optics are necessary to image the physical aperture diaphragm of the photographic lens out of the lens barrel, i.e., forming a new exit pupil at the *4*-way aperture-splitting mirror. From the conjugate relation between the object and image [34], we know that splitting the exit pupil is equivalent to splitting the physical aperture itself. By using the *4*-way aperture-splitting mirror at the new exit pupil, we reflect the incident light rays to four different directions according to where they pass through the aperture. For example, the size of the central "pinhole" mirror is equivalent to a lens aperture size of $f/8$[1]. Therefore, all rays which pass through a virtual $f/8$ aperture are steered along the optical path denoted in green, as shown in Fig. 3-3. Please note: the red, blue, green and purple colors in Fig. 3-3 are used only to distinguish the four different light propagation paths, and are not related to any real color filtering/modification. The outer radii of the other three rings are chosen to correspond to virtual aperture sizes of $f/5$, $f/3.7$ and $f/2.8$, respectively. The corresponding folding mirrors reflect the light back in the direction of the camera sensor. An imaging lens is used between the folding mirror and the camera to reduce the imaging distance and ensure that the final image size is reduced to 1/4 of the size of the camera sensor. As one can see from Fig. 3-3, the optical axes of all four optical paths deviate from the original photographic lens' optical axis. This deviation is corrected by tilting the imaging lenses according to the Scheimpflug principle [74].

The *4*-way aperture-splitting mirror used to divide the lens aperture is made by machining custom steel tubes and polishing the reflecting surfaces until they are optically flat. An image of the mirror is shown in Fig. 3-3. The angles of the directions to which light is reflected must be large enough to ensure the folding mirrors and their mounts do not interfere with the relay optics tube or block the incident light from the relay optics. However, this angle cannot be too large, as the larger the angle, the more the light deviates from the original optical axis, which can cause several field related optical aberrations such as coma and astigmatism. Additionally, large angles increase the possibility for vignetting from the camera mount opening to occur. Finally, larger reflecting angles at the aperture-splitting mirror increase the amount of occlusion due to splitting the aperture. Further details are discussed in Section 3.3.4.

We have designed the relay optics to extend the exit pupil $60mm$ behind the relay optics tube. The *4*-way aperture-splitting mirror is placed at this location. The innermost mirror and the small ring mirror are tilted 25° to the left (around the x-axis), and 18° up and down (around the y-axis) respectively. The two largest rings

[1] Ideally, the central mirror would be as small as possible to approximate the infinite depth of field of a true pinhole camera. Due to manufacturing limitations, $f/8$ was the smallest possible mirror we could build.

are tilted 18° to the right (around the x-axis), and 16° up and down (around the y-axis) respectively. The tilt angle for this arm is slightly smaller because these two rings are farther behind the relay optics. To generate the same amount of lateral shift at the position of the folding mirrors, the desired deviation angle is smaller.

The position of each folding mirror is determined by the tilting angle of the corresponding aperture-splitting mirror. The folding mirrors and imaging lenses are mounted on four, six-degree of freedom kinetic mounts, which ensure that the mirrors and lenses can be configured to the correct position and angle to form four sub-images at the four quadrants of the camera sensor (See Fig. 3-1 and 3-3).

3.3.3 Calibration

Our system needs to be both geometrically and radiometrically calibrated. Because we used stock optical elements, and built all the mounts and enclosures, there are significant distortions and aberrations in each image. We have observed spherical field curvature, radial distortion, tilt in the image plane, and variations in the focal length of each ring image (due to slight differences in the optical path lengths resulting from imprecise alignment and positioning of the mirrors and lenses). To geometrically calibrate for distortions between ring images, we photograph a calibration checkerboard and perform alignment between images. Through calibration we can alleviate some of the radial distortion, as well as find the mapping between images. In addition, imaging an LED (see Fig. 3-4) was very useful to perform fine scale adjustments of the mirror and lens angles.

We radiometrically calibrate the rings by imaging a diffuse white card. This allows us to perform vignetting correction as well as calculate the relative exposures between the different rings. Finally, we apply a weighting to each ring, proportional to its aperture size.

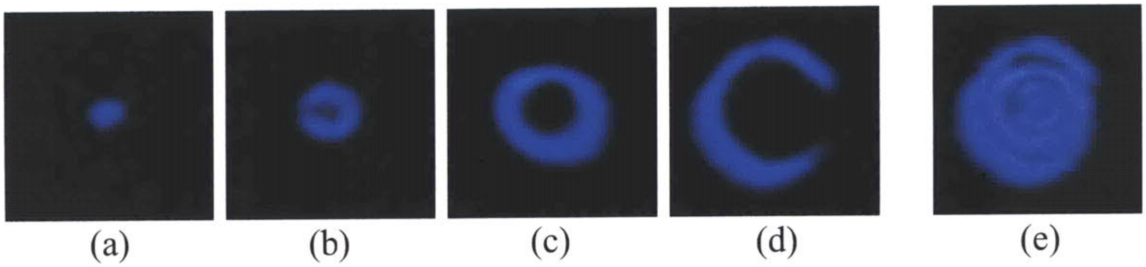


Figure 3-4: *Point Spread Functions of our system captured by imaging a defocused point source imaged through the apertures of (a) central disc, (b) the first ring, (c) the second largest ring, and (d) the largest ring. The horseshoe shape of the rings is caused by occlusion. (e) The sum of (a)-(d). Misalignment causes skew in the shape of the PSFs.*

3.3.4 Occlusion Analysis

The four reflecting surfaces on the *4*-way aperture-splitting mirror are tilted to different directions. They are placed in a spiral-step configuration as shown in Fig. 3-3. Each of the outer rings is partially occluded by its neighboring inner ring's extruded supporting base. The aperture of the central disc area is unaffected, but a small portion of each of the other three ring apertures is occluded. The occlusion can be reduced by arranging the four reflection surfaces such that the normal direction transition between each of the adjacent surface pairs is minimized. For example, as shown in Fig. 3-3, the angle between the normal direction of the central disc and that of the first ring is 36°, but the angle between that of central disc and the second largest ring is 49.1°. This arrangement produces less occlusion than if the reflection direction of the first and second rings is swapped. We captured the images of the occluded apertures by probing the camera system with an LED point source at a position off the plane of focus.

3.4 Applications

In the previous section we described an optical system to capture images, denoted as $R_0, \ldots, R_{M-1}$, taken from $M = 4$ annular apertures simultaneously. Using our representation, we can synthesize a sequence of M images, $I_0, \ldots, I_{M-1}$, of different aperture sizes by accumulating the rings, i.e., $I_j = \sum_0^j R_i$. In a single exposure, our technique can generate multiple images of the same scene, each as if taken with a different aperture setting. This set of multiple images then can be used to recover a *defocus gradient map*, which measures at each pixel the change in defocus blur as a function of aperture size. Our defocus gradient map is very similar in concept to a traditional depth map, and in fact we could compute depth from the sequence of aperture images using standard depth from defocus algorithms [13]. The defocus gradient map is integral to accomplishing sophisticated operations, such as extrapolating shallow depth of field beyond the limits of the largest aperture, changing the apparent plane of focus, and increasing image sharpness using a depth guided deconvolution scheme.

3.4.1 Defocus Gradient Map

Assuming that our scene is composed of planar patches parallel to the image plane, we can approximate defocus blur over each patch as a convolution, where the filter size is determined by the patch's distance from the plane in focus. In his original work on depth from defocus, Pentland [69] derives an equation relating the object distance d_o to internal camera parameters and the defocus blur kernel diameter σ (see Fig. 2-3):

$$d_o = \frac{fd_i}{d_i - f - \sigma N}, \tag{3.1}$$

where f is the focal length, d_i is the distance between the lens and the imager plane, and N is the f-number (the ratio of the focal length to the diameter of the lens). Solving for σ we have:

$$\sigma = \frac{(d_i - f)d_o - fd_i}{Nd_o}. \tag{3.2}$$

The sign of σ differs for points in front of $(-)$ and behind $(+)$ the in-focus plane. We assume the camera is focused on the nearest scene point to avoid the standard depth from defocus ambiguity, as well as to restrict σ to positive values. Substituting

$G = [(d_i - f)d_o - fd_i]/d_o$ and $l = 1/N$, we can rewrite Eq. 3.2 in the linear form:

$$\sigma = Gl, \tag{3.3}$$

where G is the derivative of σ with respect to the inverse f-number $1/N$. The utility of Eq. 3.3 is that if G is known, the blur kernel size can be calculated for an arbitrary f-number. We call our estimate of G at each pixel the "defocus gradient map."

The defocus gradient map measures the change in size of blurring kernels as a function of aperture size. The defocus gradient map is related to the distance of an object from the plane of focus. An object on the focus plane will always be sharp (hence its blurring kernel will be zero for all aperture sizes). An object away from the focus plane will become blurrier as the aperture size is increased, and in particular, the rate at which it becomes blurry is dependent on its distance from the in-focus plane.

It is possible to calculate the defocus gradient map by running standard depth from defocus algorithms to recover a depth map and then directly converting the depth map to a defocus gradient map. However, we do not require exact depth per se, and in fact we are more interested in the apparent change of defocus blur with respect to aperture size. The defocus gradient map is a simpler, more direct representation for our applications.

We can use Eq. 3.3 to compute the defocus gradient map. At a pixel p, the change in blur with respect to aperture size should lie on the line $\sigma_p = G_p l$. Therefore, if we can estimate σ_p in each of our aperture images, we can directly calculate G_p as the slope of the line. Unfortunately, it is difficult to directly estimate σ_p, and instead we adopt a hypothesis-and-test framework. For a set $\{G_i\}$ of discrete values of G, we hypothesize that pixel p has defocus gradient G_i, and test this hypothesis against our observed data.

In practice we use a Markov Random Field [8, 80] framework to solve for the defocus gradient map. We chose MRFs to solve for the defocus gradient map because it globally optimizes our data objective while simultaneously applying spatial regularization. We set up a MRF where the labels for each pixel are assigned from a set $\{G_i\}$ of discrete values of G. The optimization objective function is a standard combination of a data term E_i^p (Eq. 3.4) and a smoothness term, S. The penalty E_i^p for assigning a node p the label G_i is calculated as:

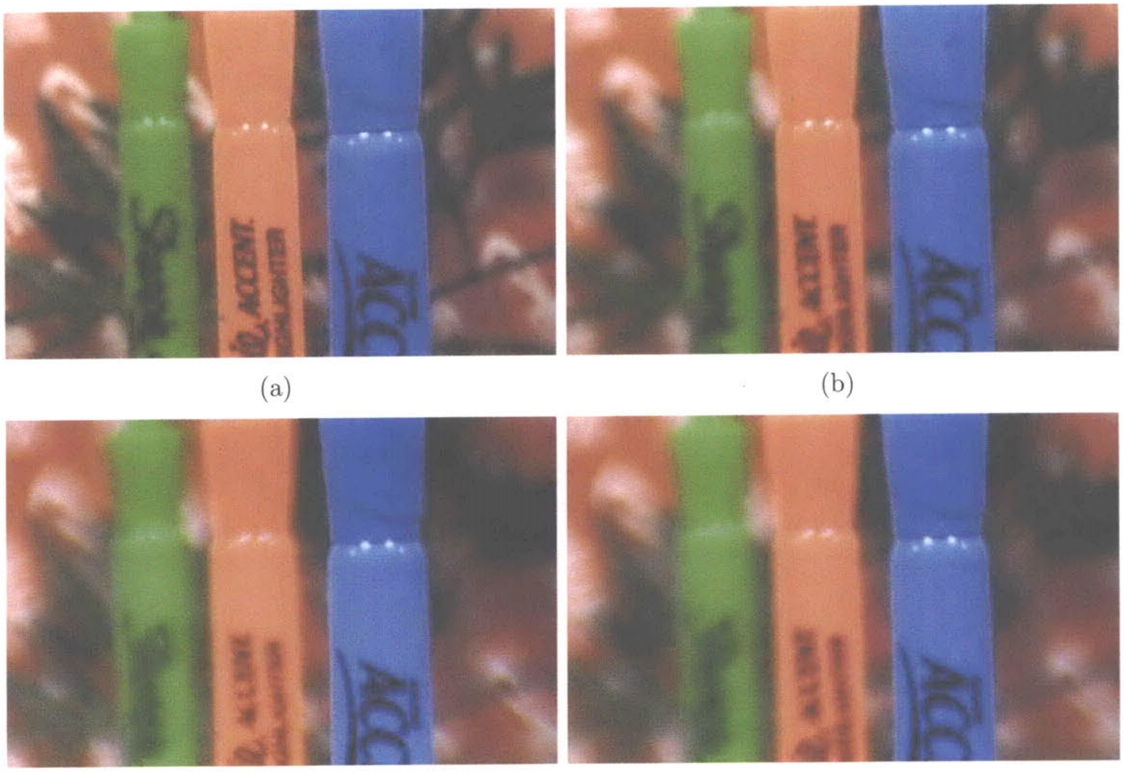
$$E_i^p = \sum_{j=1}^{M} (I_0 \otimes H(\sigma_{ij}))(p) - I_j(p). \tag{3.4}$$

Equation 3.4 measures the error at pixel p between the smallest aperture image I_0, convolved with the expected defocus PSF $H(\sigma_{ij})$ with diameter $\sigma_{ij} = G_i(1/N_j)$ and the observed blur (as measured in image I_j). We model the PSF as a disc of diameter σ_{ij}.

The smoothness (regularization) term, S, defines how similar we would like spatial neighbors to be. S is specified as horizontal S_x and vertical S_y pairwise weights between adjacent pixels. S_x is calculated as $S_x = \exp(-(I_{0x})^2 \times \alpha)$, where I_{0x} is the horizontal spatial derivative of I_0, and α is a bandwidth parameter. S_y is calculated analogously. Our assumption is that depth discontinuities often occur across intensity edges. In flat intensity regions, our smoothness term encourages nearby pixels to have the same labels. However, regions with large gradients (e.g., edges) incur a small smoothness weight, and thus are less penalized for having different depth labels. Similar discontinuity-preserving smoothness terms have been used previously [8, 45].

3.4.2 Interpolating and Extrapolating Aperture Size

Our optical system captures four images of the scene simultaneously, each from an annular section of the aperture (see Fig. 3-3). It is possible to reconstruct the four aperture images by successively accumulating rings, e.g., the third aperture image is constructed by summing the inner disc and the next two aperture rings. Furthermore, interpolating between reconstructed images approximates the effects of varying the aperture size, from the smallest to the largest captured apertures. This provides a way to continuously adjust the depth of field in an image. Figure 3-5 shows several images with varying aperture sizes constructed by summing the individual rings.



(a) (b) (c) (d)

Figure 3-5: *Images created by summing rings of the aperture. (a) Small central aperture. (b) Sum of the central disc and the first ring. (c) Sum of the central disc and the first two rings. (d) Sum of all aperture regions. Notice that the depth of field is decreased as aperture rings are added.*

Using our defocus gradient map we can extrapolate shallow depth of field beyond the physical constraints of the maximum aperture. This is accomplished by extrapolating the size of the blurring kernel (using the defocus gradient) and blurring the "pinhole" image. Figure 3-6(a) shows an image taken at $f/1.8$ and Fig. 3-6(b) shows a synthesized version computed using our defocus gradient map technique. The defocus gradient map was computed from four separate exposures ($f/\# = 22$, 13, 8, and 4, respectively). The difference image is shown in Fig. 3-6(c). Figure 3-7 shows an extrapolated image taken with our camera.

Noise Characteristics Interpolated and extrapolated images have different noise characteristics. Images created using the interpolation technique show noise characteristics similar to a standard image of the same aperture size. Interpolated images

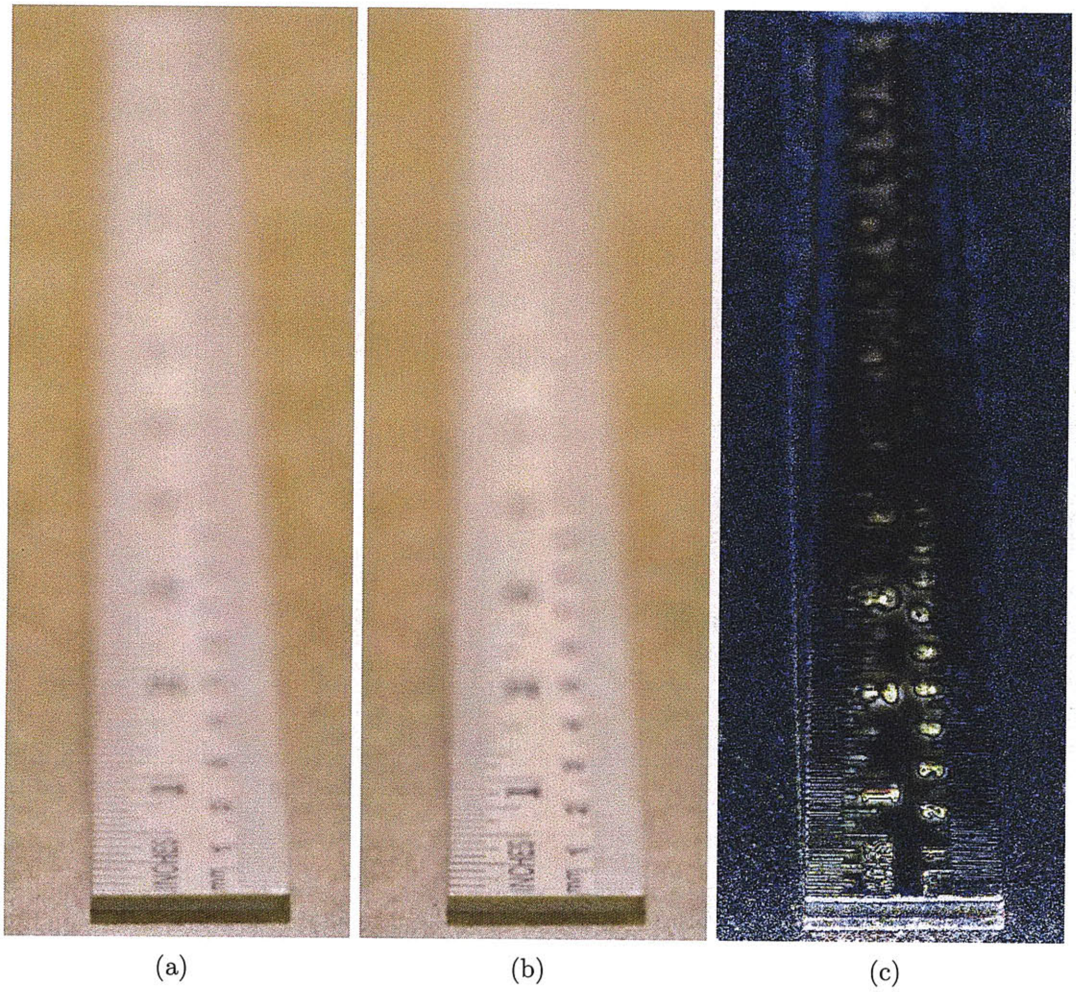


Figure 3-6: *A comparison of our extrapolation method to a reference image. (a) A reference image taken at f/1.8. (b) Our extrapolated synthetic result. (c) Difference image. The images used to compute (b) were taken in multiple exposures, without our prototype optical system. The mean error is under 5% of the average image intensity.*

have decreased shot noise due to summing multiple aperture rings. Extrapolated images use only the "pinhole" image, and thus points on the image plane exhibit the noise characteristics of the "pinhole" image. Additionally, some light efficiency is lost due to the added elements in the relay and mirror system.

3.4.3 Synthetic Refocusing and Guided Deconvolution

The defocus gradient map is an encoding of the relative distance from the focus plane at each image point. In particular, image points near the in-focus plane will have a

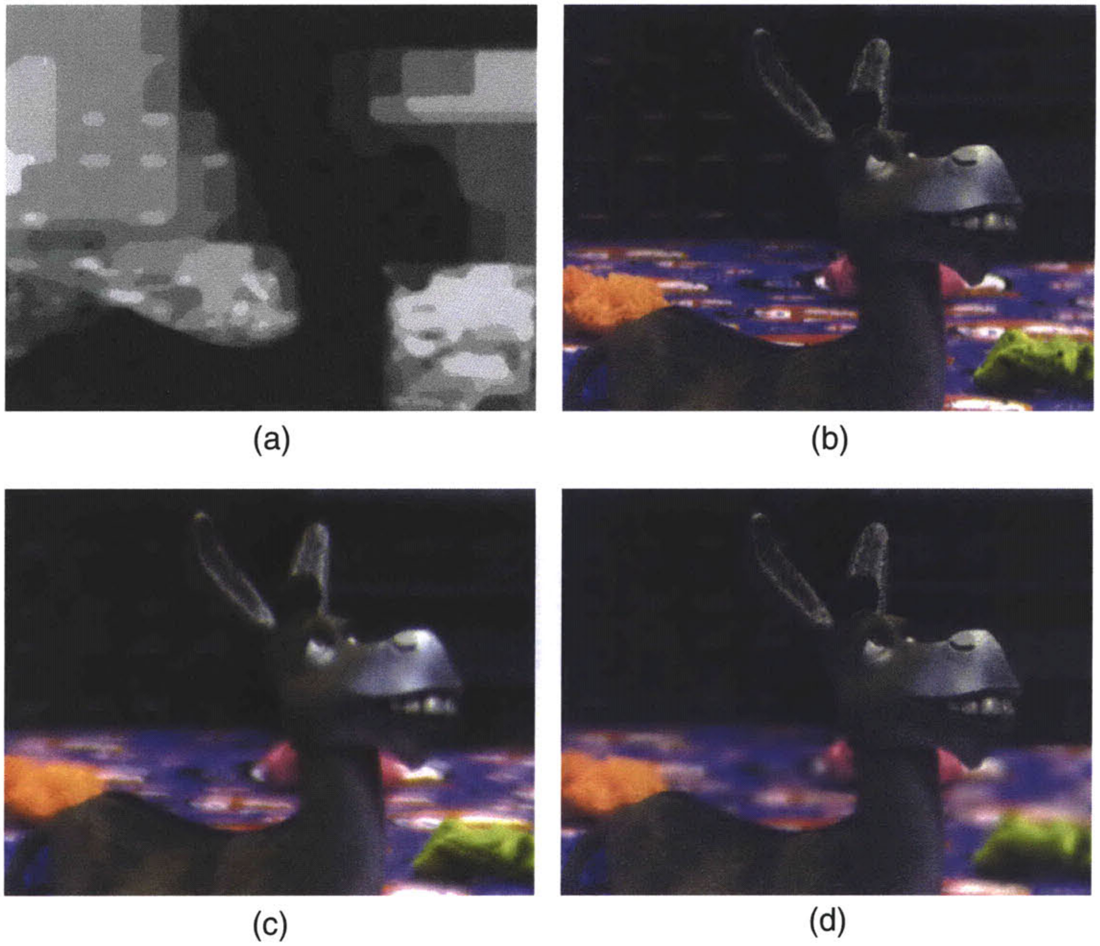
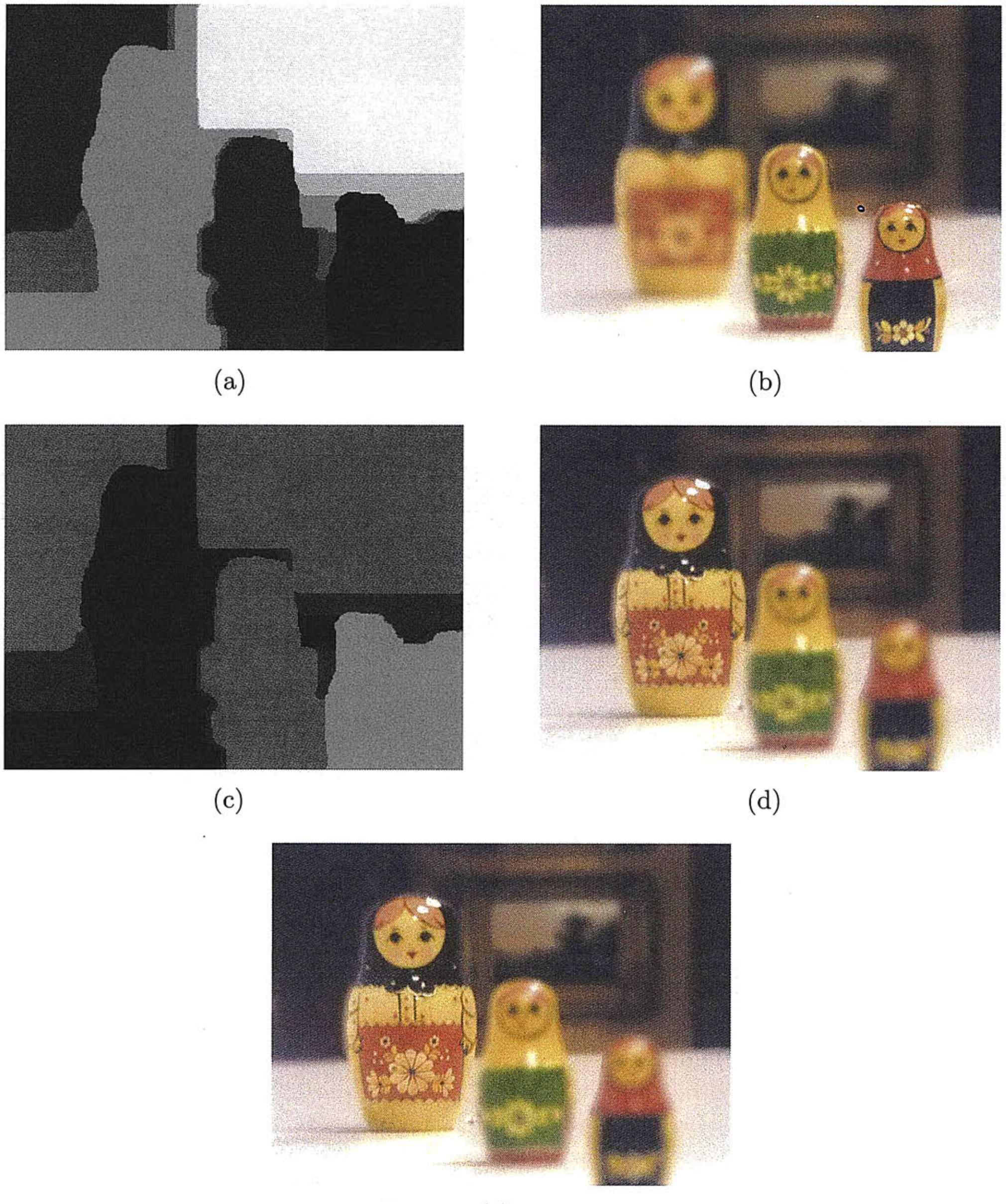


Figure 3-7: *Extrapolating the aperture to twice the area of the largest captured aperture. (a) Defocus gradient map; darker colors indicate smaller gradients (i.e., points closer to the in-focus plane). (b) f/8 image (smallest aperture). (c) Accumulated f/2.8 image. (d) Extrapolated f/2 image.*

small defocus gradient, and the defocus gradient will increase the further the point is from the in-focus plane. Since we store discrete labels in the defocus map, we can relabel, or shift, the values in the map by an offset to achieve a synthetic refocusing effect. After offsetting the labels we can perform depth of field extrapolation (Sec 3.4.2). Figure 3-8 shows an example of our synthetic refocusing method. In Fig. 3-8(a) the focus is on the doll in front. In Fig. 3-8(b) the focus has been "moved" to the doll in the back. Although directly shifting the labels in the defocus gradient map is not equivalent to moving the in-focus plane, it produces qualitatively convincing refocusing effects. An alternative approach is to convert the defocus gradient map into a



(a)

(b)

(c)

(d)

(e)

Figure 3-8: *Refocusing on near and far objects. (a) is the computed defocus gradient map. Dark values denote small defocus gradients (i.e., points closer to the in-focus plane). Using (a) we can synthesize (b) the near focus image. (c) Defocus gradient map shifted to bring the far object to focus. (d) Synthesized refocus image using (c). (e) Synthesized refocus image using our guided deconvolution method. Notice the far object is still somewhat blurry in (d), and the detail is increased in (e).*

depth map (Eq. 3.1), which can be adjusted directly and used as input to a lens blur filter (e.g., in Adobe Photoshop).

It is important to note that we cannot perform actual refocusing of the image, we can only synthesize a new shallow depth of field image where the perceived image plane has been moved. In particular, we must rely on the large depth of field present in the smallest aperture image to provide all the detail at the shifted in-focus plane.

We now describe a form of guided deconvolution to enhance details in the "pinhole" image. The defocus gradient map provides an estimate of the PSF at each pixel. This PSF estimate can be used to adjust the deconvolution kernel used at each pixel. If we use K different labels when calculating the defocus gradient map (i.e., K depth values), then we run K separate deconvolutions of the "pinhole" image, each with a different PSF to produce a set of deconvolved images $\{D_i\}$. The size and shape of each PSF used is determined by Eq. 3.3 (a different value of G for each of the K labels, l is determined by the size of the central disc aperture, e.g., $f/8$). We use Richardson-Lucy deconvolution ("deconvlucy" in Matlab).

The final output of our deconvolution method is assembled by compositing the K separate deconvolutions based on the defocus gradient map labels. For example, if a pixel p has label k ($1 \leq k \leq K$) in the defocus gradient map (i.e., pixel p is at depth k), then we copy the corresponding pixel location in the k^{th} deconvolved image D_k (which has been deconvolved with a PSF corresponding to objects at depth k) into the output image. This provides a spatially adapted deconvolution method: The PSF used to calculate the deconvolved output at a pixel is determined by the estimated depth/defocus at the pixel. In contrast, traditional deconvolution methods use a single PSF for the entire image. The main benefit we have found is that our method alleviates most of the over-sharpening artifacts that are common with deconvolution methods by spatially tailoring the PSF to the local blurriness present in the image. Figures 3-8(d) and (e) compare refocusing with and without our deconvolution, respectively. Note the improved detail in our deconvolved version.

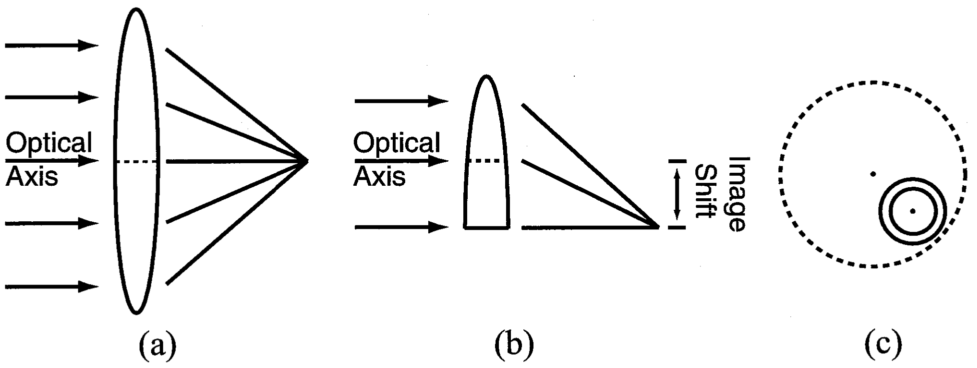


Figure 3-9: *Alternative designs. (a) Normal focusing through a lens. (b) Laterally shifting the focused image by decentering the optical axis. (c) Example of cutting an annular region from a larger theoretical lens. The distance from the optical axis of the annular region and the optical axis of the theoretical lens determines the amount of lateral shift.*

3.5 Discussion

3.5.1 Alternative Optical Designs

We have investigated several alternative optical designs in addition to the design previously described in this paper. We would like to briefly discuss two of these alternative designs, with the hope that they may inspire further research.

The first alternative design involves placing a complex refractive element (i.e., a lens with non-traditional surface curvature) at the exit pupil which is then used to divide and image concentric rings of the aperture. The surface of the refractive element is designed such that light striking an annular region of the lens forms an image shifted laterally from the center of the CCD imager. Figure 3-9 describes how it is possible to laterally shift the image formed through a lens by decentering the lens with respect to the main optical axis. Conceptually, the proposed refractive element is composed of decentered annular sections, each cut from a theoretical larger-radius lens. See Fig. 3-9(c) for an example. To mimic our current design, the refractive element would consist of four annular regions which form four images in the quadrants of the CCD. Two advantages of this design are that it could easily be extended to 9 or 16 regions and it splits the aperture at a single plane without occlusion problems. The main disadvantage we found in simulation was that because of the unusual surface shape, and the limitation to using a single lens element, the optical aberrations were unacceptably high. Additionally, it would be difficult and expensive to manufacture with glass, although it may be more practical using molded plastic optics.

The second alternative design is to place a micro-lens array over the CCD, where each lenslet is a miniaturized version of the complex refractive element just described. This is similar to the light field camera design proposed by Ng [64], however, instead of capturing a full light field, would integrate light from annular regions of the aperture, thus enabling higher spatial resolution. We believe that because the micro-lens array is responsible for a very local re-sorting of light rays, the quality would be higher than any of the previously proposed designs. Unfortunately a micro-lens array cannot be removed in order to take standard photographs.

3.5.2 Limitations

A potential drawback of our system is that our mirror design requires very precise and difficult alignment of the optical elements in order to minimize aberrations. However, precision manufacturing techniques could produce an optical system with quality comparable to standard photographic lenses. Additionally, our system has difficulty recovering accurate depth and defocus information in regions without texture. This problem is common to many depth from defocus algorithms, and we employ the standard solution of using spatial regularization. It may be possible to use the unusual shapes of each aperture ring along with coded aperture methods to further improve the depth maps.

Another limitation is that our synthetic refocus method is unable to correctly synthesize blur effects across depth discontinuities. Unlike light field cameras, we are unable to capture the subtle parallax effects that occur across occlusion boundaries.

Chapter 4

Multiplexed Flash Illumination for Relighting and Depth Extraction

4.1 Introduction

Taking good photographs in low-light situations is challenging. Standard photographic solutions for capturing low-light images include using a tripod and a long exposure, high sensitivity film (or high gains in the case of digital photography), large aperture lenses, or a flash. Using a tripod produces images with the least noise, however setting up a tripod is inconvenient and motion blur can be an issue due to the long shutter times. Using a larger aperture lens is often a good solution, however some scenes may still be too dark to be well exposed. Additionally, large aperture lenses decrease depth of field and are often very expensive. Increasing the sensitivity of the sensor is a common solution, however this can significantly increase image noise, particularly for consumer point-and-shoot cameras. For these reasons, a flash is often the most practical option for dark scenes. Unfortunately, using a flash can produce several irritating and unwanted artifacts, and it takes a skilled photographer to avoid or minimizes them. In particular, flash photographs often suffer from uneven foreground-background exposure, red eye artifacts, color casts, and strong highlights on foreheads and other glossy surfaces. All of these effects often combine to destroy the natural ambiance of the available lighting in the scene, producing harsh, unflattering pictures. Flash/no-flash methods [21, 70] combine two images of a scene, one taken with a flash and one taken without, to produce a new image with the best properties of both images. While these methods work well for static scenes, the requirement of multiple exposures is a significant barrier to the average user, and infeasible for moving scenes because of the need for multiple exposures.

We propose a method for simultaneously capturing flash and ambient lighting information in a single exposure. We use a coded flash to project a high-frequency pattern onto the scene, which spatially multiplexes flash and no-flash information (see Figure 4-1). Spatially multiplexing flash and no-flash gives information about both the detail and color in the flash regions and the ambient illumination in the no-flash regions, though with a reduced resolution and extra indirect illumination due to the flash.

We build on the idea of assorted pixels [63, 62] but extend it to computational illumination. We aim to spatially multiplex flash information into a single image. In contrast to previous work on temporal multiplexing of illumination, e.g. [15, 86, 56, 61, 76], our goal is to *simultaneously* record both types of information. Simultaneous capture is important for dynamic scenes to avoid a temporal mismatch between the images corresponding to the two lighting conditions.

Furthermore, we want to leverage the defocus information from the multiplexing light pattern in order to infer depth information. However, in contrast to previous work [58], we seek to do so in the presence of ambient illumination and with a light pattern that is not co-axial with the lens, in order to increase light efficiency.

Our main contributions are:

- The introduction of assorted flash pixels to record spatially multiplexed flash and ambient information.

- An analysis of possible sampling and reconstruction schemes.

- The estimation of a sparse depth map from flash defocus.

- Single exposure flash/no-flash applied to dynamic scenes.

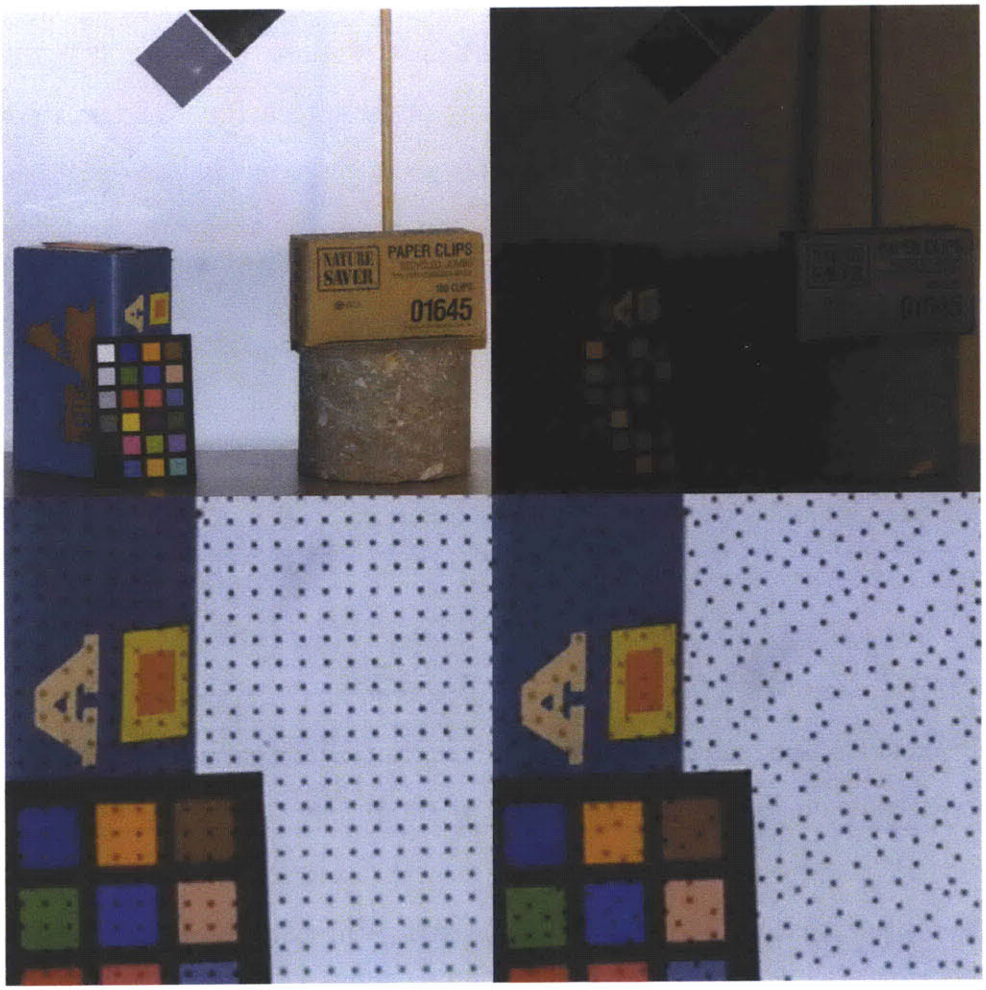


Figure 4-1: *Top: A scene photographed with and without flash. Bottom: Close-ups of two possible samplings of flash and no-flash pixels using our multiplexed flash illumination.*

4.2 Related Work

Assorted Pixels, proposed by Nayar and Narasimhan[63], introduced a method for sampling multiple dimensions of imaging (e.g. brightness, color spectrum, time, polarization) by mosaicing pixels that sample different dimensions into a single array of pixels. We extend this concept by allowing the illumnation to be mosaiced. Unlike traditional Assorted Pixels, in which the multiplexing occurs purely on the image sensor, we multiplex at the illumination source and must identify which pixels on the sensor sample which dimension.

Structured lighting has been used to accomplish a variety of tasks, including depth and shape estimation[89], refocusing [47, 58], light transport estimation [77], and direct and indirect lighting separation [61]. Many of these techniques are restricted to static scenes because they require multiple images of the scene, while our goal is to capture flash and ambient information for a scene in a single exposure. Additionally, some methods (e.g. [58]), require a coaxial camera and projector which is accomplished using a beam-splitter. Beam-splitters lose lights, and introduce glare, which is undesirable for low-light photography, our main application. Our method uses a binary transparency mask to occlude certain portions of the flash illumination and thus necessarily decreases the flash intensity.

A number of approaches seek to capture a full basis of possible illumination to enable arbitrary relighting of a scene, e.g. [15, 86]. This requires a large number of images to encode the full set of possible directions and, in the case of dynamic scenes, careful correction must be applied to warp the data [86]. In contrast, we seek a simultaneous capture but restrict ourselves to two illumination conditions.

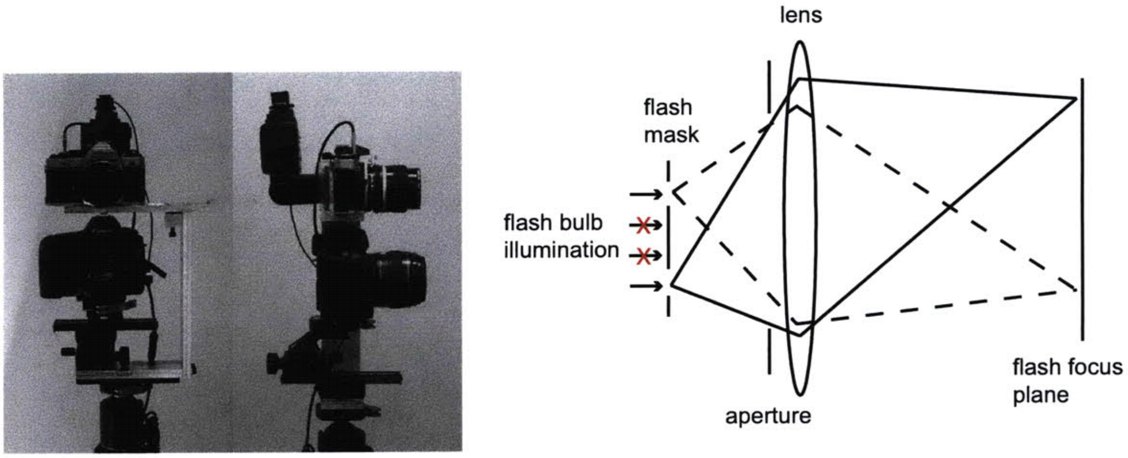
Nayar et. al. [61] describe a method for fast separation of the direct and indirect components of a scene illuminated by a single light source. This method uses a sequence of high-frequency patterns projected onto the scene to perform the separation. They also describe a single exposure version which can produce separations, albeit with a loss in resolution. We assume the scene is lit by two sources, our multiplexed flash and an ambient light source. Our goal is to separate the image into flash and ambient components by spatially multiplexing each component in a single image. We are unable to separate the indirect flash lighting from the ambient lighting, therefore our no-flash pixels capture the combined ambient plus indirect flash lighting.

We build on methods that combine a flash and no-flash image of a scene to produce a new image containing the desirable properties of both [21, 70, 6, 88, 40]. We recover a high resolution detail layer from the flash portions of the image and a large scale intensity layer from the no-flash regions. We demonstrate single exposure flash/no-flash and coarse depth map estimation as applications of our multiplexed flash illumination.

4.3 Multiplexed Illumination

We divide the flash beam into a grid of pixels and allow each pixel to be either on or off. If a flash pixel is on, light is projected onto the scene and focused at the focal plane of the camera. If a flash pixel is off, light is blocked and does not enter the scene. Figure 4-2(b) shows a diagram of our optical system. We do not assume that the flash and camera are coaxial (i.e. no beam-splitter) and instead assume that the flash and camera are only loosely aligned.

4.3.1 Hardware Prototype



(a) (b)

Figure 4-2: *Our prototype(a) consists of a DSLR camera and a film camera modified to project a high-frequency pattern through its main lens. (b) A binary mask is used to block flash rays and produce a spatially varying pattern at the flash focus plane.*

In order to achieve spatially varying flash intensities, we augment a traditional photographic flash with a binary mask pattern and focusing optics. In essence, we turn a traditional flash into a flash projector. The key distinction between our modified flash and a projector is that our flash produces a short burst of light as opposed to continously illuminating the scene, which is essential for freezing motion in photographs. While a projector can be used to simulate our flash, particularly for static scenes, we found there were a number of disadvantages to using a standard consumer projector. In particular, projectors often have low contrast, poor optics (e.g. high chromatic aberation and lens distortion), and a wide fixed aperture providing very shallow depth of field. In our design, we used printed binary transparency masks with a very high contrast ratio and the focusing optics of a high quality professional SLR camera lens with low chromatic aberation and full aperture control in order to control depth of field. An image of our system is shown in Figure 4-2(a). The film camera body on top has been transformed into our "flash projector" by removing the back and placing our mask at the original film plane. A standard flash is mounted behind the "film plane" with a diffuser separating the flash and mask. Essentially, the camera is being used in "reverse" – light is shone from the original image plane out through the lens, producing a focused version of the mask onto the scene. An additional feature of this design is that if the focusing lens is thrown completely out of focus, the flash pattern is removed (via defocus blur) and the multiplexed flash is restored back to a traditional flash, albeit with some loss in intensity due to the mask. This allows the flash to operate in two modes: traditonal and multiplexed flash.

4.3.2 Illumination Patterns

In this section we consider several possible patterns for the flash illumination, including uniform, Poisson-disk, and striped. Once a type of pattern is chosen, the main parameter we explore is the ratio of flash and no-flash pixels in a particular sampling pattern. This ratio has two direct consequences: the sampling rate (in the Nyquist sense) of the reconstructed flash and no-flash images and the total amount of flash light in the scene. In general, the ratio should be chosen such that the resulting sampling rate matches the frequency content of each component. Unfortunately, the frequency content cannot be known a priori, and we are forced to make decisions based on some estimate of expected frequency content and how important it is for the specific application. In particular, we observe that flash/no-flash techniques rely more on the high frequencies of the flash component and on the low frequencies of the no-flash one.

The total number of "on" flash pixels affects the total amount of light sent into the scene, but not the direct light received by a given illuminated point; it only increases the fraction of illuminated points. However, as the ratio of flash pixels increases, this introduces more indirect flash light, "corrupting" the no-flash pixels.

Uniform A uniform checkerboard produces an equal number of flash and no-flash samples, uniformly distributed and regularly spaced. The ratio of flash to no-flash samples can be adjusted to produce regularly spaced samples with greater or fewer no-flash samples. However, although the flash mask contains regularly spaced samples, parallax between the flash and the camera distorts the spacing of samples when imaged at the camera. This distortion makes localizing the flash and no-flash samples on the camera sensor more difficult than with traditional assorted pixel schemes.

Stripes A stripe pattern can help localize the flash and no-flash samples if the optical centers of the flash and camera are carefully aligned. In particular, we can constrain the epipolar geometry such that vertical lines in the flash mask are projected to vertical lines in the camera. A disadvantage of this pattern is that it yields a non-uniform sampling between the vertical versus horizontal dimensions.

Poisson-disk As mentioned above, applications of flash/no-flash pairs usually take their high-frequency information from the flash component. As a consequence, we may choose to undersample the no-flash component to increase the total flash intensity and record a larger number of well-exposed flash pixels. In order to hide some of the aliasing and noise that may occur, Poisson-disk distributed points can be used instead of a uniform grid when undersampling. Although they may appear randomly distributed, Poisson-disk distributed points have the property that there is a minimum distance ϵ between any pair of points. We used the method of Jones[41] to efficiently generate Poisson-disk distributed points. A disadvantage of Poisson-disk distributed points is that localizing the points is difficult, particularly in the presence of parallax and occlusions.

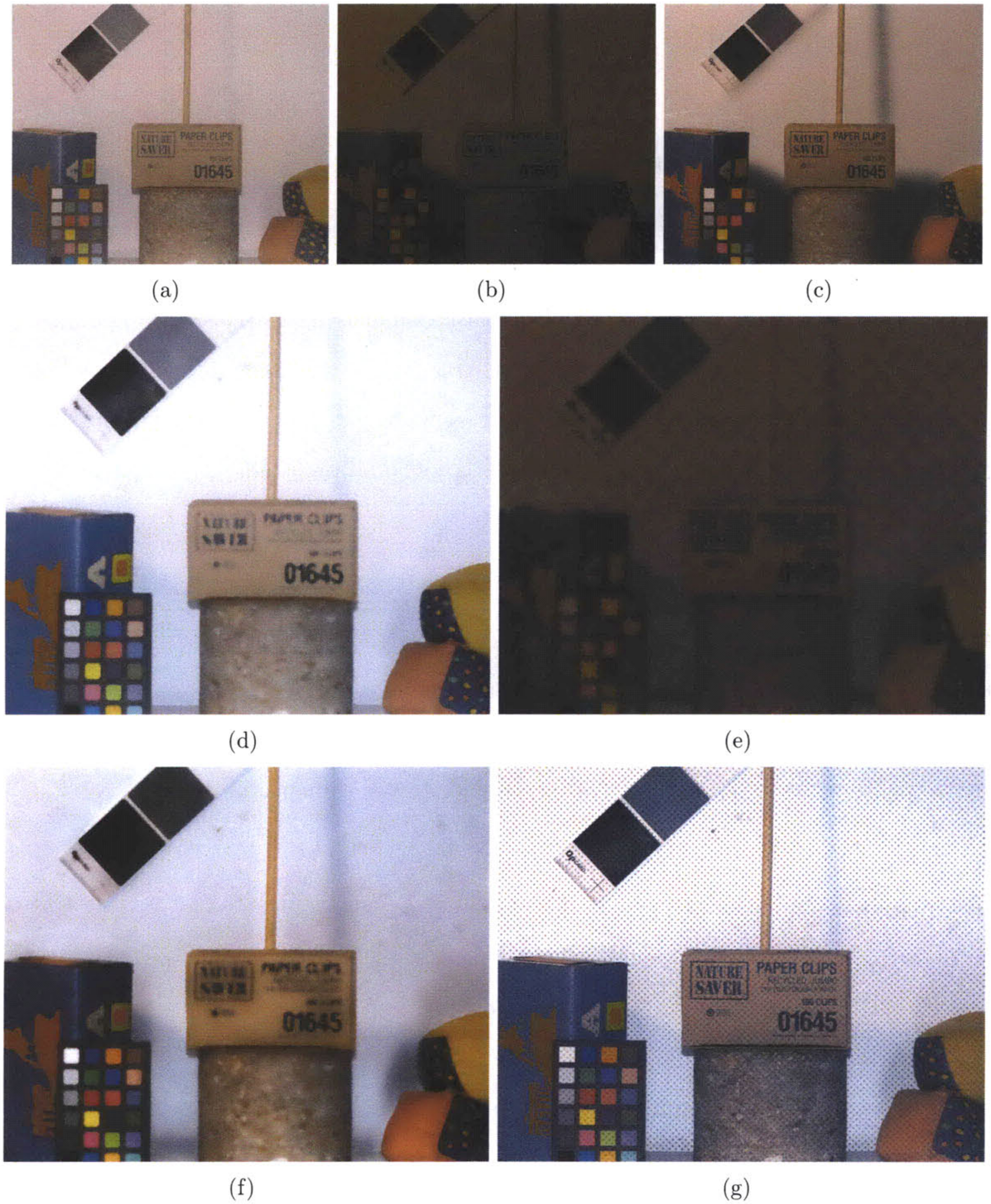


Figure 4-3: *Top row: A scene photographed with(a) and without(b) a standard flash. (c) Standard flash/no-flash image fusion. Our reconstructed flash(d) and no-flash(e) images and our single-exposure flash/no-flash reconstruction(f). The multiplexed input image is shown in (g).*

4.4 Reconstruction

Once we have captured a multiplexed flash illumination image, we must identify and separate the flash pixels from the no-flash pixels. Since we seek a direct simple extension of the traditional flash, the illumination and lens are not confocal and parallax makes it harder to identify which pixels are lit by the flash. Without geometric correspondence, we rely on statistical methods to determine flash and no-flash pixels. A simple method proposed by Nayar et. al. [61] is to choose flash pixels as the maximum pixels in some local window. Similarly, no-flash pixels are the minimum pixels in each local window. To reduce speckle noise, we compute a weighted average of the K largest and smallest pixels in a local window and use this as our estimate of flash and ambient pixels, respectively. The size of the window is chosen differently for flash and no-flash pixels and is based on the known ratio of flash to no-flash pixels.

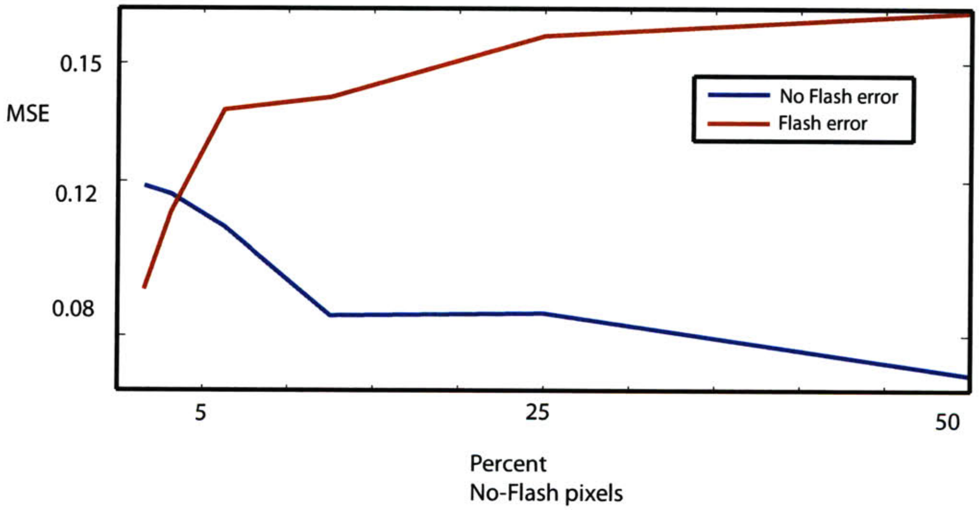
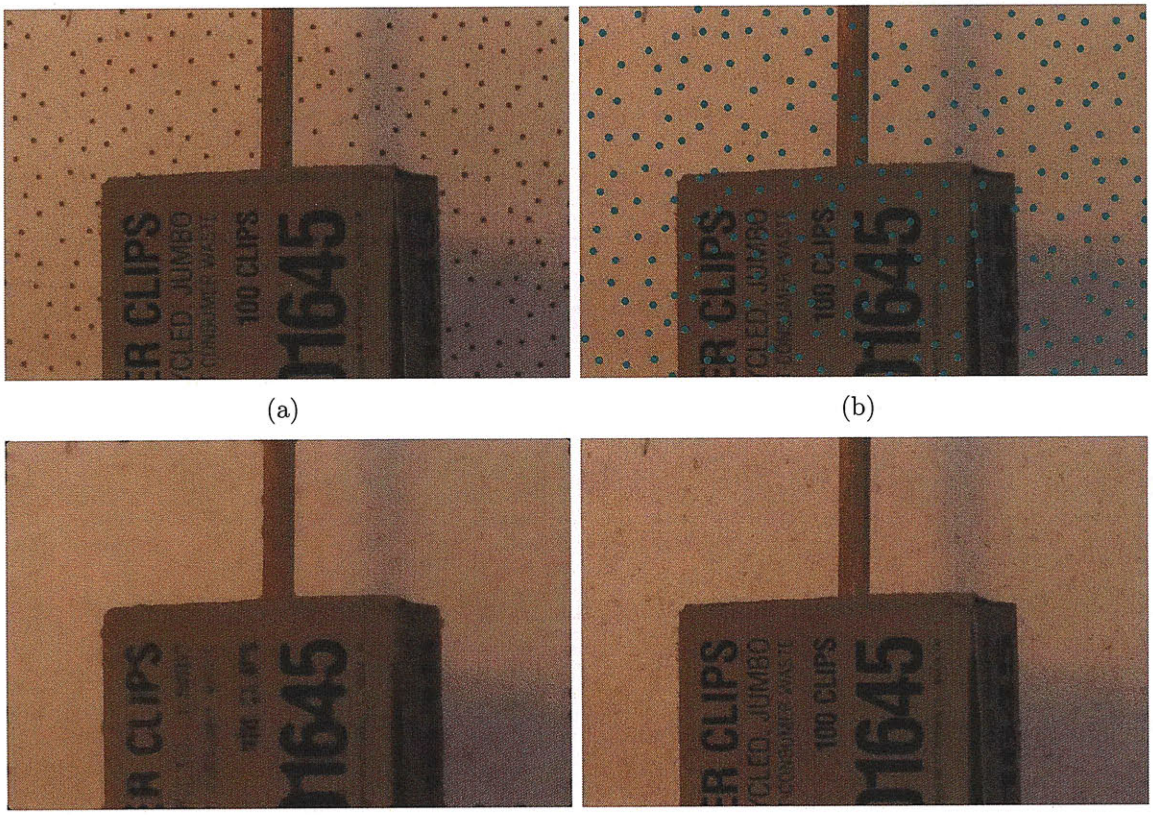


Figure 4-4: *Plot of reconstruction error (MSE) as a function of the percentage of no-flash samples used in a uniform sampling pattern. As the percentage of no-flash pixels increases, the reconstruction error of the no-flash image decreases and the error of the flash image increases. The slope of the graphs suggest using masks with between 5 and 15% no-flash pixels.*

Figure 4-4 shows a plot of the reconstruction error for the flash and no-flash components of our test scene (shown in Figure 4-3) as a function of the percentage of no-flash pixels in the flash pattern. Flash and no-flash images were taken separately and used as the ground truth. As expected, as the percentage of no-flash pixels increases, the no-flash reconstruction error decreases, and the flash reconstruction error increases. This graph suggests that there is little benefit to increasing the ratio of no-flash pixels above ≈ 20%, as the no-flash reconstruction error does not decrease significantly beyond this point. For our flash/no-flash application we use masks with ≈ 6 − 12% no-flash pixels. This trade-off between capturing flash and no-flash pixels is similar to the spatial-angular tradeoff common to many lightfield camera designs [64, 26, 28].



(a)

(b)

(c)

(d)

Figure 4-5: *Using texture synthesis to improve resolution. (a) multiplexed illumination image. (b) No-flash pixels labeled and disk of pixels around each is marked. Standard reconstruction(c) dilates and blurs features. Texture synthesis fills in missing points and avoids resolution loss.*

Improving Resolution As a consequence of using max and min operators to localize points, detail has a tendency to dilate or erode in the flash and no-flash images, depending on the local intensity gradient (see Figure 4-3(f) for an example). In order to improve sharpness and combat dilation and erosion in the flash image, we use texture synthesis to fill in missing data [20, 84, 87]. We remove a disk of pixels around each no-flash pixel location and infill these pixels with texture synthesis (see Figure 4-5). An additonal advantage of using Poisson-disk distributed no-flash samples is that the irregularity of the sampling hides artifacts that may occur when inpainting regions on a regular grid.

4.4.1 Improved Localization

We have developed an algorithm to improve localization of flash and no-flash pixels when using a uniform grid illumination pattern. Because we do not coaxially align the flash projector and the camera, there is parallax which makes localizing the no-flash pixels non-trivial. This is particularly evident across depth discontinuities and on highly curved surfaces. Depth discontinuites cause shifts in the stride between adjacent flash or no-flash pixels. Curved surfaces cause a row (or column) of points to be projected along a curve instead of along a straight line. However, locally (within a small neighborhood) the projected flash pattern is often very similar to a uniform grid. The general idea of our algorithm is to identify likely flash and no-flash pixels and then iteratively propagate local evidence to influence the estimate of nearby locations.

Initialization We initialize the estimated locations using a method similar to Nayar et. al. [61], finding the maximum or minimum pixels in non-overlapping $M \times M$ windows, where M is chosen to match the projected size (or stride) of the illumination pattern in camera pixels. We note that if the focal lengths of the flash projector and the camera are matched then the size of the projected pattern (magnification) is not affected by scene depth or parallax.

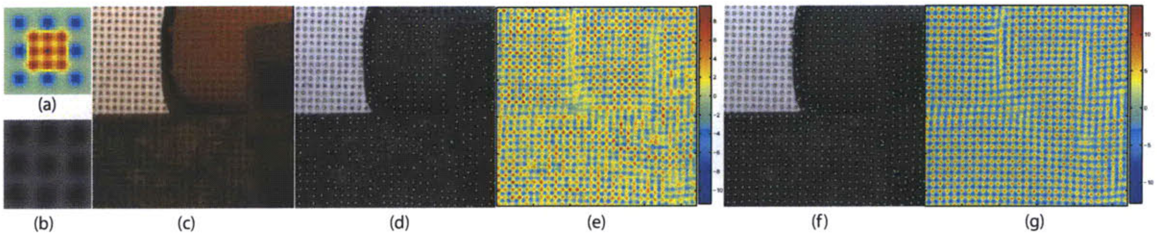


Figure 4-6: *Improving localization. (a) A typical ϕ_{min} kernel. (b) A close-up of the flash pattern projected on a scene. Notice that ϕ_{min} closely resembles (b). (c) An input scene. The inital estimate of no-flash pixel locations (d) and the corresponding P map (e). Notice that (d) has many missing pixels locations and is lacking structure. (f) shows the final estimate of no-flash pixel locations and the final P map (g) after 20 iterations. Our localization method is able to propagate local structure and accurately identifies no-flash pixels.*

Propagating local evidence Given an initial estimate of the flash and no-flash pixel locations, F and NF respectively, we wish to refine them by incorporating a local spatial model of the relative positions between adjacent flash or no-flash pixels. The intuition is that if we have found the location of one flash pixel, we can use this information to help estimate the location of neighboring flash pixels.

To propagate information we construct a map P as:

$$P = F * \phi_{max} + NF * \phi_{min}. \tag{4.1}$$

where F and NF are indicator images that have, e.g. $NF(p) = 1$ for no-flash pixels p and zero otherwise, ϕ_{max} and ϕ_{min} are kernels that encode the relative spatial locations of other flash pixel locations as signed functions and $*$ denotes convolution. For example, ϕ_{max} is positive where we expect to find flash pixels, negative where we expect to find no-flash pixels and zero otherwise. We set $\phi_{min} = -\phi_{max}$.

We iteratively perform a sequence of steps designed to find pixel locations that simultaneously agree with the input data (e.g. are local maximums or minimums) and are appropriately spaced relative to neighboring flash and no-flash pixels. First, we build

$$P_{max} = P \times I_{gray} \tag{4.2}$$

$$P_{min} = P \times (1 - I_{gray}) \tag{4.3}$$

where I is a grayscale ([0-1] normalized) version of the input image I. P_{max} and P_{min} reweight P, giving more weight to flash pixels locations that are in bright parts of the image, and more weight to no-flash pixels locations in dark parts of the image. As P_{max} and P_{min} are processed symmetrically - MAX can be substituted for MIN (and vice-versa)- the remaining steps will be described for computing P_{min} only. We find the set Q of local maxima of the laplacian $\nabla^2 P_{min}$ with response greater than a threshold τ:

$$Q = \left\{ q \middle| \nabla^2 P_{min}(q) > \tau \wedge q = \arg\max_{p \in \Omega_q} \nabla^2 P_{min}(p) \right\}. \tag{4.4}$$

In practice we use a local window Ω_q of 5×5 pixels, and a threshold $\tau = 2$. Local maxima of $\nabla^2 P_{min}$ are points where the gradient is increasing quickly (e.g. at the minimum of no-flash pixels) and we threshold to discard points with small response. We use Q to update our current estimate of no-flash pixel locations NF as:

$$\forall q \in Q, NF(q) = CLAMP(\nabla^2 P_{min}(q) - R_0)/R_1, 0, 1) \tag{4.5}$$

which linearly maps the range $[R_0,R_1]$ to [0,1] and clamps values outside the range (we found $[R_0,R_1] = [1,5]$ to work well in practice). We set $NF(p) = 0$ for all $p \notin Q$. Finally, we recalculate P (using Equation 4.1) and iterate. After K iterations we calculate the final flash and no-flash pixel positions by thresholding F and NF. In practice we run $K = 20$ iterations and use a threshold of 0.2. Figure 4-6 shows an example of P and NF before and after running our iterative estimation algorithm.

4.4.2 Improved Poisson-Disk Localization

Localizing the no-flash pixels when using a Poisson-disk pattern is even more challenging than when using a uniform grid. Unfortunately, the method presented in

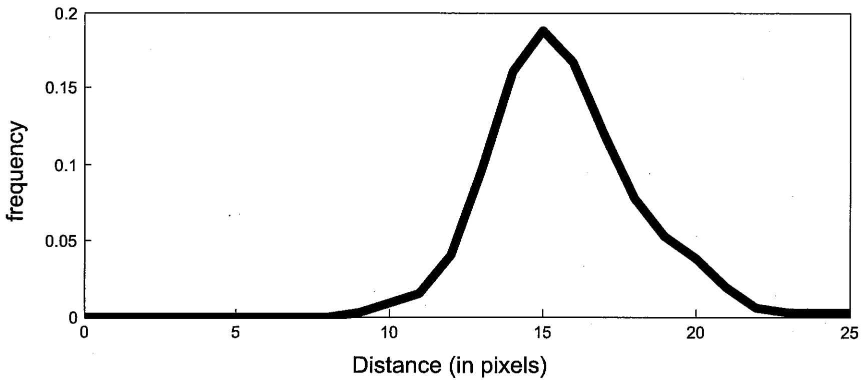


Figure 4-7: *Histogram of the distance (in image pixels) to the nearest neighbor for Poisson-disk distributed point sets projected onto a scene. Distances range between 8 and 25 pixels, and are peaked around 15 pixels.*

section 4.4.1 is only applicable for the case of uniform grids because we make explicit assumptions about the spatial location of neighboring no-flash pixels with respect to each other. Poisson-disk distributed samples do not have such strict and consistent neighborhood relationships, so it is difficult to apply the same strategy to propagate local information as we have demonstrated for the uniform case. However, Poisson-disk distributed points are not entirely random, and in fact do have some spatial relationships that we can hope to exploit. In particular, Poisson-disk points are distributed such that there is at least some minimum distance ϵ between any two points. If additionally, the points are tightly packed, then the distribution of distances between points will be peaked and narrow (see Figure 4-7). We can improve localization of Poisson-disk patterns by using knowledge about the expected distribution of distances between no-flash points. We present a two-stage algorithm for localizing Poisson-disk distributed no-flash points. In the first stage we estimate points that are very likely to be no-flash points. This first stage is designed to reliably detect no-flash points with a low rate of false positives, but as a consequence, may fail to detect many points (i.e. high precision and low recall). In the second stage we use the set of reliably labeled no-flash points, as well as the current distribution of distances between points, and our model of expected distances between points, to estimate the most likely locations for the remaining no-flash pixels.

Initialization The goal of the initialization step is to label a subset of no-flash pixels with a high confidence, while minimizing the rate of false positives. In order to detect no-flash pixels, we make an assumption about the qualitative profile of the flash intensity in the local window around each no-flash pixel. We assume that each no-flash pixel causes a single strong inverted spike in what would otherwise be a constant intensity flash. We model the recorded pixel intensity as a multiplicative modulation of the surface albedo scaled by the flash intensity (which is a function of the flash power, the distance between the flash and the surface point, and the angle between the flash and the surface normal). Under this model and our assumptions about the flash profile, each no-flash pixel will be a local minimum in an appropriately sized local window of constant albedo or texture and constant depth. This assumption fails when the local window contains multiple albedos (e.g. albedo or texture edges), depth discontinuities, and surfaces at high grazing angles. Additionally the window must be sufficiently small enough to contain a single no-flash pixel.

The general idea of our initialization algorithm is to estimate the likelihood a pixel is a no-flash pixel by counting the number of times the pixel is a local minimum in a set of overlapping windows around each pixel. To this end, we construct a map L_{nf} defined at each pixel p as:

$$L_{nf}(p) = \frac{|\{q \,|p \in \Omega_q \wedge p = \arg\min I(\Omega_q)\}|}{|\{q \,|p \in \Omega_q\}|}. \tag{4.6}$$

The numerator of equation 4.6 counts the number of windows Ω_q containing p, where p is the minimum value in Ω_q. The denominator of equation 4.6 normalizes by the total number of overlapping windows Ω_q that contain pixel p. The only parameter necessary to construct L_{nf} is the size of each neighborhood window Ω_q. In practice we use a $M \times M$ windows, chosen to be the approximate stride between neighboring no-flash pixels. Values of $L_{nf}(p)$ range continuously between 0 and 1, and give a normalized count of the number of times a pixel is a minimum in all the windows overlapping it. A value of 1 means that pixel p is the minimum in every window containing p, and a value of 0 means that p is not the minimum in any window

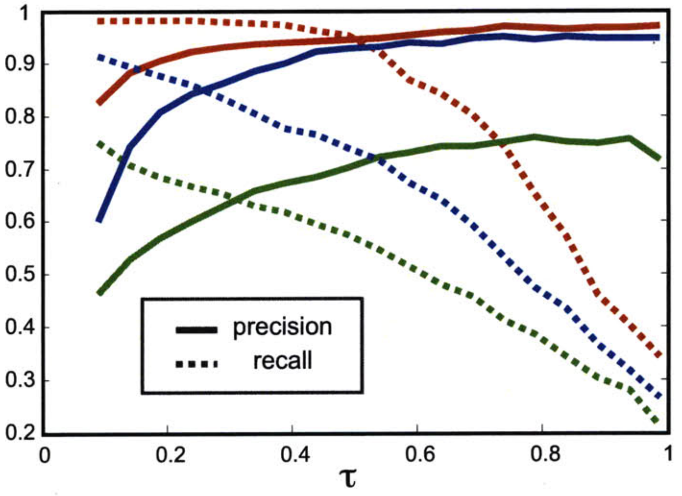


Figure 4-8: *The effect of the τ parameter on precision and recall during the initialization phase of our improved Poisson-disk localization method applied to three example patches (patches are shown in figure 4-9). Each color corresponds to a different patch; solid lines show precision curves, dashed lines show recall curves. As τ is increased precision increases and recall decreases. A value of $\tau = .75$ demonstrated mean precison rates ≈ 0.9 and mean recall rates above 0.5.*

containing p. Finally, we threshold L_{nf}, marking all pixels p with $L_{nf}(p) > \tau$ as no-flash pixels. The value of τ controls precision and recall with which we label no-flash pixels, larger τ produce higher confidence estimates but lower the recall rate. Figure 4-8 shows precision and recall graphs for the test scenes in Figure 4-9 as a function of τ. In practice we used τ values above 0.75 which exhibited mean precision rates ≈ 0.9 and mean recall rates above 0.5.

Locating Remaining Points After performing the initialization stage we have obtained a subset of reliably labeled no-flash pixels and we would like to expand this set to localize the remaining no-flash pixels. As mentioned above, we have no explicit relative or grid structure to aid our task, so instead we use the statistics of the distances between no-flash pixels to guide us. Figure 4-7 shows a representative distance distribution for Poisson-disk distributed points.

We propose an iterative algorithm to fill in the missing no-flash points, with pseudocode provided in Algorithm 1. Given an initial set of no-flash points $R =$

$\{p \,|L_{nf}(p) > \tau\}$ (line 1), we construct a distance map D_R (lines 2-4) that for each pixel p, stores the distance (in pixels) to the nearest point in R:

$$D_R(p) = \min_{q \in R} \sqrt{(p - q)^2}. \tag{4.7}$$

Algorithm 1 Poisson-disk point localization

```
1: $R \leftarrow \{p \,|L_{nf}(p) > \tau\}$
2: for all pixels $p$ do
3:     $D_R(p) \leftarrow \min_{q \in R} \sqrt{(p - q)^2}$
4: end for
5: while $\max(D_R) \geq maxdist$ and $\max(L_{nf}(R^C)) \geq \kappa$ do
6:     $Q \leftarrow \{q \,|D_R(q) \in [mindist, maxdist]\}$
7:     $q^* \leftarrow \arg\max_{q \in Q} L_{nf}(q)$
8:     $R \leftarrow R \cup \{q^*\}$
9:     for all pixels $p$ do
10:        $D_R(p) \leftarrow \min_{q \in R} \sqrt{(p - q)^2}$
11:    end for
12: end while
```

At each iteration, we add the point q^* to R with the maximum value L_{nf} among all points whose distance $D_R(q)$ is within distance bounds *mindist* and *maxdist* (lines 6-8). The distance bounds *mindist* and *maxdist* are chosen to match the statistics of the Poisson-disk distance distribution. The distance map D_R is recomputed based upon the updated set R (lines 9-11). We stop iterating when the maximum distance in D_R is below *maxdist*, or the maximum remaining values in L_{nf} are below a cut-off threshold (line 5).

4.4.3 Sampling Pattern Comparison

We compare the performance of our proposed improved localization methods with the uniform window method proposed by Nayar et. al. [61]. Figure 4-9 shows three patches from a scene illuminated with a uniform grid pattern (subfigures (a)-(c)) and a Poisson-disk pattern (subfigures (d)-(f)). The two flash patterns were chosen to have approximately the same density of no-flash pixels. Overlaid on each figure is hand-labeled ground truth no-flash pixel locations, our no-flash pixel location estimates,

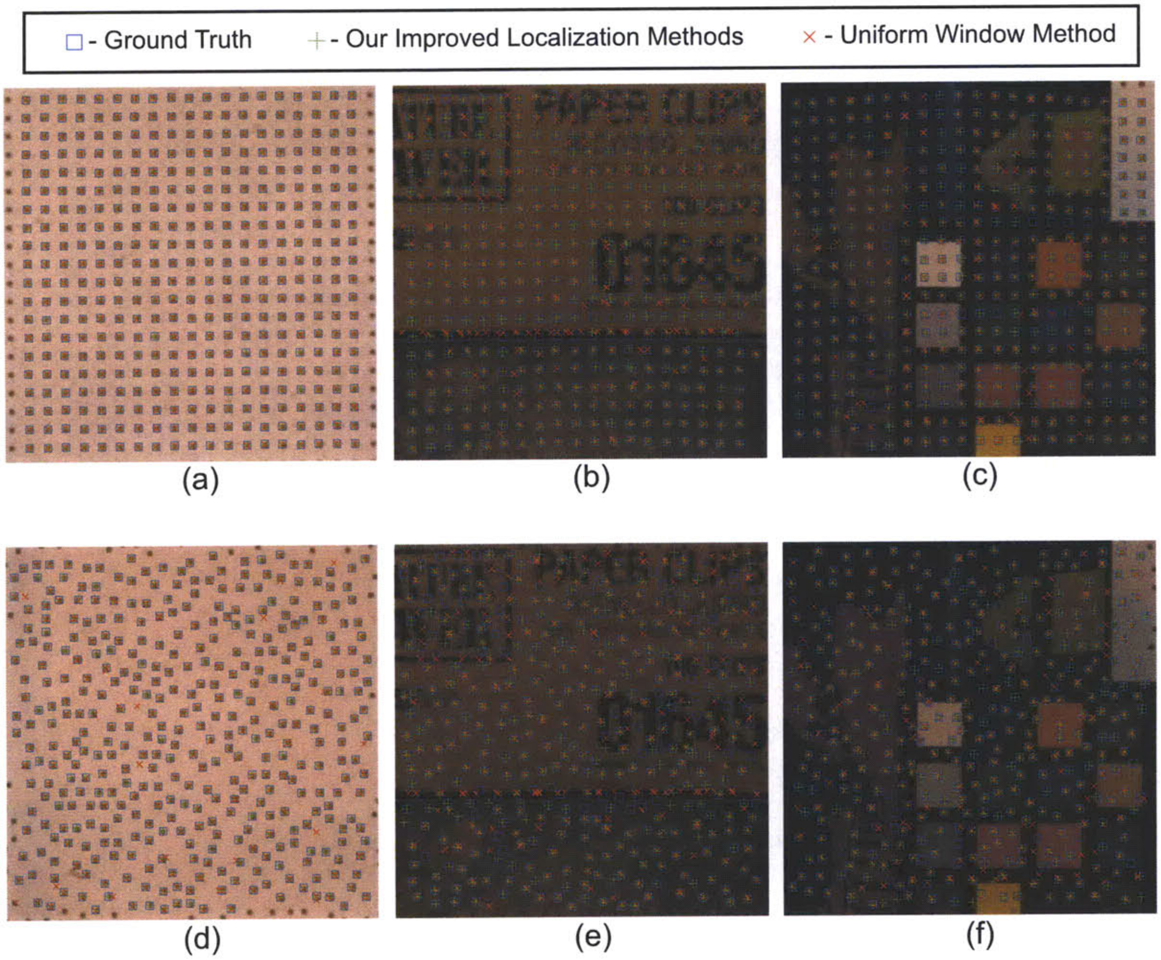

□ - Ground Truth
+ - Our Improved Localization Methods
× - Uniform Window Method
(a)
(b)
(c)
(d)
(e)
(f)

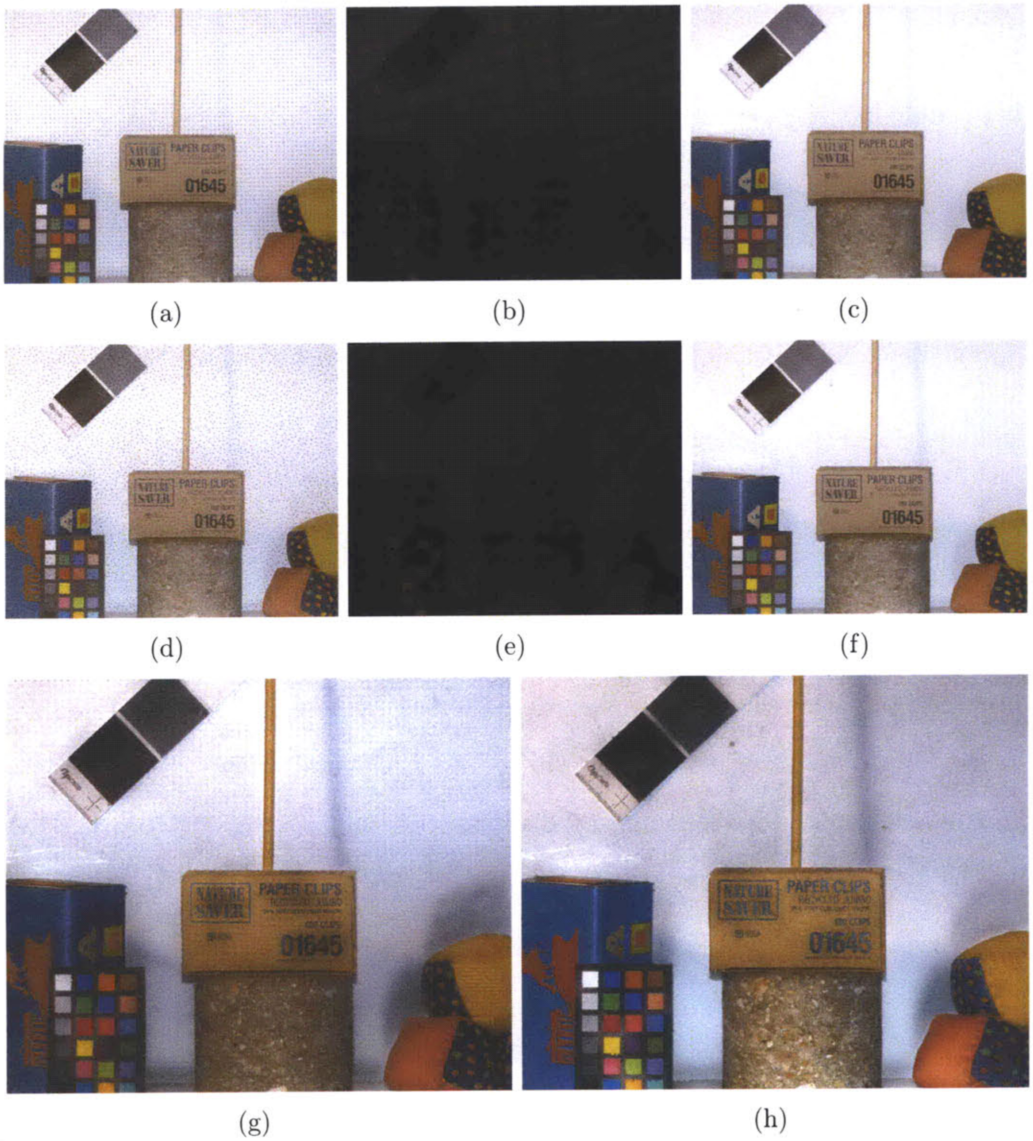
Figure	(a)	(a)	(b)	(b)	(c)	(c)	(d)	(d)	(e)	(e)	(f)	(f)
Method	Uniform	Ours	Uniform	Ours	Uniform	Ours	Uniform	Ours	Uniform	Ours	Uniform	Ours
Recall	0.9842	1.0000	0.5835	0.8869	0.7405	0.8906	0.8649	1.0000	0.6176	0.7807	0.7460	0.8984
Precision	0.9740	1.0000	0.5791	0.9225	0.7500	0.9669	0.8672	1.0000	0.5954	0.8665	0.7285	0.9628

Figure 4-9: *Results of our improved uniform and Poisson-disk localization methods compared with the uniform window method for three qualitatively different image patches: uniform depth and color (a) & (d), textured surfaces with albedo edges (b) & (e) and depth discontinuities and slanted surfaces (c) & (f). Blue boxes show hand-labeled ground truth, Green plus marks show the final result estimated by our methods, and Red exes show the estimated locations of the uniform window method. Final precision and recall values are listed in the table.*

and the locations estimated using the uniform window method. The three patches were chosen to have qualitatively different image characteristics. One patch is of a segment of a wall with approximately uniform depth and albedo. The second patch has textured surfaces and dark albedo edges, mainly caused by the printed lettering. The third patch has depth discontinuities and slanted surfaces that cause parallax effects.

We evaluated the performance of our methods by computing precision and recall statistics. In each figure, the blue boxes mark the location of the hand-labeled ground truth points, the green plus signs mark the locations of our methods estimated points, and the red exes denote the locations of the uniform window method. We removed all points within 10 pixels of the border to ignore any border effects. Our results reported for the uniform case refer to the algorithm described in section 4.4.1, and the results reported for the Poisson-disk case refer to the algorithm described in section 4.4.2. Final recall and precision results are listed in the table in figure 4-9. Our method significantly outperforms the uniform window method. The main causes of error for the uniform window method were slight shifts and fractional strides for the uniform flash pattern, and false positives due to dark albedo regions (e.g. the dark text in figures 4-9(b) and (e)). Our methods proved to be more robust to these issues by leveraging spatial information inherent in the flash patterns.

In order to perform a qualitative comparison of the uniform and Poisson-disk sampling patterns we photographed the same scene under both patterns. Figure 4-10 shows the test scene captured with uniform and Poisson-disk sampling patterns, the extracted flash and no-flash components and the fused flash / no-flash results. As expected, there are more grid artifacts in the uniform no-flash image due to undersampling, while the Poisson-disk pattern helps hide the aliasing, in exchange for increased noise. We use texture synthesis to fill in the missing flash pixels. Again, the uniform flash sampling pattern exhibits more striking and obvious artifacts due to the uniform and repetitive structure of the missing no-flash pixels. We note that the flash pattern is unintentionally slightly more defocused in the background of the uniform pattern than the Poisson-disk pattern, which decreases the contrast, creating



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 4-10: *Qualitative comparison of uniform and Poisson-disk sampling patterns. We show the same scene captured using uniform (a) and Poisson-disk (d) sampling patterns. Reconstructed uniform ambient (b), uniform flash (c), Poisson-disk ambient (e) and Poisson-disk flash (f) images. The final reconstructed outputs for uniform (g) and Poisson-disk (h). The uniform sampling pattern shows aliasing artifacts in the reconstructed ambient image (b) and texture synthesis artifacts in the reconstructed flash image (c). The Poisson-disk pattern helps hide aliasing and texture synthesis artifacts (e) & (f).*

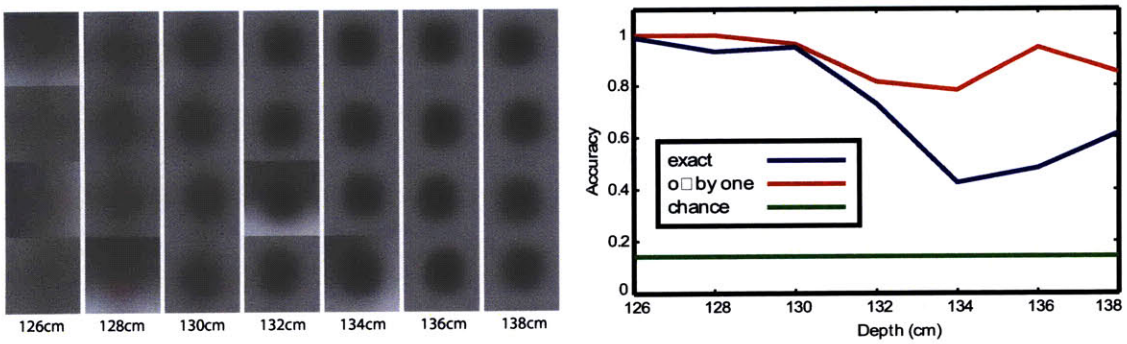
fainter shadows. The increased flash defocus causes more light to "spill" into the no-flash pixels. There is also some color casts due to the mixed lighting of the scene as well as chromatic aberration caused by the optics of the flash projector. Each final fused flash / no-flash image was white-balanced independently.

4.5 Depth from Flash Defocus

Similar to Moreno-Noguer et. al. [58], we can use the flash projector defocus to estimate a coarse depth map of the scene. However, there are several distinctions between our work and previous approaches. First, we do not assume the flash projector and the camera are coaxially aligned, and therefore must cope with parallax which makes the localization more challenging. We have described a method to improve localization in Section 4.4.1. A second fundamental difference between our setup and the one described by Moreno-Noguer and colleagues is that we aim for the flash illumination to have an infinite contrast ratio[1] between flash and no-flash pixels while they specifically illuminate the entire scene with some baseline illumination. We aim for an infinite contrast ratio because we wish to recover only no-flash illumination in the no-flash pixels. One advantage of Moreno-Noguer and colleagues approach[58] is that they are able estimate and "invert" the projector illumination blur because it is nonzero everywhere. Our goal is to estimate a sparse depth map by analyzing the blur at each no-flash pixel, and we rely on the previously mentioned methods to improve the resolution of the flash image (Section 4.4).

Defocus Patch Database Our approach is to construct a database D of examplar patches e_d that model how flash defocus changes as a function of scene depth d. In order to build our database we take multiple photographs of a planar scene containing patches with different albedos over a range of depths. The camera and flash focus remain fixed for all images, as the distance d to the planar scene is varied from d_{min}

[1] In practice this is impossible - due to indirect illumination, defocus, and the finite contrast of the occluding mask.



(a) (b)

Figure 4-11: *a) A patch database for seven depths ranging from 126cm to 138cm in 2cm increments. Each depth has $K = 4$ exemplar patches. b) Error plot of our depth estimation method. The blue curve shows the percentage of points assigned the correct depth label as a function of depth. The red curve shows the percentage of points assigned the correct depth label, or a label ±1 from the correct label. In this case a mislabeling by 1 corresponds to a 2cm error in depth estimation. The green curve shows the performance of assigning depth labels at random.*

to d_{max} producing a stack of images $\{I_d\}$. We used a relatively small aperture for the camera ($f/10$) and a large aperture for the flash projector ($f/2.8$) to ensure that most of the observed defocus is due to the flash and not the camera. From each image I_d we estimate the no-flash pixel locations and crop a $N \times N$ window around each no-flash pixel creating a large collection of example patches for each depth. We use k-means clustering to compute K examplar patches $e_d^k, k = 1 \ldots K$ for each depth d, and the set of all these examplars over all depths forms our database $D = \left\{ e_d^k | k = 1 \ldots K, d \in [d_{min}, d_{max}] \right\}$. In order to provide albedo invariance, we independantly normalize each color channel of e_d^k to have unit mean. Figure 4-11 shows a database of patches for 7 depth values ranging in 2cm increments from 126cm to 138cm. For each depth d we have computed $K = 4$ exemplar patches.

Estimating Depth Given a new scene, we would like to estimate depth at each no-flash location p. Let l_p be the $N \times N$ window of pixels centered at p, and $\hat{\mu_p}$ be the per-color channel (i.e. RGB) mean of l_p. We compute the error $E(l_p, d)$ for depth

d as:

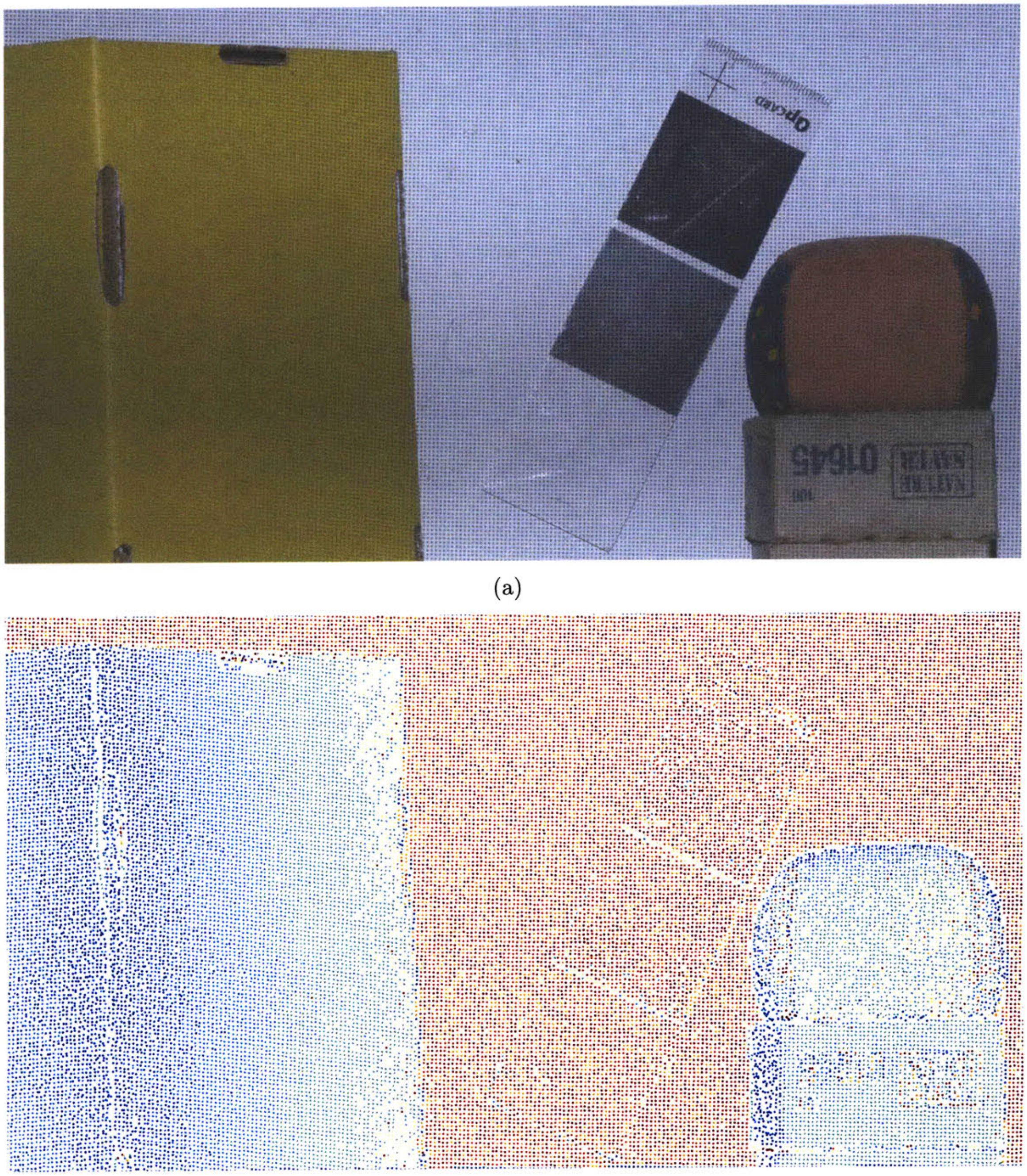
$$E(l_p, d) = \min_{k=1\ldots K} \left\| l_p - \hat{\mu_p} \cdot e_d^k \right\|^2 . \tag{4.8}$$

We rescale the examplar patches e_d^k by the RGB means $\hat{\mu_p}$, instead of normalizing l_p to unit means per channel in order to avoid amplifying noise in l_p. For example a blue object may have a very low red channel, and thus normalizing the red channel to unit mean would amplify any noise present. Conversely, weighting e_d^k by $\hat{\mu_p}$ will downweight the importance of the red channel when computing the error. We use nearest neighbor classification and select the d^* that minimizes $E(l_p, d)$ as the depth at pixel p:

$$d^* = \arg\min_d E(l_p, d) \tag{4.9}$$

Figure 4-11 shows an error plot of the number of correctly classified points as a function of depth, for a set of seven images of a planar scene, covering the depth range 126cm to 138cm in 2cm increments, using nearest neighbor classification. The seven test images were the same images used to create the patch database. Each test image contained approximately 8300 no-flash pixels. The y-axis of the plot shows the percentage of points correctly label as a function of depth. On the low end, points at 134cm were correctly identified 42% of the time, whereas on the high end, points at 126cm were correctly identified 98% of the time. Chance would correctly label points 14% of the time. In addition, the curve marked "off by one" shows the percentage of points that were assigned a depth label off by at most one from the correct label (corresponding to a depth error of 2cm in our experiment). This improves the percentage to greater than 78% of points.

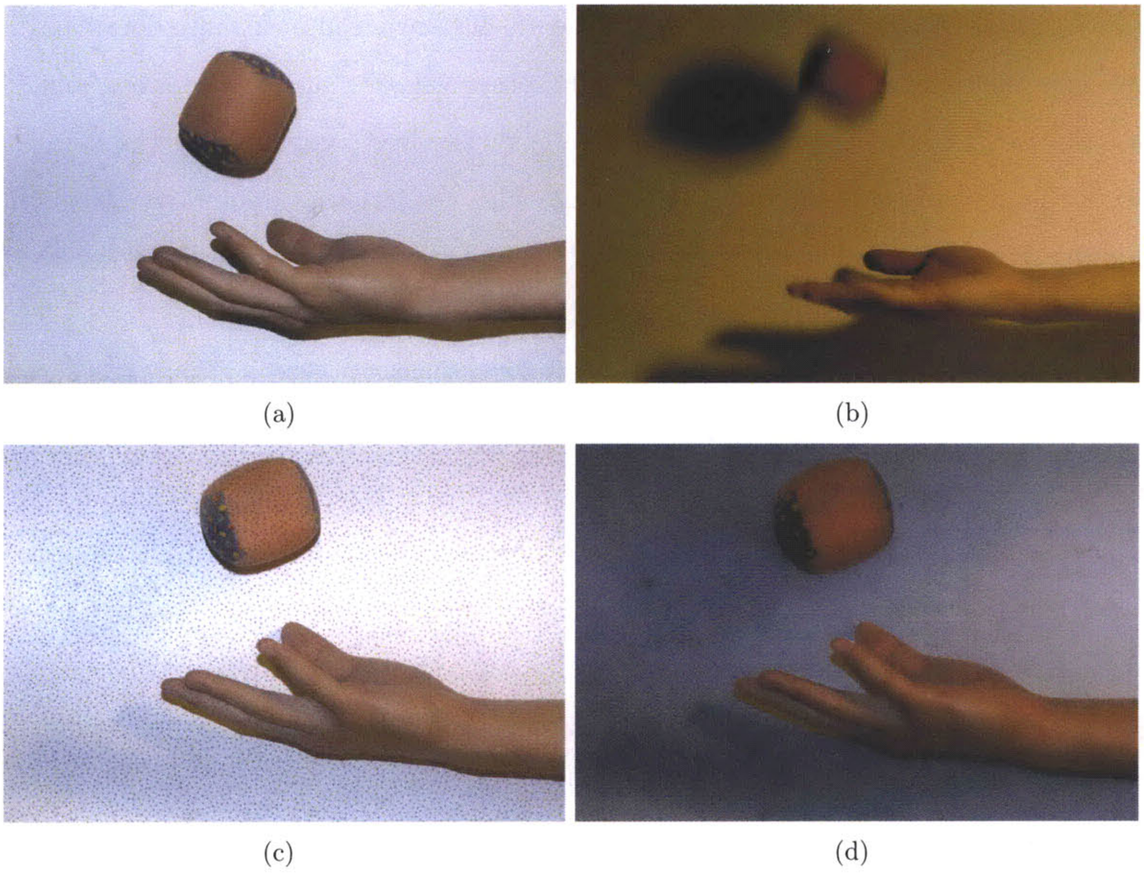
Figure 4-12 shows results for a scene with depth variation over the full working range. The yellow box on the left is slanted away and our depth map reflects this. Also note the bean bag and brown box are estimated at the same depth, as are the different segments of the gray card, disregarding the significant difference in albedos.



(a)

(b)

Figure 4-12: *a) Multiplexed flash illumination input image. b) Sparse depth map computed at each no-flash pixel. Blue values are closer to the camera. Red values are further away.*



(a)

(b)

(c)

(d)

Figure 4-13: *The motion in a dynamic scene is frozen with standard flash phography(a) but the soft ambient light is lost. Two image flash/no-flash cannot be used because the no-flash image(b) has changed and is blurry. From our multiplexed illumination image(c) we can create a new image that freezes the motion and retains the character of the ambient lighting.*

4.6 Single Exposure Flash / No-Flash

To demonstrate our multiplexed illumination, we show single exposure flash/no-flash image fusion on a dynamic scene. Traditional flash/no-flash methods[21, 70] take as input a flash and a no-flash image of the same scene. These methods assume there is minimal motion between flash and no-flash images (such that a simple alignment will produce pixel level correspondence). Next, the images are decomposed into detail and large-scale layers using the bilateral filter (and other variants such as the cross/joint bilateral filter [21, 70]). Finally a new image is synthesized by combining the detail layer of the well exposed, low noise flash image with the large-scale intensity layer of the under exposed and noisy flash image. In essence, this combines the sharp details of the flash image with the pleasing ambient lighting of the no-flash image.

Scenes with motion pose a problem for traditional flash/no-flash methods because it is no longer possible to align objects between exposures. Using our flash design, we are able to capture enough information in a single image to perform a flash/no-flash image fusion. Figure 4-13(a) shows a person tossing a bean bag, captured using a standard flash in order to freeze the motion of the object. Figure 4-13(b) shows a no-flash image taken of the same scene shortly aftwards. Objects have changed position in the no-flash image, and there is a large amount of motion blur. Figure 4-13(c) shows the results of performing flash/no-flash fusion using the components captured from a single image. In this example, we used Poisson-disk distributed no-flash points and reconstructed the flash image using texture synthesis to fill in missing data. Our result has the sharpness of the flash image, as well as the shadowing and glow of the no-flash image.

4.7 Discussion

Flash multiplexing shows promise for computational illumination in dynamic scenes because it facilitates the simultaneous capture of multiple components of illumination. In this chapter we demonstrated two possible applications of flash multiplexing:

flash/no-flash for dynamic scenes and sparse depth estimation.

However, illumination multiplexing does raise some challenging issues. A limitation of our method is the assumption that no-flash pixels capture only ambient lighting. In practice, these pixels are illuminated not only by the ambient lighting, but also by the indirect light from the flash. Additionally, there is some light spill due to defocus of nearby flash pixels and the finite contrast of the transparency mask used to create our sampling pattern. These two drawbacks limit the ability of our system to cleanly separate flash and ambient illumination.

This leakage of light due to defocus is further confounded by the effects of chromatic aberration inherent in the refractive elements used to focus the flash illumination. The chromatic aberration of the flash optics, coupled with the high-frequency occluder mask used to block the flash, combine to create a depth-dependent chromatic shift of the flash illumination that leaks into the no-flash pixels due to defocus. In effect, no-flash pixels that image points in front of the plane of focus are corrupted by light of one type of illuminant, while points behind the plane of focus are corrupted by different illuminant. While this depth dependence makes it challenging to remove the effects of the chromatic shift in the illumination, it also provides some benefit to our depth from defocus method. Unlike traditional depth from defocus methods, we are able to disambiguate points in front of versus behind the plane of focus.

Continued exploration into ways to further separate the recovered no-flash image into true ambient and flash indirect lighting would be useful. Recent work on multi-light white balance [38] may help accomplish this separation. Currently, we use texture synthesis to improve the resolution of the flash image. However texture synthesis is computationally expensive when running on large images, so research on other local methods to improve resolution would benefit the system.

Chapter 5

Spatio-Temporal Multiplexing of Flash and Ambient Illumination

5.1 Introduction

In this thesis we have investigated methods to simultaneously capture multiple photographic settings (e.g. aperture, flash) in a single exposure. In the previous chapter we proposed a method to spatially multiplex flash and ambient illumination conditions by projecting a coded flash onto the scene. This approach had the appealing property that it did not require any changes to the camera, and instead augmented the flash with simple optics and a coded mask pattern. However, performing illumination multiplexing on the flash made demultiplexing the coded information difficult. Additionally, the reconstructed no-flash signal was subject to indirect illumination leaking into it, thus a true "no-flash" image was impossible to capture.

In this chapter, we again have the goal to capture a scene under multiple illumination conditions in a single exposure. However, instead of modifying the illumination, as described in the previous chapter, or by other means such as spectral or polarization filters, we propose to modify the camera sensor itself. Although redesigning a camera sensor would require a significant investment from a manufacturer, and precludes use on current hardware, the modifications alleviate the main issues of the previous method. Specifically, because we do the multiplexing on the sensor, localization becomes trivial. Furthermore, our proposed method does not suffer from the limitation that no-flash pixels are corrupted by indirect flash illumination. Unfortunately, like the method in the previous chapter, simultaneous capture of multiple illumination conditions does come at a price: we must sacrifice image resolution to record the additional information. Single exposure methods are advantageous over multiple exposure methods because they eliminate the possibility for misalignment between exposures. Perhaps more importantly, single exposure methods save significant data bandwidth, which is a precious system resource for modern high-resolution image sensors. In particular, many applications do not require full resolution ambient and flash images, and thus the extra storage and bandwidth necessary for the high-speed capture of multiple exposures is overly costly and unnecessary.

This chapter introduces spatio-temporal multiplexing of flash and ambient illumination in a single exposure. In order to achieve a spatial multiplexing of flash and ambient illumination (i.e. pixels capture flash or ambient information, spatially mosaiced across the sensor), we temporally divide the integration time of each pixel on the sensor. Our idea is inspired by the slow-sync photography method (see figure 5-1), where a photographic flash is combined with a long exposure in order to simultaneously expose a foreground object (exposed primarily with flash lighting) and a more distant background (primarily with ambient illumination and a long exposure). We propose to augment each pixel with independent control of its integration time (i.e. a per-pixel digital shutter), in order to select whether it integrates primarily flash photons during the brief flash burst, *or* it begins integrating after the flash has fired and thus collects only ambient light. We have not constructed a physical prototype implementing per-pixel shuttering, nor do we present a detailed hardware design. Instead we focus on the high-level functional behavior of the sensor and simulate our proposed changes using multiple exposures. We describe two possible applications of spatio-temporal multiplexing: single-image flash / no-flash fusion and white balancing scenes containing two distinct illuminants (e.g. flash and fluorescent lighting).

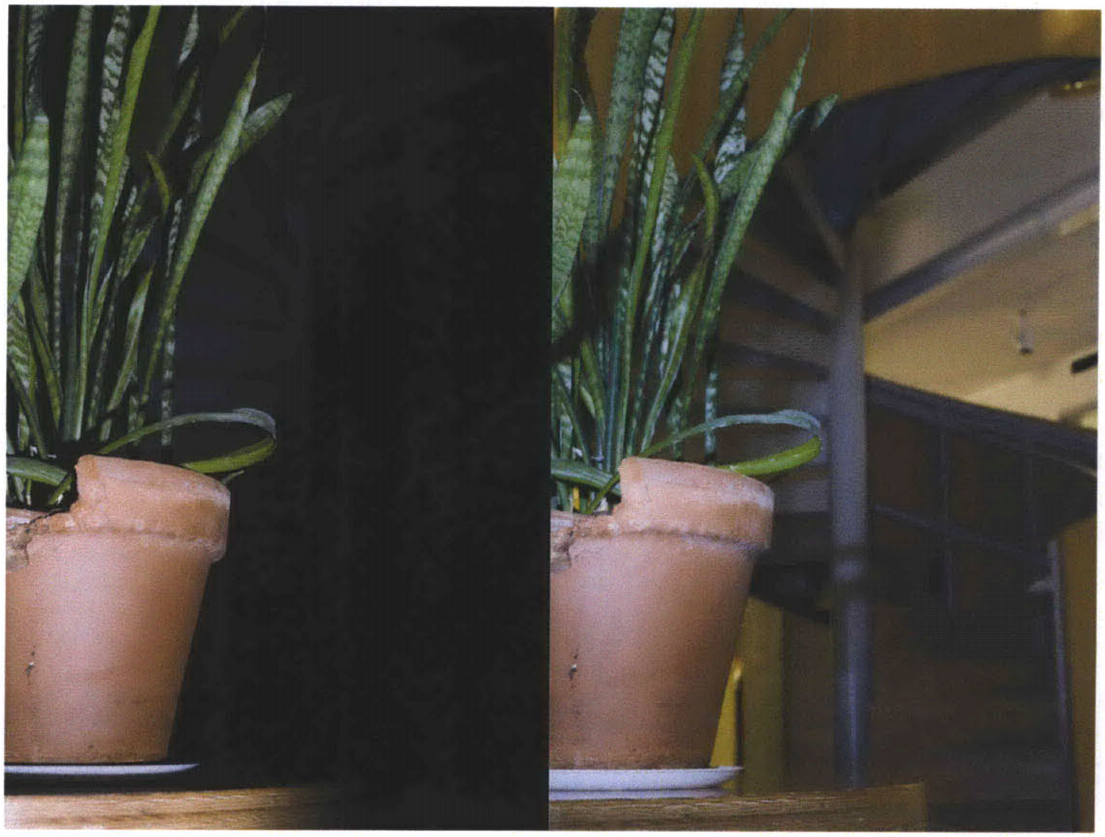


Figure 5-1: *Slow sync flash photography. Both images are captured with the same flash, aperture and ISO settings, however the image on the left is taken with a short exposure (1/250 sec) while the image on the right is taken with a long exposure time (1/2 sec). Flash intensity falls-off quadratically with depth, thus the distant background pixels to do not receive much flash light. A long exposure time allows the background to receive enough ambient lighting to be well-exposed.*

5.2 Related Work

The Assorted Pixels work proposed by Nayar and Narasimhan[63], introduced a systematic method for sampling multiple dimensions of imaging (e.g. brightness, color spectrum, polarization, etc.) by mosaicing pixels across the image sensor. In this work we consider a new sampling dimension, where the integration timing of particular pixels is synchronized to the external illumination conditions (i.e. flash).

Acosta-Serafini and colleagues [1] introduced a predictive multiple sampling method that adaptively adjusts the pixel dynamic range based on the intensity of the incident lighting. Essentially, each pixel's current value is nondestructively queried several times during the integration period to check if it is in danger of saturating. If it is expected to saturate (assuming a linear extrapolation of the current incident intensity) then the pixel's value is reset, and it begins integrating again until the full exposure time has elapsed. The ending pixel intensity, along with the number of times the pixel was reset, is used to estimate the final pixel value. We view this work as evidence that our proposed hardware modifications could indeed be physically realized using existing CMOS technology.

Raskar et. al. [71] describe a coded exposure method for motion deblurring using a so-called "fluttered shutter". A very high-speed ferro-electric shutter is "fluttered" (opened and closed in a binary psuedo-random sequence) during exposure, causing the PSF to preserve high-frequency spatial details in images with motion blur, and thus make deconvolution well-conditioned. While the authors used a high-speed physical shutter, the method could be extended to use an electronic shutter. In our work, we use a per-pixel shutter, as opposed to a global shutter which effects all pixels uniformly.

Many methods have been proposed that combine multiple exposures to produce improved photographs. High dynamic range capture [16] methods combine several images, each taken with different exposure settings, to create a single image with the full dynamic range of the scene. Tone mapping[75, 19] and exposure fusion [55, 11] take a stack of images and produce a single image suitable for a low dynamic range display. Specifically for low light situations, the flash / no-flash methods [21, 70, 37] combine a flash and no-flash image to produce a new image with the large-scale lighting properties of the no-flash image and the details of the flash image. These previous methods all required multiple exposures, while our goal is to capture enough information in a single image.

Flash photography is often the best option for low-light situations. However, when there are large depth variations in the scene it is difficult to evenly light a scene with a single flash due to the quadratic fall-off of flash intensity with distance. This produces well exposed foreground subjects with nearly black backgrounds. Slow-sync photography is a common technique used by photographers to alleviate this short-comming of standard fast-exposure flash photography (see figure 5-1). In slow-sync photography, the flash is used in tandem with a long exposure, to produce an image with a well lit foreground (from the flash) and a well lit background (from the ambient lighting and a long exposure). We expand the idea of slow-sync photography, by synchronizing the pixel exposures to capture each component separately.

We demonstrate two applications of our spatio-temporal multiplexing method: single-exposure flash / no-flash fusion and white-balancing under mixed illumination. Hsu and colleagues [38] introduced a method for performing white-balancing under mixed illumination. Their method uses an albedo-voting scheme to decompose the image into two components, and performs white-balancing on each component independently assuming known illuminant spectra. Our method allows us to decomposed the image into two components directly without needing to estimate albedos, however we trade spatial resolution in order to capture two illumination conditions (flash and ambient).

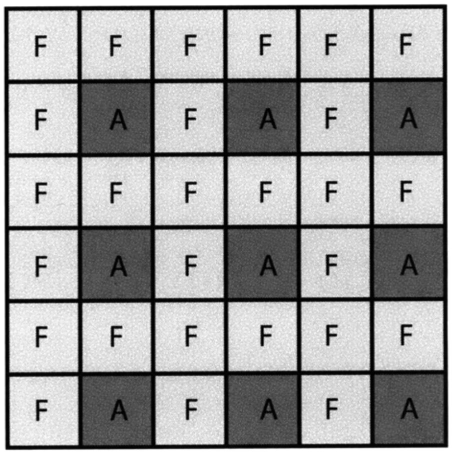
5.3 Spatio-Temporal Multiplexing

In this section we describe spatio-temporal multiplexing for simultaneously capturing a scene under flash and ambient illumination. We call our method spatio-temporal multiplexing because we rely on both spatial and temporal multiplexing. We use spatial multiplexing in the sense that different pixels on the imager will record either flash or ambient information (see figure 5-2), in a fashion similar to spectral multiplexing (i.e. traditional RGB spectral filters). We exploit the fact that a photographic flash produces a nearly instantaneous burst of light (typically lasting for one thousandth of a second) to temporally divide the pixel integration period into flash and ambient segments (see figure 5-3). By synchronizing a per-pixel digital shutter with the firing of the flash, we can configure each pixel to capture either flash only[1], ambient only, or combined flash and ambient information. A per-pixel digital shutter

[1] While technically both flash and ambient light would be captured, the magnitude of the ambient contribution typically is vanishingly small in comparison to the flash when using the short exposure time required.

can be implemented using current CMOS sensor technology, which allows transistors to be located at each photodetector (i.e. pixel). These transistors can be used to add per-pixel control logic to the sensor.

In this, and previous chapters, we argue in favor of trading spatial resolution to capture extended information. However, whenever making tradeoffs, it is always important to analyze and compare the proposed method to alternative approaches. In this case, the simplest approach is to capture multiple illumination conditions in separate exposures. In particular, it is possible to design a camera that captures two images, one with the flash, and one without, in very rapid succession. By far, the most challenging (and most costly) aspect of a high speed multi-capture camera is handling the high-bandwidth data transfer necessary to record two full resolution images rapidly. In section 5.3.1 below, we argue that for some applications (e.g. flash / no-flash image fusion) the no-flash component can be significantly lower resolution than the flash component. So, in a sense, the multi-capture camera design is "wasting" a large amount of precious bandwidth by capturing and transferring excessive ambient image pixel data. Another issue in designing a rapid multi-capture camera is creating a high-speed shutter capable of triggering twice with very little delay between exposures. Currently, most SLR cameras use mechanical two-curtain focal-plane shutters that are not as responsive as electronic shutters. Reducing delay is particularly important for capturing moving scenes and section 5.3.2 contains a more in-depth analysis of the effect of delay on reconstructed image quality.



F	F	F	F	F	F
F	A	F	A	F	A
F	F	F	F	F	F
F	A	F	A	F	A
F	F	F	F	F	F
F	A	F	A	F	A

Figure 5-2: *Spatially multiplexing Flash (F) and Ambient (A) information on the image sensor. We show a 1:4 sampling ratio of ambient to flash pixels.*

Spatial Multiplexing We propose spatially interleaving, or mosaicing, pixels across the image sensor that record either flash or ambient illumination (see figure 5-2). The main parameter to explore when spatially multiplexing is the relative sampling rates of flash and ambient pixels. We are argue that because flash pixels will in general have higher signal-to-noise ratios (because they are typically well-exposed, while ambient pixels will be under-exposed in a low-light scene), they provide higher "quality" high-frequency data and thus should be sampled higher. This argument is supported by flash / no-flash fusion methods [21, 70] which combine the detail from a flash image with the large-scale illumination from an ambient image. By definition the detail-layer contains high-frequency information, which should be sampled at a high rate in order to avoid aliasing artifacts. Furthermore, these methods have shown that high-frequency content in the large-scale layer (i.e. strong edges) is strongly correlated with the flash image, and can be inferred using joint bilateral filtering methods.

Temporal Multiplexing In order to perform temporal multiplexing the timing of the flash and the per-pixel electronic shutters must be synchronized. For simplicity of exposition, let us assume that we are using front-curtain flash firing. If the electronic shutter for pixel p "opens" slightly before the flash fires, and then "closes" just afterwards, then p will capture only flash light. If on the other hand, p "opens" just after then flash fires and remains open for the remainder of the global shutter time, then p will capture only ambient lighting. The third scenario is that p is open for the entire global shutter time and thus collects both flash and ambient lighting, which describes traditional slow-sync photography when a long shutter time is used. Although our goal is to capture flash and ambient lighting separately, which suggests using the first two temporal sampling patterns, it may in fact be beneficial to capture a combined flash and ambient image instead of a flash only image. The primary argument for capturing a combined flash-ambient image is exactly the argument for traditional slow-sync photography: the quadratic intensity fall-off with depth may leave the background image under-exposed, thus there will be very little useful information in a flash-only image. In this scenario, and with the assumption

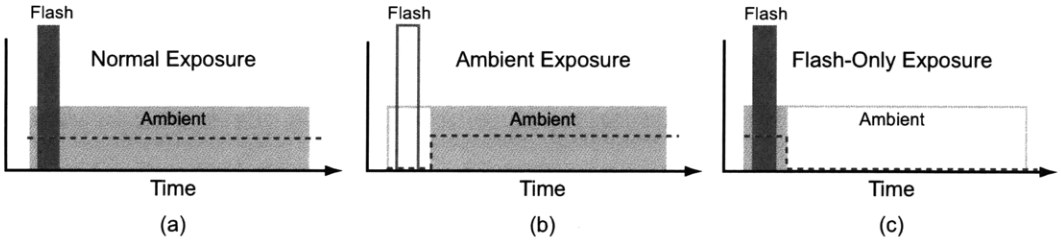


Figure 5-3: *(a) Normal exposure. (b) Ambient exposure. (c) Flash-only exposure. The blue dotted line depicts when the per-pixel shutter is "open". During a normal exposure (a) the shutter is open the entire exposure and the pixel captures both flash and ambient lighting. The flash lighting is brief but high intensity. The Ambient lighting is constant but lasts the entire exposure time. The per-pixel shutter "opens" just after the flash has fired for Ambient pixels (c) and thus only records ambient lighting. The shutter "closes" just after the flash fires for Flash-only pixels, thus recording primarily flash lighting.*

that the flash (or flash + ambient) pixels are spatially sampled at a higher resolution than the ambient-only pixels, a combined flash and ambient image will have a higher signal-to-noise ratio (albeit the "signal" will be from the ambient lighting).

If each per-pixel electronic shutter is programable then the sensor could be dynamically reconfigured to operate in either flash + ambient or flash only mode. Furthermore, the relative sampling rates of flash and ambient pixels could be changed to suit the application and scene. Finally, the sensor could even be configured to operate as a standard sensor, without any temporal or spatial multiplexing.

Spatio-Temporal Multiplexing Figure 5-4 shows a flowchart of our proposed method. The spatio-temporal multiplexed input is recorded on the sensor. Next, the individual components are demultiplexed into separate images. The flash image, which is captured at a high resolution, is only "missing" a sparse set of pixels. We use simple linear interpolation to fill in the missing pixel data; other interpolation and demosaicing methods could also be used. The ambient component is captured at a much lower resolution and thus needs to be upsampled to the resolution of the flash image. Linear, or higher-order reconstruction filters can be used to perform upsam-

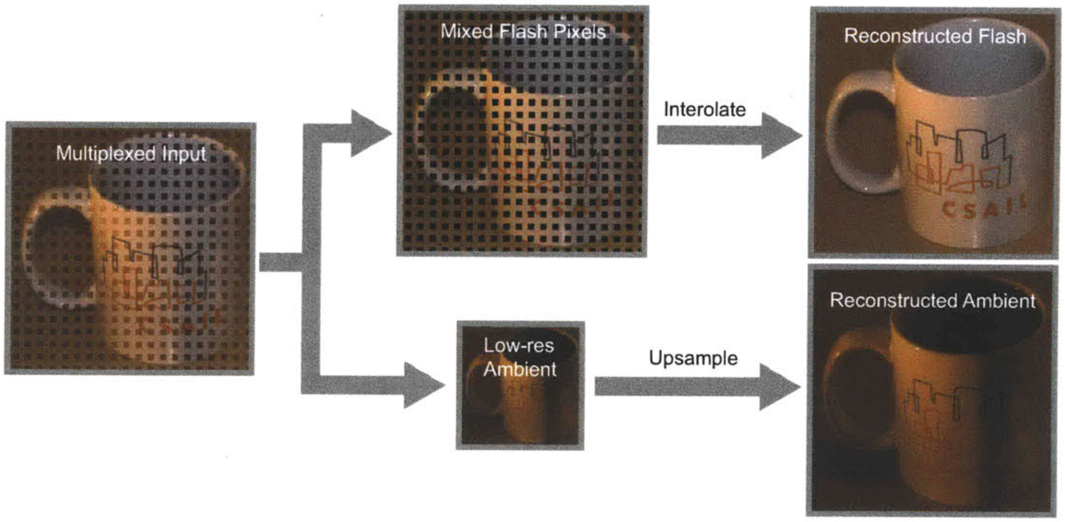


Figure 5-4: *Flowchart of the demultiplexing pipeline. The multiplexed input image contains flash and ambient data at asymmetric sampling rates. We extract the different components into separate images. The missing data from the flash image is reconstructed using linear interpolation. The ambient image is upsampled to the same resolution as the flash image using Joint Bilateral Grid Upsampling [14, 42].*

pling. Alternatively, we can leverage the fact that we already have a high resolution flash image, and perform Joint Bilateral Grid Upsampling [14, 42] to upsample the image. Joint Bilateral Grid Upsampling is a non-linear upsampling method that uses a high-resolution image to guide the upsampling of the low-resolution image by respecting edges in the high-resolution image. Joint Bilateral Grid Upsampling is most useful when the ambient image is highly sub-sampled.

Simulating Spatio-Temporal Multiplexing Hardware We are proposing a change to the hardware of the image sensor. Implementing a hardware version is expensive, and instead we simulate our proposed implementation by combining multiple exposures. We take two exposures and then digitally perform multiplexing and demultiplexing. Multiplexing is accomplished by directly subsampling the individual exposures (without any pre-filtering) at the appropriate relative sampling rates, and then compositing the pixels into a single image. We assume each pixel records RGB data (e.g. a sensor comparable to the Foveon X3 sensor [25]) and ignore demosaicing issues. Demultiplexing is described in the previous paragraph above and in figure 5-4.

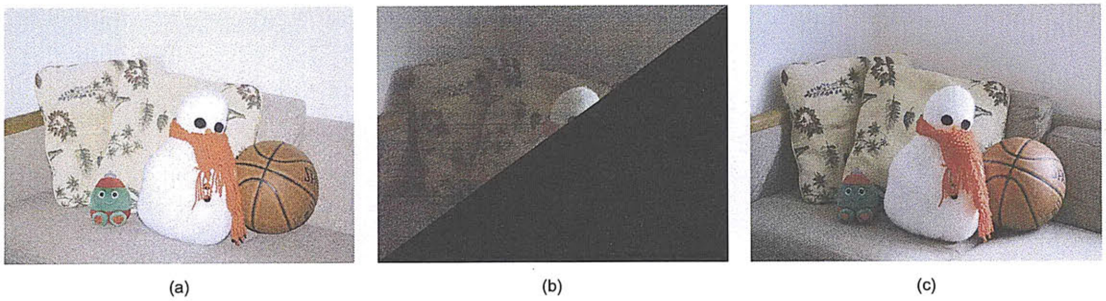


Figure 5-5: *An example scene captured with (a) and without (b) & (c) flash illumination. The flash image (a) is evenly exposed and has a high signal-to-noise ratio, but lacks the aesthetically pleasing shadowing present in a long exposure ambient-only exposure (c). A short exposure (same shutter speed as (a)) image is shown in (b). The upper half of (b) has been contrast stretch to show details, while the lower half is displayed as captured from the camera.*

5.3.1 Exposure-Resolution Analysis

In general we advocate that the flash pixels be allocated a large portion of the sensor resolution, and that the exposure be set such that flash pixels are well-exposed. As a consequence, reconstructed images will often be under-sampled and under-exposed. In this section we perform an experimental analysis of the effect of under-exposure and under-sampling on the peak signal-to-noise ratio (PSNR) of the reconstructed ambient image. In our experiment we captured a sequence of images of a scene lit under ambient illumination only, varying the exposure time between each exposure and keeping all other camera settings constant (e.g. aperture f/# and ISO). We adjusted the exposure times to bracket a seven stop range of exposure values (EV). Figure 5-5 shows a properly exposed scene as well as a the same scene severely under-exposed (-7EV). The upper triangular section of the under-exposed image has been contrast stretch to show the signal, while the lower triangle is displayed unprocessed. Additionally, we create a sequence of down-sampled images (using nearest-neighbor sampling) for each exposure. Table 5.1 shows the PSNR results for reconstructing the ambient image under each combination of under-sampling and under-exposure parameters. We treat the full-resolution well-exposed image as ground truth when computing PSNR values. Table 5.1 shows the average PSNR values computed for

		Relative Resolution						
	PSNR (dB)	1:1	1:2^2	1:4^2	1:8^2	1:16^2	1:32^2	1:64^2
Relative EV	0EV	∞	46.96	39.23	34.48	30.87	27.81	25.55
	-1EV	40.92	40.16	37.53	33.95	30.62	27.64	25.42
	-2EV	38.78	37.90	36.00	33.13	30.22	27.45	25.33
	-3EV	34.39	33.86	32.80	31.05	29.02	26.83	25.04
	-4EV	30.67	30.34	30.00	29.03	27.65	25.83	24.43
	-5EV	27.61	27.26	27.03	26.32	25.25	23.99	22.97
	-6EV	22.89	22.51	22.39	22.11	21.66	20.96	20.34
	-7EV	17.81	17.33	17.28	17.07	16.89	16.48	16.31

Table 5.1: *Effect of exposure and resolution on PSNR for the reconstructed ambient image. The horizontal axis shows the relative sampling of the input resolution to the output resolution, from full resolution to one input pixel to every* $64^2 = 4096$ *output pixels. The vertical axis shows the exposure value (EV) relative to the properly exposed ground truth image.*

three different scenes. We can use the table to determine acceptable down-sampling and under-exposure values. Examining under-exposure levels is useful because it can help in setting camera settings (such as ISO, aperture f/#, and shutter speed) to ensure that the ambient image is captured within a tolerable exposure range. For example, a threshold value of 30dB (a typical value used for image compression applications) suggests we can under-expose by 4 stops and use a 1:16 down-sampling ratio and still expect a reasonable reconstruction.

For the specific application of flash/no-flash fusion we are more concerned about a bilateral filtered version of the ambient image than the ambient image itself. In this case, we can measure the reconstructed PSNR of the bilateral filtered image instead. We use joint bilateral bilateral filtering, with a flash image of the scene as the guide image. Table 5.2 shows average PSNR results for the same three scenes used to compute Table 5.1. The obvious conclusion from Table 5.2 is that we can under-sample and under-expose more aggressively for the same PSNR, or conversely, expect better reconstruction results for the same level of under-sampling and under-exposure.

		Relative Resolution						
	PSNR (dB)	1:1	1:2^2	1:4^2	1:8^2	1:16^2	1:32^2	1:64^2
Relative EV	0EV	∞	65.72	57.75	49.09	41.35	34.78	29.86
	-1EV	48.17	48.38	48.15	45.87	40.78	34.73	29.83
	-2EV	50.22	50.11	48.77	45.26	40.40	34.59	29.72
	-3EV	39.25	39.28	39.07	38.31	36.28	32.75	29.04
	-4EV	35.64	35.66	36.00	35.59	34.23	31.42	28.39
	-5EV	37.49	37.63	37.83	37.11	34.87	31.48	27.72
	-6EV	33.15	33.47	33.22	32.56	31.31	28.89	25.37
	-7EV	29.04	29.55	29.04	28.64	27.50	24.97	21.33

Table 5.2: *Exposure and resolution analysis for the bilateral filtered large scale ambient layer. PSNR values for reconstructed bilateral filtered ambient images are significantly higher than for the full ambient image (see table 5.1).*

5.3.2 Motion Analysis

In this section we analyze the effect of scene motion on reconstructing a bilateral filtered version of the no-flash component. We also compare single-exposure capture with multi-exposure capture, where there is a nonzero delay between capturing each image. If the scene is moving this delay will cause the flash and ambient images to be misaligned. We synthetically apply a motion to the scene and measure the PSNR of the reconstructed bilateral filtered image. We model motion blur as a convolution of a static image I with a kernel $\Phi_{v,s,d}$:

$$I_m = I * \Phi_{v,s,d}, \tag{5.1}$$

where $\Phi_{v,s,d}$ encodes the velocity v (in pixels/second), camera shutter speed s (in seconds) and the delay d (in seconds) between the flash exposure. In particular, $d = 0$ if the flash and ambient images are recorded in the same exposure. We define $\Phi_{v,s,d}$ as

$$\Phi_{v,s,d}(x) = \frac{1}{vs} Rect\left(\frac{x - vd}{vs} - 1/2\right) \tag{5.2}$$

where *Rect* is the unit boxcar function. The vs term determines the width of the boxcar function, which directly corresponds to the distance the scene moves during the integration interval. Finally, the vd term causes a shift of the image, corresponding to the number of pixels the scene moves between exposures.

Tables 5.3 and 5.4 list the PSNR results for low scene motion ($v = 100$ pixels/second) and high scene motion ($v = 1000$ pixels/second) respectively. As expected, a non-zero delay between exposures ($d > 0$) has only a moderate impact on PSNR in the low motion case. The data also emphasizes the effect of the shutter speed s in the trade-off between improved signal-to-noise ratio due to longer exposure and increased motion blur. A long exposure (e.g. 1/4 of a second) properly exposes the ambient image but introduces a non-trivial amount of motion blur (≈ 25 pixels). A short exposure (e.g. 1/250 of a second) grossly under-exposes the image, but minimizes the motion blur (< 1 pixel). In these experiments, the optimal shutter speed was $s = 1/15$ sec, where the image was reasonably exposed and subject to only a small amount of motion blur (≈ 7 pixels).

		Delay (sec)						
	PSNR (dB)	0	1/60	1/30	1/15	1/10	1/5	1/2
Shutter (sec)	1/4	53.14	50.25	48.10	46.13	43.49	39.45	33.37
	1/8	57.69	54.71	52.38	49.17	44.71	38.87	32.57
	1/15	58.27	56.44	54.06	49.80	44.11	37.92	32.26
	1/30	49.24	49.39	48.86	46.54	42.35	37.42	32.31
	1/60	46.49	46.69	46.55	44.44	41.01	36.88	32.25
	1/125	38.17	38.70	38.77	38.29	36.81	34.58	31.45
	1/250	32.27	32.89	32.98	32.99	32.53	31.62	29.91

Table 5.3: *The effect of shutter speed and delay on reconstructed ambient images. Table shows average PSNR results for a set of low motion scenes ($v = 100$ pixels/second).*

The high motion case (table 5.4) underscores the importance of single-exposure capture of both ambient and flash images. Unlike the low-motion case, a non-zero delay between exposures ($d > 0$) has a much more dramatic effect on the final PSNR. For example, a drop of nearly 20dB is observed for even a 1/10 of a second delay. Current top of the line DSLR cameras have 10 fps (i.e. 1/10 of a second delay) continuous shooting modes. Although it is likely that the speed of full resolution continuous shooting modes will increase (perhaps to 60 fps or higher), capturing two images will always require twice the bandwidth. We have previously argued that

		Delay (sec)						
	PSNR (dB)	0	1/60	1/30	1/15	1/10	1/5	1/2
Shutter (sec)	1/4	34.22	33.03	32.03	31.15	30.09	28.32	25.31
	1/8	39.35	37.27	35.42	32.91	30.09	27.06	24.44
	1/15	44.32	41.55	38.00	33.04	29.09	26.34	24.06
	1/30	46.09	43.98	38.20	32.02	28.57	26.25	24.02
	1/60	46.36	43.62	36.86	31.46	28.44	26.35	24.14
	1/125	39.09	38.46	34.68	30.91	28.50	26.66	24.53
	1/250	33.11	33.10	31.70	29.60	28.02	26.72	24.87

Table 5.4: *The effect of shutter speed and delay on reconstructed ambient images. Table shows average PSNR results for a set of high motion scenes ($v = 1000$ pixels/second).*

because the sampling requirements of the flash and ambient images are asymmetric, it is not necessary to capture two full resolution images, and thus our single-exposure design may actually be cheaper to implement and manufacture due to the bandwidth savings.

5.4 Applications

We present two applications of our spatio-temporal multiplexing method: single exposure flash / no-flash image fusion and white-balancing with mixed illumination.

5.4.1 Flash / No-Flash

To apply flash / no-flash image fusion we capture two images and synthetically multiplex them into a single image as described in section 5.3. In particular, both flash and ambient images are taken with exactly the same camera settings to simulate simultaneous capture. We use a 1:4 sampling ratio of flash to ambient pixels (i.e. the ambient image is subsampled by two in each dimension, using nearest neighbor sampling). We perform demultiplexing by linearly interpolating the missing flash pixels and by using Joint Bilateral Grid Upsampling [14, 42] to upsample the ambient image to the same resolution as the flash image. The ambient image is contrast stretched and then the detail and color layers of the flash image are combined with the large-scale layer of

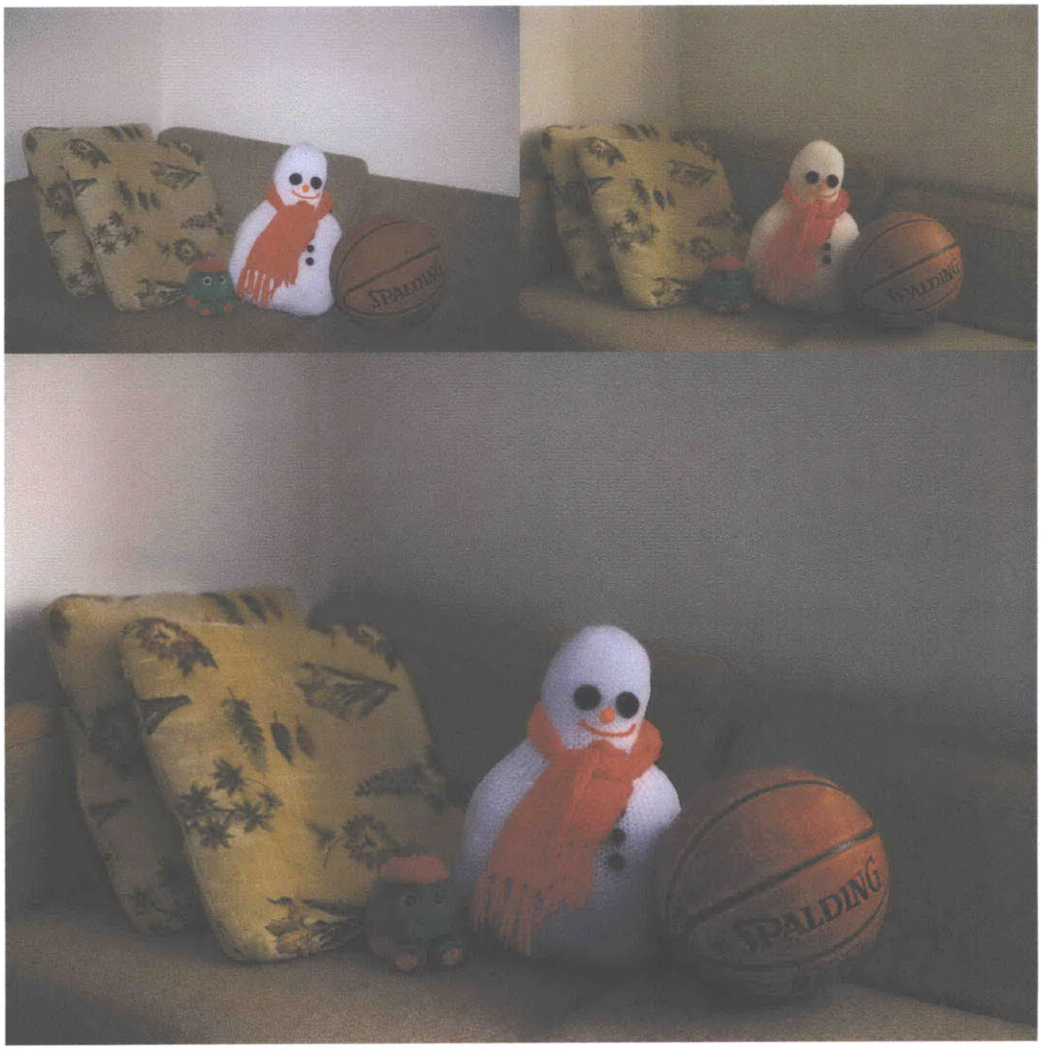


Figure 5-6: *Flash / no-flash image fusion. The reconstructed flash and no-flash (contrast stretched) images are shown on top and the final result is shown below.*

the ambient image[21]. Figures 5-6 and 5-7 show some results of our method. These results validate the claim that a low resolution ambient image is sufficient to capture the large-scale lighting of the scene. We are able to successfully extract the shading and mood of the noisy ambient image and combine it with the high-quality details of the flash image. One challenge of capturing flash and ambient images simultaneously is that we must capture both the bright flash and dim ambient lighting with the same exposure settings. In practice, we exposure for the flash image and allow the ambient image to be drastically under-exposed. These images demonstrate that

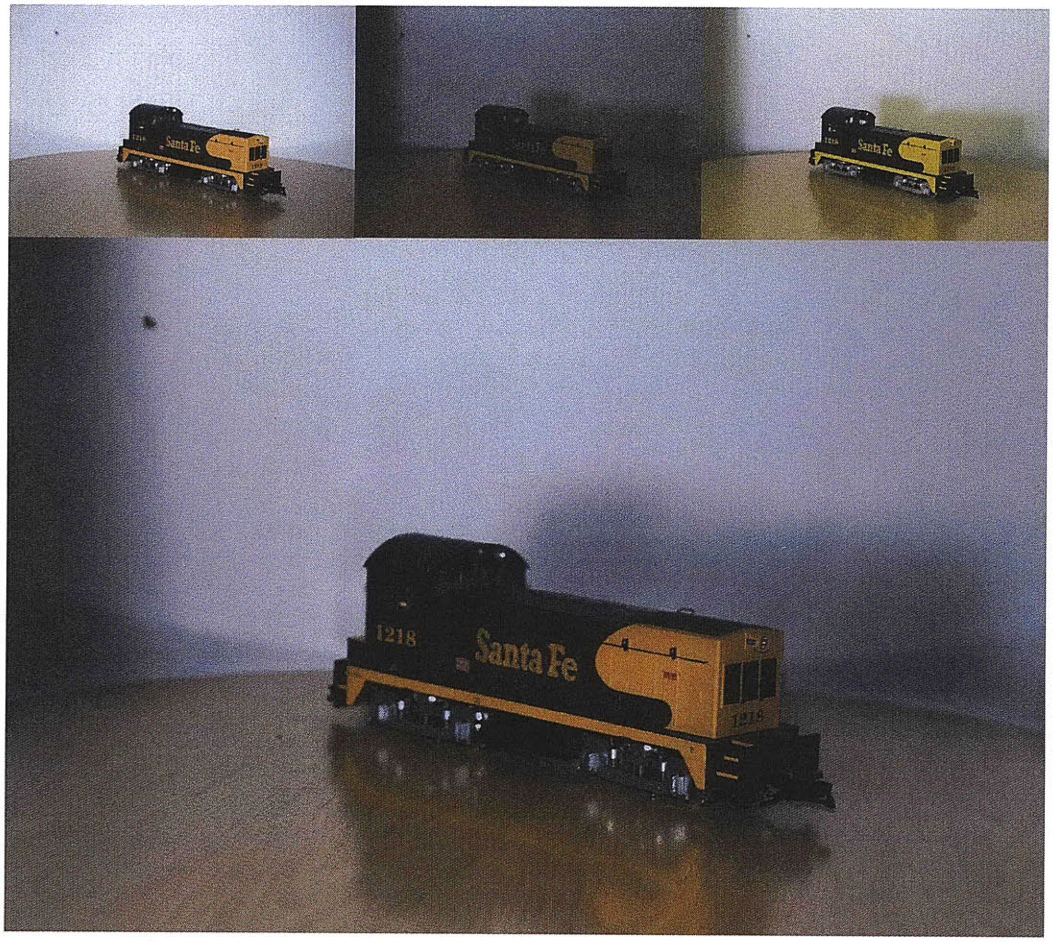


Figure 5-7: *Flash / no-flash image fusion. The reconstructed flash and no-flash (contrast stretched) images are shown at top left and center respectively and a long exposure no-flash image is shown at top right. Our result is shown below.*

there is enough dynamic range to capture useful shading information in the ambient image, even though it is severely under-exposed (see figure 5-5 for an example).

5.4.2 White Balancing with Mixed Lighting

Standard global white balancing methods (e.g. the white-patch and the gray-world [10] methods) assume there is only a single illuminant in the scene and will fail to properly white balance when this assumption is violated. Figure 5-8(e) shows the result of white balancing a scene with two distinct illuminants using the white-patch method. We have selected a patch on the white locker door in the center of the image frame as our reference and performed white balancing relative to it. While the door in the center of the image appears white, it can be seen that the wall and the white locker doors toward the edge of the frame have a distinct reddish-yellow color cast due to the spatially varying mixed illumination in the scene. Hsu and colleagues [38] introduced a method for performing white-balancing under two known illuminants. They analyze local color distributions to estimate per-pixel albedos and then infer the local mixture of each illuminant. This method works well for a large class of images, however may fail when there is not enough lighting variation to reliably estimate albedos.

We propose a simple method for white balancing with two illuminants similar to the method of DiCarlo and colleagues[17]. Using spatio-temporal multiplexing we can recover and then white balance the flash and ambient images independently. We assume the ambient lighting contains a single illuminant and therefore we can directly apply any standard global white balancing technique. The flash image may contain both flash and ambient lighting, and thus we subtract the ambient image from the flash image to obtain a flash-only image. The flash-only image can then be white balanced, again using a global white balance method.

Figures 5-8 and 5-9 show results using our proposed white balance method. Subfigures (a-c) show the reconstructed ambient, combined flash and ambient, and flash-only images, respectively. We assume for this application that the flash and ambient exposures are approximately matched, and that the flash is used primarily as a fill light. Subfigure (d) shows the final result of our method, where the ambient component (a) and the flash-only component (c) have been independently white balanced (using the white-patch method, selecting the same point in each image) and then recombined. Subfigure (e) shows the result of standard global white balance. Note the color casts on the inside of the cup and the toothbrush figure 5-9(e) are absent in our result (figure 5-9(d)).

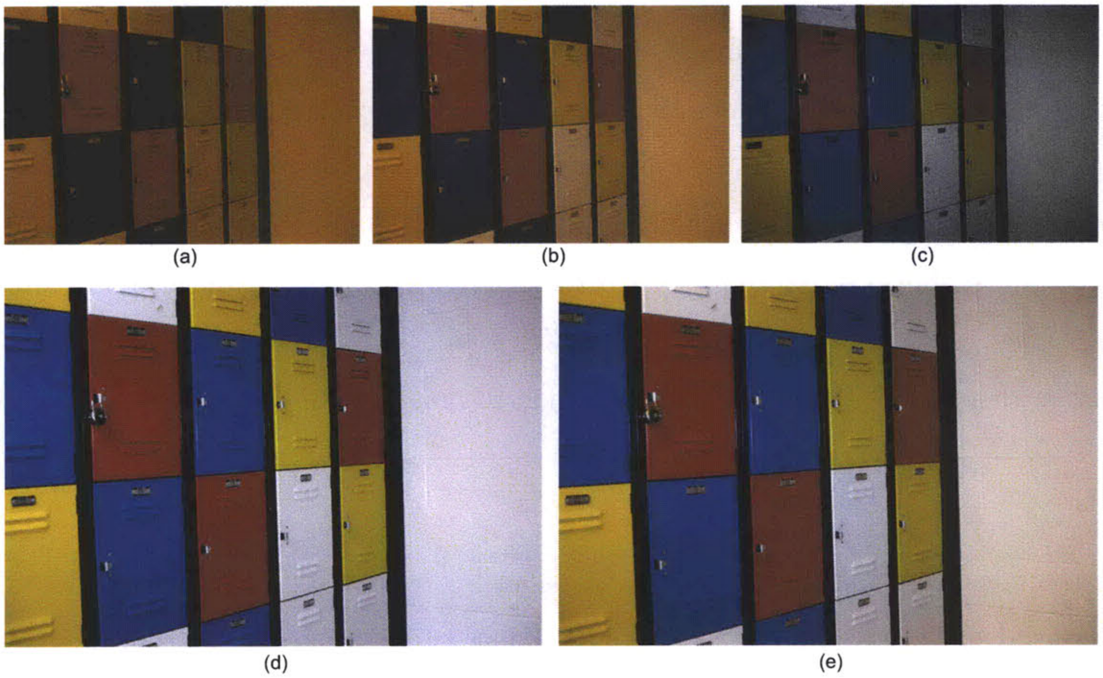


Figure 5-8: *White balancing under mixed illumination. The reconstructed ambient (a), combined flash and ambient (b) and flash-only (c) images are shown in the top row. Our white balanced result is show in (d). A simple global white balance result is shown in (e). Note the color casts on the back wall of (e), not present in our result (d).*

5.5 Discussion

In this chapter we introduced a spatio-temporal multiplexing method that can capture flash and ambient scene information in a single exposure. Our method spatially mosaiced "flash" and "ambient" pixels across the image sensor. We proposed temporally synchronizing the flash with per-pixel electronic shutters to enable capturing flash and ambient illumination separately. We demonstrated two applications of our spatio-temporal multiplexing method: single-exposure flash no-flash and white balancing with mixed illumination.

There are several drawbacks to our proposed method and analysis. The largest limitation of this work is that although we argue its feasibility, we have not demonstrated a physical prototype of our proposed system. Without a prototype, we view this work as only a preliminary investigation of spatio-temporal multiplexing. One

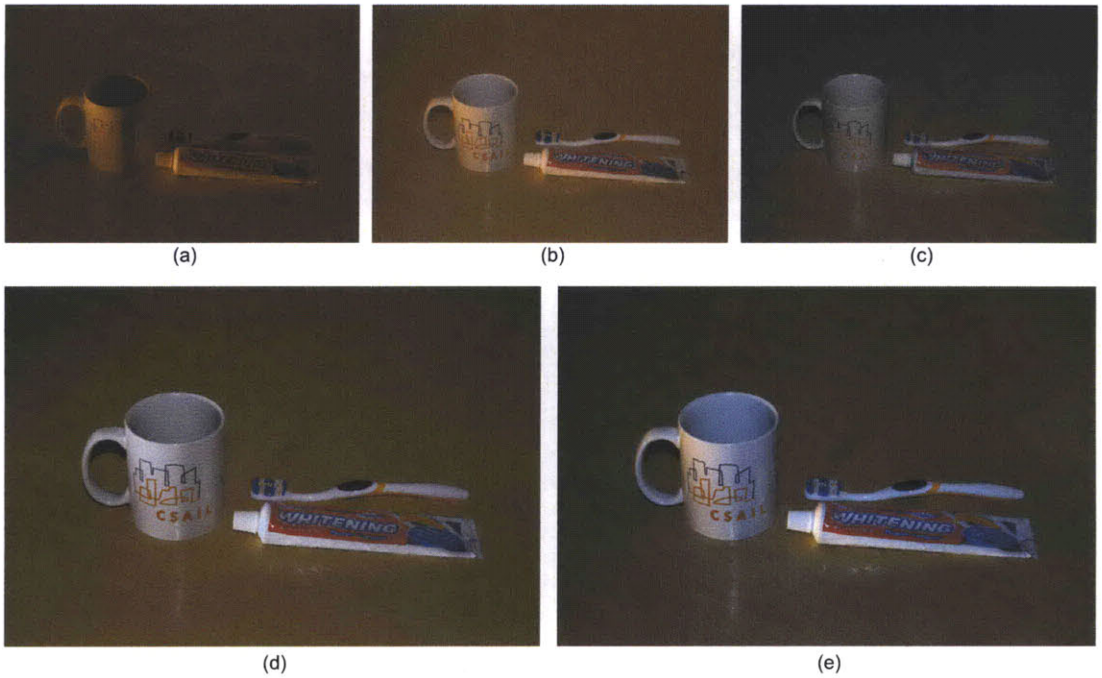


Figure 5-9: *White balancing under mixed illumination. The reconstructed ambient (a), combined flash and ambient (b) and flash-only (c) images are shown in the top row. Our white balanced result is show in (d). A simple global white balance result is shown in (e). Note the blue color casts on the inside of the cup and the toothbrush (e), not present in our result (d).*

particular simplification we have made in our simulations is to assume RGB values at each pixel (e.g. Foveon X3 sensor) and thus ignore demosaicing issues. Building a prototype would facilitate experimentation into different mosaic designs. For example, for the flash/no-flash fusion application, it may be beneficial to remove the spectral filter on some subset of pixels and record total irradiance for the ambient pixels. Another obvious limitation is that we must trade image resolution in order to multiplex information onto the sensor. Fortunately, image resolution for digital cameras has been rapidly increasing in recent years, and has arguably already surpassed the physical resolution limits of the optics and displays. Therefore this excess sensor resolution could be used for methods such as spatio-temporal multiplexing. However, it is important to note that our method requires more complex (and thus more expensive) hardware, which may significantly effect the economically-feasible sensor resolution limits.

In the future we would like to investigate other methods and applications of spatio-temporal multiplexing. One possible extension would be to divide the exposure into three or more phases, and capture the scene under many illumination conditions. For example, capturing multiple flashes placed at different locations could enable single image depth from flash [50, 49, 52].

Chapter 6

Conclusion

In this thesis, we introduced multiplexed photography, a collection of methods for simultaneously capturing multiple camera settings in a single exposure. Our goals were to extend and enhance the capabilities of digital photography, and additionally enable amateur photographers to manage the large space of camera settings. There are many different camera settings (e.g. focus, aperture, shutter speed, and flash) and erroneously setting any one can ruin an otherwise good photograph. We focused on methods and designs that allowed post-exposure editing and control of physical camera settings as well as higher-level controls such as depth of field. One common thread in all of the projects described in this thesis is that we intentionally traded image resolution to capture more information, exploiting the emerging abundance of image resolution found on modern image sensors.

This thesis comprises three projects: multi-aperture photography, multiplexed illumination, and spatio-temporal multiplexing. Each project took a different approach to the goal of capturing multiple camera settings, exploring several areas and approaches of computational photography, including new computational cameras, coded illumination methods, and computational sensors.

In the first project, multi-aperture photography, we described the design and implementation of a prototype optical system and associated algorithms to capture four images of a scene in a single exposure, each taken with a different aperture setting. A unique aspect of our design was that, unlike plenoptic cameras that capture two extra angular dimensions, we captured the one dimensional space of aperture settings directly, minimizing resolution loss. Our goal was to explore the design space of computational cameras, and examine the types of post-exposure edits that are possible without capturing a full light field. We believe that while plenoptic cameras are extremely flexible and powerful, they may be overly general for many applications. We argue it can be fruitful to examine application specific lightfield sampling strategies that are tailored to a particular task (e.g. depth of field extrapolation). Another advantage of our proposed design was that it worked with commercially available DSLR cameras, and did not require any permanent changes to the camera, allowing it to be easily removed, and a full resolution image to be captured if desired. We believe these kinds of designs may be more palatable to many photographers. Using our system we demonstrated several applications of our multi-aperture camera, including adjusting the depth of field and generating synthetically refocused images.

One avenue for future work would be to investigate different methods of coding the aperture. In particular, extending our decomposition to a spatio-temporal splitting of the aperture. This would allow us to recover frequency content lost due either from depth defocus or motion blur. It may also be possible to design an adaptive optical system that adjusts the aperture coding based on the scene. Another avenue of future work is to build a camera that simultaneously captures multiple images focused at different depths in a single exposure, using a single image sensor.

In the second project we described multiplexed flash illumination to recover both flash and ambient light information as well as extract sparse depth information in a single exposure. Traditional photographic flashes illuminate the scene with a spatially-constant light beam. By adding a mask and optics to a flash, we can project a spatially varying illumination onto the scene which allows us to spatially multiplex the flash and ambient illuminations onto the imager. We apply flash multiplexing to enable single exposure flash/no-flash image fusion, in particular, performing flash/no-flash relighting on dynamic scenes with moving objects.

Flash multiplexing demonstrates the potential of computational illumination in dynamic scenes because it enables the simultaneous capture of multiple components of illumination. Our prototype was able to multiplex flash and ambient lighting into assorted flash pixels captured at the image sensor. The defocus of the light pattern further allowed us to extract simple depth information. As an application of our multiplexed flash illumination, we demonstrate the first single-exposure flash/no-flash method suitable for dynamic scenes.

Lastly, we proposed spatio-temporal multiplexing, a novel image sensor integration strategy that enabled simultaneous capture of flash and ambient illumination. We described two possible applications of spatio-temporal multiplexing: single-image flash/no-flash relighting and white balancing scenes containing two distinct illuminants (e.g. flash and fluorescent lighting). Our method spatially mosaiced "flash" and "ambient" pixels across the image sensor. By temporally synchronizing the flash with per-pixel electronic shutters we enabled capturing flash and ambient illumination separately. We demonstrated two possible applications of our spatio-temporal multiplexing method: single-exposure flash no-flash and white balancing with mixed illumination. We have not constructed a physical prototype of our proposed system, and instead only argue its feasibility and benefits. The next step is to build a prototype, which could validate spatio-temporal multiplexing and facilitate experimentation of different mosaic designs.

As mentioned earlier, a key trend that we have exploited time after time in these projects is the excess of resolution on new image sensors. These days, it is common for consumer digital cameras to have 12 megapixel (MP) or larger sensors, which has, arguably, surpassed the needs for most consumer display and printing applications. For example, the largest monitor resolutions are only 3 or 4 MP, a professional quality (at 300dpi) 4" by 6" photo print requires only 2MP, and a fairly large print (8" by 10" at 300dpi) needs roughly 7MP. An interesting avenue for future research is to consider what other dimensions of the camera system may have "excesses of resolution" that can be exploited to capture richer photographs. For example, the dynamic range of sensors is steadily improving with each new generation of sensors, and someday soon the available precision may exceed that which is useful, thus becoming a ripe candidate for further multiplexed photography research.

Bibliography

[1] P.M. Acosta-Serafini, I. Masaki, and C.G. Sodini. Predictive multiple sampling algorithm with overlapping integration intervals for linear wide dynamic range integrating image sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 5(1):33–41, March 2004.

[2] Rolf Adelsberger, Remo Ziegler, Marc Levoy, and Markus Gross. Spatially adaptive photographic flash. Technical Report 612, ETH Zürich, December 2008.

[3] Edward H. Adelson and John Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992.

[4] Adobe. Adobe photoshop. Web page. http://www.adobe.com.

[5] Manoj Aggarwal and Narendra Ahuja. Split aperture imaging for high dynamic range. *Int. J. Comp. Vision*, 58(1):7–17, June 2004.

[6] Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 828–835, New York, NY, USA, 2005. ACM Press.

[7] Michael Ashikhmin. A tone mapping algorithm for high contrast images. In *EGRW '02: Proceedings of the 13th Eurographics workshop on Rendering*, pages 145–156, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.

[8] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11), 2001.

[9] S. Bradburn, E.R. Dowski, and W.T. Cathey. Realizations of focus invariance in optical-digital systems with wavefront coding. *Applied optics*, 36:9157–9166, 1997.

[10] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1 – 26, 1980.

[11] Peter J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

[12] W.T. Cathey and R. Dowski. A new paradigm for imaging systems. *Applied Optics*, 41:1859–1866, 1995.

[13] S. Chaudhuri and A.N. Rajagopalan. *Depth From Defocus: A Real Aperture Imaging Approach.* Springer Verlag, New York, NY, 1999.

[14] Jiawen Chen, Sylvain Paris, and Frédo Durand. Real-time edge-aware image processing with the bilateral grid. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 103, New York, NY, USA, 2007. ACM.

[15] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Computer Graphics*, SIGGRAPH 2000 Proceedings, pages 145–156, July 2000.

[16] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[17] Jeffrey M. DiCarlo, Feng Xiao, and Brian A. Wandell. Illuminating illumination. *Proceedings of Ninth Color Imaging Conference*, pages 27–34, 2001.

[18] E.R. Dowski and W.T. Cathey. Extended depth of field through wavefront coding. *Applied Optics*, 34:1859–1866, 1995.

[19] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 257–266, New York, NY, USA, 2002. ACM.

[20] Alexei A. Efros and Thomas K. Leung. Texture synthesis by non-parametric sampling. In *ICCV (2)*, pages 1033–1038, 1999.

[21] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. In *ACM Transactions on Graphics (Proceedings of Siggraph Conference)*, volume 23. ACM Press, 2004.

[22] Hany Farid and Eero Simoncelli. A differential optical range camera. In *Optical Society of America, Annual Meeting*, Rochester, NY, 1996.

[23] Raanan Fattal, Dani Lischinski, and Michael Werman. Gradient domain high dynamic range compression. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 249–256, New York, NY, USA, 2002. ACM.

[24] A. R. FitzGerrell, E.R. Dowski, and W.T. Cathey. Defocus transfer function for circularly symmetric pupils. *Applied Optics*, 36:5796–5804, 1997.

[25] Foveon. Foveon x3 direct image technology. Web page. http://www.foveon.com/article.php?a=67.

[26] Todor Georgiev, Ke Colin Zheng, Brian Curless, David Salesin, Shree Nayar, and Chintan Intwala. Spatio-angular resolution tradeoffs in integral photography. In *Proceedings of Eurographics Symposium on Rendering (2006)*, pages 263–272, June 2006.

[27] Rafael Gonzalez. *Digital Image Processing.* Pearson/Prentice Hall, Upper Saddle River, 2008.

[28] Paul Green, Wenyang Sun, Wojciech Matusik, and Frédo Durand. Multi-aperture photography. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3), July 2007.

[29] M.D. Grossberg and S.K. Nayar. High Dynamic Range from Multiple Images: Which Exposures to Combine? In *ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, Oct 2003.

[30] M.D. Grossberg, H. Peri, S.K. Nayar, and P.N. Belhumeur. Making one object look like another: controlling appearance using a projector-camera system. volume 1, pages I–452–I–459 Vol.1, June-2 July 2004.

[31] NK Guy. Flash photography with canon EOS cameras. Web page. http://photonotes.org/articles/eos-flash/.

[32] R. P. Harvey. Optical beam splitter and electronic high speed camera incorporating such a beam splitter. US 5734507, United States Patent, 1998.

[33] S.W. Hasinoff and K.N. Kutulakos. Confocal stereo. In *European Conference on Computer Vision (2006)*, pages 620–634, 2006.

[34] E. Hecht. *Optics.* Addison-Wesley, Reading, MA, 2002.

[35] Manfred Heiting, Eelco Wolf, and Vivian Walworth. *The History of Polaroid one-step photography.* Van Soest B.V., Amsterdam, Netherlands, 1978.

[36] Shinsaku Hiura and Takashi Matsuyama. Depth measurement by the multi-focus camera. In *IEEE Computer Vision and Pattern Recognition (1998)*, pages 953–961. IEEE Computer Society, 1998.

[37] Hugues Hoppe and Kentaro Toyama. Continuous flash. Technical Report MSR-TR-2003-63, Microsoft Research, One Microsoft Way, Redmond, WA 98052, Oct 2003.

[38] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédo Durand. Light mixture estimation for spatially varying white balance. *ACM Trans. Graph.*, 27(3):1–7, 2008.

[39] Aaron Isaksen, Leonard McMillan, and Steven J. Gortler. Dynamically reparameterized light fields. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, pages 297–306, New York, NY, USA, July 2000. ACM Press/Addison-Wesley Publishing Co.

[40] Jiaya Jia, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. Bayesian correction of image intensity with spatial consideration. In *8th European Conference on Computer Vision (ECCV 2004)*, pages 342–354, May 2004.

[41] Thouis R. Jones. Efficient generation of poisson-disk sampling patterns. *journal of graphics tools*, 11(2):27–36, 2006.

[42] Johannes Kopf, Michael F. Cohen, Dani Lischinski, and Matt Uyttendaele. Joint bilateral upsampling. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 96, New York, NY, USA, 2007. ACM.

[43] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 70, New York, NY, USA, 2007. ACM.

[44] Anat Levin, William T. Freeman, and Frédo Durand. Understanding camera trade-offs through a bayesian analysis of light field projections. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 88–101, Berlin, Heidelberg, 2008. Springer-Verlag.

[45] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. *ACM Transactions on Graphics*, 23(3):689–694, August 2004.

[46] Marc Levoy. New techniques in computational photography. Web page, May 2008. http://graphics.stanford.edu/talks/compphot-publictalk-may08.pdf.

[47] Marc Levoy, Billy Chen, Vaibhav Vaish, Mark Horowitz, Ian McDowall, and Mark Bolas. Synthetic aperture confocal imaging. *ACM Transactions on Graphics*, 23(3):825–834, August 2004.

[48] Marc Levoy and Pat Hanrahan. Light field rendering. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, New York, NY, USA, 1996. ACM Press.

[49] Miao Liao, Liang Wang, Ruigang Yang, and Minglun Gong. Light fall-off stereo. pages 1–8, June 2007.

[50] Miao Liao, Liang Wang, Ruigang Yang, and Minglun Gong. Real-time light fall-off stereo. pages 1380–1383, Oct. 2008.

[51] Chang Lu, Mark S. Drew, and Graham D. Finlayson. Shadow removal via flash/noflash illumination. pages 198–201, Oct. 2006.

[52] David B. Martin. Depth from Flash. Technical Report TR2000-373, Dartmouth College, Computer Science, Hanover, NH, June 2000.

[53] Morgan McGuire, Wojciech Matusik, Billy Chen, John F. Hughes, Hanspeter Pfister, and Shree Nayar. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications (2007)*, March 2007.

[54] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, John F. Hughes, and Frédo Durand. Defocus video matting. *ACM Transactions on Graphics SIGGRAPH(2005)*, 24(3):567–576, August 2005.

[55] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. *Computer Graphics and Applications, Pacific Conference on*, 0:382–390, 2007.

[56] Xioaping Miao and Terence Sim. Ambient image recovery and rendering from flash photographs. In *ICIP*, pages II: 1038–1041, 2005.

[57] Ankit Mohan, Reynold Bailey, and Jonathan Waite. Tabletop computed lighting for practical digital photography. *IEEE Transactions on Visualization and*

Computer Graphics, 13(4):652–662, 2007. Member-Tumblin,, Jack and Member-Grimm,, Cindy and Senior Member-Bodenheimer,, Bobby.

[58] Francesc Moreno-Noguer, Peter N. Belhumeur, and Shree K. Nayar. Active refocusing of images and videos. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 67, New York, NY, USA, 2007. ACM Press.

[59] T. Naemura, T. Yoshida, and H. Harashima. 3-D computer graphics based on integral photography. *Opt. Expr.*, 8:255–262, Feb 2001.

[60] 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.

[61] 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. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Jul 2006.

[62] S.K. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 472–479, Jun 2000.

[63] S.K. Nayar and S.G. Narasimhan. Assorted Pixels: Multi-Sampled Imaging With Structural Models. In *Europian Conference on Computer Vision (ECCV)*, volume IV, pages 636–652, May 2002.

[64] Ren Ng. Fourier slice photography. *ACM Transactions on Graphics SIGGRAPH (2005)*, 24(3):735–744, August 2005.

[65] Ren Ng. *Digital light field photography.* PhD thesis, Stanford, CA, USA, 2006. Adviser-Hanrahan,, Patrick.

[66] Ren Ng, Marc Levoy, Mathieu Bredif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CSTR 2005-02, Stanford University, 2005.

[67] F. Okano, J. Arai, H. Hoshino, and I. Yuyama. Three-dimensional video system based on integral photography. *Optical Engineering*, 38:1072–1077, June 1999.

[68] Sylvain Paris and Frédo Durand. A fast approximation of the bilateral filter using a signal processing approach. *Int. J. Comput. Vision*, 81(1):24–52, 2009.

[69] Alex P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, 1987.

[70] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23(3):664–672, 2004.

[71] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 795–804, New York, NY, USA, 2006. ACM.

[72] Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Trans. Graph.*, 23(3):679–688, 2004.

[73] Ramesh Raskar, Greg Welch, Kok lim Low, and Deepak Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. In *In Proceedings of Eurographics Workshop on Rendering*, pages 89–102, 2001.

[74] Sidney Ray. *Applied photographic optics.* Focal Press, Boston, MA, 1988.

[75] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267–276, 2002.

[76] Yoav Y. Schechner, Shree K. Nayar, and Peter N. Belhumeur. Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1339–1354, 2007.

[77] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R. Marschner, Mark Horowitz, Marc Levoy, and Hendrik P.A. Lensch. Dual Photography. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2005)*, 2005.

[78] Jian Sun, Sing Bing Kang, Zong-Ben Xu, Xiaoou Tang, and Heung-Yeung Shum. Flash cut: Foreground extraction with flash and no-flash image pairs. In *CVPR*, 2007.

[79] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Flash matting. *ACM Trans. Graph.*, 25(3):772–778, 2006.

[80] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall F. Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields. In *European Conference on Computer Vision (2006)*, pages 16–29, 2006.

[81] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.

[82] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 69, New York, NY, USA, 2007. ACM.

[83] M. Watanabe, S.K. Nayar, and M. Noguchi. Real-time computation of depth from defocus. In *Proceedings of The International Society for Optical Engineering (SPIE)*, volume 2599, pages 14–25, Jan 1996.

[84] Li-Yi Wei and Marc Levoy. Fast texture synthesis using tree-structured vector quantization. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 479–488. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[85] Ben Weiss. Fast median and bilateral filtering. *ACM Trans. Graph.*, 25(3):519–526, 2006.

[86] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics*, 24(3):756–764, July 2005.

[87] H. Yamauchi, J. Haber, and H.-P. Seidel. Image restoration using multiresolution texture synthesis and image inpainting. pages 120–125, July 2003.

[88] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. Image deblurring with blurred/noisy image pairs. *ACM Trans. Graph.*, 26(3):1, 2007.

[89] L. Zhang and S. K. Nayar. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Jul 2006.