DIGITAL LIGHT FIELD PHOTOGRAPHY

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Ren Ng

July 2006

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

Patrick Hanrahan   Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

Marc Levoy

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

Mark Horowitz

Approved for the University Committee on Graduate Studies.

Acknowledgments

I feel tremendously lucky to have had the opportunity to work with Pat Hanrahan, Marc Levoy and Mark Horowitz on the ideas in this dissertation, and I would like to thank them for their support. Pat instilled in me a love for simulating the flow of light, agreed to take me on as a graduate student, and encouraged me to immerse myself in something I had a passion for. I could not have asked for a finer mentor. Marc Levoy is the one who originally drew me to computer graphics, has worked side by side with me at the optical bench, and is vigorously carrying these ideas to new frontiers in light field microscopy. Mark Horowitz inspired me to assemble my camera by sharing his love for dismantling old things and building new ones. I have never met a professor more generous with his time and experience.

I am grateful to Brian Wandell and Dwight Nishimura for serving on my orals committee. Dwight has been an unfailing source of encouragement during my time at Stanford.

I would like to acknowledge the fine work of the other individuals who have contributed to this camera research. Mathieu Brédif worked closely with me in developing the simulation system, and he implemented the original lens correction software. Gene Duval generously donated his time and expertise to help design and assemble the prototype, working even through illness to help me meet publication deadlines. Andrew Adams and Meng Yu contributed software to refocus light fields more intelligently. Kayvon Fatahalian contributed the most to explaining how the system works, and many of the ray diagrams in these pages are due to his artistry.

Assembling the prototype required custom support from several vendors. Special thanks to Keith Wetzel at Kodak Image Sensor Solutions for outstanding support with the photosensor chips, Thanks also to John Cox at Megavision, Seth Pappas and Allison Roberts at Adaptive Optics Associates, and Mark Holzbach at Zebra Imaging.

⁓ For Mama and Papa ⁓

Contents

List of Figures

4 Digital Refocusing 49

5 Signal Processing Framework 79

6   Selectable Refocusing Power   113

7   Digital Correction of Lens Aberrations   131

8 Conclusion   167

A Proofs   171

1

Introduction

This dissertation introduces a new approach to everyday photography, which solves the long-standing problems related to focusing images accurately. The root of these problems is missing information. It turns out that conventional photographs tell us rather little about the light passing through the lens. In particular, they do not record the amount of light traveling along individual rays that contribute to the image. They tell us only the *sum total* of light rays striking each point in the image. To make an analogy with a music-recording studio, taking a conventional photograph is like recording all the musicians playing together, rather than recording each instrument on a separate audio track.

In this dissertation, we will go after the missing information. With micron-scale changes to its optics and sensor, we can enhance a conventional camera so that it measures the light along each individual ray flowing into the image sensor. In other words, the enhanced camera samples the total geometric distribution of light passing through the lens in a single exposure. The price we will pay is collecting much more data than a regular photograph. However, I hope to convince you that the price is a very fair one for a solution to a problem as pervasive and long-lived as photographic focus. In photography, as in recording music, it is wise practice to save as much of the source data as you can.

Of course simply recording the light rays in the camera is not a complete solution to the focus problem. The other ingredient is computation. The idea is to re-sort the recorded light rays to where they should ideally have terminated, to simulate the flow of rays through the virtual optics of an idealized camera into the pixels of an idealized output photograph.

1.1   The Focus Problem in Photography

Focus has challenged photographers since the very beginning. In 1839, the Parisian magazine *Charivari* reported the following problems with Daguerre's brand-new photographic process [Newhall 1976].

> You want to make a portrait of your wife. You fit her head in a fixed iron collar to give the required immobility, thus holding the world still for the time being. You point the camera lens at her face; but alas, you make a mistake of a fraction of an inch, and when you take out the portrait it doesn't represent your wife – it's her parrot, her watering pot – or worse.

Facetious as it is, the piece highlights the practical difficulties experienced by early photographers. In doing so, it identifies three manifestations of the focus problem that are as real today as they were back in 1839.

The most obvious problem is the burden of focusing accurately on the subject before exposure. A badly focused photograph evokes a universal sense of loss, because we all take it for granted that we cannot change the focus in a photograph after the fact. And focusing accurately is not easy. Although modern auto-focus systems provide assistance, a mistake of a "fraction of an inch" in the position of the film plane may mean accidentally focusing past your model onto the wall in the background – or worse. This is the quintessential manifestation of the focus problem.

The second manifestation is closely related. It is the fundamental coupling between the size of the lens aperture and the *depth of field* – the range of depths that appears sharp in the resulting photograph. As a consequence of the nature in which a lens forms an image, the depth of field decreases as the aperture size increases. This relationship establishes one of the defining tensions in photographic practice: how should I choose the correct aperture size? On the one hand, a narrow aperture extends the depth of field and reduces blur of objects away from the focal plane – in Figures 1.1A-C, the arches in the background become clearer as the aperture narrows. On the other hand, a narrow aperture requires a longer exposure, increasing the blur due to the natural shake of our hands while holding the camera and movement in the scene – notice that the woman's waving hand blurs out in Figures 1.1A-C.

| (A): Wide aperture | (B): Medium aperture | (C): Narrow aperture |
| $f/4$, 1/80 sec | $f/11$, 1/10 sec | $f/32$, 8/10 sec |
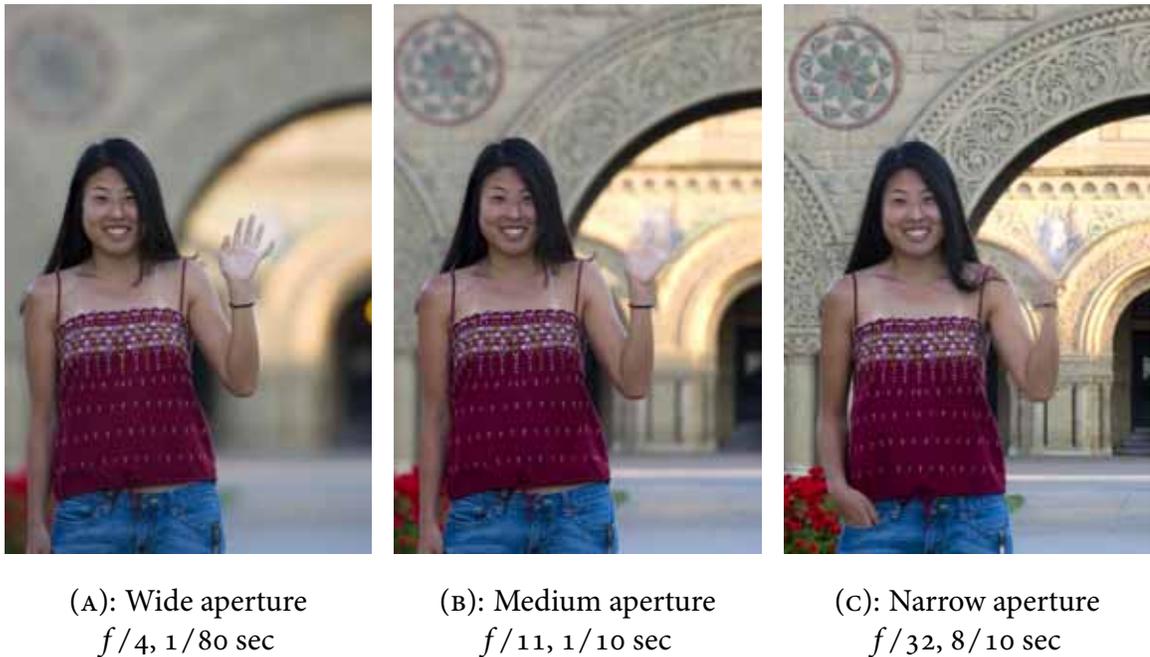
Figure 1.1: Coupling between aperture size and depth of field. An aperture of f/n means that the width of the aperture is $1/n$ the focal length of the lens.

Today's casual picture-taker is slightly removed from the problem of choosing the aperture size, because many modern cameras automatically try to make a good compromise given the light level and composition. However, the coupling between aperture size and depth of field affects the decisions made before every photographic exposure, and remains one of the fundamental limits on photographic freedom.

The third manifestation of the focus problem forces a similarly powerful constraint on the design of photographic equipment. The issue is control of lens aberrations. Aberrations are the phenomenon where rays of light coming from a single point in the world do not converge to a single focal point in the image, even when the lens is focused as well as possible. This failure to converge is a natural consequence of using refraction (or reflection) to bend rays of light to where we want them to go – some of the light inevitably leaks away from the desired trajectory and blurs the final image. It is impossible to focus light with geometric perfection by refraction and reflection, and aberrations are therefore an inescapable problem in all real lenses.

Controlling aberrations becomes more difficult as the lens aperture increases in diameter, because rays passing through the periphery of the lens must be bent more strongly to converge accurately with their neighbors. This fact places a limit on the maximum aperture of usable lenses, and limits the light gathering power of the lens. In the very first weeks of the photographic era in 1839, exposures with small lens apertures were so long (measured in minutes) that many portrait houses actually did use a "fixed iron collar" to hold the subject's head still. In fact, many portraits were taken with the subject's *eyes closed*, in order to minimize blurring due to blinking or wandering gaze [Newhall 1976]. One of the crucial developments that enabled practical portraiture in 1839 was Petzval's mathematically-guided design of a new lens with reduced aberrations and increased aperture size. This lens was 4 times as wide as any previous lens of equivalent focal length, enabling exposures that were 16 times shorter than before. Along with improvements in the sensitivity of the photographic plates, exposure times were brought down to seconds, allowing people who were being photographed to open their eyes and remove their iron collars.

Modern exposures tend to be much shorter – just fractions of a second in bright light – but the problem is far from solved. Some of the best picture-taking moments come upon us in the gentle light of early morning and late evening, or in the ambient light of building interiors. These low-light situations require such long exposures that modern lenses can seem as limiting as the portrait lenses before Petzval. These situations force us to use the modern equivalents of the iron collar: the tripod and the electronic flash.

Through these examples, I hope I've conveyed that the focus problem in photography encompasses much more than simply focusing on the right thing. It is fundamentally also about light gathering power and lens quality. Its three manifestations place it at the heart of photographic science and art, and it loves to cause mischief in the crucial moments preceding the click of the shutter.

1.2   Trends in Digital Photography

If the focus problem is our enemy in this dissertation, digital camera technology is our arsenal. Commoditization of digital image sensors is the most important recent development in the history of photography, bringing a new-found sense of immediacy and freedom to

picture making. For the purposes of this dissertation, there are two crucial trends in digital camera technology: an excess in digital image sensor resolution, and the notion that images are computed rather than directly recorded.

Digital image sensor resolution is growing exponentially, and today it is not uncommon to see commodity cameras with ten megapixels (MP) of image resolution [Askey 2006]. Growth has outstripped our needs, however. There is a growing consensus that raw sensor resolution is starting to exceed the resolving power of lenses and the output resolution of displays and printers. For example, for the most common photographic application of printing 4"×6" prints, more than 2 MP provides little perceptible improvement [Keelan 2002].

What the rapid growth hides is an even larger surplus in resolution that could be produced, but is currently not. Simple calculations show that photosensor resolutions in excess of 100 MP are well within today's level of silicon technology. For example, if one were to use the designs for the smallest pixels present in low-end cameras (1.9 micron pitch) on the large sensor die sizes in high-end commodity cameras (24 mm×36 mm) [Askey 2006], one would be able to print a sensor with resolution approaching 250 MP. There are at least two reasons that such high resolution sensors are not currently implemented. First, it is an implicit acknowledgment that we do not need that much resolution in output images. Second, decreasing pixel size reduces the number of photons collected by each pixel, resulting in lower dynamic range and signal-to-noise ratio (SNR). This trade-off is unacceptable at the high-end of the market, but it is used at the low-end to reduce sensor size and miniaturize the overall camera. The main point in highlighting these trends is that a compelling application for sensors with a very large number of small pixels will not be limited by what can actually be printed in silicon. However, this is not to say that implementing such high-resolution chips would be easy. We will still have to overcome significant challenges in reading so many pixels off the chip efficiently and storing them.

Another powerful trend is the notion that, in digital photography, images are computed, not simply recorded. Digital image sensors enabled this transformation by eliminating the barrier between recording photographic data and processing it. The quintessential example of the computational approach to photography is the way color is handled in almost all commodity digital cameras. Almost all digital image sensors sample only one of the three RGB (red, green or blue) color primaries at each photosensor pixel, using a mosaic of color filters
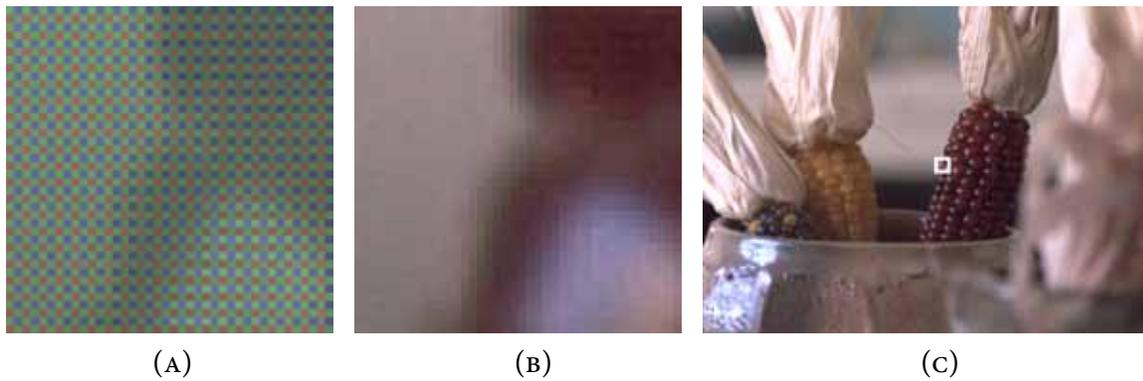
Figure 1.2: Demosaicking to compute color.

in front of each pixel as shown in Figure 1.2A. In other words, each pixel records only one
of the red, green or blue components of the incident light. Demosaicking algorithms [Ra-
manath et al. 2002] are needed to interpolate the mosaicked color values to reconstruct full
RGB color at each output image pixel, as shown in Figures 1.2B and C. This approach enables
color imaging using what would otherwise be an intensity-only, gray-scale sensor. Other
examples of computation in the imaging pipeline include: combining samples at different
sensitivities [Nayar and Mitsunaga 2000] in order to extend the dynamic range [Debevec
and Malik 1997]; using rotated photosensor grids and interpolating onto the final image
grid to better match the perceptual characteristics of the human eye [Yamada et al. 2000];
automatic white-balance correction to reduce color cast due to imbalance in the illumina-
tion spectrum [Barnard et al. 2002]; in-camera image sharpening; and image warping to
undo field distortions introduced by the lens. Computation is truly an integral component
of modern photography.

In summary, present-day digital imaging provides a very rich substrate for new photo-
graphic systems. The two key nutrients are an enormous surplus in raw sensor resolution,
and the proximity of processing power for flexible computation of final photographs.

1.3 Digital Light Field Photography

My proposed solution to the focus problem exploits the abundance of digital image sensor
resolution to sample each individual ray of light that contributes to the final image. This

Figure 1.3: Refocusing after the fact in digital light field photography.

super-representation of the lighting inside the camera provides a great deal of flexibility and control in computing final output photographs. The set of all light rays is called the *light field* in computer graphics. I call this approach to imaging *digital light field photography*.

To record the light field inside the camera, digital light field photography uses a microlens array in front of the photosensor. Each microlens covers a small array of photosensor pixels. The microlens separates the light that strikes it into a tiny image on this array, forming a miniature picture of the incident lighting. This samples the light field inside the camera in a single photographic exposure. A microlens should be thought of as an output image pixel, and a photosensor pixel value should be thought of as one of the many light rays that contribute to that output image pixel.

To process final photographs from the recorded light field, digital light field photography uses ray-tracing techniques. The idea is to imagine a camera configured as desired, and trace the recorded light rays through its optics to its imaging plane. Summing the light rays in this imaginary image produces the desired photograph. This ray-tracing framework provides a general mechanism for handling the undesired non-convergence of rays that is central to the focus problem. What is required is imagining a camera in which the rays converge as desired in order to drive the final image computation.

For example, let us return to the first manifestation of the focus problem – the burden of having to focus the camera before exposure. Digital light field photography frees us of this chore by providing the capability of refocusing photographs after exposure (Figure 1.3). The solution is to imagine a camera with the depth of the film plane altered so that it is focused

as desired. Tracing the recorded light rays onto this imaginary film plane sorts them to a different location in the image, and summing them there produces the images focused at different depths.

The same computational framework provides solutions to the other two manifestations of the focus problem. Imagining a camera in which each output pixel is focused independently severs the coupling between aperture size and depth of field. Similarly, imagining a lens that is free of aberrations yields clearer, sharper images. Final image computation involves taking rays from where they actually refracted and re-tracing them through the perfect, imaginary lens.

1.4   Dissertation Overview

Organizational Themes

The central contribution of this dissertation is the introduction of the digital light field photography system: a general solution to the three manifestations of the focus problem discussed in this introduction. The following four themes unify presentation of the system and analysis of its performance in the coming chapters.

- *System Design: Optics and Algorithms*   This dissertation discusses the optical principles and trade-offs in designing cameras for digital light field photography. The second part of the systems contribution is the development of specific algorithms to address the different manifestations of the focus problem.

- *Mathematical Analysis of Performance*   Three mathematical tools have proven particularly useful in reasoning about digital light field photography. The first is the traditional tracing of rays through optical systems. The second is a novel Cartesian ray-space diagram that unifies visualizations of light field recording and photograph computation. The third is Fourier analysis, which yields the simplest way to understand the relationship between light fields and photographs focused at different depths. These tools have proven remarkably reliable at predicting system performance.

- *Computer-Simulated Validation*   Software ray-tracing enables computer-aided plotting of ray traces and ray-space diagrams. Furthermore, when coupled with a complete computer graphics rendering system, it enables physically-accurate simulation of light fields

and final photographs from hypothetical optical designs.

- *A Prototype Camera and Experimental Validation*   The most tangible proof of system via-
  bility is a system that works, and this dissertation presents a second-generation prototype
  light field camera. This implementation provides a platform for in-depth physical valida-
  tion of theoretical and simulated performance. The success of these tests provides some
  reassurance as to the end-to-end viability of the core design principles. In addition, I have
  used the prototype to explore real live photographic scenarios beyond the reach of theo-
  retical analysis and computer simulation.

Dissertation Roadmap

I have tried to write and illustrate this dissertation in a
manner that will hopefully make it accessible and inter-
esting to a broad range of readers, including lay photog-
raphers, computer graphics researchers, optical engineers
and those who enjoy mathematics.

   Figure 1.4 is a map of some paths that one might
choose through the coming chapters, and the ter-
ritory that one would cover.   Photographers will
be most interested in the images and discussion of
Sections 4.6 – 4.7, and may wish to begin their explo-
ration there. Chapters 2 – 4 assume knowledge of calcu-
lus at the level of a first year college course, but it is not
not essential to develop the right intuition and digest the

Figure 1.4: Roadmap.

main ideas. Chapters 5 – 7 may be read in any order. They present more sophisticated anal-
ysis and variations of the system, and employ more specialized mathematics and abstract
reasoning.
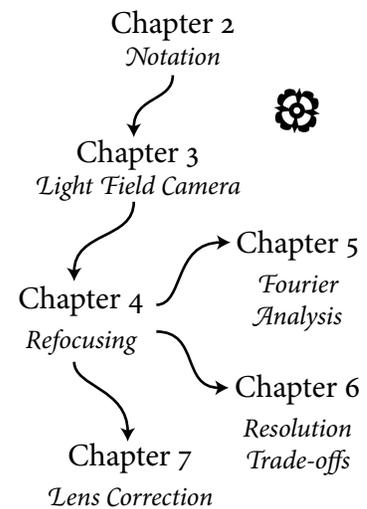
Chapter Descriptions

- Chapter 2 introduces notation and reviews the link between light fields and photographs.
- Chapter 3 presents the design principles, optics, and overall processing concepts for the

system. It also describes the prototype camera.

- Chapter 4 develops theory and algorithms to compute photographs with arbitrary focus and depth of field, and presents experimental validation of predicted performance. It contains a gallery of images shot with the prototype camera.

- Chapter 5 applies Fourier analysis to refocusing performance. This style of thinking leads to a fast Fourier-domain algorithm for certain kinds of light field post-processing.

- Chapter 6 continues the development of performance trade-offs, presenting a dynamic means to exchange image resolution and refocusing power in the light field camera.

- Chapter 7 develops processing algorithms to reduce the effects of lens aberrations, testing performance with the prototype camera and computer simulation.

- Chapter 8 summarizes lessons learned and points to future directions.

2

Light Fields and Photographs

Photographs do not record most of the information about the light entering the camera. For example, if we think about the light deposited at one pixel, a photograph tells us nothing about how the light coming from one part of the lens differs from the light coming from another. It turns out that these differences are the crucial pieces of missing information that lead to the focus problem in conventional photography.

This chapter introduces the notion of the total geometric distribution of light flowing into a camera, the *light field* inside the camera, and defines a very useful graphical representation of the light field called the *ray-space*. By plotting the footprint of a conventional photograph on the ray-space, we can visualize which parts of the light field we measure, and which parts are missing. These are the building blocks for later chapters that deal with recording and processing the full light field.

2.1   Previous Work

Thinking about the geometric distribution of light flowing in the world has a long history. Adelson and Bergen [1991] point to examples in the notebooks of Leonardo da Vinci. Levoy and Hanrahan [1996] trace the mathematical formulations of the total distribution back to early work on illumination engineering by Gershun [1939] in Moscow, and advanced by Moon and Spencer [1981] and others in America.

In the last decade and a half, thinking in terms of the total flow of light has become very

popular in vision and computer graphics, finding a well-deserved, central position in the theories of these fields. Adelson and Bergen were amongst the first in this modern intellectual movement, leading the way with an influential paper that defined the total geometric distribution of light as a *plenoptic function*[1] over the 5D space of rays – 3D for each spatial position and 2D for each direction of flow. They introduced the plenoptic function in order to systematically study how our visual systems might extract geometric information from the images that we see [Adelson and Bergen 1991]. The ray-based model is a natural abstraction for the light flowing in the world, because light is conserved as it propagates along a ray. The precise measure for the light traveling along a ray is defined in radiometry as *radiance* [Preisendorfer 1965], but for the purposes of this thesis it will largely suffice to think of light as a scalar value traveling along each ray (or a scalar value for each color channel).

The notion of the total distribution as a 4D *light field*,[2] which is the one used in this thesis, was introduced to computer graphics by Levoy and Hanrahan [1996] and Gortler et al. [1996] (who called it the *Lumigraph*). The reduction from 5D plenoptic function to 4D light field works by restricting attention to rays passing through free-space – regions free of occluders, such as opaque objects, and scattering media, such as fog. In this case, the light traveling along a ray is constant along its length, eliminating one dimension of variation. The resulting light field is closely related to the *epipolar volumes* developed by Bolles, Baker and Marimont [1987] in studying robot vision.

One reason for reducing the dimensionality was making measurement feasible. For example, the 1996 papers described methods for taking thousands of pictures of an object to sample the 4D space, allowing synthetic views of it to be computed from any viewpoint outside its convex hull. The idea of computing synthetic views from a database of images originated with Chen and Williams [1993], and was first cast in terms of the plenoptic function by McMillan and Bishop [1995]. Following this early work, research on *image-based rendering* techniques exploded in popularity, presenting an alternative to traditional methods based on explicit modeling of surface geometry, reflection properties and source lighting. Shum and Kang [2000] survey some of the earlier work, and the best place to sample the current state of the art is each year's proceedings of the SIGGRAPH conference. The next two

---

[1] From the Latin *plenus* for complete or full, they explain.
[2] *Light field* was a term first used by Gershun [1939] in Russian, translated by Moon and Timoshenko.

chapters review more related work in the context of recording light fields, refocusing and Fourier analysis.

With respect to this chapter, the one issue worth calling out is the choice of parameterization. The original parameterizations [Levoy and Hanrahan 1996; Gortler et al. 1996] were based on the intersection of rays with two planes, which had the prime advantage of simplicity. One limitation was uneven sampling density, motivating explorations in more uniform representations [Camahort et al. 1998] and representations projected onto the surface of objects [Miller et al. 1998; Wood et al. 2000; Chen et al. 2002]. Another limitation was the inability to reliably extrapolate viewing position into the scene because of occluders, motivating representations that incorporated depths or opacities along rays [Shade et al. 1998; Buehler et al. 2001; Matusik et al. 2002]. These improved performance in specific scenarios, but at the expense of increased complexity in representation and processing. Practical and general representations are still something of an unsolved problem in image-based rendering.

2.2 The Light Field Flowing into the Camera

One of the core ideas in this thesis is to restrict ourselves to looking at the light field *inside the camera body*. With this narrowing of scope, the appropriate representation becomes refreshingly simple and general. The issue of uniformity is easily solved because all the light originates from the window of the lens aperture. In addition, the problems with occlusion are gone – the inside of a camera is empty by definition. Nevertheless, I hope you'll be convinced that the light field inside the camera is a rich light field indeed, and that studying it teaches us a great deal about some of the oldest problems in photography.

In this thesis we will use the parameterization of the light field shown in Figure 2.1, which describes each ray by its intersection points with two planes: the film and the aperture inside the lens. The two-plane parameterization is a very natural fit for the light field inside the camera because every ray that contributes to a photograph passes through the lens aperture and terminates somewhere on the film.

In the *ray diagram* on the left of Figure 2.1, a single ray is shown passing through the lens aperture at u as it refracts through the glass of the lens, and terminating at position x on
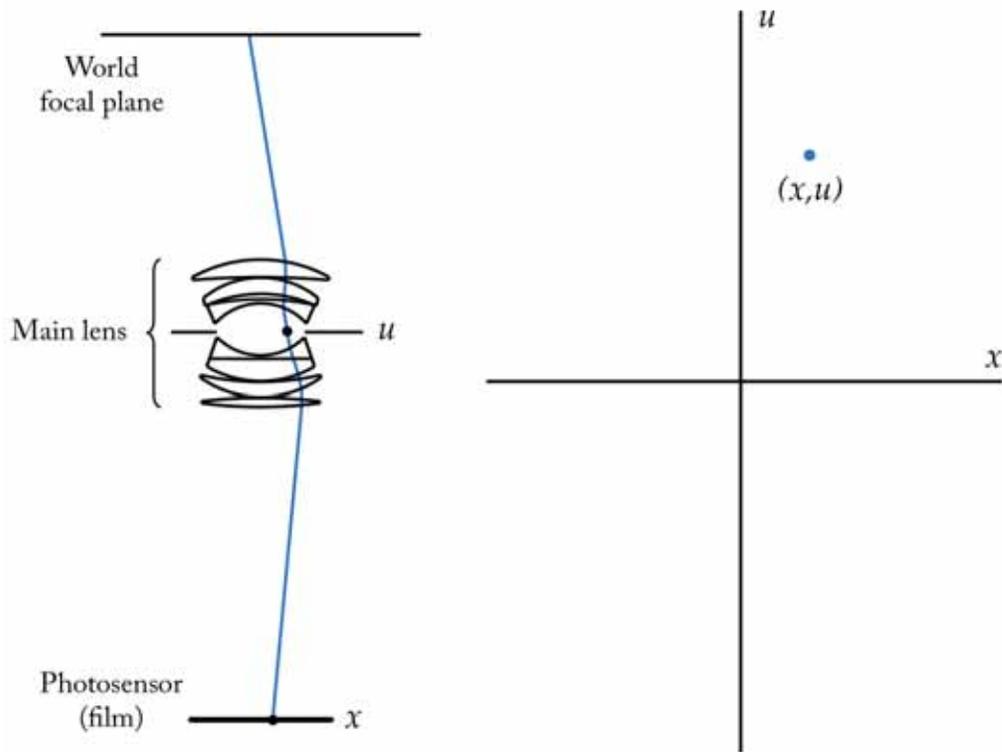
Figure 2.1: Parameterization for the light field flowing into the camera.

the photosensor. Let us refer to u as the directional axis, because the u intercept on the lens determines the direction at which the ray strikes the sensor. In addition, let us refer to x as the spatial axis. Of course in general the ray exists in 3D and we would consider intersections (u, v) at the lens and (x, y) on the film plane. Let us refer to the value of the light field along the depicted ray as $L(x, y, u, v)$, or $L(x, u)$ if we are considering the 2D simplification.

The *Cartesian ray-space diagram* on the right in Figure 2.1 is a more abstract representation of the two-dimensional light field. The ray depicted on the left is shown as a point (x, u) on the Cartesian ray-space. In general each possible ray in the diagram on the left corresponds to a different point on the ray-space diagram on the right, as suggested by Figure 2.2. The function defined over the ray-space plane is the 2D light field. Adelson and Bergen [Adelson and Bergen 1991] used these kinds of diagrams to illustrate simple features in the plenoptic function. Levoy and Hanrahan [1996] used it to visualize the density
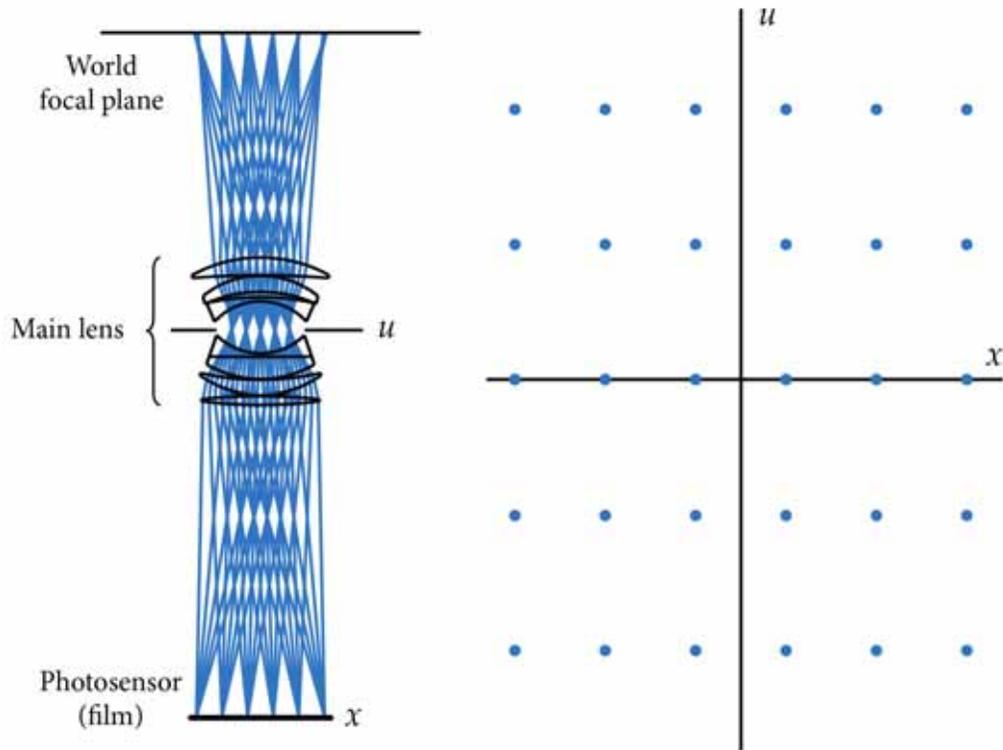
Figure 2.2: The set of all rays flowing into the camera.

of rays in a sampled light field, and it has become very common in the light field literature.

2.3   Photograph Formation

In a conventional camera, a photograph forms on a piece of photosensitive material placed inside the camera at the imaging plane. The material may be silver-halide film in traditional cameras, where photons cause the development of silver crystals, or a CCD or CMOS photosensor in a digital camera, where photons generate free electrons that accumulate in each sensor pixel. Each position on the photosensitive imaging plane sums all the rays of light that terminate there.

In general, the weight of each ray in the sum depends on its incident direction with the sensor plane. For example, radiometry predicts that rays from the periphery of the lens,
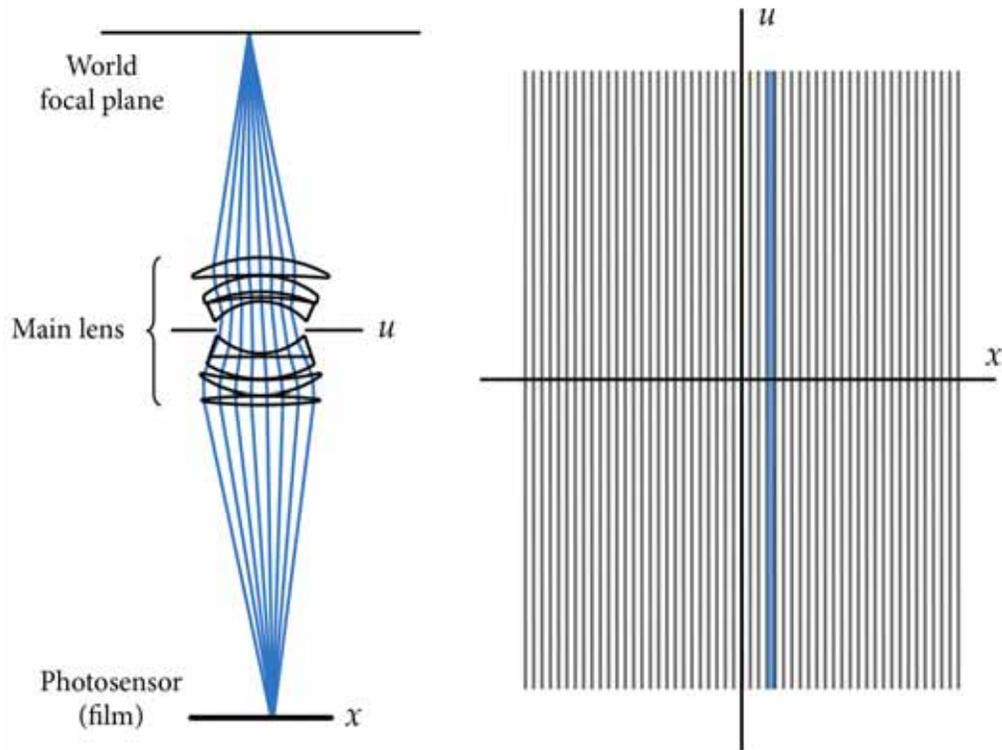
Figure 2.3: The cone of rays summed to produce one pixel in a photograph.

which arrive at the sensor from more oblique angles, contribute less energy to the value of the pixel. Another example is that the photosensitive portion of a pixel in a CMOS sensor is typically obscured by an overlay of metal wires [Catrysse and Wandell 2002], so rays from unshadowed directions will contribute more light. Nevertheless, these directional effects are in some sense undesirable artifacts due to physical or implementation limitations, and Figure 2.3 neglects them in illustrating the formation of somewhat idealized photographs.

Figure 2.3 draws in blue the cone of rays contributing to one photograph pixel value. This cone corresponds (in 2D) to the blue vertical strip on the ray-space diagram because the rays in the cone share the same x film intercept, but vary over all u positions on the lens. Of course different pixels in the photograph have different x intercepts, so they correspond to different vertical lines on the ray-space.

In fact, the ray-space drawn in Figure 2.3 is overlaid with vertical strips, where each strip

is the set of rays summed by a different photograph pixel. This drawing shows that the formation of a full photograph corresponds on the ray-space diagram to a vertical projection of the light field values. The projection preserves the spatial x location of the rays, but destroys the directional u information.

The preceding discussion is, however, limited to the photograph that forms on a piece of film that coincides with the x parameterization plane. Later chapters study the computation of photographs focused at different depths from a recording of the light field, and this kind of digital refocusing depends on an understanding of the representation of photographs focused at different depths in terms of the ray space diagram in Figure 2.3.

Figure 2.4 illustrates how the projection of the ray space changes as the camera is focused at different depths. In these diagrams, the x plane is held fixed at a canonical depth while the film plane of the camera moves. Changing the separation between the film and the lens is how we focus at different depths in a conventional camera. For example, turning the focus ring on a photographic lens simply slides the glass elements along the axis of the lens. In Figure 2.4A, the film plane is moved further from the lens, and the world focal plane moves closer to the camera. The cone of blue rays corresponds to the blue strip with positive slope on the ray-diagram. In contrast, Figure 2.4B shows that when the camera is focused further in the world, the corresponding vertical strip on the ray space has negative slope.

The slope of the ray-space strip can be understood by the fact that the convergence point of the rays moves away from the x film plane. As the intercept of a ray moves linearly across the u lens plane, the resulting intercept moves linearly across the x film plane. If the convergence point is further from the lens than the x plane, then the movement across the u plane is in the same direction as the movement across the x plane (Figure 2.4A). These two directions are opposed if the convergence point is in front of the x plane (Figure 2.4B). These figures make it visually clear that the relative rates of the movements, hence slopes on the ray-space diagram, depend on the separation between the convergence point and the x plane.

This separation is the same for every pixel in the photograph because the world focal plane is parallel to the x plane,[3] so the slant of each pixel is the same. Indeed, Figure 2.4 shows that the entire sampling grid shears to the left if the focus is closer than the x plane,

---

[3] This is limited to the most common type of cameras where the film and the lens are parallel, but not to view cameras where they may be tilted relative to one another. In a view camera, the projection of the ray-space is fan-shaped.
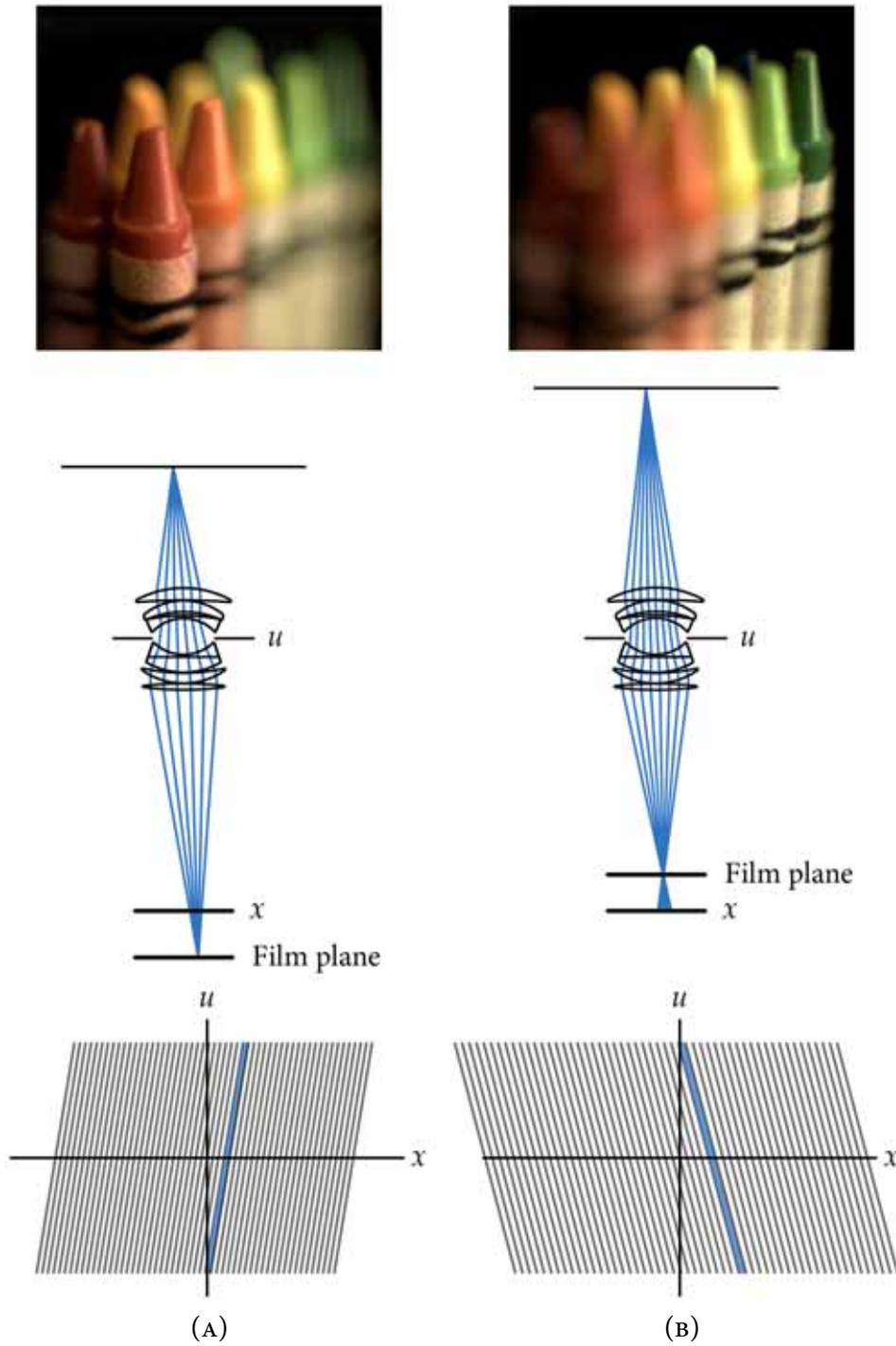
Figure 2.4: The projection of the light field corresponding to focusing further and closer than the chosen x parameterization plane for the light field.

and to the right if the focus is further. This is the main point of this chapter: a photograph is an integral projection of the canonical light field, where the trajectory of the projection depends on the depth at which the photograph is focused.

2.4 Imaging Equations

The 2D simplification above is well suited to visualization and high-level intuition, and will be used for that purpose throughout this thesis. However, a formal mathematical version of the 4D representation is also required for the development of analysis and algorithms. To conclude this introduction to photographs and light fields, this section derives the equations relating the canonical light field to photographs focused at different depths. As we will see in Chapter 4, this mathematical relationship is a natural basis for computing photographs focused at different depths from the light fields recorded by the camera introduced in the next chapter.

The image that forms inside a conventional camera, as depicted in Figure 2.3, is proportional to the irradiance on the film plane. Classical radiometry shows that the irradiance from the aperture of a lens onto a point on the film is equal to the following weighted integral of the radiance coming through the lens [Stroebel et al. 1986]:

$$E_F(x,y) = \frac{1}{F^2} \iint L_F(x,y,u,v) \cos^4\theta \, du \, dv, \tag{2.1}$$

where F is the separation between the exit pupil of the lens and the film, $E_F(x,y)$ is the irradiance on the film at position (x,y), L_F is the light field parameterized by the planes at separation F, and θ is the angle between ray (x,y,u,v) and the film plane normal. The $\cos^4\theta$ term is a well-known falloff factor sometimes referred to as *optical vignetting*. It represents the reduced effect of rays striking the film from oblique directions. However, Equation 2.1 ignores all other directional dependences such as surface microstructure in CMOS sensors.

For simplicity, Equation 2.1 also assumes that the uv and xy planes are infinite in extent, and that L is simply zero beyond the physical bounds of the lens and sensor. To further simplify the equations in the derivations throughout the thesis, let us also absorb the $\cos^4\theta$ into the definition of the light field itself, by re-defining $L(x,y,u,v) := L(x,y,u,v) \cos^4\theta$.
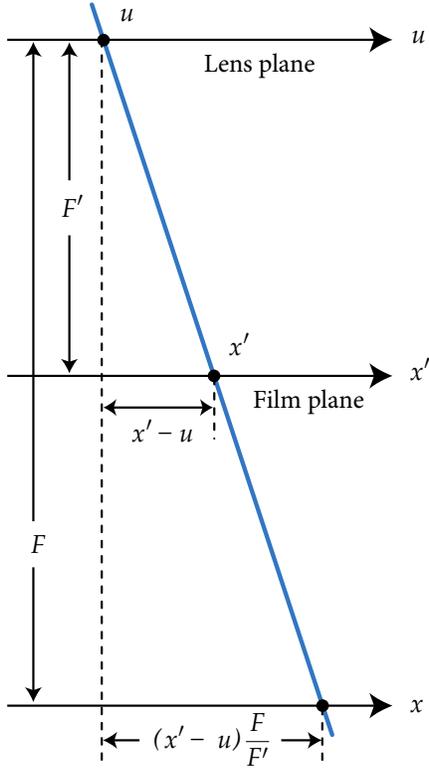
Figure 2.5: Transforming
ray-space coordinates.

This re-definition is possible without reducing ac-
curacy for two reasons. First we will only be deal-
ing with re-parameterizations of the light field that
change the separation between the parameteriza-
tion planes. Second, θ depends only on the angle
that the ray makes with the light field planes, not
on their separation.

Let us now turn our attention to the equations
for photographs focused at depths other than the x
parameterization plane. As shown in Figure 2.4, fo-
cusing at different depths corresponds to changing
the separation between the lens and the film plane,
resulting in a shearing of the trajectory of the inte-
gration lines on the ray-space. If we consider the
photograph focused at a new film depth of F', then
deriving its imaging equation is a matter of express-
ing $L_{F'}(x', u)$ in terms of $L_F(x, u)$ and then applying
Equation 2.1.

The diagram above is a geometric construction
that illustrates how a ray parameterized by the x'
and u planes for $L_{F'}$ may be re-parameterized by its intersection with planes x and u for L_F.
By similar triangles, the illustrated ray that intersects the lens at u and the film plane at x',
also intersects the x plane at $u + (x' - u)\frac{F}{F'}$. Although the diagram only shows the 2D case
involving x and u, the y and v dimensions share an identical relationship. As a result, if we
define $\alpha = F'/F$ as the relative depth of the film plane,

$$L_{F'}(x', y', u, v) = L_F\left(u + \frac{x' - u}{\alpha},\ v + \frac{y' - v}{\alpha},\ u,\ v\right)$$

$$= L_F\left(u\left(1 - \frac{1}{\alpha}\right) + \frac{x'}{\alpha},\ v\left(1 - \frac{1}{\alpha}\right) + \frac{y'}{\alpha},\ u,\ v\right). \tag{2.2}$$

This equation formalizes the 4D shear of the canonical light field that results from focusing at
different depths. Although the shear of the light field within the camera has not been studied

in this way before, it has been known for some time that the two-plane light field shears when one changes its parameterization planes. It was observed by Isaksen et al. [2000] in a study of dynamic reparameterization and synthetic focusing, and is a basic transformation in much of the recent research on the properties of the light field [Chai et al. 2000; Stewart et al. 2003; Durand et al. 2005; Vaish et al. 2005; Ng 2005]. The shear can also be seen in the fundamental matrix for light propagation in the field of matrix optics [Gerrard and Burch 1975].

Combining Equations 2.1 and 2.2 leads to the final equation for the pixel value (x', y') in the photograph focused on a piece of film at depth $F' = \alpha \cdot F$ from the lens plane:

$$E_{(\alpha \cdot F)}(x', y') = \frac{1}{\alpha^2 F^2} \int\int L_F\left(u\left(1 - \frac{1}{\alpha}\right) + \frac{x'}{\alpha}, \; v\left(1 - \frac{1}{\alpha}\right) + \frac{y'}{\alpha}, \; u, \; v\right) du \, dv. \qquad (2.3)$$

This equation formalizes the notion that imaging can be thought of as shearing the 4D light field, and then projecting down to 2D.

3

Recording a Photograph's Light Field

This chapter describes the design of a camera that records the light field on its imaging plane in a single photographic exposure. Aside from the fact that it records light fields instead of regular photographs, this camera looks and operates largely like a regular digital camera. The basic idea is to insert an array of microlenses in front of the photosensor in a conventional camera. Each microlens covers multiple photosensor pixels, and separates the light rays that strike it into a tiny image on the pixels underneath (see, for example, Figure 3.3).

The use of microlens arrays in imaging is a technique called *integral photography* that was pioneered by Lippmann [1908] and greatly refined by Ives [1930; 1931]. Today microlens arrays are used in many ways in diverse imaging fields. Parallels can be drawn between various branches of engineering, optics and the study of animal vision, and the end of this chapter surveys related work.

If the tiny images under each microlens are focused on the main lens of the camera, the result is something that Adelson and Wang [1992] refer to as a *plenoptic camera*. This configuration provides maximum directional resolution in recorded rays. Section 3.5 of this chapter introduces a *generalized light field camera* with configurable microlens focus. Allowing the microlenses to defocus provides flexible trade-offs in performance.

Section 3.6 describes the design and construction of our prototype light field camera. This prototype provides the platform for the experiments in subsequent chapters, which elaborate on how to post-process the recorded light field to address different aspects of the focus problem.

3.1   A Plenoptic Camera Records the Light Field

The plenoptic camera comprises a main photographic lens, a microlens array, and a digital photosensor, as shown in Figure 3.1. The scale in this figure is somewhat deceptive, because the microlenses are drawn artificially large to make it possible to see them and the overall camera at the same scale. In reality they are microscopic compared to the main lens, and so is the gap between the microlenses and the photosensor. In this kind of camera, the microlens plane is the imaging plane, and the size of the individual microlenses sets the spatial sampling resolution.

A grid of boxes lies over the ray-space diagram in Figure 3.1. This grid depicts the sampling of the light field recorded by the photosensor pixels, where each box represents the bundle of rays contributing to one pixel on the photosensor. To compute the sampling grid, rays were traced from the boundary of each photosensor pixel out into the world through its parent microlens array and through the glass elements of the main lens. The intercept of the ray with the microlens plane and the lens plane determined its (x, u) position on the ray-space diagram. As an aside, one may notice that the lines in the ray-space diagram are not perfectly straight, but slightly curved. The vertical curvature is due to aberrations in the optical design of the lens. Correcting such defects is the subject of Chapter 7. The curvature of the boundary at the top and bottom is due to movement in the exit pupil of the aperture as one moves across the x plane.

Each sample box on the ray-space corresponds to a narrow bundle of rays inside the camera. For example, Figure 3.1 depicts two colored sample boxes, with corresponding ray-bundles on the ray diagram. A column of ray-space boxes corresponds to the set of all rays striking a microlens, which are optically sorted by direction onto the pixels underneath the microlens, as shown by the gray boxes and rays in Figure 3.2. If we summed the gray photosensor pixel samples on Figure 3.2 B1, we would compute the value of an output pixel the size of the microlens, in a photograph that were focused on the optical focal plane.

These examples highlight how different the plenoptic camera sampling grid is compared to that for a conventional camera (compare Figure 2.3). In the conventional camera, all the grid cells extend from the top to the bottom of the ray space, and their corresponding set of rays in the camera is a cone that subtends the entire aperture of the lens. In the conventional
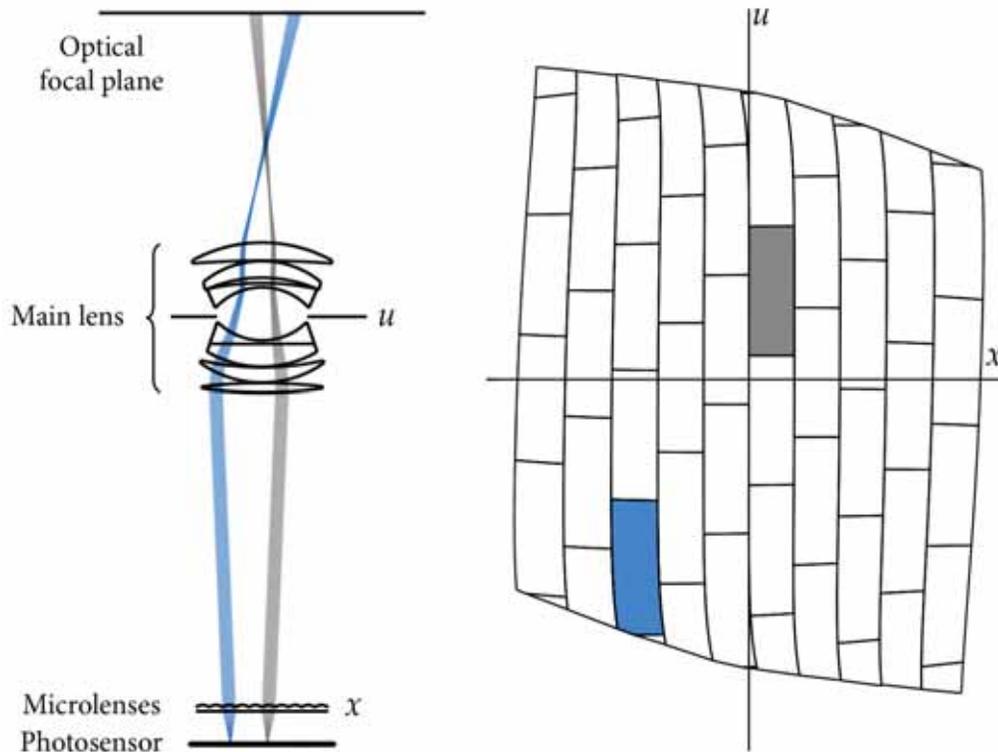
Figure 3.1: Sampling of a photograph's light field provided by a plenoptic camera.

camera the width of a grid column is the width of a photosensor pixel. In the plenoptic camera, on the other hand, the grid cells are shorter and wider. The column width is the width of a microlens, and the column is vertically divided into the number of pixels across the width of the microlens. In other words, the plenoptic camera sampling grid provides more specificity in the u directional axis but less specificity in the x spatial axis, assuming a constant number of photosensor pixels.

This is the fundamental trade-off taken by the light field approach to imaging. For a fixed sensor resolution, collecting directional resolution results in lower resolution final images, with essentially as many pixels as microlenses. On the other hand, using a higher-resolution sensor allows us to add directional resolution by collecting more data, without necessarily sacrificing final image resolution. As discussed in the introductory chapter, much higher resolution sensors may be possible in today's semiconductor technology. Finally, Section 3.5

shows how to dynamically configure the light field camera to record a different balance of spatial and directional resolution.

3.2   Computing Photographs from the Light Field

This thesis relies on two central concepts in processing the light field to produce final photographs. The first is to treat synthesis of the photograph as a simulation of the image formation process that would take place in a desired virtual camera. For example, Figure 3.2A illustrates in blue all the rays that contribute to one pixel in a photograph refocused on the indicated virtual focal plane. As we know from Chapter 2, this blue cone corresponds to a slanted blue line on the ray-space, as shown in Figure 3.2 B1. To synthesize the value of the pixel, we would estimate the integral of light rays on this slanted line.

The second concept is that the radiance along rays in the world can be found by ray-tracing. Specifically, to find the radiance along the rays in the slanted blue strip, we would geometrically trace the ideal rays from the world through the main lens optics, through the microlens array, down to the photosensor surface. This process is illustrated macroscopically by the blue rays in Figure 3.2A. Figure 3.2 B2 illustrates a close-up of the rays tracing through the microlenses down to the photosensor surface. Each of the shaded sensor pixels corresponds to a shaded box on the ray-space of Figure 3.2 B3. Weighting and summing these boxes estimates the ideal blue strip on Figure 3.2 B1. The number of rays striking each photosensor pixel in B2 determines its weight in the integral estimate. This process can be thought of as rasterizing the strip onto the ray-space grid and summing the rastered pixels.

There are more efficient ways to estimate the integral for the relatively simple case of refocusing, and optimizations are presented in the next chapter. However, the ray-tracing concept is very general and subsumes any camera configuration. For example, it handles the case of the generalized light field camera model where the microlenses defocus, as discussed in Section 3.5 and Chapter 6, as well as the case where the main lens exhibits significant optical aberrations, as discussed in Chapter 7.

A final comment regards the first concept of treating image synthesis as simulating the flow of light in a real camera. This approach has the appealing quality of producing final images that look like ordinary, familiar photographs, as we shall see in Chapter 4. However,
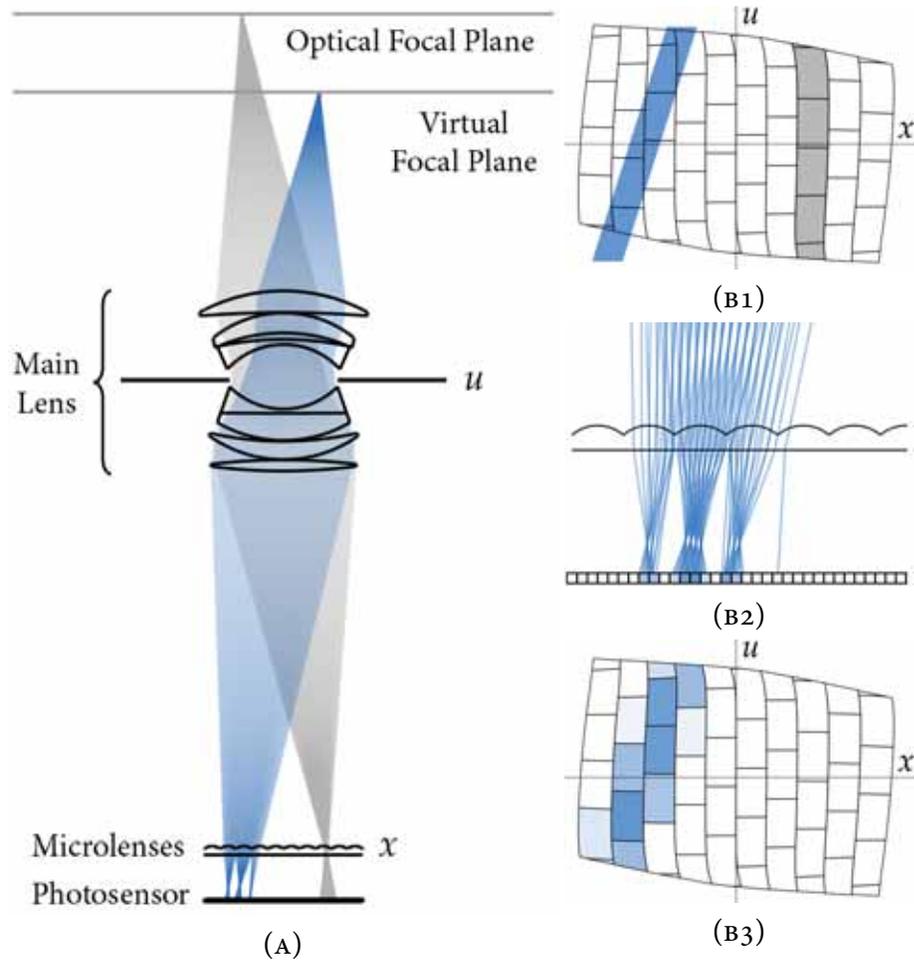
Figure 3.2: Overview of processing the recorded light field.

it should be emphasized that the availability of a light field permits ray-tracing simulations of much more general imaging configurations, such as view cameras where the lens and sensor are not parallel, or even imaging where each pixel is focused with a different depth of field or a different depth. This final example plays a major role in the next chapter.

3.3 Three Views of the Recorded Light Field

In sampling the light traveling along all rays inside the camera, the light field provides rich information about the imaged scene. One way to think of the data is that it captures the

(z1)   (z2)
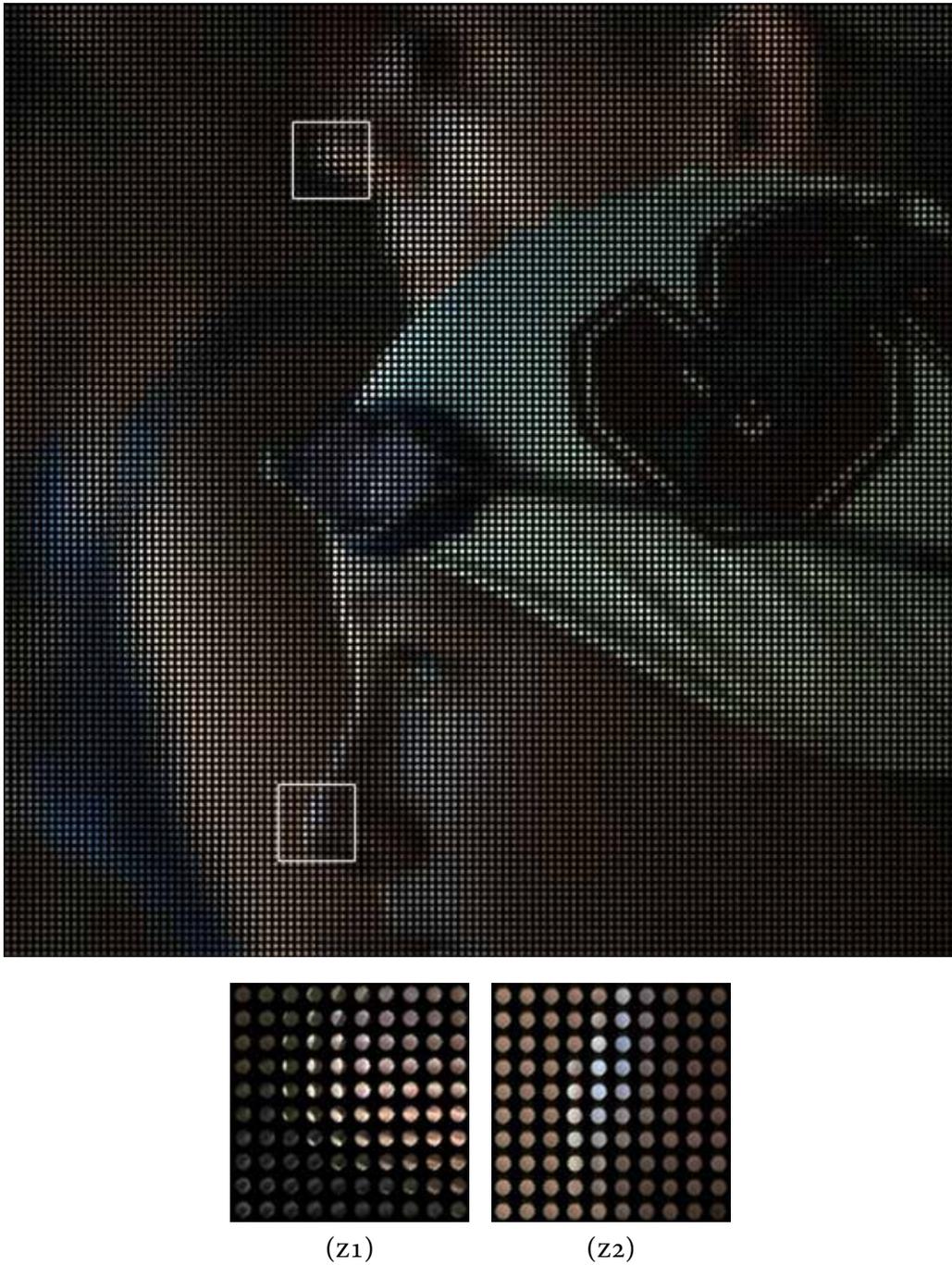
Figure 3.3: Raw light field photograph read off the photosensor underneath the microlens array. The figure shows a crop of approximately one quarter the full image so that the microlenses are clearly visible in print.

lighting incident upon each microlens. A second, equivalent interpretation is that the light field provides pictures of the scene from an array of viewpoints spread over the extent of the lens aperture. Since the lens aperture has a finite area, these different views provide some parallax information about the world. This leads to the third property of the light field—it provides information about the depth of objects in the scene.

The sections below highlight these different interpretations with three visualizations of the light field. Each of these visualizations is a flattening of the 4D light field into a 2D array of images. Moving across the array traverses two dimensions of the light field, and each image represents the remaining two dimensions of variation.

Raw Light Field Photograph

The simplest view of the recorded light field is the raw image of pixel values read off the photosensor underneath the microlens array, as shown in Figure 3.3. Macroscopically, the raw image appears like a conventional photograph, focused on the girl wearing the white cap, with a man and a girl blurred in the background. Looking more closely, the raw image is actually composed of an array of disks, where each disk is the image that forms underneath one microlens. Each of these microlens images is circular because it is a picture of the round aperture of the lens viewed from that position on the film. In other words, the raw light field photograph is an (x, y) grid of images, where each image shows us the light arriving at that film point from different (u, v) positions across the lens aperture.

The zoomed images at the bottom of Figure 3.3 show detail in the microlens images in two parts of the scene: the nose of the man in the background who is out of focus (Image z1), and the nose of the

Figure 3.4: Conventional photograph computed from the light field photograph in Figure 3.3.

girl in the foreground who is in focus (Image z2). Looking at Image z1, we see that the light coming from different parts of the lens are not the same. The light coming from the
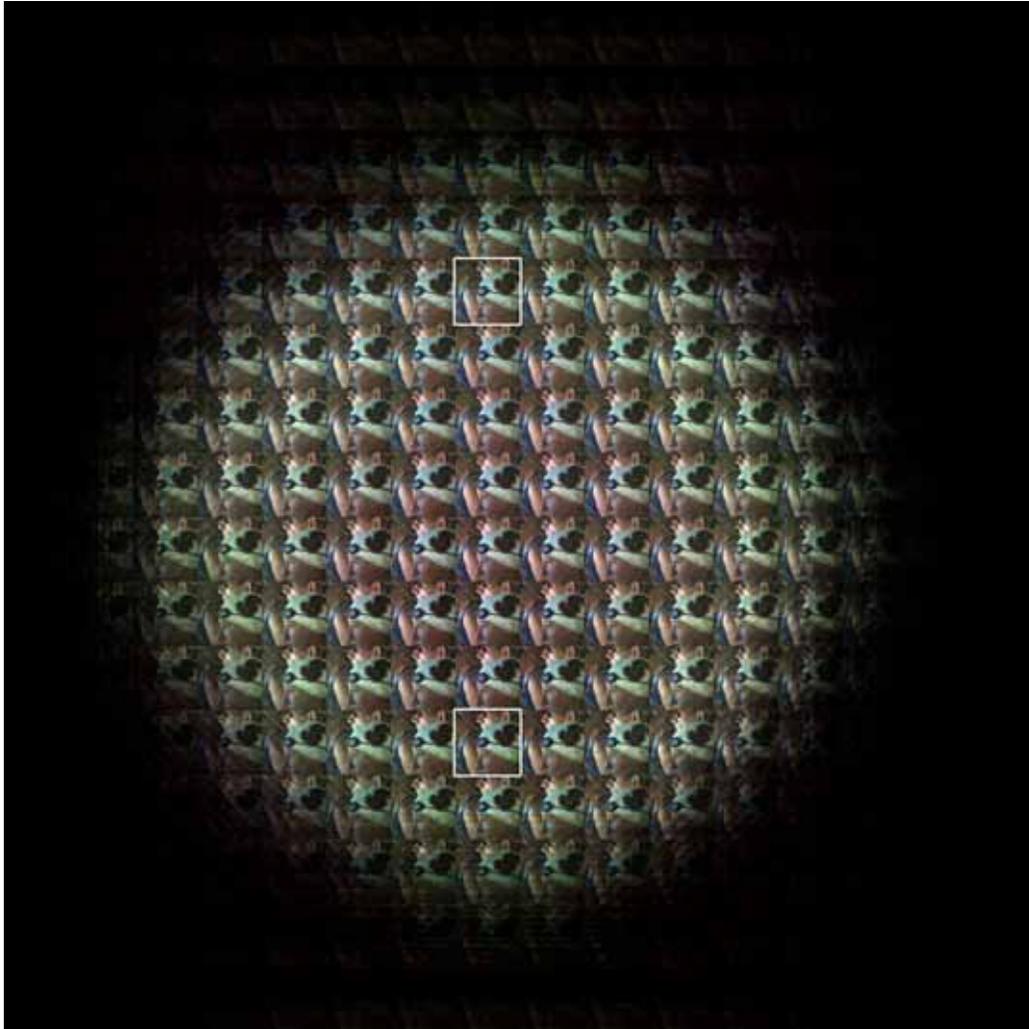
(z1)   (z2)

Figure 3.5: Sub-aperture images of the light field photograph in Figure 3.3. Images z1 and z2 are close-ups of the indicated regions at the top and bottom of the array, respectively.

left side of the microlens images originate on man's nose. The light coming from the right side originate in the behind the man. This effect can be understood by thinking of the rays tracing from different parts of the aperture out into the world. They pass through a point on the focal plane in front of the man, and diverge — some striking the man and others missing him to strike the background. In a conventional photograph, all the light would be summed up, leading to a blur around the profile of the man's nose as in Figure 3.4.

In contrast, the zoomed image of the girls' nose in Figure 3.3 z2 reveals microlens images of constant color. Since she is on the world focal plane, each microlens receives light coming from a single point on her face. Since her skin reflects light diffusely, that is, equally in all directions, all the rays have the same color. Although not all materials are diffuse, most reflect the same light over the relatively small extent of a camera aperture in most photographic scenarios.

Sub-Aperture Images

The second view of the recorded light field is an array of what I call its *sub-aperture images*, as shown in Figure 3.5. I computed this view by transposing the pixels in the raw light field photograph. Each sub-aperture image is the result of taking the same pixel underneath each microlens, at the offset corresponding to (u, v) for the desired sub-region of the main lens aperture.

Macroscopically, the array of images is circular because the aperture of the lens is circular. Indeed, each image in the array is a picture of the scene from the corresponding (u, v) position on the circular aperture. The zoomed images at the bottom of Figure 3.5 show sub-aperture images from the top and bottom of the lens aperture. Although these images look similar, examining the differences between the two images, as shown on the right, reveals a parallax shift between

Difference in pixel values between the two zoomed images in Figure 3.5.

the girl in the foreground and the man in the background. The man appears several pixels higher in Figure 3.5 z2 from the bottom of the array, because of the lower viewpoint.

This disparity is the source of blurriness in the image of the man in the conventional

(z1)   (z2)
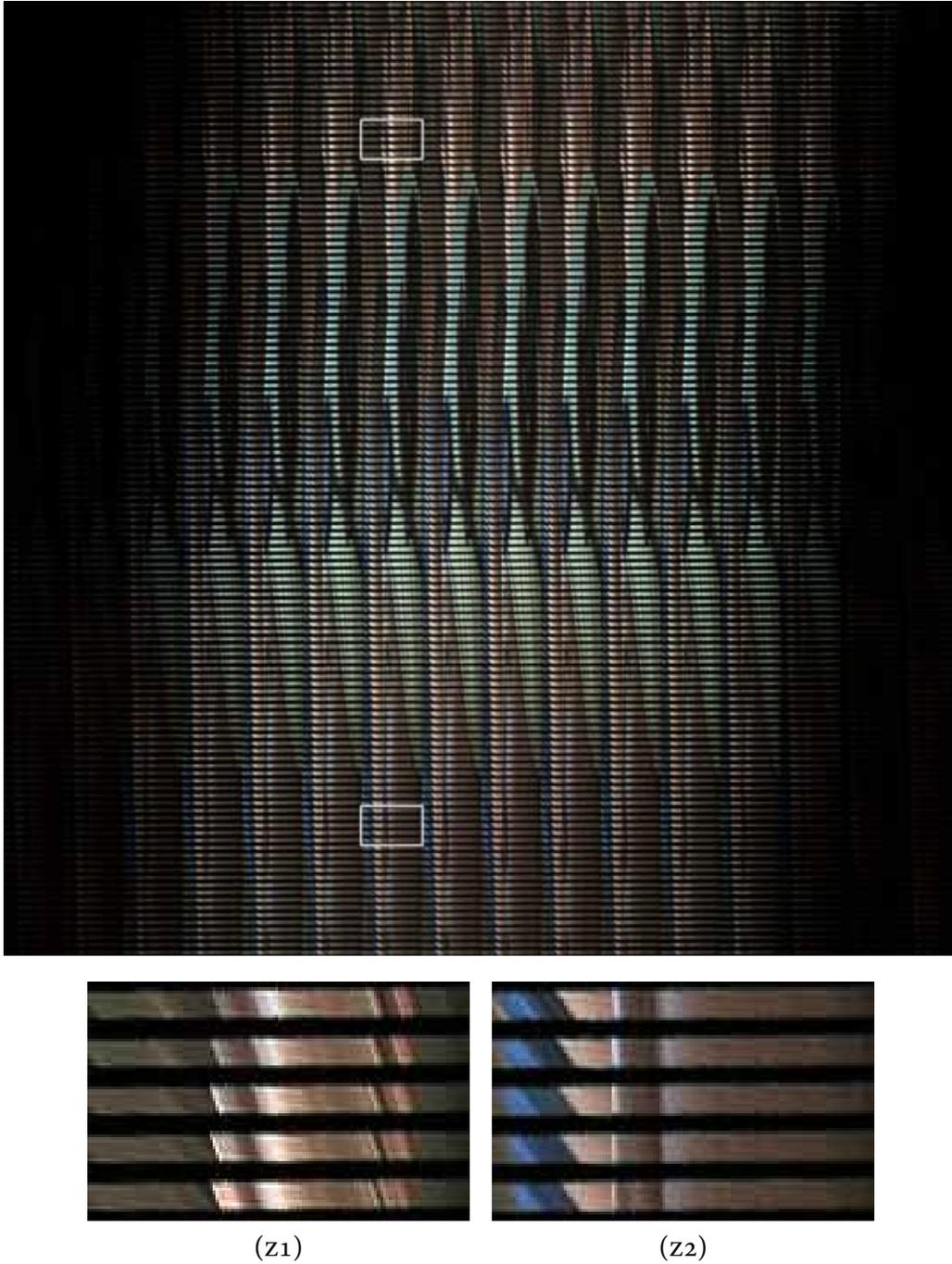
Figure 3.6: Epipolar images of the light field photograph in Figure 3.3. Images z1 and z2 are close-ups of the indicated regions at the top and bottom of the array, respectively.

photograph of Figure 3.4. Computing the conventional photograph that would have formed with the full lens aperture is equivalent to summing the array of sub-aperture images, summing all the light coming through the lens.

Epipolar Images

The third view is the most abstract, presenting an array of what are called *epipolar images* of the light field (see Figure 3.6). Each epipolar image is the 2D slice of the light field where y and v are fixed, and x and u vary. In Figure 3.6, y varies up the array of images and v varies to the right. Within each epipolar image, x increases horizontally (with a spatial resolution of 128 pixels), and u varies up the image (with a directional resolution of about 12 pixels). Thus, the zoomed images in Figure 3.6 show five (x, u) epipolar images, arrayed vertically.

These zoomed images illustrate the well-known fact that depth of objects in the scene can be estimated from the slope of lines in the epipolar images [Bolles et al. 1987; Forsyth and Ponce 2002]. The greater the slope, the greater the disparity as we move across u on the lens, indicating a further distance from the world focal plane. For example, the zoomed image of Figure 3.6 z2 corresponds to five rows of pixels in a conventional photograph that cut

Rows of pixels shown in epipolar images of Figures 3.6 z1 and z2.

through the nose of the girl in the foreground and the arm of the girl in blue in the background, as shown on the image on this page. In Image z2, the negative slope of the blue lines correspond to the further distance of the girl in blue. The vertical lines of the nose of the girl in the foreground show that she is on the focal plane. As another example, Figure 3.6 z1 comes from the pixels on the nose of the man in the middle ground. The intermediate slope of these lines indicates that the man is sitting between the two girls.

An important interpretation of these epipolar images is that they are graphs of the light field in the parameterization of the 2D ray-space diagrams such as Figure 3.2 B1. Figure 3.6 provides a database of such graphs for different (y, v) slices of the 4D light field.

3.4   Resolution Limits of the Plenoptic Camera

As we have seen, the size of the microlenses determines the spatial resolution of the light field sampling pattern. This section describes how to optimize the microlens optics to maximize the directional resolution, how diffraction ultimately limits the 4D resolution, and concludes with a succinct way to think about how to distribute samples in space and direction of a diffraction-limited plenoptic imaging system.

Microlens Optics

The image that forms under a microlens dictates the directional u resolution of the light field camera system. In the case of the plenoptic camera, optimizing the microlens optics to maximize the directional resolution means producing images of the main lens aperture that are as sharp and as large as possible.

    Maximizing the sharpness requires focusing the microlenses on the aperture plane[1] of the main lens. This may seem to require dynamically changing the separation between the photosensor plane and microlens array in order to track the aperture of the main lens as it moves during focusing and zooming. However, the microlenses are vanishingly small compared to the main lens (for example the microlenses are 280 times smaller than the main lens in the prototype camera described below), so regardless of its zoom or focus settings the main lens is effectively fixed at the microlenses' optical infinity. As a result, focusing the microlenses in the plenoptic camera means cementing the microlens plane one focal length away from the photosensor plane. This is a very convenient property, because it means that the light field sensor comprising the microlens array and the photosensor can be constructed as a completely passive unit if desired. However, there are benefits to dynamically changing the separation, as explored in Section 3.5.

    The directional resolution relies not only on the clarity of the image under each microlens, but also on its size. We want it to cover as many photosensor pixels as possible. The idea here is to choose the relative sizes of the main lens and microlens apertures so that the images are as large as possible without overlapping. This condition means that the f-numbers of the main lens and microlens array must be matched, as shown in Figure 3.7.

---

[1]Principal plane of the lens to be exact.

$f/2.8$ $f/4$ $f/8$
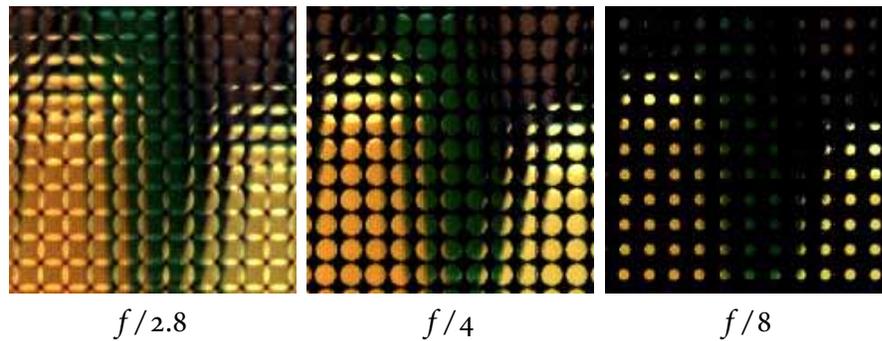
Figure 3.7: Close-up of microlens images formed with different main lens aperture sizes. The microlenses are $f/4$.

The figure shows that increasing or decreasing the main lens aperture simply increases or decreases the field of view of each microlens image. This makes sense because the microlens image is just a picture of the back of the lens. The f-number of the microlenses in these images is $f/4$. At $f/4$ the field of view are maximal without overlapping. When the aperture width is reduced by half to $f/8$, the images are too small, and resolution is wasted. When the main lens is opened up to $f/2.8$, the images are too large and overlap.

In this context, the f-number of the main lens is not simply its aperture diameter divided by its intrinsic focal length. Rather, we are interested in the *image-side f-number*, which is the aperture diameter divided by the separation between the principal plane of the main lens and the microlens plane. This separation is larger than the intrinsic focal length in general, when focusing on subjects that are relatively close to the camera.

As an aside, Figure 3.7 shows that a significant fraction of the photosensor is black because of the square packing of the microlens disks. The square packing is due to the square layout of the microlens array used to acquire the light field. The packing may be optimized by using different microlens array geometries such as a hexagonal grid, which is fairly common in integral cameras used for 3D imaging [Javidi and Okano 2002]. The hexagonal grid would need to be resampled onto a rectilinear grid to compute final images, but implementing this is easily understood using the ray-tracing approach of Section 3.2. A different approach to reducing the black regions would be to change the aperture of the camera's lens from a circle to a square, allowing tight packing with a rectilinear grid.

Diffraction-Limited Resolution

If we were to rely purely on ray-based, geometrical arguments, it might seem that arbitrarily high 4D resolution could be achieved by continually reducing the size of the microlenses and pixels. However, at sufficiently small scales the ray-based model breaks down and the wave nature of light must be considered. The ultimate limit on image resolution at small scales is diffraction. Rather than obtaining a geometrically sharp image on the imaging plane, the effect of diffraction is to blur the 2D signal on the imaging plane, which limits the effective resolvable resolution.

In the plenoptic camera, the diffraction blur reduces the clarity of the circular images formed on the photosensor under each microlens. Assuming that the microlenses are larger than the diffraction spot size, the blur appears as a degradation in the resolution along the *directional* axes of the recorded light field. In other words it corresponds to a vertical blurring of the ray-space diagram within each column of the light field sampling grid in Figure 3.8A, for example.

Classical wave optics predicts that the micron size of the diffraction blur on an imaging plane is determined by the aperture of highest f-number (smallest size) in the optical train [Hecht 2002; Goodman 1996]. This is based on the principle that light spreads out more in angle the more it is forced to pass through a small opening. The exact distribution of the diffraction blur depends on such factors as the shape of the aperture, the wavelength of light, and whether or not the light is coherent. Nevertheless, the dominant sense of scale is set by the f-number. Assuming that the highest f-number aperture in the optical system is f/n, then a useful rule of thumb is that the blur spot size (hence resolvable resolution) is roughly n microns on the imaging plane.

Let us apply this rule of thumb in considering the design of spatial and angular resolutions in a hypothetical light field camera. Assume that the lens and microlenses are $f/2$ and diffraction limited, and that the sensor is a standard 35 mm format-frame measuring 24 mm × 36 mm. The f-number means that the diffraction-limited resolution is roughly 2 microns, so let us assume that the sensor will contain pixels that are 2 microns wide. This provides a raw sensor resolution of 18,000 × 12,000.

Our control over the final distribution of spatial and angular resolution is in the size of the microlenses. If we choose a microlens size of 20 microns that covers 10 pixels, then we

will obtain a spatial microlens resolution of 1800×1200 (roughly 2.2 MP) with a directional resolution of about 10×10. Alternatively, we could choose a smaller microlens size, say 10 microns, for roughly 8.6 MP of spatial resolution, with just 5×5 directional resolution.

3.5 Generalizing the Plenoptic Camera

As discussed in the previous section, the plenoptic camera model requires the microlenses to focus on the aperture of the main lens, which means that the photosensor is fixed at the focal plane of the microlenses. This section introduces a generalization of the camera model to allow varying the separation between the microlenses and sensor from one focal length down to zero as shown in Figure 3.8 on page 38. The motivating observation is that a magnifying glass stops magnifying when it is pressed flat against a piece of paper. This intuition suggests that we can effectively transform the plenoptic camera back into a conventional camera by pressing the photosensor against the microlens surface.

On first consideration, however, this approach may seem problematic, since decreasing the separation between the microlenses and photosensor corresponds to defocusing the images formed by the microlenses. In fact, analysis of microlens focus in integral photography generally does not even consider this case [Arai et al. 2003]. The macroscopic analogue would be as if one focused a camera further than infinity (real cameras prevent this, since nothing would ever be in focus in this case). I am not aware of any other researchers who have found a use for this kind of camera in their application domain.

Nevertheless, defocusing the microlenses in this way provides tangible benefits for our application in digital photography. Figure 3.8 illustrates that each level of defocus provides a different light field sampling pattern, and these have different performance trade-offs. Figure 3.8A illustrates the typical plenoptic separation of one microlens focal length, where the microlenses are focused on the main lens. This configuration provides maximal directional u resolution and minimal x spatial resolution on the ray-space.

Figure 3.8B illustrates how this changes when the separation is halved. The horizontal lines in the ray-space sampling pattern tilt, becoming more concentrated vertically and less so horizontally. This effect can be intuitively understood in terms of the shearing of the light field when focusing at different depths. In Chapter 2, this shearing was described in
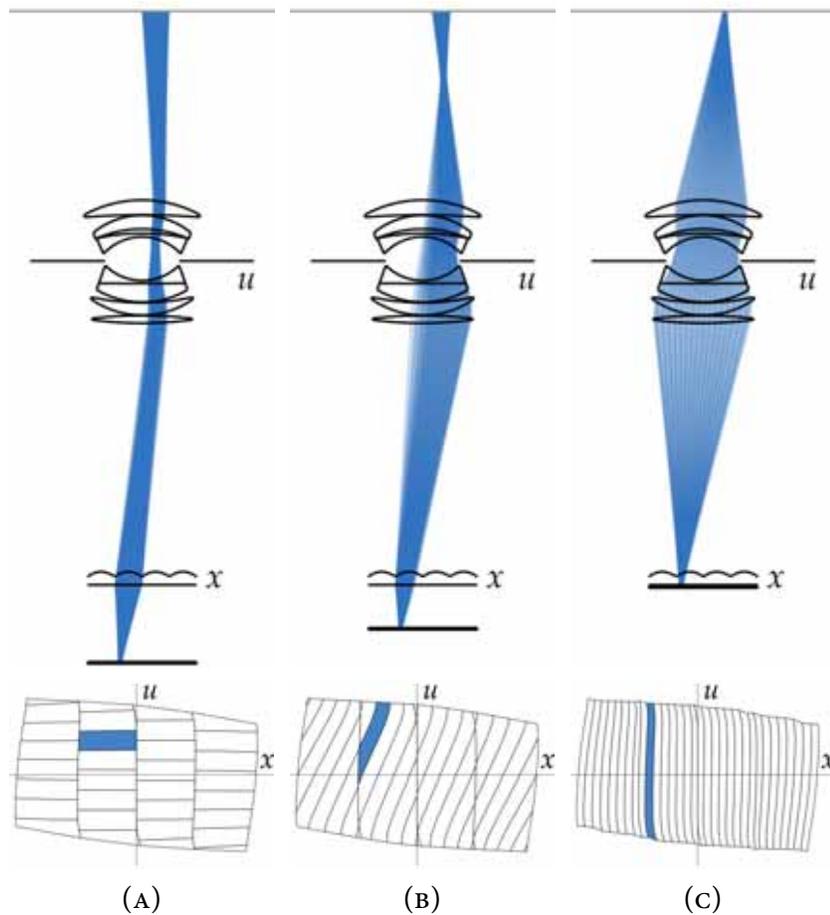
Figure 3.8: Generalized light field camera: changes in the light field sampling pattern due to reducing the separation between the microlenses and photosensor.

terms of focus of the main lens. Here the focus change takes place in the microscopic camera composed of each microlens and its patch of pixels, and the shearing takes place in the microscopic light field inside this microlens camera. For example, Figure 3.8A shows 4 such microlens cameras. Their microscopic light fields correspond to the 4 columns of the ray-space diagram, and the sampling pattern within these columns is what shears as we change the focus of the microlenses in Figure 3.8B and C.

Figure 3.8C illustrates how the ray-space sampling converges to a pattern of vertical columns as the separation between microlenses and photosensor approaches zero. In this

case the microlenses are almost completely defocused, and we obtain no directional resolution, but maximal spatial resolution. That is, the values read off the photosensor for zero separation approach the values that would appear in a conventional camera in the absence of the microlens array.

(A)   (B)   (C)
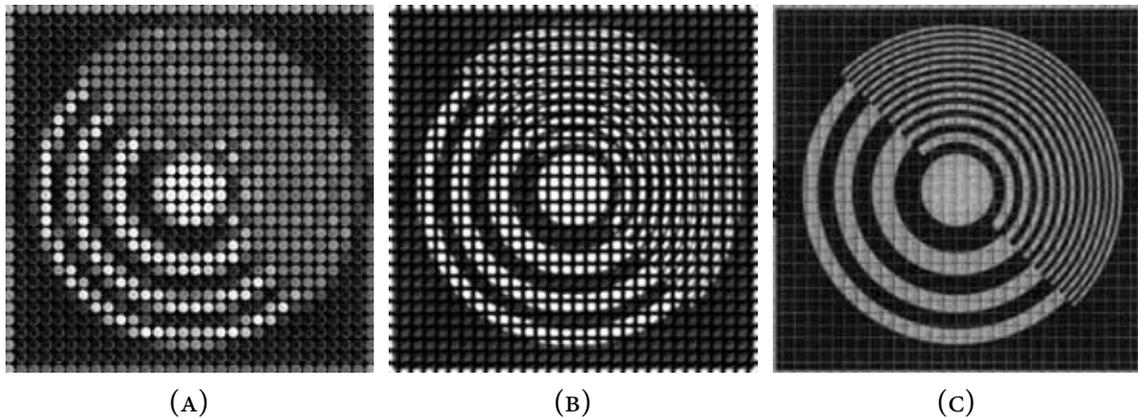
Figure 3.9: Simulated raw image data from the generalized light field camera under the three configurations of microlens focus in Figure 3.8.

Figure 3.9 shows a portion of simulated raw light field photographs of a resolution chart for the three configurations of the generalized camera shown in Figure 3.8. These images illustrate how the microlens images evolve from sharp disks of the back of the main lens at full separation (Figure 3.9A), to filled squares of the irradiance at the microlens plane when the separation reaches zero (Figure 3.9C). Notice that the finer rings in the resolution chart only resolve as we move towards smaller separations.

The prototype camera provides manual control over the separation between the microlenses and photosensor. It enables a study of the performance of the generalized light field camera. An analysis of its properties is the topic of Chapter 6.

3.6 Prototype Light Field Camera

We had two goals in building a prototype light field camera. The first was to create a device that would allow us to test our theories and simulations regarding the benefits of recording and processing light field photographs. Our second goal was to build a camera that

could be used as much like an ordinary camera as possible, so that we would be able to explore applications in a range of traditional photographic scenarios, such as portraiture, sports photography, macro photography, etc.

Our overall approach was to take an off-the-shelf digital camera and attach a microlens array in front of the photosensor. The advantage of this approach was simplicity. It allowed us to rapidly prototype a working light field camera by leveraging the mechanical and electronic foundation in an existing system. A practical issue with this approach was the restricted working volume inside the camera body. Most digital cameras are tightly integrated devices with little room for adding custom mechanical parts and optics. As discussed below, this issue affected the choice of camera and the design of the assembly for attaching the microlens array in front of the photosensor.

The main disadvantage of using an off-the-shelf camera was that we were limited to a prototype of modest resolution, providing final images with 296×296 pixels, with roughly 12×12 directional resolution at each pixel. The reason for this is that essentially all existing sensors are designed for conventional imaging, so they provide only a moderate number of pixels (e.g. 10 MP) that match current printing and display resolutions. In contrast, the ideal photosensor for light field photography is one with a very large number (e.g. 100 MP) of small pixels, in order to match the spatial resolution of conventional cameras while providing extra directional resolution. As discussed in the introductory chapter, such high sensor resolutions are theoretically possible in today's VLSI technology, but it is not one of the goals of this thesis to address the problem of constructing such a sensor.

Components

The issues of resolution and working volume led us to choose a medium format digital camera as the basis for our prototype. Medium format digital cameras provide the maximum sensor resolution available on the market. They also provide easiest access to the sensor because the digital "back," which contains the sensor, detaches completely from the body, as shown in Figure 3.10B.

Our digital back is a Megavision FB4040. The image sensor that it contains is a Kodak KAF-16802CE color sensor, which has effectively 4096×4096 pixels that are 9.25 microns wide. For the body of our camera we chose a medium-format Contax 645. We used a variety

(A)   (B)   (C)

Figure 3.10: The medium format digital camera used in our prototype.

of lenses, including a 140 mm $f/2.8$ and an 80 mm $f/2.0$. The wide maximum apertures on these lenses meant that, even with extension tubes attached for macro photography, we could achieve an $f/4$ image-side f-number to match the f-number of the microlenses.

In choosing our microlens array we focused on maximizing the directional resolution, in order to best study the post-processing advantages of having this new kind of information. This goal meant choosing microlenses that were as large as possible while still allowing us to compute final images of usable resolution.

We selected an array that contains 296×296 lenslets that are 125 microns wide (Figure 3.11). Figure 3.11C is a micrograph showing that the microlenses are square shaped, and densely packed in a rectilinear array. The fill-factor is very close to 100%. The focal length of the microlenses is 500 microns; their f-number is $f/4$.

(A)   (B)   (C)
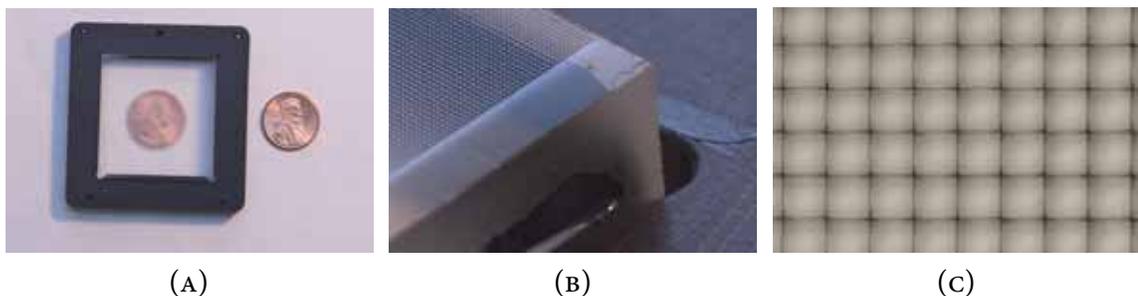
Figure 3.11: The microlens array in our prototype camera.

A good introduction to microlens fabrication technology can be found in Dan Daly's book [2001], although the state of the art has advanced rapidly and a wide variety of different

techniques are used in the industry today. Our microlens array was made by Adaptive Optics Associates (part 0125-0.5-s). They first manufactured a nickel master for the array using a diamond-turning process to carve out each microlens. The master was then used to mold our final array of resin lenses on top of a square glass window.

Assembly

The two main design issues in assembling the prototype were positioning the microlenses close enough to the sensor surface, and avoiding contact with the shutter of the camera body.

The required position of the microlenses is 500 microns away from the photosensor surface: the focal length of the microlenses. This distance is small enough that we were required to remove the protective cover glass that usually seals the chip package. While this was a challenge for us in constructing this prototype, it illustrates the desirable property that a chip for a light field sensor is theoretically no larger than the chip for a conventional sensor.

To avoid contact with the shutter of the camera we had to work within a volume of only a few millimeter separation, as can be seen in Figure 3.10B. This is the volume within which we attached the microlens array. To maximize this volume, we removed the infrared filter that normally covers the sensor. We replaced it with a circular filter that screws on to the front of the main camera lens.

Figure 3.12 illustrates the assembly that we used to control the separation between the microlens array and the photosensor. We glued the microlens array window to a custom lens holder, screwed a custom base plate to the digital back over the photosensor, and then attached the lens holder to the base plate with three screws separated by springs. Adjusting the three screws provided control over separation and tilt. The bottom of Figure 3.12 shows a cross-section through the assembled parts.

We chose adjustment screws with 56 threads per inch to ensure sufficient control over the separation. We found that adjusting the screws carefully provided a mechanical resolution of roughly 10 microns. The accuracy required to focus the microlens images accurately is on the order of 70-80 microns, which is the maximum error in separation for which the circle of confusion falls within one microlens pixel. The formula for this error is just the usual depth of focus [Smith 2005] given by twice the product of the pixel width (9.25 micron) and the f-number of the microlenses ($f/4$).
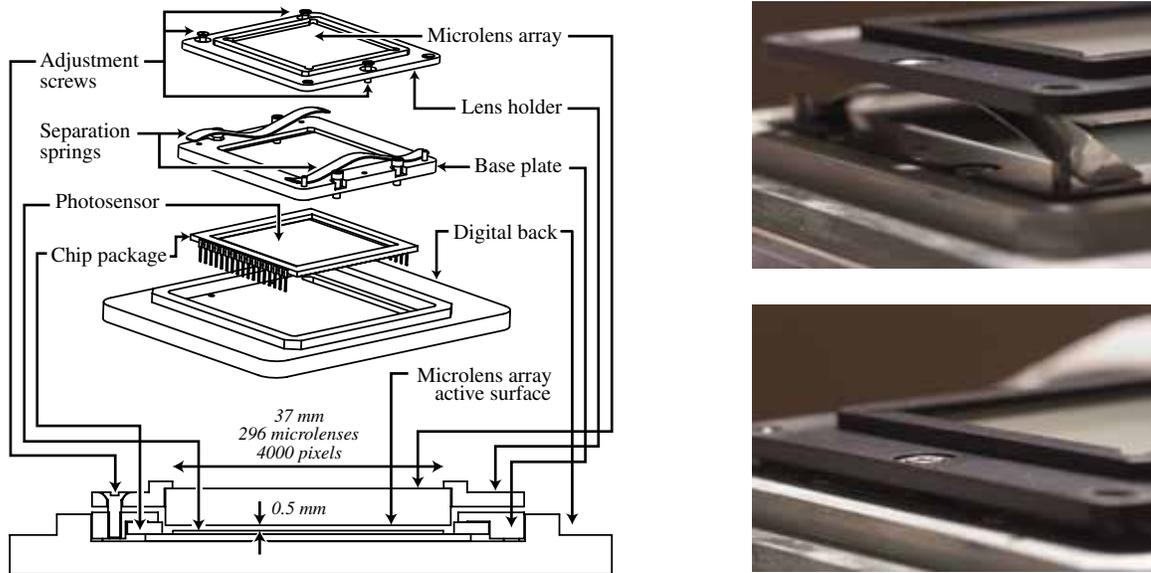
Figure 3.12: *Left*: Schematic of light field sensor assembly for attaching microlens array to digital back. Exploded view on top, and cross-sectional view on bottom. *Right*: Photographs of the assembly before and after engaging adjustment screws.

To calibrate the separation at one focal length for the plenoptic camera configuration, we took images of a pin-hole light source with the bare light field sensor. The image that forms on the sensor when illuminated by such a light is an array of sharp spots (one under each microlens) when a focal length of one microlens is achieved. The procedure took 10–20 iterations of screw adjustments. The separation was mechanically stable after calibration and did not require re-adjustments.

We created a high contrast pin-hole light source by stopping down the 140 mm main lens to its minimum aperture and attaching 78 mm of extension tubes. This created an aperture of approximately $f/50$, which we aimed at a white sheet of paper. The resulting spot images subtend about 1 sensor pixel underneath each microlens image.

Operation

Unlike most digital cameras, the Megavision digital back does not read images to on-camera storage such as Flash memory cards. Instead, it attaches to a computer by Firewire IEEE-1394,

and reads images directly to the computer. We store raw light field photographs and use custom software algorithms, as described in later chapters, to process these raw files to produce final photographs.

Two subsystems of our prototype require manual adjustments where an ideal implementation would use electronically-controlled motors. The first is control over the separation between the microlens array and the photosensor. To test the performance of the generalized camera model in Chapter 6, we configured the camera with different separations by manually adjusting the screws in the lens holder for the microlens array.

The second system requiring manual control is the aperture size of the main lens relative to the focal depth. In the plenoptic camera, the optimum choice of aperture size depends on the focal depth, in contrast to a conventional camera. The issue is that to maintain a constant image-side f-number of $f/4$, the aperture size must change proportional to the separation between the main lens and the imaging plane. This separation changes when focusing the camera. For example, consider an aperture width that is image-side $f/4$ when the camera is focused at infinity (with a separation of one focal length). The aperture must be doubled when the camera is focused close-by in macro photography to produce unit magnification (with a separation of two focal lengths).

3.7   Related Work and Further Reading

Integral Photography, Plenoptic Cameras and Shack-Hartmann Sensors

Over the course of the last century, cameras based on integral photography techniques [Lippmann 1908; Ives 1931] have been used in many fields, under different names. In the area of 3D imaging, many "integral camera" variants have been studied [Okoshi 1976; Javidi and Okano 2002]. In the last two decades especially, the advent of cheap digital sensors has seen renewed interest in using such cameras for 3D imaging and stereo display applications [Davies et al. 1988; Okano et al. 1999; Naemura et al. 2001; Yamamoto and Naemura 2004]. This thesis takes a different approach, focusing on enhanced post-processing of ordinary 2D photographs, rather than trying to provide the sensation of 3D viewing.

Georgiev et al. [2006] have also explored capturing directional ray information for enhanced 2D imaging. They use an integral camera composed of a coarse array of lenses and

prisms in front of the lens of an ordinary camera, to capture an array of sub-aperture images similar to the one showed in Figure 3.5. Although their angular resolution is lower than used here, they interpolate in the directional space using algorithms from computer vision.

A second name for integral cameras is the "plenoptic camera" that we saw earlier in this chapter. Adelson and Wang introduced this kind of camera to computer vision, and tried to estimate the shape of objects in the world from the stereo correspondences inherent in its data [Adelson and Wang 1992]. In contrast, our application of computing better photographs turns out to be a fundamentally simpler problem, with a correspondingly more robust solution. As an extreme example, it is impossible to estimate the depth of a red card in front of a red wall using the plenoptic camera, but it is easy to compute the correct (all-red) photograph. This is a general theme that the computer graphics and computer vision communities have discovered over the last thirty years: programming a computer to interpret images is very hard, but programming it to make images turns out to be relatively easy.

A third name for integral cameras is the "Shack-Hartmann sensor" [Shack and Platt 1971; Platt and Shack 2001]. It is used to measure aberrations in optical systems by analyzing what happens to a laser beam as it passes through the optics and the microlens array. In astronomy, Shack-Hartmann sensors are used to measure distortions in the atmosphere, and deformable mirrors are used to optically correct for these aberrations in real-time. Such adaptive optics took about twenty years from initial proposal [Babcock 1953] to the first working systems [Hardy et al. 1977]), but today they appear in almost all new land-based telescopes [Tyson 1991]. Adaptive optics have also been applied to compensate for the aberrations of the human eye, to produce sharp images of the retina [Roorda and Williams 1999]. In opthalmology, Shack-Hartmann sensors are used to measure aberrations in the human eye for refractive surgery planning [Liang et al. 1994; Haro 2000]. Chapter 7 gives these approaches a twist by transforming the Shack-Hartmann sensor from an optical analysis tool into a full-fledged imaging device that can digitally correct final images collected through aberrated optics.

Other Related Optical Systems

To acquire light fields, graphics researchers have traditionally taken many photographs sequentially by stepping a single camera through known positions [Levoy and Hanrahan 1996;

Isaksen et al. 2000]. This is simple, but slow. Instantaneous capture, allowing light field videography, is most commonly implemented as an array of (video) cameras [Yang et al. 2002; Wilburn et al. 2005]. Sampling the light field via integral photography can be thought of as miniaturizing the camera array into a single device. This makes acquisition as simple as using an ordinary camera, but sacrifices the large-baseline and flexibility of the array.

I would like to make comparisons between the light field camera and three other optical systems. The first is the modern, conventional photosensor array that uses microlenses in front of every pixel to concentrate light onto the photosensitive region [Ishihara and Tanigaki 1983; Gordon et al. 1991; Daly 2001]. One can interpret the optical design in this paper as an evolutionary step in which we use not one detector underneath each microlens, but rather an array of detectors capable of forming an image.

The second comparison is to artificial compound eye sensors (insect eyes) composed of a microlens array and photosensor. This is akin to the light field sensor in our prototype without a main lens. The first 2D version of such a system appears to have been built by Ogata et al. [1994], and has been replicated and augmented more recently using updated microlens technology [Tanida et al. 2001; Tanida et al. 2003; Duparré et al. 2004]. These projects endeavor to flatten the traditional camera to a plane sensor, and have achieved thicknesses as thin as a sheet of paper. However, the imaging quality of these optical designs is fundamentally inferior to a camera system with a large main lens; the resolution past these small lens arrays is severely limited by diffraction, as noted in comparative studies of human and insect eyes [Barlow 1952; Kirschfeld 1976].

As an aside from the biological perspective, it is interesting to note that our optical design can be thought of as taking a human eye (camera) and replacing its retina with an insect eye (microlens / photosensor array). No animal has been discovered that possesses such a hybrid eye [Land and Nilsson 2002], but this dissertation, and the work of Adelson and Wang, shows that such a design possesses unique and compelling capabilities when coupled with sufficient processing power.

The fourth comparison is to holographic stereograms, which are holograms built up by sequentially illuminating from each direction with the appropriate view. Halle has studied such systems in terms of discrete sampling [Halle 1994], identifying the cause of aliasing artifacts that are also commonly seen in multi-camera light field sampling systems. The

similarities are more than passing, and Camahort showed that holographic stereograms can be synthesized from light fields [Camahort 2001]. In fact, I have worked with Zebra Imaging to successfully print a hologram from one of the light fields taken with our prototype light field camera. These results open the door to single-shot holography using light field cameras, although the cost of printing is still an issue.

4

Digital Refocusing

This chapter explores the simplest application of recording the light field: changing the focus of output photographs after a single exposure. In Figure 4.1, A1-A5 are five photographs computed from the raw light field discussed in Section 3.3. That is, these five images were computed from a single 1/125 second exposure of the prototype light field camera. They show that we can refocus on each person in the scene in turn, extracting a striking amount of detail that would have been irretrievably lost in a conventional photograph. Image A3, focused on the girl wearing the white cap, is what I saw in the camera viewfinder when I clicked the shutter. It is the photograph a conventional camera would have taken.

People who see these images for the first time are often surprised at the high fidelity of such *digital refocusing* from light fields. Images A1-A5 look like the images that I saw in the viewfinder of the camera as I turned the focus ring to focus on the girl in the white cap. The underlying reason for this fidelity is that digital refocusing is based on a physical simulation of the way photographs form inside a real camera. In essence, we imagine a camera focused at the desired depth of the output photograph, and simulate the flow of rays within this virtual camera. The software simulates a real lens, tracing the recorded light rays to where they would have terminated in the virtual camera, and simulates a real sensor, summing the light deposited at each point on the virtual imaging plane.

Figure 4.1B is a different kind of computed photograph, illustrating *digitally extended depth of field*. In this image, every person in the scene is in focus at the same time. This is the image that a conventional camera would have produced if we had reduced the size of the

Figure 4.1: Examples of refocusing (A1–A5) and extended depth of field (B).

lens aperture in the classical photographic method of optically extending the depth of field. Image B was computed by combining the sharpest portions of images A1-A5 [Agarwala et al. 2004], and can be thought of as refocusing each pixel at the depth of the closest object in that direction.

A crucial advantage of the digitally extended depth of field photograph over the classical photograph is that the former uses the light coming through a larger lens aperture. This means that recording a light field to obtain high depth of field captures light more efficiently, allowing less grainy images with higher signal to noise ratio (SNR). Sections 4.3–4.5 study these improvement theoretically and through numerical experimentation, demonstrating linear improvement with the directional u resolution.

This chapter assumes that the light fields are recorded with a light field camera configured in the plenoptic configuration that provides maximum directional resolution. This simplification enables easier discussion of the refocusing algorithm and performance characteristics. The added complexity of general light field camera configurations is considered in Chapter 6.

4.1 Previous Work

Refocusing

Isaksen, McMillan and Gortler [2000] were the first to demonstrate virtual refocusing from light fields. It was proposed in the original graphics paper on light field rendering [Levoy and Hanrahan 1996], and sometimes goes by the name of synthetic aperture photography in more recent work [Yang et al. 2002; Levoy et al. 2004; Vaish et al. 2004; Wilburn et al. 2005]. Vaish et al. [2005] consider the interesting variation of a tilted focal plane, such as one might achieve in a view camera where the film plane may be set at an angle to the main lens [Adams 1995].

These demonstrations of refocusing suffer from two problems, however. First, it is difficult to capture the required light field datasets, requiring lengthy scanning with a moving camera, or large arrays of cameras that are not suitable for the spontaneous shooting that we associate with regular photography. Second, the results tend to exhibit high aliasing in blurred regions due to incomplete sampling of the virtual lens aperture (e.g. due to gaps between cameras). Light field photography addresses both these issues: the light field camera is as easy to use as a conventional hand-held camera. In addition, the optical design reduces aliasing drastically by integrating all the rays of light passing through the aperture.

A more theoretical point is that the method described here is the first attempt to accurately formulate refocusing in terms of the image formation process that occurs inside a real camera. This was probably of greater interest to me because this dissertation shares a much closer relationship with conventional photography than previous work. In the past, the relationship was much more abstract (the virtual lens diameters were sometimes over a meter across), and the practitioners' primary goal was simply to qualitatively reproduce the visual effect of finite depth of field.

Focus and Depth from Defocus and Auto-Focus

Other researchers have studied the problem of trying to refocus from two images focused at different depths [Subbarao et al. 1995; Kubota and Aizawa 2005]. These methods are usually based on a class of computer vision algorithms called *depth from defocus*, which estimate the depth of objects based on the relative blur in two images focused at different depths [Krotkov 1987; Pentland 1987; Subbarao 1988]. Depth from defocus eventually led to impressive systems for estimating depth from video in real-time [Nayar et al. 1995; Pentland et al. 1994]. However the systems for refocusing from two images were less successful. Although they could generate reasonable images for virtual focal depths close to the optical focal planes in the input images, artifacts quickly increased and resolved detail decreased at further depths. The fundamental problem is that mis-focus is a low-pass filter that powerfully truncates high-frequency detail for objects at increasing distances from the optical focal plane. Basic signal processing principles make it clear that it is unrealistic to recover the attenuated high frequencies with good fidelity.

Extended Depth of Field

Like the algorithms discussed later in this chapter, *Wavefront Coding* system of Dowski and Johnson [1999] can be used to decouple the trade-off between aperture size and depth of field. In their system they use aspheric lenses that produce images with a depth-independent blur. Deconvolution of these images retrieves image detail at all depths. In contrast, as described below, the approach in light field photography is to record a light field, process it to refocus at all depths, and combine the sharpest parts of these images to extend the depth of field, albeit with a trade-off in image resolution or amount of data collected. Dowski and Johnson's method may permit higher resolution at the expense of noise at edge boundaries, but light field photography avoids the problems of deconvolving blurry images and provides greater flexibility in image formation.

The algorithms for extending the depth of field described below rely on methods of of producing extended depth of field images from a set of images focused at different depths. This problem has been studied by many researchers, including Ogden et al. [1985], Haeberli [1994], Eltoukhy and Kavusi [2003] and Agarwala et al. [2004] (whose software I use).

More recently there has been some attention towards maximizing depth of field in classical light field rendering [Takahashi et al. 2003].

4.2 Image Synthesis Algorithms

The first half of this section describes how to compute refocused photographs from the light field by numerical integration of the imaging equation derived in Chapter 2. The second half describes how to use collections of refocused images to extend the depth of field so that the entire image is as sharp as possible.

Refocusing

The ideal set of rays contributing to a pixel in digital refocusing is the set of rays that converge on that pixel in a virtual conventional camera focused at the desired depth. Chapter 2 derived Equation 2.3, which formally specifies in its integral the set of light rays for the pixel at position (x', y'):

$$E_{(\alpha \cdot F)}(x', y') = \frac{1}{\alpha^2 F^2} \int \int L_F \left(u(1-1/\alpha) + x'/\alpha, \ v(1-1/\alpha) + y'/\alpha, \ u, \ v \right) \, du \, dv. \quad (4.1)$$

Recall that in this equation L_F is the light field parameterized by an xy plane at a depth of F from the uv lens plane, α is the depth of the virtual film plane relative to F, and $E_{(\alpha \cdot F)}$ is the photograph formed on virtual film at a depth of $(\alpha \cdot F)$

One way to evaluate this integral is to apply numerical quadrature techniques, such as sampling the integrand for different values of u and v and summing them. The ray-tracing procedure described in Section 3.2 is used to evaluate the integrand for these different samples of u and v. To idea is to trace the ray $(u(1-1/\alpha) + x'/\alpha, \ v(1-1/\alpha) + y'/\alpha, \ u, \ v)$ through the microlens array and down to the photosensor. The intersection point is where the ray deposited its energy in the camera during the exposure, and the value of L_F is estimated from the photosensor values near this point.

However, a more efficient method is suggested by the linearity of the integral with respect to the underlying light field. Examining Equation 4.1 reveals the important observation that refocusing is conceptually a summation of dilated and shifted versions of the sub-aperture

images over the entire uv aperture.

This point is made clearer by explicitly defining the sub-aperture image at lens position (u, v) in the light field L_F. Let us represent this sub-aperture image by the 2D function $L_F^{(u,v)}$, such that the pixel at position (x, y) in the sub-aperture image is given by $L_F^{(u,v)}(x, y)$. With this notation, we can re-write Equation 4.1 as:

$$E_{(\alpha \cdot F)}(x', y') = \frac{1}{\alpha^2 F^2} \iint L_F^{(u,v)} \left(u(1-1/\alpha) + x'/\alpha, \ v(1-1/\alpha) + y'/\alpha \right) \ du \, dv, \qquad (4.2)$$

where $L_F^{(u,v)} \left(u(1-1/\alpha) + x'/\alpha, \ v(1-1/\alpha) + y'/\alpha \right)$ is simply the sub-aperture image $L_F^{(u,v)}$, dilated by a factor of α and shifted by a factor of $(u(1-1/\alpha), v(1-1/\alpha))$. In other words, digital refocusing can be implemented by shifting and adding the sub-aperture images of the light field. This technique has been applied in related work on *synthetic aperture imaging* using light fields acquired with an array of cameras [Vaish et al. 2004; Levoy et al. 2004]. Let us take a closer look at the dilation and the shift in the present case of refocusing ordinary photographs.

The dilation factor α in Equation 4.2 actually plays no real part in computing final images—it can simply be ignored. The reason for this is that the dilation factor is the same for all images. Scaling all images and summing does not change the ultimate output resolution of the synthesized photograph.

The real meaning of the dilation factor has to do with the property that digital refocusing does not alter the field of view on the subject. With most photographic lenses, optically focusing the lens closer causes the field of view of the subject to decrease, because the magnification of the subject increases. In contrast, with digital refocusing the field of view stays constant regardless of chosen focal depth, because the synthesized image has the field of view of the original sub-aperture images. Only the focus changes. In other words, the dilation factor in Equation 4.2 represents the dilation in field of view relative to the field of view that would have been obtained if we had optically focused the lens at the desired depth. With respect to the characteristic of changing focus without changing magnification, digital refocusing is functionally similar to a *telecentric* lens [Smith 2000], although the underlying causes are quite different.

The shift, $(u(1-1/\alpha), v(1-1/\alpha))$, of each sub-aperture image in Equation 4.2 increases

(A): No refocus   (B): Refocus closer   (C): Refocus further
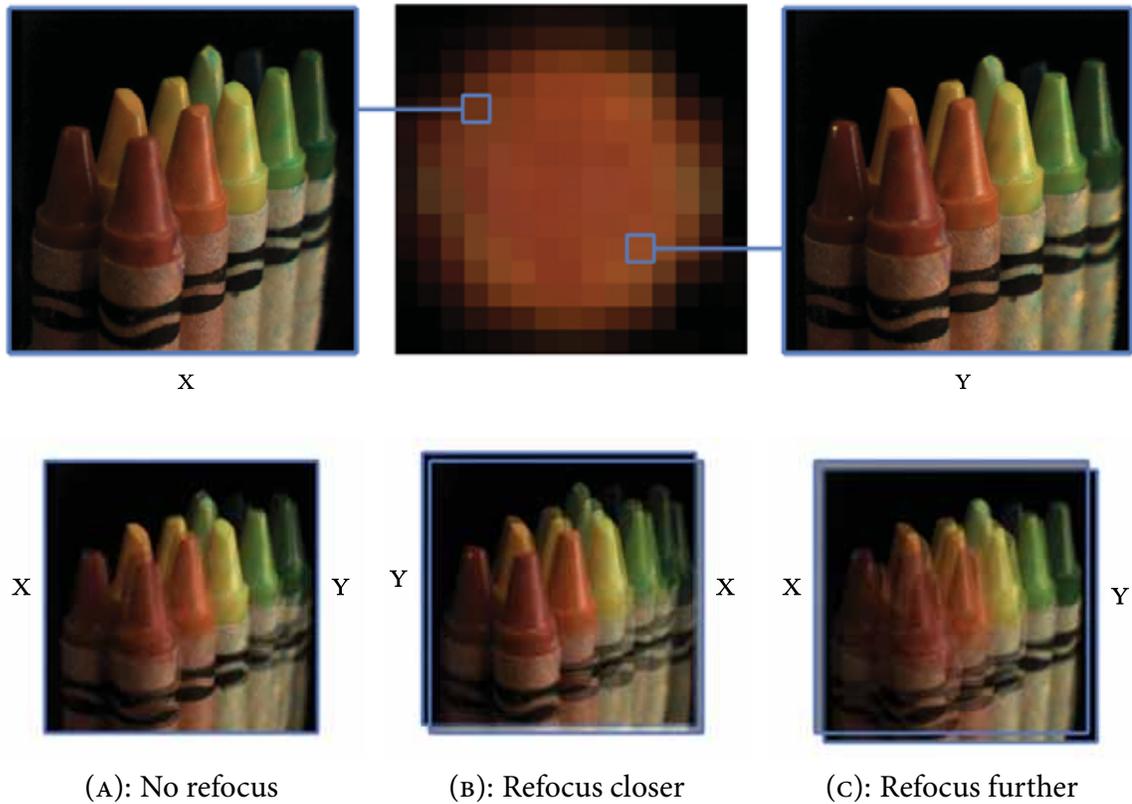
Figure 4.2: Shift-and-add refocus algorithm, illustrated with just two sub-aperture images for didactic purposes.

with both the distance of the sub-aperture from the center of the lens (u, v), and the relative extent, α, to which we want to refocus away from the optical focal plane. Figure 4.2 visualizes the shifts for three different virtual film planes, summing just two sub-aperture images, X and Y, for illustrative purposes. Using two sub-aperture images causes out-of-focus regions to appear as twice-repeated edges rather than a uniform blur, making it easier to see the shift effect. Figure 4.2A corresponds to no refocusing, with $\alpha = 1$, and a shift of 0 for both sub-aperture images. The remaining two images show that the direction of shifts depends on whether we are focusing closer or farther to align features at the desired depth.

The minimal discretization of this algorithm is to shift-and-add just the 12×12 sub-aperture images in the raw light-field shown in Figure 3.5. For most applications, the quality of the resulting photographs is quite good. However, this process may generate undesired

(A): Unrefocused

(B): Sub-aperture   (C1): Undersampled, aliased   (C2): Adequately sampled
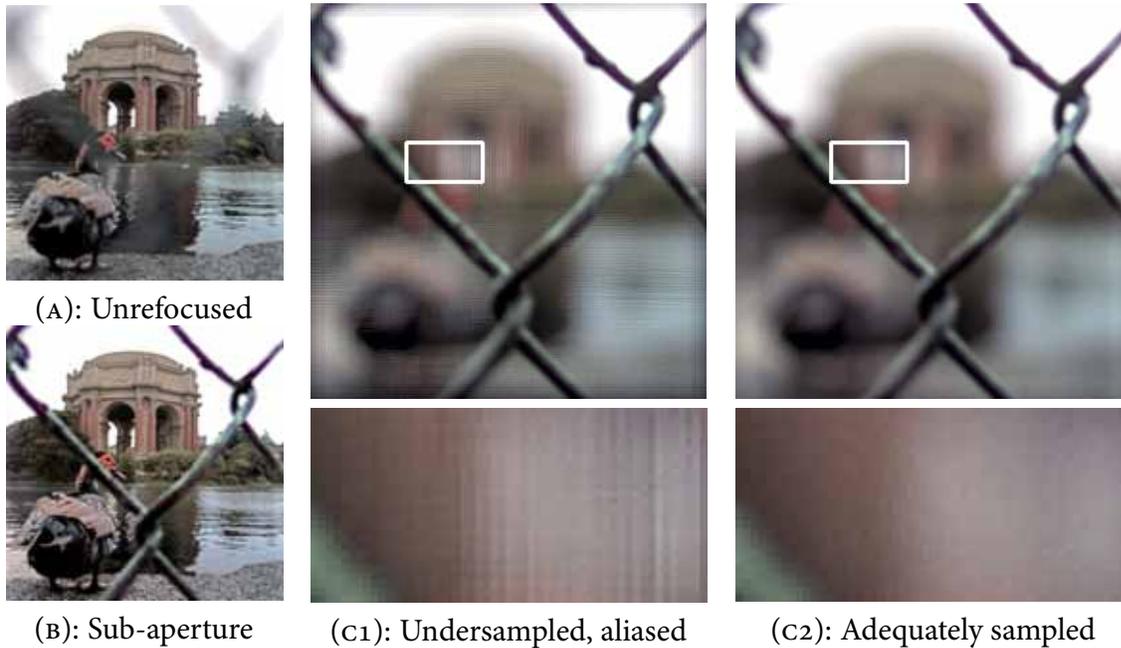
Figure 4.3: Aliasing of blurred regions in under-sampled shift-and-add refocus algorithm.

step edges in out-of-focus regions when focusing at greater distances from the optical focal plane. These kinds of artifacts can be seen in Figure 4.3-C1, which illustrates digital refocusing of a photograph onto a chain-link fence in the extreme foreground. The step-edge artifacts are visible as banding in the close-up of the out-of-focus region. While these artifacts are relatively subtle in still images, they become much more apparent in animations of continuous refocusing. For reference, the unrefocused, conventional photograph is shown in Figure 4.3-A and one sub-aperture image is shown in Figure 4.3-B.

These step-edge artifacts are a form of aliasing due to undersampling the (u, v) aperture in numerical integration of Equation 4.2. The problem occurs when the shift in neighboring sub-aperture images differs by more than one pixel, so edges in one sub-aperture may not be blended smoothly into the neighbor image during the summation process.

A solution is to super-sample the aperture plane, interpolating sub-aperture images at a resolution finer than 12×12. The super-sampling rate is chosen so that the minimum shift is less than one output pixel. The resulting artifact-free image is shown in Figure 4.3 C2. The extra sub-aperture images are interpolated from the nearest values in the light field.

Quadrilinear interpolation in the 4D space performs well. This process may be interpreted as a higher-order quadrature method for numerically integrating Equation 4.2.

Another kind of artifact is a darkening around the borders of the image as we refocus away from the focal plane. *Vignetting* of this kind is visible in Figure 4.3 C1. The cause of this artifact is that some of the shifted sub-aperture images do not cover this border region (see Figure 4.2B and C). Another way to interpret the problem is that some of the rays required to estimate Equation 4.1 are missing. These rays fell outside the physical boundary of the light field sensor, were never measured by the camera, and the corresponding values in the recorded L_F function are zero.

A solution for vignetting is to normalize the values of the border pixels by the fraction of rays that were actually found in the recorded light field. For example, this fraction will be the smallest in the most extreme border pixels. Dividing out by this fraction normalizes the value so that its intensity matches neighboring pixels. Figure 4.3 C2 was computed with this normalization procedure, eliminating the darkening evident in Figure 4.3 C1.

Extending the Depth of Field

There are many ways to compute an image with large depth of field from a light field. Perhaps the simplest is to simply extract one of the sub-aperture images. This approach can be thought of as *digitally stopping down the lens*, because it corresponds to producing the image that results from light coming through a reduced size aperture. "Stopping down the lens" is photography jargon for selecting a smaller aperture size. The problem with digital stopping down, as in its physical counterpart, is that it wastes the majority of the light that passes through the full aperture. The result is grainier images with lower SNR (see Figure 4.4B).

It is useful to think concretely in terms of the number of photons involved. Let us assume that each microlens in the prototype camera collects 10,000 photons during a particular exposure. These 10,000 photons are distributed amongst the microlens image on the 12×12 patch of pixels underneath the microlens. A sub-aperture image uses only one of the pixels in this patch, so it uses only about 90 photons out of the 10,000.

It is possible to extend the depth of field in a far superior manner by using all the information in the light field. The main concept is to refocus each pixel, assuming a full lens aperture, on the depth of the closest object in that direction. By using a full aperture in the

(A): Unrefocused   (B): Sub-aperture image   (C): Extended DOF

Figure 4.4: Comparison of a sub-aperture image and an image computed with digitally extended depth of field.

numerical integration of Equation 4.1, we obtain high SNR by combining the contributions of the 10,000 photons from all over the aperture. The resulting image, shown in Figure 4.4C, matches the depth of field of the sub-aperture image, but is far less grainy. In this dissertation, the phrase *digitally extending the depth of field* will be reserved for the process of computing high depth of field in this high SNR manner.

The epipolar images in Figure 3.6 z2 provide a visual way to conceptualize this process. Ordinary refocusing corresponds to projecting the entire epipolar image at a specific trajectory, as described in Chapter 2. This is how the images in Figure 4.1A were produced. In contrast, digitally extending the depth of field corresponds to projecting each column of the epipolar image along an individual trajectory that best aligns with the local features of the light field. Figure 4.1B can be thought of as projecting the blue pixels in Figure 3.6 z2 along a trajectory of negative slope and the brown pixels along a vertical trajectory.

The implementation of digitally extending the depth of field used in this dissertation begins by refocusing at all depths in the scene to create a focal stack of images. For example,

Figure 4.1 shows five of the frames in such a focal stack. Computing an extended depth of field image from a focal stack is a well studied problem, as reviewed in Section 4.1, and I make use of the *digital photomontage* algorithm developed by Agarwala et al. [2004] to compute the final extended depth of field image.

Digital photomontage accepts a set of images of a scene and an objective function defined on its pixels. It computes an output image that combines segments of the input images to optimize the objective function over the output image. It uses an iterative algorithm called *graph-cut optimization* [Boykov et al. 2001] to compute where to cut the input images, and *gradient-domain diffusion* [Pérez et al. 2003] to blend across the seams. For extending the depth of field, the objective function maximizes local contrast and minimizes cuts across image boundaries. By choosing other objective functions, some defined through interactive painting, the algorithm can be used to produce a wide variety of montages.

4.3 Theoretical Refocusing Performance

The method of estimating the refocus imaging equation by shifting and adding sub-aperture images provides intuition about the theoretical performance of digital refocusing. The shift-and-add method suggests that we should be able to render a desired focal plane as sharp as it appears in the sub-aperture images. Assuming that the images underneath each microlens are N pixels across, the sub-aperture images are ideally N times sharper than the full aperture images, since they correspond to imaging through apertures that are N times smaller.

Chapter 5 formalizes this intuition by applying Fourier analysis. The assumption underlying the analysis is that the recorded light fields are band-limited in the 4D space. Band-limited assumptions are very common in signal processing analysis, and such an assumption is a very reasonable model for the plenoptic camera. The mathematical details will be deferred until Chapter 5, but it is useful to state the theoretical performance in three different but equivalent ways.

First, digital refocusing allows us to reduce the amount of blur anywhere in the output photograph by a factor of up to N compared to the optical mis-focus blur. As an example consider our prototype camera where $N = 12$. If the optical blur in a particular region were less than 12 pixels, then we would be able to refocus the region "exactly" in the sense that

we could drive the blur down to at most one output image pixel. If the blur were more than 12 pixels, refocusing would make the region 12 times sharper, but still leave a residual blur.

A second way to state the performance is to quantify the range of desired focal depths in the world for which we can compute an exact refocused photograph. Since we can refocus exactly in regions where the optical mis-focus is less than N pixels, this range of depths is equal to the depth of field of an aperture N times smaller than the physical aperture of the lens. In terms of our prototype, where the main lens is $f/4$ and $N = 12$, we would ideally be able to refocus exactly on any depth within the depth of field of an $f/48$ lens. Put in these terms, photographers will recognize that digital refocusing presents a very significant extension in the effective depth of field.

Figure 4.3 makes these concepts clearer visually. In Image B, the depth of field includes the building in the background and the duck in the middle ground, which are crisp. In contrast, the chain-link fence in the foreground is out of the depth of field: it appears slightly blurry. Thus, we can refocus exactly onto the building and the duck, but we cannot refocus onto the chain link fence perfectly sharply. In Image C2, note that the refocused chain link fence is slightly blurry. Of course this is not to say that refocusing has failed—a striking amount of detail has been retrieved over the conventional photograph in Image A. The point is simply that the chain link fence cannot be rendered as sharp as if we had optically focused onto it. Interestingly, at just 2 cm from the camera, the fence was closer than the minimum focusing distance of the lens, so it would not have been possible to optically focus on it. Digital refocusing can improve the minimum focusing distance of lenses.

The third way of stating the performance is in terms of an *effective depth of focus* for the light field camera. The traditional depth of focus is the very narrow range of sensor depths inside the camera for which a desired plane will be rendered crisply. It is usually defined as the range for which the optical mis-focus falls within a single pixel. It is the job of the auto-focus mechanism in a conventional camera to ensure that the sensor is placed within the depth of focus of the desired subject before releasing the shutter of the camera. In the case of the light field camera, let us define the *effective* depth of focus as the range of sensor depths for which we can compute a crisp photograph of the desired subject plane, using digital refocusing if necessary. For this comparison let us assume that the output photographs of the light field camera have the same resolution as the conventional camera being compared
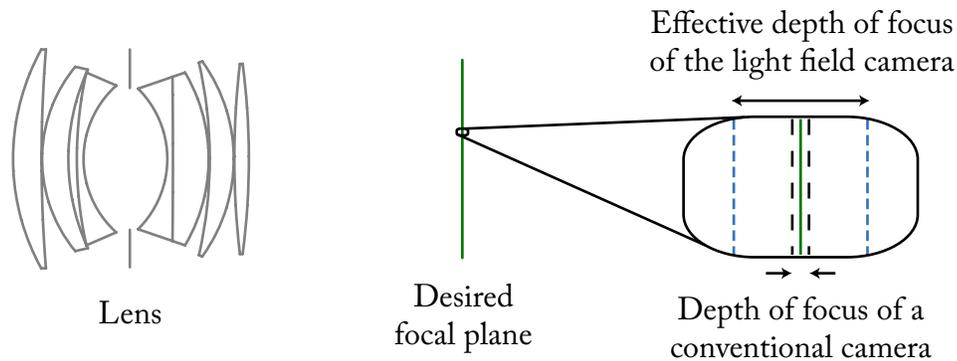
Figure 4.5: Improvement in effective depth of focus in the light field camera compared to a conventional camera. The range is 8 times wider in the light field camera, assuming that the directional resolution of the light field is 8×8.

against (i.e. that the microlenses in the light field camera are the size of the conventional camera sensor pixels, and the light field camera pixels are N times narrower). The first characterization of performance above states that we can refocus exactly if the blur radius is less than N pixels. This is N times larger than the one-pixel tolerance in the definition of the conventional depth of focus.

This observation directly implies that the effective depth of focus in the light field camera is N times larger than the regular depth of focus in the conventional camera, as shown in Figure 4.5. This means that the auto-focus mechanism can be N times less precise in the light field camera for equal output image quality, because digital refocusing can be used after the fact. This statement of performance makes very intuitive sense: the post-exposure capabilities of digital refocusing reduce the pre-exposure requirements on optical focusing.

4.4 Theoretical Noise Performance

A subtle but important implication of the theoretical refocusing performance is that it provides a superior way to produce images with high depth of field. For the following discussion, let us again assume that we have a plenoptic camera and conventional camera that have the same sensor size and that produce the same resolution output images (i.e. the plenoptic camera microlenses are the same size as the conventional camera pixels). However, assume that

the light field camera has $N \times N$ pixels across each microlens' image (i.e. it pays the price of collecting N^2 times more data than the conventional camera).

Let us compare the clarity of an image produced by a conventional camera with aperture stopped down from f/A to $f/(A \cdot N)$, and an image produced by a light field camera with the full f/A aperture. According to the theoretical refocusing performance, the light field camera can match the sharpness of the conventional photograph on any plane by digitally refocusing onto that plane. Furthermore, by digitally extending the depth of field the light field camera can match the entire depth of field of the conventional $f/(A \cdot N)$ camera. However, the light field camera would collect N^2 times more light per unit time because of its larger aperture. It would only require an exposure duration $1/N^2$ times as long to collect as much light. Alternatively, for equal exposure duration, the light field camera would collect N^2 times higher signal.

A critical question is, how much higher will the SNR be in the case of the same exposure duration? That is, how much improvement would the N^2 times higher signal provide relative to the inherent noise? This is a complicated question in general, depending on the characteristics of the sensor, but the worst-case performance is easy to understand, and that is what we test in the experiments below. The most important concept is that the standard deviation in the number of photons collected at a pixel follows Poisson statistics [Goodman 1985], such that the standard deviation is proportional to the square root of the mean light level. For example, if an average of 10,000 photons would be collected, then the standard deviation for the detected value would be 100 photons. Given this fact, the least improvement occurs when the light level is sufficiently high that the Poisson noise exceeds any fixed, constant sources of noise in the sensor. Under this worst-case assumption, the SNR of the light field camera is N times higher, because its signal is N^2 times higher, but the Poisson noise is only N times higher.

That said, it is worth addressing a point in the original assumption that sometimes causes confusion. Looking back at the original assumptions, one may wonder how the analysis would change if we had assumed the conventional camera and the light field camera had photosensors with the same resolution, rather than the light field camera having $N \times N$ higher resolution. In other words, let us now consider the case where the conventional camera pixels are N times narrower. There is an incorrect tendency to suspect that the noise of

the conventional camera would somehow improve under this assumption. The reason that it cannot improve is that the same total number of photons enter the camera and strike the sensor. This number is fixed by the size of the lens aperture and the area of the sensor, not by the resolution of the pixels. In fact, since there are N^2 more pixels on the sensor under the current assumptions, each pixel now has $1/N^2$ the area and collects $1/N^2$ as many photons as before. The SNR would be worse by an additional factor of N. Of course this factor would be eliminated by down-sampling the higher-resolution image to the resolution of the light field camera, converging on the analysis above.

4.5 Experimental Performance

The following subsections present experimental data from the prototype camera to test both the theoretical refocusing and noise reduction predictions. In summary, the results show that the prototype can refocus to within a factor of 2 of the theoretically ideal performance, and that the measured noise closely corroborates the theoretical discussion above.

Experimental Method

The basic experimental approach is to choose a fixed a main lens focus, and compare the prototype camera with digital refocusing against a conventional camera with various main lens aperture sizes. To make this comparison, I designed a scene with a single plane target at a small distance *away* from the optical focus of the main lens. In this regime, the sharpness of conventional photographs increases as the aperture size decreases.

In these experiments I chose to explicitly compare a hypothetical conventional camera with 296×296 125-micron pixels against our prototype with microlenses of the same size and 12×12 pixels across each microlens image. Thus, the question that is addressed is, how much improvement is obtained by capturing 12×12 directional samples at each output pixel instead of a single light value?

A difficulty in implementing this comparison is how to obtain the images for the conventional camera, since no real device exists with 125-micron wide pixels. I approximated the output of such a hypothetical device by summing all the pixels in each microlens image. Since all the light that strikes a 125-micron microlens is deposited in some pixel in the image

underneath it, summing this light counts all the photons that would have struck the desired 125-micron conventional pixel.

An important advantage of this approach to estimating the conventional photographs is that it allows use of exactly the same set-up for acquiring the light fields for digital refocusing and the conventional photographs for comparison. The only change that occurs is in the aperture size of the camera, which is easily adjusted without disturbing the position or focus of the camera. If I had used different cameras for each, meaningful comparison would require factoring out differences in position and focus of the two cameras, as well as differences in the noise characteristics of two separate sensors.

A potential disadvantage of the chosen approach is that it could result in artificially higher sources of constant sensor noise due to the summation of read noise from multiple pixels. Nevertheless, as already discussed, this increase is insignificant if the Poisson noise dominates the constant sources of noise. The experiments below test the change in SNR with light level to ensure that this is the case.

Refocusing Results

I recorded an $f/4$ light field photograph, and a series of conventional photographs in half-stop increments from $f/4$ to $f/45$. These images were shot without changing the focus setting on the lens. The higher f-number photographs were sharper because they have smaller aperture sizes and the resulting circle of confusion is smaller. The light field photograph was sharpened by digitally refocusing onto the resolution chart as a post-process.

Figure 4.6 presents a visual comparison of the light field camera performance compared to select conventional images. Note how much more blurry the $f/4$ conventional photograph is compared to the $f/4$ light field camera image. The light field image seems sharper than the conventional $f/16$ photograph, but not quite as sharp as the $f/32$. It appears to most closely match the sharpness of the $f/22$ image.

Figure 4.7 is a numerical version of the same experiment, comparing experimentally measured modulation transfer function (MTF) curves. I computed the MTFs by shooting photographs of a point light source and computing their Fourier transforms [Williams 1915]. For the light field camera, I shot the light field, computed the sharpest possible refocused photograph of the point light, and computed the Fourier transform of the refocused image.

$f/4$ Refocused Light Field $f/4$ Conventional

$f/16$ Conventional $f/22$ Conventional $f/32$ Conventional
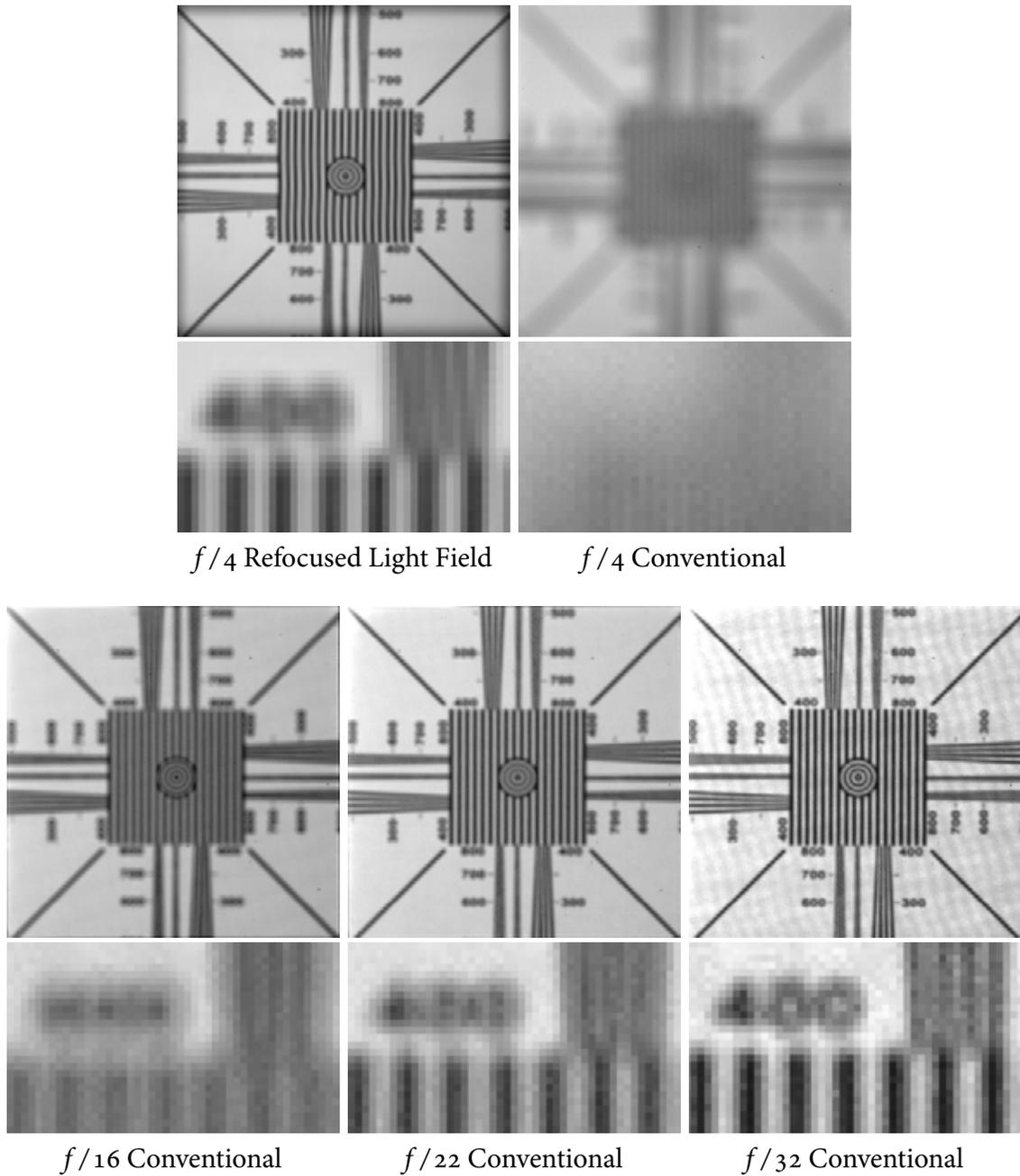
Figure 4.6: Experimental test of refocusing performance: visual comparison.
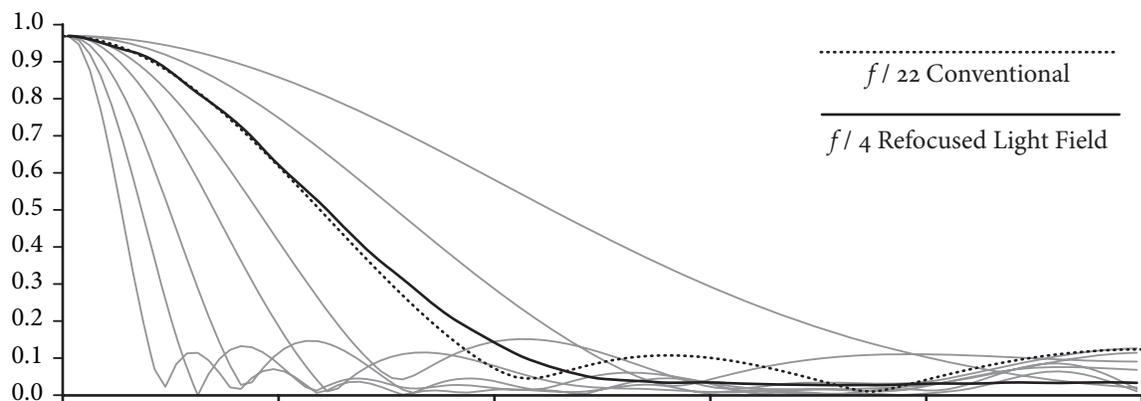
Figure 4.7: Experimental test of refocusing performance: MTF analysis. Horizontal axis plots spatial frequency from 0 up to the Nyquist rate of the sensor.

Figure 4.7 illustrates the classical increase in MTF curve with aperture number. The graph contains a series of MTF plots for a conventional camera with different aperture sizes. The unlabeled gray plots from left to right are for f-numbers 4, 5.6, 8, 11, 16, 32 and 45. As the f-number increases, the aperture size decreases, leading to sharper images and a higher MTF. The two plots in black are of primary interest, being the plots for the $f/22$ conventional image and the $f/4$ refocused light field image. The MTF of the light field camera most closely matches that of the $f/22$ conventional camera, providing numerical corroboration for the visual comparison in Figure 4.6.

According to theory, the light field camera, with 12 pixels across each microlens image, would have matched the $f/45$ conventional camera in ideal conditions. In reality, the experiment shows that ability to refocus within the depth of field of an $f/22$ aperture, approximately a loss of a factor of 2.

The main source of loss is the resampling that takes place to estimate the slanted strip of rays corresponding to each pixel value. As shown in Figure 3.2, estimating the slanted strip with the sample boxes of the light field introduces error, because the boxes exceed the boundaries of the strip by some amount. Another, relatively minor source of loss is diffraction, as discussed in Section 3.4. While diffraction is not the limiting factor on resolution in this camera given the $f/4$ aperture of our microlenses, the diffraction blur has a width of

approximately 2/3 of a sensor pixel, which contributes slightly to the loss in effective directional resolution.

Noise Results

I measured the improvement in noise when the exposure duration is kept constant. As predicted by theory, I found that the SNR of the $f/4$ light field refocused photograph was higher than the $f/22$ conventional photograph with equivalent depth of field. The computed increase in SNR was 5.8, which is close to the square root of the increase in light level (the lens aperture is approximately 5.5 times wider in the $f/4$ light field camera). This square root scaling of SNR indicates that Poisson sources of noise are dominant, as discussed in Section 4.4. I calculated the SNR of each image by comparing the deviations of the photograph against a noise-free standard, which I obtained by averaging multiple photographs shot under identical conditions. The SNR is the root-mean-square of the difference between a given image and the noise-free standard.

$f/4$ Conventional $f/22$ Conventional $f/4$ Refocused Light Field
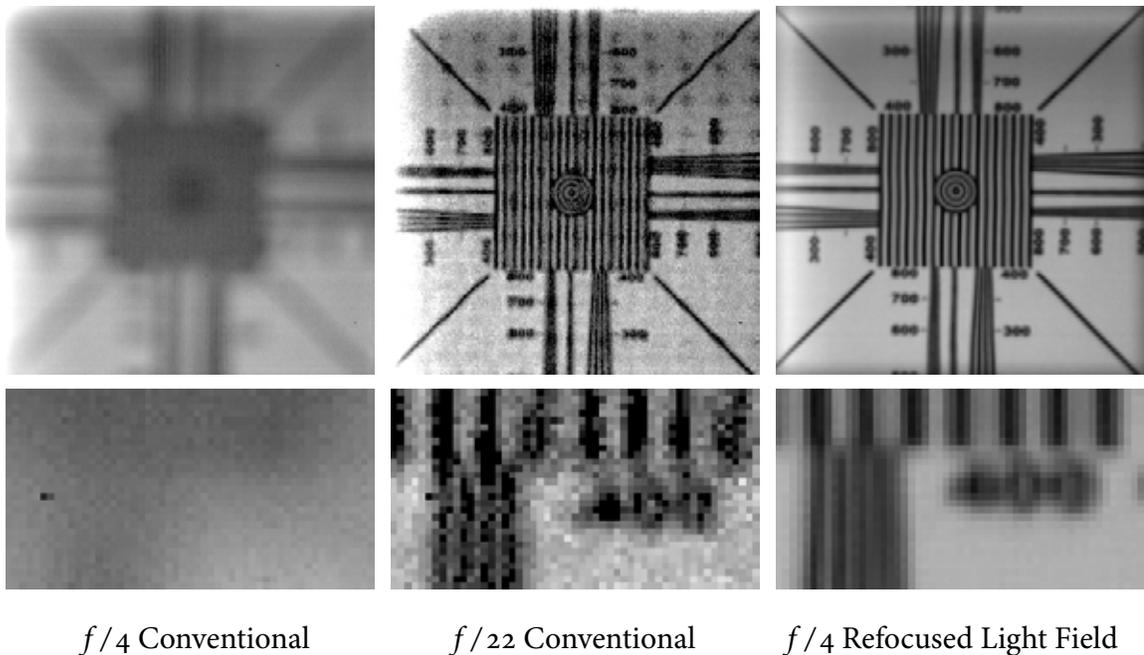
Figure 4.8: Experimental test of noise reduction using digital refocusing.
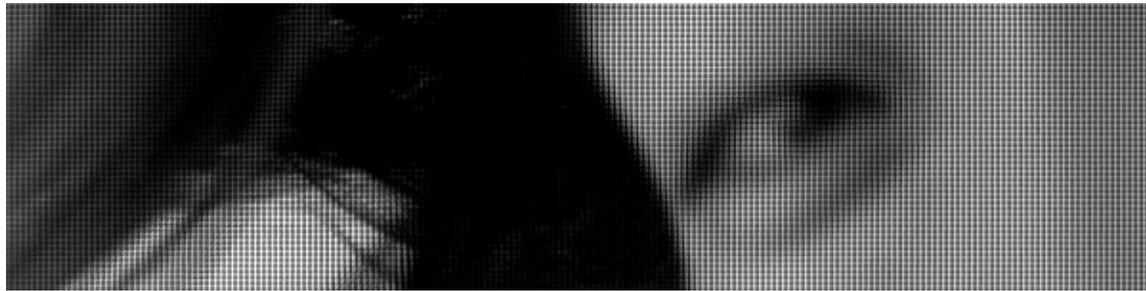
(A): Closeup of raw light field photograph

(B1): Unrefocused   (B2): Refocused on face   (B3): Extended DOF

Figure 4.9: Refocusing and extending the depth of field.

Figure 4.8 presents some of the images used in this numerical analysis. The three photographs were produced under the same conditions as in Figure 4.6, except that the light level was deliberately kept low enough to make noise visible, and the shutter duration was held constant regardless of aperture size. In Figure 4.8, the conventional $f/4$ photo is completely blurred out due to optical mis-focus of the main lens. The $f/22$ image is relatively sharp, but much noisier. Note that the $f/22$ image is scaled by a factor of approximately 30

(A): $f/4$, 1/125 sec
Light Field, Extended DOF

(B): $f/22$, 1/125 sec
Conventional Photograph

(C): $f/22$, 1/4 sec
Conventional Photograph
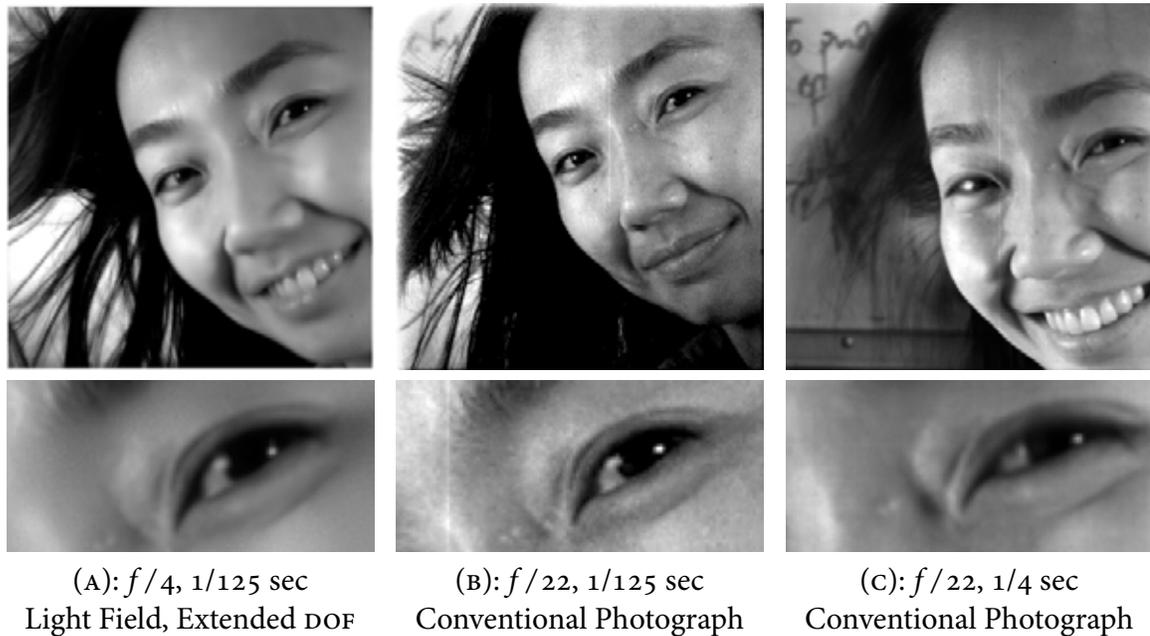
Figure 4.10: Comparison between a light field photograph with digitally extended depth of field and conventional photographs with optically extended depth of field.

to match the brightness of the $f/4$ image. The third image is an $f/4$ light field photograph, with digital refocusing onto the resolution chart. Note that this image did not have to be scaled, because every pixel integrated all the light coming through the $f/4$ aperture. The SNR increase of 5.8 times is between the $f/22$ conventional image and the $f/4$ light field photograph.

4.6   Technical Summary

This section presents an experiment that visually summarizes the main improvements in flexibility and light-gathering performance provided by digital refocusing as compared to conventional imaging.

Figure 4.9 illustrates a light field recorded with the prototype camera, and a set of flexible photographic edits applied to this single exposure in computing final photographs. Note that the unrefocused photograph computed in Image B1 is optically focused on the woman's hair, and the motion of her hair has been frozen by the short 1/125 second exposure enabled by

the relatively fast light-gathering power of the $f/4$ lens. However, her facial features are blurred because of the shallow $f/4$ depth of field. In Image B2, the photograph has been digitally refocused onto the woman's face to bring her eyes into focus. Note, however, that her hair is now out of focus. Whether this is desirable depends on the artist's intentions. Image B3 illustrates the digitally extended depth of field image in which both her hair and features are sharp.

Figure 4.10 compares the digitally extended depth of field image against comparable photographs taken with a conventional camera. The digitally extended depth of field image of Figure 4.9 is replicated in Figure 4.10A. In Image B, the exposure duration was kept at 1/125 second, and I used the classical method of reducing the aperture size (down to $f/22$) to extend the depth of field. Although the woman's hair and face are both in focus and sharp, the resulting image was over 30 times darker because of the 30 times smaller aperture. I scaled the image to match the brightness of Image A, revealing much higher graininess due to a relative increase in Poisson photon noise. In contrast, in Image C, I kept the $f/22$ aperture, but used a 30-times longer exposure of 1/4 second to capture the same amount of light as in Image A. This is the exposure setting that would be chosen by an auto-exposure program on most conventional cameras. The resulting image contains as many photons as the light field photo and matches its low noise grain. However, in Image C the woman's hair is blurred out due to motion over the relatively long exposure time. Figure 4.10 makes it visually clear why light field photography may have significant implications for reducing noise in low-light photography.

In summary, this experiment highlights two improvements of the light field paradigm over conventional imaging. The first is the unprecedented flexibility of choosing what is in focus in final photographs after a single photographic exposure. The second is the capability of shooting with significantly shorter exposure durations or lower image noise.

4.7   Photographic Applications

To conclude this chapter, this section presents examples of applying digital refocusing in common photographic scenarios. The focus is on improvements in flexibility and performance enabled by digital refocusing and extending the depth of field.

Portraiture

Good photographs of people commonly have blurry backgrounds in order to concentrate attention on the human subject of interest. In conventional photography, a blurry background is achieved by selecting a large lens aperture to obtain a shallow depth of field. For example, in some point-and-shoot cameras, the camera may be able to detect that the scene is a picture of a person, in which case it automatically selects the largest available aperture size. In other cameras, the photographer may have to choose the correct aperture manually.

In any case, use of a large aperture presents certain practical difficulties. One example is that because the depth of field is shallow, it becomes essential to focus very accurately. Usually the correct focal point is on the eyes of the subject. However, with a large aperture, the depth of field may cover only a few centimeters. When photographing a moving subject or taking a candid picture, it is quite easy to accidentally focus on the ear or the hair of the subject, leaving the eyes slightly out of focus. Figure 4.11A illustrates this kind of accident, where the camera is mis-focused by just 10 cm on the girl's hair, rendering her features blurry. Since the photograph was captured with the light field camera, digital refocusing may be applied, bringing the focal plane onto the girls' eyes in Figure 4.11C.

(A): Mis-focused original (B): Raw light field of eye (C): Refocusing touch-up

Figure 4.11: Fixing a mis-focused portrait

A different challenge is having more than one subject, such as a portrait of two people offset in depth. In this case there is no single focal plane of interest, so using a large aperture

means that only one face will be in focus (see, for example, Figure 4.12A). With a conventional camera, it would be impossible to achieve a nicely blurred background and yet capture both faces in sharp focus. Of course one could stop the aperture of the main lens down to extend the depth of field, but this would make the background sharp and distracting (see Figure 4.12B.

(A): Unrefocused   (B): Extended DOF   (C): Partially extended DOF
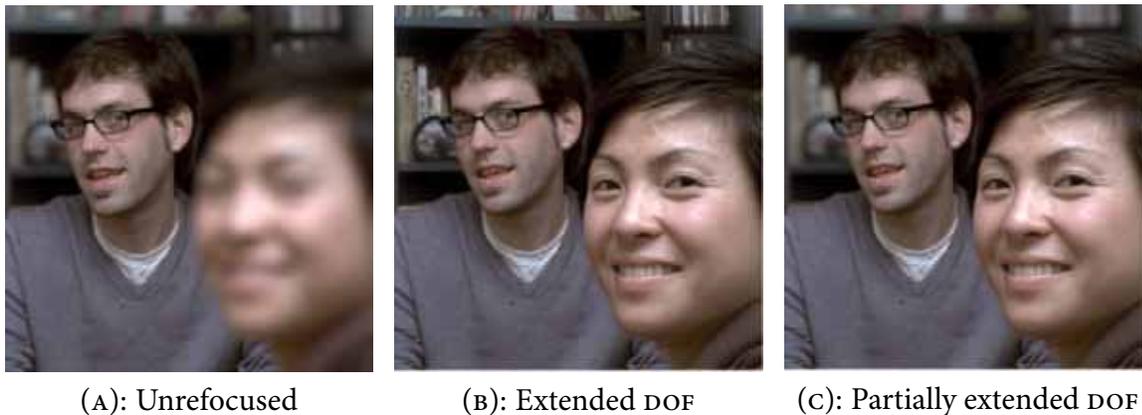
Figure 4.12: Maintaining a blurred background in a portrait of two people.

With the light field camera, a superior solution is to only partially extend the depth of field, to selectively include only the human subjects while leaving the background blurry. This can be achieved by digitally refocusing onto each person in turn, and combining the sharpest portions of these images. Figure 4.12C illustrates this approach. Notice that both faces are sharp, but the blur in the background is maintained.

Figure 4.13 illustrates a novel application of digital refocusing in portraiture. This figure contains three frames of a refocus movie that sweeps the focal plane from front to back in the scene. The entire movie was computed from one light field, and is really just another light field output format. The result is a digital snapshot that presents a new kind of interactivity. A common experience in viewing such movies is a sensation of discovery as a subject of interest "pops" into focus and can finally be recognized. Many of the people who have seen these movies find this effect quite absorbing, and replay the movies many times.

Figure 4.13: The sensation of discovery in refocus movies.

Action Photography

Action photography is the most demanding in terms of light capture efficiency and focus accuracy. Large apertures are usually needed in all but the brightest conditions, in order to enable short shutter durations to freeze the motion in the acquired photograph. The resulting shallow depth of field, coupled with moving targets, makes it challenging to focus accurately.

Figure 4.14 illustrates that our camera can operate with very short exposures. The light field photographs in A1-A3 and B1-B3 illustrate refocusing photographs of swimmers diving into a pool at the start of a race. The exposure duration in these images is 1/500th of a second. Images C1-C3 show refocusing through water that is splashing out of a broken wine glass. This photograph was shot in a dark room with a flash, for an effective exposure duration of 1/3000th of a second.

Landscape Photography

While sports photography is a natural fit for the capabilities provided by digital refocusing, landscape photography may at first seem like it presents relatively few opportunities. The reason for this is that in landscape photography it is typically desirable for everything in the scene to be in focus at the same time, from the flowers in the foreground to the mountains in the background. It is very common to shoot landscapes using the smallest aperture available on a lens to obtain this large depth of field.
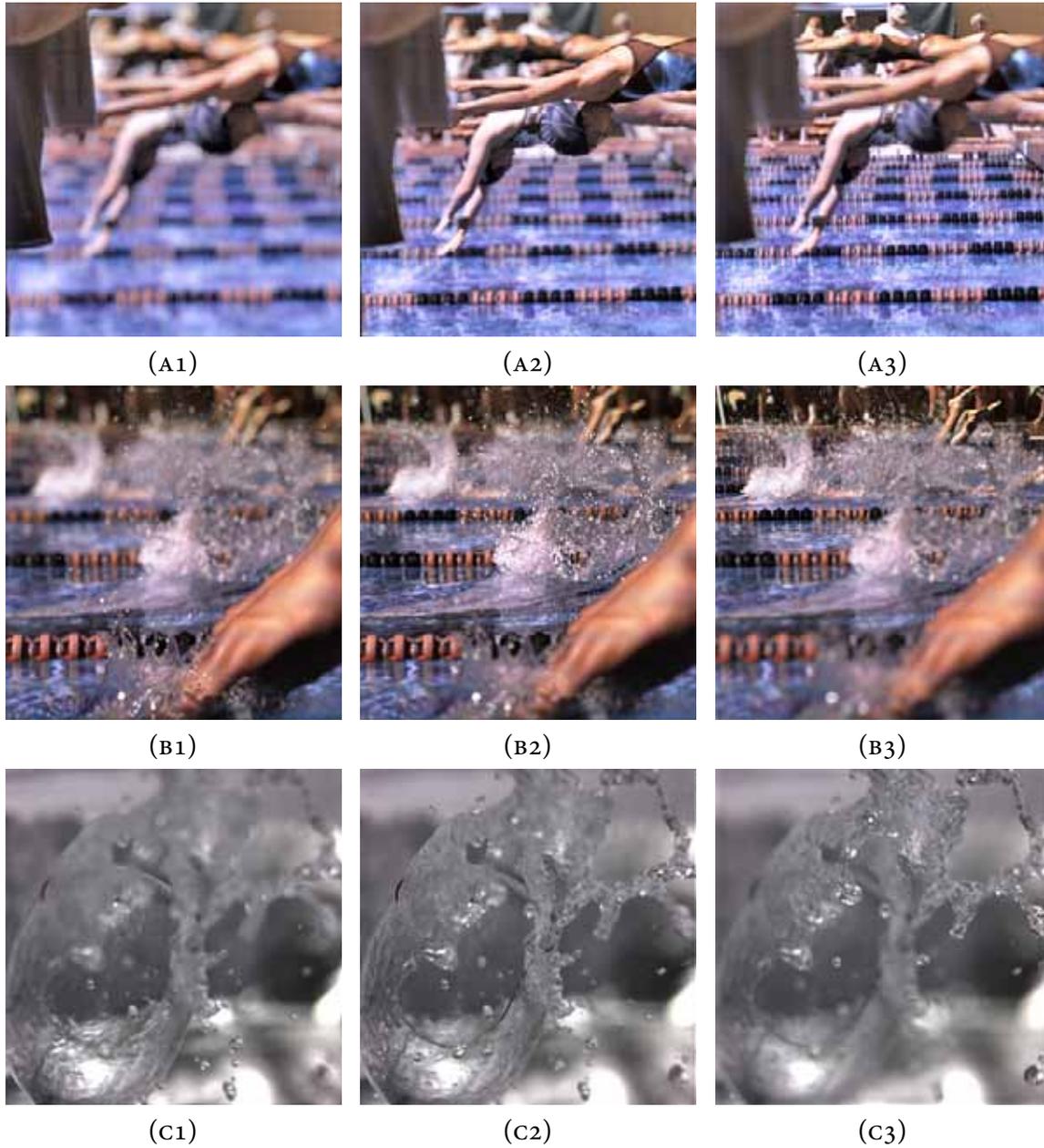
Figure 4.14: Digital refocusing of high-speed light field photographs.

At the same time, the best landscapes are often taken in the hours closest to sunrise and sunset, when the light is most dramatic and colorful. Shooting at these times means that the light level is relatively low. Coupled with the use of a small aperture for large depth of field, exposure durations are often quite long. For this reason, use of a tripod is usually considered mandatory to hold the camera still, which is unfortunate because it limits freedom of composition.

The light field approach to photography is useful when taking landscapes, because it allows for shorter exposure durations using digitally extended depth of field, enabling a much greater range of conditions over which the camera may be hand-held and better compositions achieved.

| (A1): Refocus far | (A2): Refocus close | (A3): Extended DOF |
| (B1): Refocus far | (B2): Refocus close | (B3): Extended DOF |
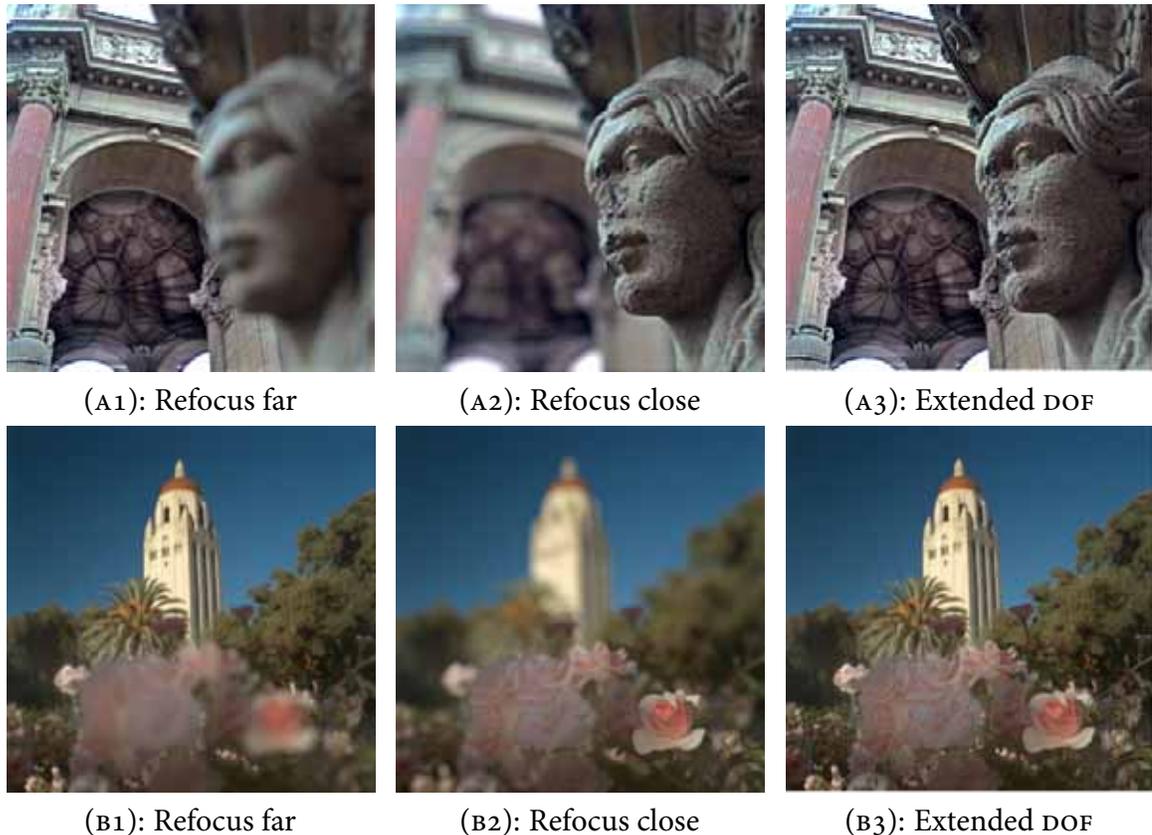
Figure 4.15: Extending the depth of field in landscape photography.

Figure 4.15 illustrates the benefits with respect to two light field landscapes. In each case,

(A1)   (A2)   (A3)

(B1)   (B2)   (B3)
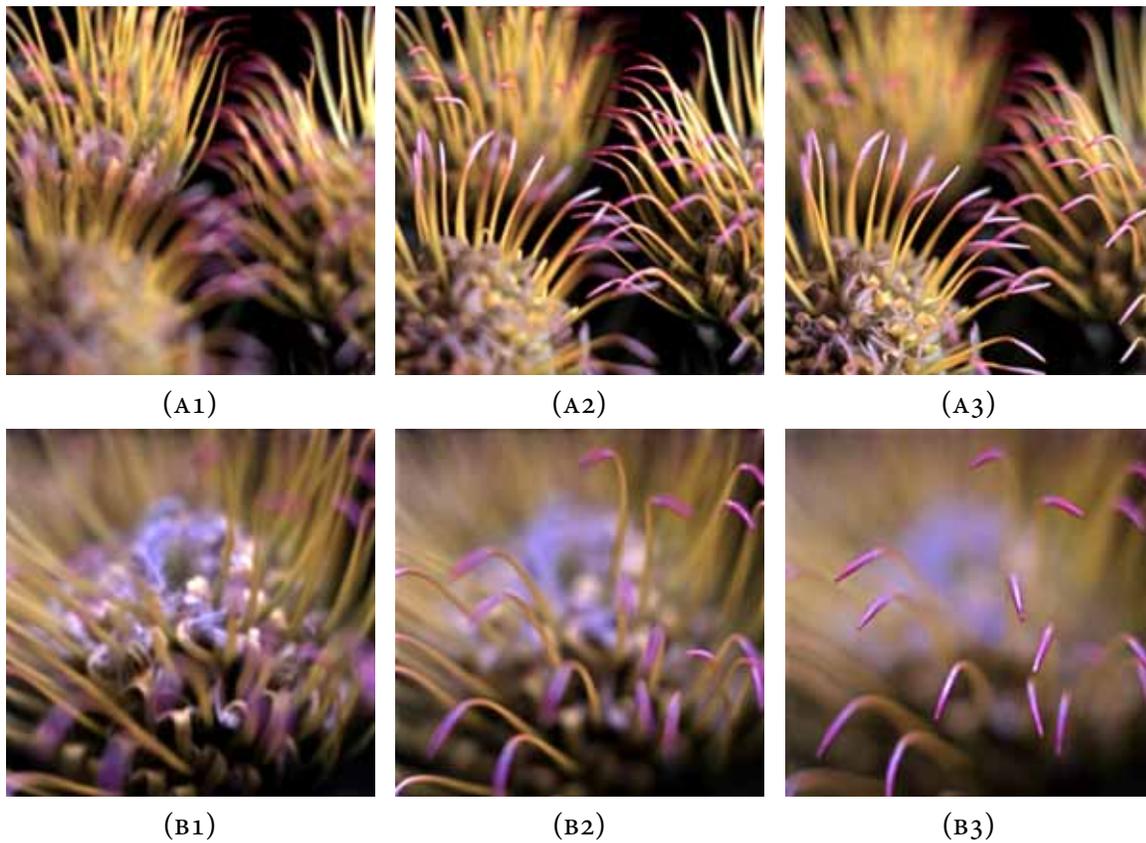
Figure 4.16: Digital refocusing of light field macro photographs of proteas flowers.

using a conventional camera with a large aperture would result in either the foreground or background being blurry, as shown by the first two images in each sequence. The results in the right-most images are the result of digitally extending the depth of field, providing large depth of field with a 30 times shorter exposure than would be possible with a conventional camera.

Macro photography

The depth of field decreases with the proximity of the focal plane to the camera, so macro photography, where the subject is extremely close to the camera lens, places heavy demands on focus accuracy. To give a sense of scale, to achieve 1:1 magnification, where one millimeter on the real subject is imaged to one millimeter on the sensor inside the camera, the

(A): High viewpoint   (B): Low viewpoint

(C): Further viewpoint   (D): Closer viewpoint
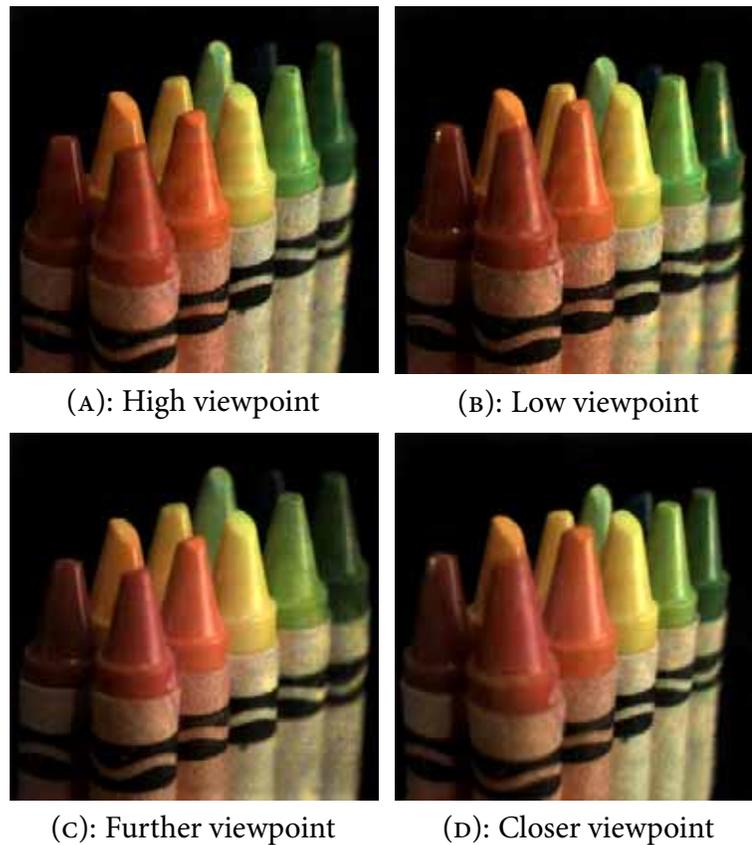
Figure 4.17: Moving the viewpoint in macro photography

subject needs to be positioned at a distance of just two focal lengths away from the camera. For example, some of the macro photographs shown here were shot with an 80 mm lens, where the subject was placed just 160 mm away from the camera.

Because of the proximity of the subject, macro photography is very sensitive to the exact placement of the camera. Movement of a few centimeters can drastically chance the composition of the scene. In addition, unlike most other kinds of photography, focusing in macro photography is often achieved by moving the entire camera rather than turning the focus ring. The photographer selects the desired magnification, and moves the camera back and forth until the desired subject intersects the optical focal plane. As a result, getting the correct composition by moving parallel to the subject plane, and the correct focus, by moving

perpendicularly, can be a challenging three-dimensional motion on the part of the photographer. The ability to digitally refocus reduces the requirements in at least the perpendicular axis of motion, making it easier to compose the photograph as desired. Figure 4.16 illustrates digital refocusing of two close-up shots of proteas flowers.

A different advantage afforded by light field macro photography is being able to virtually move the camera position by synthesizing pin-hole images rather than full-aperture images. Since the lens is large compared to the subject in macro photography, the cone of rays that enters the camera from each point subtends a relatively wide angle. By moving the virtual pin-hole camera within this cone of rays, one can obtain changes in parallax or perspective, as illustrated in Figure 4.17.

The top row of Figure 4.17 illustrates movement of the viewpoint laterally within the plane of the lens aperture, to produce changes in parallax. The bottom row illustrates changes in perspective by moving along the optical axis, away from the scene to produce a near-orthographic rendering (Image C) and towards the scene to produce a medium wide angle (Image D).

5

Signal Processing Framework

As it has with so many scientific phenomena and engineering systems, Fourier analysis provides an entirely different perspective on photograph formation and digital light field photography. This chapter applies signal processing methods based on Fourier theory to provide fresh insight into these matters. For example, this style of thinking allows more rigorous analysis of the performance of digital refocusing. It also leads directly to the design of a very different kind of algorithm for computing refocused photographs, which is asymptotically faster than the integral projection methods described in Chapter 4.

Section 5.2 is an overview of the chapter, keeping the Fourier interpretation at a high level. The focus is on intuition and geometry rather than equations, but basic exposure to Fourier transforms is assumed. Later sections delve into the formal mathematical details, and assume familiarity with Fourier transforms and linear systems theory at the level of a first year graduate level course, such as a course based on Bracewell's book [1986].

5.1   Previous Work

The closest related Fourier analysis is the plenoptic sampling work of Chai et al. [2000]. They show that, under certain assumptions, the angular band-limit of the light field is determined by the closest and furthest objects in the scene. They focus on the classical problem of rendering pin-hole images from light fields, whereas this thesis analyzes the formation of photographs through lenses.

Imaging through lens apertures was first demonstrated by Isaksen et al. [2000]. They qualitatively analyze the reconstruction kernels in Fourier space, showing that the kernel width decreases as the aperture size increases. This chapter continues this line of investigation, explicitly deriving the equations for full-aperture imaging from the radiometry of photograph formation.

More recently, Stewart et al. [2003] have developed a hybrid reconstruction kernel that combines full-aperture imaging with band-limited reconstruction. This allows them to optimize for maximum depth-of-field without distortion. In contrast, this chapter focuses on fidelity with full-aperture photographs that have finite depth of field. As we have seen in the previous chapter, narrow depth of field is often purposely used in photography to visually isolate a subject and direct the viewers gaze.

Finally, Durand et al. [2005] have constructed a framework for analyzing the Fourier content of the light field as it propagates through a scene and is transformed by reflections, propagation, and other phenomena. Where their analysis focuses on the transformations in the light field's frequency spectrum produced by the world, this chapter focuses on the transformations imposed by the detector and the ramifications on final image quality.

5.2   Overview

The advantage of moving into the Fourier domain is that it presents a simpler view of the process of photographic imaging. While the spatial relationship between photographs and light fields (Chapter 2) can be be understood fairly intuitively, the relationship itself—integral projection—is a relatively heavyweight mathematical operation. In contrast, in the Fourier domain the relationship is simpler—a photograph is simply a 2D slice of the 4D light field. That is, the values of the Fourier transform of a photograph simply lie along a 2D plane in the 4D Fourier transform of the light field. This "Fourier Slice Photograph Theorem," which is derived in Section 5.3, is one of the main theoretical contributions of this thesis.

Figure 5.1 is a graphical illustration of the relationship between light fields and photographs, in both the spatial and Fourier domains. The graphs in the middle row of the figure review the fact that, in the spatial domain, photographs focused at different depths
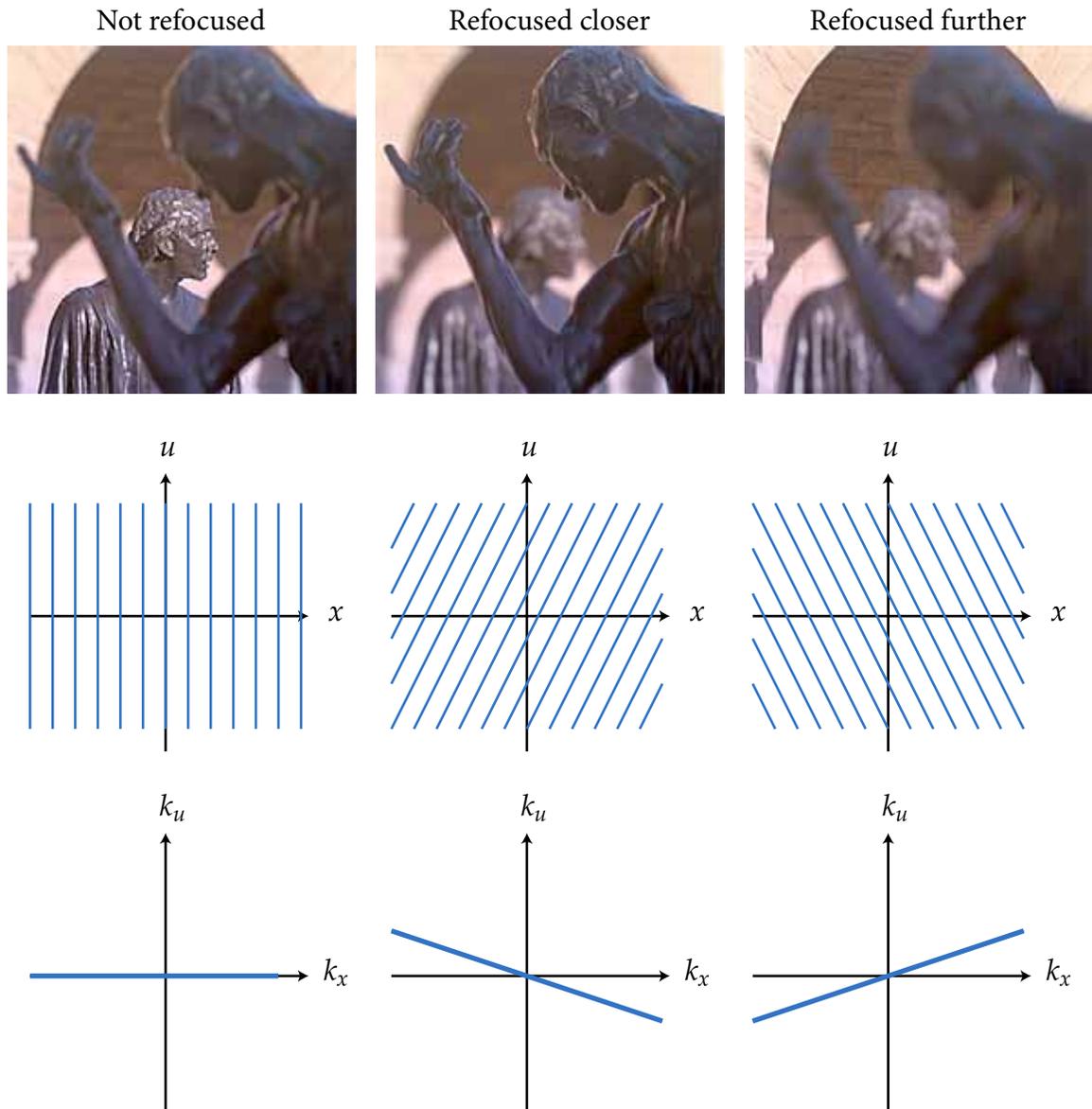
Figure 5.1: The relationship between photographs and light fields is integral projection in the spatial domain (middle row) and slicing in the Fourier domain (bottom row).

correspond to slices at different trajectories through the ray-space. The graphs in the bottom row of Figure 5.1 visualize the situation in the Fourier domain. These graphs illustrate the use of "k-space" notation for the Fourier axes, a practice borrowed from the magnetic resonance (MR) literature and used throughout this chapter. In this notation, k_x and k_u are the Fourier-domain variables that correspond to the x and u spatial variables (similarly for k_y and k_v). The graphs in the bottom row illustrate the Fourier-domain slices that provide the values of the photograph's Fourier transform. The slices pass through the origin. The slice trajectory is horizontal for focusing on the optical focal plane, and as the chosen virtual film plane deviates from this focus, the slicing trajectory tilts away from horizontal.

The underlying reason for the simplification when transforming into the Fourier domain is a consequence of a well-known result due to Ron Bracewell [1956] called the Fourier Slice Theorem (also known as the Central Section Theorem and the Fourier Projection-Slice Theorem). Bracewell originally discovered the theorem while studying reconstruction problems in synthetic aperture radar. However, it has grown to make its greatest contributions in medical imaging, where it is of fundamental theoretical significance for reconstruction of volumes in computed tomography (CT) and positron emission tomography (PET) [Macovski 1983], and the design of excitation pulse sequences in (MR) imaging [Nishimura 1996].

The classical theorem applies to 1D projections of 2D functions, so applying it to photographs and light fields requires generalization to the relatively unusual case of 2D projections from 4D functions. Nevertheless, the connection with the classical theorem remains strong, and provides an important channel for cross-fertilization between light field photography and more mature imaging communities. For example, mining the MR literature produced some of the algorithmic improvements described in Section 5.5 for computing refocused photographs in the Fourier-domain.

Perhaps the most important theoretical application of the Fourier Slice Photograph Theorem is a rigorous derivation of the performance of digital refocusing from light fields sampled with a plenoptic camera. Section 4.3 presented geometric intuition in terms of sub-aperture images, but it is difficult to make those arguments formal and exact. In this chapter, Section 5.4 presents a formal derivation under the assumption that the light fields recorded by the plenoptic camera are band-limited.

Figure 5.2c illustrates what it means for the recorded light field to be band-limited: it

(A) Continuous light field (B) Perfect photograph (C) Recorded light field

(D) Photograph computed
without refocusing

(E) Refocused photograph
with full spatial resolution

(F) Refocused too far with
loss in spatial resolution
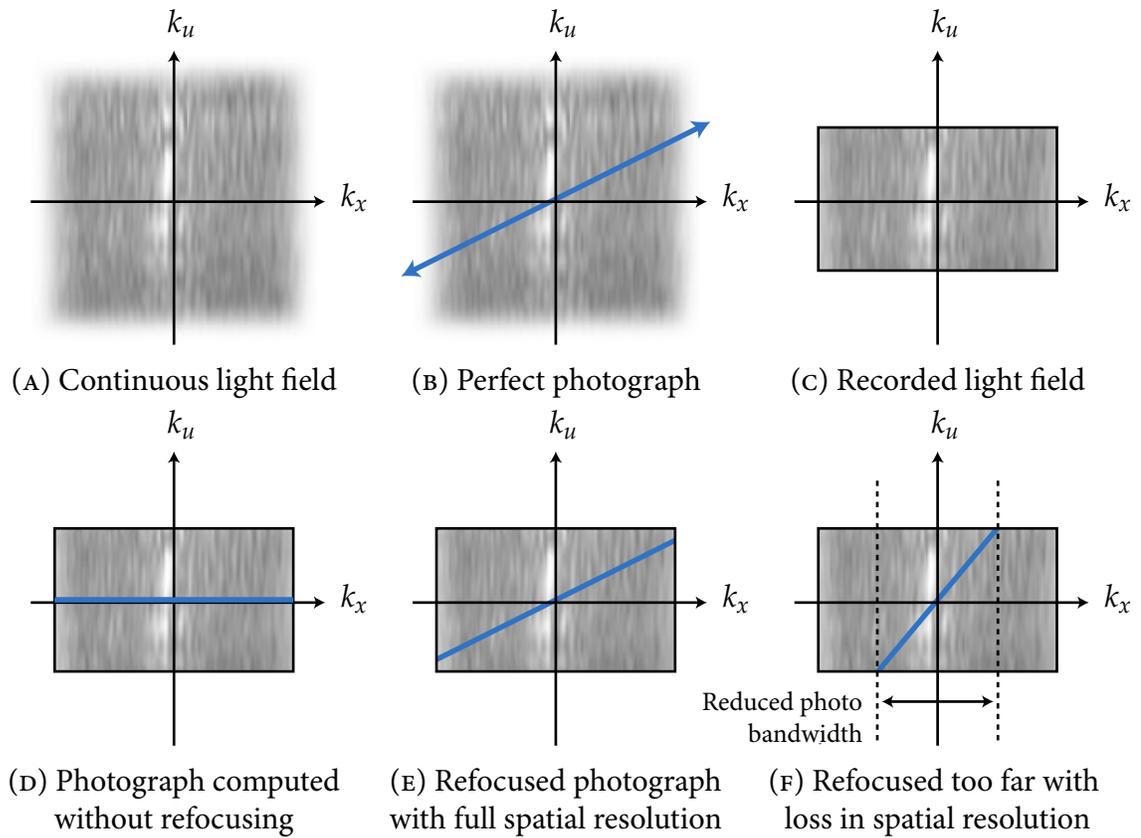
Figure 5.2: Fourier-domain intuition for theoretical refocusing performance.

provides perfect information about the continuous light field within a box about the origin, and is zero outside this box. The texture shown in the background of the graphs of Figure 5.2 is for illustrative purposes only. In the spatial domain, which is not shown, the band-limited assumption means that the signal is sufficiently blurred that the finest details match the sampling rate of the microlenses and pixels in the camera. While perfect band-limiting is physically impossible, it is a plausible approximation in this case because the camera system blurs the incoming signal through imperfections in its optical elements, through area integration over the physical extent of microlenses and photosensor pixels, and ultimately through diffraction.

With the band-limited assumption, analysis of the performance of digital refocusing becomes exact. Figure 5.2 illustrates the basic concepts, for the case of simplified 2D light

fields. Since photographs are simply slices at different trajectories through the origin of the ray-space, a computed photograph is the line segment obtained when we intersect the slice with the band-limited box of the recorded light field (Figures 5.2D–F).

The complete slice, which extends to arbitrarily high frequencies away from the origin, corresponds to a perfect photograph (Figure 5.2B). The crucial observation is that the computed line segment is therefore a band-limited version of the perfect photograph. By calculating the band-limit of this segment, we can make precise statements about the effective resolution of the refocused photograph.

The mathematical details of this analysis are left to Section 5.4, but Figures 5.2E and F illustrate the critical turning point. They show that, when focusing at large distances from the optical plane, the slice trajectory tilts sufficiently far from the horizontal that it crosses the corner of the band-limited box. Until this point, the slice intersects the vertical border of the band-limit, and the resulting photograph has full resolution (limited by the spatial resolution of the recorded light field) . However, for extreme tilts the slice intersects the horizontal border, and the resulting photograph has reduced resolution (partially limited by directional resolution). The range of slice trajectories that provide full resolution is shown in blue on Figure 5.3. Analyzing the angular bounds of this region provides an expression for the range of virtual film depths that permit exact refocusing. Section 5.4 performs this analysis, corroborating the intuition developed in Chapter 4 that the range increases linearly with directional resolution.

Figure 5.3: The range of slice trajectories in the Fourier space for exact refocusing.

Section 5.5 applies the Fourier Slice Photograph Theorem in a very different manner to derive a fast Fourier Slice Digital Refocusing algorithm. This algorithm computes photographs by extracting the appropriate 2D slice of the light field's Fourier transform and performing an inverse 2D Fourier transform. The asymptotic complexity of this algorithm is $O(n^2 \log n)$, where n is the resolution of the light field in each of its four dimensions. This

complexity compares favorably to the $O(n^4)$ approach of existing algorithms, which are essentially different approximations of numerical integration in the 4D spatial domain.

5.3 Photographic Imaging in the Fourier Domain

Chapter 2 introduced the imaging integral in Equation 2.3, which relates light fields and photographs focused at different depths. Our first step here is to codify Equation 2.3 in the operator notation that will be used throughout this chapter.

Operator notation provides a higher level of mathematical abstraction, allowing the theorems derived below to express the relationship between transformations that we are interested in (e.g. image formation) rather than being tied up in the underlying functions being acted upon (e.g. light fields and photographs). Throughout this chapter, calligraphic letters, such as $\mathcal{A}$, are reserved for operators. If f is a function in the domain of $\mathcal{A}$, then $\mathcal{A}[f]$ denotes the application of $\mathcal{A}$ to f.
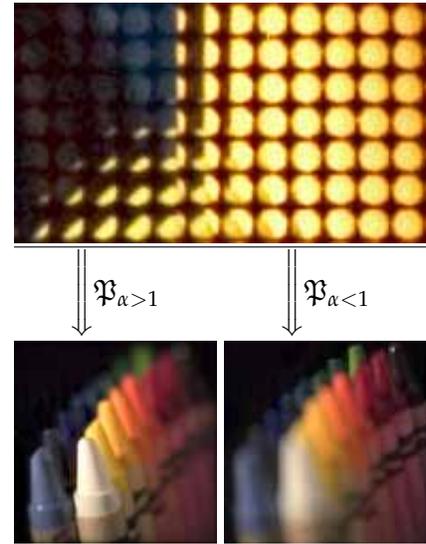
Figure 5.4: Photographic Imaging Operator.

Photographic Imaging Operator Let $\mathcal{P}_\alpha$ be the operator that transforms an in-camera light field parameterized by a separation of depth F into the photograph formed on film at depth $(\alpha \cdot F)$:

$$\mathcal{P}_\alpha [L_F] (x', y') =$$
$$\frac{1}{\alpha^2 F^2} \iint L_F \left(u \left(1 - \frac{1}{\alpha} \right) + \frac{x'}{\alpha}, \ v \left(1 - \frac{1}{\alpha} \right) + \frac{y'}{\alpha}, \ u, \ v \right) \, du \, dv. \qquad (5.1)$$

This operator is what is implemented by a digital refocusing code that accepts light fields and computes refocused photographs (Figure 5.4). As noted in Chapter 2, the operator can be thought of as shearing the 4D space, and then projecting down to 2D.

As described in the chapter overview, the key to analyzing the imaging operator is the Fourier Slice Theorem. The classical version of the Fourier Slice Theorem [Deans 1983] states that a 1D slice of a 2D function's Fourier spectrum is the Fourier transform of an orthographic integral projection of the 2D function. The slicing line is perpendicular to the projection lines, as illustrated in Figure 5.5. Conceptually, the theorem works because the value at the origin of frequency space gives the DC value (integrated value) of the signal, and rotations do not fundamentally change this fact. From this perspective, it makes sense that the theorem generalizes to higher dimensions. It also makes sense that the theorem works for shearing operations as well as rotations, because shearing a space is equivalent to rotating and dilating the space.

These observations mean that we can expect that photographic imaging, which we have observed is a shear followed by projection, should be proportional to a dilated 2D slice of the light field's 4D Fourier transform. With this intuition in mind, Sections 5.3.1 and 5.3.2 are simply the mathematical derivations in specifying this slice precisely, culminating in Equations 5.6 and 5.7.

5.3.1   Generalization of the Fourier Slice Theorem

Let us first digress to study a generalization of the theorem to higher dimensions and projections, so that we can apply it in our 4D space. A closely related generalization is given by the partial Radon transform [Liang and Munson 1997], which handles orthographic projections from N dimensions down to M dimensions. The generalization here formulates a broader class of projections and slices of a function as canonical projection or slicing following an appropriate *change of basis* (e.g. an N-dimensional rotation or shear). This approach is embodied in the following operator definitions.

> **Integral Projection** Let $\mathcal{I}_M^N$ be the canonical projection operator that reduces an N-dimensional function down to M-dimensions by integrating out the last $N - M$ dimensions: $\mathcal{I}_M^N [f] (x_1, \ldots, x_M) = \int f(x_1, \ldots, x_N) \, dx_{M+1} \ldots dx_N$.

> **Slicing** Let $\mathcal{S}_M^N$ be the canonical slicing operator that reduces an N-dimensional function down to an M dimensional one by zero-ing out the last $N - M$ dimensions: $\mathcal{S}_M^N [f] (x_1, \ldots, x_M) = f(x_1, \ldots, x_M, 0, \ldots, 0)$.

Change of Basis Let $\mathcal{B}$ denote an operator for an arbitrary change of basis of an N-dimensional function. It is convenient to also allow $\mathcal{B}$ to act on N-dimensional column vectors as an $N \times N$ matrix, so that $\mathcal{B}[f](\mathbf{x}) = f(\mathcal{B}^{-1}\mathbf{x})$, where $\mathbf{x}$ is an N-dimensional column vector, and $\mathcal{B}^{-1}$ is the inverse of $\mathcal{B}$.

Scaling Let any scalar a be used to denote the operator that scales a function by that constant, so that $a[f](\mathbf{x}) = a \cdot f(\mathbf{x})$.

Fourier Transform Let $\mathcal{F}^N$ denote the N-dimensional Fourier transform operator, and let $\mathcal{F}^{-N}$ be its inverse. $\mathcal{F}^N[f](\mathbf{u}) = \int f(\mathbf{x}) \exp\left(-2\pi i \left(\mathbf{x} \cdot \mathbf{u}\right)\right) d\mathbf{x}$, where $\mathbf{x}$ and $\mathbf{u}$ are N-dimensional vectors, and the integral is taken over all of N-dimensional space.

With these definitions, we can state a generalization of the classical theorem as follows:

GENERALIZED FOURIER SLICE THEOREM *Let f be an N-dimensional function. If we change the basis of f, integral-project it down to M of its dimensions, and Fourier transform the resulting function, the result is equivalent to Fourier transforming f, changing the basis with the normalized inverse transpose of the original basis, and slicing it down to M dimensions. Compactly in terms of operators, the theorem says:*

$$\mathcal{F}^M \circ \mathcal{I}_M^N \circ \mathcal{B} = \mathcal{S}_M^N \circ \frac{\mathcal{B}^{-T}}{|\mathcal{B}^{-T}|} \circ \mathcal{F}^N, \tag{5.2}$$

where the transpose of the inverse of $\mathcal{B}$ is denoted by $\mathcal{B}^{-T}$, and $|\mathcal{B}^{-T}|$ is its scalar determinant.

A proof of the theorem is presented in Appendix A.1.

Figure 5.6 summarizes the relationships that are implied by the theorem between the N-dimensional signal, M-dimensional projected signal, and their Fourier spectra. One point to note about the theorem is that it reduces to the classical version (compare Figures 5.5 and 5.6) for $N = 2$, $M = 1$ and the change of basis being a 2D rotation matrix ($\mathcal{B} = \mathcal{R}_\theta$). In this case, the rotation matrix is its own inverse transpose ($\mathcal{R}_\theta = \mathcal{R}_\theta{}^{-T}$), and the determinant $\left|\mathcal{R}_\theta{}^{-T}\right|$ equals 1. As a result, the basis change in the Fourier domain is the same as in the spatial domain.
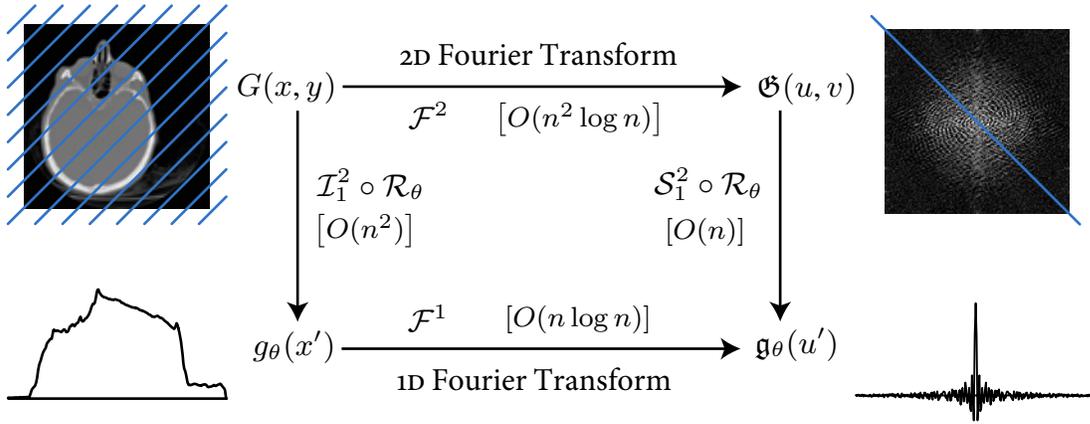
Figure 5.5: Classical Fourier Slice Theorem, in terms of the operator notation used in this chapter. Computational complexities for each transform are given in square brackets, assuming n samples in each dimension.
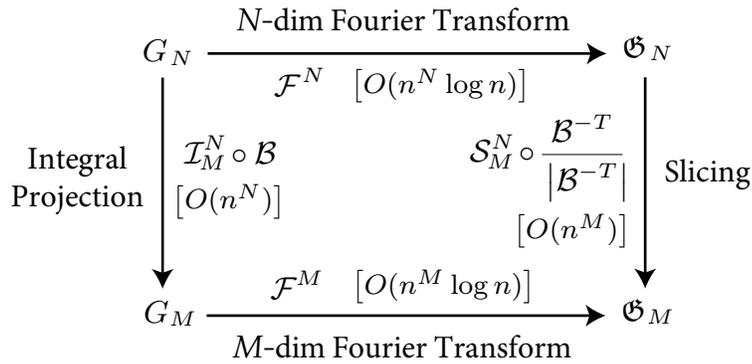
Figure 5.6: Generalized Fourier Slice Theorem (Equation 5.2). Transform relationships between an N-dimensional function G_N, an M-dimensional integral projection of it, G_M, and their respective Fourier spectra, $\mathfrak{G}_N$ and $\mathfrak{G}_M$.

This is a special case, however, and in general the Fourier slice is taken with the normalized transpose of the inverse basis, $(\mathcal{B}^{-T} / |\mathcal{B}^{-T}|)$. In 2D, this fact is a special case of the so-called Affine Theorem for Fourier transforms [Bracewell et al. 1993]. It is also related to the well-known fact in geometry that transforming a surface by a particular matrix means transforming the surface normals by the transpose of the inverse matrix. This is an important factor that is taken into account in computing the shaded color of surfaces in computer

graphics, for example. Readers interested in understanding why this occurs in the theorem may consult the proof in Appendix A.1. As a final point, dividing out by the (scalar) determinant of the matrix normalizes the equations for changes in the integrals due to dilations of the space caused by the basis change.

The theorem provides both computational and theoretical advantages. From the computational perspective, the theorem leads directly to a fast method for calculating orthographic projections of an N-dimensional function, G_N, if we have its Fourier spectrum, $\mathfrak{G}_N$. For example, assume we are trying to compute a projection down to M dimensions, to obtain function G_M. As shown in Figure 5.6, naïve projection via numerical integration (left downward arrow) takes $O(n^N)$ time, where we assume that there are n samples in each of the N dimensions. A faster approach is extracting a slice of $\mathfrak{G}_N$ (right downward arrow) and applying an inverse Fourier transform (bottom leftward arrow), which takes only $O(n^M \log n)$ via the Fast Fourier Transform algorithm.

From the theoretical perspective, the theorem is an analytic tool for characterizing the spectral information content of a particular projection of a function. For example, in the case of CT scanning, it allows us to analyze whether a family of x-ray projections fully characterizes the densities of the tissue volume that is being scanned. In general, the theorem implies that a family of projections fully defines a function if the family of corresponding Fourier slices spans the Fourier transform of the function. We will return to this issue later in considering whether the set of all photographs focused at all depths captures the same information as a light field.

5.3.2 Fourier Slice Photograph Theorem

This section derives the Fourier Slice Photograph Theorem, which lies at the heart of this signal processing framework for light field photography. This theorem factors the Imaging Operator using the Generalized Fourier Slice Theorem.

The first step is to recognize that the Imaging Operator (Equation 5.1) indeed corresponds to integral projection of the light field following a change of basis (shear):

$$\mathcal{P}_\alpha \left[L_F \right] = \frac{1}{\alpha^2 F^2} \, \mathcal{I}_2^4 \circ \mathcal{B}_\alpha \left[L_F \right],$$

(5.3)

which relies on the following specific change of basis:

Imaging Change of Basis $\mathcal{B}_\alpha$ is a 4D change of basis defined by the following matrices:

$$\mathcal{B}_\alpha = \begin{bmatrix} \alpha & 0 & 1-\alpha & 0 \\ 0 & \alpha & 0 & 1-\alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathcal{B}_\alpha{}^{-1} = \begin{bmatrix} 1/\alpha & 0 & 1-1/\alpha & 0 \\ 0 & 1/\alpha & 0 & 1-1/\alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that $\mathcal{B}_\alpha$ and $\mathcal{B}_\alpha^{-1}$ should not be confused with the unscripted symbol, $\mathcal{B}$, used for a *generic* change of basis in the statement of the Generalized Fourier Slice Theorem. Directly applying these definitions and the definition for $\mathcal{I}_2^4$ verifies that Equation 5.3 is consistent with Equation 5.1.

We can now apply the Generalized Fourier Slice Theorem (Equation 5.2) to turn the integral projection in Equation 5.3 into a Fourier-domain slice. First, let us re-write Equation 5.3 as

$$\mathcal{P}_\alpha = \frac{1}{\alpha^2 F^2}\, \mathcal{F}^{-2} \circ (\mathcal{F}^2 \circ \mathcal{I}_2^4 \circ \mathcal{B}_\alpha).$$

In this form, the Generalized Fourier Slice Theorem (Equation 5.2) applies directly to the terms in brackets, allowing us to write

$$\mathcal{P}_\alpha = \frac{1}{\alpha^2 F^2}\, \mathcal{F}^{-2} \circ \left(\mathcal{S}_2^4 \circ \frac{\mathcal{B}_\alpha{}^{-T}}{\left| \mathcal{B}_\alpha{}^{-T} \right|} \circ \mathcal{F}^4 \right).$$

Finally, noting that $\left| \mathcal{B}_\alpha{}^{-T} \right| = \left| \mathcal{B}_\alpha^{-1} \right| = 1/\alpha^2$, we arrive at the following result:

$$\mathcal{P}_\alpha = \frac{1}{F^2}\, \mathcal{F}^{-2} \circ \mathcal{S}_2^4 \circ \mathcal{B}_\alpha{}^{-T} \circ \mathcal{F}^4, \tag{5.4}$$

namely that a photograph ($\mathcal{P}_\alpha$) is obtained from the 4D Fourier spectrum of the light field by: extracting an appropriate 2D slice ($\mathcal{S}_2^4 \circ \mathcal{B}_\alpha{}^{-T}$), applying an inverse 2D transform ($\mathcal{F}^{-2}$), and scaling the resulting image ($1/F^2$).

Before stating the final theorem, let us define one last operator that combines all the action of photographic imaging in the Fourier domain:

Fourier Photographic Imaging Operator

$$\mathfrak{P}_\alpha = \frac{1}{F^2} \, \mathcal{S}_2^4 \circ \mathcal{B}_\alpha^{-T}. \tag{5.5}$$

It is easy to verify that $\mathfrak{P}_\alpha$ has the following explicit form, directly from the definitions of $\mathcal{S}_2^4$ and $\mathcal{B}_\alpha$. This explicit form is required for calculations:

$$\mathfrak{P}_\alpha[G](k_x, k_y) = \frac{1}{F^2} \, G(\alpha \cdot k_x, \; \alpha \cdot k_y, \; (1-\alpha) \cdot k_x, \; (1-\alpha) \cdot k_y). \tag{5.6}$$

Applying Equation 5.5 to Equation 5.4 brings us, finally, to our goal:

FOURIER SLICE PHOTOGRAPH THEOREM

$$\mathcal{P}_\alpha = \mathcal{F}^{-2} \circ \mathfrak{P}_\alpha \circ \mathcal{F}^4. \tag{5.7}$$

A photograph is the inverse 2D Fourier transform of a dilated 2D slice in the 4D Fourier transform of the light field.

Figure 5.7 illustrates the relationships between light fields and photographs that are implied by this theorem. The figure makes it clear that $\mathfrak{P}_\alpha$ is the Fourier-dual to $\mathcal{P}_\alpha$. The left half of the diagram represents quantities in the spatial domain, and the right half is the Fourier domain. In other words, $\mathfrak{P}_\alpha$ acts exclusively in the Fourier domain to produce the Fourier spectrum of a refocused photograph from the Fourier spectrum of the light field. It is worth emphasizing that the derivation of the theorem is rooted in geometrical optics and radiometry, and it is consistent with the physics of image formation expressed by these models of optics.

The value of the theorem lies in the fact that $\mathfrak{P}_\alpha$, a slicing operator, is conceptually simpler than $\mathcal{P}_\alpha$, an integral operator. This point is made especially clear by reviewing the explicit definitions of $\mathfrak{P}_\alpha$ (Equation 5.6) and $\mathcal{P}_\alpha$ (Equation 5.1). By providing a Fourier-based interpretation, the theorem provides two equivalent but very different perspectives on image formation. In this regard, the Fourier Slice Photograph Theorem is not unlike the Convolution Theorem, which provides different viewpoints on filtering in the two domains.

From a practical standpoint, the theorem provides a faster computational pathway for certain kinds of light field processing. The computational complexities for each transform
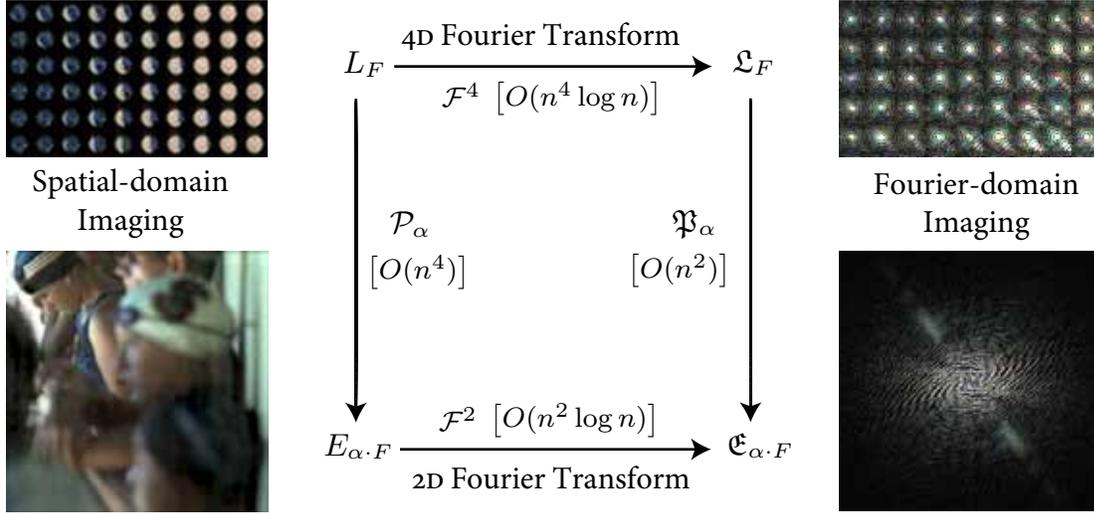
Figure 5.7: Fourier Slice Photograph Theorem. Transform relationships between the 4D light field L_F, a lens-formed 2D photograph $E_{\alpha \cdot F}$, and their respective Fourier spectra, $\mathfrak{L}_F$ and $\mathfrak{E}_{\alpha \cdot F}$.

are illustrated in Figure 5.7, assuming a resolution of n in each dimension of the 4D light field. The most salient point is that slicing via $\mathfrak{P}_\alpha$ ($O(n^2)$) is asymptotically faster than integration via $\mathcal{P}_\alpha$ ($O(n^4)$). This fact is the basis for the algorithm in Section 5.5.

5.3.3   Photographic Effect of Filtering the Light Field

A light field produces exact photographs focused at various depths via Equation 5.1. If we distort the light field by filtering it, and then form photographs from the distorted light field, how are these photographs related to the original, exact photographs? The following theorem provides the answer to this question.

FILTERED LIGHT FIELD IMAGING THEOREM   *A 4D convolution of a light field results in a 2D convolution of each photograph. The 2D filter kernel is simply the photograph of the 4D filter kernel focused at the same depth. Compactly in terms of operators,*

$$\mathcal{P}_\alpha \circ \mathcal{C}_h^4 = \mathcal{C}_{\mathcal{P}_\alpha[h]}^2 \circ \mathcal{P}_\alpha, \tag{5.8}$$

where we have expressed convolution with the following operator:

Convolution $\mathcal{C}_h^N$ is an N-dimensional convolution operator with filter kernel h, such that $\mathcal{C}_h^N[F](\mathbf{x}) = \int F(\mathbf{x} - \mathbf{u})\,h(\mathbf{u})\,d\mathbf{u}$ where $\mathbf{x}$ and $\mathbf{u}$ are N-dimensional vector coordinates, and F and h are N-dimensional functions.

Figure 5.8 illustrates the theorem diagrammatically. On the diagram, L_F is the input 4D light field, and $\overline{L}_F$ is a 4D filtering of it with 4D kernel h. $E_{\alpha \cdot F}$ and $\overline{E}_{\alpha \cdot F}$ are the photographs formed from the two light fields, respectively. The theorem states that $\overline{E}_{\alpha \cdot F}$ is a 2D filtering of $E_{\alpha \cdot F}$, where the 2D kernel is the photograph of the 4D kernel, h.

In spite of its plausibility, the theorem is not obvious, and proving it in the spatial domain is quite difficult. Appendix A.2 presents a proof of the theorem. At a high level, the approach is to apply the Fourier Slice Photography Theorem and the Convolution Theorem to move the analysis into the Fourier domain. In that domain, photograph formation turns into a simpler slicing operator, and convolution turns into a simpler multiplication operation.

In the next section we will use this theorem to theoretically analyze a simple model of
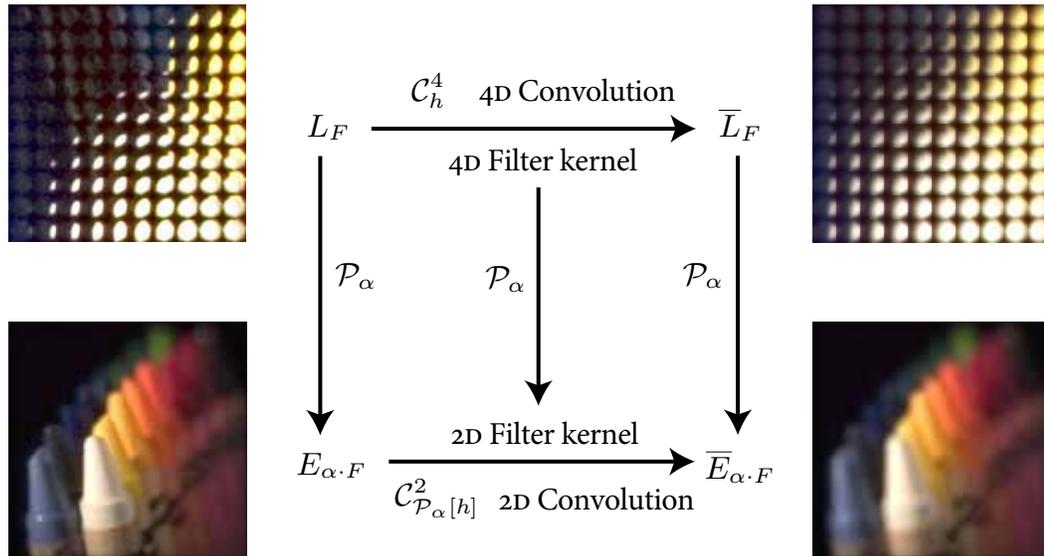
Figure 5.8: Filtered Light Field Photography Theorem. Transform relationships between a 4D light field L_F, a filtered version of the light field, $\overline{L}_F$, and photographs $E_{\alpha \cdot F}$ and $\overline{E}_{\alpha \cdot F}$.

digital refocusing using the plenoptic camera. As an example of how to apply the theorem, let us consider trying to compute photographs from a light field that has been convolved by a box filter that is of unit value within the unit box about the origin. In 4D, the box function is defined by $\sqcap(x, y, u, v) = \sqcap(x) \sqcap(y) \sqcap(u) \sqcap(v)$, where the 1D box is defined by:

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

With $\sqcap(x, y, u, v)$ as the filter, the theorem shows that

$$\begin{aligned} \mathcal{P}_\alpha \left[L_F **** \sqcap \right] &= \left(\mathcal{P}_\alpha \circ \mathcal{C}_\sqcap^4 \right) [L_F] \\ &= \left(\mathcal{C}_{\mathcal{P}_\alpha[\sqcap]}^2 \circ \mathcal{P}_\alpha \right) [L_F] \\ &= \mathcal{P}_\alpha [\sqcap] ** \mathcal{P}_\alpha [L_F]. \end{aligned}$$

For explicitness, these equations use the "star" notation for convolution, such that $f **** g$ represents the 4D convolution of 4D functions f and g, and $a ** b$ represents the 2D convolution of 2D functions a and b.

The left hand side of the equation is the photograph computed from the convolved light field. The right hand side is the exact photograph from the unfiltered light field ($\mathcal{P}_\alpha [L_F]$) convolved by the 2D kernel function $\mathcal{P}_\alpha [\sqcap]$. This kernel is the 2D photograph of the box filter treated as a 4D light field. What exactly is a photograph of a 4D box light field? The following diagram visualizes things in terms of a 2D box light field.
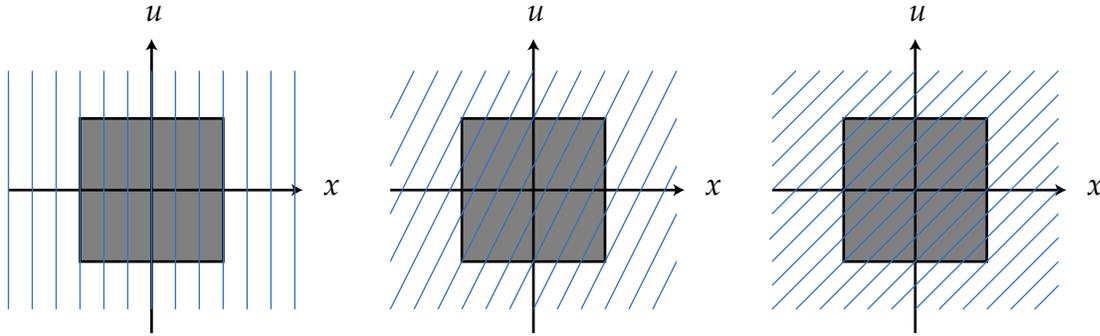
The blue lines show the projection trajectories for focusing photographs at three different

depths. The resulting 1D projected photographs are a box function, a flat-top pyramid, and a triangle function, as shown in the following diagram.
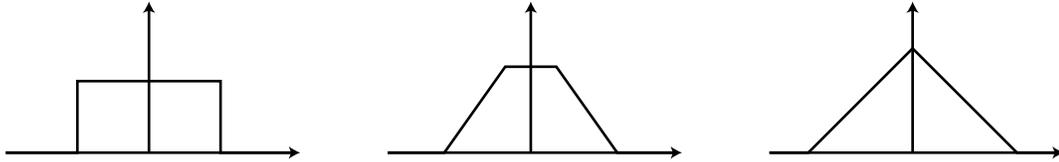
By analogy in the 4D light field, the 2D blur kernel over the exact photograph is a 2D box function, a tent function or a flat-top tent function. Exactly which depends on the amount of refocusing, α, as in the 2D versions above.

This example illustrates how the Filtered Light Field Imaging Theorem can be used as a tool for assessing the design of light field imaging systems. For example, in studying light field acquisition devices, such as the plenoptic camera, the impulse response of the recording system is the 4D filter kernel in the theorem statement, $h(x, y, u, v)$. As another example, in processing the light field, the resampling strategy (e.g. quadrilinear or windowed sinc interpolation) defines $h(x, y, u, v)$. Given this kernel, the theorem shows that the resulting filter over the output photograph is simply $\mathcal{P}_\alpha[h]$. Computing this 2D kernel is very practical: it uses the same code as computing refocused photographs from the light field. Analyzing the changes in the filter over output photographs provides a simple and practical procedure for optimizing the design of light field camera systems.

Of course this analysis is an idealized view, and the model does not capture all the features of real systems. First, real systems are unlikely to be completely linear, and one can draw reasonable conclusions from this kind of Fourier analysis only if a meaningful approximation to a system impulse response exists. This limitation makes it difficult to apply these techniques to analysis of digital lens correction (Chapter 7), for example, where the resampling strategy can be highly spatially variant. A second limitation is that real discrete systems are not band-limited, so there will inherently be some aliasing which is not properly modelled by the convolution framework above. In spite of these limitations, the signal-processing style of thinking developed here provides a new perspective to study light field imaging, and reveals many insights that are not at all clear in the spatial domain.

5.4   Band-Limited Analysis of Refocusing Performance

This section applies the Filtered Light Field Imaging Theorem to digital refocusing from a plenoptic camera, to answer the following questions. What is the quality of a photograph refocused from a recorded light field? How is this photograph related to the exact photograph, such as the one that might have been taken by a conventional camera that were optically focused at the same depth?

The central assumption here, as introduced in Section 5.2, is that the plenoptic camera captures band-limited light fields. Section 5.2 summarized the intuition in terms of the intersection of the ideal refocusing lines with the rectangular bounds of the light field's bandwidth. This section presents the algebraic details of this analysis, working in the spatial domain and using the Filtered Light Field Imaging Theorem.

The band-limited assumption means that the Fourier spectrum of the recorded light field is multiplied by a dilated version of the 4D box filter described earlier, $\sqcap(x, y, u, v)$. By the Convolution Theorem, multiplication by the box function in the Fourier domain means that the signal is convolved by the Fourier transform of the box function in the spatial domain. It is well known that the Fourier transform of the box function is the perfect low-pass filter, the sinc function. Let us adopt multi-dimensional notation for the sinc function also, so that $\text{sinc}(x, y, u, v) = \text{sinc}(x)\,\text{sinc}(y)\,\text{sinc}(u)\,\text{sinc}(v)$, with

$$\text{sinc}(x) = \frac{\sin \pi x}{\pi x}.$$

In other words, the band-limited assumption means that the recorded light field, $\hat{L}_F$, is simply the *exact* light field, L_F, convolved by a 4D sinc:

$$\hat{L}_F = \mathcal{C}^4_{\text{lowpass}}[L_F], \text{ where}$$

$$\text{lowpass}(x, y, u, v) = \frac{1}{(\Delta x \Delta u)^2} \cdot \text{sinc}\left(\frac{x}{\Delta x}, \frac{y}{\Delta x}, \frac{u}{\Delta u}, \frac{v}{\Delta u}\right).$$

In this equation, Δx and Δu are the linear spatial and directional sampling rates of the plenoptic camera, respectively. The $1/(\Delta x \Delta u)^2$ is an energy-normalizing constant to account for dilation of the sinc.

Analytic Form for Refocused Photographs

Our goal is an analytic solution for the digitally refocused photograph, $\hat{E}_{F'}$, computed from the band-limited light field, L_F. This is where we apply the Filtered Light Field Photography Theorem. Letting $\alpha = F'/F$,

$$\hat{E}_{F'} = \mathcal{P}_\alpha \left[\hat{L}_F \right] = \mathcal{P}_\alpha \left[\mathcal{C}^4_{\text{lowpass}} \left[L_F \right] \right]$$
$$= \mathcal{C}^2_{\mathcal{P}_\alpha [\text{lowpass}]} \left[\mathcal{P}_\alpha \left[L_F \right] \right] = \mathcal{C}^2_{\mathcal{P}_\alpha [\text{lowpass}]} \left[E_{F'} \right],$$

where $E_{F'}$ is the *exact* photograph at depth F'. This derivation shows that the digitally refocused photograph is a 2D-filtered version of the exact photograph. The 2D kernel is simply a photograph of the 4D sinc function interpreted as a light field, $\mathcal{P}_\alpha [\text{lowpass}]$.

It turns out that photographs of a 4D sinc light field are simply 2D sinc functions:

$$\mathcal{P}_\alpha [\text{lowpass}] = \mathcal{P}_\alpha \left[\frac{1}{(\Delta x \Delta u)^2} \cdot \text{sinc} \left(\frac{x}{\Delta x}, \frac{y}{\Delta x}, \frac{u}{\Delta u}, \frac{v}{\Delta u} \right) \right]$$
$$= \frac{1}{D_x^2} \cdot \text{sinc} \left(\frac{x}{D_x}, \frac{y}{D_x} \right), \tag{5.9}$$

where the Nyquist rate of the 2D sinc depends on the amount of refocusing, α:

$$D_x = \max(\alpha \Delta x, |1 - \alpha| \Delta u). \tag{5.10}$$

This fact is difficult to derive in the spatial domain, but applying the Fourier Slice Photograph Theorem moves the analysis into the frequency domain, where it is easy (see Appendix A.3).

The critical point here is that since the 2D kernel is a sinc, the digitally refocused photographs are just band-limited versions of the exact photographs. The performance of digital refocusing is therefore defined by the variation of the 2D kernel bandwidth (Equation 5.10) with the extent of refocusing.

Interpretation of Refocusing Performance

Recall that the spatial and directional sampling rates of the camera are Δx and Δu. Let us further define the width of the camera sensor as W_x, and the width of the lens aperture as

W_u. With these definitions, the spatial resolution of the sensors is $N_x = W_x/\Delta x$ and the directional resolution of the light field camera is $N_u = W_u/\Delta u$.

Since $\alpha = (F'/F)$ and $\Delta u = W_u/N_u$, it is easy to verify that

$$|\alpha\Delta x| \geq |(1-\alpha)\Delta u|$$
$$\Leftrightarrow \quad |F'-F| \leq \frac{\Delta x N_u F'}{W_u}. \tag{5.11}$$

The claim here is that this is the range of focal depths, F, where we can achieve "exact" refocusing, i.e. compute a sharp rendering of the photograph focused at that depth. What we are interested in is the Nyquist-limited resolution of the photograph, which is the number of band-limited samples within the field of view.

Precisely, by applying Equation 5.11 to Equation 5.10, we see that the bandwidth of the computed photograph is $(\alpha\Delta x)$. Next, the field of view is not simply the size of the light field sensor, W_x, but rather (αW_x). This dilation is due to the fact that digital refocusing scales the image captured on the sensor by a factor of α in projecting it onto the virtual focal plane (see Equation 5.1). If $\alpha > 1$, for example, the light field camera image is zoomed in slightly compared to the conventional camera, the telecentric effect discussed in Section 4.2.

Thus, the Nyquist resolution of the computed photograph is

$$\frac{\alpha W_x}{\alpha\Delta x} = \frac{W_x}{\Delta x}. \tag{5.12}$$

This is simply the spatial resolution of the camera, the maximum possible resolution for the output photograph. This justifies the assertion that digital refocusing is "exact" for the range of depths defined by Equation 5.11. Note that this range of exact refocusing increases linearly with the directional resolution, N_u, as described in Section 4.3.

If we exceed the exact refocusing range, i.e.

$$|F'-F| > \frac{\Delta x N_u F'}{W_u},$$

then the band-limit of the computed photograph, $\hat{E}_{F'}$, is $|1-\alpha|\Delta u$, which will be larger than

$\alpha\Delta x$ (see Equation 5.10). The resulting resolution is not maximal, but rather

$$\frac{\alpha W_x}{|1 - \alpha|\Delta u},$$

which is less than the spatial resolution of the light field sensor, $W_x/\Delta x$. In other words, the resulting photograph is blurred, with reduced Nyquist-limited resolution.

Re-writing this resolution in a slightly different form provides a more intuitive interpretation of the amount of blur. Since $\alpha = F'/F$ and $\Delta u = W_u/N_u$, the resolution is

$$\frac{\alpha}{|1 - \alpha|}\frac{W_x}{\Delta u} = \frac{W_x}{\frac{W_u}{N_u \cdot F'} \cdot |F' - F|}. \tag{5.13}$$

Because $\frac{N_u F'}{W_u}$ is the f-number of a lens N_u times smaller than the actual lens used on the camera, we can now interpret $\frac{W_u}{N_u \cdot F'} \cdot |F' - F|$ as the size of the conventional circle of confusion cast through this smaller lens when the film plane is mis-focused by a distance of $|F' - F|$.

In other words, when refocusing beyond the exact range, we can only make the desired focal plane appear as sharp as it appears in a conventional photograph focused at the original depth, with a lens N_u times smaller, as described in Section 4.3. Note that the sharpness increases linearly with the directional resolution, N_u. During exact refocusing, it simply turns out that the resulting circle of confusion falls within one pixel and the spatial resolution completely dominates.

In summary, a band-limited assumption about the recorded light fields enables a mathematical analysis that corroborates the geometric intuition of refocusing performance presented in Chapter 4.

5.5 Fourier Slice Digital Refocusing

This section applies the Fourier Slice Photograph Theorem in a very different way, to derive an asymptotically fast algorithm for digital refocusing. The presumed usage scenario is as follows: an in-camera light field is available (perhaps having been captured by a plenoptic camera). The user wishes to digitally refocus in an interactive manner, i.e. select a desired

focal plane and view a synthetic photograph focused on that plane.

In previous approaches to this problem [Isaksen et al. 2000; Levoy et al. 2004; Ng et al. 2005; Vaish et al. 2004], spatial integration as described in Chapter 4 results in an $O(n^4)$ algorithm, where n is the number of samples in each of the four dimensions. The algorithm described in this section provides a faster $O(n^2 \log n)$ algorithm, with the penalty of a single $O(n^4 \log n)$ pre-processing step.

Algorithm

The algorithm follows trivially from the Fourier Slice Photograph Theorem. Figure 5.9 illustrates the steps of the algorithm.

PRE-PROCESS

- Prepare the given light field, L_F, by pre-computing its 4D Fourier transform, $\mathcal{F}^4[L_F]$, via the Fast Fourier Transform. This step takes $O(n^4 \log n)$ time.

REFOCUSING   For each choice of desired virtual film plane at a depth of F'

- Extract the Fourier slice (via Equation 5.6) of the pre-processed Fourier transform, to obtain $(\mathfrak{P}_\alpha \circ \mathcal{F}^4)[L_F]$, where $\alpha = F'/F$. This step takes $O(n^2)$ time.
- Compute the inverse 2D Fourier transform of the slice, to obtain $(\mathcal{F}^{-2} \circ \mathfrak{P}_\alpha \circ \mathcal{F}^4)[L_F]$. By the theorem, this final result is $\mathcal{P}_\alpha[L_F] = E_{F'}$ the photo focused at the desired depth. This step takes $O(n^2 \log n)$ time.

This approach is best used to quickly synthesize a large family of refocused photographs, since the $O(n^2 \log n)$ Fourier-slice method of producing each photograph is asymptotically much faster than the $O(n^4)$ method of brute-force numerical integration via Equation 5.1.

Implementation and Results

The complexity in implementing this simple algorithm has to do with ameliorating the artifacts that result from discretization, resampling and Fourier transformation. Unfortunately, our eyes tend to be very sensitive to the kinds of ringing artifacts that are easily introduced by Fourier-domain image processing. These artifacts are conceptually similar to the issues tackled in Fourier volume rendering [Levoy 1992; Malzbender 1993], and Fourier-based medical reconstruction techniques [Jackson et al. 1991] such as those used in CT and MR.
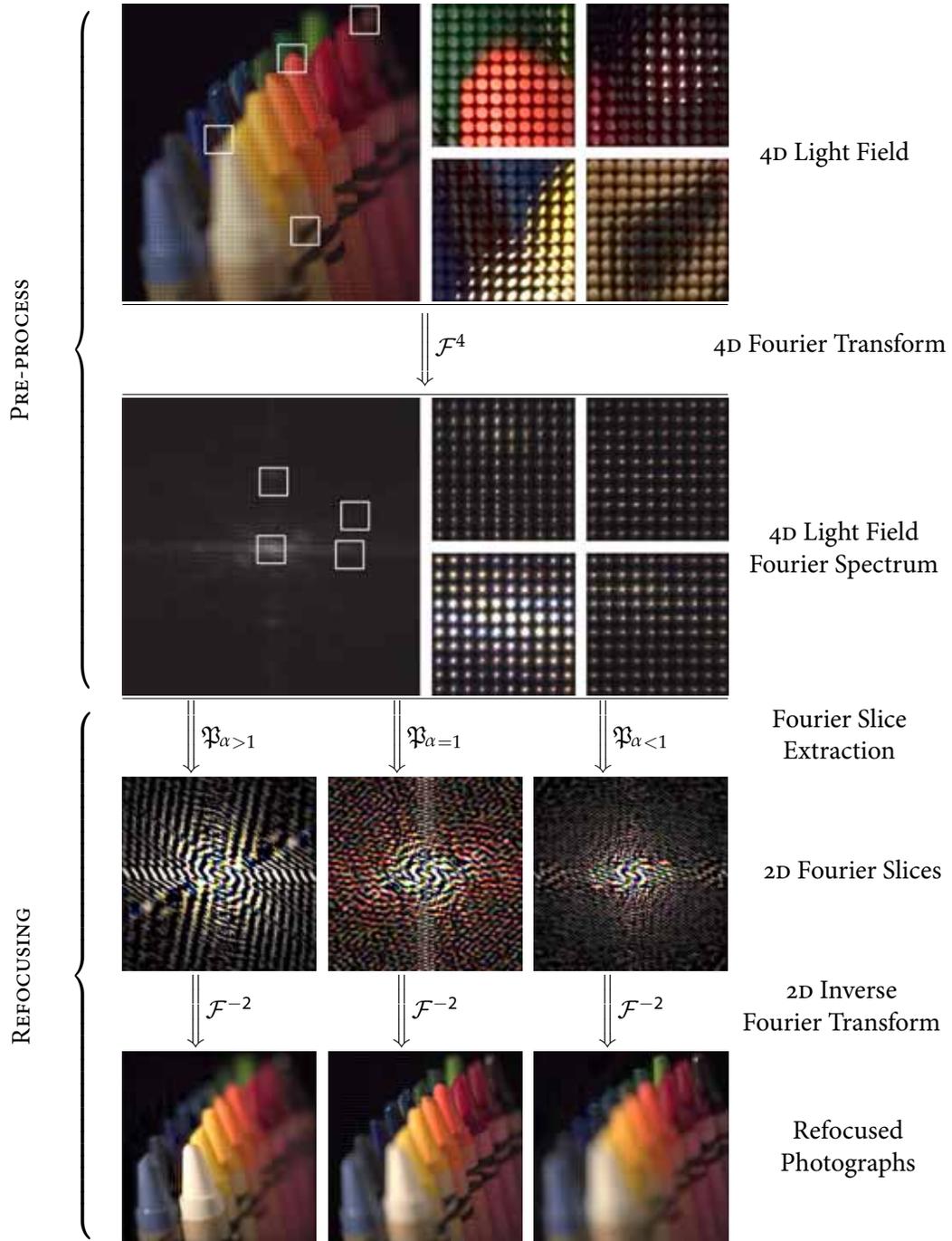
Figure 5.9: Fourier Slice Digital Refocusing algorithm.

Sophisticated signal processing techniques have been developed by these communities to address these problems. The sections below describe the most important issues for digital refocusing, and how to address them with adaptations of the appropriate signal processing methods.

Sources of Artifacts

In general signal-processing terms, when we sample a signal it is replicated periodically in the dual domain. When we reconstruct this sampled signal with convolution, it is multiplied in the dual domain by the Fourier transform of the convolution filter. The goal is to perfectly isolate the original, central replica, eliminating all other replicas. This means that the ideal filter is band-limited: it is of unit value for frequencies within the support of the light field, and zero for all other frequencies. Thus, the ideal filter is the sinc function, which has infinite extent.

(A): Spatial domain   (B): Fourier domain
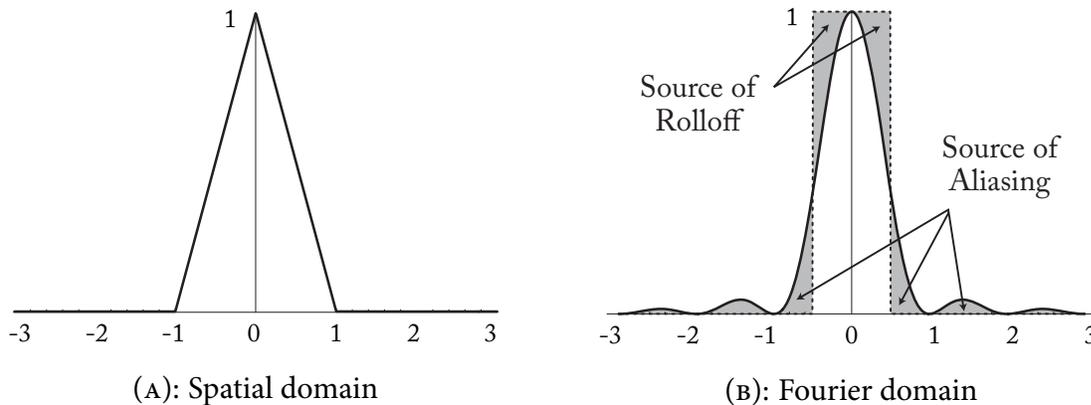
Figure 5.10: Source of artifacts. Example bilinear reconstruction filter (A), and frequency spectrum (solid line on B) compared to ideal spectrum (dotted line).

In practice we must use an imperfect, finite-extent filter, which will exhibit two important defects (Figure 5.10). First, the filter will not be of unit value within the band-limit, instead gradually decaying to smaller fractional values as the frequency increases. Second, the filter will not be truly band-limited, containing energy at frequencies outside the desired stop-band. Figure 5.10B illustrates these deviations as shaded regions compared to the ideal filter spectrum.

| (A): Reference image | (B): Rolloff artifacts | (C): Aliasing artifacts |
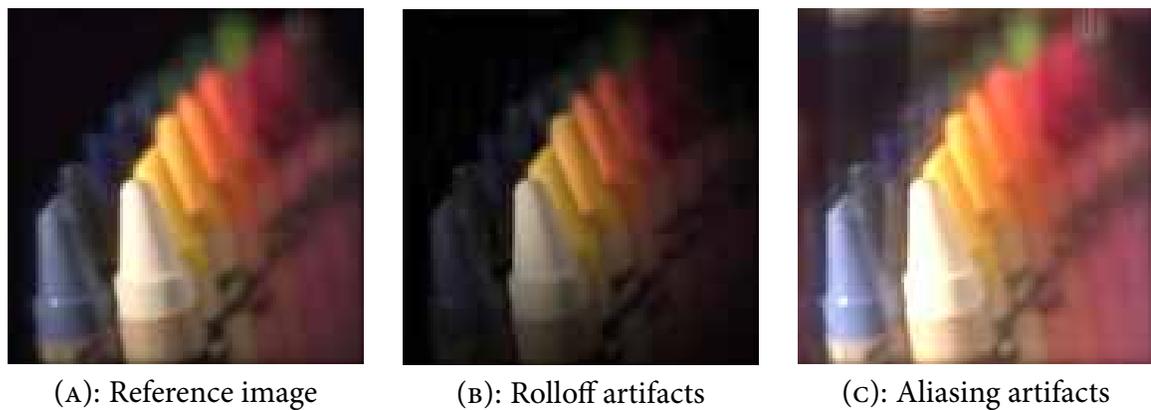
Figure 5.11: Two main classes of artifacts.

The first defect leads to so-called *rolloff artifacts* [Jackson et al. 1991]. The most obvious manifestation is a darkening of the borders of computed photographs. Figure 5.11B illustrates this roll-off with the use of the Kaiser-Bessel filter described below. Decay in the filter's frequency spectrum with increasing frequency means that the spatial light field values, which are modulated by this spectrum, also "roll off" to fractional values towards the edges. The reference image in Figure 5.11A was computed with spatial integration via Equation 5.1.

The second defect, energy at frequencies above the band-limit, leads to *aliasing artifacts* (post-aliasing, in the terminology of Mitchell and Netravali [1998]) in computed photographs. The non-zero energy beyond the band-limit means that the periodic replicas are not fully eliminated, leading to two kinds of aliasing. First, the replicas that appear parallel to the slicing plane appear as 2D replicas of the image encroaching on the borders of the final photograph. Second, the replicas positioned perpendicular to this plane are projected and summed onto the image plane, creating ghosting and loss of contrast. Figure 5.11 illustrates these artifacts when the filter is quadrilinear interpolation.

Correcting Rolloff Error

Rolloff error is a well understood effect in medical imaging and Fourier volume rendering. The standard solution is to multiply the affected signal by the reciprocal of the filter's inverse Fourier spectrum, to nullify the effect introduced during resampling. In our case, directly

(A): Without pre-multiplication   (B): With pre-multiplication
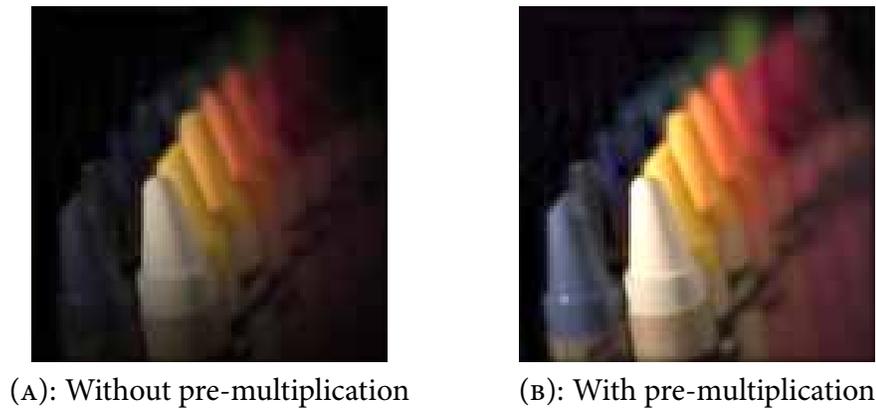
Figure 5.12: Rolloff correction.

analogously to Fourier volume rendering [Malzbender 1993], the solution is to spatially pre-multiply the input light field by the reciprocal of the filter's 4D inverse Fourier transform. This is performed prior to taking its 4D Fourier transform in the pre-processing step of the algorithm. Figure 5.12 illustrates the effect of pre-multiplication for the example of a Kaiser-Bessel resampling filter, described in the next subsection.

Unfortunately, this pre-multiplication tends to accentuate the energy of the light field near its borders, maximizing the energy that folds back into the desired field of view as aliasing.

Suppressing Aliasing Artifacts

The three main methods of suppressing aliasing artifacts are oversampling, superior filtering and zero-padding. Oversampling means drawing more finely spaced samples in the frequency domain, extracting a higher-resolution 2D Fourier slice $(\mathfrak{P}_F \circ \mathcal{F}^4)[L_F]$. Increasing the sampling rate in the Fourier domain increases the replication period in the spatial domain. This means that less energy in the tails of the in-plane replicas will fall within the borders of the final photograph.

Exactly what happens computationally will be familiar to those experienced in discrete Fourier transforms. Specifically, increasing the sampling rate in one domain leads to an increase in the field of view in the other domain. Hence, by oversampling we produce an image that shows us more of the world than desired, not a magnified view of the desired

portion. Aliasing energy from neighboring replicas falls into these outer regions, which we crop away to isolate the central image of interest.

(A): 2× oversampling (Fourier domain)

(B): Inverse Fourier transform of (A)

(C): No oversampling

(D): Cropped version of (B)
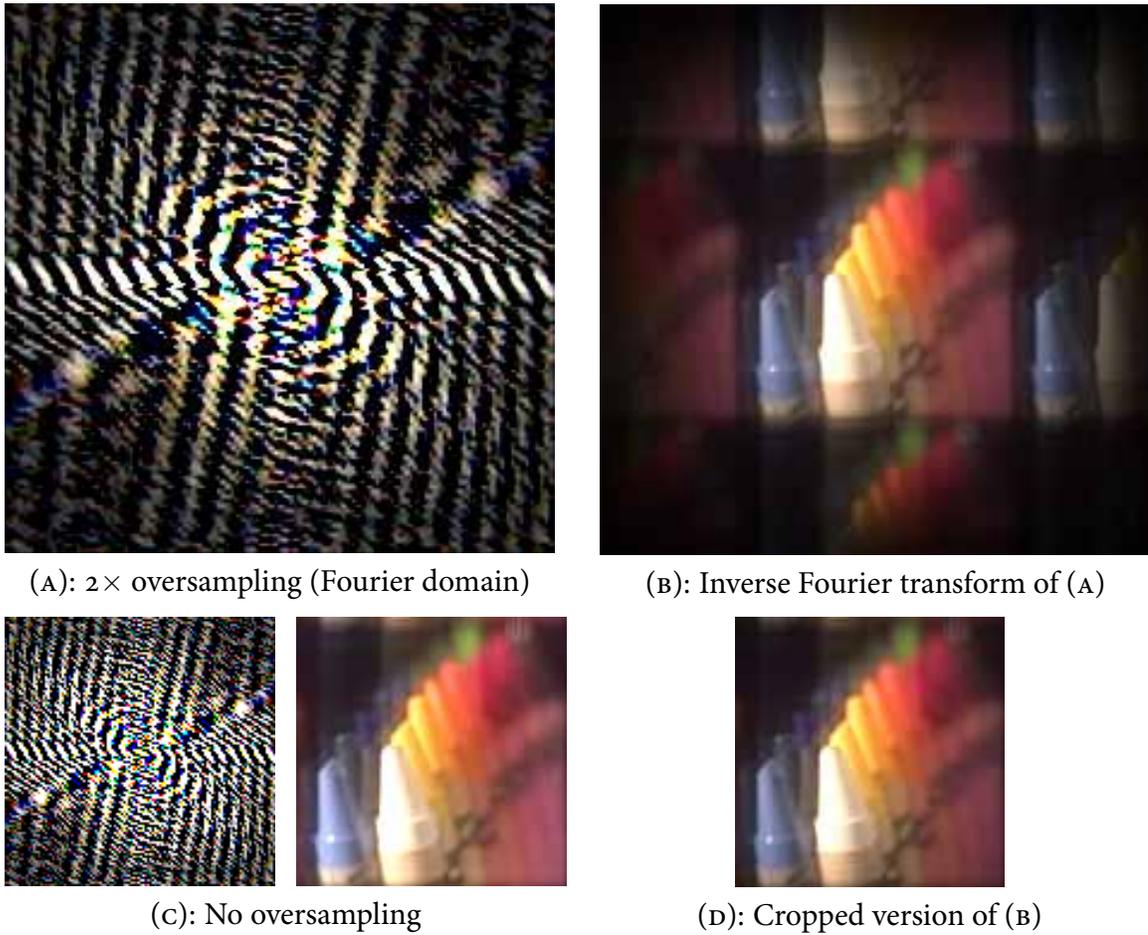
Figure 5.13: Reducing aliasing artifacts by oversampling in the Fourier domain.

Figure 5.13 illustrates this approach. Image A illustrates an extracted Fourier slice that has been oversampled by a factor of 2. Image B illustrates the resulting image with twice the normal field of view. Some of the aliased replicas fall into the outer portions of the field of view, which are cropped away in Image D. For comparison, the image pair in C illustrates the results with no oversampling. The image with oversampling contains less aliasing in the desired field of view.

Oversampling is appealing because of its simplicity, but oversampling alone cannot produce good quality images. The problem is that it cannot eliminate the replicas that appear perpendicular to the slicing plane, which are projected down onto the final image as described in the previous section.

This brings us to the second major technique of combating aliasing: superior filtering. As already stated, the ideal filter is a sinc function with a band-limit matching the spatial bounds of the light field. Our goal is to use a finite-extent filter that approximates this perfect spectrum as closely as possible. The best methods for producing such filters use iterative techniques to jointly optimize the band-limit and narrow spatial support, as described in Jackson et al. [1991] in the medical imaging community, and Malzbender [1993] in the Fourier volume rendering community.

| (A): Quadrilinear filter | (B): Kaiser-Bessel filter | (C): Kaiser-Bessel filter |
| (width 2) | (width 1.5) | (width 2.5) |
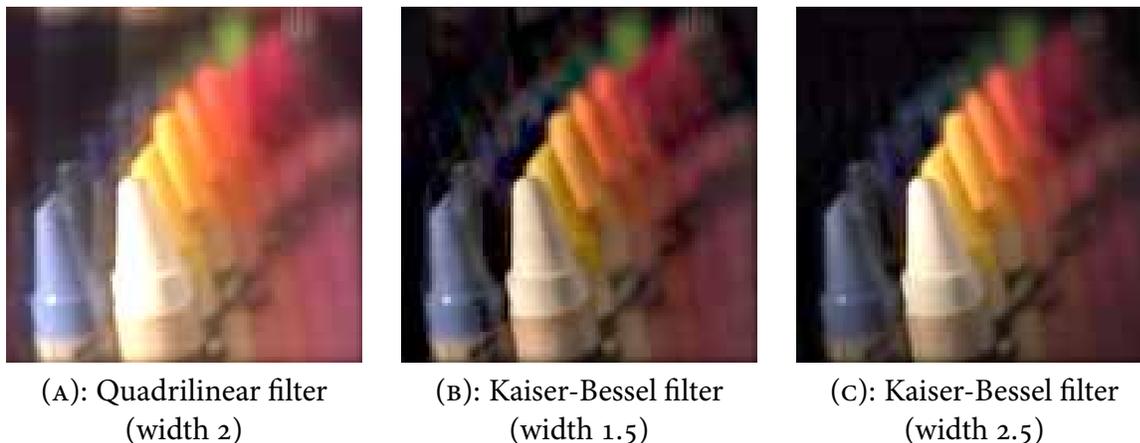
Figure 5.14: Aliasing reduction by superior filtering. Rolloff correction is applied.

Jackson et al. show that a much simpler, and near-optimal, approximation is the Kaiser-Bessel function. They also provide optimal Kaiser-Bessel parameter values for minimizing aliasing. Figure 5.14 illustrates the striking reduction in aliasing provided by such optimized Kaiser-Bessel filters compared to inferior quadrilinear interpolation. Surprisingly, a Kaiser-Bessel window of just width 2.5 suffices for excellent results.

As an aside, it is possible to change the aperture of the synthetic camera and *bokeh* of the resulting images by modifying the resampling filter. Using a different aperture can be

(A): Without padding   (B): With padding
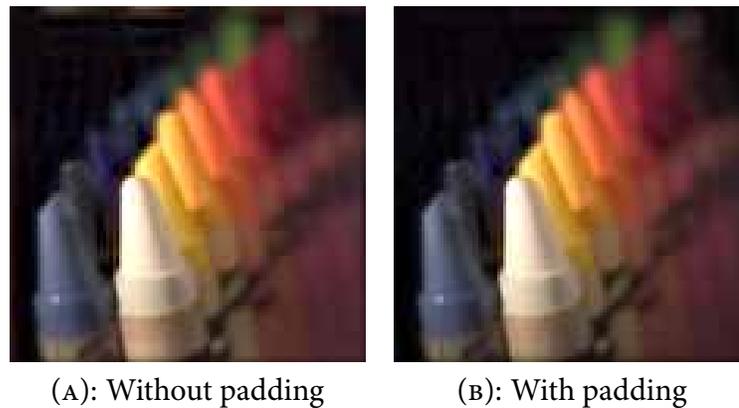
Figure 5.15: Aliasing reduction by padding with a border of zero values. Kaiser-Bessel re-sampling filter used.

thought of as multiplying $L_F(x, y, u, v)$ by an aperture function $A(u, v)$ before image forma-tion. For example, digitally stopping down would correspond to a mask $A(u, v)$ that is one for points (u, v) within the desired aperture and zero otherwise. In the Fourier domain, this multiplication corresponds to convolving the resampling filter by the Fourier spectrum of $A(u, v)$. This is related to work in shading in Fourier volume rendering [Levoy 1992].

The third and final method to combat aliasing is to pad the light field with a small bor-der of zero values before pre-multiplication and taking its Fourier transform [Levoy 1992; Malzbender 1993]. This pushes energy slightly further from the borders, and minimizes the amplification of aliasing energy by the pre-multiplication described in 5.5. In Figure 5.15, notice that the small amount of aliasing present near the top left border of Image A is elimi-nated in Image B with the use zero-padding.

Implementation Summary

Implementing the algorithm proceeds by directly discretizing the algorithm presented at the beginning of Section 5.5, applying the following four techniques to suppress artifacts. In the pre-processing phase,

1. Pad the light field with a small border (e.g. 5%) of zero values.
2. Pre-multiply the light field by the reciprocal of the Fourier transform of the resampling filter.

Fourier slice algorithm

Spatial integration
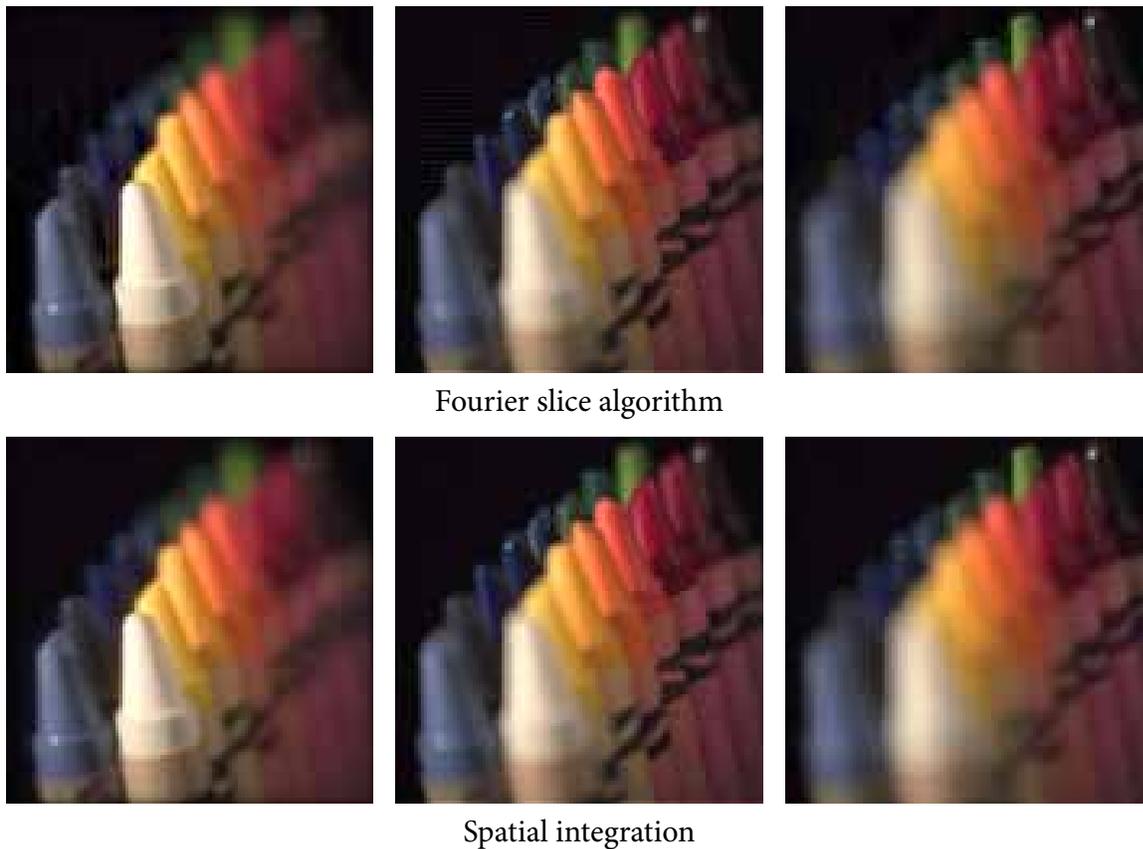
Figure 5.16: Comparison of refocusing in the Fourier and spatial domains.

In the refocusing step, which involves extracting the 2D Fourier slice,

3. Use a linearly-separable Kaiser-Bessel resampling filter. A filter width of 2.5 produces excellent results. For fast previewing, an extremely narrow filter of width 1.5 produces results that are superior to (and faster than) quadrilinear interpolation.

4. Oversample the 2D Fourier slice by a factor of 2. After Fourier inversion, crop the resulting photograph to isolate the central quadrant.

Performance Summary

This section compares the image-quality and efficiency of Fourier Slice algorithm for digital refocusing against the spatial-domain methods described in Chapter 4.

Dealing first with image quality, Figures 5.16 and 5.17 compare images produced with

Fourier slice algorithm

Spatial integration
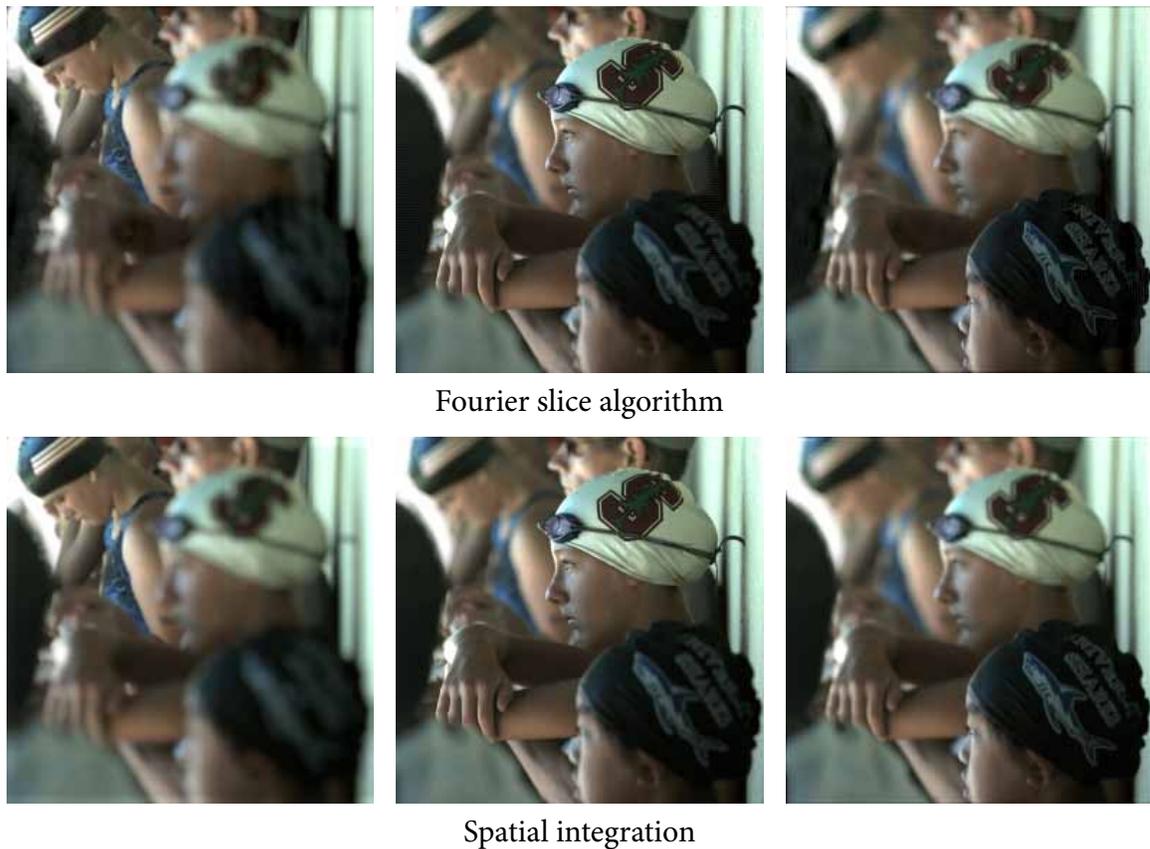
Figure 5.17: Comparison of refocusing in the Fourier and spatial domains II.

the Fourier domain algorithm and spatial integration. The images in the middle columns are the ones that correspond to no refocusing ($\alpha = 1$). Figure 5.17 illustrates a case that is particularly difficult for the Fourier domain algorithm, because it has bright border regions and areas of over-exposure that are 4 times as bright as the correct final exposure. Ringing and aliasing artifacts from these regions, which are relatively low energy compared to the source regions, are more easily seen when they overlap with dark regions.

Nevertheless, the Fourier-domain algorithm performs well even with the filter kernel of width 2.5. Although the images produced by the two methods are not identical, the comparison shows that the Fourier-domain artifacts can be controlled with reasonable cost.

In terms of computational efficiency, my CPU implementation of the Fourier and spatial algorithms run at the same rate with 16×16 directional resolution. With 32×32 directional resolution, the Fourier algorithm is an order of magnitude faster [Ng 2005]. The Fourier

Slice method outperforms the spatial methods as the directional *uv* resolution increases, because the number of light field samples that must be summed increases for the spatial integration methods, but the cost of slicing in the Fourier domain stays constant per pixel.

5.6   Light Field Tomography

The previous section described an algorithm that computes refocused photographs by extracting slices of the light field's Fourier spectrum. This raises the intriguing theoretical question as to whether it is possible to invert the process and reconstruct a light field from sets of photographs focused at different depths. This kind of *light field tomography* would be analogous to the way CT scanning reconstructs density volumes of the body from x-ray projections.

Let us first consider the reconstruction of 2D light fields from 1D photographs. We have seen that, in the Fourier domain of the 2D ray-space, a photograph corresponds to a line passing through the origin at an angle that depends on the focal depth, α. As α varies over all possible values, the angle of the slice varies from $-\pi/2$ to $\pi/2$. Hence, the set of 1D photographs refocused at all depths is equivalent to the 2D light field, since it provides a complete sampling of its Fourier transform. Another way of saying this is that the set of all 1D refocused photographs is the Radon transform [Deans 1983] of the light field. The 2D Fourier Slice Theorem is a classical way of inverting the 2D Radon transform to obtain the original distribution.

An interesting caveat to this analysis is that it is not physically clear how to acquire the photographs corresponding to negative α, which are required to acquire all radial lines in the Fourier transform. If we collect only those photographs corresponding to positive α, then it turns out that we omit a full fourth of the 2D spectrum of the light field.

In any case, this kind of thinking unfortunately does not generalize to the full 4D light field. It is a direct consequence of the Fourier Slice Photograph Theorem (consider Equation 5.6 for all α) that the footprint of all full-aperture 2D photographs lies on the following 3D manifold in the 4D Fourier space:

$$\left\{ (\alpha \cdot k_x, \; \alpha \cdot k_y, \; (1-\alpha) \cdot k_x, \; (1-\alpha) \cdot k_y) \text{where } \alpha \in [0, \infty), \text{ and } k_x, k_y \in \mathbb{R} \right\}. \qquad (5.14)$$

In other words, a set of conventional 2D photographs focused at all depths is not equivalent to the 4D light field. It formally provides only a small subset of its 4D Fourier transform.

One attempt to extend this footprint to cover the entire 4D space is to use a slit aperture in front of the camera lens. Doing so essentially masks out all but a 3D subset of the light field inside the camera. One can tomographically reconstruct this 3D *slit light field* from 2D slit aperture photographs focused at different depths. This works in much the same way that one can build up the 2D light field from 1D photographs (although the same caveat about negative α holds). By moving the slit aperture over all positions on the lens, this approach builds up the full 4D light field by tomographically reconstructing each of its constituent 3D slit light fields.

6

Selectable Refocusing Power

The previous two chapters concentrated on an in-depth analysis of the plenoptic camera where the microlenses are focused on the main lens, since that is the case providing maximal directional resolution and differs most from a conventional camera. The major drawback of the plenoptic camera is that capturing a certain amount of directional resolution requires a proportional reduction in the spatial resolution of final photographs. Chapter 3 introduced a surprisingly simple way to dynamically vary this trade-off, by simply reducing the separation between the microlenses and photosensor. This chapter studies the theory and experimental performance of this *generalized light field camera*.

Of course the more obvious way to vary the trade-off in space and direction is to exchange the microlens array for one with the desired spatial resolution. The disadvantage of this approach is that replacing the microlens array is typically not very practical in an integrated device like a camera, especially given the precision alignment required with the photosensor array. However, thinking about a family of plenoptic cameras, each customized with a different resolution microlens array, provides a well-understood baseline for comparing the performance of the generalized light field camera. Figure 6.1 illustrates such a family of customized plenoptic cameras. Note that the resolution of the photosensor is only 32 pixels in these diagrams, and the microlens resolutions are similarly low for illustrative purposes.

The ray-trace diagrams in Figure 6.1 illustrate how the set of rays that is captured by one photosensor pixel changes from a narrow beam inside the camera into a broad cone as the

(A): 4 microlenses   (B): 8 microlenses   (C): 16 microlenses   (D): 32 microlenses
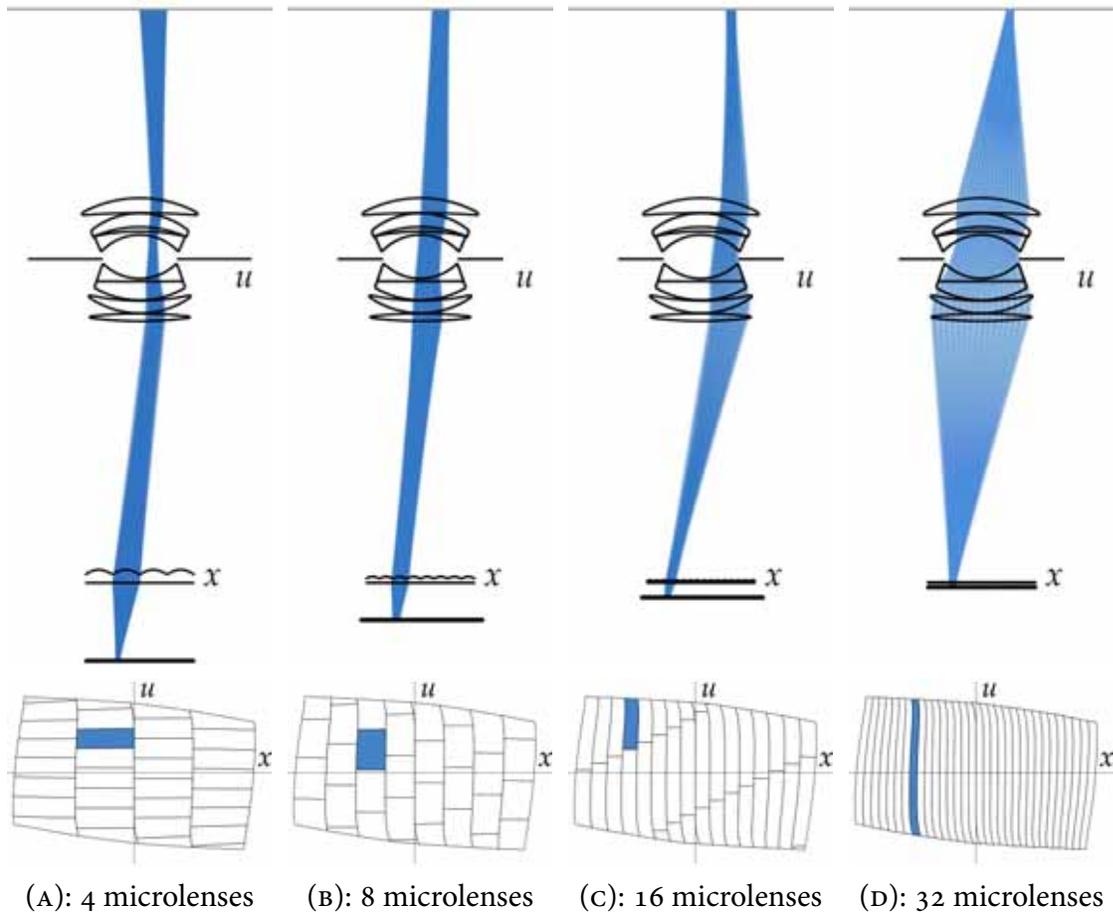
Figure 6.1: Plenoptic cameras with custom microlens arrays of different resolutions.

microlens resolution increases. The ray-space diagrams show how the ray-space cell for this set of rays transforms from a wide and short rectangle into a tall skinny one – this is the ray-space signature of trading directional resolution for spatial resolution.

6.1   Sampling Pattern of the Generalized Light Field Camera

Each separation of the microlens array and photosensor is a different configuration of the generalized light field camera. Let us define β to be the separation as a fraction of the depth that causes the microlenses to be focused on the main lens. For example, the typical plenoptic camera configuration corresponds to $\beta = 1$, and the configuration where the microlenses

(A): $\beta = 1$ (B): $\beta = 0.75$ (C): $\beta = 0.5$ (D): $\beta = 0$
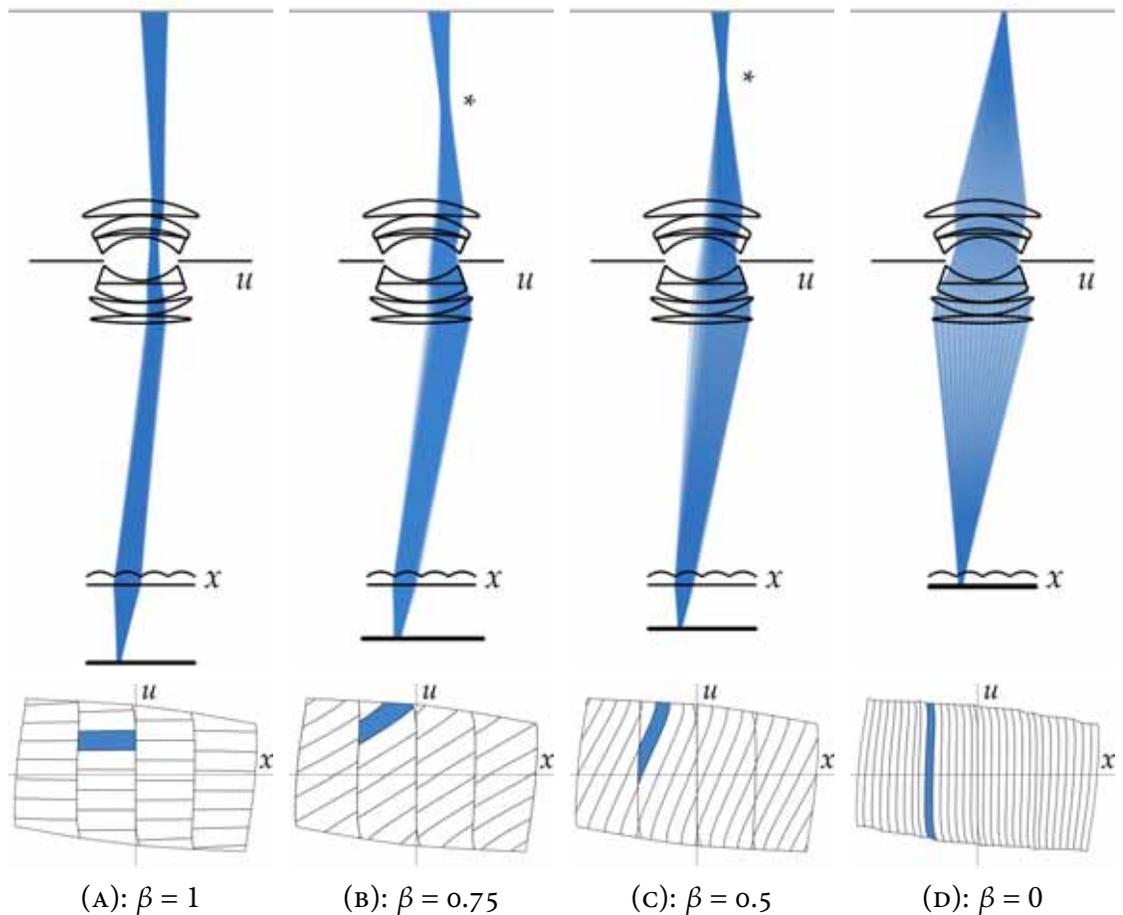
Figure 6.2: Different configurations of a single generalized light field camera.

are pressed up against the photosensor is $\beta = 0$. As introduced in Section 3.5, decreasing the β value defocuses the microlenses by focusing them *beyond* the aperture of the main lens.

Figure 6.2 illustrates four β-configurations. The configurations were chosen so that the effective spatial resolution matched the corresponding plenoptic camera in Figure 6.1. Note that the changes in beam shape between Figures 6.1 and 6.2 are very similar at a macroscopic level, although there are important differences in the ray-space. As the highlighted blue cells in Figure 6.2 show, reducing the β value results in a shearing of the light field sampling pattern within each column. The result is an increase in effective spatial resolution (reduction in x extent), and a decrease in directional resolution (increase in u extent). An important drawback of the generalized camera's ray-space sampling pattern is that it is more

anisotropic, which causes a moderate loss in effective directional resolution. However, experiments at the end of the chapter suggest that the loss is not more than a factor of 2 in effective directional resolution.

Recovering higher resolution in output images is possible as the β value decreases, and works well in practice. However, it requires changes in optical focus of the main lens and final image processing as discussed below. At $\beta = 0$ the microlenses are pressed up against the sensor, lose all their optical power and the effective spatial resolution is that of the sensor. The effective resolution decreases linearly to zero as the β value increases, with the resolution of the microlens array setting a lower bound. In equations, if the resolution of the sensor is $M_{\text{sensor}} \times M_{\text{sensor}}$ and the resolution of the microlens array is $M_{\text{lenslets}} \times M_{\text{lenslets}}$, the output images have effective resolution $M_{\text{effective}} \times M_{\text{effective}}$, where

$$M_{\text{effective}} = \max((1 - \beta)M_{\text{sensor}},\ M_{\text{lenslets}}). \tag{6.1}$$

Deriving the Sampling Pattern

Figure 6.2 was computed by ray-tracing through a virtual model of the main lens, procedurally generating the ray-space boundary of each photosensor pixel. This section provides additional insight in the form of a mathematical derivation of the observed sampling pattern.

An important observation in looking at Figure 6.2 is that the changes in the sampling pattern of the generalized light field camera are localized within the columns defined by the microlens array. Each column represents the microscopic ray-space between one microlens and the patch of photosensors that it covers. Defocusing the microlens by reducing the β value shears the microscopic ray-space, operating under the same principles discussed in Chapter 2 for changing the focus of the main lens. The derivation below works from the microscopic ray-space, where the sampling pattern is trivial, moving out into the ray-space of the full camera.

Figure 6.3A is a schematic for the light field camera, showing a microlens, labeled i, whose microscopic light field is to be analyzed further. Figure 6.3B illustrates a close-up of the microlens, with its own local coordinate system. Let us parameterize the microscopic light field's ray-space by intersection of rays with the three illustrated planes: the microlens plane, x^i, the sensor plane, s^i, and the focal plane of the microlens, w^i. In order to map

(A): Light field camera   (B): Zoom of microlens i (C)   (D)

(E)   (F): Macroscopic ray-space
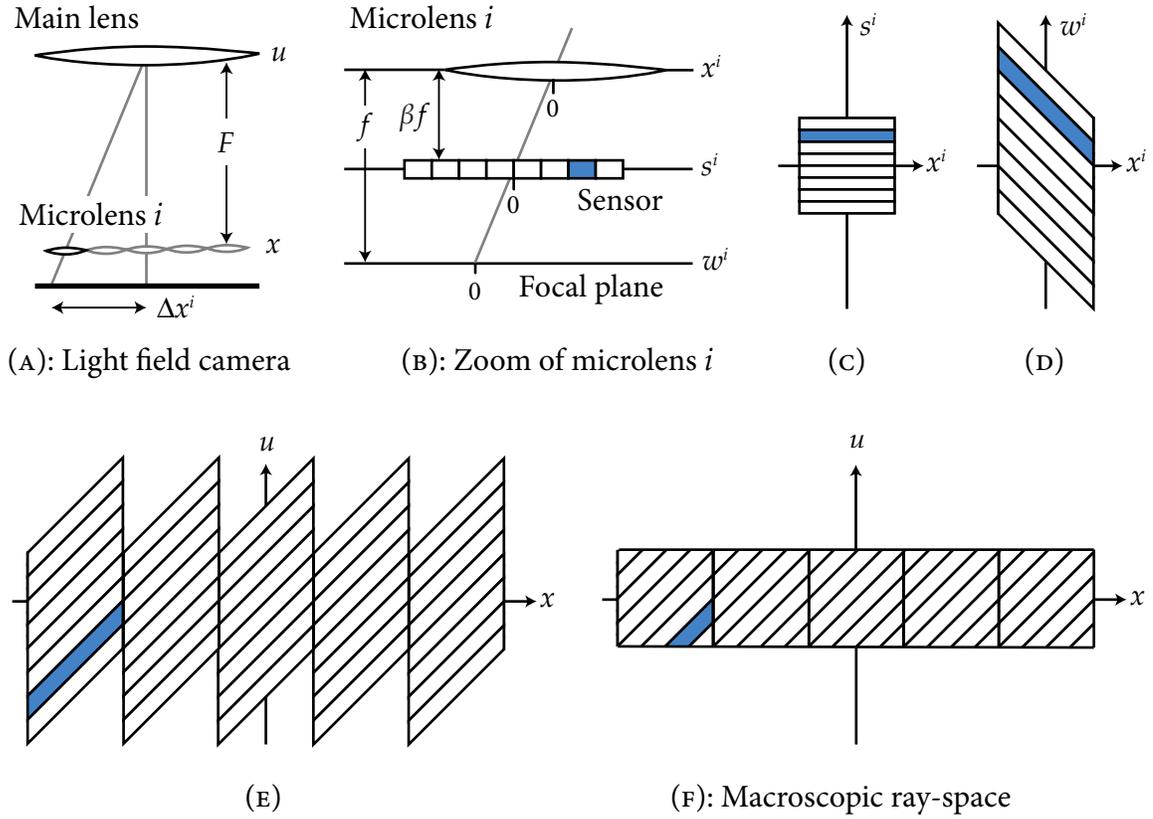
Figure 6.3: Derivation of the sampling pattern for the generalized light field camera.

this microscopic ray-space neatly into a column of the macroscopic ray-space for the whole camera, it is convenient to choose the origins of the three planes to lie along the line passing through the center of the main lens and the center of microlens i, as indicated on Figure 6.3B. Also note on the figure that the focal length of the microlenses is f, and the separation between the microlenses and the sensor is βf.

Figure 6.3C illustrates the shape of the sampling pattern on the ray-space parameterized by the microlens x^i plane and the photosensor s^i plane. This choice of ray-space parameterization makes it is easy to see that the sampling is given by a rectilinear grid, since each photosensor pixel integrates all the rays passing through its extent on s^i, and the entire surface of the microlens on x^i. Let us denote the microscopic light field as $l^i_{\beta f}(x^i, w^i)$, where the subscript βf refers to the separation between the parameterization planes. The eventual

goal is to transform this sampling pattern in this ray-space into the macroscopic ray-space, by a change in coordinate systems.

Figure 6.3D illustrates the first transformation: re-parameterizing $l^i_{\beta f}$ by changing the lower parameterization plane from the sensor plane s^i to the microlens focal plane w^i. Let us denote the microscopic light field parameterized by x^i and w^i as $l^i_f(x^i, w^i)$, where the f subscript reflects the increased separation of one microlens focal length. Re-parameterizing into this space is the transformation that introduces the shear in the light field. It is directly analogous to the transformation illustrated in Figure 2.5. Following the derivation described there it is easy to show that

$$l^i_{\beta f}(x^i, s^i) = l^i_f \left(x^i, x^i \left(1 - \frac{1}{\beta} \right) + \frac{s^i}{\beta} \right). \tag{6.2}$$

The transformation from the microscopic light fields under each microlens into the macroscopic ray-space of the camera is very simple. It consists of two steps. First, there is a horizontal shift of Δx^i, as shown on Figure 6.3A, to align their origins. The second step is an inversion and scaling in the vertical axis. Since the focal plane of the microlens is optically focused on the main lens, every ray that passes through a given point on l^i_f passes through the same, conjugate point on the u. The location of this conjugate point is opposite in sign due to optical inversion (the image of the main lens appears upside down under the microlens). It is also scaled by a factor of $\frac{F}{f}$ because of optical magnification. Combining these transformation steps,

$$l^i_f(x^i, w^i) = L \left(\Delta x^i + x^i, \ -\frac{F}{f} w^i \right). \tag{6.3}$$

Combining Equations 6.2 and 6.3 gives the complete transformation from $l^i_{\beta f}$ to the macroscopic space:

$$l^i_{\beta f}(x^i, s^i) = L \left(\Delta x^i + x^i, \ x^i \frac{F}{f} \left(\frac{1}{\beta} - 1 \right) - \frac{F}{f} \frac{s^i}{\beta} \right). \tag{6.4}$$

In particular, note that this equation shows that the slope of the grid cells in Figure 6.3E is

$$\frac{F}{f} \left(\frac{1}{\beta} - 1 \right), \tag{6.5}$$

a fact that will be important later.

Note that the preceding arguments hold for the microscopic light fields under each microlens. Figure 6.3E illustrates the transformation of all these microscopic light fields into the macroscopic ray space, showing how they pack together to populate the entire space. As the final step in deriving the macroscopic sampling pattern, Figure 6.3F illustrates that the main lens truncates the sampling pattern vertically to fall within the range of u values passed by the lens aperture.

As a final note, the preceding analysis assumes that the main lens is ideal, and that the f-numbers of the system are matched to prevent cross-talk between microlenses. The main difference between the idealized pattern derived in Figure 6.3F and the patterns procedurally generated in Figure 6.2 is a slight curvature in the grid lines. These are real effects due to aberrations in the main lens, which are the subject of the next chapter.

The Problem with Focusing the Microlenses Too Close

The previous section examines what happens when we allow the microlenses to defocus by focusing beyond the main lens ($\beta < 1$). An important question is whether there is a benefit to focusing *closer* than the main lens, corresponding to moving the photosensor plane *further* than one focal length ($\beta > 1$). The difficulty with this approach is that the the microlens images would grow in size and overlap. The effect could be balanced to an extent by reducing the size of the main lens aperture, but this cannot be carried very far. By the time the separation increases to two (microlens) focal lengths, the micro-images will have defocused to be as wide as the microlenses themselves, even if the main lens is stopped down to a pin-hole. For these reasons, this chapter concentrates on separations less than or equal to one focal length.

6.2 Optimal Focusing of the Photographic Lens

An unusual characteristic of the generalized camera is that we must focus its main lens differently than in a conventional or plenoptic camera. In the conventional and plenoptic cameras best results are obtained by optically focusing on the subject of interest. In contrast, for intermediate β values the highest final image resolution is obtained if we optically focus slightly

beyond the subject, and use digital refocusing to pull the virtual focal plane back onto the subject of interest.

Figure 6.2B and C illustrate this phenomenon quite clearly. Careful examination reveals that maximal spatial resolution of computed photographs would be achieved if we digitally refocus slightly closer than the world focal plane, which is indicated by the gray line on the ray-trace diagram. The refocus plane of greatest resolution corresponds to the plane passing through the point where the convergence of world rays is most concentrated (marked by asterisks). The purpose of optically focusing further than usual would be to shift the asterisks onto the gray focal plane. Notice that the asterisk is closer to the camera for higher β values.

The ray-space diagrams shown in Figure 6.2B and C provide additional insight. Recall that on the ray-space, refocusing closer means integrating along projection lines that tilt clockwise from vertical. Visual inspection makes it clear that we will be able to resolve the projection line integrals best when they align with the diagonals of the sampling pattern – that is, the slope indicated by the highlighted, diagonal blue cell on each ray-space. From the imaging equation for digital refocusing (Equation 4.1), the slope of the projection lines for refocusing is

$$\left(1 - \frac{1}{\alpha}\right). \tag{6.6}$$

Recall that $\alpha = F'/F$, where F' is the virtual film depth for refocusing, and F is the depth of the microlens plane. The slope of the ray-space cells in the generalized light field sampling pattern was calculated above, in Equation 6.5. Equating these two slopes and solving for the relative refocusing depth, α,

$$\alpha = \frac{(1 - \beta)F}{(1 - \beta)F - \beta f}. \tag{6.7}$$

This tells us the relative depth at which to refocus to produce maximum image resolution. If we wish to eventually produce maximum resolution at depth F, we should therefore optically focus the main lens by positioning it at depth

$$F_{\text{opt}} = \frac{F}{\alpha} = \frac{(1 - \beta)F - \beta f}{(1 - \beta)} = F + \frac{\beta}{(\beta - 1)} f. \tag{6.8}$$

For the range of β values that we are looking at ($0 < \beta \leq 1$), $F_{\text{opt}} < F$, indicating that the microlens plane should be brought closer to the main lens. This means optically focusing

further in the world, which is consistent with shifting the asterisks in the ray-trace view in Figure 6.2 onto the desired world plane. The *optical mis-focus* is the difference between F_{opt} and F, given by $\frac{\beta}{(\beta-1)} f$.

Note that as β approaches 1, the optical mis-focus asymptotes to negative infinity, which is meaningless. The reason for this is that the slope of the sampling grid cells become too horizontal, and the optimal resolution is dominated by the vertical columns of the the sampling pattern set by the resolution of the microlens array, not by the slope of the cells within each column. When this occurs, it is best to set the optical focal depth at the desired focal depth (i.e. $F_{opt} = F$), to provide the greatest latitude in refocusing about that center. In practice it makes sense to stop using Equation 6.8 once the effective resolution for that β configuration falls to less than twice the resolution of the microlens array.

Figure 6.4: Predicted effective resolution and optical mis-focus as a function of β for the prototype camera.

This cutoff is shown as a dotted vertical line on Figure 6.4. The graphs in this figure plot effective resolution and optical mis-focus for the prototype camera described in Chapter 3, where $M_{sensor} = 4096$, $M_{lenslets} = 296$, and $f = 0.5$ mm.

Recall that the predicted effective resolution $M_{effective} \times M_{effective}$ of the output images is

$$M_{effective} = \max((1 - \beta)M_{sensor}, \; M_{lenslets}). \tag{6.9}$$

As with F_{opt}, the predicted value for $M_{effective}$ derives from an analysis of the ray-space sampling pattern. Refocusing optimally aligns the imaging projection lines with the slope of the grid cells, enabling extraction of the higher spatial resolution inherent in the sheared sampling pattern. By visual inspection, the effective resolution of the computed image is equal to the number of grid cells that intersect the x axis. Within each microscopic light
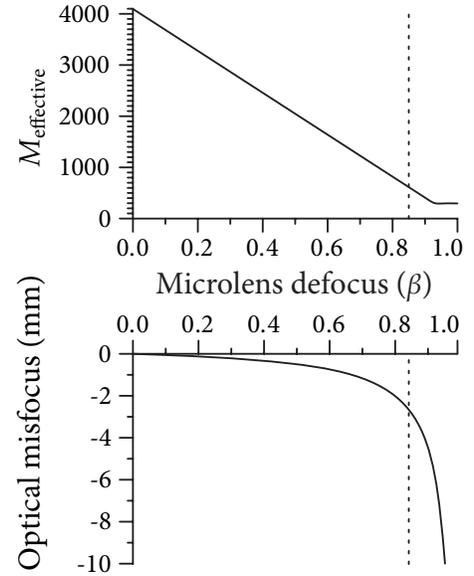
field, Equation 6.2 shows that the number of grid cells crossed is proportional to $(1 - \beta)$, because of the shearing of the microscopic light field sampling patterns. Hence the overall resolution is proportional to $(1 - \beta)$. The maximum possible resolution is the resolution of the sensor, and the minimum is the resolution of the microlens array. Experiments below test this predicted linear variation in effective resolution with $(1 - \beta)$.

In summary, when recording light fields with intermediate β, if the auto-focus sensor indicates that the subject is at depth F, the sensor should be positioned at F_{opt}. Digital refocusing onto depth F after exposure will then produce the image with the maximum possible effective resolution of $\max((1 - \beta)M_{\text{sensor}}, M_{\text{lenslets}})$.

6.3   Experiments with Prototype Camera

The prototype camera allowed photographic testing of performance for $\beta \geq 0.6$, by manually adjusting the screws on the microlens array to choose different separations between the microlens and photosensor. It was not possible to screw down to smaller β values because the separation springs bottomed out. The overall set-up of our scene was similar to that in Section 4.5: I shot light fields of a resolution chart at varying levels of main lens mis-focus, and tested the ability to digitally refocus to recover detail. In order to examine exactly how much the effective spatial resolution changes with β, I computed final photographs with 4096×4096 pixels to match the full resolution of the photosensor.

Figure 6.5 illustrates the major trade-off that occurs when decreasing β: maximum spatial resolution increases (potentially allowing sharper final images), and directional resolution decreases (reducing refocusing power). The images show extreme close-up views of the center 1/100th of final images computed from recorded light fields of the ISO 12233 image resolution chart. World focus was held fixed at a depth of approximately 1.3 meters from the camera. Optical mis-focus refers to the distance that the target was moved closer to the camera than this optical focal depth.

Notice that at $\beta = 1.0$ the resolution is never enough to resolve the finer half of rings on the chart, but it is able to resolve the coarser rings over a wide range of main lens mis-focus out to at least 17 cm. In contrast, at $\beta = 0.6$, it is possible to resolve all the rings on the chart at a mis-focus of 5 cm. However, notice that resolution is much more sensitive to mis-focus,
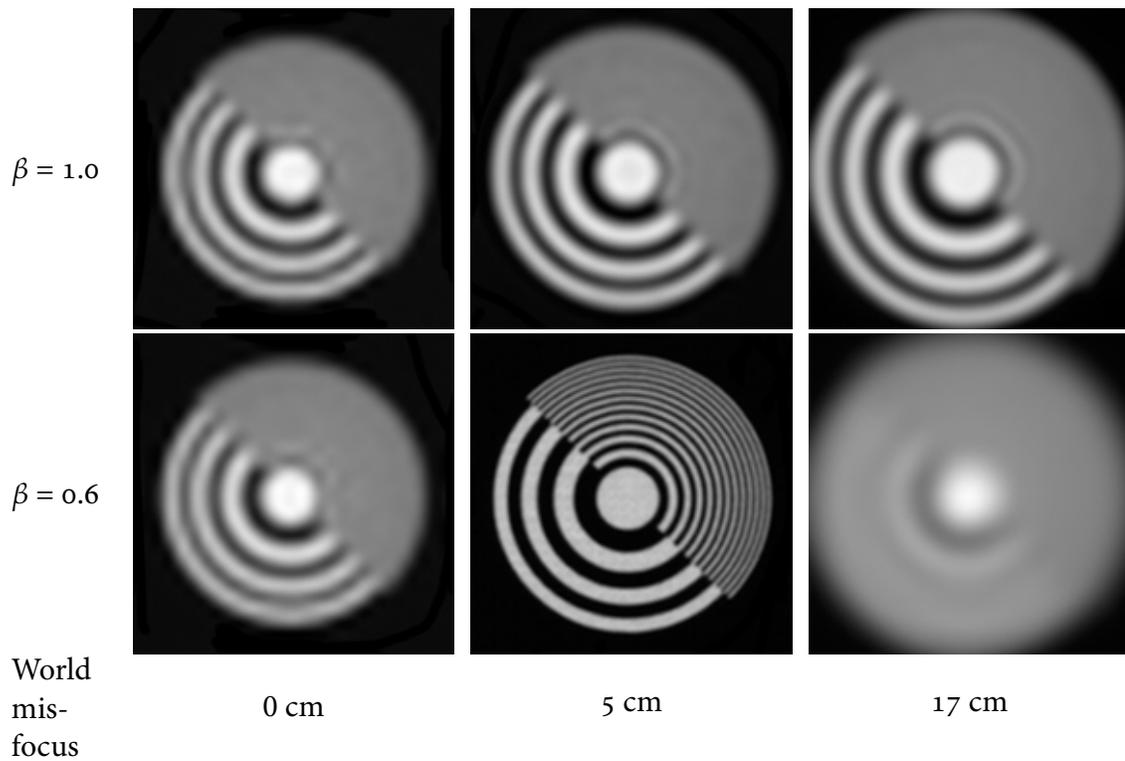
Figure 6.5: Decreasing β trades refocusing power for maximum image resolution.

blurring completely by 17 cm. Indeed, the images for optical mis-focus of 2.5 cm and 7.5 cm (not shown) were noticeably more blurry than for 5 cm.

The fact that the highest image resolution is achieved when digitally refocusing 5 cm closer is numerically consistent with theory. The lens of the camera has a focal length of 140 mm, so the world focal depth of 1.3 meters implies $F \approx 156.9$ mm. Equation 6.7 therefore implies an optimal refocus separation of 157.7 mm, with a corresponding world refocus plane at 1.25 meters, which is indeed a mis-focus of 5 cm closer than the optical world focal plane, as observed on Figure 6.5.

As an aside, the slight ringing artifacts visible in the images with 0 optical mis-focus are due to pre-aliasing [Mitchell and Netravali 1998] in the raw light field. The high frequencies in the resolution chart exceed the spatial resolution of these β configurations, and the square microlenses and square pixels do not provide an optimal low-pass filter. Note that digital refocusing has the desirable effect of anti-aliasing the output images by combining samples

| Camera | Ray-Tracer | Camera | Ray-Tracer |
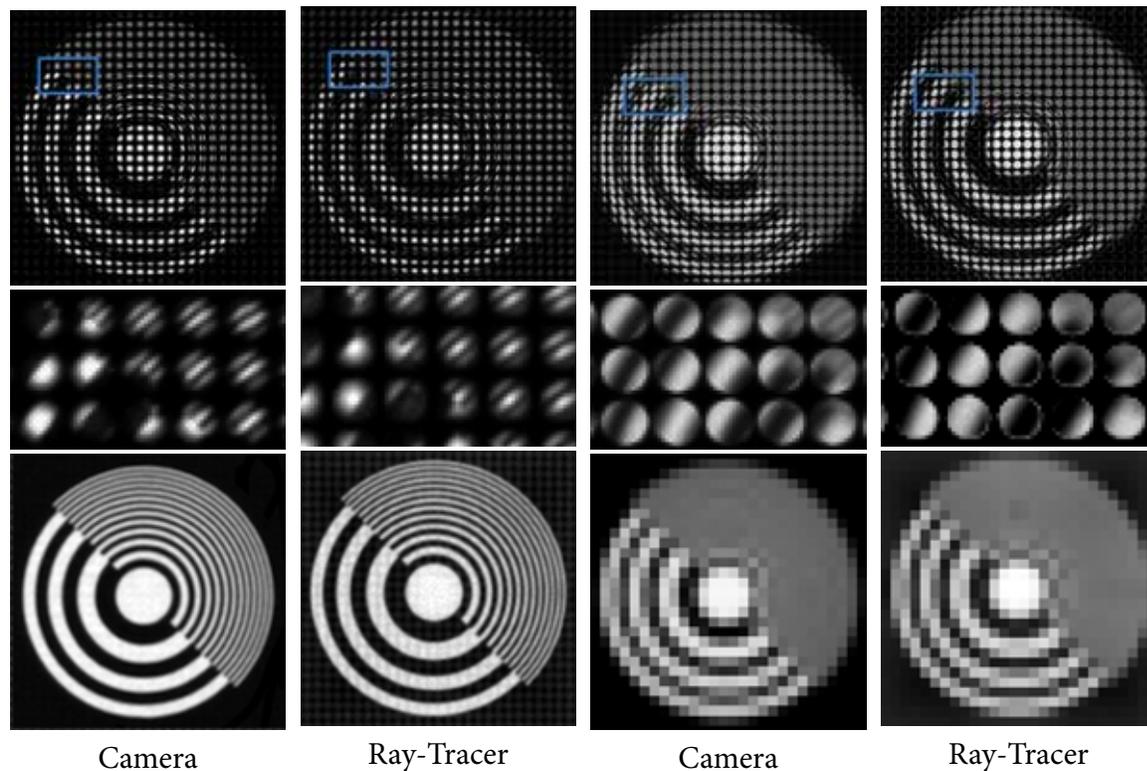
Figure 6.6: Comparison of data acquired from the prototype camera and simulated with a ray-tracer, for $\beta = 0.6$ (left half), and $\beta = 1$.

from multiple spatial locations, so the images with non-zero mis-focus actually produce higher-quality images.

6.4   Experiments with Ray-Trace Simulator

To explore the performance of the generalized camera for low β configurations, we enhanced Pharr and Humphreys' physically-based rendering system [2004] to compute the irradiance distribution that would appear on the photosensor in our prototype. Figure 6.6 illustrates the high fidelity that can be achieved with such a modern ray-tracer. The images presented in this figure correspond to the set-up for the middle row of Figure 6.5, that is, world mis-focus of 5 cm. The left half of Figure 6.6 is for $\beta = 0.6$, the right half for $\beta = 1.0$. The top two rows illustrate the raw light field data. The bottom row shows final images refocused onto

$\beta = 0.4$ $\beta = 0.2$ $\beta = 0.01$
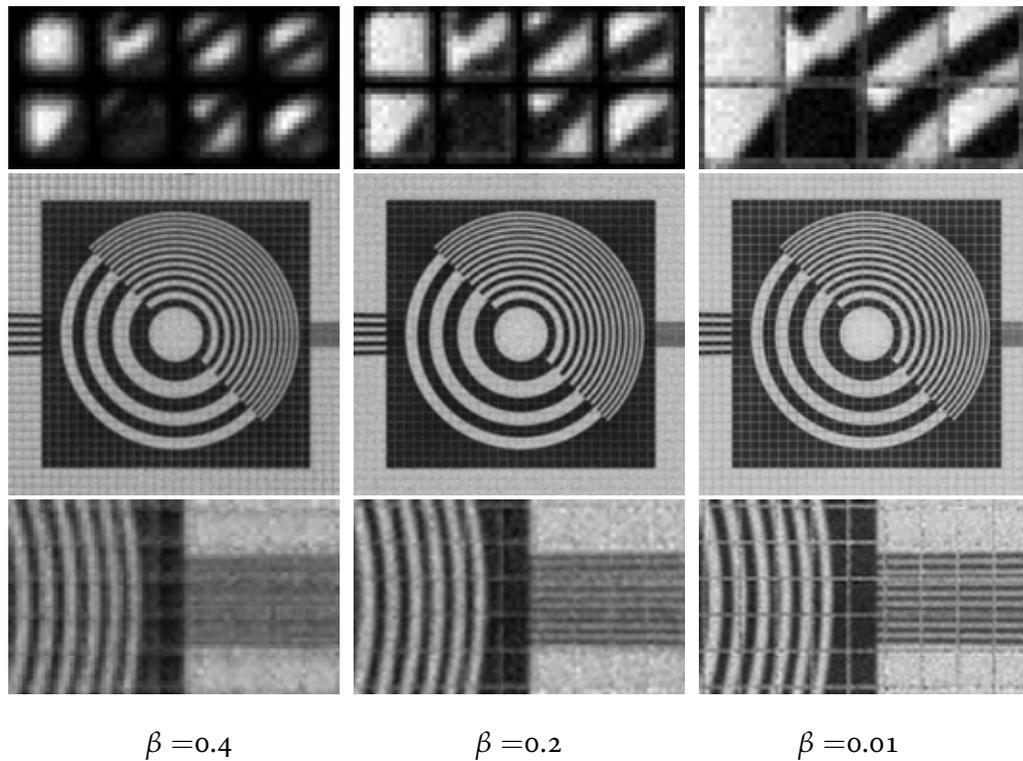
Figure 6.7: Simulation of extreme microlens defocus.

the resolution chart. These images reveal very good agreement between the simulation and our physically-acquired data, not only in final computed photographs but in the raw data itself.

Figure 6.7 presents purely simulated data, extrapolating performance for lower β values of 0.4, 0.2 and close to 0, which could not be physically acquired with the prototype. Each of the light fields was simulated and resampled assuming that the optimal main lens focus for each β was achieved according to Equation 6.8. In other words, the optical focus for $\beta = 0.4$ and 0.2 were slightly further than the target. The top row of images are zoomed views of a 4×2 section of microlens-images, illustrating how decreasing β causes these micro-images to evolve from blurred images of the circular aperture to filled squares containing the irradiance striking each microlens square.

The bottom two rows of images illustrate how resolution continues to increase as β decreases to 0. The area shown in the extreme close-up in the bottom row contains the finest lines on the ISO 12233 chart to the right of the right border of the black box. These lines project onto the width of roughly 3 photosensor pixels per line-pair. As the right most column shows, as β converges on zero separation, final photographs are recorded at close to the full spatial resolution of the photosensor.

MTF Analysis

The ray-tracing system enabled synthetic MTF analysis, which provided further quantitative evidence for the theory predicted by ray-space analysis. The overall goal was to visualize the decay in refocusing power and increase in maximum image resolution as β decreases. Another goal was to compare the performance of each β configurations against a light field camera customized with a microlens array of equivalent spatial resolution. Based on earlier discussion, we would expect the customized light field camera to provide slightly more refocusing power due to its isotropic sampling grid.

The analysis consisted of computing the variation in MTF for a number of different β configurations of the prototype camera. For each configuration, a virtual point light source was held at a fixed depth, and the optical focus was varied about that depth to test the ability of the system to compensate for mis-focus. The resulting light fields were processed to compute a refocused photograph of the point light source that was as sharp as possible. The Fourier transform of the resulting photograph provided the MTF for the camera for that configuration and that level of mis-focus.

The following graphs summarize the performance of the MTF for each of these tests as a single number. The summary measure that was chosen is the spatial frequency at which the computed MTF first drops below 50%. Although this measure is a simplistic summary of the full MTF, it is still well correlated with the ability of the system to resolve fine details, rising as the refocused image of the point light source becomes sharper and its MTF increases.

Figure 6.8A illustrates the plots of the described measure for five β configurations of the prototype camera. The horizontal axis is the relative optical mis-focus on the image-side, such that a deviation of 0.02 means that the separation between the main lens and the microlens plane was 2% greater than the depth at which the point light source forms an ideal

image. Note that the vertical axis is plotted on a log scale.

Figure 6.8A illustrates three effects clearly. First, the maximum image resolution (the maximum height of each plot) decreases as β increases, as expected by theory. The β configurations that are plotted move half-way closer to $\beta = 1$ with each step ($\beta = 0$, 0.5, 0.75, 0.875 and 1), and the maximum spatial frequency halves as well, indicating roughly linear variation with $(1 - \beta)$ as predicted by theory. The second effect is a broadening of the peak with increasing β, indicating greater tolerance to optical mis-focus. The breadth of the peak is a measure of the range of depths for which the camera can produce sharp images. The third effect is clear evidence for the shift in optimal optical focal plane, as predicted by Equation 6.8. As β increases, the maximum of the plot migrates to the left (the microlens plane moves closer to the lens). This indicates optical focus further than the point light source, corroborating the discussion earlier in the chapter.

Figure 6.8B re-centers each graph about the optimal optical depth predicted by Equation 6.8, in order to ease comparison with the customized plenoptic cameras, which are presented in Figure 6.8C. The resolution of the microlenses in the customized cameras were chosen to match the maximum spatial resolution of the generalized camera configurations, as predicted by Equation 6.9. The legend in Figure 6.8C gives the width of these microlenses.

Figure 6.9 re-organizes the curves in Figure 6.8B and Figure 6.8C for more direct comparison. The curve for each β value of the generalized camera is plotted individually, on a graph with the closest curves from the customized plenoptic cameras. These graphs make three technical points. The first point is that the overlap of the two curves in Figure 6.9A for $\beta = 0.01$ corroborates the prediction that the generalized light field camera can be made to have the performance of a conventional camera with full spatial resolution.

The second point is corroboration for the predicted effective resolution in Equation 6.9. This is seen in the matching peaks of the two curves with the same color in Figures 6.9B, 6.9C and 6.9D. Recall that the dotted curves of the same color represents the performance of a customized plenoptic camera with a microlens array resolution given by Equation 6.9.

The third point is that the effective refocusing power of the generalized camera is less than the customized camera of equivalent spatial resolution, but that the reduction in power is quite moderate. In Figures 6.9B, 6.9C and 6.9D, the generalized light field camera is compared not only against the plenoptic camera with equivalent spatial resolution, but also the
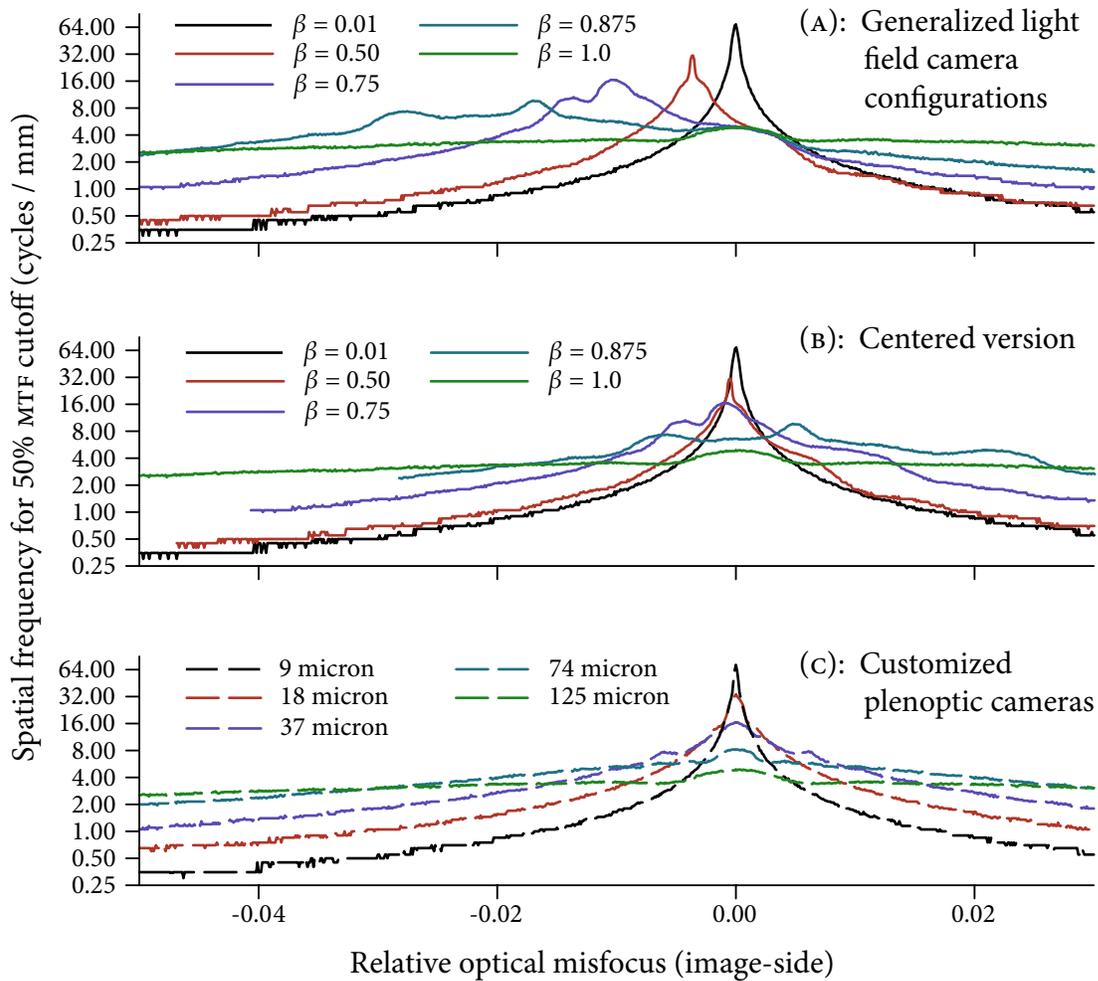
Figure 6.8: MTF comparison of trading refocusing power and image resolution.

one that has twice as much spatial resolution (hence half the directional resolution). In these graphs, the plot for the generalized camera lies *between* the two plenoptic cameras, bounding its effective refocusing power within a factor of 2 of the ideal performance given by a customized plenoptic camera. The loss disappears, of course, as β increase to 1 and the generalized camera converges on an ordinary plenoptic camera, as shown in Figure 6.9E.
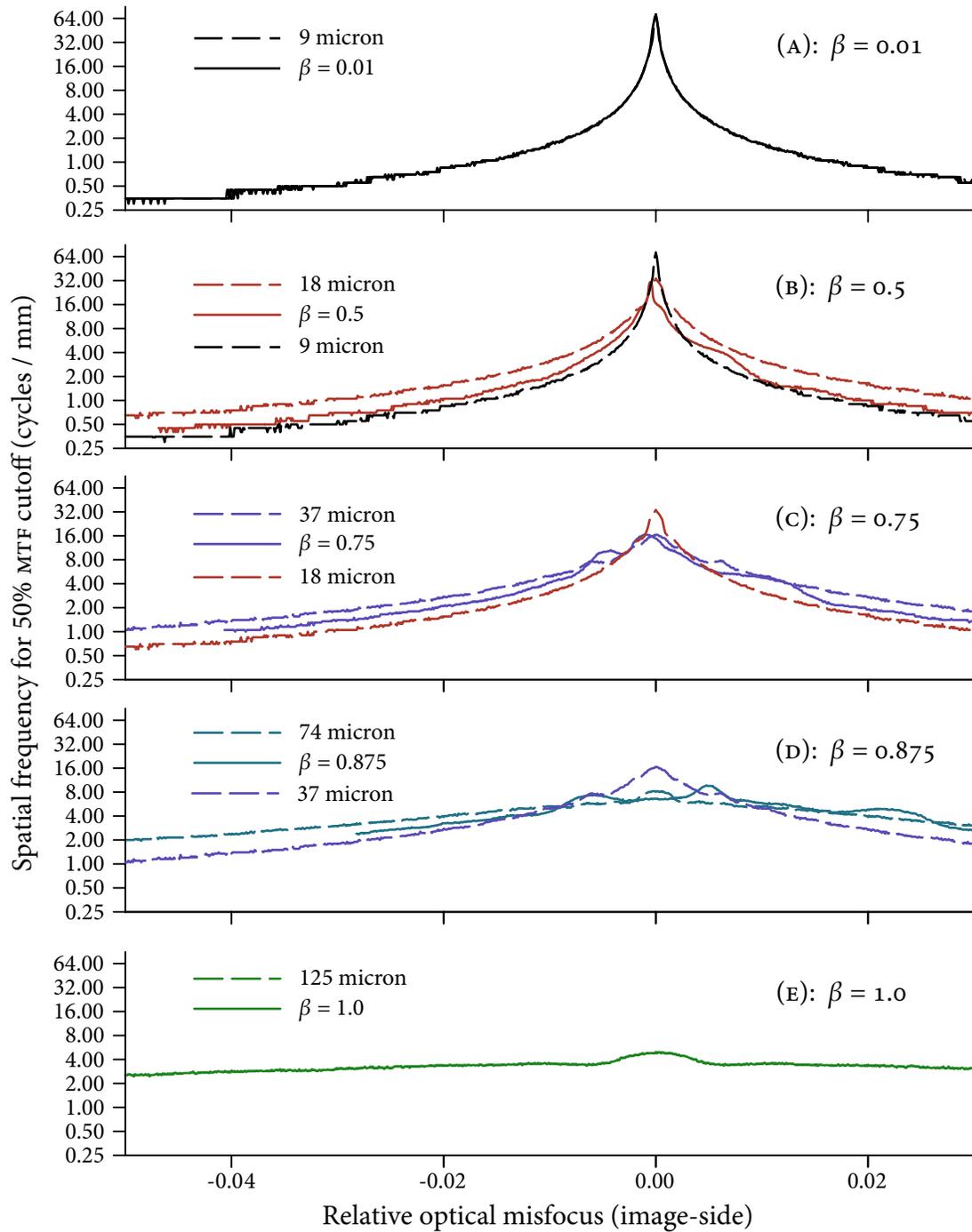
Figure 6.9: MTF comparison of trading refocusing power and image resolution II.

Summary

This chapter shows that defocusing the microlenses by moving the photosensor plane closer is a practical method for recovering the full spatial resolution of the underlying photosensor. This simple modification causes a dramatic change in the performance characteristics of the camera, from a low-resolution refocusing camera to a high-resolution camera with no refocusing. A significant practical result is that decreasing the separation between the microlenses and photosensor by only one half recovers all but 2×2 of the full photosensor resolution. This could be important in practice, as it means that it is not necessary for the photosensor to be pressed directly against the microlenses, which would be mechanically challenging.

For intermediate separations, the spatial resolution varies continuously between the resolution of the microlens array and that of the photosensor. A very nice property is that the refocusing power decreases roughly in proportion to the increase in spatial resolution. There is some loss in the effective directional resolution compared to the ideal performance of a plenoptic camera equipped with a custom microlens array of the appropriate spatial resolution, but simulations suggest that the loss in directional resolution is well contained within a factor of 2 of the ideal case.

These observations suggest a flexible model for a generalized light field camera that can be continuously varied between a conventional camera with high spatial resolution, and a plenoptic camera with more moderate spatial resolution but greater refocusing power. The requirement would be to motorize the mechanism separating the microlenses and the photosensor and provide a means to select the separation to best match the needs of the user for a particular exposure. For example, the user could choose high spatial resolution and put the camera on a tripod for a landscape photograph, and later choose maximal directional resolution to maximize the chance of accurately focusing an action shot in low light.

This approach greatly enhances the practicality of digital light field photography by eliminating one of its main drawbacks: that one must trade spatial resolution for refocusing power. By means of a microscopic adjustment in the configuration of the light field camera, the best of all worlds can be selected to serve the needs of the moment.

7

Digital Correction of Lens Aberrations

A lens creates ideal images when it causes all the rays that originate from a point in the world to converge to a point inside the camera. *Aberrations* are imperfections in the optical formula of a lens that prevent perfect convergence. Figure 7.1 illustrates the classical case of spherical aberration of rays refracting through a plano-convex lens, which has one flat side and one convex spherical side. Rays passing through the periphery of the spherical interface refract too strongly, converging at a depth closer to the lens than rays that pass close to the center of the lens. As a result, the light from the desired point is blurred over a spot on the image plane, reducing contrast and resolution.

Maxwell established how fundamental a problem aberrations are in the 1850s. He proved that no optical system can produce ideal imaging at all focal depths, because such a system would necessarily violate the basic mechanisms of reflection and refraction [Maxwell 1858]. Nevertheless, the importance of image quality has motivated intense study and optimization over the last 400 years, including contributions from such names as Gauss, Galileo, Kepler, Newton, and innumerable others. A nice introduction to the history of aberration theory is presented in a short paper by Johnson [1992], and Kingslake's classic book [1989] presents greater detail in the context of photographic lenses.

Correction of aberrations has traditionally been viewed as an optical design problem. The usual approach has been to combine lens elements of different shapes and glass types, balancing the aberrations of each element to improve the image quality of the combined

system. The most classical example of this might be the historical sequence of improvements in the original photographic objective, a landscape lens designed by Wollaston in 1812 [Kingslake 1934]. It consisted of a single-element meniscus lens with concave side to an aperture stop. In 1821, Chevalier improved the design by splitting the meniscus into a cemented doublet composed of a flint glass lens and a crown glass lens. Finally, in 1865, Dallmeyer split the crown lens again, placing one on either side of the central flint lens.

Today the process of correcting aberrations by combining glass elements has been carried to remarkable extremes. Zoom lenses provide perhaps the most dramatic illustration of this phenomenon. Zooming a lens requires a non-linear shift of at least three groups of lens elements relative to one another, making it very challenging to maintain a reasonable level of aberration correction over the zoom range. However, the convenience of the original zoom systems was so desirable that it quickly launched an intense research effort that led to the extremely sophisticated, but complex design forms that we see today [Mann 1993]. As an example, commodity 35 mm zoom lenses contain no fewer than 10 different glass elements, and some have as many as 23 [Dickerson and Lepp 1995]! Today, all modern lens design work is computer-aided [Smith 2005], where design forms are iteratively optimized by a computer. One reason for the large numbers of lens elements is that they provide greater degrees of freedom for the optimizer to achieve the desired optical quality [Kingslake 1978].

Figure 7.1: Spherical aberration.

This chapter introduces a new pure-software approach to compensating for lens aberrations after the photograph is taken. This approach complements the classical optical techniques. The central concept is simple: since a light field camera records the light traveling along all rays inside the camera, we can use the computer to re-sort aberrated rays of light to where they should ideally have converged. Digital correction of this kind improves the quality of final images by reducing residual aberrations present in any given optical recipe.

For simplicity, this chapter assumes a light field camera configured with the plenoptic

separation ($\beta = 1$) for maximum directional resolution. In addition, the analysis assumes monochromatic light – that is, light of a single wavelength. This simplification neglects the important class of so-called chromatic aberrations, which are due to wavelength-dependent refraction of lens elements. The techniques described here may be extended to colored light, but this chapter focuses on the simplest, single-wavelength case.

7.1 Previous Work

Ray-tracing has a long history in lens design. Petzval's design of his famous portrait lens was the first example of large-scale computation. In order to compete in the lens design competition sponsored by the Société d'Encouragement in Paris, he recruited the help of "Corporals Löschner and Haim [of the Austrian army] and eight gunners skilled in computing" [Kingslake 1989]. After about six months of human-aided computation, he produced a lens that, at $f/3.6$, was 16 times brighter than any other lens of its time. The lens was "revolutionary," and along with the use of "quickstuff," new chemical coatings designed to increase the sensitivity of the silver-coated photographic plates, "exposures were reduced to seconds" from the minutes that were required previously [Newhall 1976].

Kingslake reports from the perspective of 70 years in studying lens design that "by far the most important recent advance in lens-design technology has been the advent of the digital computer" [Kingslake 1989]. The reason for this is that lens design involves the tracing of a large number of rays to iteratively test the quality of an evolving design. One of the earliest uses of computer-aided ray-tracing in optimizing lenses seems to have been the LASL program at Los Alamos [Brixner 1963].

In computer graphics, ray-tracing of camera models has progressed from simulation of the computationally efficient pin-hole camera to one with a real lens aperture [Potmesil and Chakravarty 1981; Cook et al. 1984], to simulation of multi-element lenses with a quantitative consideration of radiometry [Kolb et al. 1995]. The method of Cook et al. differs from Potmesil and Chakravarty in the sense that it is based on a numerically unbiased Monte-Carlo evaluation of the rendering equation [Kajiya 1986]. It is one of the great algorithms of computer graphics, forms the basis of most modern ray-tracing techniques, and lies at the heart of the ray-tracing system [Pharr and Humphreys 2004] that I used to compute the

simulated light fields in this chapter and the previous one. One of the limitations of these kinds of ray-tracing programs is that they do not take into account the wave-nature of light. In particular, the simulated images in this chapter are free of diffraction effects – incorporating these requires a more costly simulation of the optical transfer function of the imaging system [Maeda et al. 2005]. The implicit assumption here is that the aberrations under study dominate the diffraction blur.

7.2   Terminology and Notation

This chapter introduces an extra level of detail to the ray-space notation used in previous chapters. The new concept is the notion of *two* sets of ray-spaces inside the camera: the *ideal* ray-space, which is the one we encountered in previous chapters, and the *aberrated* ray-space, which is composed of the rays physically flowing inside the camera body. Ideal rays are what we wish we had recorded with the light field camera, and aberrated rays are what we actually recorded. This subtlety was not necessary in previous chapters, where the implicit assumption was that the main lens of the camera was aberration-free. Let us differentiate between these two spaces by denoting an ideal ray as (x, y, u, v) and an aberrated ray as (x', y', u', v').

The two ray-spaces are connected by the common space of rays in the world. An aberrated camera ray maps to a world ray via geometric refraction through the glass elements of the main lens. In contrast, an ideal camera ray maps to a world ray via tracing through an idealized approximation of the lens' optical properties that is free of aberrations. In this chapter, we will use the standard Gaussian idealization of the lens based on paraxial optics, which is also known as the thick lens approximation [Smith 2005]. The Gaussian approximation is the linear term in a polynomial expansion of the lens' properties, derived by considering the image formed by rays passing an infinitesimal distance from the center of the lens. The process of transforming a ray through these ideal optics is sometimes referred to as Gaussian conjugation.

These two mappings into the world space define a mapping, C, directly from the aberrated space to the ideal space:

$$C : \mathbb{R}^4 \to \mathbb{R}^4$$

$$C(x', y', u', v') = (x, y, u, v). \tag{7.1}$$

I call this map the *ray correction function*, and its inverse the *ray distortion function*. These are the fundamental mappings that must be calculated in computing digitally corrected images. Conceptually, C results from composing the mapping from aberrated rays to world rays with the inverse of the mapping from ideal rays to world rays. A procedure to compute this mapping, as illustrated by Figure 7.2 is to take the input aberrated camera ray, trace it out into the world through the real optics of the lens, and then compute its Gaussian conjugate back into the camera

The ray correction function encodes the extent to which a real lens deviates from paraxial imaging. In a well-corrected lens where the residual aberrations are small, the ray-correction function is close to the identity mapping.

The light field sampling grid that is recorded by the light field camera is rectilinear in the
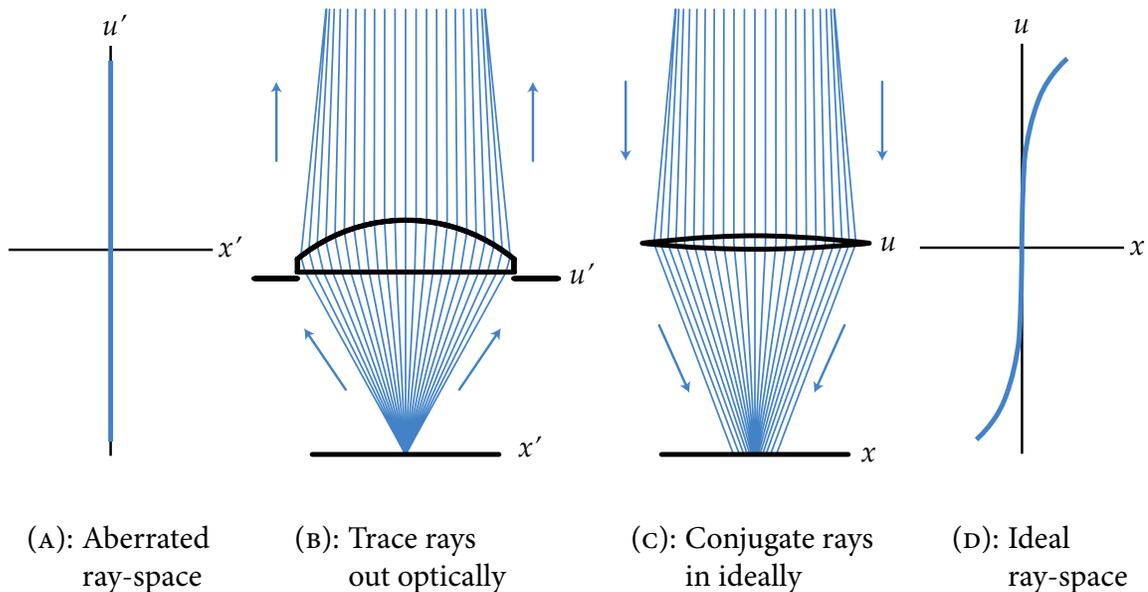

(A): Aberrated ray-space (B): Trace rays out optically (C): Conjugate rays in ideally (D): Ideal ray-space

Figure 7.2: Ray correction function.

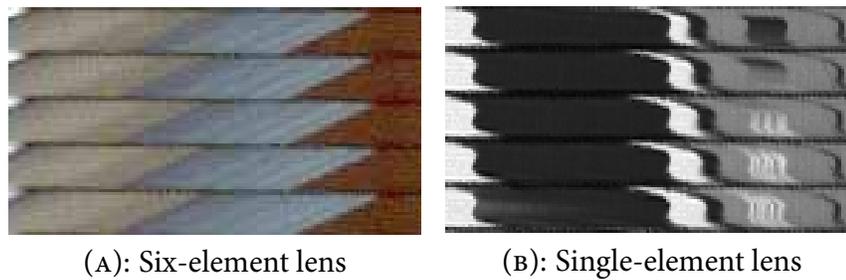(A): Six-element lens   (B): Single-element lens

Figure 7.3: Comparison of epipolar images with and without lens aberrations.

aberrated ray-space. Projection into the ideal ray-space warps the grid. The light field sampling diagrams throughout this thesis visualize this warped grid. In earlier chapters where aberrations were minimal and the ray correction function was close to the identity map, the resulting sampling grid was close to rectilinear. In this chapter, the grid lines become curved (Figure 7.2D illustrates the curving of one vertical grid line). The footprint of each photosensor pixel in the light field camera is, in general, a curved quadrilateral in the ray-space, rather than a simple box.

7.3   Visualizing Aberrations in Recorded Light Fields

Given the notion of aberrations as curvature in the ray-space, it is natural to look at the epipolar images of the light field for manifestations of such curves. Figure 7.3 illustrates epipolar images from two light fields recorded with the prototype camera. Figure 7.3A was shot with a consumer-grade lens that contains six glass elements and is well-corrected for aberrations. This lens is close to Gaussian ideal, and the boundaries between objects are straight lines as one expects in perfect light fields. In contrast, Figure 7.3B was shot with a single-element, plano-convex lens that exhibits heavy aberrations. The boundaries between objects are 'S' shaped curves, corresponding to the warping of ray-space lines due to aberrations in the lens.

One might observe that the S-shaped curves curves in Figure 7.3B are flipped horizontally compared to Figure 7.2D. This is because the epipolar images are a map from the ideal rays in the world into the aberrated space inside the camera. In other words, the epipolar images are plotted in the *aberrated* ray-space, and are actually a view of the distortion function
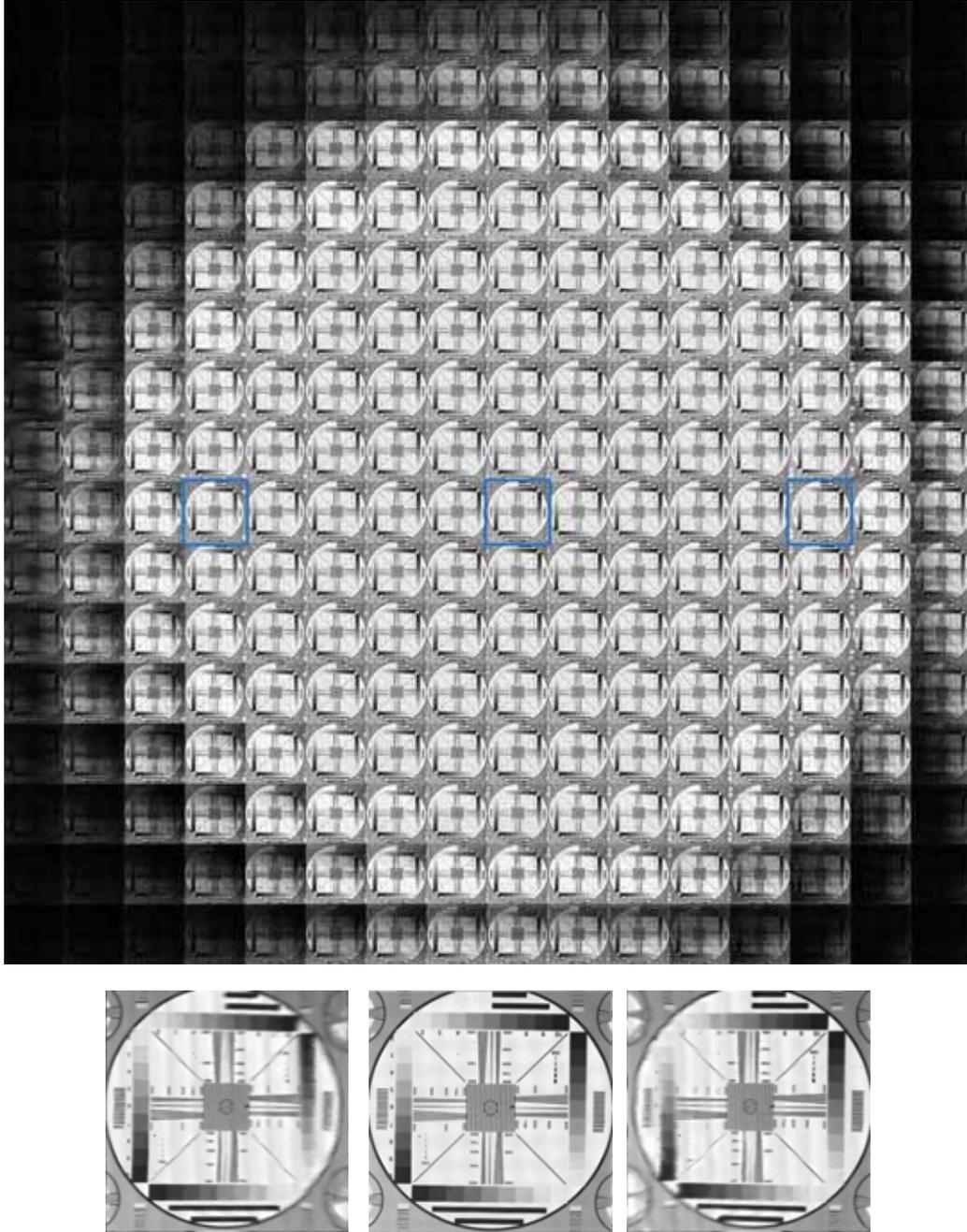
Figure 7.4: Aberrations in sub-aperture images caused by a plano-convex lens.

– the inverse of the correction function visualized in Figure 7.2D.

Tracing across a row of pixels in Figure 7.3B reveals an interesting feature of the aberrated epipolar images: the slopes of the curves are different for different points along the same row. This effect appears in rows away from the center of the epipolar image. A moment's consideration will convince the reader that a row of pixels in an epipolar image corresponds to a row of pixels in a sub-aperture image. Furthermore, from the discussion in Section 3.3, we know that the depth of an object in the world is related to its slope on the epipolar image. These two facts imply that in the aberrated sub-aperture images, different pixels will appear focused at different depths in the scene.

Evidence for this effect is visible in the zoomed sub-aperture images of the light field in Figure 7.4. The scene being photographed is a resolution test chart. Figure 7.4 shows that in sub-aperture images from the periphery of the lens, part of the image is blurred and part is in focus. It is easy to verify that the pixels which are blurry are focused closer in the world – moving the resolution chart closer to the camera brings these pixels into focus, while defocusing the rest of the resolution chart.

7.4   Review of Optical Correction Techniques

The visualization of the correction function on the ideal ray-space provides a different way to visualize the action of traditional optical correction techniques. Two classical techniques are stopping down the lens, and adding lens elements to balance aberrations.

Figure 7.5 illustrates stopping down the lens. The ray-space shown is ideal, and the curved columns are the rays integrated by pixels in a conventional camera. The curved vertical lines separating columns are referred to as *ray-intercept curves* in traditional optical engineering, although it is not common practice to illustrate their variation across the field in the same diagram, as shown on the ray-space diagrams here. The curves in Figure 7.5 show that each pixel collects light from a broad spatial range. The diagram shows that the the most highly aberrated regions come from the edges of the lens aperture (extremal u values), where the slope is greatest. Stopping down the aperture prevents these extremal rays from reaching the sensor pixel, reducing its x extent and image blur. Of course the price is reduced light sensitivity – much fewer rays are captured, so longer exposures are required.
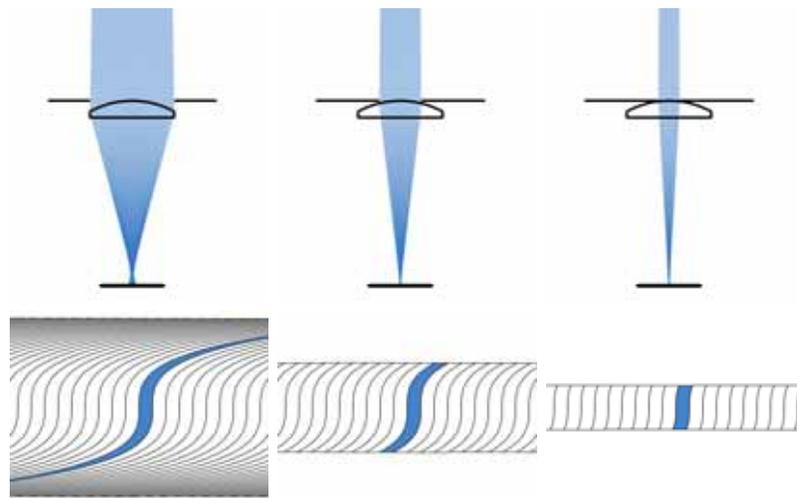
Figure 7.5: Classical reduction in spherical aberration by stopping down the lens.
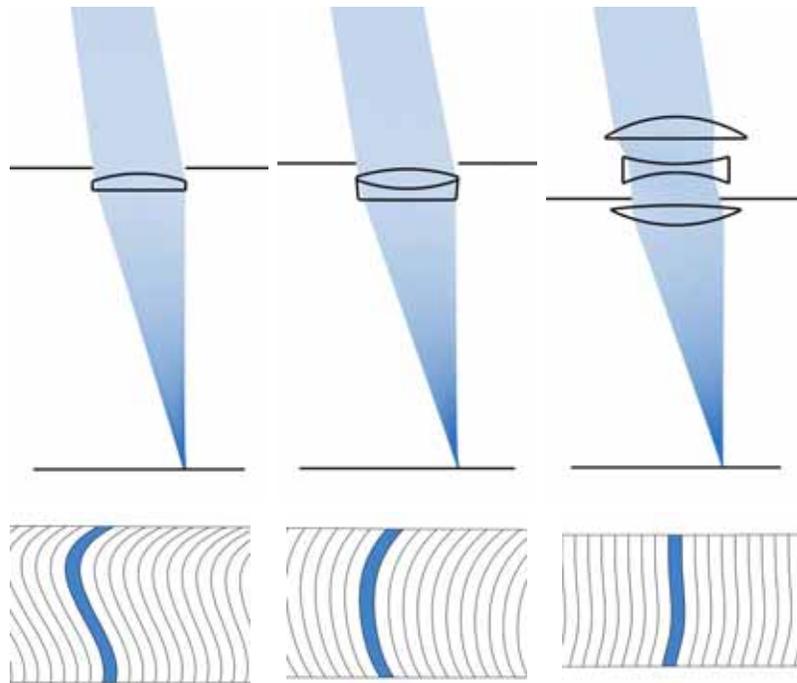

Figure 7.6: Classical reduction in aberrations by adding glass elements to the lens.

Figure 7.6 illustrates the second class of optical corrections – adding lens elements and tuning their respective curvatures and glass types. The ray-space diagrams show how adding elements provides the ability to shape the ray-space curves so that they are closer to vertical.

In contrast to these optical techniques, digital correction is an attempt to "straighten the curves" in software. This is possible because, in collecting ray information, the light field camera essentially splits the vertical curved columns in Figures 7.5 and 7.6 into multiple cells (Figure 7.7).

7.5 Digital Correction Algorithms

At a high level, digital correction of lens aberrations is simply a repetition of the basic concept at the heart of this thesis: resorting the rays in the recorded light field to where we ideally wanted them to converge. To determine where we want the rays to converge we will ray-trace a paraxial idealization of the lens, and to determine where the rays actually went in the recorded light field we will ray-trace an optical model of the real lens. In the latter case, we must accurately model the geometry of all the lens' curved glass elements, as in optical engineering.

Figure 7.7 is an overview of digital correction in terms of the ray-space, illustrating how to compute a single output pixel in a corrected image. Figure 7.7A illustrates a set of rays from a single point in the world, tracing into the camera through a double-convex lens. This highly aberrated lens was chosen for illustrative purposes. Figure 7.7B illustrates the ideal (x, u) ray-space inside the camera, with the aberrated (x', u') light field sampling grid superimposed. Each cell in the grid represents the rays integrated by a single photosensor pixel inside the camera. The vertical blue strip represents the set of rays shown on Figure 7.7A. Figure 7.7C illustrates estimation of the desired vertical strip using the recorded photosensor values. The procedure can be thought of as rasterizing the vertical strip onto the warped grip and summing the rasterized pixels. In contrast, Figure 7.7D illustrates all the rays collected by a single microlens in the camera – this is the pixel value that would have been recorded in a conventional photograph without digital correction. Note that the spatial extent of the curved strip is wider, hence more blurry, than the digitally-corrected estimate in Figure 7.7C.

Figures 7.7A–C illustrate the first implementation of digital correction: the *pixel-order* method, which involves iteration over the pixels of the output image. A second implementation is the *ray-order* method, involving iteration over the samples in the recorded light field.
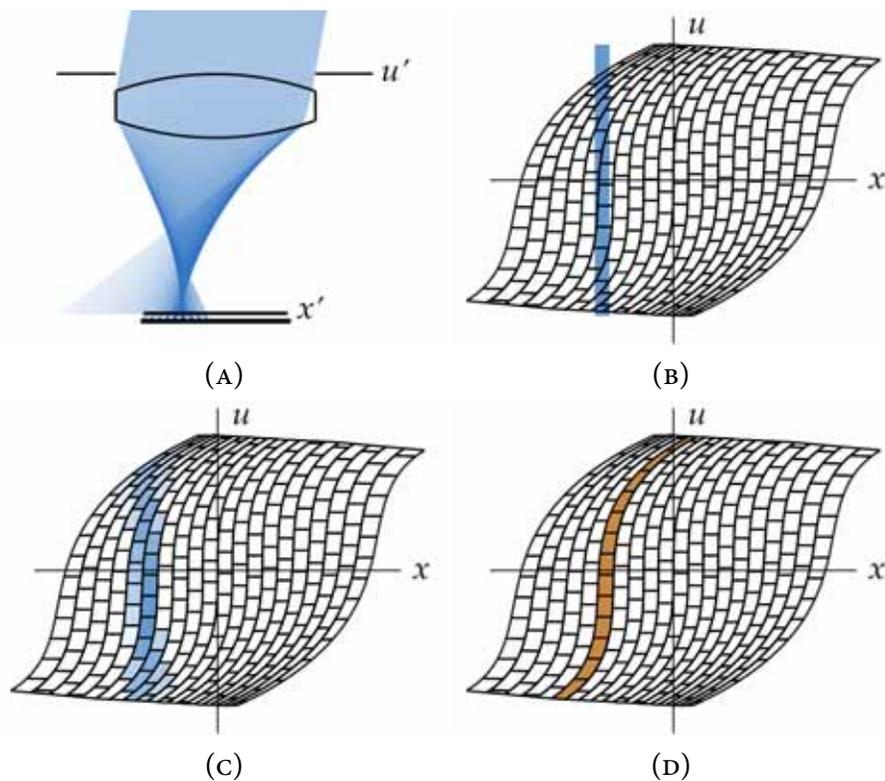
Figure 7.7: Ray-space illustration of digital correction of lens aberrations.

These are similar to the gather and scatter methods of texture resampling in computer graphics. The operations at the core of the two correction methods are the same: tracing through real optics with aberrations and tracing through idealized paraxial optics without aberrations. The two methods differ in the order that they apply these operations. The images in the remainder of the chapter are computed with the pixel-order algorithm. The ray-order algorithm is more convenient in numerical analysis of performance later in the chapter.

Pixel-Order Image Synthesis

The pixel-order method can be thought of as extracting the unaberrated energy for an output image pixel from different cells in the aberrated light field. It comprises the following steps to compute the value of each output image pixel.

1. Sample all the ideal camera rays converging to that output pixel. A Monte-Carlo method is to draw random samples distributed over the corresponding sensor pixel's area and

over the aperture of the lens.

2. Compute the world-space conjugates of the rays using the ideal paraxial approximation for the camera lens.

3. Reverse the direction of the world rays and ray-trace them back into the camera through the geometrically accurate model of the camera's lens, through the microlens array and down to the sensor surface.

4. Estimate the radiance along each ray from the neighborhood of sensor pixel values in the recorded light field. The images below use quadrilinear interpolation of the nearest 16 samples in the 4D space. Lower-quality nearest-neighbor interpolation can be used for speed. Slower, wider reconstruction filters can be used for higher image quality.

5. Average the radiance estimates to compute the final output pixel value.

Ray-Order Re-Projection of the Light Field

The ray-order method can be thought of as re-projecting the aberrated energy in the light field into an unaberrated output photograph. It comprises the following steps for each cell in the recorded light field.

1. Sample the bundle of rays inside the camera that would converge to the corresponding sensor pixel in the light field camera. A simple Monte-Carlo method for sampling this bundle of rays is to draw random samples over the area of the sensor pixel, and random directions over the pixel's parent microlens.

2. Trace these rays away from the sensor surface, through the microlenses, through the geometrically accurate model of the camera's lens and out into the world.

3. Reverse the direction of the world rays and compute their optical conjugates back into the camera using the ideal paraxial approximation of the camera's lens.

4. Intersect these rays with the imaging plane. At each location, add the light field sample value into a running sum of the values at the pixel in the corresponding location.

After this process concludes, normalize the value of each output image pixel, dividing by the number of rays summed there over the course of processing the entire light field.

Confidence Weighting for Increased Contrast Enhancement

The non-linear distortions introduced by aberrations mean that some light field cells pollute the corrected photograph more than others. We have seen this effect in two different ways so far. First, in looking at aberrated sub-aperture images in Figure 7.4, we saw that the same region of the scene can appear with very different amounts of blur when viewed from different parts of the lens. Second, in looking at the projection of ideal vertical strips of ray-space onto the aberrated light field sampling grid, we saw that some grid cells could be much wider than the ideal strip, leading to larger amounts of blur. For example, in Figure 7.7D the widest grid cells contributing to the estimate are at the top of the grid.

These observations motivate an optional enhancement in the resampling process for digital correction, designed to further raise the contrast and clarity of the corrected image. The idea is to weight the contribution of each photosensor pixel in inverse proportion to its spatial extent when projected onto the output image plane. This modification means computing a *weighted* average of light field sample values in the final step of the pixel-order algorithm.

In the corrected images shown later in this chapter, I used the following weight function, where Δx and Δy are the projected width and height of the light field cell in the output image. For convenience, the units are in terms of output pixel widths.

$$w(\Delta x, \Delta y) = h(\Delta x) \cdot h(\Delta y), \quad \text{where}$$

$$h(x) = \begin{cases} 1, & x \leq 1 \\ \exp\left(-\frac{(1-x)^2}{2\sigma^2}\right), & x > 1 \end{cases}. \tag{7.2}$$

In words, the weighting function decreases according to a Gaussian fall-off as the projected width of the cell increases beyond one output image pixel. The x and y dimensions are treated separately, with the overall weight being the product of the weights for each dimension. I used a standard deviation of $\sigma = 2$ for the Gaussian fall-off. Calculation of Δx and Δy, which varies as a function of (x, y, u, v), is discussed at the end of this section.

Figure 7.8 visualizes the weighting function of the aberrated light field cells. Each pixel's weight is proportion to how blue it appears in this figure. The figure illustrates that the weight tends to be higher for rays passing through the center of the lens, where the aberrations are least. A more subtle and interesting phenomenon is that the weight varies across the
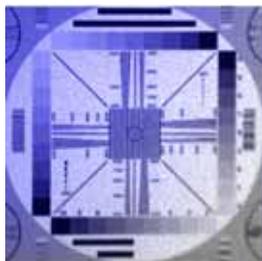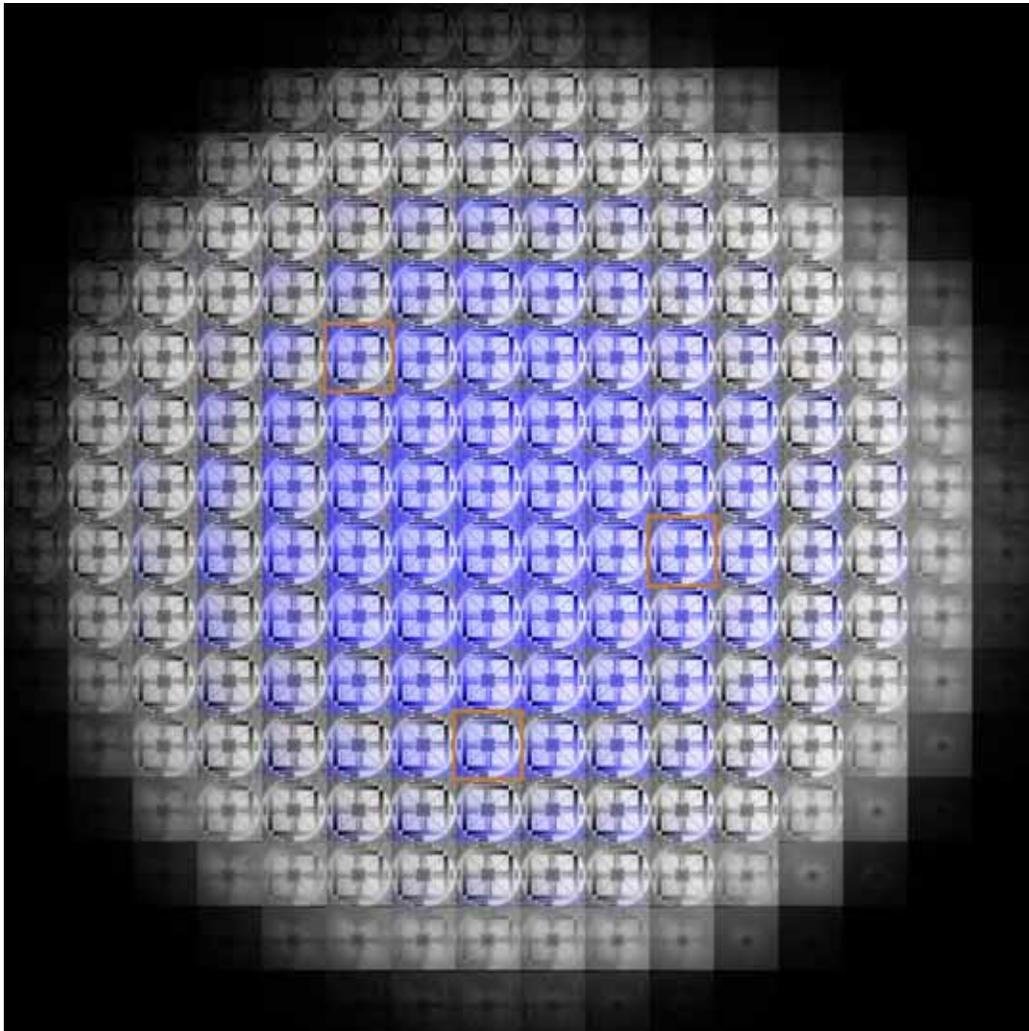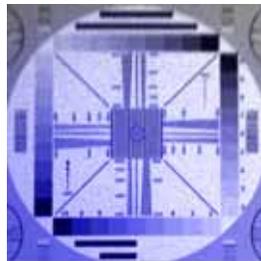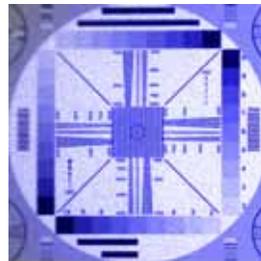
Figure 7.8: Weighting of rays (light field cells) in weighted correction. Each ray's weight is proportional to how blue it appears.

pixels in the same sub-aperture image, as shown in the three zoomed images ($z_1 - z_3$). Close examination reveals that the weights are higher for areas in sharp focus, exactly as the weighting function was designed to do.

Reducing the weight of the blurry samples reduces residual blur in the corrected photograph. Equation 7.2 defines one weighting function, but of course we are free to design others. Choosing a weighting function that reduces the weight of cells with larger projected area more aggressively results in greater contrast and resolution. The trade-off, however, is that reducing the average weight (normalized to a maximum weight of 1) decreases the effective light gathering power of each output pixel. For example, the average weight of the cells in Figure 7.8 is 32%, which in some sense matches the light gathering power of a conventional camera with an aperture reduced to 32% area. However, stopping down the lens imposes the same sub-aperture on every output image pixel. Weighted correction provides the extra freedom of varying the aperture across the image plane. As shown in the experiments below, this allows weighted correction to produce a sharper image.

Computing the Projected 2D Size of Light Field Samples

Computing Δx and Δy for the weighting function in Equation 7.2 involves projecting the aberrated light field cell onto the output image plane and calculating its 2D size. In practice, it is sufficient to approximate the projected size by assuming that the correction function, C, is locally linear over the light field cell. In this case, Δx can be approximated using the first-order partial derivatives of the correction function:

$$\Delta x \approx \frac{1}{\Delta x'} \left(\left| \frac{\delta C_x}{\delta x'} \Delta x' \right| + \left| \frac{\delta C_x}{\delta y'} \Delta y' \right| + \left| \frac{\delta C_x}{\delta u'} \Delta u' \right| + \left| \frac{\delta C_x}{\delta v'} \Delta v' \right| \right), \tag{7.3}$$

where I have defined the four components of C explicitly:

$$\begin{aligned} C\left(x',y',u',v'\right) = \Big(\ & C_x(x',y',u',v'), \\ & C_y(x',y',u',v'), \\ & C_u(x',y',u',v'), \\ & C_v(x',y',u',v') \ \Big) = (x,y,u,v) \,. \end{aligned} \tag{7.4}$$

The analogous equation for Δy is

$$\Delta y \approx \frac{1}{\Delta y'} \left(\left| \frac{\delta C_y}{\delta x'} \Delta x' \right| + \left| \frac{\delta C_y}{\delta y'} \Delta y' \right| + \left| \frac{\delta C_y}{\delta u'} \Delta u' \right| + \left| \frac{\delta C_y}{\delta v'} \Delta v' \right| \right). \tag{7.5}$$

Let me focus your attention on three features of these equations. First, dividing by $\Delta x'$ and $\Delta y'$ normalizes the units so that they are relative to the size of output image pixels, as required by the weighting function in Equation 7.2.

The second point is that the partial derivatives in these equations vary as a function of the light field cell position (x, y, u, v). For example in Figure 7.7, $\frac{\delta C}{\delta x'}$ and $\frac{\delta C}{\delta u'}$ are the vectors parallel to the distorted horizontal and vertical lines of the sampling grid, and the distortion varies over the ray-space. I compute the value of the partial derivatives using simple finite differences of the sampled correction function, C. Recall that computing $C(x', y', u', v')$ is a matter of tracing ray (x', y', u', v') out of the camera into the world using a model of the real optics, then ideally conjugating it back into the camera using idealized paraxial optics.

The third point to note is that $\Delta x'$, $\Delta y'$, $\Delta u'$ and $\Delta v'$ are constants in Equations 7.3 and 7.5. $\Delta x'$ and $\Delta y'$ are the width and height of the microlenses in the light field camera (9.25 microns in the prototype). $\Delta u'$ and $\Delta v'$ represent the projected size of the sensor pixels on the (u', v') lens plane. For example, the experiment described in the next section uses a plano-convex lens with a clear aperture diameter of approximately 40 mm. With a directional resolution of 12×12, $\Delta u'$ and $\Delta v'$ are approximately 3.33 mm for that experiment.

7.6   Correcting Recorded Aberrations in a Plano-Convex Lens

There were two over-arching goals to the experiments in this section. The first was to visually demonstrate that digital correction could raise contrast and resolution in real images acquired with the prototype camera. The second goal was to use the prototype camera data to provide a measure of validation for our simulation software, used to compute raw light fields and digital corrections. This software is used in the last part of the chapter in quantitative performance tests of a wider range of lenses at much higher light field resolutions.

The lens tested in this section is a plano-convex lens with a focal length of 100 mm. It is made out of standard BK7 glass. It is similar to the one illustrated in Figure 7.1, and it

(A)   (B)
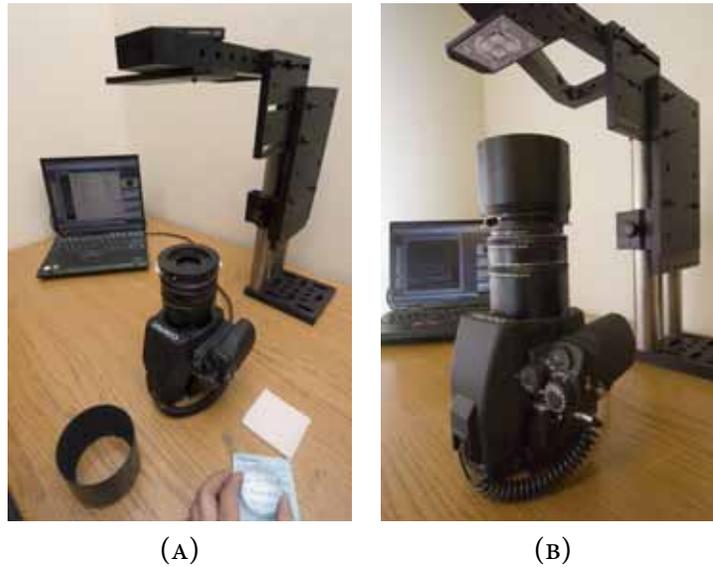
Figure 7.9: Set-up for plano-convex lens prototype.

appears in the bottom right of the photograph in Figure 7.9A. This simple lens was chosen because it produces aberrations extreme enough to be visible and correctable in the relatively low-resolution 296×296 photographs produced by our prototype.

I recorded an aberrated light field using the prototype camera by replacing its usual photographic lens with this plano-convex lens (convex side up). A manual aperture was placed against the planar side of the lens, and stopped down to achieve an $f/4$ aperture. I set the separation between lens and image plane to focus on a resolution test-chart approximately 24 cm away, as shown in Figure 7.9, and tuned focus by adjusting the height of the target until maximum sharpness was achieved.

I simulated a matching raw light field using Monte-Carlo ray-tracing. The computer model of the lens, microlens array and sensor were matched to the manufacturer's physical specifications. The separation between the main lens and the microlens plane was matched to measurements on the prototype set-up. As with the physical set-up, I tuned focus by adjusting the distance of the virtual resolution chart until maximum sharpness was achieved.

(A) No correction
(Conventional photograph)

(B) Digital correction
No confidence weighting

(C) Digital correction
Confidence weighting
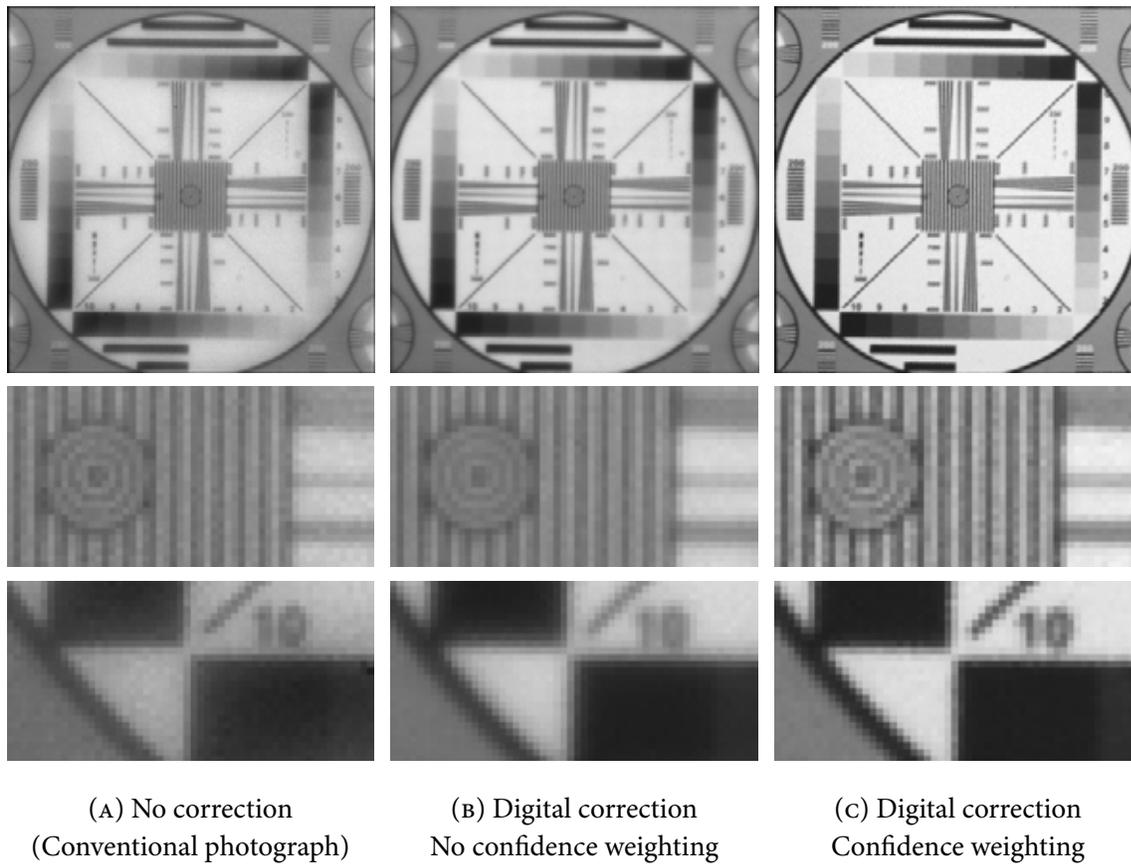
Figure 7.10: Comparison of uncorrected and corrected images from a light field recorded with the prototype camera.

Results

Figure 7.10 visually compares the quality of images computed from the recorded light field with and without correction. Column A, computed without correction, is equivalent to a conventional photograph. It exhibits the classical softness in resolution across the image plane due to spherical aberration. In addition, the zoomed image at the bottom of Figure 7.10A illustrates significant loss in contrast at the edge of the frame, where regions that should be black and white appear as gray due to cross-pollution. Column B illustrates that correction raises the contrast, particularly along the edges of the image but less so in the center of the frame where aberrations are less. Column C illustrates that weighted correction

(A1): Recorded light field   (B1): Simulated light field

(A2): Corrected photograph from A1   (B2): Corrected photograph from B1
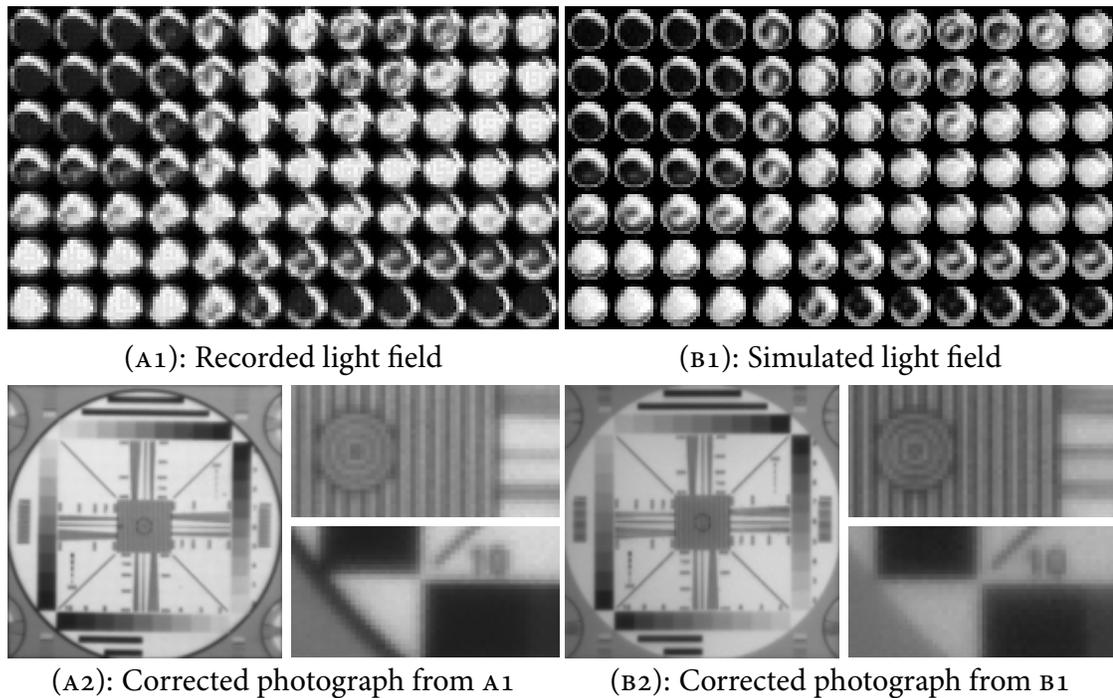
Figure 7.11: Comparison of recorded and simulated data for digital lens correction.

raises the contrast and resolution further still.

Figure 7.11 compares the recorded data with the simulated version. Images A1 and B1 compare close-ups of the raw light field data. Even at this extreme level of zoom, the overall match is quite good, although small differences are visible due to error in calibrating the physical and virtual geometry. Figure 7.11 A2 and B2 illustrate that these calibration errors cause only small differences in output images. These two images are for correction without weighting, and similarly good agreement is found in uncorrected images and correction with weighting.

Comparison of Weighted Correction with Reduced-Aperture Conventional Imaging

From the discussion in Section 7.5, we know that the weighted correction used on the plano-convex experiment results in an average light usage of 32% of the light field.

Figure 7.12 compares the corrected image with weighting against a conventional image
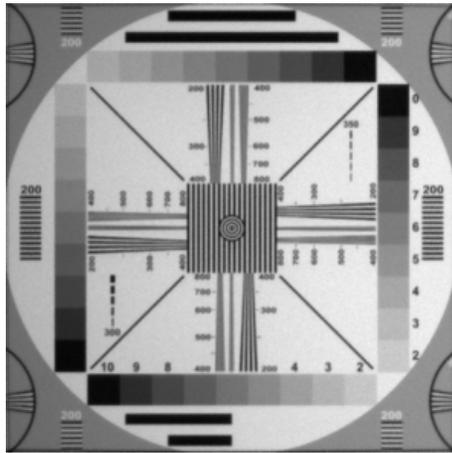
Digitally corrected image with weighting   Conventional photograph, 56% aperture
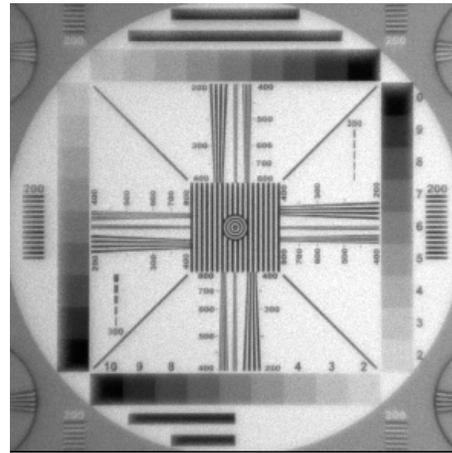
Figure 7.12: Comparison of weighted correction with conventional imaging where the lens aperture is stopped down for equivalent light gathering power.

where the aperture is reduced to 32% area (56% diameter). The conventional image was computed by only summing the rays under each microlens that passed through the reduced aperture, without resorting of rays. Although the aberrations in the stopped-down conventional image are reduced compared to the full-aperture version in Figure 7.10, the weighted correction still provides significantly better contrast. For example, the black bars at the top and bottom are much darker in the corrected image.

Weighted correction produces a superior image for two reasons. First, it resorts rays, which improves in the convergence of rays that are used. Second, weighted correction has greater flexibility in choosing to use rays that converge well on the resolution chart. In the conventional case, stopping down the lens excludes sub-aperture images from the periphery of the lens that tend to contain a larger fraction of blurry pixels, but even the reduced aperture contains some of these artifacts. In contrast, weighted correction can use an effectively larger aperture and discard the worst rays, as shown in the zoomed images of Figure 7.8.

7.7 Simulated Correction Performance

In contrast to the previous section, the results in this section derive purely from computer simulation. The experiments here apply traditional, numerical analyses [Smith 2005] of the ray-traced point spread function (PSF) to compare the performance of various lenses with and without digital correction. The PSF is the spot of energy that appears in an output image response to a point source of light in the world. Ideal imaging produces a diffraction-limited spot, but in practice aberrations usually result in a larger blur. One of the main goals of this section was to explore how digital correction works across a range of lenses. Is it likely to be a general-purpose technique for improving the quality of lenses in optical engineering? To keep the comparison as simple as possible, this section examines only digital correction without weighting.

7.7.1 Methods and Image Quality Metrics

The cameras simulated in this section assume the following geometry and resolutions. A 35 mm format sensor is assumed – that is, the microlens array and photosensor both measure 36 mm $\times$ 24 mm. The spatial resolution, that is the resolution of the microlens array, is assumed to be constant at 1800$\times$1200 (2.1 MP). A range of $N \times N$ directional resolutions are tested, from $N = 1$ (uncorrected), up to $N = 16$. Since the spatial resolution and sensor size are fixed, increasing N assumes increasing photosensor resolution. $N = 10$, requiring 1.9 micron pixels, lies at the limit of technology currently shipping in commodity cameras [Askey 2006]. 1.7 micron pixels have been demonstrated in the laboratory [Micron 2005], and the maximum resolution simulated here, $N = 16$, assumes a further reduction by 26% down to 1.25 micron.

The imaging performance of various lenses was quantified by computing PSFs to analyze imaging performance at different points on the imaging plane. Computing a PSF means tracing rays from a point light source in the world through all parts of the lens, down to the imaging plane to produce an image. In the case of simulating a light field camera, the rays are traced through the microlens array down to the photosensor surface to produce a raw light field. Final corrected photographs of the PSF were computed using the ray-order method, which is more efficient than the pixel-order method in this case, because so few

photosensor pixels are illuminated by the PSF calculation.

Although the PSF can be used to provide very detailed analysis of imaging performance, it can be cumbersome for comparison across lenses because it is very high-dimensional. The PSF is a 2D function that varies with position on the imaging plane, as well as focal depth. For example, it is very common for the PSF to become broader at the edge of the image. As another example, most lenses do not focus well at the short focal distances required for high magnification (except for specially-designed macro lenses). For such close-focusing, the PSF tends to spread out.

(A): Spherical aberration
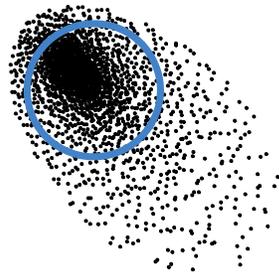
(B): Coma

Figure 7.13: PSF and RMS measure.

In any case, to compare performance across lenses, I used a series of summary statistics derived from the PSF. The first level of summary is provided by the root mean square (RMS) spot radius, which is sometimes used in optical design texts. The RMS spot radius can be thought of as the standard deviation of the PSF interpreted as a probability distribution function. As with all statistics, the RMS measure is a double-edged sword. On the one hand, it provides a very compact summary of one of the most important characteristic of the point spread: it's gross size. Figure 7.13A is a spot diagram illustrating the density of a PSF exhibiting spherical aberration, overlaid with a circle of one RMS radius.

On the other hand, the RMS measure discards information about the shape of the spot, which can greatly affect the qualitative appearance of aberrations in the final image. For example, Figure 7.13B illustrates a PSF exhibiting the classical aberration known as *coma*. This comet-shaped PSF is oriented radially away from the center of the image – the illustrated PSF comes from the bottom right of an aberrated image. Coma is well known to cause objects to appear as if they are "flying out of the field of view" [Smith 2005]. The overlaid RMS circle clearly does not capture this visual characteristic, giving only the standard deviation.

The RMS spot radius is a large reduction in information, but it is still a 2D function over

the image plane. The next level of summary is provided by averaging the RMS measure across the 2D plane, producing a single-number summary of the entire imaging system's performance. Although it is somewhat crude, this measure does track performance trends faithfully, improving with the quality of the photographic lens and with the amount of directional resolution available in digital correction.

An equivalent, but slightly more intuitive measure of average RMS radius is *effective resolution*. This measure is designed to give a rough idea of the "resolution" of output images. I have, somewhat arbitrarily, defined the effective pixel size as the square that can be inscribed within a circle of one average RMS spot radius (Figure 7.14). The effective resolution is then just the number of squares that fit within the sensor. In the experiments below, the effective resolution is therefore defined to be

$$\frac{24 \text{ mm} \times 36 \text{ mm}}{(\sqrt{2}R)^2},$$

Figure 7.14: Effective pixel size.

where R is the average RMS spot radius (in mm), and $\sqrt{2}R$ is the width of the inscribed square. The thought underlying the concept of effective resolution is that the total computed resolution of the output image is irrelevant if the PSF is much broader than one output image pixel. Effective resolution provides a measure of the number of PSF-limited spots that we can discriminate in the output image.

The final measure of image quality used in this section is the MTF. The MTF is one of the most popular measures of image quality, providing a very sensitive measure of the ability of the imaging system to resolve contrast and fine details in the output image. It is defined as the magnitude of the Fourier transform of the point spread function. Indeed, that is how I computed the MTF in the experiments below: by first computing the PSF and then computing its discrete Fourier transform. A detail of this implementation is that it incorporates the effective MTF due to the finite sized microlenses and pixels of the sensor [Park et al. 1984]. This means that the computed MTF does not exceed the Nyquist sampling rate of the 20 micron spatial sampling grid.

Since the MTF is the Fourier transform of the PSF, it is as high-dimensional as the PSF.
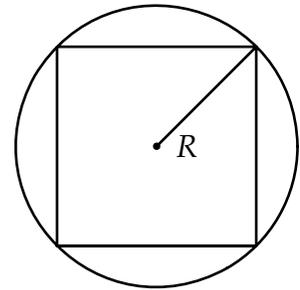
This chapter employs one of the most common summaries of the MTF, which is to plot a small number of spatial frequencies as a function of distance on the imaging plane away from the optical axis. If image quality degrades towards the edge of the image, the reduction in quality would be visible as a decay in the plotted MTF. Additional features of the MTF are described in relation to the plots presented below.

7.7.2   Case Analysis: Cooke Triplet Lens

Let us first analyze the PSF of a particular lens in some detail. The chosen lens is an $f/2.9$ Cooke triplet [Tronnier and Eggert 1960]. Figure 7.15 illustrates the tracing of rays through the lens' three glass elements in the simulation of the light field camera's PSF. Columns A, B and C show rays converging on three different positions in the image: at the optical axis, half-way towards the edge of the image, and at the edge.

The middle row zooms in on a 200 micron patch of the imaging plane at the convergence of rays. Ten microlenses are shown across the field of view, and the rays terminate at the bottom on the photosensor surface. These close-up views illustrate the complexity of the ray structure inside the camera. The shape of rays converging on the microlens plane is far from a simple cone as assumed in ideal imaging. In addition, the shape of rays and quality of convergence change across the imaging plane. For example, Diagram A2 shows that the rays in the center of the frame converge onto 2 microlenses. In contrast, Diagram C2 shows worse convergence near the edge of the image, with rays spread over approximately 6 microlenses.

The bottom row of Figure 7.15 provides a different view of the aberrations, in ray-space. The horizontal, x, field of view is 200 microns, but the vertical, u, field of view spans the entire lens plane. The directional resolution is $N = 16$. The distortion of the sampling grid provides more information about the nature of the aberrations at different parts of the field. In Diagram C1, one can see how optical correction has been used to force rays coming from the edge of the lens to converge reasonably well. In fact, the worst aberrations come from the parts of the lens midway between the optical axis and the edge of the lens. In contrast, Diagram C1 shows that at the right edge of the image, the worst aberrations come from the left-most portion of the lens, since the grid is most distorted near the bottom.

These ray-diagrams highlight the crucial concept of scale relative to image resolution
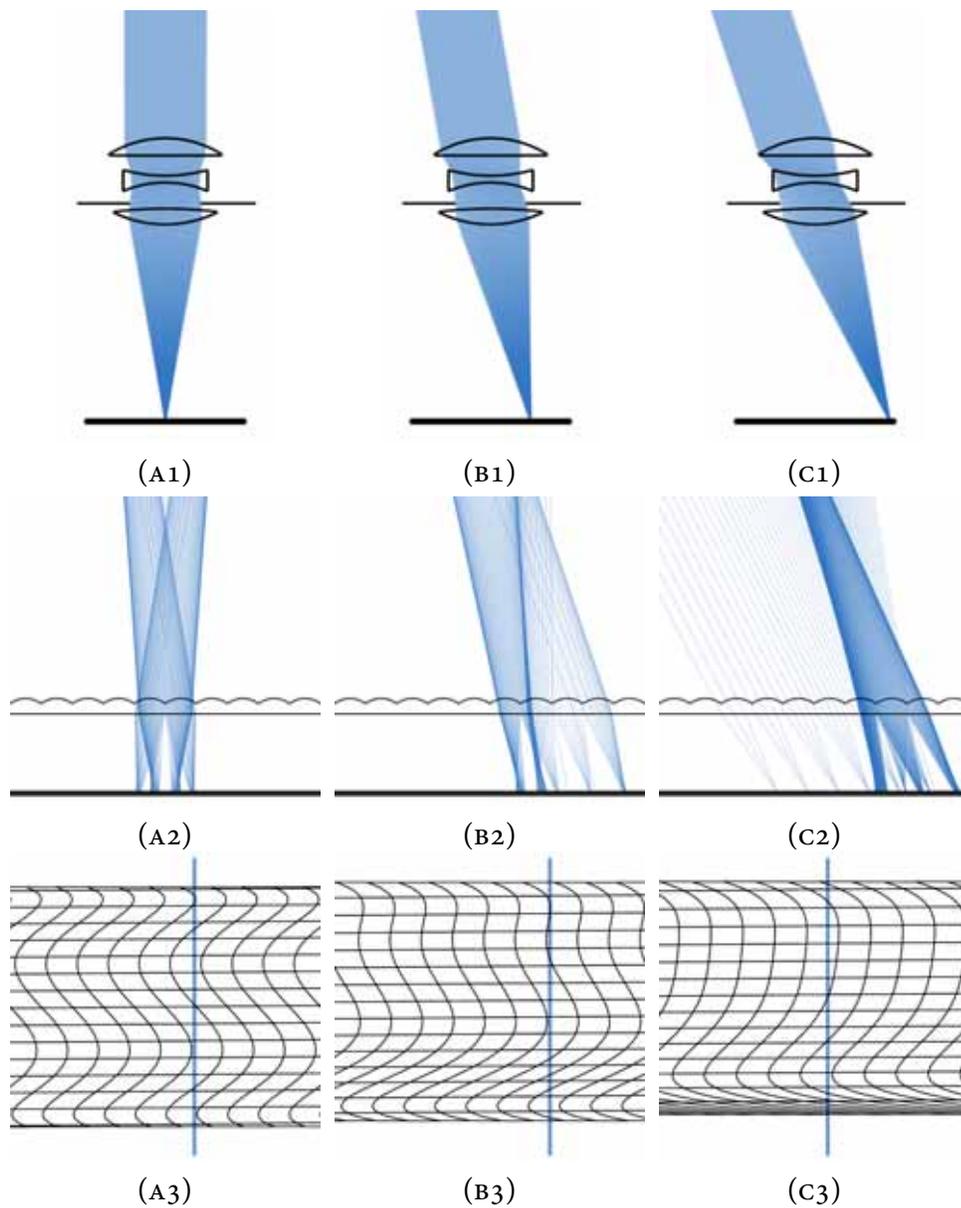
Figure 7.15: Aberrated ray-trace and ray-space of a Cooke triplet lens.

when considering the seriousness of aberrations. If the spatial resolution were less, the curvature would be negligible relative to the spacing of the grid columns. Conversely, any residual aberration will exhibit significant curvature if the spatial resolution is high enough. This means that to correct for aberrations, the number of vertical cuts needed in each column,

⊢—⊣ 20 microns   ⊢—⊣ 20 microns   ⊢—⊣ 20 microns

(A): Center   (B): Middle   (C): Edge

Uncorrected spot diagrams for Cooke triplet PSF.

⊢—⊣ 20 microns   ⊢—⊣ 20 microns   ⊢—⊣ 20 microns

(D): Center   (E): Middle   (F): Edge
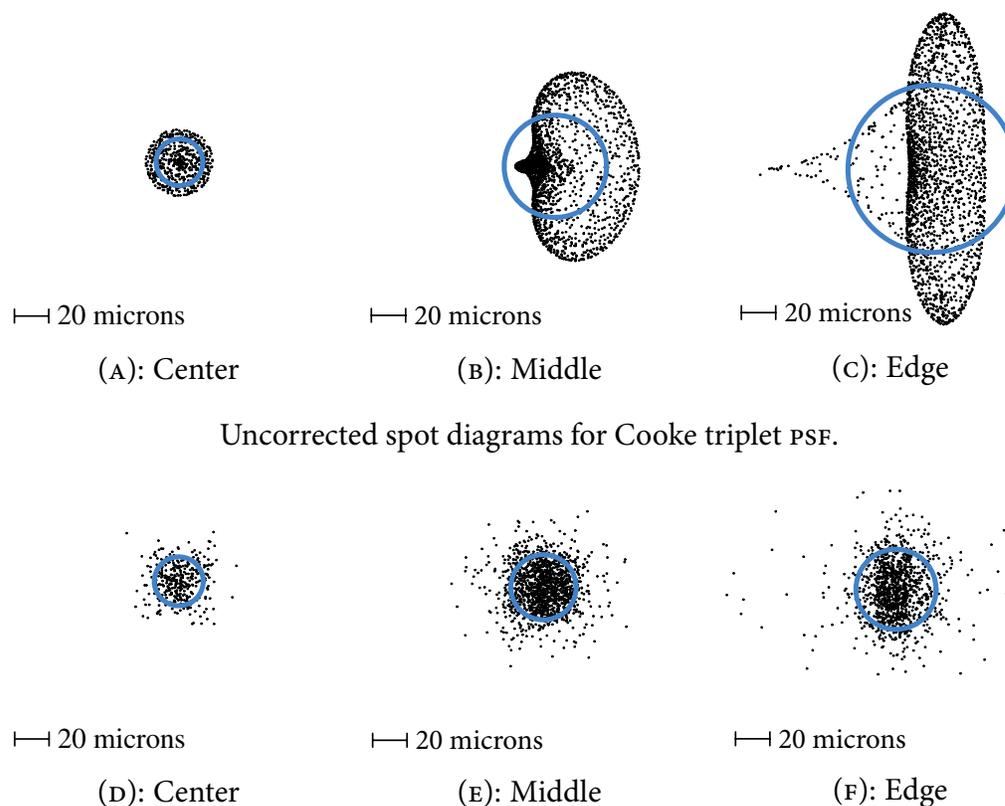
Figure 7.16: Digitally corrected spot diagrams for Cooke triplet PSF.

corresponding to the directional resolution, is proportional to the spatial resolution. This makes intuitive sense: increasing the spatial resolution increases our sensitivity to aberrations in the output image, and raises the amount of directional resolution required to correct for the residual aberrations.

Figure 7.16 illustrates the uncorrected and corrected PSF for the triplet lens at the three positions shown in Figure 7.15. The overlaid circles show the RMS spot. With this amount of directional resolution (16×16), the average RMS spot radius is roughly 14 microns, and the effective resolution of output images is close to the full 2.1 MP of the microlens array. Figure 7.16 illustrates an important characteristic of digital correction that is not captured by the summary statistics. It shows that digital correction tends to make the PSF more radially symmetric, reducing qualitative problems introduced by aberrations such as coma.

Figure 7.17 illustrates the improvement in correction performance as the directional
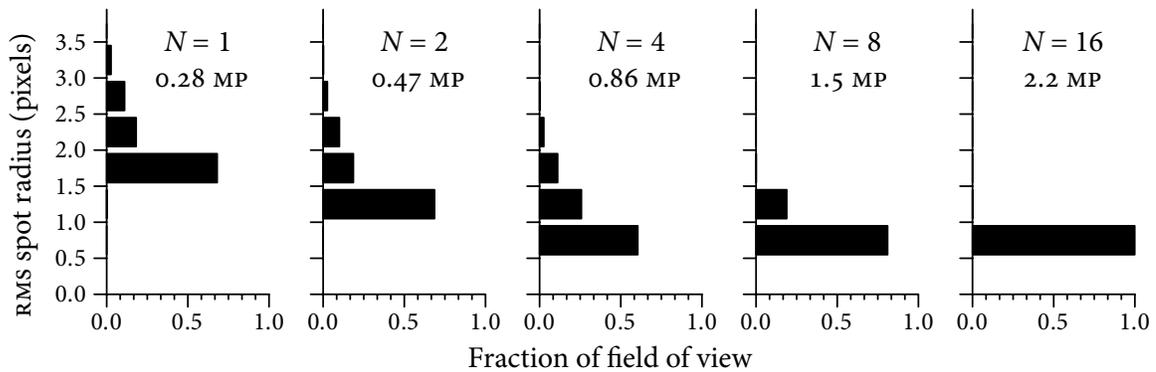
Figure 7.17: Histogram of triplet-lens PSF size across imaging plane.

resolution ($N \times N$) increases from $N = 1$ (uncorrected conventional imaging) to 16. The histograms for each value of N illustrate the distribution of RMS radii across the imaging plane. The vertical axis of the graph is measured in terms of output image pixels. For example, for $N = 1$, 68% of the imaging plane has an RMS spot radius of approximately 2 pixel widths. These histograms were computed by sampling the PSF at over 1000 different positions distributed evenly but randomly (using stratified sampling) across the 36 mm $\times$ 24 mm imaging plane. 65,536 rays were cast in estimating the RMS radius for each PSF.

The histograms illustrate how the corrected RMS spot size converges as the directional resolution increases. The rate of convergence depends on how much aberration is present in the lens, with greater distortions requiring larger amounts of directional resolution for accurate correction. The effective resolution of the output image grows from 0.28 MP to the full 2.2 MP as the directional resolution increases from $N = 1$ to 16. The starting, uncorrected resolution may seem surprisingly low. One factor is that we are using the lens with its aperture wide-open, where aberrations are worst. Second, effective resolution decays rapidly as the spot size increases. If the effective resolvable spot covers just 3 pixels, as it almost does in the triplet lens without correction, then the effective resolution decreases by almost an order of magnitude.

The histograms give a sense for how the average performance increases across the image, but not how the performance varies from the center to the edge. Figure 7.18 measures variation along that axis, in the form of MTF plots. With microlenses of 20 micron width,
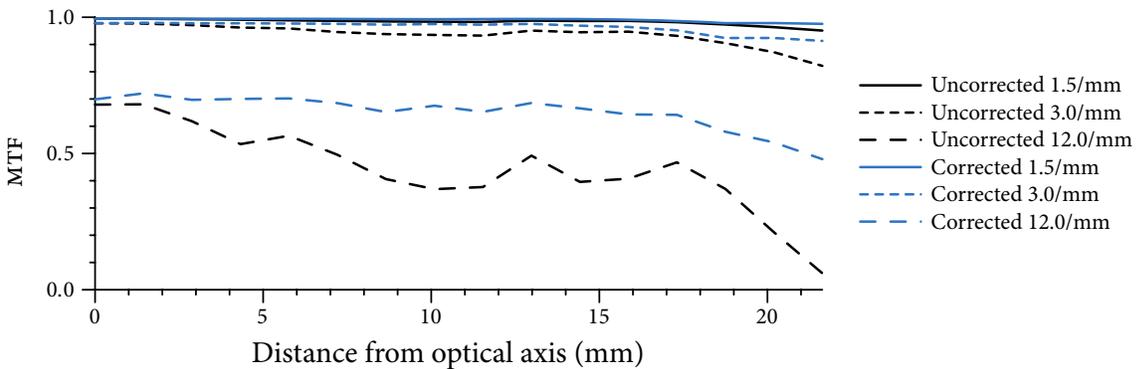
Figure 7.18: MTF of triplet lens with and without correction (infinity focus).
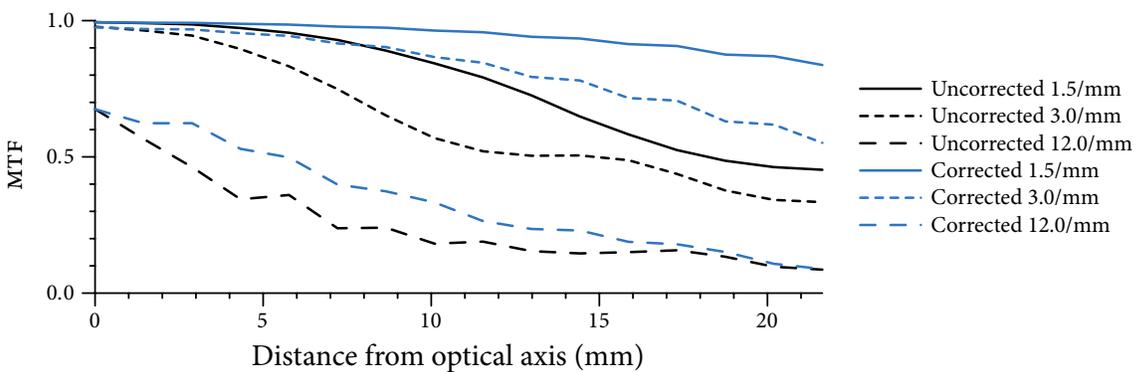
Figure 7.19: MTF of triplet lens with and without correction (macro focus).

the Nyquist spatial sampling rate is 25 cycles per mm. The graphs plot the MTF at three spatial frequencies: 1/2, 1/4 and 1/8 the Nyquist rate. The curve for 1/2 the Nyquist rate is generally correlated with "sharpness," ability of the system to resolve fine details. Curves above 70% are extremely rare at these higher frequencies. The curves for lower frequencies generally correlate with how "contrasty" the final image is. Good lenses have very high MTF at these frequencies, and conventional wisdom is that differences of a few percentage points are visible in final images.

Figure 7.18 shows that, even without digital correction, the contrast of the triplet lens is quite good across the imaging plane, dipping only slightly at the very edge of the image. Sharpness, however, is relatively poor at these resolutions, especially out towards the edge.

Digital correction improves sharpness and contrast across the field of view, and makes performance more even.

Figure 7.19 illustrates the triplet used to focus much closer, at a distance of 100 mm for 1:1 magnification (i.e. macro photography). The lens was not designed for such close focusing, and its performance is quite poor at this depth. This phenomenon reflects the kind of trade-off that is inherent in all real optical engineering, as a consequence of Maxwell's principle that perfect imaging at all focal depths cannot be achieved using only refraction and reflection. The designers of the triplet lens prioritized simplicity of the optical recipe (using only three elements), and performance when focused at infinity. Those constraints were incompatible with good imaging performance in the macro range.

The corrected MTF in Figure 7.19 illustrates dramatic improvement in the contrast of the system. This example illustrates that digital correction *extends the useful focal range* of a lens. However, the improvement in the sharpness is moderate, because the directional resolution of the system was not sufficient relative to the distortion in the light field at this focal depth. The reason for this is two-fold. First, the light field distortion is much greater at the macro focal depth. Figure 7.20 illustrates this fact by comparing the ray-space for infinity and macro focus at the three image positions shown in Figure 7.15. From the discussion previously in this chapter, it is evident that a higher directional resolution would be required to correct the greater aberrations at the macro focus.

Unfortunately, the second factor is that the directional resolution is halved at macro focal depths. This can be seen in Figure 7.20 by the fact that there are only half as many vertical cuts in the grid columns of D-F as there are in A-C. The underlying cause of the reduction is that the separation between the main lens and the imaging plane increases by a factor of 2 when changing focus from infinity to macro. As a result, the aperture appears half as large radially from the perspective of the microlenses, and the images that appear under each microlens span only half as many pixels.

These observations suggest that in designing lenses to be used with a light field camera, the designer should optimize optical image quality for close focus distances rather than infinity. Higher directional resolution at further focal distances can be used to offset slightly worse optical performance in that range.

(A): Center   (B): Middle   (C): Edge

Ray-space of triplet lens, focusing at infinity.

(D): Center   (E): Middle   (F): Edge
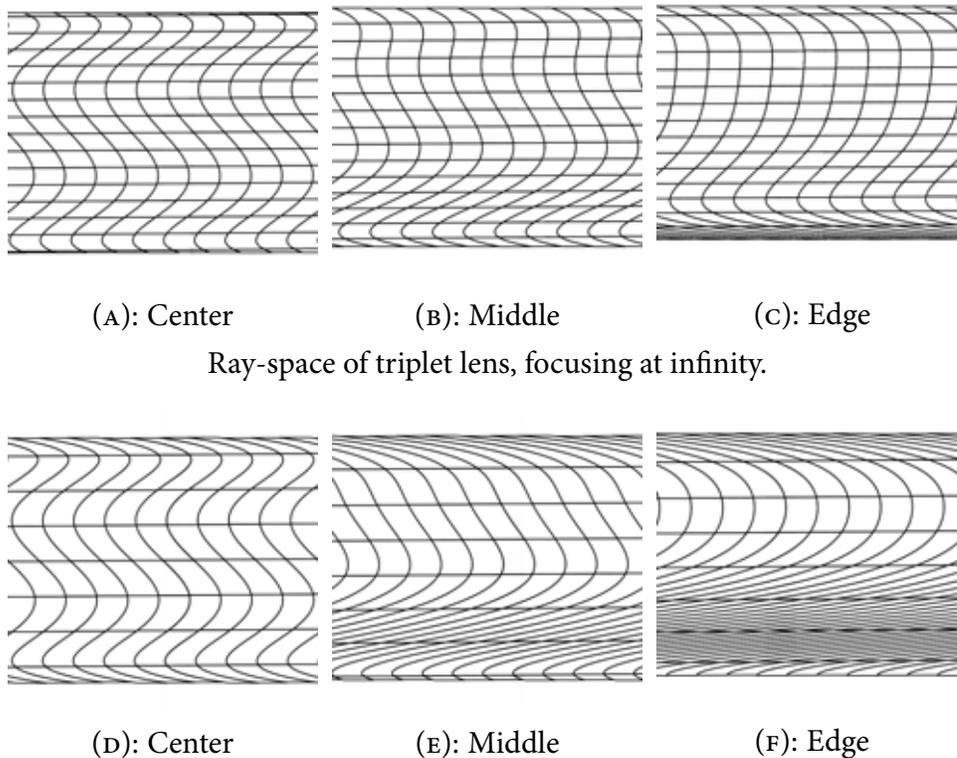
Figure 7.20: Ray-space of triplet lens at macro focal depth.

7.7.3   Correction Performance Across a Database of Lenses

The previous section illustrated how digital correction could be used to improve the performance of one multi-element lens. It should be clear from the ray-traces diagrams and visualizations of the aberrated ray-space that the amount of improvement will depend on the exact formula of the lens and the shape of the distortions in the recorded light field.

To provide some feeling for the performance of digital correction across a range of lenses, this chapter concludes with a a summary of simulated correction performance for a database of 22 lenses. The optical formulas for these lenses were obtained by manually extracting every fixed-focal-length lens in the Zebase [Zemax 2003] database for which the description implied that photographic application was possible.

The initial set of lenses selected from Zemax was modified in two ways. First, the set

was pruned to exclude lenses that contained aspheric elements, because the simulation system does not currently support analytic ray-tracing of such aspheric surfaces. Second, the remaining lenses were linearly scaled in size so that their projected image circles matched the diagonal length of the 35 mm format sensor in our virtual camera. The issue is that the scaling of the original lens formulas was quite arbitrary. For example, many formulas were normalized in size to a nominal focal length of 100 mm, regardless of the effective focal length in 35 mm format. Figure 7.21 lists the final set of lenses used in this experiment, and some of their basic properties.

The database spans an eclectic range of design forms, including triplets, double Gauss variations, telephotos and wide-angle lenses. They also span a fairly wide range of focal lengths and f-numbers. Many of these lens recipes originated in classical optical engineering textbooks, or from expired patents. Nevertheless, many of the designs are quite old, and this may best be thought of as an experiment on variations of some classic design patterns. There are many lens design forms used in diverse imaging applications today, and this database does not attempt to sample the breadth of current art.

In spite of these caveats, the 22-lens database provides useful experimental intuition. The results imply that digital correction will be quite generally applicable at resolutions that are achievable in present or near-term technology.

Figure 7.22 presents the experimental data. Each lens appears on a different row, sorted in order of increasing uncorrected performance. For each lens, effective resolution with unweighted correction is plotted across the row as a function of directional resolution. The directional resolution is $N \times N$, beginning at $N = 1$ (conventional uncorrected), and increasing up to $N = 16$, as shown by the key near the bottom of the figure. Note that the horizontal axis is plotted on a log scale.

The most striking feature of the data is the regular improvement provided with digital correction. Effective resolution rises steadily with directional resolution for all the lenses, indeed as one would expect from the interpretation of the algorithm in ray-space. In other words, the experiment provides evidence that digital correction is a robust and general-purpose technique for improving image quality.

A second important point is that the data show that in many cases the amount of improvement is roughly proportional to the recorded directional resolution. In particular, the

Lens File	Description	Glass elements	Focal length	Aperture	Row*
E-001	Cooke triplet anastigmat	3	135 mm	$f/3.5$	13
E-002	Triplet	3	130 mm	$f/2.0$	1
E-006	Triplet objective USP 2,453,260	3	130 mm	$f/2.7$	5
E-009	Cooke triplet USP 3,176,582	3	52 mm	$f/2.9$	15
F-020	Wide angle USP 3,976,465	10	23 mm	$f/1.8$	12
F-025	Wide-angle objective	6	36 mm	$f/3.0$	17
G-001	Telephoto	5	130 mm	$f/5.8$	18
G-002	SLR telephoto	9	191 mm	$f/2.8$	22
I-014	Photographic objective	5	63 mm	$f/3.4$	20
L-001	Double Gauss	6	63 mm	$f/2.5$	16
L-002	Double Gauss	7	63 mm	$f/1.8$	7
L-008	Photographic lens USP 2,350,035	7	85 mm	$f/1.4$	2
L-009	Photographic objective USP 2,379,392	7	85 mm	$f/1.5$	9
L-010	Large aperture objective USP 2,701,982	8	85 mm	$f/1.0$	4
L-011	Modified Gauss USP 3,743,387	7	58 mm	$f/1.2$	6
L-012	Large aperture photo lens USP 3,938,884	7	58 mm	$f/1.2$	3
L-013	Lrg. aper. photographic USP 4,364,643	7	54 mm	$f/1.2$	8
L-014	Gauss type objective USP 4,390,252	6	85 mm	$f/2.5$	21
L-015	Double gauss objective USP 3,552,829	7	63 mm	$f/1.6$	14
L-020	8-glass double Gauss USP 2,701,982	8	63 mm	$f/1.1$	11
L-033	Momiyama USP 4,364,643	7	54 mm	$f/1.2$	10
L-035	Double Gauss objective	6	63 mm	$f/2.5$	9

* In Figure 7.22.

Figure 7.21: Database of lenses.

experiment indicates that relatively moderate amount of directional data can provide significant improvements in image quality. For example, a directional resolution of 8×8, which is viable in current VLSI technology, produces increases in effective resolution by an average factor of 5.4 across the database, and by up to a factor of 15 (for rows 1 and 7). Taking a step back, this is an important validation of the required directional resolution for useful digital correction. From first principles it was not immediately clear whether correcting for aberrations across the lens aperture would work without extremely high resolution images of the lens aperture. The experiment here shows that it does, although of course more resolution provides greater effective resolution up to the limiting resolution of the microlens array.
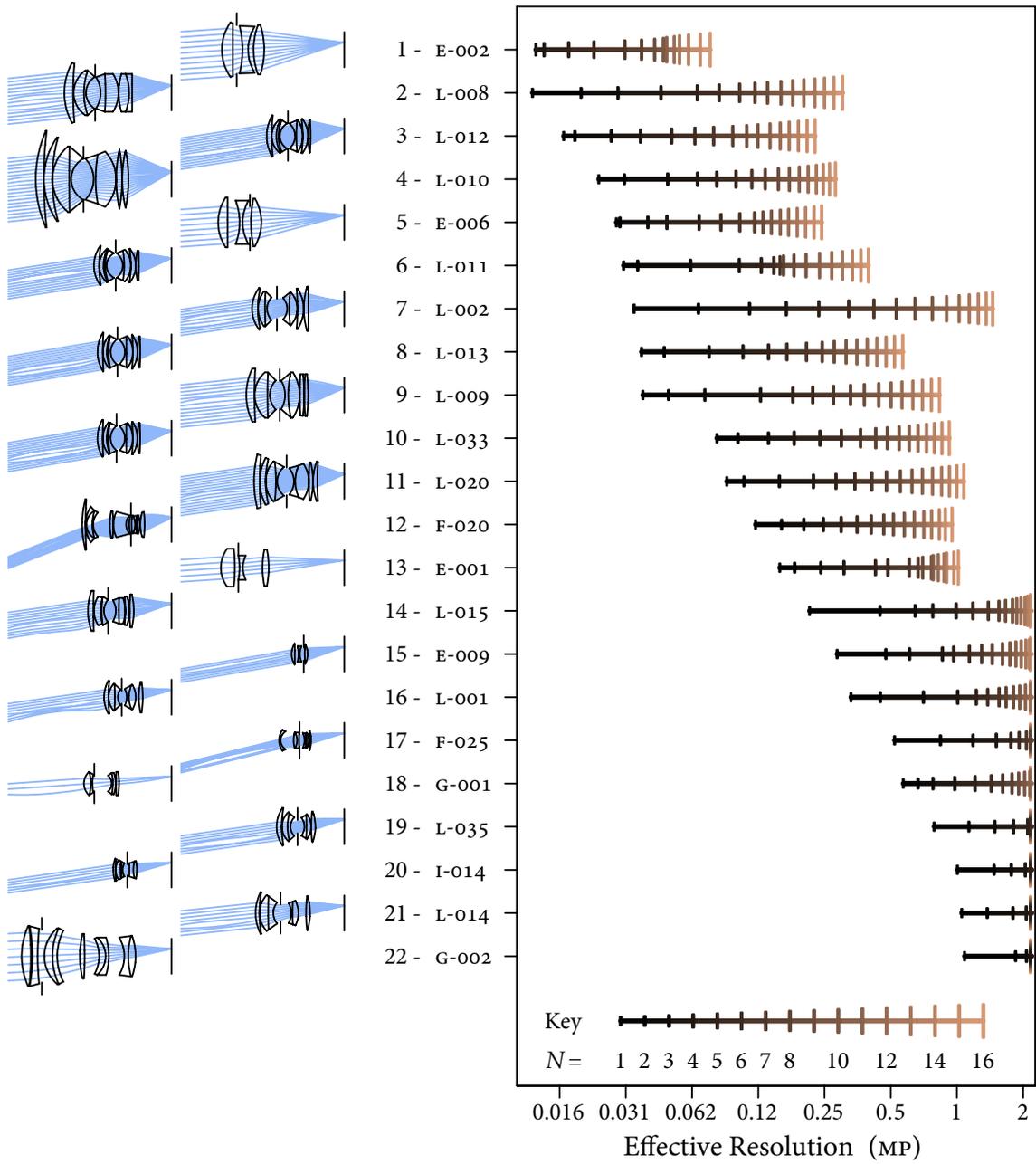
Figure 7.22: Improvement in effective resolution with digital correction.

Another feature of the data is that the effective resolution without digital correction (the left-most data point for each row) is quite low for many lenses. This corresponds to the performance of conventional photographic imaging. For 14 out of the 22 lenses the effective resolution is actually below one quarter of the assumed 2.1 MP output image resolution. Although some lenses are unusually low, and probably reflect a prioritization of maximum light gathering power (aperture size) over image quality, it is not surprising that many lenses would suffer from low effective resolution with their apertures wide-open, as tested here. It is common photographic wisdom that in order to obtain the sharpest images from any lens, one should stop the lens down to half its maximum diameter or less [Adams 1995]. Of course this reduces the light gathering power, requiring longer exposures. The results here re-iterate that digital correction can raise the image quality without having to pay as great a penalty in light gathering power.

In Closing

To exploit digital correction to maximum advantage, it must be considered during the design phase of the lens, while determining the optical requirements. One example was already discussed earlier in the chapter: lenses designed to be used with digital correction should prioritize optical performance at focal planes close to the camera. A different consideration is that digital correction expands the space of possible solutions given a set of image quality requirements. It will allow the use of larger apertures than before, or it may enable simplification of the optical formula (e.g. a reduction in the number of glass elements) required to satisfy the target requirements.

In order to reach the full potential of the technique, further study is needed into methods for joint optimization of the the optical recipe and digital correction software. As one example of the complexity of the design task, rows 5 – 9 in Figure 7.22 illustrate very different growth rates with directional resolution, even though they all start with similar uncorrected effective resolutions. From a qualitative examination of the shape of the distortions in the ray-space, it seem plausible that a major limitation on the rate of improvement are regions of extremely high slope that are localized in u, such as uncorrected spherical aberration. On the one hand, this suggests that unweighted digital correction would work best on lenses

without this characteristic. On the other hand, localized regions of high curvature in the ray space are especially amenable to weighted correction, which can eliminate them without greatly reducing the collecting power.

The concept of weighted correction raises an important question about aberrated light fields: are cells of low weight fundamentally less useful? To answer this question, recall that the discussion of aberrated sub-aperture images in Section 7.3 showed that different light field cells are maximally focused at different depths in the world. What this means is that the weight of a light field cell depends on the refocus depth. Maximum weight will result when refocusing at the cell's plane of maximal focus. The warped ray-space sampling grid of Figure 7.7D provides one way to visualize this process. Recall that refocusing corresponds to straight, parallel projection lines, which will be overlaid on the warped grid. When the slope of the lines match the slope of the grid cells, then the weight will be highest. A different way to visualize the changing weight is in terms of the blue-tinted sub-aperture images of Figure 7.8. In this kind of diagram, the distribution of blue weight shifts as we refocus at different depths. In other words, aberrations do not result in useless light field samples, but rather in samples that extend the range of sharp refocusing at the price of reduced effective light usage at any one refocus depth.

These fundamental non-linearities are what make the study of aberrations and their correction such a challenging subject. Continued study of digital correction will have to address much of the complexity of traditional lens design, and add to it the additional nuances of weighting function design.

8

Conclusion

Through a pioneering career as one of the original photojournalists, Henri Cartier-Bresson (1908 – 2004) inspired a generation of photographers, indeed all of us, to seek out and capture the *Decisive Moment* in our photography.

> The creative act lasts but a brief moment, a lightning instant of give-and-take, just long enough for you to level the camera and to trap the fleeting prey in your little box.

Armed only with his manual Leica, and "no photographs taken with the aid of flashlight, either, if only out of respect for the actual light," Cartier-Bresson made it seem as if capturing the decisive moment were as easy as turning one's head to casually observe perfection.

But most of us do not find it that easy. I love photography, but I am not a great photographer. The research in this dissertation grew out of my frustration at losing many shots to mis-focus. One of the historical lines in photography has been carried by generations of camera engineers. From the original breakthrough by Daguerre in 1839, which was instantly popular in spite of the toxicity of the chemical process, to the development of the hand-camera, derided by even the great Alfred Stieglitz before he recognized its value, to the rise of digital photography in the last ten years, we have seen continuous progress in the photographic tools available to us. But picture-making science is still young, and there are still many problems to be solved.

The main lesson that I have learned through my research is that significant parts of the physical process of making photographs can be executed faithfully in software. In particular, the problems associated with optical focus are not fundamental characteristics of photography. The solution advanced in this dissertation is to record the light field flowing into conventional photographs, and to use the computer to control the final convergence of rays in our images. This new kind of photography means unprecedented capabilities after exposure: refocusing, choosing a new depth of field, and correcting lens aberrations.

Future cameras based on these principles will be physically simpler, capture light more quickly, and provide greater flexibility in finishing photographs. There is a lot of work to be done on re-thinking existing camera components in light of these new capabilities. The last chapter discussed how lens design will change to exploit digital correction of aberrations. With larger-aperture lenses, it may be possible to use a weaker flash system or do away with it in certain scenarios. Similarly, the design of the auto-focus system will change in light of digital refocusing and the shift in optimal lens focus required by selectable refocusing power. Perhaps the greatest upheaval will be in the design of the photosensor. We need to maximize resolution with good noise characteristics – not an easy task. And the electronics will need to read it out at reasonable rates and store it compactly. This is the main price behind this new kind of photography: recording and processing a lot more data. Fortunately, these kinds of challenges map very well to the exponential growth in our capabilities for electronic storage and computing power.

In photography, one of the most promising areas for future work is developing better processing tools for photo-finishing. In this dissertation, I chose methods that stayed close to physical models of image formation in real cameras. Future algorithms should boldly pursue non-physical metaphors, and should actively interpret the scene to compute a final image with the best overall composition. The quintessential example would be automatic refocusing of the people detected in the picture while softening focus on the background, as I tried to suggest in the treatment of the two-person portrait in Figure 4.12. Such automatic photo-finishing would be a boon for casual photographers, but it is inappropriate for the professional photographer or serious enthusiast. Experts like these need interactive tools that give them artistic control. A simple idea in this vein is a virtual brush that the user would "paint" over the photograph on the computer to push the local focus closer or further

– analogous to dodging and burning in the old darkroom. Having the lighting at every pixel in a photograph will enable all kinds of new computer graphics like this.

The ideas in this dissertation have already begun to make an impact in scientific imaging. A light field camera attached to a microscope enables 3D reconstruction of the specimen from a single photographic exposure [Levoy et al. 2006], because it collects rays passing through the transparent specimen at different angles. Telescopes present another interesting opportunity. Would it be possible to discard the real-time deformable mirror used in modern adaptive-optics telescopes [Tyson 1991], instead recording light field video and correcting for atmospheric aberrations in software? In general, every imaging device that uses optics in front of a sensor may benefit from recording and processing ray directional information according to the principles described in this dissertation.

This is a very exciting time to be working in digital imaging. We have two powerful evolutionary forces acting: an overabundance of resolution, and processing power in close proximity. I hope I have convinced you that cameras as we know them today are just an evolutionary step in where we are going, and I feel that we are on the verge of an explosion in new kinds of cameras and computational imaging.

But thankfully, some things are sure to stay the same. Photography will celebrate its 167th birthday this year, and photographs are much older even than that – we had them floating on our retinas long before we could fix them on metal or paper. To me, one of the chief joys in light field photography is that it feels like photography – it very much *is* photography as we know it. Refocusable images are compelling exactly because they look like the images we've always seen, except that they retain a little more life by saving the power to focus for later. I find that this new kind of photography makes taking pictures that much more enjoyable, and I hope you will too. I look forward to the day when I can stand in the tall grass and learn from fellow light field photographers shooting in the field.

A

Proofs

A.1 Generalized Fourier Slice Theorem

GENERALIZED FOURIER SLICE THEOREM

$$\mathcal{F}^M \circ \mathcal{I}_M^N \circ \mathcal{B} = \mathcal{S}_M^N \circ \left(\mathcal{B}^{-T} / \left| \mathcal{B}^{-T} \right| \right) \circ \mathcal{F}^N.$$

Proof. The following proof is inspired by one common approach to proving the classical 2D version of the theorem. The first step is to note that

$$\mathcal{F}^M \circ \mathcal{I}_M^N = \mathcal{S}_M^N \circ \mathcal{F}^N, \tag{A.1}$$

because substitution of the basic definitions shows that for an arbitrary function, f, both $\left(\mathcal{F}^M \circ \mathcal{I}_M^N \right) [f] (u_1, \ldots, u_M)$ and $\left(\mathcal{S}_M^N \circ \mathcal{F}^N \right) [f] (u_1, \ldots, u_M)$ are equal to

$$\int f(x_1, \ldots, x_N) \exp\left(-2\pi i \left(x_1 u_1 + \cdots + x_M u_M \right) \right) dx_1 \ldots dx_N.$$

The next step is to observe that if basis change operators commute with Fourier transforms

via $\mathcal{F}^N \circ \mathcal{B} = (\mathcal{B}^{-T} / |\mathcal{B}^{-T}|) \circ \mathcal{F}^N$, then the proof of the theorem would be complete because for every function f we would have

$$(\mathcal{F}^M \circ \mathcal{I}_M^N \circ \mathcal{B}) [f] = (\mathcal{S}_M^N \circ \mathcal{F}^N \circ \mathcal{B}) [f]$$

by Equation A.1, and the commutativity relation would give us the final theorem:

$$(\mathcal{F}^M \circ \mathcal{I}_M^N \circ \mathcal{B}) [f] = \left(\mathcal{S}_M^N \circ \left(\mathcal{B}^{-T} / \left|\mathcal{B}^{-T}\right|\right) \circ \mathcal{F}^N\right) [f]. \tag{A.2}$$

Thus, the final step is to show that $\mathcal{F}^N \circ \mathcal{B} = (\mathcal{B}^{-T} / |\mathcal{B}^{-T}|) \circ \mathcal{F}^N$. Directly substituting the operator definitions establishes these two equations:

$$(\mathcal{F}^N \circ \mathcal{B}) [f] (\mathbf{u}) = \int f(\mathcal{B}^{-1}\mathbf{x}) \exp\left(-2\pi i \left(\mathbf{x}^T \mathbf{u}\right)\right) d\mathbf{x}; \tag{A.3}$$

$$\left(\frac{\mathcal{B}^{-T}}{|\mathcal{B}^{-T}|} \circ \mathcal{F}^N\right) [f] (\mathbf{u}) = \frac{1}{|\mathcal{B}^{-T}|} \int f(\mathbf{x}') \exp\left(-2\pi i \left((\mathbf{x}')^T (\mathcal{B}^T \mathbf{u})\right)\right) d\mathbf{x}'. \tag{A.4}$$

In these equations, $\mathbf{x}$ and $\mathbf{u}$ are N-dimensional column vectors (so $\mathbf{x}^T\mathbf{u}$ is the dot product of $\mathbf{x}$ and $\mathbf{u}$), and the integral is taken over all of N-dimensional space.

Let us now apply the change of variables $\mathbf{x} = \mathcal{B}\mathbf{x}'$ to Equation A.4. The first substitution is $\mathbf{x}' = \mathcal{B}^{-1}\mathbf{x}$. Furthermore, $d\mathbf{x} = |\mathcal{B}| \, d\mathbf{x}'$. However, since $|\mathcal{B}| = 1/ |\mathcal{B}^{-1}| = 1/ |\mathcal{B}^{-T}|$ by basic properties of determinants, $d\mathbf{x} = 1/ |\mathcal{B}^{-T}| \, d\mathbf{x}'$, which is the second substitution. Making these substitutions,

$$\left(\frac{\mathcal{B}^{-T}}{|\mathcal{B}^{-T}|} \circ \mathcal{F}^N\right) [f] (\mathbf{u}) = \int f(\mathcal{B}^{-1}\mathbf{x}) \exp\left(-2\pi i \left(\left(\mathcal{B}^{-1}\mathbf{x}\right)^T \left(\mathcal{B}^T \mathbf{u}\right)\right)\right) d\mathbf{x}$$

$$= \int f(\mathcal{B}^{-1}\mathbf{x}) \exp\left(-2\pi i \left(\left(\mathbf{x}^T \mathcal{B}^{-T}\right) \left(\mathcal{B}^T \mathbf{u}\right)\right)\right) d\mathbf{x}$$

$$= \int f(\mathcal{B}^{-1}\mathbf{x}) \exp\left(-2\pi i \left(\mathbf{x}^T \mathbf{u}\right)\right) d\mathbf{x}, \tag{A.5}$$

where the second line relies on the linear algebra rule for transposing matrix products. Equations A.3 and A.5 show that $\mathcal{F}^N \circ \mathcal{B} = (\mathcal{B}^{-T} / |\mathcal{B}^{-T}|) \circ \mathcal{F}^N$, completing the proof. □

A.2 Filtered Light Field Imaging Theorem

FILTERED LIGHT FIELD IMAGING THEOREM

$$\mathcal{P}_F \circ \mathcal{C}_h^4 = \mathcal{C}_{\mathcal{P}_F[h]}^2 \circ \mathcal{P}_F.$$

To prove the theorem, let us first establish a lemma involving the closely-related modulation operator:

Modulation $\mathcal{M}_\beta^N$ is an N-dimensional modulation operator, such that $\mathcal{M}_\beta^N [F] (\mathbf{x}) = F(\mathbf{x}) \cdot \beta(\mathbf{x})$ where $\mathbf{x}$ is an N-dimensional vector coordinate.

Lemma. *Multiplying an input 4D function by another one, h, and transforming the result by $\mathfrak{P}_F$, the Fourier photography operator, is equivalent to transforming both functions by $\mathfrak{P}_F$ and then multiplying the resulting 2D functions. In operators,*

$$\mathfrak{P}_F \circ \mathcal{M}_h^4 = \mathcal{M}_{\mathfrak{P}_F[h]}^2 \circ \mathfrak{P}_F \tag{A.6}$$

Algebraic verification of the lemma is direct given the basic definitions, and is omitted here. On an intuitive level, however, the lemma makes sense because $\mathfrak{P}_\alpha$ is a slicing operator: multiplying two functions and then slicing them is the same as slicing each of them and multiplying the resulting functions.

Proof of theorem. The first step is to translate the classical Fourier Convolution Theorem (see, for example, Bracewell [1986]) into useful operator identities. The Convolution Theorem states that a multiplication in the spatial domain is equivalent to convolution in the Fourier domain, and vice versa. As a result,

$$\mathcal{F}^N \circ \mathcal{C}_h^N = \mathcal{M}_{\mathcal{F}^N[h]}^N \circ \mathcal{F}^N \tag{A.7}$$

$$\text{and} \quad \mathcal{F}^N \circ \mathcal{M}_h^N = \mathcal{C}_{\mathcal{F}^N[h]}^N \circ \mathcal{F}^N. \tag{A.8}$$

Note that these equations also hold for negative N, since the Convolution Theorem also applies to the inverse Fourier transform.

With these facts and the lemma in hand, the proof of the theorem proceeds swiftly:

$$
\begin{aligned}
\mathcal{P}_F \circ \mathcal{C}_h^4 &= \mathcal{F}^{-2} \circ \mathfrak{P}_F \circ \mathcal{F}^4 \circ \mathcal{C}_h^4 \\
&= \mathcal{F}^{-2} \circ \mathfrak{P}_F \circ \mathcal{M}_{\mathcal{F}^4[h]}^4 \circ \mathcal{F}^4 \\
&= \mathcal{F}^{-2} \circ \mathcal{M}_{(\mathfrak{P}_F \circ \mathcal{F}^4)[h]}^2 \circ \mathfrak{P}_F \circ \mathcal{F}^4 \\
&= \mathcal{C}_{(\mathcal{F}^{-2} \circ \mathfrak{P}_F \circ \mathcal{F}^4)[h]}^2 \circ \mathcal{F}^{-2} \circ \mathfrak{P}_F \circ \mathcal{F}^4 \\
&= \mathcal{C}_{\mathcal{P}_F[h]}^2 \circ \mathcal{P}_F,
\end{aligned}
$$

where we apply the Fourier Slice Photography Theorem (Equation 5.7) to derive the first and last lines, the Convolution Theorem (Equations A.7 and A.8) for the second and fourth lines, and the lemma (Equation A.6) for the third line. □

A.3   Photograph of a Four-Dimensional Sinc Light Field

This appendix derives Equation 5.9, which states that a photograph from a 4D sinc light field is a 2D sinc function. The first step is to apply the Fourier Slice Photographic Imaging Theorem to move the derivation into the Fourier domain.

$$
\begin{aligned}
&\mathcal{P}_\alpha \left[1/(\Delta x \Delta u)^2 \cdot \mathrm{sinc}(x/\Delta x, y/\Delta x, u/\Delta u, v/\Delta u) \right] \\
=\,&1/(\Delta x \Delta u)^2 \cdot (\mathcal{F}^{-2} \circ \mathfrak{P}_\alpha \circ \mathcal{F}^4) \left[\mathrm{sinc}(x/\Delta x, y/\Delta x, u/\Delta u, v/\Delta u) \right] \\
=\,&(\mathcal{F}^{-2} \circ \mathfrak{P}_\alpha) \left[\sqcap(k_x \Delta x, k_y \Delta x, k_u \Delta u, k_v \Delta u) \right]. \quad\quad\quad\quad (\text{A.9})
\end{aligned}
$$

Now we apply the definition for the Fourier photography operator $\mathfrak{P}_\alpha$ (Equation 5.5), to arrive at

$$
\begin{aligned}
&\mathcal{P}_\alpha \left[1/(\Delta x \Delta u)^2 \cdot \mathrm{sinc}(x/\Delta x, y/\Delta x, u/\Delta u, v/\Delta u) \right] \quad\quad\quad\quad (\text{A.10}) \\
&= \mathcal{F}^{-2} \left[\sqcap(\alpha k_x \Delta x, \, \alpha k_y \Delta x, \, (1-\alpha)k_x \Delta u, \, (1-\alpha)k_y \Delta u) \right].
\end{aligned}
$$

Note that the 4D rect function now depends only on k_x and k_y, not k_u or k_v. Since the product of two dilated rect functions is equal to the smaller rect function,

$$\mathcal{P}_\alpha \left[1/(\Delta x \Delta u)^2 \cdot \text{sinc}(x/\Delta x, y/\Delta x, u/\Delta u, v/\Delta u) \right]$$
$$= \mathcal{F}^{-2} \left[\sqcap(k_x D_x, k_y D_x) \right] \tag{A.11}$$
$$\text{where } D_x = \max(\alpha \Delta x, |1 - \alpha| \Delta u). \tag{A.12}$$

Applying the inverse 2D Fourier transform completes the proof:

$$\mathcal{P}_\alpha \left[1/(\Delta x \Delta u)^2 \cdot \text{sinc}(x/\Delta x, y/\Delta x, u/\Delta u, v/\Delta u) \right]$$
$$= 1/D_x^2 \cdot \text{sinc}(k_x/D_x, k_y/D_x). \tag{A.13}$$

Bibliography

ADAMS, A. 1995. *The Camera*. Bulfinch Press.

ADELSON, E. H., AND BERGEN, J. R. 1991. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing, edited by Michael S. Landy and J. Anthony Movshon*. Cambridge, Mass.: MIT Press.

ADELSON, E. H., AND WANG, J. Y. A. 1992. Single lens stereo with a plenoptic camera. IEEE *Transactions on Pattern Analysis and Machine Intelligence 14*, 2 (Feb), 99–106.

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. ACM *Transactions on Graphics (Proceedings of* SIGGRAPH *2004) 23*, 3, 292–300.

ARAI, J., HOSHINO, H., OKUI, M., AND OKANO, F. 2003. Effects of focusing on the resolution characteristics of integral photography. *J. Opt. Soc. Am. A 20*, 6 (June 2003), 992–1004.

ASKEY, P., 2006. Digital cameras timeline: 2006. Digital Photography Review, http://www.dpreview.com/reviews/timeline.asp?start=2006.

BABCOCK, H. W. 1953. The possibility of compensating astronomical seeing. *Publications of the Astronomical Society of the Pacific 65*, 386, 229–.

BARLOW, H. B. 1952. The size of ommatidia in apposition eyes. *Journal of Experimental Biology 29*, 4, 667–674.

BARNARD, K., CARDEI, V., AND FUNT, B. 2002. A comparison of computational color constancy algorithms – part I: Methodology and experiments with synthesized data. IEEE *Transactions on Image Processing 11*, 9, 972–983.

BOLLES, R., BAKER, H., AND MARIMONT, D. 1987. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision 1*, 1, 7–55.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. IEEE *Transactions on Pattern Analysis and Machine Intelligence 23*, 11, 1222–1239.

BRACEWELL, R. N., CHANG, K.-Y., JHA, A. K., AND WANG, Y. H. 1993. Affine theorem for two-dimensional fourier transform. *Electronics Letters 29*, 304–309.

BRACEWELL, R. N. 1956. Strip integration in radio astronomy. *Aust. J. Phys. 9*, 198–217.

BRACEWELL, R. N. 1986. *The Fourier Transform and Its Applications, 2nd Edition Revised*. New York: McGraw-Hill.

BRIXNER, B. 1963. Automatic lens design for nonexperts. *Applied Optics 2*, 12, 1281 –1286.

BUEHLER, C., BOSSE, M., MCMILLAN, L., GORTLER, S., AND COHEN, M. 2001. Unstructured lumigraph rendering. In SIGGRAPH '01: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 425–432.

CAMAHORT, E., LERIOS, A., AND FUSSELL, D. 1998. Uniformly sampled light fields. In *Proceedings of Eurographics Rendering Workshop 1998*, 117–130.

CAMAHORT, E. 2001. 4D *Light Field Modeling and Rendering*. PhD thesis, The University of Texas at Austin.

CATRYSSE, P. B., AND WANDELL, B. A. 2002. Optical efficiency of image sensor pixels. *J. Opt. Soc. Am. A 19*, 8, 1610–1620.

CHAI, J., TONG, X., CHAN, S.-C., AND SHUM, H.-Y. 2000. Plenoptic sampling. In SIGGRAPH 00, 307–318.

CHEN, S. E., AND WILLIAMS, L. 1993. View interpolation for image synthesis. *Computer Graphics 27*, 279–288.

CHEN, W., BOUGUET, J., CHU, M., AND GRZESZCZUK, R. 2002. Light field mapping: Efficient representation and hardware rendering of surface light fields. In SIGGRAPH 02, 447–456.

COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed ray tracing. In SIGGRAPH '84, 137–145.

DALY, D. 2001. *Microlens Arrays*. New York: Taylor & Francis Inc.

DAVIES, N., MCCORMICK, M., AND YANG, L. 1988. 3D imaging systems: a new development. *Appl. Opt.*, 27, 4520–4528.

DEANS, S. R. 1983. *The Radon Transform and Some of Its Applications*. New York: Wiley-Interscience.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In SIGGRAPH 97, 369–378.

DICKERSON, J., AND LEPP, G. 1995. *Canon Lenses*. Sterling Pub. Co.Inc.

DOWSKI, E. R., AND JOHNSON, G. E. 1999. Wavefront coding: A modern method of achieving high performance and/or low cost imaging systems. In SPIE *Proceedings*, vol. 3779.

DUPARRÉ, J., DANNBERG, P., SCHREIBER, P., AND BRÄUER, A. 2004. Micro-optically fabricated artificial apposition compound eye. In *Electronic Imaging – Science and Technology, Proc.* SPIE *5301*.

DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. 2005. A frequency analysis of light transport. ACM *Transactions on Graphics (Proceedings of* SIGGRAPH *2005) 24*, 3, 1115–1126.

ELTOUKHY, H., AND KAVUSI, S. 2003. A computationally efficient algorithm for multi-focus image reconstruction. In *Proceedings of* SPIE *Electronic Imaging*.

FORSYTH, D. A., AND PONCE, J. 2002. *Computer Vision: A Modern Approach*. Upper Saddle River, N.J.: Prentice Hall.

GEORGIEV, T., ZHENG, C., CURLESS, B., SALESIN, D., NAYAR, S., AND INTWALA, C. 2006. Spatio-angular resolution tradeoff in integral photography. To appear in *Proceedings of the Eurographics Symposium on Rendering 2006*.

GERRARD, A., AND BURCH, J. M. 1975. *Introduction to Matrix Methods in Optics*. New York: Wiley.

GERSHUN, A. 1939. The light field. Moscow, 1936. *Journal of Mathematics and Physics* XVIII, 51–151. Translated by P. Moon and G. Timoshenko.

GOODMAN, J. W. 1985. *Statistical Optics*. New York: Wiley-Interscience.

GOODMAN, J. W. 1996. *Introduction to Fourier Optics, 2nd edition*. New York: McGraw-Hill.

GORDON, N. T., JONES, C. L., AND PURDY, D. R. 1991. Application of microlenses to infrared detector arrays. *Infrared Physics 31*, 6, 599–604.

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The Lumigraph. In SIGGRAPH *96*, 43–54.

HAEBERLI, P., 1994. A multifocus method for controlling depth of field. Graphics Obscura, URL: http://www.sgi.com/grafica/depth.

HALLE, M., 1994. Holographic stereograms as discrete imaging systems.

HARDY, J. W., LEFEBVRE, J. E., AND KOLIOPOULOS, C. K. 1977. Real-time atmospheric compensation. *J. Opt. Soc. Am. 67*, 3 (March 1997), 360–369.

HARO, L. D. S. 2000. *Wavefront Sensing in the Human Eye with a Shack-Hartmann Sensor*. PhD thesis, University of London.

HECHT, E. 2002. *Optics, 4th edition*. Reading, Mass.: Addison-Wesley.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In SIGGRAPH *2000*, 297–306.

ISHIHARA, Y., AND TANIGAKI, K. 1983. A high photosensitivity IL-CCD image sensor with monolithic resin lens array. In *Proceedings of* IEEE IEDM, 497–500.

Ives, H. E. 1930. Parallax panoramagrams made with a large diameter lens. *J. Opt. Soc. Amer. 20*, 332–342.

Ives, H. E. 1931. Optical properties of a Lippmann lenticulated sheet. *J. Opt. Soc. Amer. 21*, 171–176.

Jackson, J. I., Meyer, C. H., Nishimura, D. G., and Macovski, A. 1991. Selection of convolution function for Fourier inversion using gridding. IEEE *Transactions on Medical Imaging 10*, 3 (September), 473–478.

Javidi, B., and Okano, F., Eds. 2002. *Three-Dimensional Television, Video and Display Technologies.* New York: Springer-Verlag.

Johnson, R. B. 1992. A historical perspective on understanding optical aberrations. In *Lens design : proceedings of a conference held 22-23 January 1992, Los Angeles, California, Warren J. Smith, editor*, Bellingham, Wash., USA: SPIE Optical Engineering Press, 18–29.

Kajiya, J. 1986. The rendering equation. In SIGGRAPH *86*, 143–150.

Keelan, B. W. 2002. *Handbook of Image Quality : Characterization and Prediction.* New York: Marcel Dekker.

Kingslake, R. 1934. Development of the photographic objective. *J. Opt. Soc. Am. 24*, 73–.

Kingslake, R. 1978. *Lens Design Fundamentals.* New York: Academic Press.

Kingslake, R. 1989. *A History of the Photographic Lens.* Boston: Academic Press.

Kirschfeld, K. 1976. The resolution of lens and compound eyes. In *Neural Principles in Vision edited by F. Zettler and R. Weiler*. New York: Springer, 354–370.

Kolb, C., Mitchell, D., and Hanrahan, P. 1995. A realistic camera model for computer graphics. In SIGGRAPH *1995*, 317–324.

Krotkov, E. 1987. Focusing. *International Journal of Computer Vision 1*, 3, 223–237.

KUBOTA, A., AND AIZAWA, K. 2005. Reconstructing arbitrarily focused images from two differently focused images using linear filters. IEEE *Transactions on Image Processing 14*, 11 (Nov 1995), 1848–1859.

LAND, M. F., AND NILSSON, D.-E. 2002. *Animal Eyes*. New York: Oxford University Press.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In SIGGRAPH 96, 31–42.

LEVOY, M., CHEN, B., VAISH, V., HOROWITZ, M., MCDOWALL, I., AND BOLAS, M. 2004. Synthetic aperture confocal imaging. ACM *Transactions on Graphics (Proceedings of* SIGGRAPH *2004) 23*, 3, 822–831.

LEVOY, M., NG, R., ADAMS, A., FOOTER, M., AND HOROWITZ, M. 2006. Light field microscopy.

LEVOY, M. 1992. Volume rendering using the fourier projection-slice theorem. In *Proceedings of Graphics Interface '92*, 61–69.

LIANG, Z.-P., AND MUNSON, D. C. 1997. Partial Radon transforms. IEEE *Transactions on Image Processing 6*, 10 (Oct 1997), 1467–1469.

LIANG, J., GRIMM, B., GOELZ, S., AND BILLE, J. F. 1994. Objective measurement of wave aberrations of the human eye with use of a hartmann-shack wave-front sensor. *J. Opt. Soc. Am. A 11*, 7 (July 1994), 1949–1957.

LIPPMANN, G. 1908. La photographie intégrale. *Comptes-Rendus, Académie des Sciences 146*, 446–551.

MACOVSKI, A. 1983. *Medical Imaging Systems*. Englewood Cliffs, N.J.: Prentice Hall.

MAEDA, P., CATRYSSE, P., AND WANDELL, B. 2005. Integrating lens design with digital camera simulation. In *Proceedings of the* SPIE *Electronic Imaging '2005 Conference, Santa Clara, CA*.

MALZBENDER, T. 1993. Fourier volume rendering. ACM *Transactions on Graphics 12*, 3 (July), 233–250.

MANN, A., Ed. 1993. *Selected Papers on Zoom Lenses*, vol. MS85. SPIE Optical Engineering Press.

MATUSIK, W., PFISTER, H., NGAN, A., BEARDSLEY, P., ZIEGLER, R., AND MCMILLAN, L. 2002. Image-based 3D photography using opacity hulls. In *Proceedings of* SIGGRAPH *2002*, 427–437.

MAXWELL, J. C. 1858. On the general laws of optical instruments. *The Quarterly Journal of Pure and Applied Mathematics 2*, 233–246.

MCMILLAN, L., AND BISHOP, G. 1995. Plenoptic modeling: an image-based rendering system. In *Proceedings of* SIGGRAPH *1995*, 39–46.

MICRON, 2005. Micron Technology, Inc. demonstrates leading imaging technology with industry's first 1.7-micron pixel CMOS sensor. http://www.micron.com/news/product/2005-05-16_first_1-7_pixel-cmos_sensor.html.

MILLER, G., RUBIN, S., AND PONCELON, D. 1998. Lazy decompression of surface light fields for precomputed global illumination. In *Proceedings of Eurographics Rendering Workshop 1998*, 281–292.

MITCHELL, D. P., AND NETRAVALI, A. N. 1998. Reconstruction filters in computer graphics. In SIGGRAPH *98*, 221–228.

MOON, P., AND SPENCER, D. E. 1981. *The Photic Field*. Cambridge, Mass.: MIT Press.

NAEMURA, T., YOSHIDA, T., AND HARASHIMA, H. 2001. 3-D computer graphics based on integral photography. *Optics Express 8*, 2, 255–262.

NAYAR, S. K., AND MITSUNAGA, T. 2000. High dynamic range imaging: Spatially varying pixel exposures. In IEEE *Conference on Computer Vision and Pattern Recognition (*CVPR*)*, vol. 1, 472–479.

NAYAR, S. K., WATANABE, M., AND NOGUCHI, M. 1995. Real-Time Focus Range Sensor. In IEEE *International Conference on Computer Vision (*ICCV*)*, 995–1001.

NEWHALL, B. 1976. *The Daguerreotype in America, 3rd rev. ed.* New York: Dover.

NG, R., LEVOY, M., BRÉDIF, M., DUVAL, G., HOROWITZ, M., AND HANRAHAN, P. 2005. Light field photography with a hand-held plenoptic camera. Tech. Rep. CTSR 2005-02, Stanford University Computer Science.

NG, R. 2005. Fourier slice photography. ACM *Transactions on Graphics (Proceedings of* SIG-GRAPH *2005) 24*, 3, 735–744.

NISHIMURA, D. G. 1996. *Principles of Magnetic Resonance Imaging.* Department of Electrical Engineering, Stanford University.

OGATA, S., ISHIDA, J., AND SASANO, T. 1994. Optical sensor array in an artifical compound eye. *Optical Engineering 33*, 11, 3649–3655.

OGDEN, J. M., ADELSON, E. H., BERGEN, J. R., AND BURT, P. J. 1985. Pyramid-based computer graphics. *RCA Engineer 30*, 5, 4–15.

OKANO, F., ARAI, J., HOSHINO, H., AND YUYAMA, I. 1999. Three-dimensional video system based on integral photography. *Optical Engineering 38*, 6 (June 1999), 1072–1077.

OKOSHI, T. 1976. *Three-Dimensional Imaging Techniques.* New York: Academic Press.

PARK, S. K., SCHOWENGERDT, R., AND KACZYNSKI, M.-A. 1984. Modulation-transfer-function analysis for sampled image systems. *Applied Optics 23*, 15.

PENTLAND, A., SCHEROCK, S., DARRELL, T., AND GIROD, B. 1994. Simple range cameras based on focal error. *J. Opt. Soc. Am. A 11*, 11 (Nov), 2925–.

PENTLAND, A. P. 1987. A new sense for depth of field. IEEE *Transactions on Pattern Analysis and Machine Intelligence 9*, 4, 523–531.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. ACM *Transactions on Graphics (Proceedings of* SIGGRAPH *2003) 22*, 3, 313–318.

PHARR, M., AND HUMPHREYS, G. 2004. *Physically Based Rendering: From Theory to Implementation.* Morgan Kaufmann.

PLATT, B., AND SHACK, R. 2001. History and principles of Shack-Hartmann wavefront sensing. *Journal of Refractive Surgery*, 17 (Sept/Oct 2001).

POTMESIL, M., AND CHAKRAVARTY, I. 1981. A lens and aperture camera model for synthetic image generation. In SIGGRAPH '81: *Proceedings of the 8th annual conference on Computer graphics and interactive techniques*, 297–305.

PREISENDORFER, R. W. 1965. *Radiative Transfer on Discrete Spaces*. Oxford: Pergamon Press.

RAMANATH, R., SNYDER, W. E., BILBRO, G. L., AND SANDER, W. A. 2002. Demosaicking methods for Bayer color arrays. *Journal of Electronic Imaging 11*, 3, 306–315.

ROORDA, A., AND WILLIAMS, D. R. 1999. The arrangement of the three cone classes in the living human eye. *Nature 397*, 10 (11 Feb 1999), 520–522.

SHACK, R. V., AND PLATT, B. C. 1971. Production and use of a lenticular hartmann screen (abstract). *J. Opt. Soc. Am. 61*, 656.

SHADE, J., GORTLER, S., HE, L.-W., AND SZELISKI, R. 1998. Layered depth images. In SIGGRAPH '98: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 231–242.

SHUM, H.-Y., AND KANG, S. B. 2000. A review of image-based rendering techniques. In IEEE/SPIE *Visual Communications and Image Processing (VCIP) 2000*, 2–13.

SMITH, W. J. 2000. *Modern Optical Engineering, 3rd edition*. New York: McGraw-Hill.

SMITH, W. J. 2005. *Modern Lens Design, 2nd edition*. New York: McGraw-Hill.

STEWART, J., YU, J., GORTLER, S. J., AND MCMILLAN, L. 2003. A new reconstruction filter for undersampled light fields. In *Proceedings of the Eurographics Symposium on Rendering 2003*, 150 – 156.

STROEBEL, L., COMPTON, J., CURRENT, I., AND ZAKIA, R. 1986. *Photographic Materials and Processes*. Boston: Focal Press.

SUBBARAO, M., WEI, T.-C., AND SURYA, G. 1995. Focused image recovery from two defocused images recorded with different camera settings. IEEE *Transactions on Image Processing 4*, 11 (Dec 1995), 1613–1628.

SUBBARAO, M. 1988. Parallel depth recovery by changing camera parameters. In IEEE *International Conference on Computer Vision (ICCV)*, 149–156.

TAKAHASHI, K., KUBOTA, A., AND NAEMURA, T. 2003. All in-focus view synthesis from under-sampled light fields. In *Proc. VRSJ Intl. Conf. on Artificial Reality and Telexistence (ICAT)*.

TANIDA, J., KUMAGAI, T., YAMADA, K., MIYATAKE, S., ISHIDA, K., MORIMOTO, T., KONDOU, N., MIYAZAKI, D., AND ICHIOKA, Y. 2001. Thin observation module by bound optics (TOMBO): concept and experimental verification. *Applied Optics 40*, 11 (April), 1806–1813.

TANIDA, J., SHOGENJI, R., KITAMURA, Y., YAMADA, K., MIYAMOTO, M., AND MIYATAKE, S. 2003. Color imaging with an integrated compound imaging system. *Optics Express 11*, 18 (September 8), 2109–2117.

TRONNIER, E., AND EGGERT, J., 1960. Three lens photographic objective. United States Patent 3,176,582.

TYSON, R. K. 1991. *Principles of Adaptive Optics*. Boston: Academic Press.

VAISH, V., WILBURN, B., JOSHI, N., AND LEVOY, M. 2004. Using plane + parallax for calibrating dense camera arrays. In *Proceedings of* CVPR, vol. 1, 2–9.

VAISH, V., GARG, G., TALVALA, E.-V., ANTUNEZ, E., WILBURN, B., HOROWITZ, M., AND LEVOY, M. 2005. Synthetic aperture focusing using a shear-warp factorization of the viewing transform. In *Workshop on Advanced 3D Imaging for Safety and Security (in conjunction with CVPR 2005)*.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., LEVOY, M., AND HOROWITZ, M. 2005. High performance imaging using large camera arrays. 765–776. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2005).

WILLIAMS, C. S. 1915. *Introduction to the Optical Transfer Function.* New York: Wiley.

WOOD, D., AZUMA, D., ALDINGER, K., CURLESS, B., DUCHAMP, T., SALESIN, D., AND STUET-
ZLE, W. 2000. Surface light fields for 3D photography. In SIGGRAPH *00,* 287–296.

YAMADA, T., IKEDA, K., KIM, Y.-G., WAKOH, H., TOMA, T., SAKAMOTO, T., OGAWA, K.,
OKAMOTO, E., MASUKANE, K., ODA, K., AND INUIYA, M. 2000. A progressive scan CCD
image sensor for DSC applications. IEEE *Journal of Solid-State Circuits 35,* 12, 2044–2054.

YAMAMOTO, T., AND NAEMURA, T. 2004. Real-time capturing and interactive synthesis of 3D
scenes using integral photography. In *Stereoscopic Displays and Virtual Reality Systems* XI,
Proceedings of SPIE, 155–166.

YANG, J., EVERETT, M., BUEHLER, C., AND MCMILLAN, L. 2002. A realtime distributed light
field camera. In *Proceedings of Eurographics Workshop on Rendering 2002,* 77–86.

ZEMAX. 2003. *Zebase, Optical Design Database, Version 5.0.* San Diego: Zemax Development
Corp.