RANGEFINDING WITH A PLENOPTIC CAMERA

THESIS

Robert A. Raynor, 2nd Lieutenant, USAF

AFIT-ENP-14-M-29

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENP-14-M-29

RANGEFINDING WITH A PLENOPTIC CAMERA

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Robert A. Raynor, BS

2nd Lieutenant, USAF

March 2014

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

AFIT-ENP-14-M-29

RANGEFINDING WITH A PLENOPTIC CAMERA

Robert A. Raynor, BS
2nd Lieutenant, USAF

Approved:

//signed//	4 March 2014
Col Karl C. Walli, PhD (Chair)	Date
//signed//	7 March 2014
Col Matthew D. Sambora, PhD (Member)	Date
//signed//	7 March 2014
Lt Col Anthony L. Franz, PhD (Member)	Date

AFIT-ENP-14-M-29

Abstract

The plenoptic camera enables simultaneous collection of imagery and depth information by sampling the 4D light field. The light field is distinguished from data sets collected by stereoscopic systems because it contains images obtained by an N by N grid of apertures, rather than just the two apertures of the stereoscopic system. By adjusting parameters of the camera construction, it is possible to alter the number of these 'subaperture images,' often at the cost of spatial resolution within each. This research examines a variety of methods of estimating depth by determining correspondences between subaperture images. A major finding is that the additional 'apertures' provided by the plenoptic camera do not greatly improve the accuracy of depth estimation. Thus, the best overall performance will be achieved by a design which maximizes spatial resolution at the cost of angular samples. For this reason, it is not surprising that the performance of the plenoptic camera should be comparable to that of a stereoscopic system of similar scale and specifications. As with stereoscopic systems, the plenoptic camera has immediate applications in the domains of robotic navigation and 3D video collection, though these domains may be expanded in the future as technological advances extend the range over which the camera accurately recovers depth.

To Mom and Dad

Acknowledgements

Much thanks to my research advisor, Col Karl Walli, for launching the project and providing insight and guidance throughout. I'd also like to thank the members of my research committee, Col Matthew Sambora and Lt Col Anthony Franz, for valuable discussions and feedback throughout the course of the research. Finally, I'm grateful to my friends and family for helping me get to where I am and for supporting me throughout the writing of this thesis.

Robert A. Raynor

Table of Contents

Page

Abstract iv

Acknowledgements vi

List of Figures ix

List of Tables xii

List of Symbols xiii

I. Introduction 1

II. Background 6

III. Imaging Theory 9

3.1 Introduction 9
3.2 The Light Field within a Simple Imaging System 9
3.3 Light Field Sampling with a Plenoptic Camera 14
Discretized Light Field Representation 19
3.4 Light Field Subspaces and Image Formation 20
Image Formation 23
Sampling Effects 26
3.5 Diffraction Effects 28
3.6 The Fourier Transformed Light Field 41
3.7 Focused Plenoptic Camera Sampling 47

IV. Plenoptic Ranging 53

4.1 Introduction 53
4.2 3D Point Clouds using Feature Matching 58
Quantization Error 59
Normally Distributed Error 60
Stereo Ranging with Normally Distributed Error 63
Estimator Comparison 65
4.3 The Scale Invariant Feature Transform 67
SIFT Feature Detection 67
SIFT Descriptor 69
SIFT Implementation 70
Feature Matching 71
Stereo Ranging using SIFT 71
Light Field Ranging using SIFT 75
4.4 Range Finding using Epipolar Plane Images 79

Page

Slope estimation using Light Field Photo-consistency . . . 79
Sampled Light Field Slope Uncertainty . . . 84
Continuous Light Field Slope Analysis . . . 92
Experimental Results . . . 93
Simulated Camera Analysis: Varying Lens Diameter . . . 95
Simulated Camera Analysis: Varying Detector Size . . . 98
Simulated Camera Analysis: Spatial/Angular Trade-off . . . 98
4.5 Range Finding via Refocusing . . . 104
Depth through Refocusing . . . 104
Fourier Domain Ranging . . . 106
Fourier Ranging Resolution . . . 107
Camera Calibration . . . 111
4.6 Summary . . . 116

V. Plenoptic Camera Utility . . . 119

5.1 Remote Sensing Application . . . 122
5.2 Autonomous Navigation . . . 123
5.3 3D Video . . . 124
5.4 Future Development . . . 125

VI. Conclusion . . . 127

6.1 Contributions . . . 127
6.2 Future Work . . . 128
6.3 Final Remarks . . . 129

Appendix A. Projection Slice Theorem . . . 131

References . . . 134

List of Figures

Figure		Page
1	The Light Field as a Radiance Distribution	10
2	The Light Field within a Camera	12
3	Ray Mapping in an Imaging System	13
4	Comparison between Plenoptic and Conventional Cameras	15
5	Plenoptic Camera Radiometry	17
6	Plenoptic Camera Sampling	18
7	Subaperture Image Formation	21
8	Light Field Slices	22
9	Point Source Moving Away from Camera	27
10	Geometry for Diffraction Analysis	29
11	Plenoptic Camera OTF: Relative Scale	34
12	Plenoptic Camera OTF: Absolute Scale	37
13	Plenoptic Camera OTF and Sampling Cutoff Frequencies	39
14	Plenoptic Camera Minimum Detector Size	40
15	The Projection Slice Theorem	42
16	Fourier Slice Imaging	44
17	Interpolation Filter Performance	45
18	Refocused Image Comparison	46
19	Focused Plenoptic Camera	48
20	Conventional Plenoptic Camera Comparison	48
21	Focused Plenoptic Camera Geometry	49
22	Conventional Plenoptic Camera Sampling	50

Figure Page

23 Focused Plenoptic Camera Sampling 51

24 Matched Sampling Performance for Focused and Traditional Plenoptic Cameras 52

25 An HCI Light Field 56

26 Plenoptic Camera Tradeoffs 57

27 Tradeoff Sampling 57

28 Quantization Error Visualization 60

29 A Simple Stereo Ranging Setup 63

30 Equal Baseline for Stereoscopic and Plenoptic Systems 64

31 Slope Estimator Performance 66

32 Difference of Gaussians Feature Detector. 69

33 Stereo Matching Performance using Modified SIFT (Author's Implementation) 73

34 Stereo Matching Performance using SIFT (VLFeat) 74

35 Maps from Stereo Matching using SIFT 75

36 Feature Matching Framework 76

37 Simulated Camera Performance with SIFT 77

38 SIFT Localization Error 78

39 Calculation of Photo-Consistency 80

40 Depth Estimation Using Photo-Consistency 81

41 DSI Shadow 82

42 Photo-Consistency of Gradient without Noise 86

43 Photo-Consistency of Gradient with Noise 87

44 Simulated Photo-Consistency of Gradient with Noise 88

45 Photo-Consistency of Gaussian Noise 89

Figure Page

46 Estimation Uncertainty . . . 90

47 Estimation Uncertainty (cont.) . . . 91

48 Slope Maps From Noisy Light Fields . . . 94

49 Photo-Consistency with Noise . . . 95

50 Experimental Slope Uncertainty, Varying Lens Diameter . . . 96

51 The Effect of Changing Number of Subapertures on Photo-Consistency . . . 98

52 Experimental Slope Uncertainty, Varying Detector Size . . . 99

53 Comparison of DSIs Generated from Noisy EPIs . . . 100

54 Experimental Slope Uncertainty, Varying Microlens Size . . . 101

55 Experimental Slope Uncertainty, Varying Microlens Size (cont.) . . . 102

56 Point Spread Function, 1D . . . 105

57 Point Spread Function . . . 106

58 Sparrow Resolvability Criterion . . . 110

59 Fourier Ranging Test . . . 112

60 Camera Calibration Target . . . 113

61 Slope Estimation Results . . . 114

62 Camera Calibration Plots . . . 115

63 Plenoptic Camera Performance Regimes . . . 120

64 Logarithmic Scale Uncertainties . . . 121

65 Uncertainty Nomograph . . . 121

List of Tables

Table		Page
1	LF Dimension Intervals	19
2	Conventional and Focused Plenoptic Camera Sampling Densities	51
3	Conventional and Focused Plenoptic Camera Equivalents	51
4	HCI Light Field Camera Parameters	56
5	Empirical Values of Sampled Slope Uncertainty, $\sigma_{\bar{m}}$, for Zero Added Noise.	117
6	Operator Definitions	132

List of Symbols

Symbol		Page
(s,t)	Coordinate defining a location on the focal plane of a conventional camera or the microlens plane of a plenoptic camera.	11
(u,v)	Coordinate defining a point on the main lens plane of a camera.	11
$L(s,t,u,v)$	The radiance along a ray traveling from (u,v) to (s,t).	11
D	Diameter of main collecting lens.	11
W_s	Width of the rectangular collecting area containing detectors (conventional camera) or microlenses (plenoptic camera) in s dimension.	11
z_o	Distance from object to main lens plane.	11
z_i	Distance from main lens plane to image of object.	11
f	Focal length of main collecting lens.	11
(x,y)	Transverse location of image relative to optical axis.	12
m	Derivative of s with respect to u for the mapping associated with a point source.	13
Δs	Detector width in a conventional camera, or microlens width in plenoptic camera.	16
Δt	Detector height in a conventional camera, or microlens width in plenoptic camera.	16
Δq	Detector width.	16
Δu	Width of detector instantaneous field of view (IFOV).	16
l_m	Spacing between main lens plane and microlens plane.	16
l_d	Spacing between microlens plane and detector plane. Equal to focal length of microlens for plenoptic camera.	16
$\bar{m}$	Sampled light field slope, in s samples per u sample.	18

Symbol		Page
γ	Factor relating sampled light field slope to continuous light field slope, defined as $\gamma = \Delta u / \Delta s = \bar{m}/m$	18
N_s	The number of microlenses in the s dimension.	19
N_u	The number of subapertures in the u dimension.	19
σ_m	Uncertainty in slope m as either a standard deviation or root mean square error (RMSE).	53
$\sigma_{\bar{m}}$	Uncertainty in sampled slope $\bar{m}$ as either a standard deviation or root mean square error (RMSE).	53
σ_z	Uncertainty in object distance z_o as either a standard deviation or root mean square error (RMSE).	53
σ_n^2	Variance of the random variable associated with normally distributed registration error. Used in context of feature matching.	60
σ^2	Variance of the random variable associated with Gaussian image noise. Used within the context of photo-consistency.	84
g	The strength of the image gradient within a subaperture image.	84

I. Introduction

Even amidst the technological marvels of the 21st century, the human vision system remains arguably the most impressive of its kind known to man. Our eyes are integrated together with a multitude of systems and procedures which give us a visual awareness of our surroundings, encompassing aspects like structure, depth, motion, contiguousness, texture, and more. Though no computer vision system may ever perfectly mimic these capabilities without major breakthroughs in artificial intelligence, certain isolated aspects continue to move within the grasp of modern technology.

Depth perception is one of these aspects. In the human vision system, depth information is obtained through a variety of means. Some of these means, such as the intelligent evaluation of the apparent size of recognized objects or other aspects of a scene, are well beyond the scope of this thesis. However, other methods make use of a simpler, more attainable mechanism. For example, the displacement between a person's two eyes means that each eye renders a slightly different view of a scene. The brain integrates these views together to provide a single image, along with a sense of depth.

The information afforded in this manner plays an important part in how we interact with the world, as is clear when one simply closes an eye. Upon doing so, it immediately becomes more difficult to ascertain spatial relationships among objects and textures in a scene and, in short, to interact with one's surroundings.

This same is true with respect to platforms employed by the U.S. Air Force. The coupling of depth information with imagery makes each more usable. In the context of

remote sensing, depth information can be useful in the automated analysis of a scene, as it provides an additional channel for use in image registration and segmentation. Depth information can also be critical for understanding the dynamics of a region and preparing operators who will be deploying to that location. Just as humans employ depth information to assist in movement and collision avoidance, so depth information can assist mobile Air Force systems in performing navigation. Indeed, such information is critical for any autonomous system using imagery to interact with its environment. The plenoptic camera is of interest to the U.S. Air Force because it stands to provide a cheap and accurate means of supplying depth information in real time to this wide variety of systems.

Like the human vision system, the plenoptic camera relies on the phenomenon of parallax. Parallax refers to the apparent shift in an object's location with respect to its background and foreground when viewed along different lines of sight. In the context of computer vision, parallax can be understood as the fact that, given two cameras having optical axes subject to relative translation and rotation, an object's imaged location will shift relative to the optical axis in a depth-dependent manner.

In its dependence on the parallax effect, the plenoptic camera is comparable to a slew of other passive ranging technologies. In stereovision systems, the parallax effect is quantified in the disparity between the pixel location of an object between two images taken from slightly different angles [1]. The process of triangulation is used in conjunction with this information to provide a depth estimate for an object. In a similar fashion, structure from motion (SFM) techniques determine disparity between images taken as a camera moves relative to a scene in order to estimate depth [2]. Insofar as more than two views of an object are available, structure from motion techniques provide denser sampling than stereoscopic devices. However, knowledge of the change in camera position and orientation for a moving platform may be

unknown or known with less precision than the relative orientation of the cameras in stereoscopic systems, and uncertainties in camera location will introduce error into the final depth estimation.

A less intuitive manifestation of the parallax phenomenon involves depth estimation from defocus. In this problem, an image is analyzed to determine the depth-dependent circle of confusion causing blurring at each image point. This circle of confusion can be thought of as resulting from the difference in the appearance of the scene (resulting from the parallax effect) when viewed through different portions of a camera's lens.

Thinking about defocus in this manner is a helpful primer for consideration of the plenoptic camera. The plenoptic camera incorporates an array of microlenses in front of a detector array in order to separate rays incident from different portions of the main lens [3]. Isolating the light from one portion of the main lens allows for creation of a 'subaperture image,' i.e., an image appearing as if taken from a small subaperture of the main lens [4]. The collection of subaperture images can be arranged into a 2D array, and disparities between successive images used to calculate depth in a manner similar to structure from motion, but without the camera position uncertainties associated with that technique. Thus, plenoptic camera ranging can be usefully thought of alternately as a constrained form of structure from motion, or as a variant of the depth from defocus problem.

The dense data set captured by the plenoptic camera, consisting of a collection of images of an object, captured from an N by M grid of locations, is a construct that developed within the image-based rendering community under the title of the 'Light Field' [5]. As the availability of plenoptic cameras for easy recording of the light field has increased, the concept of light field has seen growing interest within the computer

vision community, and numerous approaches and algorithms have been developed for estimating depth from the sampled light field [3][6][7][8].

The purpose of this research is to provide a framework for thinking about and quantifying the ranging capabilities of a plenoptic camera. The plenoptic camera design contains numerous degrees of freedom which affect different aspects of its performance. Some of these effects are very pronounced and others subtle. Depth estimation accuracy is also dependent on the content of the scene being imaged. Passive ranging systems typically have trouble with regions of a scene barren of features, like walls in a building. The same is true for the plenoptic camera, which gives best performance where image gradient magnitudes are high.

The goal of this research is to provide a description of the impact of these various factors on the plenoptic camera's depth resolving performance. This involves two major areas of investigation. The first area of investigation concerns the sampling characteristics of a plenoptic camera. Given a particular plenoptic camera geometry, how does this geometry sample the continuous light field? What will the sampled light field look like for a point at a known location relative to the camera? Answering these questions requires a detailed look at the plenoptic camera geometry and sampling characteristics.

Once the forward process of light field sampling has been defined, the range finding operation is simply the reverse process of backing out the location of a point responsible for the captured light field. In this domain we ask the question, how well can the characteristic shape of a point source within the sampled light field be identified and fit to a model which then yields depth information? Answering this question requires that we engage with the modern image processing techniques which have been applied to light field imaging and ranging, and seek to understand the sources of error and uncertainty within these techniques.

The overall contribution of the thesis is a set of equations which define the performance of the plenoptic camera and its dependence of various parameters of interest, as well as empirical testing which confirms and/or defines the scope of these equations.

II. Background

Its ability to perform stereo ranging was the main feature noted by Adelson and Wang when they first created the plenoptic camera [3]. Observing that the plenoptic camera allowed for an object to be viewed through different sections or sub-apertures of the main camera lens, they developed an algorithm to determine depth from the resulting parallax shift in object location. Since this shift is manifest as a sloping of lines when neighboring subaperture images are stacked on top of each other (see Fig. 8), their algorithm incorporated image gradient to estimate the slope direction within a region of the light field.

The potential to perform refocusing using light fields was first explored by Isaksen et al. in the context of image-based rendering [9]. The refocusing operation consists of nothing more than a shifted superposition of subaperture images in a manner which counteracts the parallax effect for a given object depth. Not until the construction of a hand-held plenoptic camera by Ng et al. was this capability demonstrated in the context of light field photography with a plenoptic camera [10]. In principle, range finding via refocusing involves the construction of a stack of refocused images followed by a search for sharp features within each image to isolate depths which contain objects.

The sheared projection which constitutes this refocusing operation bears strong similarity to certain computed tomographic techniques employed in medical imaging. In that context, projections of a density distribution obtained by radiographic techniques such as x-ray scanning are used to recreate the original density distribution. Here, the density distribution plays a role analogous to that of the light field, and the projections, a role equivalent to that of the refocused images. Often, a useful relationship exists between the distribution and its projections in a transformed domain such as the Fourier domain. In the medical imaging domain, such relationships

can help simplify the process of reconstructing the density distribution, and subsequently rendering views of the object from different perspectives. In the domain of light field imaging, this second aspect is most readily applicable. Ng et al. show that image rendering performed in this manner can provide a significant reduction in the computational cost of refocusing [11]. This is because refocused images are obtained in the Fourier domain by extracting a 2D slice from the light field, as opposed to the projection operation required in the spatial domain. A Fourier domain approach to the depth-through-refocusing technique, demonstrated in [12], searches for images having high spatial frequencies of large magnitude, as this suggests the presence of sharp features associated with in-focus objects.

Plenoptic camera range finding also benefits from research performed on light fields generated by methods predating the plenoptic camera. Light fields have been traditionally collected using 2D arrays of cameras. A single camera mounted on a gantry allowing for translation in two dimensions also allows for scanning light field collection. Some of the first work with such data used edge detection and line fitting to estimate light field slope, as in [6]. This technique is comparable to the section concerning the application of the SIFT algorithm to light field ranging within this thesis (See Section 4.3). More recent work with light fields gathered from track-mounted cameras estimates local light field orientation by finding the slope along which the light field shows high consistency (low variance) [7]. The uncertainty of this approach is assessed in detail within this report. Some of the most sophisticated techniques for employing light fields from plenoptic cameras involve the combination a local slope estimator with a system of global constraint enforcement. The global optimization framework employed in [8] employs the structure tensor to provide a local slope estimate. The structure tensor, derived in [13], starts with the principle that a region having a particular orientation should contain energy concentrated along

a line in the Fourier domain. Orientation detection can be achieved by least squares fitting to this line, which is an operation able to be performed entirely within the spatial domain. The structure tensor itself achieves good local estimates compared to other methods like the gradient method in [3]. The estimate is improved by employing an optimization framework in which the cost of assigning a given depth takes into account the ordering of objects, evidenced by occlusions, as well as certain other features, such as the location of edges yielded via edge detection.

III. Imaging Theory

3.1 Introduction

The dataset captured by a plenoptic camera is known as a light field [3]. The light field is based on a geometric optics formulation of light by which it describes the propagation of light energy in a space. Some of the earliest work with light fields appears in the context of image based rendering. See [14] and [5] for detailed descriptions of the light field within that context. The goal of this chapter is to describe how the light field is captured by the plenoptic camera, and to examine what the sampled light field will look like for a point source at some known location. To this end, different subspaces of the light field allowing for easy visualization of its structure will be discussed. The process of generating refocused images from the light field will be considered within both the spatial and frequency domains. Finally, the traditional plenoptic camera sampling geometry will be compared with that of the 'focused' plenoptic camera, and a scheme for creating a focused plenoptic camera with equivalent sampling characteristics to that of a traditional plenoptic camera will be presented.

3.2 The Light Field within a Simple Imaging System

In perhaps its most basic form, the light field is nothing more than the radiance distribution in a 2D plane. The radiance along a ray at the point (s,t) and in the direction (θ,ϕ), defined according to Fig. 1, is given by [15]

$$L(s,t,\theta,\phi) = \frac{d^2\Phi}{d\Omega dA_1 \cos\theta} \tag{1}$$

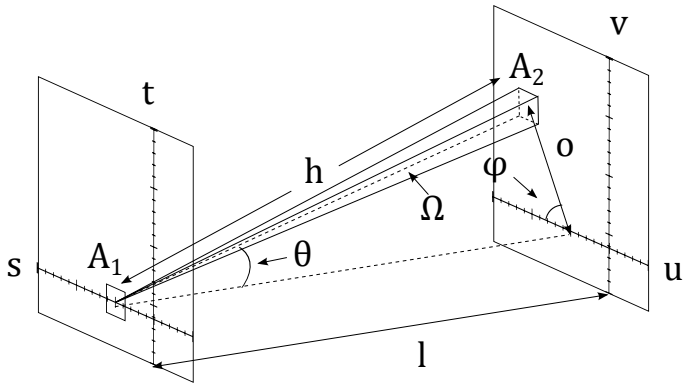


Figure 1. The Light Field as a Radiance Distribution. The light field can be thought of as the radiance distribution along a 2D plane. $L(s,t,\theta,\phi)$ gives the radiance at the position (s,t) and in the direction (θ,ϕ). This direction can be conveniently defined in terms of a second point, (u,v), on a plane parallel to the original plane. Likewise, the solid angle, Ω, used to normalize the radiance, can be given in terms of an area on the second plane, A_2, and the distance between the two planes, l.

where L has units of $[W/cm^2sr]$. The angles in this equation can be defined conveniently by considering the intersection of the ray with a second plane positioned parallel to the first. The geometry needed for performing this parametrization is pictured in Fig. 1. We define $\Phi(s,t,\theta,\phi)$ as the radiant flux exiting the area A_1 in the (s,t) plane into the solid angle Ω, defined by the pyramid-shaped region in the figure.

The length, h of this region is also the hypotenuse of the right triangle with base l and height o. The length of the hypotenuse is determined via the Pythagorean theorem as

$$h = \sqrt{l^2 + (s-u)^2 + (t-v)^2}. \tag{2}$$

This allows for the angle, θ, to be defined implicitly as

$$\cos(\theta) = \frac{\sqrt{l^2 + (s-u)^2 + (t-v)^2}}{l} = \sqrt{1 + \frac{(s-u)^2 + (t-v)^2)}{l^2}}. \tag{3}$$

At this point, we make the assumption that l is sufficiently large compared to the other dimension that $\cos(\theta) \approx 1$.

Since the angle spanned by A_2 is small, the solid angle Ω can be approximated by

$$\Omega \approx A_2 \cos(\theta)/l^2 \approx \Delta u \Delta v / l^2 \tag{4}$$

where the cosine is approximated to equal one as discussed above. Using these definitions, along with $A_1 = \Delta s \Delta t$, Eq. 1 can be written in an alternate parametrization as

$$L(s,t,u,v) = \frac{l^2 d^4 \Phi(s,t,u,v)}{dsdtdudv}. \tag{5}$$

We are interested in using this parametrization of the light field to describe the radiance distribution inside of a simple camera. We let the (s,t) plane represent the focal plane or detector plane of the camera, and the (u,v) plane represent the plane of the collecting lens. The function $L(s,t,u,v)$ gives the radiance along a ray traveling from the point (u,v) on the main lens plane to the point (s,t) on the focal plane. The variables u and v are defined as $\{(u,v) \in \Re^2 : |u| \leq R \ \wedge |v| \leq \sqrt{R^2 - u^2}\}$ where $R = \mathrm{D}/2$ is the radius of the main collecting lens. The variables s and t are likewise defined as $\{(s,t) \in \Re^2 : |s| \leq W_s/2 \wedge |t| \leq W_t/2\}$ where W_s and W_t are the widths of the rectangular collecting area (which we will later define as the area containing microlenses in the case of a plenoptic camera) in the s and t dimensions. The zero of each axis is located at the optical axis of the camera.

We assume an ideal imaging relationship between an object point located outside of the camera a distance z_o from the main lens plane and an image point located inside of the camera a distance z_i from the main lens, where z_o and z_i are related by the imaging relation

$$\frac{1}{z_o} + \frac{1}{z_i} = \frac{1}{f} \tag{6}$$

where f is the focal length of the lens. The assumption of an ideal imaging relationship implies that all rays coming from the object point and passing through the

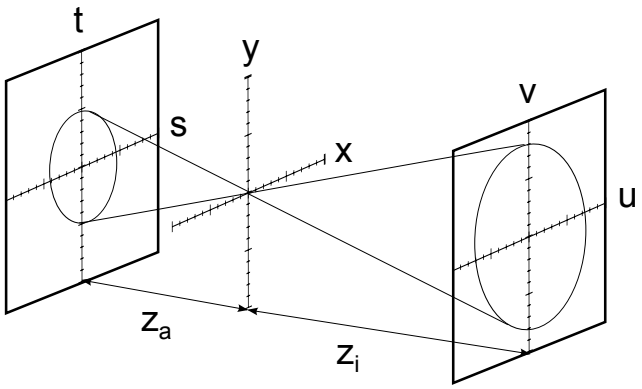


Figure 2. The Light Field within a Camera. The light field can be used to describe the radiance distribution within an imaging system. Under the condition of ideal imaging, all of the light from a given object point will pass through a point (x, y) on a plane located some distance z_i from the main lens.

aperture stop represented by the main collecting lens will also pass exactly through the image point defined by the distance z_i and the coordinate (x, y), which specifies the transverse location of the image relative to the optical axis. Fig. 2 provides a visualization of this scenario.

The requirement that all rays pass through a specified image point for a given object point implies a mapping between the uv plane and the st plane that is unique for each object point. This mapping specifies the region of the light field containing the radiance from the point source. Fig. 3 provides the geometry for understanding this mapping in two dimensions. In passing through the image point, the ray forms two similar triangles, one on each side of the image plane. Their dimensions are related by

$$\frac{s-x}{z_a} = \frac{x-u}{z_i}. \tag{7}$$

This equation is solved for s in terms of u by

$$s = x\left(1+\frac{z_a}{z_i}\right) - u\frac{z_a}{z_i} = mu + x(1-m) \tag{8}$$

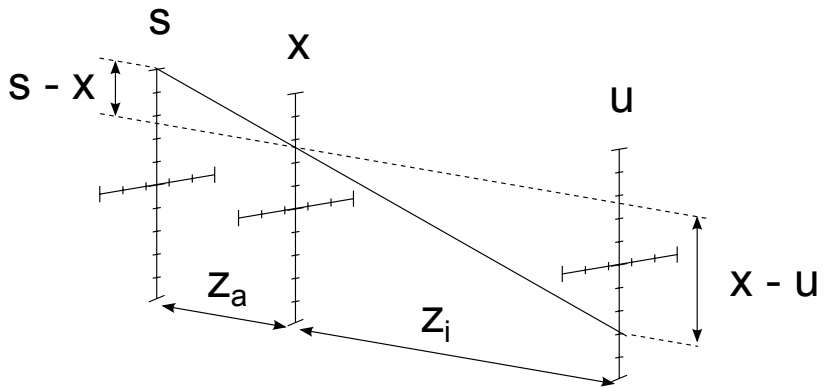


Figure 3. Ray Mapping in an Imaging System. The condition of ideal imaging, that all rays from a given object pass through the same point within the camera, imposes a mapping between the main lens position (u, v) and the focal plane position (s, t) The ray in the figure forms two similar triangles, one anterior to the image plane and one posterior. Equating the ratio of each triangle's dimensions yields an equation that can be solved to establish the mapping between (s, t) and (u, v) unique to a given image location.

where $m = -z_a/z_i$. Substituting $z_a = l - z_i$ and letting $\alpha = z_i/l$, we see that $m = 1 - 1/\alpha$, in agreement with the format used in [10]. The equation relating t and v can be derived in the same manner. Eq. 8 is nothing more than the equation of a line with slope m and an u-intercept related to x. Since m is a function of z_i, it is implicitly a function of object distance. This fact will become of much importance as we move forward.

Ignoring the effects of reflection and absorption by optical elements, the radiance within the conic regions of Fig. 2 will be equal to the radiance at the source object. We can use the mapping defined in Eq. 8 to state this mathematically, as in

$$L\left(mu + x(1-m), mv + y(1-m), u, v\right) = L_0(x, y, m) \tag{9}$$

where $L_0(x, y, m)$ is the radiance at the object point associated with (x, y, m). Identifying an object point in this manner is appropriate since each of these values can be determined directly from the location of the object in world space. The equation is simplified if we identify the object point, not by the location of its image, (x, y),

but by the center of its circle of confusion at the microlens plane, (s, t). This location is determined by setting $u = 0$ in Eq. 8, whereupon we see that $s = x(1 - m)$ and $t = x(1 - m)$. Upon making this substitution, we obtain the simplified form

$$L\left(s + mu, t + mv, u, v\right) = L_0(s, t, m). \tag{10}$$

For a given object point, the values of m, s, and t in this equation will be fixed. Allowing u and v to vary across their respective domains, the equation assigns the radiance L_0 to the points on a 2D plane within the 4D space of the light field. In the 2D subspace of the light field defined by fixing t and v, this plane will appear as a 1D line of slope $ds/du = m$ (See Fig. 6). More detailed interpretations of the equation are explored in future sections.

3.3 Light Field Sampling with a Plenoptic Camera

Fig. 4 provides a comparison of a conventional camera and a plenoptic camera. A detector element of a conventional camera captures only information about the irradiance (power per unit area) at the focal plane. The irradiance is determined by integrating the radiance function, L, over the solid angle defined by the main lens. Using Eq. 5, this integration can be performed instead over the lens area,

$$E(s, t) = \frac{1}{l^2} \int_u \int_v L(s, t, u, v) dv du \tag{11}$$

where u and v are integrated over their domains, as defined earlier.

The radiant flux (power) captured by a detector will be equal to the integrated irradiance over the surface of the detector. We represent this integration by convolving

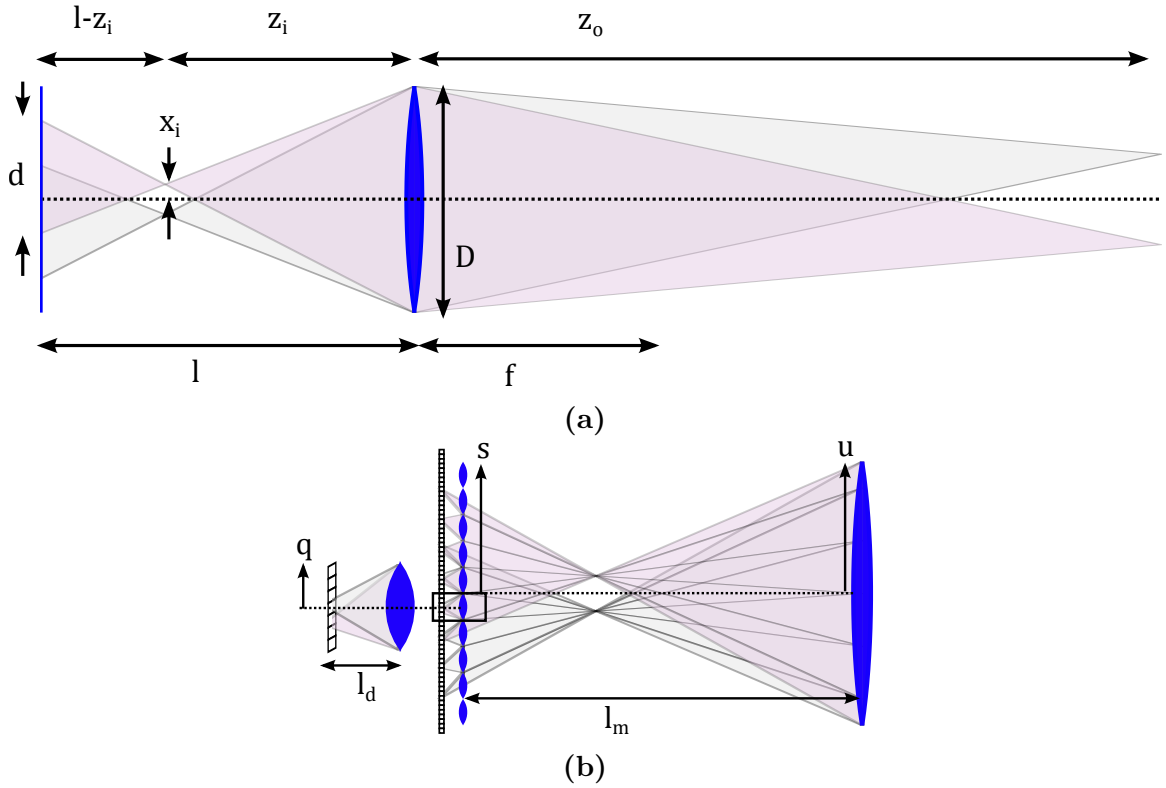


Figure 4. Comparison between Plenoptic and Conventional Cameras. (a) shows the overlapping circles of confusion for two out-of-focus point sources in a traditional camera. The insertion of a microlens array before the detector array in (b) allows the light from 'out-of-focus' point sources to be sampled separately. The microlenses are placed one focal length away from the detector plane. This allows for each microlens to be envisioned as an angularly sensitive pixel, and makes possible mapping between each microlens subpixel and a region of the main lens.

the irradiance distribution with a detector-shaped kernel, $h_d(s,t)$, given by

$$h_d(s,t) = RECT(s/\Delta s, t/\Delta t) \tag{12}$$

where here Δs and Δt are detector sizes. Sampling is then achieved via multiplication with a comb function, having a periodicity matched to the detector pitch, such that [16]

$$\Phi(s,t) = [E(s,t) * h_d(s,t)] \frac{1}{\Delta s \Delta t} COMB\left(\frac{s}{\Delta s}, \frac{t}{\Delta t}\right). \tag{13}$$

Since the camera detects only irradiance, it will not distinguish between the overlapping circles in Fig. 4a. Where the cones of light from the two point sources overlap, they both contribute to the irradiance integrated by the detector.

The plenoptic camera introduces a microlens array in front of the detector array of a conventional camera. In this section, we consider the case where the microlenses are separated from the detector array by a microlens focal length. Since the microlenses are small relative to the separation, l, between the microlens plane and the main lens, the situation of the main lens approximates 'optical infinity' [10]. This means that the detectors behind each microlens image the back of the main lens, with each detector integrating up the radiance from the region of its instantaneous field of view (IFOV). This allows for the light from the two point sources in Fig. 4b to be recorded separately, since each point source inhabits a different region of the light field.

Fig. 5 shows how the detector size Δq and the microlens size Δs stop the region of the light field, L, integrated by each pixel. The image of the detector at the main lens plane gives the IFOV, $\Delta u = \Delta q \; l_m/l_d$. In order to obtain a radiant flux quantity associated with each detector, Eq. 5 must be integrated over the regions defined by the microlens in front of the detector as well as the detector IFOV at the main lens

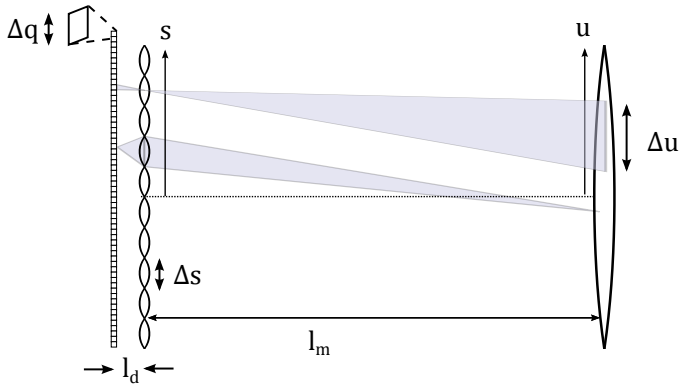


Figure 5. Plenoptic Camera Radiometry. Each microlens performs stopping of the radiance reaching the detectors beneath it. The radiance is also stopped at the at detector plane by the extent of the detector itself. Projecting the detector dimensions forward onto the main lens allows us to picture this stopping as occurring at the main lens plane according to the dector IFOV, Δu. More precisely, the dimensions of each microlens and the detector IFOV provide the limits of integration needed for obtaining a radiant flux by integrating Eq. 5

plane. Analogously to the conventional camera, we represent this integration as a convolution with a 4D kernel whose dimensions are determined by the size of these regions, given by

$$h(s,t,u,v) = RECT(s/\Delta s, t/\Delta t, u/\Delta u, v/\Delta v). \tag{14}$$

Once again, sampling is achieved via multiplication with the appropriate comb function,

$$S(s,t,u,v) = \frac{1}{\Delta s \Delta t \Delta u \Delta v} COMB\left(\frac{s}{\Delta s}, \frac{t}{\Delta t}, \frac{u}{\Delta u}, \frac{v}{\Delta v}\right), \tag{15}$$

such that

$$\Phi(s,t,u,v) = \frac{1}{l^2}[L(s,t,u,v) * h(s,t,u,v)]S(s,t,u,v). \tag{16}$$

Note that, although it is likely that microlenses will be circular in shape and arranged in a non-rectangular grid, our sampling equations assume rectangular lenses and a rectangular grid for the sake of simplicity. Even where a close-packed hexagonal grid of circular microlenses is used to maximize fill-factor, resampling of the light field

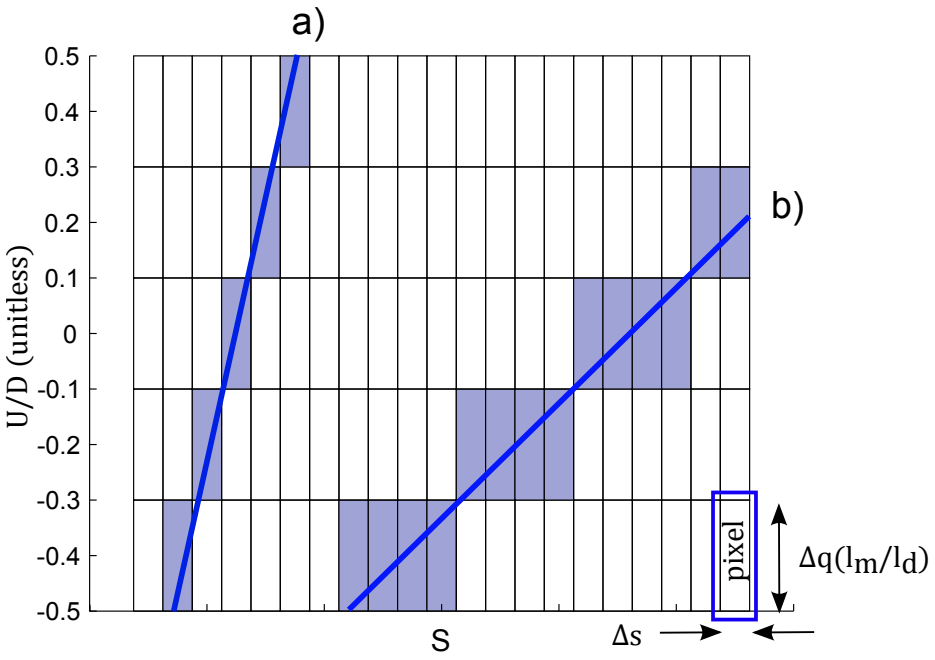


Figure 6. Plenoptic Camera Sampling: How the plenoptic camera samples the sloped line associated with a point source. The process of convolution followed by sampling with the comb function in Eq. 16 collects all the energy associated with the portion of the line within the rectangular region and attributes it to the discrete sample at the center of the region. Each row of the figure represents a row of pixels within a different subaperture image. Each column is a row of pixels within a microlens image.

after data collection can be used to give effective sampling characteristics similar to those indicated here [17].

Fig. 6 illustrates how the plenoptic camera samples the sloped line associated with a point source, as discussed in the previous section. The process of convolution followed by sampling with the comb function in Eq. 16 collects all the energy associated with the portion of the line within the rectangular region and attributes it to the discrete sample. The slope, $\bar{m}$, in s samples per u sample is related to the continuous slope m by the relation

$$\bar{m} = \frac{ds/\Delta s}{du/\Delta u} = m\frac{\Delta u}{\Delta s} = m\gamma \tag{17}$$

where $\gamma = \Delta u/\Delta s$. Note that there is a critical point at $|\bar{m}| > 1$, where the point source illuminates multiple microlenses per subaperture (multiple pixels per subaperture image, represented by a row within the figure).

Discretized Light Field Representation.

We find that when working with actual sampled light fields, it is very convenient to work with a normalized form of the light field dimensions. To this end, we define the function

$$K(\bar{s}, \bar{t}, \bar{u}, \bar{v}) = \Phi(\bar{s}\Delta s, \bar{t}\Delta t, \bar{u}\Delta u, \bar{v}\Delta v) \tag{18}$$

where the variable ranges are defined in Table 1. N_s and N_t give the number of microlenses in each respective dimension. N_u and N_v give the number of aperture regions in each dimension, where this number is related to the number of pixels beneath each microlens by $N_u = \Delta s/\Delta q$.

Table 1. LF Dimension Intervals

Variable	is a member of the set	Variable	is a member of the set
s	$[-W_s/2, W_s/2]$	$\bar{s}$	$[-(N_s-1)/2, (N_s-1)/2]$
t	$[-W_t/2, W_t/2]$	$\bar{t}$	$[-(N_t-1)/2, (N_t-1)/2]$
u,v	$\{(u,v) \in \Re^2 : \lvert u\rvert \leq R \ \wedge \lvert v\rvert \leq \sqrt{R^2-u^2}\}$	$\bar{u}$	$[-(N_u-1)/2, (N_u-1)/2]$
		$\bar{v}$	$[-(N_v-1)/2, (N_v-1)/2]$

We allow the normalized variables to take on real values. However, where non-integer values are used, it is implied that some form of interpolation must be employed in order to provide the function value. In most cases within this document, interpolation will be discussed explicitly. When a normalized variable is used as the index of a summation, the variable is assumed to span the set of integers contained in the interval defined by Table 1.

Eq. 8, which specifies the region of the radiance distribution populated with the radiance from a single point source, must be modified for use with these normalized coordinates. The modified equation is given by

$$K\left(\bar{s} + \bar{m}\bar{u}, \bar{t} + \bar{m}\bar{v}, \bar{u}, \bar{v}\right) = K_0(\bar{s}, \bar{t}, \bar{m}). \tag{19}$$

3.4 Light Field Subspaces and Image Formation

It is useful to gain a sense of the information contained in a number of 2D subspaces of the light field formed by fixing two of its parameters. In its two-plane parametrization the light field contains a certain degree of symmetry, as it can be used to describe the radiance at either of the two planes used in the parametrization. Nonetheless, the structure of the information contained in the light field is very much asymmetric when the light field is used to describe the radiance distribution within an imaging system.

The coordinates u and v of the sampled light field specify the location of the subaperture of size $\Delta u \Delta v$ which crops the radiance integrated to give the radiant flux $\Phi(s, t, u, v)$ collected by a detector. Likewise, the coordinates s and t specify the location of a microlens of size $\Delta s \Delta t$ which crops the radiance at the microlens plane. Comparing Eq. 16 and 13 shows that this results in spatial sampling in the microlens plane that depends on microlens pitch in the same way that focal plane sampling depends on pixel pitch in a conventional camera.

From these considerations, we expect that by fixing u and v to some value, we will obtain an image very comparable to that taken with a conventional camera of pixel size $\Delta s \Delta t$, but with a lens that is masked to only allow light through the region indicated by the coordinate (u, v). This 2D light field slice is what is known for this reason as a 'subaperture image' [4]. As illustrated in Fig. 7, the subaperture image will have reduced blur for defocused objects due to the higher f-number associated with the smaller aperture size. The figure also illustrates how the mapping identified in Eq. 19 results in a u-dependent shift in the image location.

Fig. 8 shows the 3D subspace of the light field obtained by fixing v. The representation of the subspace is built by vertically stacking all the subaperture images having the same value of v. As expected, the top surface of the structure has the

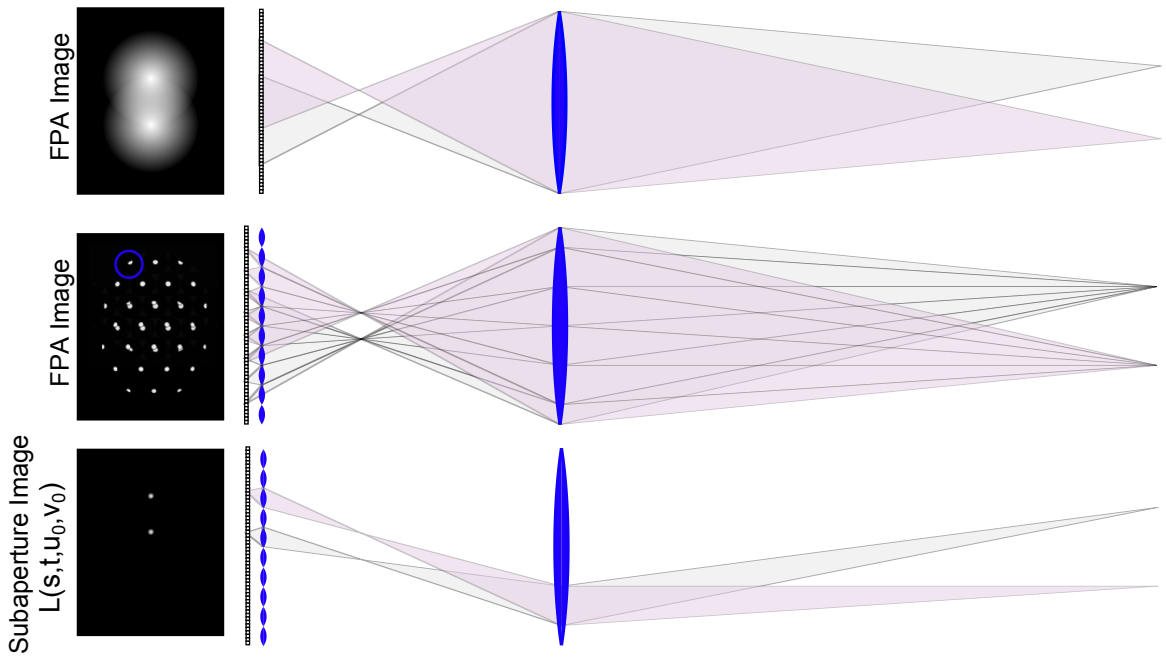


Figure 7. Subaperture Image Formation. The top row of the Figure illustrates the image collected by a conventional camera for two defocused point sources. A subaperture image (bottom row) is obtained from the light field (middle row) by fixing u and v (i.e. by taking all of the pixels 'looking' at a particular subaperture). The subaperture image is sharper than the conventional image due to its higher f-number. Also, the location of the point sources within the image is shifted due to the mapping of Eq. 19.

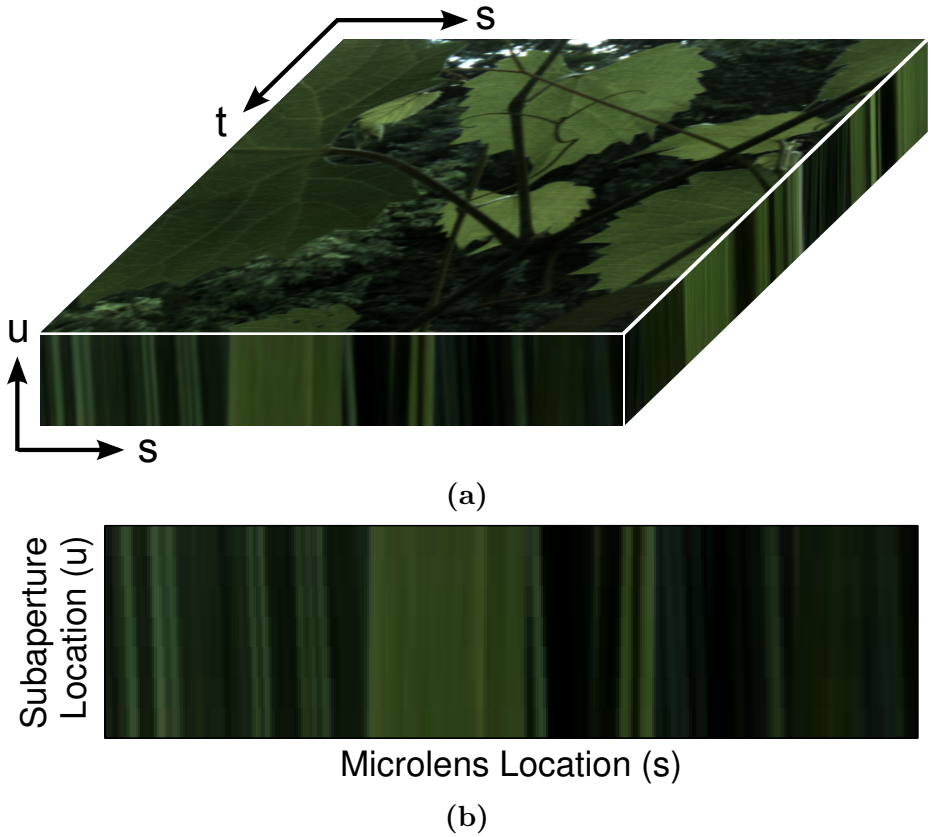


Figure 8. Light Field Slices. (a) shows a 3D subspace of the light field obtained by fixing v. The representation shown here is formed by stacking a row of subaperture images into a 3D cube. The slice of this cube obtained by fixing the spatial coordinate, t, is known as an epipolar plane image (EPI), and is shown in (b). In the 4D light field, a point source in object space is represented by a 2D plane. In an EPI, this plane appears as a line with slope $\bar{m}$.

characteristics of an image taken by a conventional camera. The front face of the structure is the 2D plane formed by fixing t and v. This subspace is known as an Epipolar plane image or EPI [6]. The sloped lines visible in this image result from difference in the apparent location of objects when viewed through different apertures of the camera under the parallax effect. The slope of each line, given by Eq. 17, is related to the distance from the camera to the point responsible for the line.

Image Formation.

By integrating Eq. 16 over u and v, we are able to recover the image produced by the conventional camera in Eq. 13. This is not surprising, as the summation of subaperture images is nothing more than the recombination of the information shown to be separately captured in Fig. 4. The image is given by

$$img(\bar{s}, \bar{t}) = \sum_{\bar{u}} \sum_{\bar{v}} K(\bar{s}, \bar{t}, \bar{u}, \bar{v}) \tag{20}$$

where the summations are over all integer values within the intervals defined for $\bar{u}$ and $\bar{v}$.

The summation over $\bar{u}$ and $\bar{v}$ will result in the 2D lines in Fig. 6 and Fig. 8b being projected down into one dimension. It is evident that, if the line is initially vertical, its projection will fill only one pixel in the the final image, $img(\bar{s}, \bar{t})$. Conversely, if the line is sloped such that it spans multiple s samples, it will impact multiple pixels of the projection $img(\bar{s}, \bar{t})$. This spreading out of sloped lines is nothing more than the circle of confusion associated with out-of-focus points in a conventional photograph.

When an image is formed via projection, vertical lines ($\bar{m} = 0$) appear as in-focus points, while increasing slope leads to an increasing degree of defocus in the generated image. This suggests that some degree of refocusing may be performed by 'shearing' the light field by some amount, Δm, prior to projecting, such that points originally having slope $-\Delta m$ will appear to be in focus.

Though we arrive at this result from an intuitive consideration of the light field structure, it is possible to achieve the same result more formally by looking at the re-parametrization of the light field necessary to simulate a conventional camera with varying focal length, as in [10].

In order to represent the shearing operator, we find it convenient to define the vectors $\bar{\mathbf{x}} = [\bar{s}, \bar{t}, \bar{u}, \bar{v}]^T$, $\bar{\mathbf{s}} = [\bar{s}, \bar{t}]^T$, and $\bar{\mathbf{u}} = [\bar{u}, \bar{v}]^T$. We then allow each of our functions to accept these vectors as arguments, as in $f(\bar{\mathbf{x}}) = f(\bar{s}, \bar{t}, \bar{u}, \bar{v})$. Under this convention, the shearing operator can be defined by a matrix multiplication of the argument vector, as in

$$\mathcal{B}[f(\bar{\mathbf{x}})](\bar{\mathbf{x}}) = f(\mathcal{B}^{-1}\bar{\mathbf{x}}), \tag{21}$$

where the matrix needed to shear the light field by the slope $\bar{m}$ is given by [11]

$$\mathcal{B}_{\bar{m}} = \begin{bmatrix} 1 & 0 & -\bar{m} & 0 \\ 0 & 1 & 0 & -\bar{m} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathcal{B}_{\bar{m}}^{-1} = \begin{bmatrix} 1 & 0 & \bar{m} & 0 \\ 0 & 1 & 0 & \bar{m} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{22}$$

To confirm that this operator has the desired effect, we write Eq. 19 in terms of the new vector coordinates introduced here.

$$K([\bar{s} + \bar{m}\bar{u}, \bar{t} + \bar{m}\bar{v}, \bar{u}, \bar{v}]^T) = K_0(\bar{s}, \bar{t}, \bar{m}). \tag{23}$$

We wish to shear the light field such that the object identified by $(\bar{s}, \bar{t}, \bar{m})$ is represented within the light field by a vertical line. To do this we apply the operator $\mathcal{B}_{-\bar{m}}$. Under the effects of this transformation, the value $K_0(\bar{s}, \bar{t}, \bar{m})$ is remapped to a new region of the light field:

$$K_0(\bar{s}, \bar{t}, \bar{m}) = \mathcal{B}_{-\bar{m}}[K(\bar{\mathbf{x}})](\bar{\mathbf{x}}) = K(\mathcal{B}_{-\bar{m}}^{-1}\bar{\mathbf{x}}). \tag{24}$$

The matrix product $\mathcal{B}_{-\bar{m}}^{-1}\bar{\mathbf{x}}$ evaluates as

$$\begin{bmatrix} 1 & 0 & -\bar{m} & 0 \\ 0 & 1 & 0 & -\bar{m} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{s} + \bar{m}\bar{u} \\ \bar{t} + \bar{m}\bar{u} \\ \bar{u} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} \bar{s} \\ \bar{t} \\ \bar{u} \\ \bar{v} \end{bmatrix} \tag{25}$$

such that the final result of the shearing operation is the updated mapping,

$$K([\bar{s}, \bar{t}, \bar{u}, \bar{v}]^T) = K_0(\bar{s}, \bar{t}, \bar{m}). \tag{26}$$

This equation confirms that, under the impact of the shearing operator, $\mathcal{B}_{-\bar{m}}$, a point source which originally mapped to a region of the light field with slope $\bar{m}$, now maps to a region having zero slope.

The use of operators in this section was based on a similar use of operators in [11]. As we will rely on operator notation in the next section, it will continue to be useful to adopt conventions and operator definitions similar to those employed in that paper.

The projection used earlier to form a conventional image from the light field is given its own operator, defined as

$$\mathcal{P}[f(\bar{\mathbf{x}})](\bar{\mathbf{s}}) = \sum_{\bar{u}} \sum_{\bar{v}} f(\bar{s}, \bar{t}, \bar{u}, \bar{v}). \tag{27}$$

The composition of the shearing and projecting operators can be defined as an imaging operator, since it results in the generation of a refocused image,

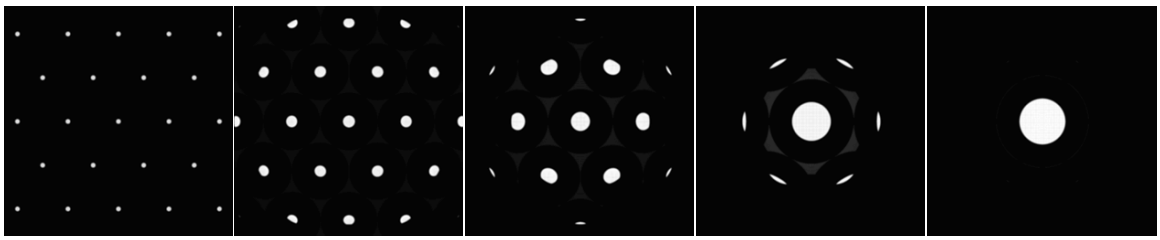
$$\mathcal{I}_{\bar{m}}[f(\bar{\mathbf{x}})](\bar{\mathbf{s}}) = (\mathcal{P} \circ \mathcal{B}_{\bar{m}})[f(\bar{\mathbf{x}})](\bar{\mathbf{s}}) \tag{28}$$

where $\circ$ indicates functional composition.

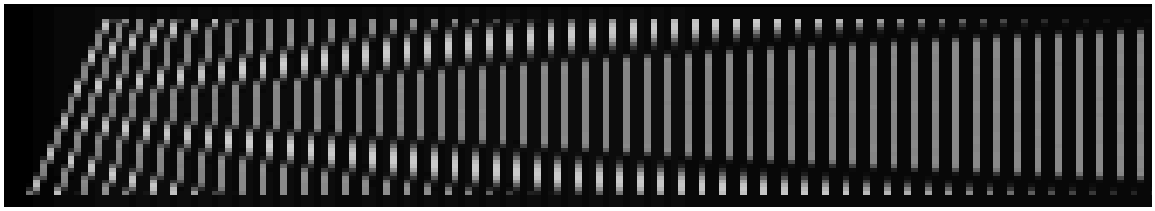
Sampling Effects.

Now that we have identified a procedure for generating refocused images from the light field, it is appropriate to examine the features of images rendered by this procedure. The representation of the plenoptic camera light field sampling in Fig. 6 illustrates that, where $|\bar{m}| < 1$, the radiance from a single point source is localized to one pixel per subaperture image. However, where this condition is not met, the number of pixels per subaperture image increases to the order of $n = \text{round}(\bar{m}^2)$ (Fig. 6 shows a number of pixels per row of $n = \text{round}(\bar{m})$, but each row is only one dimension of a subaperture image). This is an important observation because it indicates that there is a limit to the plenoptic camera's ability to produce refocused imagery. Even in the absence of an optical point spread function due to aberrations or diffraction, it is not always possible to generate from the light field imagery in which the circle of confusion due to defocus is contained within a single image pixel.

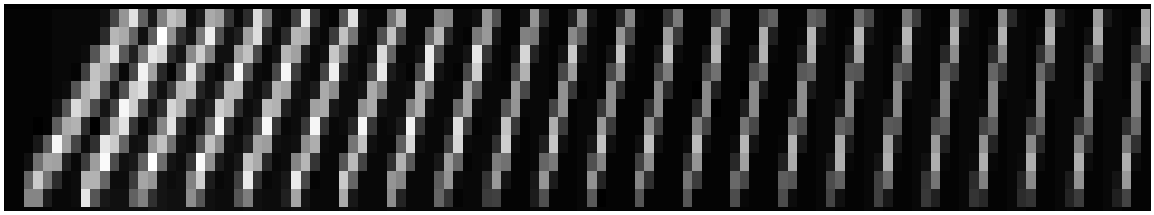
Fig. 9 supports this result by means of a plenoptic camera simulated via geometric raytracing. The first row of the figure shows the distribution of irradiance at the detector plane of the camera at intervals as a single point source is repositioned successively further from the camera. The second row shows lines of diminishing slope as the object becomes more distant, as would be seen in an EPI slice of the light field. The slopes, $\bar{m}$, in this case are much smaller than unity, so only one pixel is illuminated per row for each line. The third row shows the effect of increasing the detector size of the camera, Δq. Since $\Delta u = \Delta q l_m / l_d$, this leads to an increase in Δu which in turn leads to an increase in $\gamma = \Delta u / \Delta s = \bar{m}/m$ as well as $\bar{m}$. This brings some of the slopes on the left side of the image near to the threshold value of unity, and these lines begin to shown a certain amount of spreading. The spreading in



(a)



(b)



(c)

Figure 9. Point Source Moving Away from Camera. (a) shows simulated raw detector data as a point source moves away from the camera at intervals. (b) shows 2D slices of the light fields obtained from the sensor data in (a), but at smaller intervals. (c) illustrates the spread of the sampled light field as subpixel size increases.

the figure is actually somewhat greater than would be expected for an ideal imaging system, and this is because of spherical aberration induced by the main camera optic.

3.5 Diffraction Effects

Analysis of light field sampling by a plenoptic camera up to this point has assumed a geometric optics framework free of aberrations or effects of diffraction. This approach is helpful for setting the stage for the plenoptic camera, but the effects of these factors can play a role in the camera's design and performance. This section will provide a framework for considering the effects of diffraction.

Diffraction is a phenomenon with origins outside of the ray model of light. The discussion here follows a far more detailed treatment of the subject in [18]. Early speculation concerning the nature of light propagation proposed that each point on an expanding wavefront would expand into a secondary spherical wavefront, such that the envelope of these secondary wavefronts constituted the new wavefront. This concept is known as the Huygens principle. By allowing these secondary wavefronts to interfere with each other, Fresnel was able to account for many observed diffraction effects. It was not until later on that this approach was to some degree grounded mathematically in Maxwell's equations, first by Gustav Kirchoff. Modifications of his original assumptions led to the Rayleigh-Sommerfield formula, as commonly used today [18],

$$U(s,t) = \frac{1}{i\lambda} \iint U(u,v) \frac{\exp(i\frac{2\pi}{\lambda} r_{01})}{r_{01}} \cos\theta du dv. \tag{29}$$

where r_{01} is defined in Fig. 10. This equation gives mathematical expression to the notion of secondary spherical wavefronts expanding from an initial wavefront. The scalar field U (a non-physical quantity introduced in place of the vector field to make

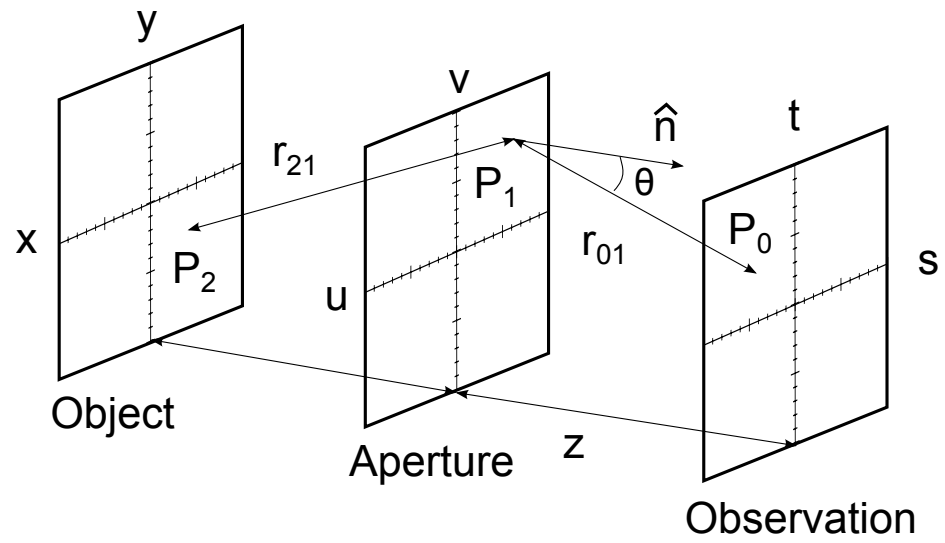


Figure 10. Geometry for Diffraction Analysis: Geometry to be used in discussing diffraction from an aperture or lens. P_2 is a point on the object, P_1 is a point on the aperture, and P_0 is point on the observation plane. In an imaging system, a lens is placed at the aperture plane.

the problem tractable) at the plane (s,t) is given by the superposition of spherical waves originating from each point on (u,v).

In the case of light passing through an aperture, various approximations to this equation can be made depending on the scale of the aperture compared to the distance to the second plane. These approximations primarily involve the number of terms retained in the Taylor series expansion of r_{01}. For example, in the Fraunhofer region, such eliminations lead to the form [18]

$$U(s,t) = \frac{A}{i\lambda z}\iint U(u,v)\exp\left[-i\frac{2\pi}{\lambda z}(su+tv)\right]dudv, \tag{30}$$

where A is a phase factor dependent on z. It is worth noting that the field at (s,t) is simply the scaled Fourier transform of the field at (u,v). The distances needed for this approximation are very large at optical wavelengths.

The Fraunhofer approximation is noted here because a similar result is obtained for the impulse response of a simple imaging system, i.e. the field found at the image

plane for a point source at the object plane. Such a system is modeled by starting with a spherical wave to represent the point source. The effect of a thin lens can be represented as a phase transformation which converts this diverging wavefront into a spherical wave centered on the image point. Eq. 29 is then used to model the propagation of this field at the lens aperture to the image plane. Upon making certain substitutions described in [18], it is seen that the impulse response is of the form [18]

$$h(s,t) = \frac{A}{\lambda z_i} \iint P(u,v) \exp\left[-i\frac{2\pi}{\lambda z_i}(su+tv)\right] dudv \tag{31}$$

This impulse response can be used as a Point Spread Function (PSF) to relate the diffraction-limited image to the ideal image predicted by geometric optics only for coherent, monochromatic illumination. Only under this condition do field strengths add linearly in order to make this approach valid. For this case, the diffraction-limited and geometric images are related by [18]

$$U_i(s,t) = \iint h(s-\xi, t-\eta) U_g(\xi,\eta) d\xi d\eta = h(s,t) * U_g(s,t). \tag{32}$$

For an incoherent imaging system, the property of linearity is observed by intensity rather than field strength. Therefore, it is necessary to employ a PSF which operates on intensity images, which is the square of the field impulse response [18]

$$I_i(s,t) = \iint |h(s-\xi, t-\eta)|^2 I_g(\xi,\eta) d\xi d\eta = |h(s,t)|^2 * I_g(s,t). \tag{33}$$

For the case of a circular aperture, the PSF is given by the airy disc pattern [19]

$$|h(s,t)|^2 = 4J_1^2\left(\frac{\pi\sqrt{s^2+t^2}}{\lambda f/\#}\right) \Bigg/ \left(\frac{\pi\sqrt{s^2+t^2}}{\lambda f/\#}\right)^2 \tag{34}$$

where J_1 is the first order Bessel function. Associated with the point spread function is its Fourier transform, known as the Optical Transfer Function (OTF). By the convolution theorem, the OTF operates by multiplying the spectrum of the ideal geometric image to give the spectrum of the diffraction limited image. For a circular aperture, the OTF is given by [19]

$$OTF^{1D}(k,k_0) = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}\left(\frac{k}{k_0}\right) - \frac{k}{k_0}\sqrt{1-\left(\frac{k}{k_0}\right)^2}\right] & : k \leq k_0 \\ 0 & : k > k_0 \end{cases} \tag{35}$$

where $k_0 = 1/(\lambda f/\#)$ is the cutoff frequency. Because the OTF for a circular aperture falls to zero for frequencies beyond this cutoff, these frequencies will not appear within the final image.

Within a conventional digital camera, high spatial frequencies are also filtered as a result of sampling of the image by discrete detector elements in the focal plane array. According to the Nyquist sampling theory, the samples of a signal spaced at p are sufficient for exactly reproducing a signal composed of frequencies lower than $1/2p$ [18]. If the original signal has frequencies higher than this cutoff, those frequencies will be 'folded' into lower ones as aliasing.

It is common to match the OTF cutoff frequency to the sampling cutoff frequency in order to avoid aliasing as well as oversampling [20]. Oversampling increases noise without, in most cases, providing an improvement in resolution. For a conventional camera, the matching condition is given by

$$k_{NYQ} = \frac{1}{2\Delta q} = k_0 = \frac{1}{\lambda f/\#}. \tag{36}$$

For a central wavelength, this equation defines a relationship between the lens diameter, focal length, and pixel pitch. Given the additional layer of microlenses within a

plenoptic camera, it is not surprising that the role of diffraction in general, and the interplay between sampling and MTF cutoffs in particular, is more complicated than for a conventional camera. Within the context of the plenoptic camera, we would like to match the MTF cutoff of the main lens to the cutoff of the microlens sampling, and the MTF cutoff of each microlens to the sampling cutoff of the underlying pixels.

As a first order approach to this problem, we assume that the two diffraction effects are decoupled, i.e., that spreading at the microlens plane does not impact spreading at the detector plane. Under this approximation, the effects of diffraction can are modeled via a 4D point spread function, given by the convolution of the 2D PSF associated with the main lens with the 2D PSF associated with the microlenses. Since the two PSFs are functions of independent variables, this results in

$$|h(s,t,u,v)|^2 = 16 \left[\frac{J_1^2\left(\frac{\pi D\sqrt{s^2+t^2}}{l_m\lambda}\right)}{\left(\frac{\pi D\sqrt{s^2+t^2}}{l_m\lambda}\right)^2} \right] \left[\frac{J_1^2\left(\frac{\pi \Delta s\sqrt{u^2+v^2}}{l_d\lambda}\right)}{\left(\frac{\pi \Delta s\sqrt{u^2+v^2}}{l_d\lambda}\right)^2} \right]. \tag{37}$$

The 4D optical transfer function is likewise given by

$$OTF^{4D}(k_s,k_t,k_u,k_v) = OTF^{1D}(\sqrt{k_s^2+k_u^2},k_{st}^0)OTF^{1D}(\sqrt{k_s^2+k_u^2},k_{uv}^0) \tag{38}$$

where 1D OTF is as defined in Eq. 35, and the cutoff frequencies are defined as $k_{st}^0 = D/l_m\lambda$ and $k_{uv}^0 = \Delta s/l_d\lambda$, according to the general definition $k_0 = 1/(\lambda f/\#)$. Ideally, these cutoff frequencies should be matched to the Nyquist cutoff frequencies associated sampling rates implied in Eq. 15, as in the following:

$$k_{st}^0 = k_{st}^{NYQ}, \quad k_{uv}^0 = k_{uv}^{NYQ}. \tag{39}$$

To allow for the possibility that these two constraints may not be simultaneously achievable, we introduce coefficients c_1 and c_2,

$$k_{st}^0 = c_1 k_{st}^{NYQ}, \quad k_{uv}^0 = c_2 k_{uv}^{NYQ}, \tag{40}$$

were $c_1 > 1$ implies undersampling at the microlens plane, which may lead to aliasing, and $c_1 < 1$ implies oversampling. The same holds true for c_2 with regard to the detector plane. Evaluating for the various cutoff frequencies, we get

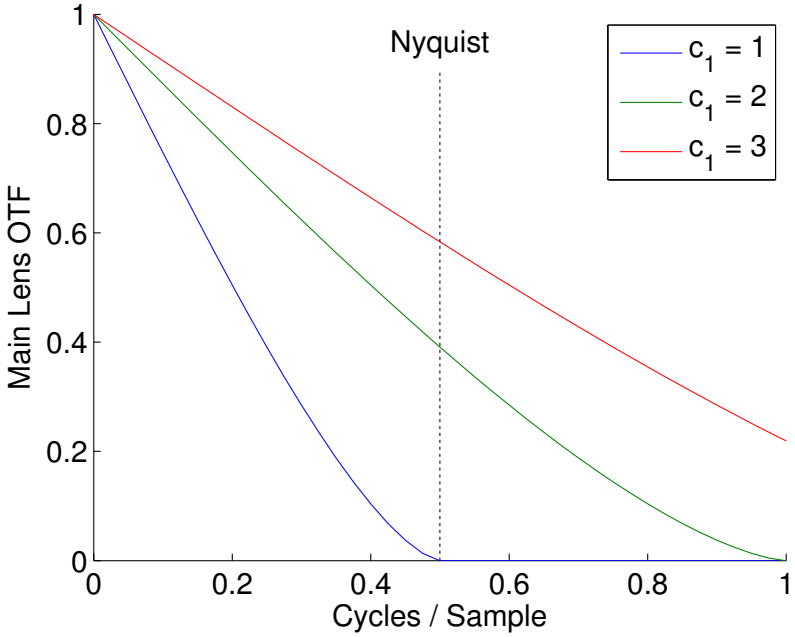
$$\frac{D}{l_m \lambda} = \frac{c_1}{2\Delta s}, \quad \frac{\Delta s}{l_d \lambda} = \frac{c_2}{2\Delta q} \tag{41}$$

Dividing the two equations gives, after canceling like terms and rearranging,

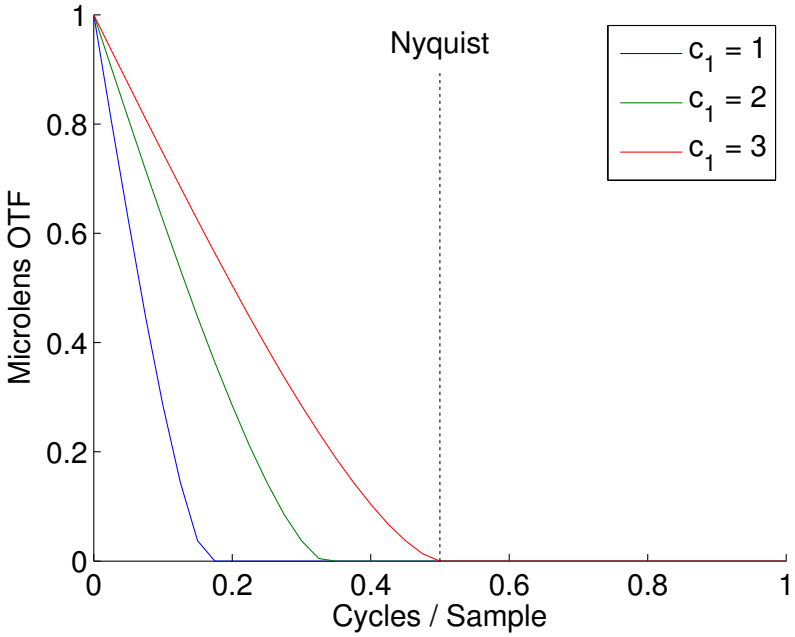
$$\frac{c_1}{c_2} = \frac{D}{\Delta q}\frac{l_d}{l_m} = \frac{D}{\Delta u} = N_u \tag{42}$$

which follows because $\Delta u = \Delta q l_m / l_d$ and $D = N_u \Delta u$. This result is very interesting because it means that, for a plenoptic camera, since $N_u > 1$, it is impossible for both c_1 and c_2 to equal unity, and thus it is impossible to simultaneously match the OTF cutoff of the main lens to the cutoff of the microlens sampling, and the OTF cutoff of each microlens to the sampling cutoff of the underlying pixels. Rather, it is necessary that there be some amount of undersampling at the microlens plane, oversampling at the detector plane, or both.

Fig. 11 graphically illustrates the problem for a plenoptic camera having $N_u = 3$ subapertures. At $c_1 = 1$, the main lens OTF cutoff is perfectly matched to the sampling cutoff (Nyquist) of the microlenses. However, the microlens OTF cutoff is short of the Nyquist rate for the detector array, indicating oversampling at the



(a)



(b)

Figure 11. Plenoptic Camera OTF: Relative Scale. The figure depicts how the main lens OTF and microlens OTF changes with respect to the Nyquist frequency as the parameter c_1 is altered for the a camera having $N_u = 3$ subapertures.

detector plane. At $c_1 = 3$, the microlens OTF is matched to the detector sampling cutoff. However, for this case, there is undersampling at the microlens plane.

Fig. 11 is not helpful for assessing what is the optimal value of c_1 and c_2 since it masks the fact that, in absolute terms, changing these parameters will impact the value of the Nyquist frequency for either the main lens or microlens, depending on how the change is effected. In order to examine actual performance, it is necessary to introduce the concepts of ground sampled distance (GSD) and ground spot size (GSS) [20]. These concepts reflect the fact that it is not spatial frequencies resolvable at the image that are important, per se, but rather the spatial frequencies at the object. In other words, these concepts account for the magnification of the imaging system.

In general, the GSD is defined as

$$\text{GSD} = p\frac{h}{f} \tag{43}$$

where p is the size of the pixel or whatever is performing the sampling, h is the distance to the object, and f is the focal length. Corresponding to the GSD is a ground sampled Nyquist frequency, k_N^G, defined as

$$k_N^G = \frac{1}{2\text{GSD}} = \frac{f}{2hp} = \frac{f}{h}k_N. \tag{44}$$

This relationship also applies for the ground sampled OTF cutoff frequency,

$$k_0^G = \frac{f}{h}f_0 = \frac{f}{h}\frac{1}{\lambda f/\#} = \frac{D}{\lambda h}. \tag{45}$$

For the plenoptic camera, cutoff frequencies at the target relate to cutoffs at the microlens plane resulting from microlens sampling and the main lens OTF. Matching cutoff frequencies gives

$$k_0^G = c_1 k_N^G, \quad \text{or} \quad \frac{D}{\lambda h} = c_1 \frac{l_m}{2h\Delta s}. \tag{46}$$

In the same way, cutoff frequencies at the detector plane brought about by detector sampling and the microlens OTF are related to spatial frequencies at the main lens plane. The L superscript is introduced in order to refer to this case:

$$k_0^L = c_2 k_N^L, \quad \text{or} \quad \frac{\Delta s}{\lambda l_m} = \frac{c_2}{2\Delta u} \tag{47}$$

.

We now wish to consider the case of a camera with fixed lens diameter. For this case, as illustrated in the equations that follow, the optical cutoff at the ground plane and the sampling cutoff at the main lens plane are fixed. Changing c_1 effects a change in the sampling cutoff at the ground plane and the optical cutoff at the main lens plane, and is achieved by altering the ratio of l_m to Δs.

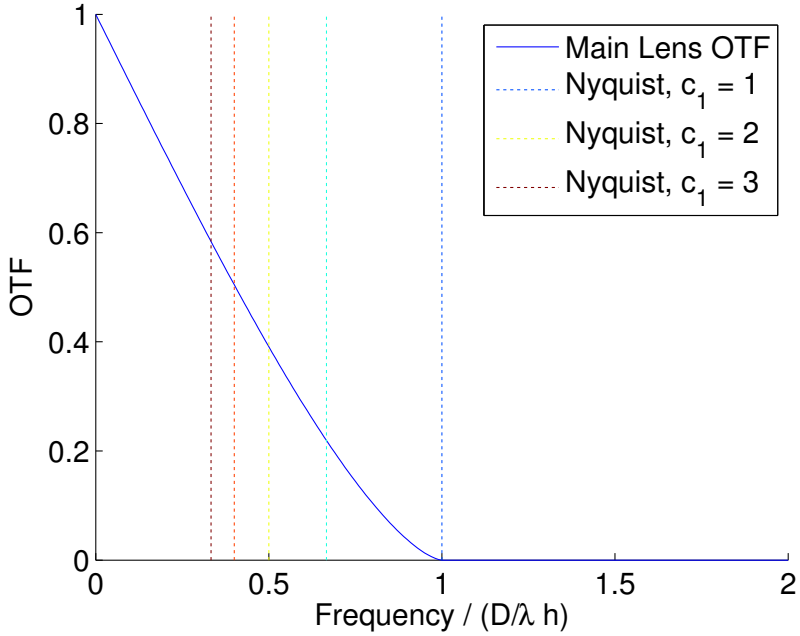
$$k_0^G = \frac{D}{\lambda h} \tag{48}$$

$$k_N^L = \frac{N_u}{2D} \tag{49}$$

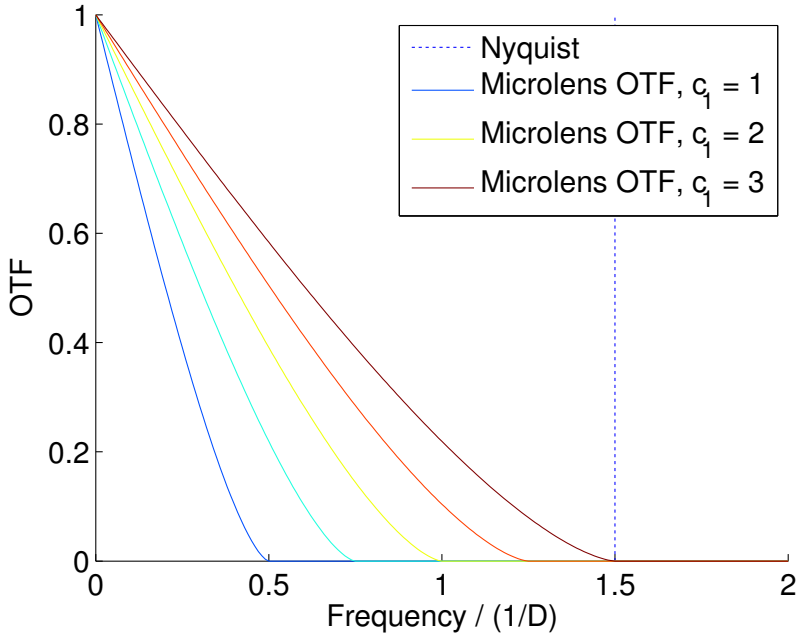
$$k_N^G = \frac{l_m}{2h\Delta s} = \frac{1}{c_1}\frac{D}{\lambda h} \tag{50}$$

$$k_0^L = \frac{\Delta s}{\lambda l_m} = \frac{c_1}{2D} \tag{51}$$

Fig. 12 shows how the various cutoffs vary in terms of the ground sampled frequency for a plenoptic camera with fixed lens diameter and $N_u = 3$. The figure



(a)



(b)

Figure 12. Plenoptic Camera OTF: Absolute Scale. The figure depicts how the Nyquist frequency and OTF change with the parameter c_1 for a camera with $N_u = 3$ subapertures and a fixed lens diameter D.

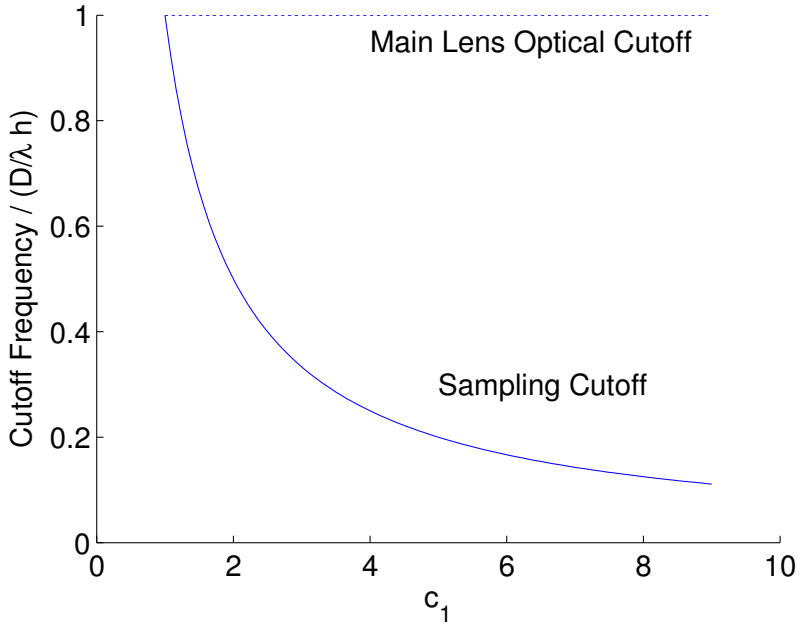
confirms that it is impossible to improve frequency in one domain without reducing it in the other.

Finally, Fig. 13 extends the picture for varying subaperture numbers (N_u). The figure illustrates that it is possible to achieve high angular sampling by increasing the number of subapertures and setting c_1 equal to the number of subapertures. However, as c_1 increases, the Nyquist frequency of the microlens sampling drops far below that of the main lens OTF cutoff. Because of these opposing behaviors, there is no preferred value of c_1 that gives overall best performance.

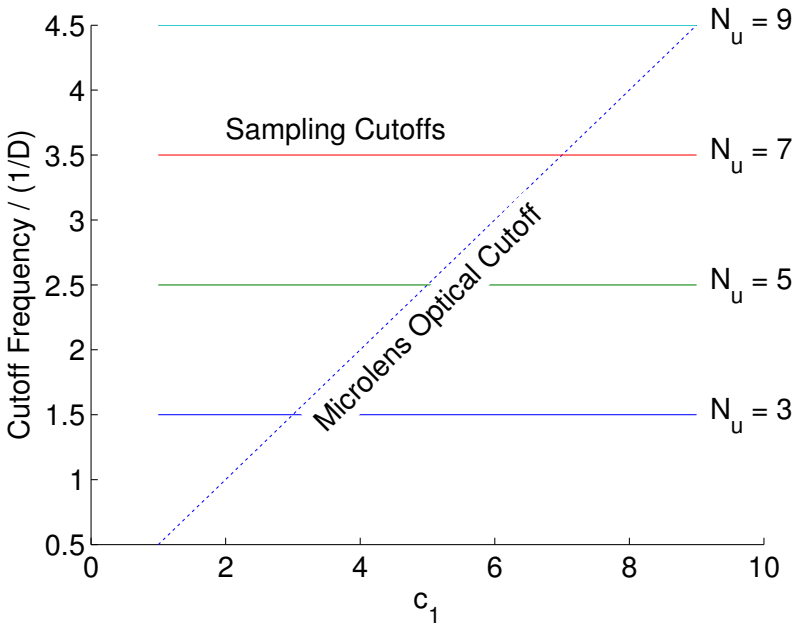
In future sections, it will be convenient to ignore the effects of diffraction, and to assume that sampling by the detector elements and microlenses constitutes the limiting factor impacting the precision of the camera's ranging performance. The preceding figures illustrate that as long as $c_1 \geq N_u$ or equivalently $c_2 \geq 1$ this approach is warranted, since where this is true, all Nyquist frequencies fall below OTF cutoff frequencies. Requiring that $c_2 \geq 1$ imposes a constraint on the relationship between the plenoptic camera $f/\#$ and the detector size. Namely, for the condition to be met, it must be true that

$$\Delta q \geq (f/\#)\frac{\lambda}{2} = \frac{l_m}{D}\frac{\lambda}{2}. \tag{52}$$

Fig. 14 provides a nomograph relating main lens diameter, focal length (l_m), and wavelength (λ) to the minimum pixel size satisfying Eq. 52. For optical wavelengths at f/#'s of interest, the mininum pixels sizes are small enough so as not be be a concern. This analysis does not deal with optical aberrations, whose impact is likely more critical in an optical system. However, it is worth noting that optical aberrations do not effect the location of the cutoff frequency of the OTF [18].



(a)



(b)

Figure 13. Plenoptic Camera OTF and Sampling Cutoff Frequencies. An illustration of how the optical and sampling cutoffs depend on c_1 for different numbers of subapertures N_u. If $c_1 \geq N_U$, oversampling will be avoided and the effects of diffraction can be safely ignored.

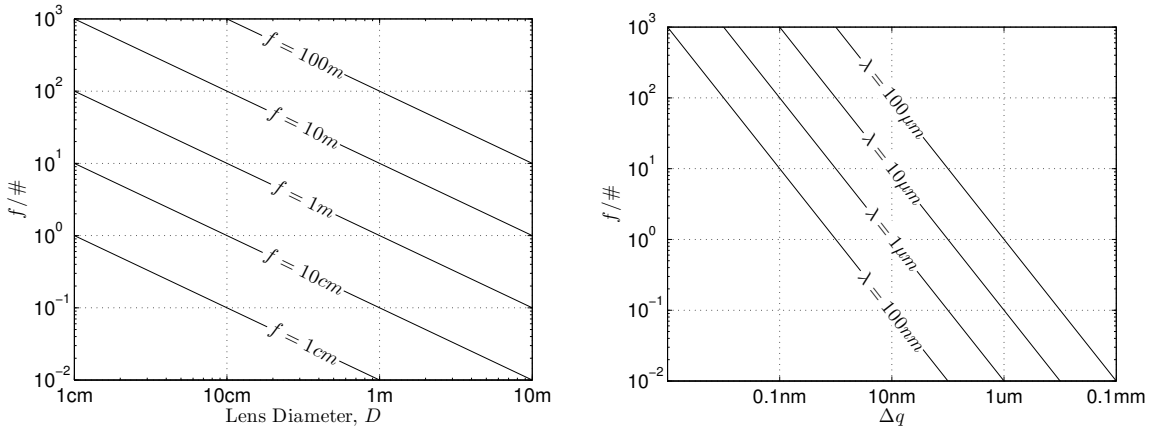


Figure 14. Plenoptic Camera Minimum Detector Size. Detectors smaller than the scale indicated by this nomograph will result in oversampling of the optical point spread function.

3.6 The Fourier Transformed Light Field

As noted previously, focused images of the type generated by a conventional camera can be obtained by projecting the light field down to two spatial dimensions by integrating over the angular dimensions, u and v (See Eq. 20). This projection may be preceded by a shearing step to control the depth at which objects appear in focus. This relationship bears a strong similarity to certain forms of medical imaging, in which x-ray attenuation provides a projection of the density distribution of a bone or tissue. Computed tomography is the process of using projections along different directions to reconstruct the original 3D density distribution. A common approach to this problem involves utilizing useful relationships between the density distribution and its rotated projections within various transformed domains.

The projection slice theorem defines this relationship for the Fourier domain. In its most basic 2D form, the theorem states that the sequence of projecting a 2D function along a line and then taking the 1D Fourier transform along that line is equivalent to the sequence of taking the 2D Fourier transform of the function and then extracting the 1D slice along the same line (see Fig. 15). An intuitive basis for the theorem is explained by Malzbender in [21]:

> Any point in the frequency domain corresponds to a sinusoid with some amplitude, phase,and orientation. If the sinusoid is not aligned with the projection direction, its projection will sum to zero. However, those components aligned with the projection direction sum to some finite value. This set of components with nonzero projections can be found in the frequency domain along a line perpendicular to the projection direction.

Ng et al. in [11] were the first to demonstrate the projection slice theorem's extension for use with the higher dimensionality of the light field. They discuss refocusing in the Fourier domain with continuous variables. While their discussion is useful for proving the validity of the concept, it does not address some of the details of a

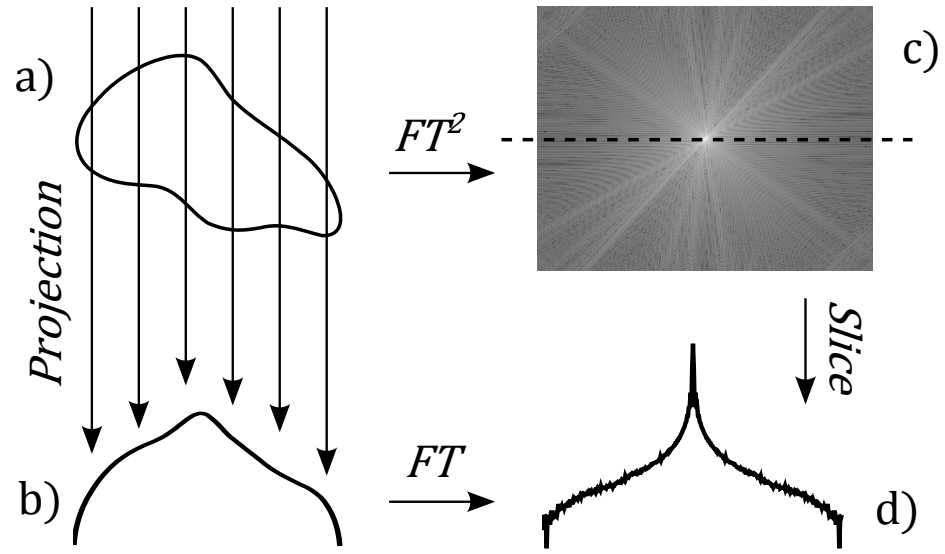


Figure 15. The Projection Slice Theorem. The projection slice theorem states that the projection operation in the spatial domain has a Fourier domain equivalent of taking the central slice perpendicular to the direction of projection. The spectrum in the lower right (d) can be obtained either by taking the 1D Fourier transform of the projection (b), or by extracting the slice along the dotted line in the spectrum (c).

practical implementation, which calls for the use of summations and discrete Fourier transforms. Here, we walk through the essentials of the math for the the discrete case, and show where it is important to modify the results provided in [11].

In order to proceed, we must augment the list of operators defined in the previous section. To begin, we define the spatial frequency variables, k_s, k_t, k_u, and k_v, and their normalized equivalents, $\bar{k}_s$, $\bar{k}_t$, $\bar{k}_u$, and $\bar{k}_v$, where $k = \bar{k}\Delta k = \bar{k}/(N-1)$. Nyquist for the two cases is defined as $k_N = \pm 1/2$ and $\bar{k}_N = \pm(N-1)/2$, respectively. We also allow for vector indexing using the definitions $\bar{\mathbf{k}} = [\bar{k}_s, \bar{k}_t, \bar{k}_u, \bar{k}_v]^T$ and $\bar{\mathbf{k}}_s = [\bar{k}_s, \bar{k}_t]^T$.

The 4D Discrete Fourier Transform is defined as

$$\mathcal{FT}^4[f(\bar{\mathbf{x}})](\bar{\mathbf{k}}) = \sum_{\bar{s},\bar{t},\bar{u},\bar{v}} f(\bar{s},\bar{t},\bar{u},\bar{v}) \exp\left[-2\pi i\left(\bar{k}_s\frac{\bar{s}}{N_s} + \bar{k}_t\frac{\bar{t}}{N_t} + \bar{k}_u\frac{\bar{u}}{N_u} + \bar{k}_v\frac{\bar{v}}{N_v}\right)\right]. \tag{53}$$

We use the letter G to refer to the Fourier transformed light field, as in

$$G(\bar{\mathbf{k}}) = \mathcal{FT}^4[K(\bar{\mathbf{x}})](\bar{\mathbf{k}}). \tag{54}$$

We also define a slice operator in the Fourier domain which returns the subspace obtained by setting $\bar{k}_u = 0$ and $\bar{k}_v = 0$, as in

$$\mathcal{S}[f(\bar{\mathbf{k}})](\bar{\mathbf{k}}_s) = f(\bar{k}_s, \bar{k}_t, 0, 0). \tag{55}$$

Appendix A shows that the sequence of shearing, projecting, and Fourier transforming is equivalent to the sequence of Fourier transforming, shearing, and slicing, i.e.

$$(\mathcal{FT}^2 \circ \mathcal{P} \circ \mathcal{B}_{\bar{m}})[f(\bar{\mathbf{x}})] = (\mathcal{S} \circ \bar{\mathcal{B}}_{\bar{m}}^{-T} \circ \mathcal{FT}^4)[f(\bar{\mathbf{x}})] \tag{56}$$

where

$$\bar{\mathcal{B}}_{\bar{m}} = \begin{bmatrix} 1 & 0 & -\bar{m}(N_u/N_s) & 0 \\ 0 & 1 & 0 & -\bar{m}(N_v/N_t) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{57}$$

From this, it follows by direct evaluation that

$$\mathcal{FT}^2[img(\bar{\mathbf{s}})](\bar{k}_s) = G\left(\bar{k}_s, \bar{k}_t, -\bar{m}\frac{N_u}{N_s}\bar{k}_s, -\bar{m}\frac{N_v}{N_t}\bar{k}_t\right). \tag{58}$$

This means that, in the frequency domain, a refocused image is formed simply by taking a 2D slice from the transformed light field, in contrast to the projection operation required in the spatial domain. The ideal image formation criterion derived in the previous section can be easily shown within the frequency domain. Outside a range of alpha values, k_s is cropped leading to lost high spatial frequency information.

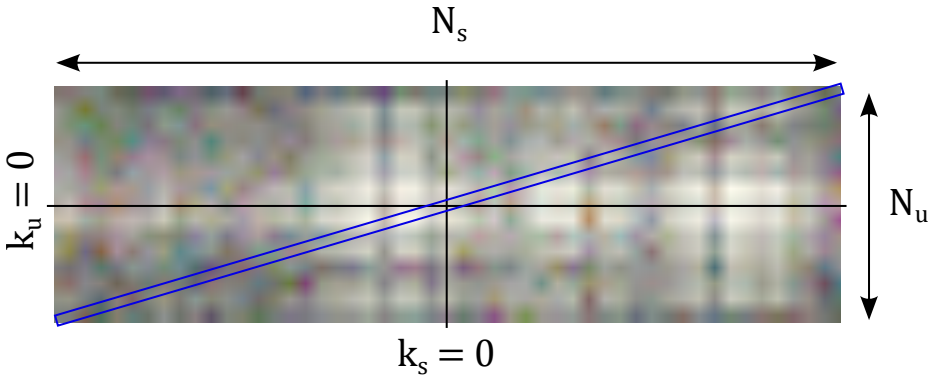


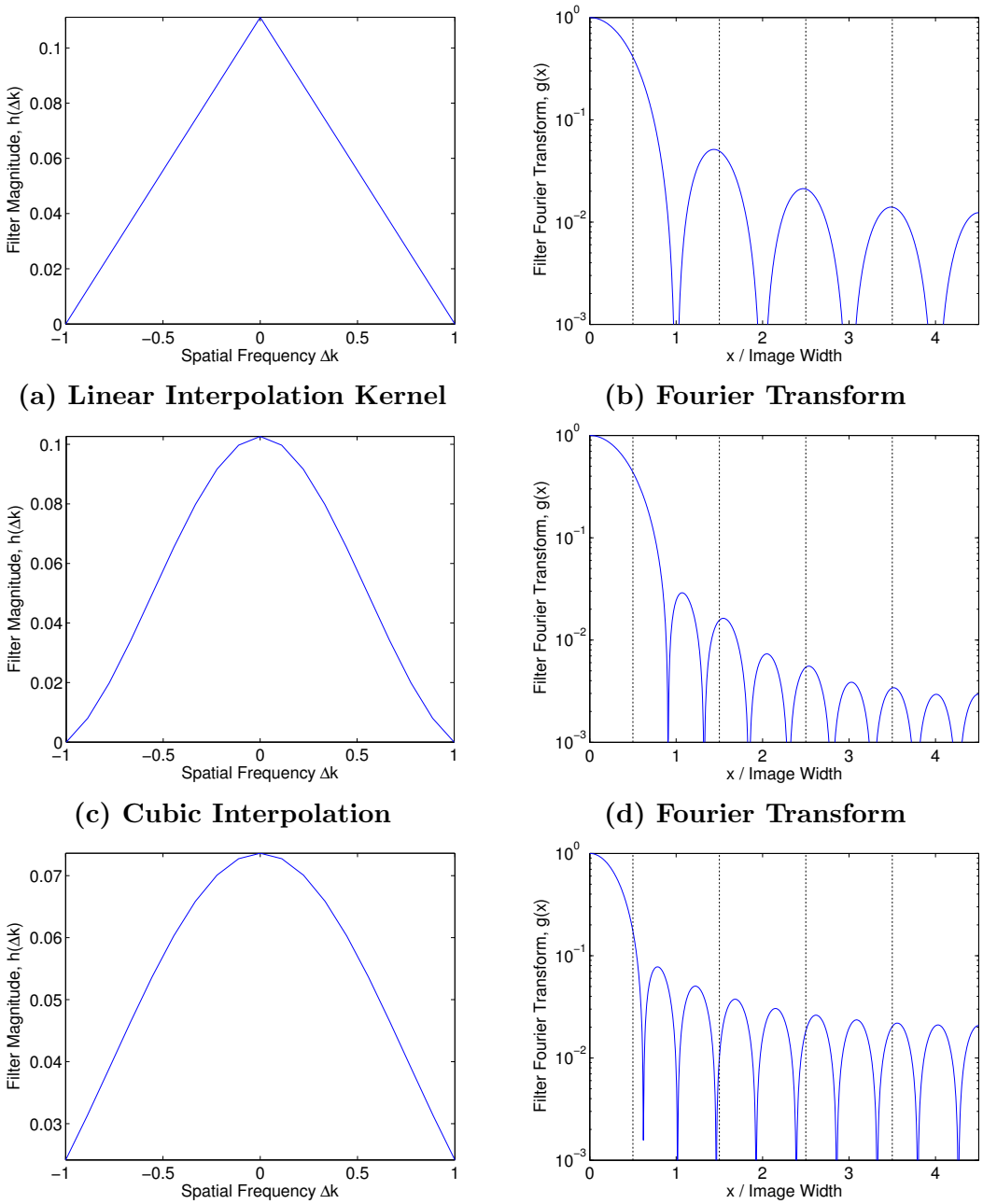
Figure 16. Fourier Slice Imaging. In the Fourier domain, an image is contained within a central slice of the light field. The figure illustrates the steepest slope at which slicing can occur before cropping of high spatial frequencies takes place.

In order to avoid cropping, k_u must remain less than the Nyquist limit of $N_u/2$ when $k_s = N_s/2$, as indicated in Fig. 16. By Eq. 58, this leads directly to the requirement that $\bar{m} \leq 1$, as discussed previously. The same result is derived with reference to the continuous light field in [11] under the assumption of band limited performance.

Interpolation is required in order to extract an arbitrary angled slice from the evenly sampled 4D light field space. This interpolation is best thought of as a reconstruction of the original continuous light field function from the sampled points, which can be represented as weighted delta functions within the continuous space. Reconstruction is achieved by convolving the gridded delta functions with some manner of interpolation filter.

Any finite impulse response (FIR) filter will have a Fourier domain transfer function of infinite extent. Fig. 17 shows the Fourier transform of some common interpolation kernels. When these kernels are used as interpolation filters in the Fourier domain, the tiled spatial image is multiplied by this Fourier transform. Regions where the transfer function is non-zero outside of the central tile of the spatial domain tend to show up as a faint shadowing or aliasing effect in the final image [22].

In principle, this problem is resolved by using the ideal SINC interpolation filter whose Fourier transform is a RECT function. The use of such a filter would eliminate



(a) Linear Interpolation Kernel

(b) Fourier Transform

(c) Cubic Interpolation

(d) Fourier Transform

(e) Kaiser-Bessel Reconstruction

(f) Fourier Transform

Figure 17. Interpolation Filter Performance. Interpolation can be envisioned as using a reconstruction filter to recreate the original continuous function from the sampled function, and then resampling at the new rate. Convolution in one domain is equivalent to multiplying by the Fourier transform of the convolution filter in the alternate domain. When the Fourier transform is nonzero outside of the central tile (indicated by the leftmost hashed line in plots b, d, and f), information from those regions appear as ghosting or aliasing within the final alternate domain image. Kaiser-Bessel interpolation is good for reducing aliasing because the central lobe of its Fourier transform can be adjusted to falloff close to the boundary of the central tile.

(a) **(b)**

Figure 18. Refocused Image Comparison. The image formed using cubic interpolation in (a) shows noticeable aliasing near the dots and edges of the die. These artifacts are reduced considerably by using Kaiser-Bessel interpolation (b).

aliasing by cropping out only the central tile of the spatial domain. To imitate SINC interpolation with a FIR filter, [22] describes a process of iteratively truncating within both domains until the resulting filter is sufficiently localized in each. The result of this process is known as the prolate spheroidal wave function (PSWF), and can be approximated by the Kaiser-Bessel function, which has the form

$$h(x) = \frac{I_0(\beta\sqrt{(}1 - (2x/w)^2))}{wI_0(\beta)}, \tag{59}$$

where β is an attenuation factor, w is the window width, and I_0 is the modified zero order Bessel function of the first kind [22]. Trial and error showed that a window of $w = 3$ and $\beta = 5$ gave good results for this application.

Fig. 18 compares the results of Fourier domain refocusing using Kaiser-Bessel reconstruction compared to cubic interpolation. Since the Fourier transform of the Kaiser-Bessel function drops off within the central image tile of the spatial domain, the refocused image will show a drop-off in intensity away from the center. Though [21] implies that this must be corrected by premultiplication of the Light Field by

the inverse of the filter Fourier transform, it was found that multiplication following image formation provided equal results and greater flexibility.

A worthwhile note with respect to implementation is that, in applications where the zero element of the image is located at the upper-left corner rather than the image center, it is often necessary to rearrange quadrants of the spatial domain prior to Fourier transforming [21].

3.7 Focused Plenoptic Camera Sampling

The previous sections deal with a camera in which the detector array is separated from the microlens array by one microlens focal length, so that the microlenses are focused at infinity. Since the main lens/microlens separation is large compared to the scale of the microlenses, this distance approximates optical infinity, and the microlenses can be thought of as being focused on the main lens. Thus, this arrangement results in a direct mapping between position on the detector array and position on the main lens plane.

If the detector array is placed at some distance from the microlens other than the microlens focal length, then the microlenses will image a plane other than that of the main lens. The arrangement has been referred to as the 'focused' plenoptic camera configuration [23].

Fig. 19 gives a diagram of a focused plenoptic camera, in which the image of the object (the arrow) exists at the same location as the conjugate plane of the detector array. In this case, each microlens reimages a region of the primary image. The spacing of the pixels beneath the microlens will determine how densely the primary image is spatially sampled within each microlens image. For the case of Fig. 19, we can imagine that the pixels are spaced so as to sample the primary image at the two locations shown.

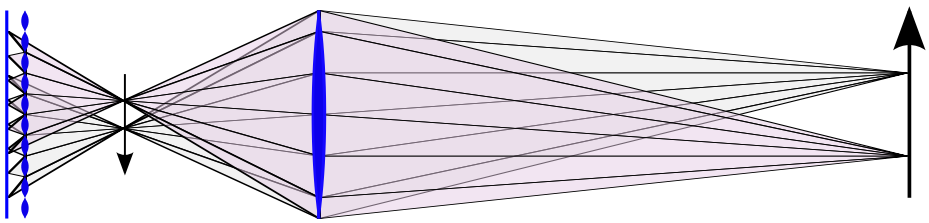


Figure 19. Focused Plenoptic Camera. In a focused plenoptic camera, the microlenses do not focus to the back of the main lens, but to a plane within the camera. If the main lens produces an image at this point, it will reimage to the detector plane.

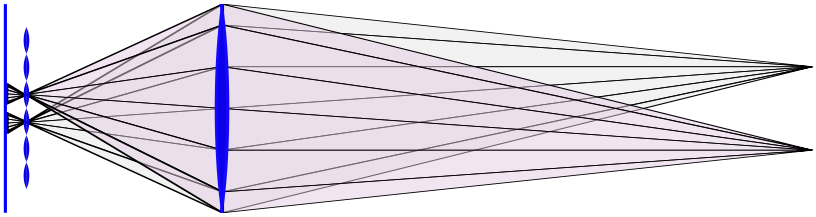


Figure 20. Conventional Plenoptic Camera Comparison. The sampling characteristics of a focused plenoptic camera can be closely mimicked by a conventional plenoptic camera with appropriately sized and positioned microlenses and detector elements.

Fig. 20 shows the setup of a conventional plenoptic camera, where the microlens plane has been placed at the plane that was conjugate to the detector plane of the focused plenoptic camera. Here, the size of the microlenses determines spatial sampling of the image and the subpixel spacing determines angular sampling. By choosing the correct microlens and pixel sizes, the figure suggests that a conventional plenoptic camera can achieve the sampling characteristics of a focused plenoptic camera, though the subsequence image formation process from the raw sensor data will be quite different.

In order to formalize this suggestion of equivalence, we examine the sampling patterns for the two types of cameras. First, we need to determine how the conventional plenoptic camera samples the light field at the main lens (u) plane and microlens (s) plane. For equivalence, this sampling must be matched (in terms of sampling density) by the sampling of the light field by the focused plenoptic camera at the main lens (u) plane and the plane conjugate to the detector plane (the s' plane in Fig. 21).

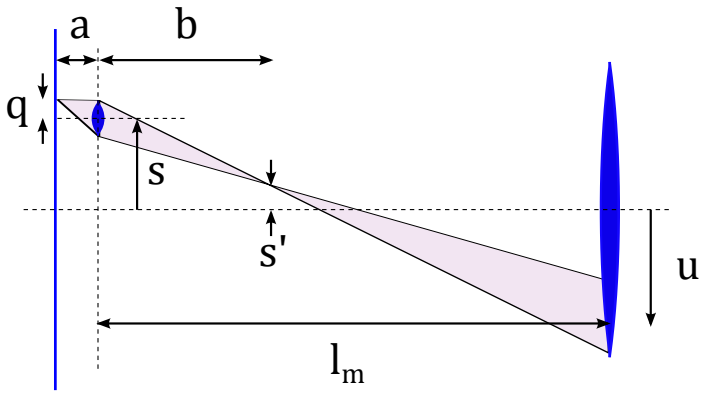


Figure 21. Focused Plenoptic Camera Geometry. Geometry of a focused plenoptic camera. The s' plane is the conjugate plane of the detector array.

Fig. 22 shows the sampling pattern for a conventional plenoptic camera. The light received by a pixel is constrained first by the stopping performed by the microlens at the microlens plane, and next by the spatial extent of the detector itself—or equivalently, by the projection of the detector at the main lens plane. The figure illustrates how sampling is performed at the microlens plane and at the main lens plane.

Fig. 21 provides the geometry necessary for determining the focused plenoptic camera sampling. Here, a is the distance from the microlens array to the detector array, and a and b are related by the lens equation. The conjugate plane of the detector plane is designated the s' plane. The figure illustrates a number of similar triangles formed by a ray of light passing through the camera. The dimensions of the triangles are related by

$$\frac{s - s'}{b} = \frac{s' - u}{l_m - b} = \frac{q}{a} = \frac{s - u}{l_m}. \tag{60}$$

For a single microlens (s fixed), we see that sampling in u (angular sampling) is dependent on pixel size, i.e. $\Delta u_s = l_m / a \Delta q$, where we use the s subscript to indicate that the s (the microlens location) does not change. Likewise for sampling in the s' plane (spatial sampling): $\Delta s'_s = b / a \Delta q$. To see how s' changes as we translate across

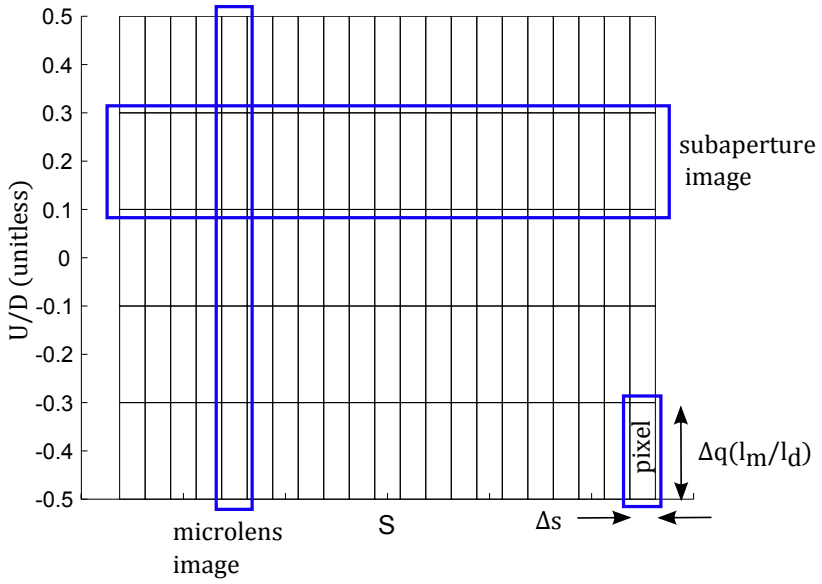


Figure 22. Conventional Plenoptic Camera Sampling. In a conventional plenoptic camera, sampling in the microlens plane is determined by the microlens size, Δs, while sampling in the main lens plane is determined by the magnified detector size.

microlenses, we solve Eq. 60 for s' in terms of s and u, to give

$$s' = u\frac{a}{l_m} + s\left(1 - \frac{a}{l_m}\right) \tag{61}$$

from which we see that $\Delta s'_u = \Delta s(1 - a/l_m)$. Fig. 23 shows the sampling pattern for the focused plenoptic camera, which illustrates these relationships.

To mimic the performance of a focused plenoptic camera with a traditional plenoptic camera with microlenses placed in the s' plane, it is simply necessary to ensure that its sampling density is the same as that of the focused plenoptic camera. Table 2 shows the sampling rates for the two variants, and Table 3 shows how parameters must be set within a conventional plenoptic camera to mimic the performance of a focused camera with given parameters.

Fig. 24 shows a possible matching of sampling patterns between a focused plenoptic camera and a conventional plenoptic camera. Notice how each conventional camera sample contains exactly one focused camera sample.

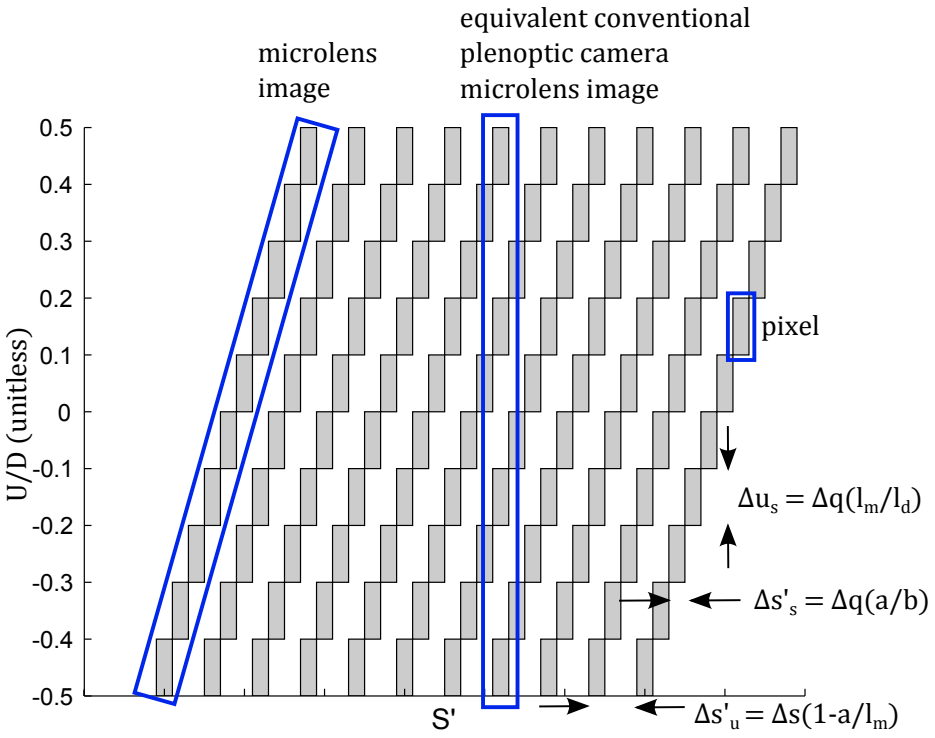


Figure 23. Focused Plenoptic Camera Sampling. If the light field is parameterized in terms of s' and u, focused plenoptic camera sampling is dependent on a number of parameters which can be adjusted to mimic traditional plenoptic camera sampling.

Table 2. Conventional and Focused Plenoptic Camera Sampling Densities

	Conventional Plenoptic Camera	Focused Plenoptic Camera
Spatial Sampling	$1/\Delta s$	$\frac{1}{\Delta s'_s} = \frac{b}{a}\frac{1}{\Delta q}$
Angular Sampling	N_u/D	$\frac{N_u}{D}\frac{\Delta s'_s}{\Delta s'_u} = \frac{N_u}{D}\frac{\Delta q}{\Delta s}\frac{a}{b}/\left(1-\frac{a}{l_m}\right)$

Table 3. Conventional and Focused Plenoptic Camera Equivalents

	Conventional Plenoptic Camera	Focused Plenoptic Camera
Microlens Size	$\Delta q\frac{a}{b}$	Δs
Numer of Subpixels	$N_u\frac{\Delta q}{\Delta s}\frac{a}{b}/\left(1-\frac{a}{l_m}\right)$	N_u
Subpixel Size	$\frac{\Delta s}{N_u}\left(1-\frac{a}{l_m}\right)$	Δq

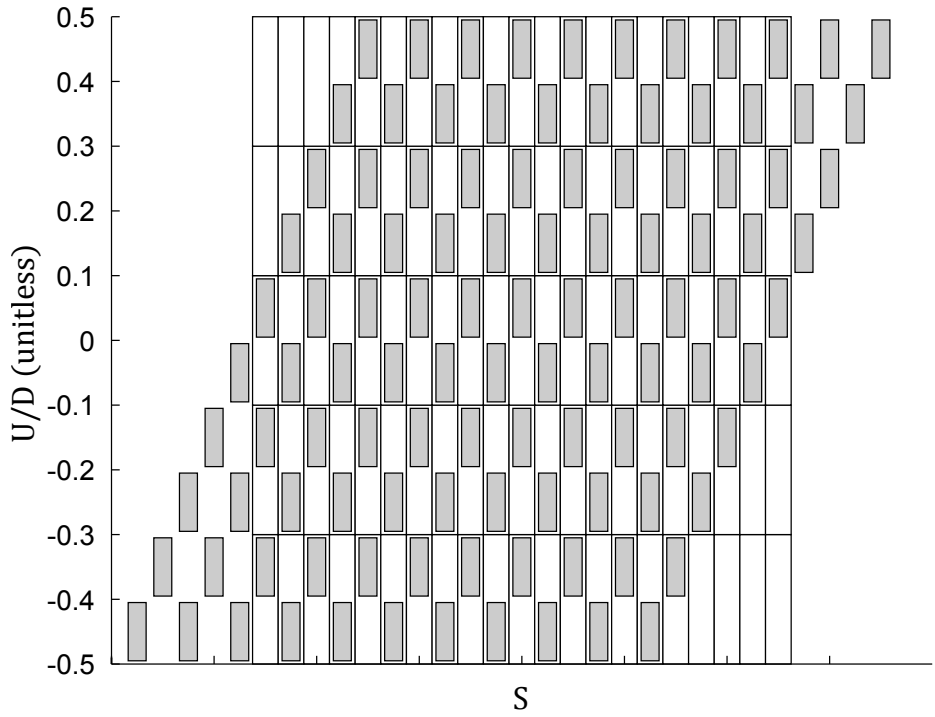


Figure 24. Matched Sampling Performance for Focused and Traditional Plenoptic Cameras: Traditional plenoptic camera and conventional plenoptic camera sampling patterns, matched via the equivalences in Table 3.

It is worth noting that the sampling of the two cameras is not identical in that the mapping that relates a sensor array location to a parameterized light field coordinate is much different for the two cameras. Techniques for generating refocused images directly from focused plenoptic camera data are discussed in [23] and [24]. Generation of a light field with a 2-plane parametrization is presented in [25]. Despite differences in data processing, the ability to create a conventional plenoptic camera which samples the light field with the same angular and spatial sampling densities as a focused plenoptic camera means that the focused plenoptic camera need not be treated as a separate case in the analysis presented in this paper. Rather, the expected performance for a given focused plenoptic camera may be determined by using the equations given here to determine the equivalent conventional camera, which can be used for further analysis.

IV. Plenoptic Ranging

4.1 Introduction

The fundamental result of the previous chapter is that a point in object space is represented by a 2D plane within the 4D light field. The orientation of the plane is directly related to the object's distance from the camera, as well as other fixed camera parameters. For the plenoptic camera, range finding is the operation of identifying these region and determining its orientation.

Eq. 8 shows that for an image located $z_i = \alpha l_m$ from the main lens, the light field will have a slope $m = ds/du$ given by

$$m = \frac{z_a}{z_i} = \frac{l_m - z_i}{z_i} = 1 - \frac{l_m}{f} + \frac{l_m}{z_o} \tag{62}$$

where z_o is the object distance, which is related to z_i by the lens equation (Eq. 6). We use the term 'sampled light field slope' to refer to the slope in terms of s samples per u sample, $\bar{m} = d\bar{s}/d\bar{u} = m\gamma$, where $\gamma = \Delta u/\Delta s$. By these relationships, a difference in distance δ_z is related to a difference in slope δ_m or in sampled slope $\delta_{\bar{m}}$ by

$$\delta_z = \frac{z_o^2}{l_m}\delta_m = \frac{z_o^2}{l_m\gamma}\delta_{\bar{m}}. \tag{63}$$

Uncertainty is related in the same manner,

$$\sigma_z = \frac{z_o^2}{l_m}\sigma_m = \frac{z_o^2}{l_m\gamma}\sigma_{\bar{m}}, \tag{64}$$

where σ_m and $\sigma_{\bar{m}}$ are the uncertainties associated with m and $\bar{m}$, respectively, and σ_z is the uncertainty associated with object distance.

In dealing with experimental results, uncertainty is determined by finding the mean square error (MSE) or root mean square error (RMSE) of the estimated quantity across the sample represented by a particular light field. Mean Squared Error is well understood to be defined as

$$\text{MSE}(\hat{x}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2 \tag{65}$$

where $\hat{x}_i$ is the estimated value and x_i is the true value. On the other hand, in the context of uncertainty modeling, variance is used to quantify uncertainty. The variance of an estimator is calculated typically in terms of the variance of some random variable incorporated into a simple light field model. Variance is defined as

$$\text{var}(\hat{x}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - \bar{x})^2 \tag{66}$$

where $\bar{x}$ is the mean value of the sample set. The two metrics are related by [26]

$$\text{MSE}(\hat{x}) = \text{var}(\hat{x}) + (Bias(\hat{x}, x))^2 \tag{67}$$

indicating that, for an unbiased estimator, the metrics should be equivalent. For this reason, the symbol σ is used in each context, whether to refer to RMSE or standard deviation. Throughout the chapter we will have occasion to employ a few properties of the variance. The first is a scaling property,

$$\text{var}(ax) = a^2\text{var}(x), \tag{68}$$

which follows directly from the definition of variance. The second is known as the Bienayme formula [27], which states that

$$\text{var}\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{N} \text{var}(x_i) \tag{69}$$

when the values of x_i are uncorrelated. We combine these two properties to say that $\text{var}(ax+by+c) = a^2\text{var}(x)+b^2\text{var}(y)$ where x and y are uncorrelated random variables and a, b, and c are constants.

The goal of this chapter is to obtain an expression for the uncertainty in the sampled light field slope, $\sigma_{\bar{m}}$, in terms of parameters intrinsic to the sampled light field, i.e. the number of angular samples, N_u, or the gradient of the sampled light field. An analysis of $\sigma_{\bar{m}}$ is particularly useful because the quantity should be independent of camera parameters such as microlens size, main lens diameter, etc. Thus, the analysis can be performed on a light field recorded by an arbitrary camera, and then extrapolated to other constructions via Eq. 64. In this chapter, we examine light fields from cameras having a variety of sampling characteristics to test whether the sampled slope uncertainty can truly be decoupled from camera parameters not intrinsic to the sampled light field.

Synthetic light fields are utilized extensively within this chapter due to the ease of obtaining ground truth depth and disparity information. Synthetic light fields are typically generated by some type of 3D rendering software, by translating a camera through a grid of positions to obtain the plenoptic camera's 'subaperture images.' One disadvantage of this method of simulation is that it does not naturally account for the spreading effects discussed in section 3.4.

The synthetic light fields utilized in this paper are generated using the 3D modeling software, Blender, and made available by the Heidelberg Collaboratory for Image Processing (HCI) [28]. A sample light field is shown in Fig. 25. Though vary-

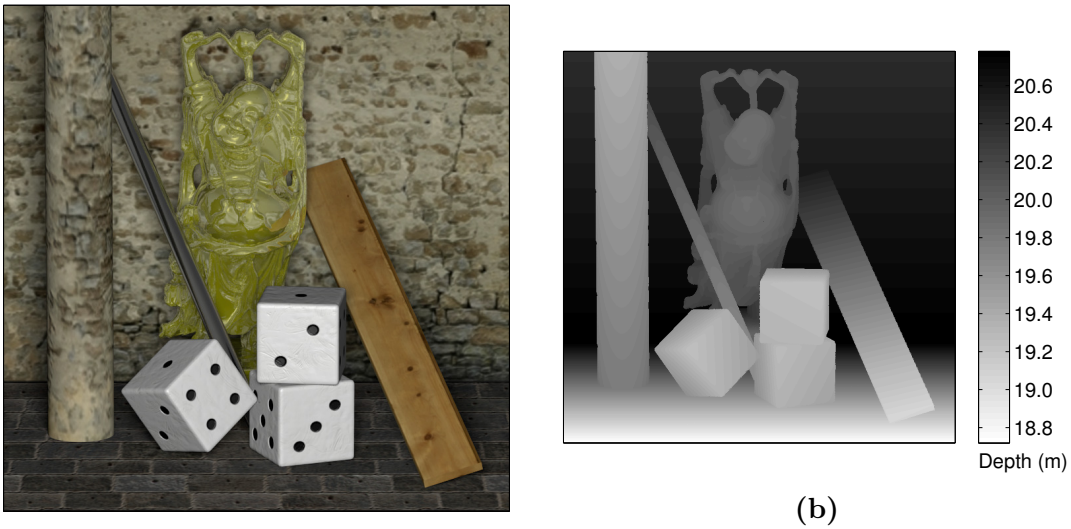


Figure 25. An HCI Lightfield. Depths are assigned physical units corresponding to the camera parameters given in Table 4.

ing camera sampling characteristics would be performed most naturally and without constraints by returning to the original Blender scenes, within this work this option was forgone for the simplicity of simulating tradeoffs by performing resampling of the full light fields. This method does not allow for the addition of information, so any tradespaces explored must involve courser sampling than the original rendered light field.

The HCI light fields have an angular resolution of 9×9 and spatial resolution of 768×768. Based on information provided about the setup of the Blender rendering environment, the light field sampling can be related to that of a plenoptic camera with the characteristics given in Table 4.

Table 4. HCI Light Field Camera Parameters

D	f	l_m	Δq	Δs	Δu	N_u	N_S	Ws
0.56m	0.95m	1m	15.5μm	138.9μm	6.25cm	9	768	10.7cm

Figs. 26 and 27 illustrate three ways in which the light field can be resampled in order to investigate the role of different parameters. 1) Simply cropping the light

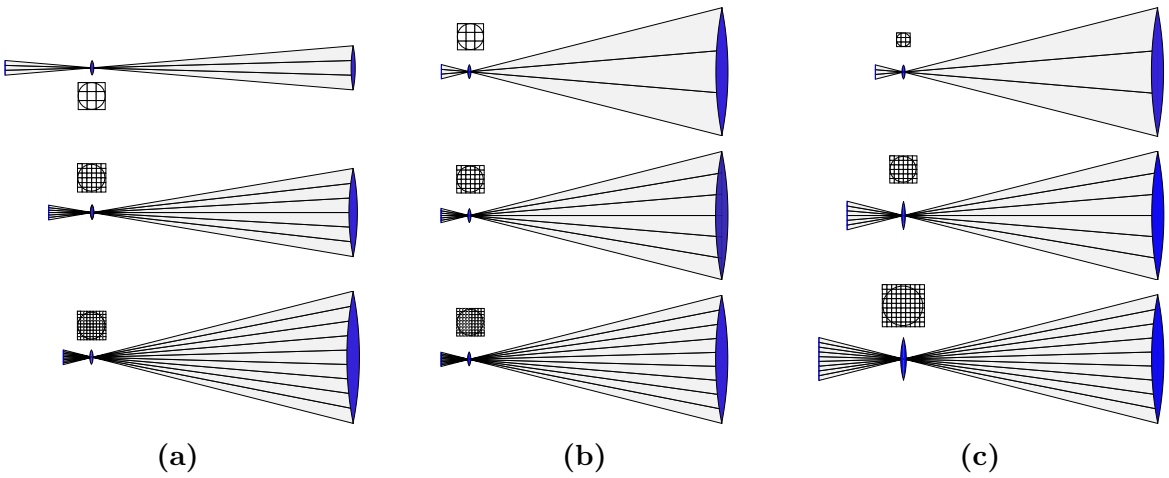


Figure 26. Plenoptic Camera Tradeoffs. The HCI lightfield can be resampled to investigate the role of various camera parameters. In (a), lens diameter and detector size are varied to increase N_u while keeping Δu and microlens size constant. In (b) N_u is increased by changing only detector size. (c) involves the tradeoff in spatial and angular resolution achieved by varying microlens size.

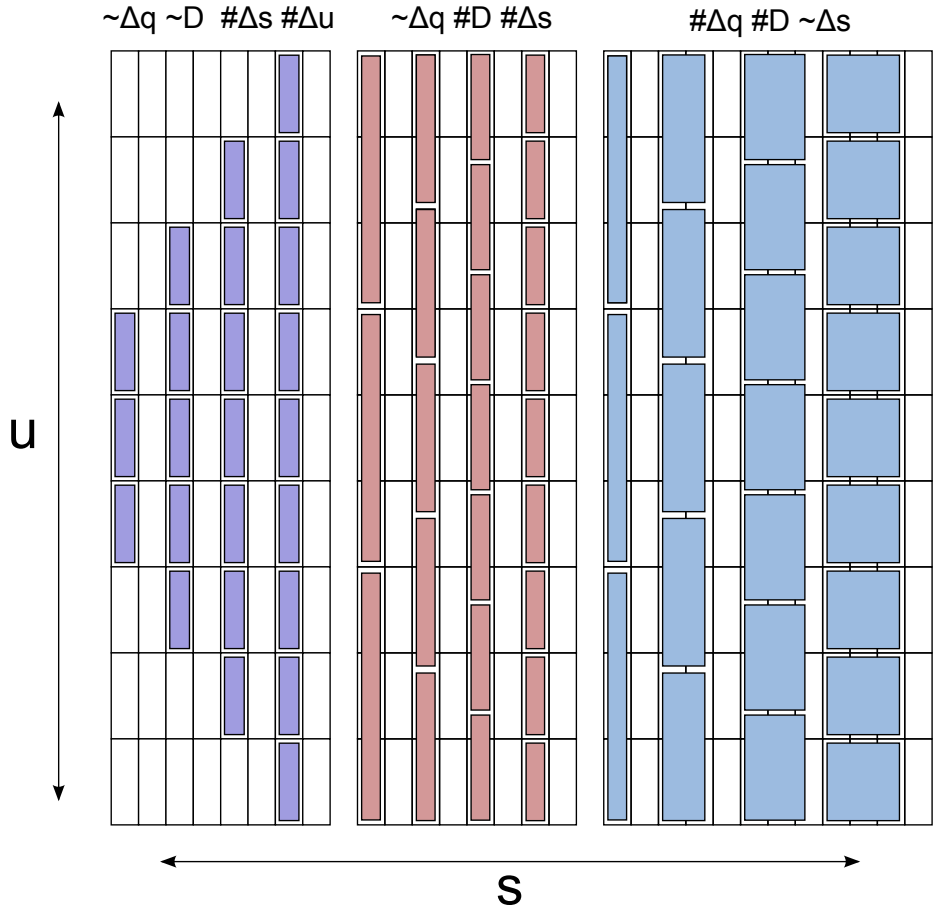


Figure 27. Tradeoff Sampling. Resampling required for the simulated cameras in Fig. 26.

field in the angular dimensions by successively decreasing amounts is equivalent to expanding the camera diameter while keeping microlens size constant and decreasing detector size. 2) Holding all parameters constant while changing pixel size can be simulated by resampling within only the angular dimensions. 3) Changing the microlens diameter results in a tradeoff between angular and spatial sampling rates. This tradeoff is simulated by resampling in both the spatial and angular dimensions. In resampling, it is crucial that proper interpolation be employed to eliminate aliasing. Here, low pass Gaussian filtering was employed to remove all frequencies above Nyquist prior to downsampling via nearest-neighbor interpolation.

Three slope estimation frameworks are examined in the chapter. The first utilizes a feature matching algorithm to determine correspondences between images, resulting in a sparse 3D point cloud. The second approach can be thought of as an extension of traditional image correlation techniques to the expanded light field space. It looks for minima in the variance calculated along different slopes within the light field. Finally, a Fourier domain ranging technique is explored, and its performance is evaluated.

4.2 3D Point Clouds using Feature Matching

Image registration provides one avenue of approach to the depth estimation problem. The major elements of an image registration algorithm are a feature detector and descriptor. The existence of a feature detection step sets this method apart from many of the others to be discussed. Having such a step means that the resulting depth map will be to some degree sparse–i.e., a depth estimate will not be generated for every pixel in a rendered image of the scene. The advantage closely related to this is that, once a feature has been detected, it is typically a small matter to estimate its location with a sub-pixel level of accuracy. For example, given a line of pixels identified by an edge detector to constitute an edge, the location of the edge in sub-

pixel space can be estimated with a simple linear fit. An approach similar to this is described below for the feature detection employed in this section.

Once a collection of features has been identified, a descriptor vector is generated for each feature using the region surrounding the feature within the image. The descriptor must capture the salient attributes of the surrounding to give a distinctive description, capable of distinguishing the feature from all others detected within the scene. These descriptor vectors are then matched with other descriptor vectors using Euclidean distance, spectral angle, or some other classifier, to establish a mapping between the two images. Image registration algorithms are often designed to be robust to translations, rotations, and scalings of an original image. Thus, it is very important that the detector be able to identify the same features within an image under these effects.

This section provides a theoretical framework for predicting the expected uncertainty within the context of feature matching. Features within separate images are assumed to be correctly detected and matched, such that any error results from feature localization error. Feature localization is treated alternately with the assumption of pixel level accuracy and the assumption of error with a normally distributed probability density function.

Quantization Error.

The case of feature localization to within pixel accuracy can be treated by the model,

$$s_i = mu_i + e, \tag{70}$$

where e is uniformly distributed over $(-\Delta s/2, \Delta s/2)$.

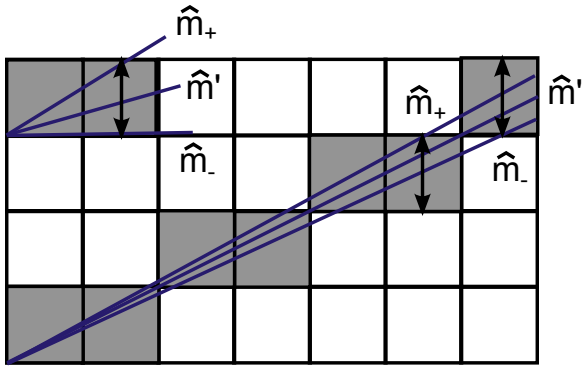


Figure 28. Quantization Error Visualization. The figure shows the range of possible slopes when a line is known with pixel-level precision.

For such a case, $\hat{m}$ is a maximum likelihood estimator of m if and only if $\hat{m}_- < \hat{m} < \hat{m}_+$, where

$$\hat{m}_+ = \min\left(\frac{s_i + \Delta u/2}{u_i}\right) \tag{71}$$

and

$$\hat{m}_- = \max\left(\frac{s_i - \Delta u/2}{u_i}\right) \tag{72}$$

give the extrema of slopes falling within the bounds of the error distributions, as shown in Fig. 28. The estimator $\hat{m}' = (\hat{m}_+ + \hat{m}_-)/2$ has certain optimality properties, discussed in [29]. Its uncertainty is given by $\Delta m = \hat{m}_+ - \hat{m}_-$.

Normally Distributed Error.

The assumption of normally distributed error is useful because it leads to a clean analytic result. The model underlying this section is given by

$$s_i = mu_i + n, \tag{73}$$

where n is a zero-mean normally distributed random variable with variance σ_n^2, and $u_i = i\Delta u$, where $i \in \{1, N_u\}$. Given this model, simple linear regression provides the optimal estimator for m, with $\hat{m} = \mathrm{cov}(s,u)/\mathrm{var}(u)$, where the variance and

covariance are well understood to be defined as

$$\mathrm{cov}(s,u) = \frac{1}{N_u}\sum_{i=1}^{N_u}(u_i - < u_i >)(s_i - < s_i >) \tag{74}$$

and

$$\mathrm{var}(u) = \frac{1}{N_u}\sum_{i=1}^{N_u}(u_i - < u_i >)^2 \tag{75}$$

where the angular brackets are used to denote the mean of the enclosed variable. The covariance can be rewritten by substituting in Eq. 73, as in

$$\mathrm{cov}(s,u) = \frac{1}{N_u}\sum_{i=1}^{N_u}(u_i - < u_i >)(mu_i + n_i - m < u_i >). \tag{76}$$

Upon factoring, this gives

$$\mathrm{cov}(s,u) = \frac{1}{N_u}\left[m\sum_{i=1}^{N_u}(u_i - < u_i >)^2 + \sum_{i=1}^{N_u}(u_i - < u_i >)n_i\right]. \tag{77}$$

Substituting these expressions for covariance and variance into the equation for $\hat{m}$, we obtain an updated expression for the slope estimator:

$$\hat{m} = \mathrm{cov}(s,u)/\mathrm{var}(u) = m + \frac{\sum_{i=1}^{N_u}(u_i - < u_i >)n_i}{\sum_{i=1}^{N_u}(u_i - < u_i >)^2}. \tag{78}$$

Employing the fact that $\mathrm{var}(ax+by+c) = a^2\sigma_x^2 + b^2\sigma_y^2$ if a, b, and c are constants and x and y are random variables, as discussed in the chapter introduction, the variance of the estimator directly reduces to

$$\mathrm{var}(\hat{m}) = \frac{\sigma_n^2}{N_u \mathrm{var}(u)}. \tag{79}$$

Assuming that N_u is odd, the denominator can be written as

$$N_u \text{var}(u) = \sum_{i=1}^{N_u} \left(u_i^2 - < u_i >\right)^2 = 2\Delta u^2 \sum_{i=1}^{(N_u-1)/2} i^2 \tag{80}$$

where we have used the definition of u_i and a reindexing to provide an equivalent expression. The series $\sum_{i=1}^{n} i^2$ is a square pyramidal number having a known analytic sum of $n(n+1)(2n+1)/6$ given by Faulhaber's formula [30]. This substitution allows for more convenient expression as

$$N_u \text{var}(u) = \Delta u^2 N_u(N_u - 1)(N_u + 1)/12 \approx D^2 N_u/12. \tag{81}$$

This definition can be substituted back into Eq. 79 to obtain a final expression for the uncertainty in m:

$$\sigma_m = \frac{\sigma_n}{D}\sqrt{\frac{12}{N_u}}. \tag{82}$$

The result in terms of the sampled light field error is obtained via $\sigma_{\bar{m}} = \gamma\sigma_m = \sigma_m \Delta u/\Delta s$.

$$\sigma_{\bar{m}} = \frac{\sigma_n/\Delta s}{N_u}\sqrt{\frac{12}{N_u}} = \frac{\bar{\sigma}_n}{N_u}\sqrt{\frac{12}{N_u}} \tag{83}$$

where $\bar{\sigma}_n$ becomes the registration error as a fraction of pixel size.

The preceding calculations apply to the case where the slope is estimated from a single 2D slice of the 4D light field where t and v are fixed. In evaluating the fundamental performance limitation for estimating from the full 4D space, we assume that N_u samples in u are available for each of $N_v = N_u$ values of v. Working through the same process for $u_i = \Delta u\{1, 1..., 2, 2..., N_u, N_u...\}$ for a total of N_u^2 samples gives an improved uncertainty,

$$\sigma_m = \frac{\sigma_n}{D}\frac{\sqrt{12}}{\sqrt{N_u^2 - 2}} \approx \frac{\sigma_n}{D}\frac{\sqrt{12}}{N_u}. \tag{84}$$

Once again, the error in terms of the sampled light field is given by

$$\sigma_{\bar{m}} = \bar{\sigma}_n \frac{\sqrt{12}}{N_u^2}. \tag{85}$$

Stereo Ranging with Normally Distributed Error.

This section provides a simple framework for comparing predicted performance between a plenoptic and stereoscopic system. Fig. 29 shows the geometry of the system to be considered. Within such a stereo vision system, the location of an object within each camera's image specifies a line traveling out from the camera into object space. These two lines form a simple linear system which can be solved to give a depth estimate in terms of the disparity, d, in the image location between the two cameras,

$$z_{est} = Bf/d = f/m \tag{86}$$

where B is the baseline separating the two camera axes and f is the focal length of the pinhole cameras [1]. For simpler comparison with the plenoptic camera, we have reformulated the result in terms of a slope, $m = d/B$.

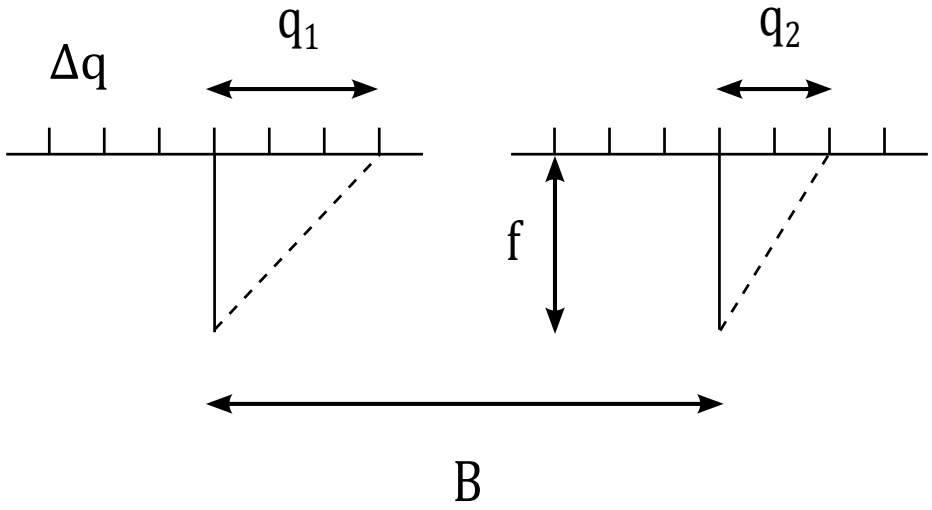


Figure 29. A Simple Stereo Ranging Setup. The diagram represents two pinhole cameras with parallel optical axes separated by distance B.

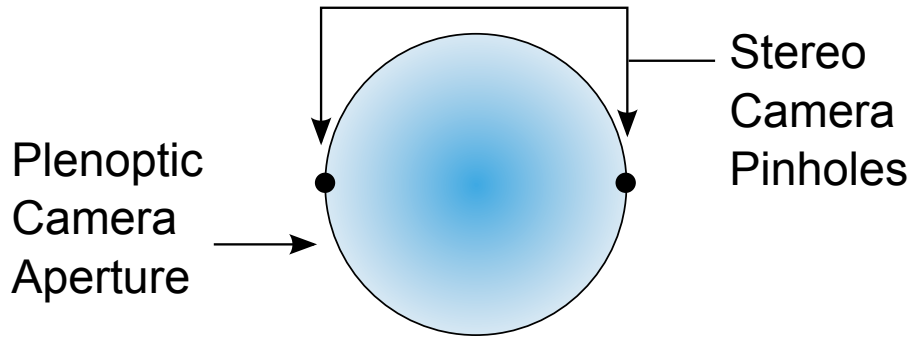


Figure 30. Equal Baseline for Stereoscopic and Plenoptic Systems.

We assume that feature registration between the two images is performed with some normally distributed error. That is, $d = q_1 - q_2$, where q_1 and q_2 are random variables with normally distributed probability density functions (PDF) having standard deviation σ. The PDF of the sum of two random variables is equal to the convolution of the original PDFs. Since the convolution of two Gaussian distributions is a third Gaussian distribution, having $\sigma_3^2 = \sigma_1^2 + \sigma_2^2$, it follows that the PDF of d is normally distributed with a variance of $\sigma_d^2 = 2\sigma^2$. The uncertainty in the slope, m, expressed as a standard deviation, is then given by

$$\sigma_m = \frac{\sigma_d}{B} = \sqrt{2}\frac{\sigma}{B}. \tag{87}$$

This equation will be useful in evaluating results within the next section.

An interesting comparison involves the case where the stereo baseline, B, is equal to the plenoptic lens diameter D (See Fig. 30), and both systems have the same pixel size, Δq. We assume that the feature localization error, σ is proportional to Δq for the stereo system and Δs for the plenoptic camera. Using Eqs. 82 and 87, the uncertainties are related by

$$\frac{(\sigma_m)_{plen}}{(\sigma_m)_{ster}} = \sqrt{6}. \tag{88}$$

This comparison gives some sense of the advantages and disadvantages of each system. In general, both types of systems appear to operate on the same general playing field. The plenoptic camera has the advantage of being a monocular system with a minimal hardware requirement. Its primary disadvantage is that its performance is coupled to lens size, which is more limited than the camera baseline of the stereo system. Since this equation is derived within the context of feature matching, it also does not take into account the benefits of other approaches to light field ranging to be discussed in later sections, which improve upon feature matching performance.

Estimator Comparison.

Fig. 31 compares the estimators described in this section for the cases of quantization error and normally distributed error. The figure illustrates that, for the case when feature location is known within pixel accuracy, the maximum likelihood estimator is superior until there are more than about 20 angular samples. In general, the stereo estimator is much worse for this case.

When the feature location estimate is subject to a normally distributed error term, there is very good agreement between the error obtained via simulation and the expected error derived previously. The error resulting from SLR estimation falls off noticeably faster with the number of angular samples than the stereo estimation.

An unexpected result is that, in the case of normally distributed error, it is possible to achieve better performance than obtained using simple linear regression by using a modification of the maximum likelihood estimator for the case of uniform error. The limits in Eqs. 71 and 72 represent the constraints on possible slopes imposed by the collective uncertainty limits of the data points. In the context of normally distributed error, this is not a meaningful concept, as no slope is impossible, however improbable. By scaling the Δu term in each of the equations, we can select by trial

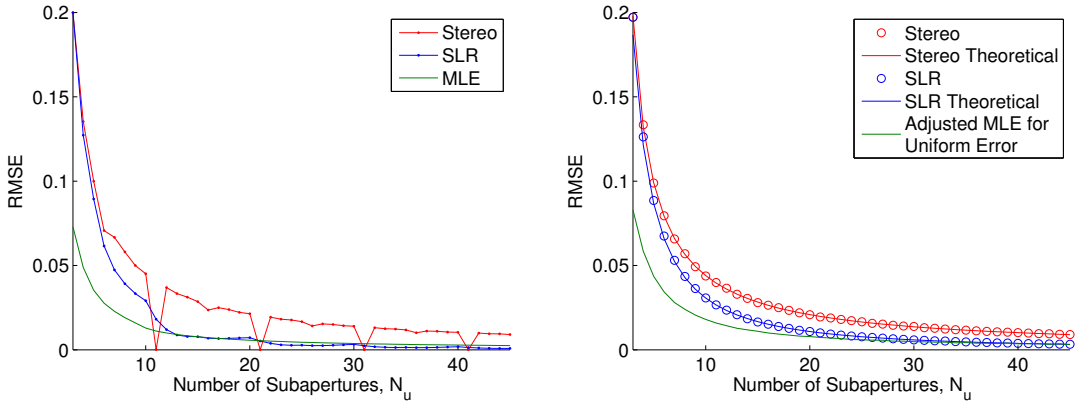


Figure 31. Slope Estimator Performance. Plot (a) applies the Maximum Likelihood Estimator (MLE), Simple Linear Regression (SLR), and Stereo estimator to the case of a rasterized line $s = round(mu + b)$, where m was sampled at 1000 evenly spaced points between 0 and 100, and b was sampled at 1000 points between -1 and 1. The MLE is superior at small angular sample numbers, but increasingly gives out to the SLR at higher numbers. Plot (b) gives the theoretical and simulated Root Mean Square Error for the case of a line with normally distributed error, $s = mu+n$, where $\sigma_n = 0.28$ in order to match the standard deviation of a uniform distribution of unit width. Since changing m was not found to affect results, m was fixed at 0.1. The RMSE was calculated over 10000 samples for each point. The agreement between the simulation and theory is strong for both the SLR and stereo case, indicating that the formulas derived in the previous sections are valid. An unexpected result is that a modification of the MLE for uniform error gives improved performance over the Simple Linear Regression.

and error the range which gives the best slope estimation performance. The improved performance in Fig. 31 (b) was obtained by using a multiplicative factor of 2. The improvement in performance over the SLR estimator is likely due to the fact that the MLE estimator as we have presented it uses knowledge about the zero u-intercept of both of the models discussed in this section, whereas the SLR method assumes that the intercept is unknown. In reality, the intercept will be some unknown, non-zero value. We expect that, if an estimator for the case of unknown intercept, as described in [29], were to be employed, its performance would not exceed that of the simple linear regression. For this reason, the simple linear regression estimator is employed within the next section.

4.3 The Scale Invariant Feature Transform

The Scale Invariant Feature Transform (SIFT), as described by David Lowe, is one common algorithm for matching features between images taken of the same subject [31], and has become the present 'gold bar' standard for image registration within the computer vision community. This section describes the SIFT algorithm.

SIFT Feature Detection.

In order to achieve scale invariant feature detection, SIFT first generates a scale space representation of an image. Scale space adds an additional parameter, σ, to an image, $im(x, y)$, such that $im(x, y, \sigma)$ gives a blurring of the original image to the point where only features on the scale of σ can be discerned. Blurring is performed using a Gaussian filter, which allows repeated convolutions to be efficiently utilized to keep kernel size small even as σ grows large. Downsampling at each doubling of σ is also used for the same purpose.

Feature detection within scale space is performed using a edge detector known as the Difference of Gaussians (DoG). The DoG representation of an image is obtained by subtracting the image at one scale, $im(x, y, \sigma)$, from an image at a larger scale, $im(x, y, k\sigma)$.

The Gaussian convolution involved in the generation of scale space is equivalent to low pass filtering of the image. When one low pass filter is subtracted from another, wider low pass filter, the result is a band pass filter. Thus, the response of the DoG filter will be highest to those features whose scale puts them within the passband of this filter. A point is considered a local extreme when it is uniformly larger or smaller than all 26 of its nearest neighbors within scale space.

For a simple analytic example, we can consider a feature consisting of a Gaussian blob with a standard deviation σ. The Gaussian filter used in scale space generation starts at σ_0 and grows by a factor of $k = 2^{1/S}$, where S is the number of steps per octave. The scale space representation is then given by a Gaussian with variance $\sigma_s^2 = \sigma^2 + \sigma_0^2 k^{2n}$, which at its peak has a value of

$$S_n = \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_0^2 k^{2n})}}. \tag{89}$$

Setting the second derivative of this expression equal to zero gives the location where difference $S_n - S_{n+1}$ reaches an extrema. This is easily shown to result in

$$\sigma_0 k^n = \sqrt{2}\sigma, \tag{90}$$

indicating that the DoG will reach an extrema where the scale space scale is on the order of a feature scale. This is illustrated graphically in Fig. 32. The first column shows different layers of the scale space representation of a Gaussian blob with a standard deviation of 10. The second column shows the difference taken between

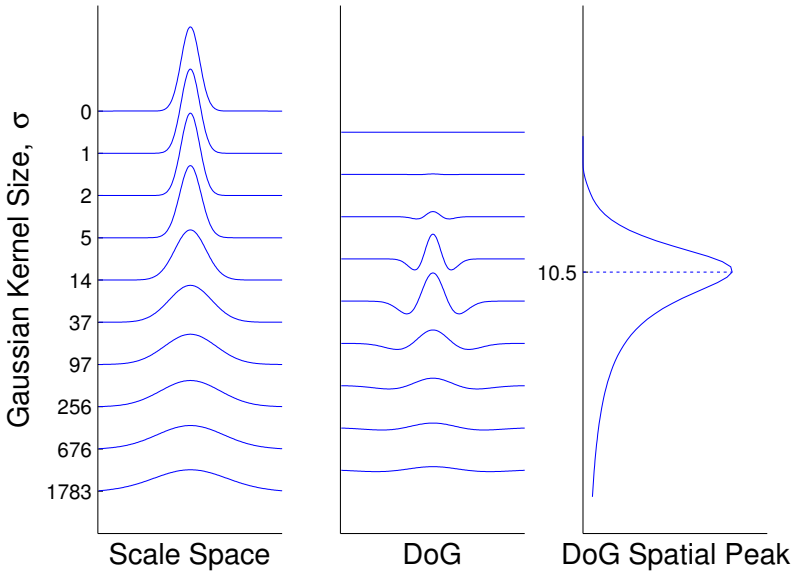


Figure 32. Difference of Gaussians Feature Detector. The first column shows the scale space representation of a feature consisting of a Gaussian 'blob' with standard deviation 10. The DoG spatial peak reaches a maximum when scale spaces used to form the DoG are near the scale of the original feature.

successive layers of scale space. The peak of the difference is plotted in the final column, and is seen to reach a maxima where the scale space σ is near to that of the original feature.

SIFT Descriptor.

When features are detected at a particular scale, a feature descriptor is compiled based on the feature's surroundings at that scale. This ensures that those same surroundings will be used to build a descriptor in any other image at an arbitrary scale where the same feature is detected. Since SIFT is meant to accommodate the possibility of such changes in scale between images, all features detected in one image, regardless of scale, will traditionally be compared with all features of another image during the matching stage.

The SIFT descriptor uses image gradient information to characterize the local neighborhood of a feature. The gradients calculated within an N_x by N_y region are

first multiplied by a Gaussian weighting function, and then sliced into any number of smaller subregions (The original SIFT implementation described by Lowe used a 4 by 4 grid of subregions). Histograms of the image gradient in each region are then concatenated to form a descriptor vector for the feature. To achieve rotation invariance, the direction of the maximum gradient is first subtracted from all gradient orientations prior to histogram formation.

SIFT Implementation.

Though SIFT is designed to perform image registration in the presence of image scaling, rotation, and translation, not all of these factors are present within the plenoptic ranging problem. Eliminating these extra degrees of flexibility allows for the development of a modified feature matching algorithm, which should outperform a full application of SIFT. A few of these changes are listed here.

1) Since neighboring subaperture images are not rotated from each other, feature vectors do not need to use orientations relative to the gradient of greatest magnitude. This stands to eliminate errors resulting when two gradients are of nearly the same magnitude.

2) To achieve scale invariance, SIFT builds feature vectors from the scale space layer at which the feature was detected. Descriptor vectors are thus 'scale normalized,' and can be compared to features detected at any scale within a separate image. Since no rescaling occurs between subaperture images, in principle it should be possible to only compare a feature in one image with features detected at the same scale within a neighboring image. In practice, we allow all features within the same octave to be compared, as the DoG detector will not with perfect consistency detect a feature at the same scale under image translation.

3) To accommodate arbitrary translation, rotation, and scaling, SIFT compares all features detected within one image to all features within a second image. This means that the location of a feature must be definite in all dimensions. For example, edges whose location along the edge is difficult to define, must be culled by the SIFT algorithm. In plenoptic range finding, the transformation between neighboring subapture images is constrained to a translation along a known direction. This has two consequences. Firstly, features need only to be matched with features in the second image along the known line of translation. Secondly, detection of edges can be allowed and encouraged by removing the requirement for the DoG to be a maximum in the direction transverse to translation.

Feature Matching.

Feature matching is performed by searching for the least Euclidean distance between feature vectors. In cases where a feature in one image does not have an equivalent within the second image, due to either detection failure or false detection, it is expected that the minimum Euclidean distance will not deviate far from the next-smallest distance. In order to remove such cases, only matches are retained where the minimum distance is smaller than all other distances by at least some specified factor which we will refer to as a matching threshold.

Stereo Ranging using SIFT.

In this section we examine the results of stereo matching using SIFT. Stereo matching is performed using the two extreme subaperture images of a synthetic light field. Since ground truth for depths within the synthetic light field is known, this can be used to calculate the actual disparity for each image point. The two performance

metrics examined in this section are the variance of the estimated disparity from the actual disparity across the entire imag, and the number of matches.

SIFT performance is based on a number of different parameters alluded to in the previous sections. The starting scale of the first octave of scale space, which gives the amount of initial blurring of the image, determines the scale of the smallest features that will be detected. A detection threshold determines how large a local DoG extrema must be in absolute value in order to be classified as a feature. Finally, the matching threshold determines how close a match must be, compared to other close matches, in order to be retained as a true match. Each of these parameters was varied across a range of values, using both the author's modified implementation of SIFT and a standard implementation called VLFeat, meant to closely mimic the specifications of Lowe's paper [32].

Fig. 33 shows the results for the authors SIFT implementation, while Fig. 34 shows the results for VLFeat. As expected, the author's implementation does provide a much greater volume of matches, with all parameters being equal. However, the variance achieved with VLFeat is also considerably lower than that of modified SIFT.

Across the board, increasing the matching threshold leads to better performance at the cost of match count. Both of these effects are expected. Though the number of matches continues to drop as the matching threshold is increased, the falloff in variance diminishes quickly after a value of about 2, making this an optimal choice. Increasing the detection threshold, while reducing the number of matches, does not seem to result in better accuracy. This may indicate that a non-zero detection threshold leads to detection failure (a feature detected in one image, but not in a second image). The trend with respect to initial blurring scales is slightly more difficult to interpret. Nonetheless, in both cases, an initial blurring scale of $\sigma = 0.8$ provides the best results.

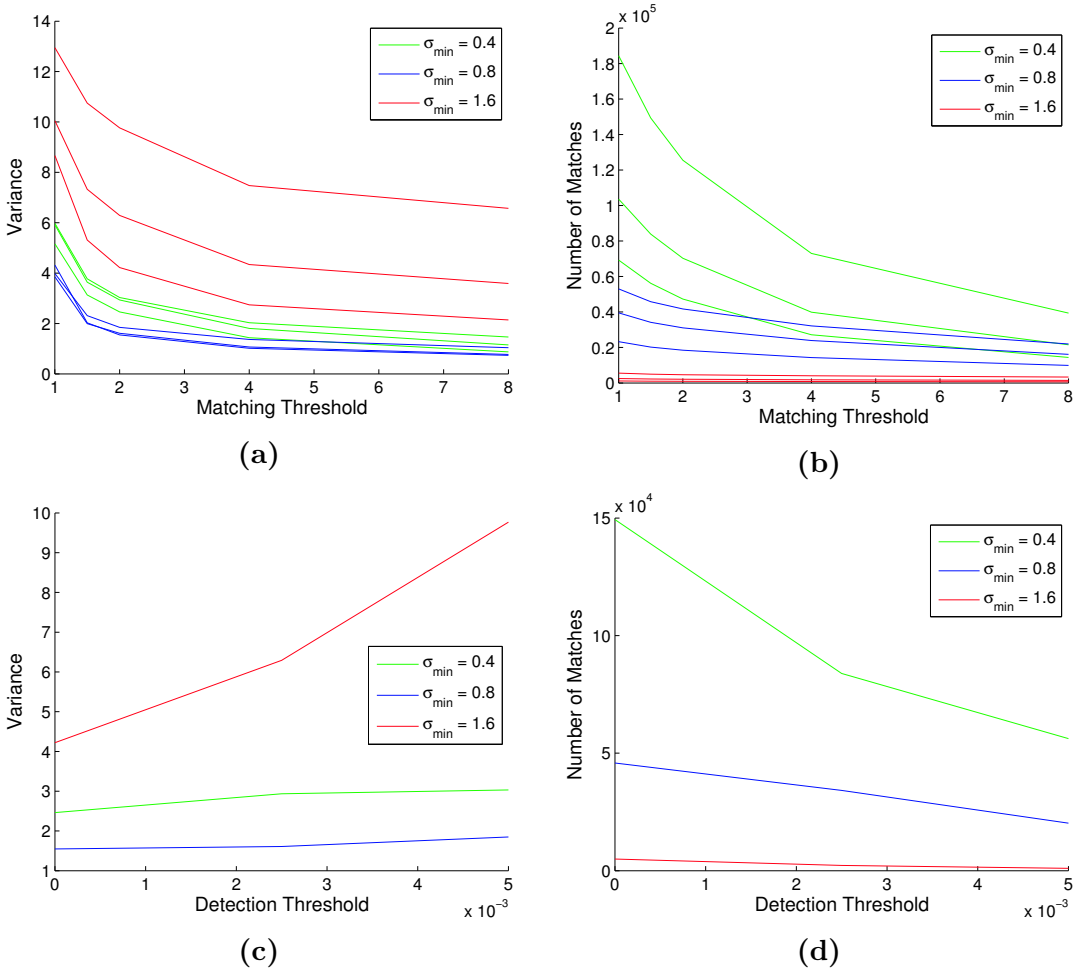


Figure 33. Stereo Matching Performance using Modified SIFT (Author's Implementation). (a) and (b) show the variance from ground truth and number of matches, respectively, in terms of matching threshold and level of initial blurring. The multiple lines at each blurring level correspond to different detection thresholds. (c) and (d) explicitly show the dependence on detection threshold, at a matching threshold of two.

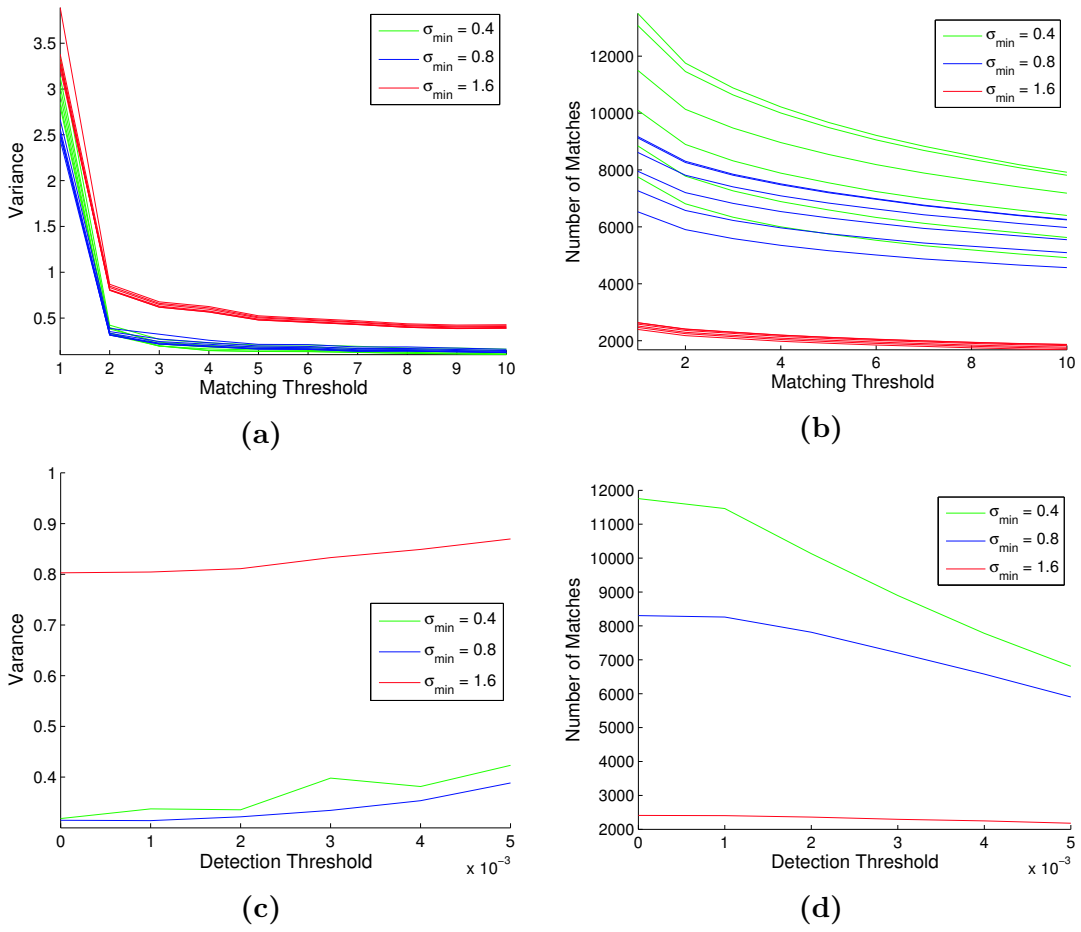
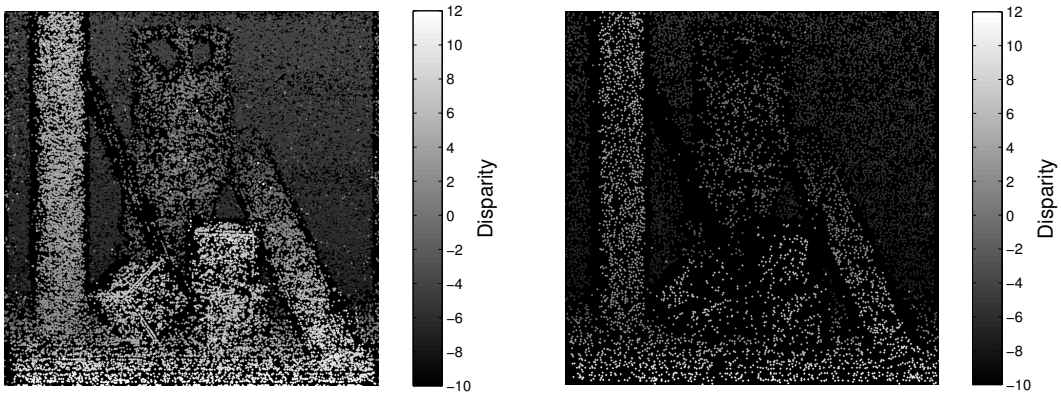


Figure 34. Stereo Matching Performance using SIFT (VLFeat). (a) and (b) show the variance from ground truth and number of matches, respectively, in terms of matching threshold and level of initial blurring. The multiple lines at each blurring level correspond to different detection thresholds. (c) and (d) explicitly show the dependence on detection threshold, at a matching threshold of two.



(a) **(b)**

Figure 35. Maps from Stereo Matching using SIFT. (a) and (b) show the maps yielded by the author's SIFT implementation and by VLFeat, respectively, using the optimal conditions determined from Figs. 33 and 34, namely, with $\sigma_{min} = 0.8$, Matching Threshold = 2, and Detection Threshold = 0. VLFeat yields better accuracy at the cost of match count.

Fig. 35 shows the depth maps generated by each method using optimum parameters. In each case, the minimum blurring was chosen as $\sigma_{min} = 0.8$, the detection threshold was set to zero, and the matching threshold was set to 2. Though the author's implementation provides a much greater number of matches, the variance from the true disparity is large compared to that achieved using VLFeat. Since the intent of this research is to assess the performance limits of the plenoptic camera, VLFeat is used in further analysis. Future work might further explore the tradeoffs existing between the two implementations, and how overall performance could be optimized.

Light Field Ranging using SIFT.

Improved performance over the stereo-matching results presented in the previous section should be possible by using the intermediate subaperture images, in addition to those located at the two extremes, to estimate disparity. In its fullest application, this would entail feature matching between each subaperture image and its 4 nearest neighbors. In this section, we deal with the simplified case where one angular

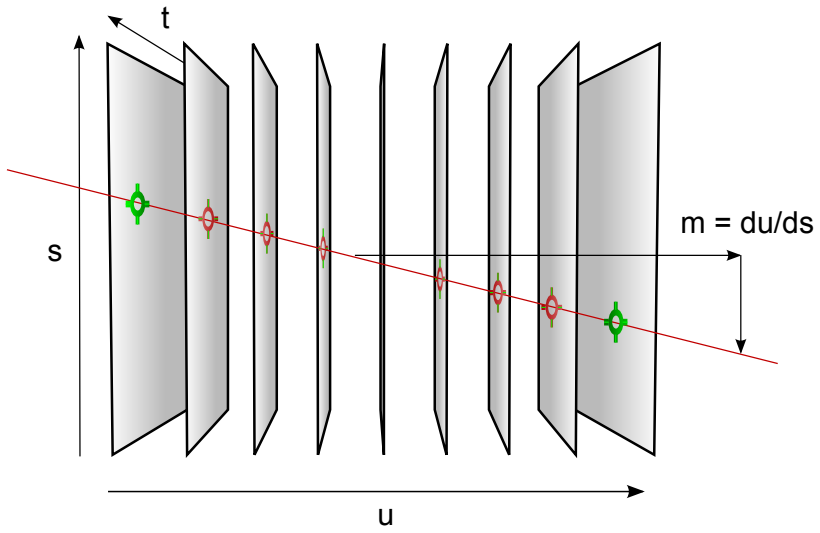


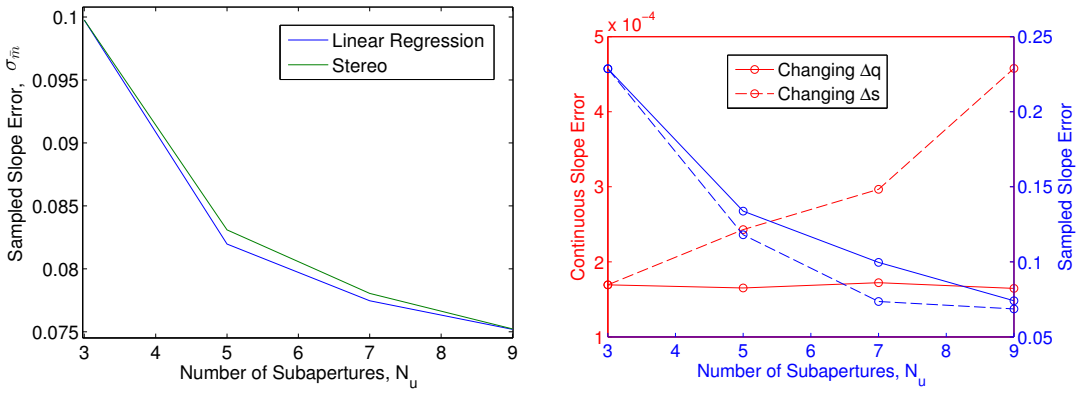
Figure 36. Feature Matching Framework. In stereo ranging, slope estimates are formed using only two images from the stack. When the entire light field is available, intermediate images can be used to achieve better estimate via simple linear regression.

coordinate, u, is varied while the other, v, remains fixed, thus giving N_u different subaperture images.

The approach employed is to establish feature matches between each subaperture image and its two neighbors at $u + 1$ and $u - 1$. These matches are sorted in order to produce a matrix in which each column corresponds to a feature and each row corresponds to a subaperture image, cell values giving the location of the feature within the image.

If the same feature is being accurately detected and matched within each image, it should follow that the disparity between successive images will be nearly the same size. To remove cases where features are improperly matched, we calculate the variance of the disparity in each column, and throw out columns in which the variance exceeds a specified threshold.

Fig. 36 contrasts the approach taken here with the stereo based approach of the previous section. Given increased number of sample points, a simple linear regression becomes an appropriate approach to determining the light field slope. Fig. 37a shows,



(a) (b)

Figure 37. Simulated Camera Performance with SIFT. In (a), N_u is increased according to the first scheme in Fig. 27, such that γ stays constant. The improvement in accuracy with N_u falls short of that predicted in Eq. 85. Also, the results obtained using linear regression do not significantly improve upon results obtained via stereo. (b) shows the uncertainty for the other two schemes in Fig. 27. Increasing N_u in either of these manners does not lead to an overall improvement in uncertainty, represented by the continuous light field slope error.

for both the case of slope estimated using the two extreme images (stereo) and the case of slope estimated from the entire range of images (linear regression), how slope uncertainty diminishes as angular samples are added in a manner corresponding to the first scheme in Fig. 27.

Interestingly, the performance of the stereo estimator is remarkably close to that of the linear regression estimator. A comparison of Eqs. 82 and 87 indicates that the error when slope is calculated using the linear regression should drop off faster than the stereo case by an additional factor of $1/\sqrt{N_u}$.

A plausible explanation for this discrepancy might be that the feature localization error does not match the assumption of a normally distributed probability density function. However, Fig. 38 illustrates that the localization error distribution is fairly well approximated by a normal distribution. The localization error in the figure was calculated by using the average of the feature locations across all subaperture images as a true location for the central subaperture image. True locations for the other images were then calculated by using the known light field slope obtained from the

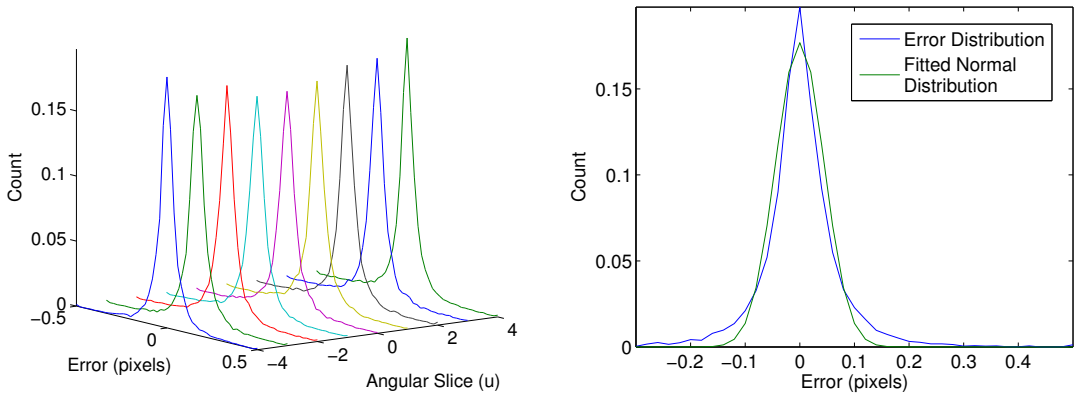


Figure 38. SIFT Localization Error. Corresponds to the case pictured in Fig. 37a. The shape of the error distribution remains largely constant across the range of subaperture images. The distribution is well approximated by a normal fit. The fit shown in (b) has a coefficient of determination, r^2, of 0.96.

ground truth depth map. Further analysis is necessary to determine if the slight deviation from the distribution and its normal fit shown in the figure is sufficient to account for the failure of Fig. 37a to match with theory.

In Fig. 37a, angular samples are added by increasing the simulated camera diameter D and decreasing the detector size Δq in such a way that the factor $\gamma = \bar{m}/m$ remains constant. Fig. 37b shows the cases where angular resolution is added in accordance with the two other schemes in Fig. 27. Since these tradeoffs do not maintain a constant γ, the sampled light field slope error and continuous light field slope error are shown separately.

Though all three cases show a fall-off in sampled light field slope error with increasing number of angular samples, the dependence falls short of that described in equation 83. This, in turn, leads to unexpected behavior for the continuous light field slope error. For example, though changing the detector size to increase N_u while keeping D constant should lead to a decreased uncertainty according to Eq. 84, the behavior in Fig. 37b is constant with N_u.

4.4 Range Finding using Epipolar Plane Images

The feature matching framework is useful for a number of reasons. Feature matching using SIFT-like algorithms is a very commonly employed technique for determining structure from imagery within the computer vision community. Therefore, depth estimation using SIFT represents an obvious first approach to the plenoptic ranging problem. A second advantage of feature matching is the straightforward uncertainty analysis available via the simple linear regression estimator. Finally, as discussed in the previous section, when registration between images is linked to an entity having an existence within the scene itself (namely, a feature), this entity can be localized to within subpixel precision, allowing for highly accurate depth estimates.

For these reasons, it makes sense to employ feature matching as a first look at plenoptic rangefinding. However, the high dimensionality of the light field also allows for other more direct methods which, in their simplicity, afford considerable advantages over the use of SIFT. These methods operate directly on either the light field itself or on Epipolar Plane Images (See section 3.4).

Since a point source must appear as a sloped line within an EPI, one simple approach is to search for this line by looking for slopes along which the EPI has low variance or photo-consistency. This can be thought of as the equivalent of image registration through correlation, applied to the light field.

Slope estimation using Light Field Photo-consistency.

Fig. 39 shows an Epipolar Plane Image (EPI). As discussed in section 3.2, the EPI is composed of sloped lines, each of which maps to a single point within the scene corresponding to the light field, such that the slope of the line relates to the distance to the point. Now, imagine calculating the variance of an EPI along each of its vertical columns. In areas containing vertical lines, the values contained within a

Figure 39. Calculation of Photo-Consistency. The figure shows a portion of an EPI under varying degrees of shearing. The variance along the dotted white line is minimized with the shearing slope matches the slope of the lines in the light field.

single column would stay consistent, leading to a low variance. On the other hand, in areas where a column is crossed by multiple slanted lines, the variance will be higher.

This suggests the approach of estimating slope by shearing the light field by different amounts, and looking for vertical lines identified by low variance at each degree of shearing. Following this approach for a given light field slice will result in an N_s by N_m matrix of variance values, where N_s is the width of the slice and N_m is the number of slopes used for shearing the EPI.

This matrix can be visualized as a disparity space image, or DSI, as in [7]. Fig. 40 shows a DSI and depth map built from the same light field. Each row of the image corresponds to a different degree of shearing of the light field slice. The value of each pixel gives the variance calculated from the vertical columns within the sheared slice. To avoid confusion with the concept of variance within the context of random variables, the term photo-consistency is henceforth used in place of variance for this value.

Close observation reveals that the DSIs in Fig. 40 (b) and (c) are composed of a fundamental unit shaped something like a bundle of lines passing through a central minimum. The bundles are especially prominent in two locations, corresponding to the edges of the ball seen in the depth map shown in the figure. These bundles are known as DSI shadows, since they seem to shadow certain points on the DSI. The

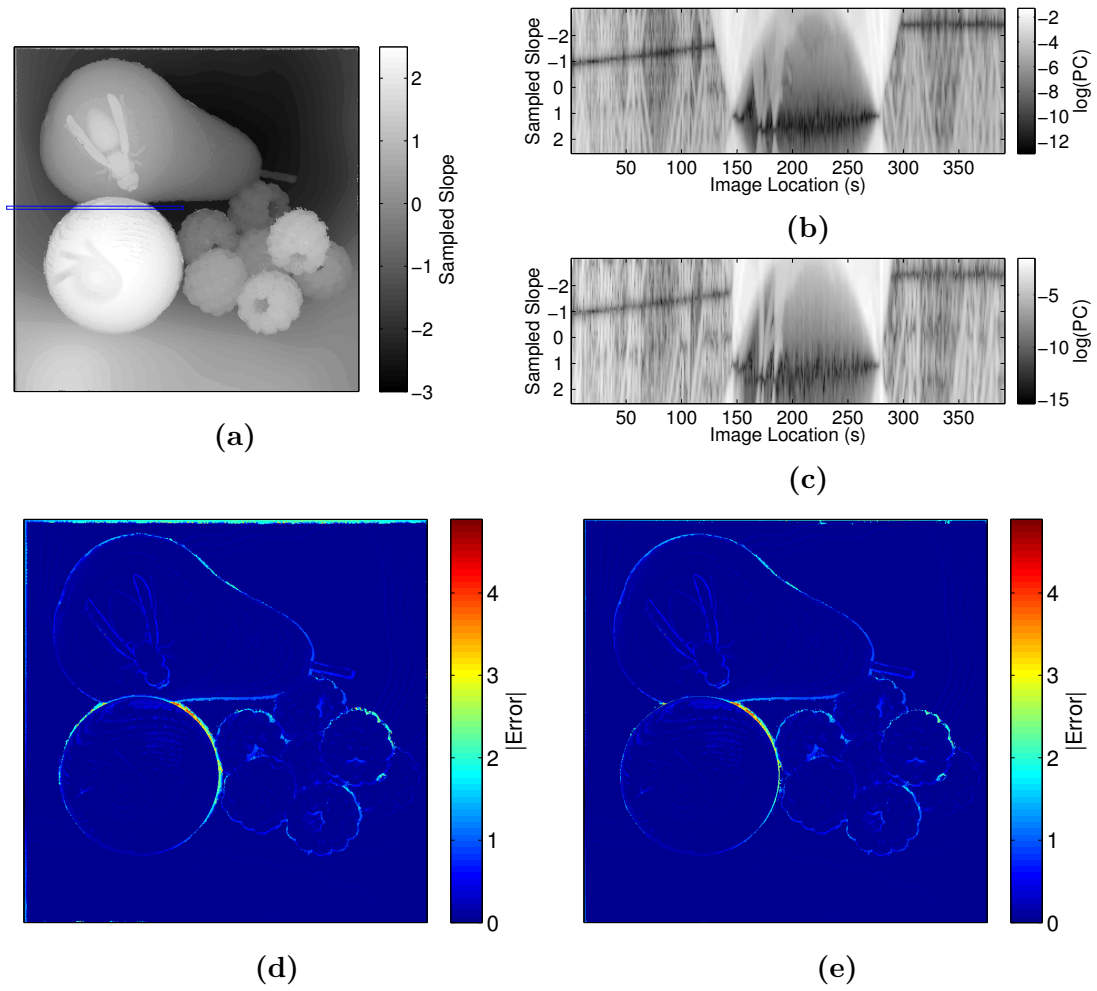


Figure 40. Depth Estimation Using Photo-Consistency. (a) shows a slope map generated using the photo-consistency technique. (b) shows a DSI corresponding to the slice of (a) outlined in blue. Note the shadows located at the edges of the ball, which cover a small portion of the line of minimum values on each side. (c) gives the same DSI, but generated using only the lower angular half of the sheared light field. This causes the shadows to tilt inward above, allowing for a better estimate of the occluded region. (d) and (e) show how the error changes when this extra measure is taken in regions where occlusions are detected. The improvement is most visible in the upper left corner of the ball.

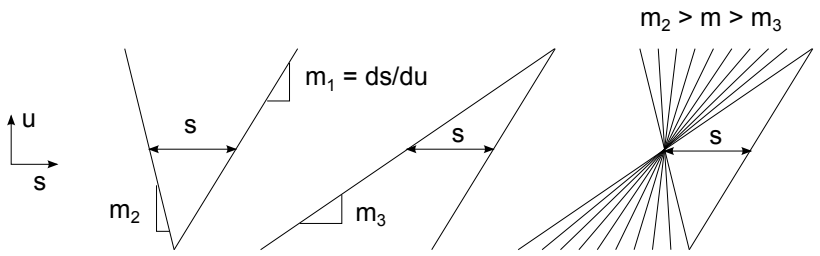


Figure 41. DSI Shadow. Each point on a DSI maps to a line in the Light Field along with the variance is taken to provide the value of the DSI point. The shadow of a DSI point of interest is composed of all those points whose corresponding lines on the light field overlap the line represented by the point of interest. The figure illustrates that there will be a range of slopes outside of this shadow, which increases with spatial separation s.

shadowed points are those corresponding to the actual slope of the light field (vertical position on DSI) at a given location (horizontal DSI position). These appear as minima within the DSI column. The shadow of a DSI minimum consists of all those points whose corresponding lines on the light field cross the line corresponding to the DSI minimum. The shadow will be prominent when the light field contains an edge or strong gradient, as in the two points in the figure.

Fig. 41 shows the range of sloped lines $\bar{m} \in (\bar{m}_2, \bar{m}_3)$ which will not intersect with a line at slope $\bar{m}_1$ located a distance $\bar{s}$ away. Solving the simple system yields

$$\bar{m}_{2,3} = \bar{m}_1 \mp 2\bar{s}/N_q. \tag{91}$$

The shadow region consists of those lines outside of the range $(\bar{m}_2, \bar{m}_3)$, or those satisfying

$$|\bar{m} - \bar{m}_1| > 2\frac{\bar{s}}{N_u} \tag{92}$$

This equation defines two fingers which meet at $\bar{m}_1$ where $\bar{s} = 0$, and recede toward the DSI edges as $\bar{s}$ increases. This is the shape seen in Fig. 40.

DSI shadows can have unwanted effects near object edges. In this context, the DSI shadow represents the effect of one object occluding another over a range of

subaperture images. Notice in the first DSI in Fig. 40 that the line of minimum points on the left of the image seems to disappear behind the shadow corresponding to the edge of the ball. Various sophisticated approaches are possible for dealing with such cases. The approach employed here is a modification of that described in [7], wherein successively distant 'tubes' of equidistant portions of the light field are extracted prior to recalculating the DSI in an iterative process. In the context of a plenoptic camera, the extent of occlusions is generally more limited than for the EPIs in [7], which were collected by a track mounted camera. Even in the presence of occlusions, we can normally count on having at least one half of the light field occlusion free. Therefore, a possible approach is to 1) detect occlusions, 2) determine the 'directionality' of the occlusion, and 3) use a precalculated DSI generated from the occlusion-free half of the light field to estimate slope.

A region is considered to be occluded if the variance generated from a given half of the EPI is less than the variance generated from the full EPI by a specified factor, and if the location of the two minima are separated by more than a specified offset. A factor of 10 and a 0.1 offset threshold yielded good results.

The second DSI in Fig. 40 was generated by using the bottom half of the DSI. Notice how the DSI shadows bend inward, exposing minima which were previously covered. The two error maps in the figure show how this correction leads to reduced error at object boundaries, although the degree of the improvement is variable.

Though DSI shadows can cause problems at object boundaries, in general, the shadow helps 'frame' the DSI minimum, ensuring that it will be easily detected. This is especially true when the image gradient is high. In general, we expect that slope estimation performance will be dependent on gradient scale, especially in the presence of noise.

Sampled Light Field Slope Uncertainty.

To formalize these observations, we consider the case of a light field slice composed of a horizontal gradient with additive Gaussian noise,

$$L(s,u) = gs + n(s,u) \tag{93}$$

where $n \sim N(0,\sigma^2)$ is a zero mean, normally distributed random variable having variance σ^2, and g is the spatial gradient (gradient within a subaperture image):

$$g = \frac{dL}{ds}. \tag{94}$$

For the sampled light field, $\bar{L}(\bar{s},\bar{u}) = \bar{g}\bar{s} + n$, we assume that the gradient is scaled according to the simple relation,

$$\bar{g} = \frac{dL}{ds}\Delta s = \frac{dL}{ds}\Delta q N_u, \tag{95}$$

although we will later see that this assumption is not completely valid, and that a robust treatment of gradient scaling under the effect of sampling is a difficult problem worthy of greater attention. Next, we define a sloped slice of the sampled light field, $S_m(u)$, as

$$S_m(u) = \bar{L}(\bar{m}\bar{u},\bar{u}), \quad \bar{u} \in [-(N_u-1)/2, (N_u-1)/2]. \tag{96}$$

A method of interpolation, discussed in more detail later, is used to provide values of L at non-integer values of $\bar{s} = \bar{m}\bar{u}$. We want the photo-consistency of this slice, which we define as $\mathcal{P}_{\bar{m}}$,

$$\mathcal{P}_{\bar{m}} = \mathrm{var}(S_m) = \mathrm{var}(\bar{g}\bar{m}\bar{u})) + \mathrm{var}(n) = \mathrm{var}(\bar{g}\bar{m}\bar{u}) + \sigma_s^2, \tag{97}$$

which follows because variance is a linear operator when the variables being summed are uncorrelated. The s subscript is added to σ to annotate that this is the sample variance, and not the variance of the normal distribution used to define the random variable, n. This will become of importance later on. The variance of $\bar{g}\bar{m}\bar{u}$ is calculated as

$$\mathrm{var}(\bar{g}\bar{m}\bar{u}) = \frac{1}{N_u - 1} \sum_{-(N_u-1)/2}^{(N_u-1)/2} (\bar{g}\bar{m}\bar{u} - < \bar{g}\bar{m}\bar{u} >)^2 = \frac{2\bar{m}^2\bar{g}^2}{N_u - 1} \sum_{\bar{u}=1}^{(N_u-1)/2} \bar{u}^2 \tag{98}$$

which results because the mean value $< \bar{u} >= 0$. Since $\sum_{x=1}^{n} x^2 = n(n+1)(2n+1)/6$ [30], it follows that

$$\mathrm{var}(\bar{g}\bar{m}\bar{u}) = \frac{\bar{m}^2\bar{g}^2}{12} N_u(N_u + 1) \tag{99}$$

and

$$\mathcal{P}_{\bar{m}} = \frac{\bar{m}^2\bar{g}^2}{12} N_u(N_u + 1) + \sigma_s^2. \tag{100}$$

Where $\sigma = 0$ this equation defines a parabola in m, where the concavity of the parabola increases with the number of angular slices N_u and decreases as the gradient becomes shallower. Fig. 42 compares two parabolas generated using Eq. 100 and by actually shearing a gradient image and calculating photo-consistency. The two parabolas are visually identical.

When $\sigma > 0$, the parabola will be affected by noise related to variation in the sample variance of n, σ_s^2:

$$\mathrm{var}\left(\mathcal{P}_{\bar{m}}\right) = \mathrm{var}\left(\frac{\bar{m}^2\bar{g}^2}{12} N_u(N_u + 1)\right) + \mathrm{var}(\sigma_s^2) = \mathrm{var}(\sigma_s^2) \tag{101}$$

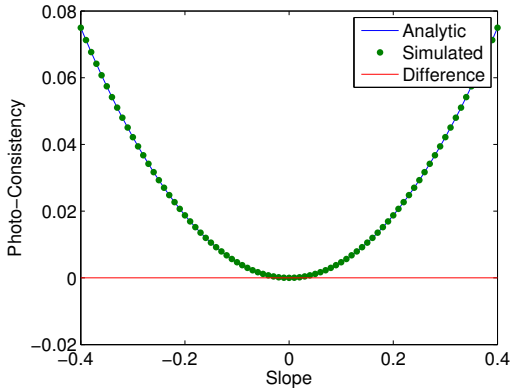


Figure 42. Photo-Consistency of Gradient without Noise. The plots were generated using Eq. 100 and through direct simulation. The plots are visually identical.

where the sample variance, σ_s^2 (equivalently understood as the photo-consistency of the noise component of the light field), expands to

$$\text{var}(\sigma_s^2) = \text{var}\left(\frac{1}{N_u - 1}\sum_{i=1}^{N_u}(n_i - \mu)^2\right). \tag{102}$$

We approximate that the sample mean, μ, is equal to zero. Since $\text{var}(ax + by) = a^2\text{var}(x) + b^2\text{var}(y)$, where a and b are constants and x and y are uncorrelated random variables, the variance operator can be brought inside of the summation:

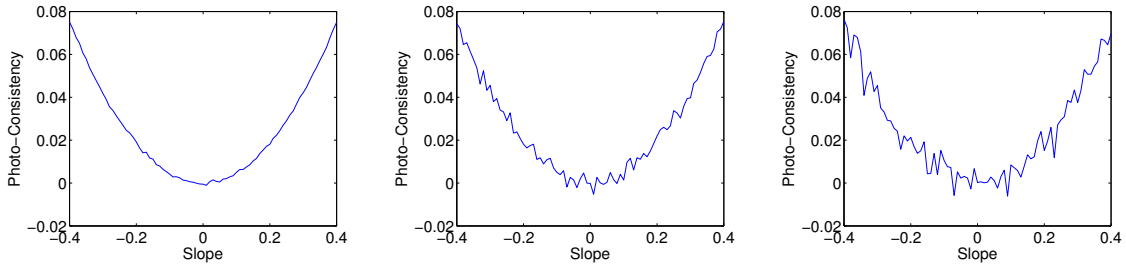
$$\text{var}\left(\frac{1}{N_u - 1}\sum_{i=1}^{N_u} n_i^2\right) = \frac{1}{(N_u - 1)^2}\sum_{i=1}^{N_u}\text{var}(n_i^2) = \frac{N_u}{(N_u - 1)^2}\text{var}(n^2) \tag{103}$$

where

$$\text{var}(n^2) = \int n^2 f(n) dn = 3\sigma^4 \tag{104}$$

and $f(n) = N(0, \sigma)$ is the probability density function for n. The integral is solved via integration by parts. Combining Eqs. 101, 103, and 104 gives the variance of the photo-consistency,

$$\text{var}\left(\mathcal{P}_{\bar{m}}\right) = \frac{N_u}{(N_u - 1)^2} 3\sigma^4 \approx \frac{3}{N_u}\sigma^4. \tag{105}$$



(a) $\sigma^2 = 0.001$ **(b)** $\sigma^2 = 0.005$ **(c)** $\sigma^2 = 0.01$

Figure 43. Photo-Consistency of Gradient with Noise. Plots were generated using Eq. 106 with increasing amounts of Gaussian noise.

This allows us to modify Eq. 100 by expanding the sample variance, as in

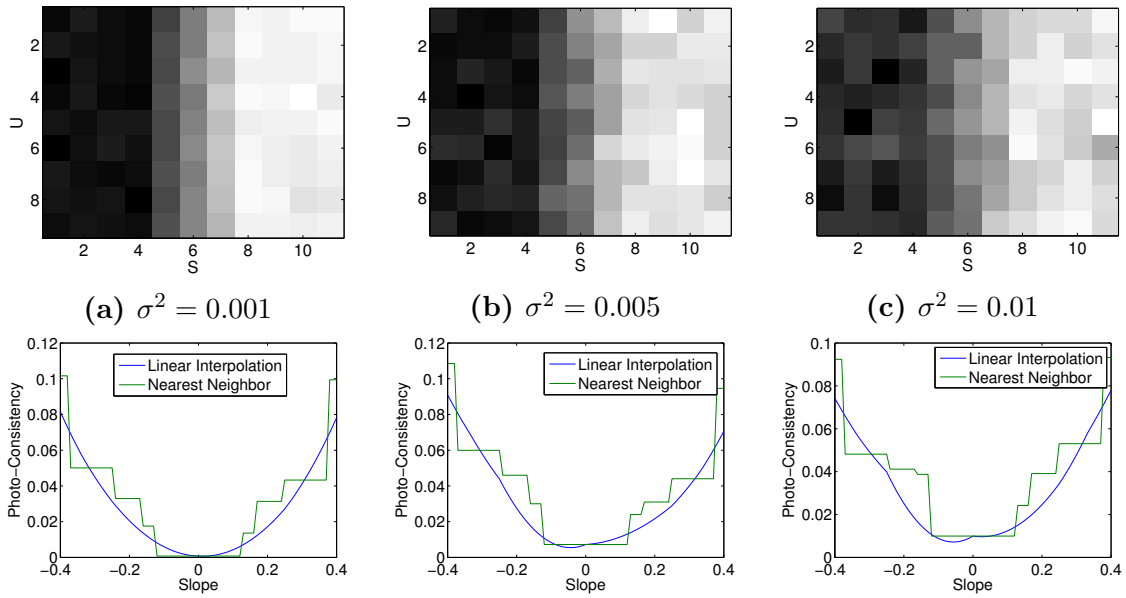
$$\mathcal{P}_{\bar{m}} = \frac{\bar{m}^2\bar{g}^2}{12}N_u(N_u+1) + \sigma^2 + p \tag{106}$$

where $p \sim N(0, \sigma_p^2)$ is a zero mean, normally distributed random variable with standard deviation $\sigma_p = \sqrt{3/N_u}\sigma^2$.

Fig. 43, shows a range of DSI slices generated using Eq. 106. Though, as will be shown below, this is not an accurate representation of the results of shearing a noisy gradient, it does provide a simple framework for thinking about the effects of noise on slope estimation. From the figure, it is clear that, in the presence of too much noise, the minimum of the DSI may no longer exist at the vertex of the parabola, leading to a faulty estimate.

In order to quantify this effect, we determine how far the slope $\bar{m}$ must change in order for the corresponding change in $\mathcal{P}_{\bar{m}}$ to equal the standard deviation σ_p of $\mathcal{P}_{\bar{m}}$:

$$\mathcal{P}_{\Delta\bar{m}} - \mathcal{P}_0 = \sigma_p. \tag{107}$$



(a) $\sigma^2 = 0.001$ **(b)** $\sigma^2 = 0.005$ **(c)** $\sigma^2 = 0.01$

(d) $\sigma^2 = 0.001$ **(e)** $\sigma^2 = 0.005$ **(f)** $\sigma^2 = 0.01$

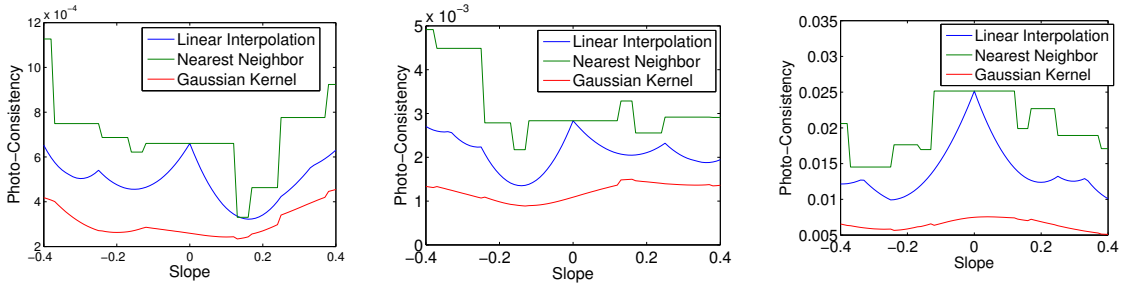
Figure 44. Simulated Photo-Consistency of Gradient with Noise. Plots were generated via direct simulation using noisy gradients. Deviation from the noiseless parabola in Fig. 42 is not uncorrelated with changing slope, as in Fig. 43.

Beyond this point, we consider that it is unlikely for noise to produce a false minimum. The equation solved by

$$\Delta\bar{m} = \frac{1}{\bar{g}}\sqrt{\frac{12\sigma_p}{N_u(N_u+1)}} = \frac{\sigma}{\bar{g}}\sqrt{\frac{12}{N_u(N_u+1)}}\sqrt{\frac{3}{N_u}} \approx \frac{4.5\sigma}{N_u^{5/4}\bar{g}}. \tag{108}$$

This indicates a strong dependence in ranging uncertainty on the intensity of noise, the scale of any gradient features, and the number of angular samples.

Fig. 44 shows some photo-consistency curves resulting from shearing a noisy gradient image over a range of slopes. There is little resemblance to the plots in Fig. 43. The reason for this is that Eq. 106 is derived from the assumption that both $\bar{g}(\bar{s})$ and $n(\bar{s},\bar{u})$ exist on a continuous space in s. This is to allow for the shearing operation, since the angled slice $S_m = \bar{L}(\bar{m}\bar{u},\bar{u})$ calls for evaluation of $\bar{L}$ at arbitrary real values of $\bar{s}$. In reality, $\bar{L}(\bar{s},\bar{u})$ is an image having a finite number of samples, and interpolation is required to shear at arbitrary slopes. The end result is that the

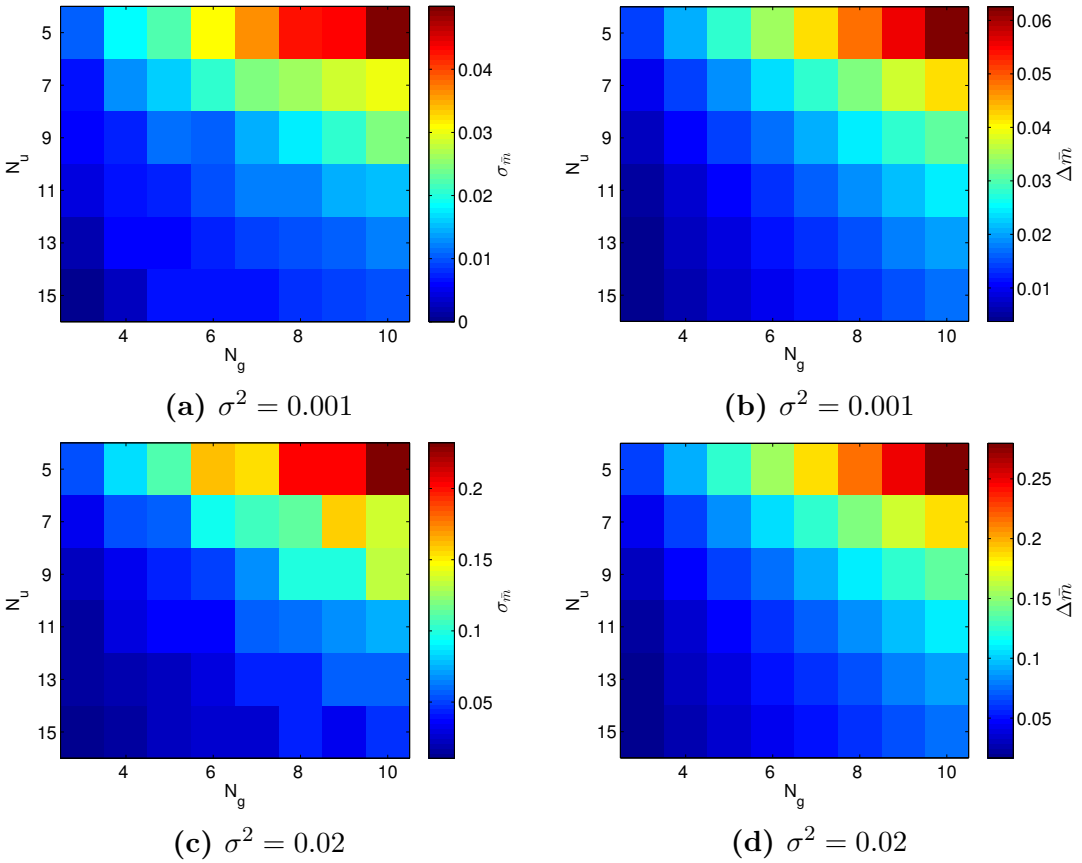


(a) $\sigma^2 = 0.001$ **(b)** $\sigma^2 = 0.005$ **(c)** $\sigma^2 = 0.01$

Figure 45. Photo-Consistency of Gaussian Noise. Plots were generated using varying amounts of Gaussian noise, with no gradient. Using linear interpolation to perform shearing results in cusps at certain points, particularly $m = 0$, where the effective size of the interpolation kernel goes to unity. Using an interpolation kernel that maintains a more constant size helps eliminate these cusps.

variance of the photo-consistency of any given slice of an image will obey Eq. 106 across a set of images having different noise content. However, in a particular image, the photo-consistency of a slice at one slope will not be uncorrelated from the photo-consistency at nearby slopes. This explains why the plots in Fig. 44 show deviation from the ideal parabola, but not in the quickly varying manner of Fig. 43.

A second noteworthy feature of the plots in Fig. 44 is the cusp appearing at $\bar{m} = 0$ in the two rightmost plots. This cusp can be seen more clearly in Fig. 45, where the photo-consistency of only the noise component is plotted. The central maximum for the curves generated using linear interpolation is explained by the fact that at $\bar{m} = 0$, no interpolation is needed. Since interpolation has the effect of convolving the noise and thus reducing its variance, the absence of interpolation at $\bar{m} = 0$ leads to a heightened variance compared to points where interpolation is employed. The other cusps are likely located at points where the need for interpolation is minimal. In order to ameliorate this effect, we employ a method of interpolation which uses a convolution kernel that maintains a constant size. The figure illustrates how this is effective in removing the cusps, although it also has the effect, mostly benign, of reducing the variance everywhere by some factor.



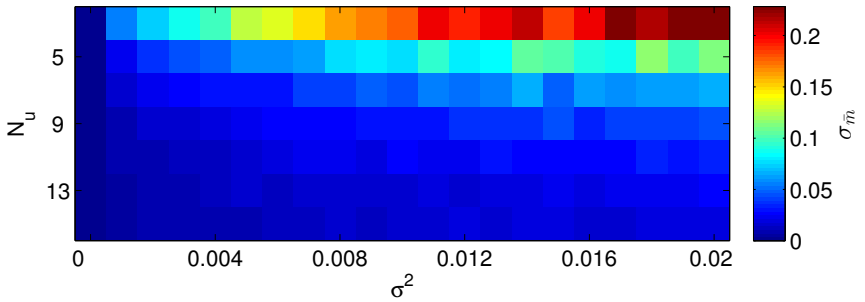
(a) $\sigma^2 = 0.001$ **(b)** $\sigma^2 = 0.001$

(c) $\sigma^2 = 0.02$ **(d)** $\sigma^2 = 0.02$

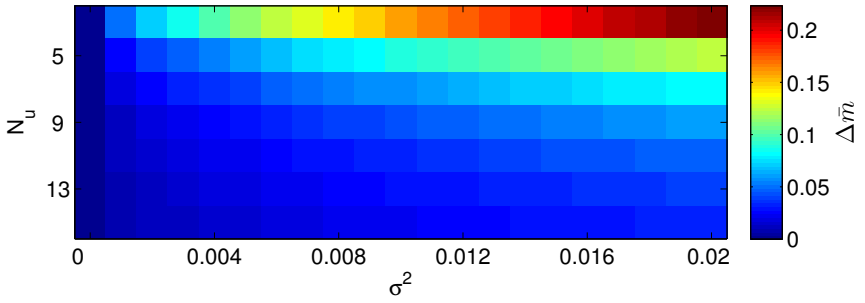
Figure 46. Estimation Uncertainty: Standard deviation of slope estimate from true value as a function of Gradient Scale (N_g) and Number of Angular Samples (N_u), at two different noise levels. (a) and (c) give simulation results, while (b) and (d) give the results predicted by Eq. 108. The analytic result is not formulated as a standard deviation, but it scales in the same way. The values in (b) and (d) in this figure and in Fig 47 haven been divided by a factor of 2.5 in order to make comparison easier.

In order to verify slope uncertainty defined by Eq. 108, simulations were performed by shearing a gradient of variable scale, having a variable number of angular slices, and subject to Gaussian noise of varying standard deviation.

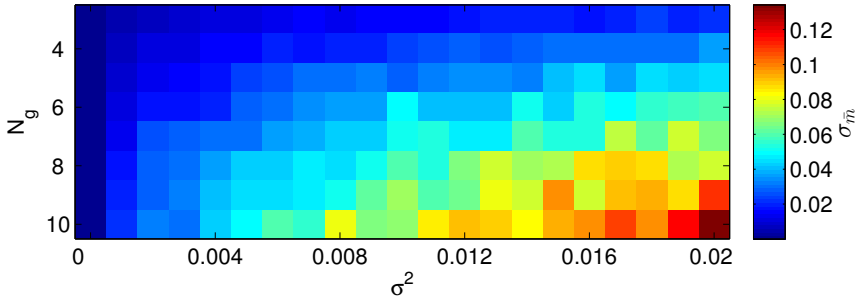
Figs. 46 and 47 illustrate that the relationship in Eq. 108 remains valid, despite the differences between the assumptions made in its derivation and the actual behavior of the sheared, noisy gradient illustrated in Fig. 45. The figures illustrate that, although the number of angular samples does not appear to be highly important at



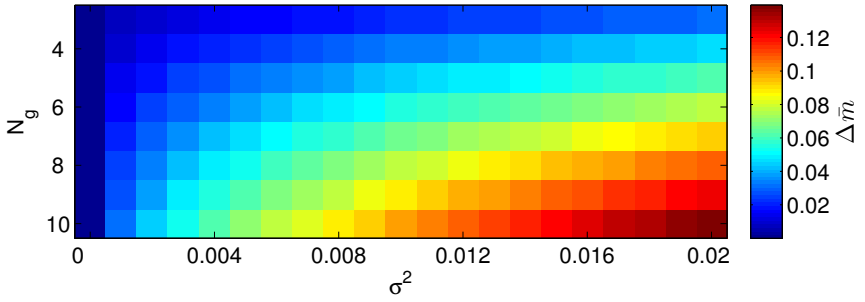
(a)



(b)



(c)



(d)

Figure 47. Estimation Uncertainty (cont.): Standard deviation of slope estimate from true value. (a) and (b) give analytic and simulation results, respectively, as the number of angular samples (N_u) and noise (σ) are varied, and with a gradient scale, $N_g = 5$. (c) and (d) give the results at $N_u = 9$, while varying noise and gradient scale.

low noise conditions, at higher noise levels, an increased number of angular samples results in much better accuracy.

Continuous Light Field Slope Analysis.

The analysis of the previous section focuses on uncertainty in estimating the slope of the sampled light field. Though this analysis is valid within that context, it is insufficient to assess trade-offs in plenoptic camera construction since it ignores the way that the sampled light field slope itself is affected by the changes in sampling characteristics brought about by changing camera parameters. In this section, those effects are considered.

The sampled light field slope and the continuous light field slope are related by the ratio γ, as in

$$\bar{m} = m\gamma = m\frac{\Delta u}{\Delta s} = m\frac{D/N_u}{N_u \Delta q} = m\frac{D}{\Delta q}\frac{1}{N_u^2}. \tag{109}$$

We assume that a sampled spatial gradient is scaled by the sampling interval, Δs, as in

$$\bar{g} = \frac{dL}{ds}\Delta s = \frac{dL}{ds}\Delta q N_u. \tag{110}$$

Upon making these substitutions into the photo-consistency equation, Eq. 106, the continuous-slope photo-consistency, $\mathcal{P}_m$, is given by

$$\mathcal{P}_m = \frac{g^2 m^2 D^2}{12} + \sigma^2 + p. \tag{111}$$

The slope uncertainty, Δm is then given by

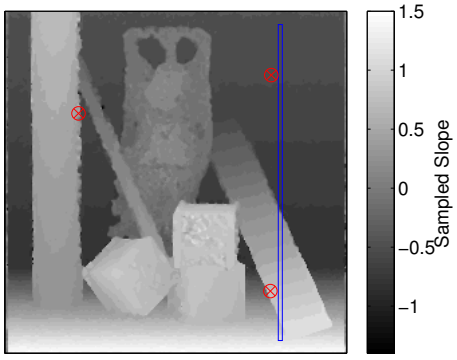
$$\Delta m \approx \frac{4.5}{N_u^{1/4}}\frac{\sigma}{gD}. \tag{112}$$

The most important difference between this equation and Eq. 108 is the dependence on N_u. If the main lens diameter D is held constant, the improvement gained by adding angular samples is only the small $N_u^{-1/4}$ dependence related to an improved signal to noise ratio. The equation implies that the pixel size Δq is not directly related to uncertainty. Thus, the impact of increasing N_u by expanding the microlens size Δs or by decreasing the pixel size Δq should be similar. In the next section, these dependencies are verified experimentally within the context of a synthetic light field.

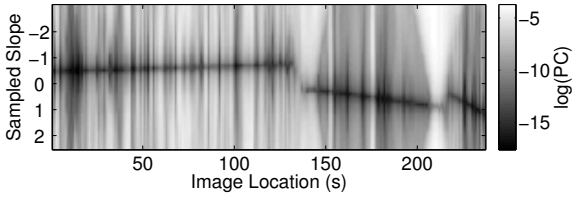
Experimental Results.

Fig. 48 shows depth maps generated using the photo-consistency technique under differing noise conditions, along with associated DSI images for a portion of the scene. Notice that, as the noise level increases, the DSI minima become less distinct until reaching a point where they are difficult to identify. This leads to the noisy behavior seen in the depth maps themselves.

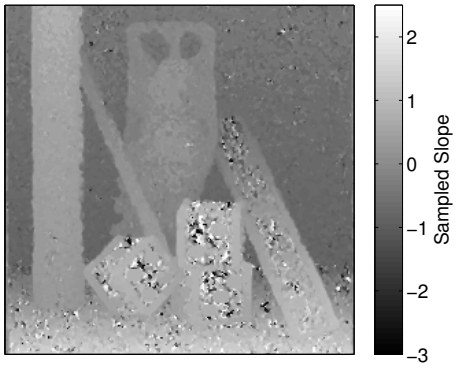
Fig. 49 shows the photo-consistency as noise increases in a non-logarithmic scale for three different points selected from the scene in Fig. 48. In all three cases, the plot starts out having a parabolic shape. This validates the previous section's prediction that photo-consistency curves should take on the shape of a parabola in the vicinity of a minimum. As the amount of image noise increases, the photo-consistency curve itself becomes noisy, leading to detection of false minima. Where the image gradient is stronger (leftmost plot), the parabola is steeper, making it harder for variations due to image noise to create a false minimum of significant magnitude. This is why, in Fig. 48, even under large noise conditions, strong edges like those of the die and plank remain well resolved.



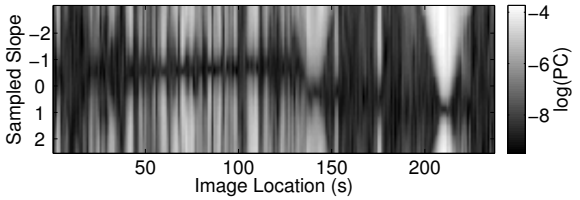
(a) $\sigma^2 = 0$



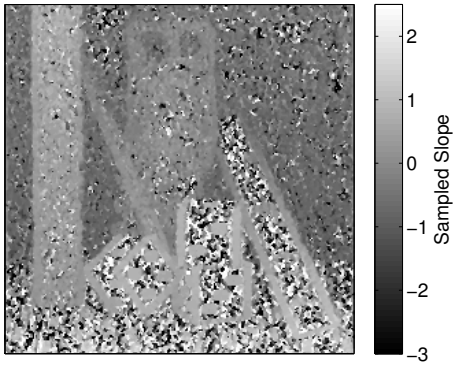
(b) $\sigma^2 = 0$



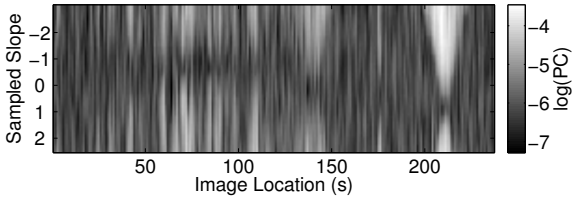
(c) $\sigma^2 = 0.0004$



(d) $\sigma^2 = 0.0004$



(e) $\sigma^2 = 0.005$



(f) $\sigma^2 = 0.005$

Figure 48. Slope Maps From Noisy Light Fields. As noise increases, the DSI minimum becomes increasingly poorly defined, leading to the noisy slope estimates seen in the maps.

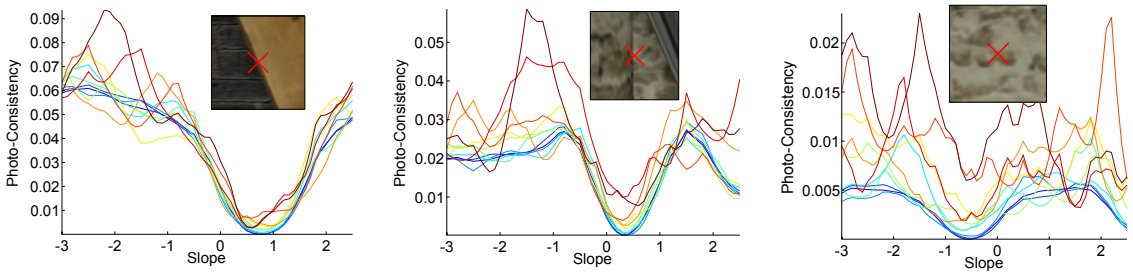


Figure 49. Photo-Consistency with Noise. Plots are associated with the indicated points of the light field in Fig. 48 (a). The plots grow more erratic as noise increases. Where the image gradient is higher, photo-consistency minimums stay more localized under the effects of noise.

Simulated Camera Analysis: Varying Lens Diameter.

This chapter's introduction outlines a number of ways in which the synthetic light fields provided by HCI can be resampled to simulate changes to the plenoptic camera configuration. The simplest of these is illustrated in the first column of Fig. 27. This corresponds to increasing the camera lens diameter while decreasing the detector sizes. The transformation increases camera lens diameter while decreasing the detector sizes in a manner that maintains the ratio $\gamma = \Delta u / \Delta s$. Since γ relates the continuous and sampled light field slopes, keeping γ constant means that the behavior of σ_m will mimic that of $\sigma_{\bar{m}}$ under changes in N_u brought about by this transformation.

Fig. 50 shows $\sigma_{\bar{m}}$ under the impact of varying noise and changing number of angular samples in the manner described above. According to the model derived above, uncertainty should grow linearly with noise and fall off with $1/N_u^{5/4}$. However, performing an exponential fit to the results shows that the growth with noise is closer to $\sqrt{\sigma}$ and the fall-off under changing N_u closer to $1/\sqrt{N_u}$.

A clue to the reason for these discrepancies is found within the slope maps in Fig. 48. Certain regions of the map quickly display estimation noise having an amplitude that spans the entire possible range (from the smallest possible to greatest possible

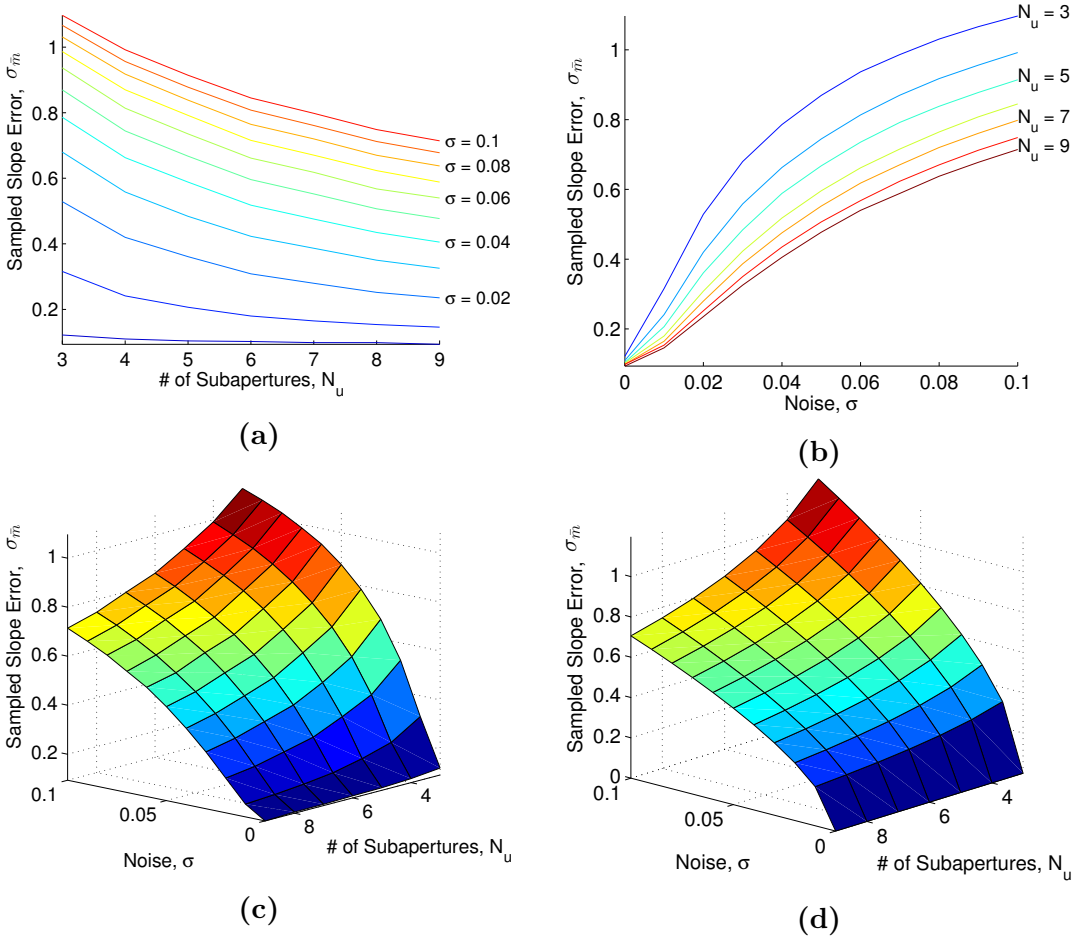


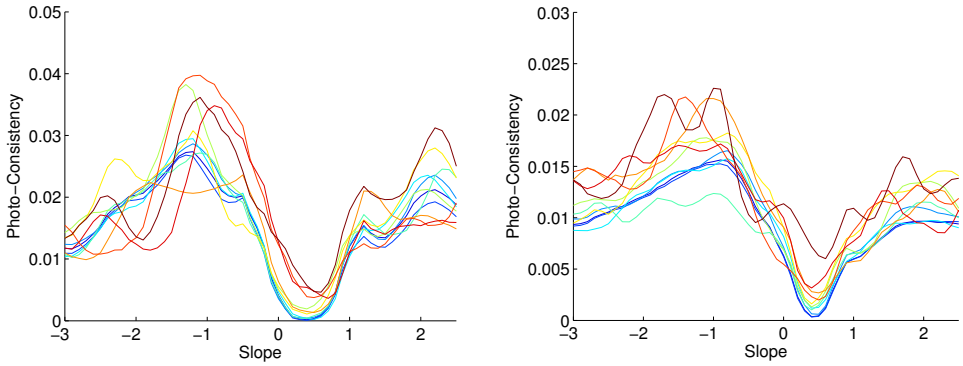
Figure 50. Experimental Slope Uncertainty, Varying Lens Diameter: Effects of increasing the number of angular samples while holding Δs and Δu constant. This requires simultaneously increasing the lens diameter and decreasing the detector size. The exponential fit in (d) follows $N_u^{-0.48}\sigma^{-0.56}$, rather than the expected $N_u^{-1.25}\sigma^{-1}$. Discrepancies with theory are likely due to the faulty assumption of 'infinite' gradients.

slope). Adding more noise to the light field therefore makes no difference in the estimation noise seen in the depth map for such regions.

This observation is explained in the following manner: in developing the analytic model, we treat the gradients involved as having indefinite extent. Throughout the section, photo-consistency plots are shown as parabolas of indefinite extent, corresponding to the assumption of an indefinitely extensive gradient. However, in a true scene or image, most gradients exist as part of edges or patterns of textures. For such cases, as the shearing slope moves away from actual light field slope, the photo-consistency will typically reach a noise floor at which parabolic behavior ceases. Once the variation in the photo-consistency induced by light field noise reaches a magnitude exceeding this floor, the photo-consistency minimum is liable to jump outside of the parabolic region, resulting in an error which spans the entire possible range.

It follows that, as noise levels increase, a large component of the increase in error is due to additional samples entering this regime of where error is more or less uniformly distributed across the entire possible range. At some point, a saturation-like behavior must take place as the number of locations displaying this behavior approaches the total number of locations, and the number of opportunities for new instances of the behavior to come about diminishes. This likely explains why, in Fig. 50, uncertainty appears to grow linearly with noise for a while before reaching a point where growth diminishes.

Fig. 51 illustrates that, though adding additional angular samples increases the concavity of the parabola within the parabolic region of the photo-consistency curve, it does not significantly impact height of the floor surrounding the parabolic region. This explains why the fall-off in N_u is less than expected: once a location begins to display unbounded behavior, increasing N_u does little to improve estimation accuracy.



(a) $N_q = 3$ **(b)** $N_q = 9$

Figure 51. The Effect of Changing Number of Subapertures on Photo-Consistency. Increasing the number of subapertures causes the parabolic region of the curve to become tighter. But the floor outside of the parabolic region actually drops.

Simulated Camera Analysis: Varying Detector Size.

Resampling the light field according to the second scheme in Fig. 27 corresponds to changing the size of the detector elements while keeping all other parameters constant. The effect of this variation is shown in Fig. 52. The continuous light field slope error appears to decrease linearly as N_u increases. However, when an exponential fit is performed, the falloff comes close to the predicted $(1/N_u)^{0.25}$ dependence, with the best fit falling off with $(1/N_u)^{0.17}$. Similarly, the sampled light field slope error falls of as $(1/N_u)^{1.15}$ compared to the theoretical prediction of $(1/N_u)^{1.25}$.

Simulated Camera Analysis: Spatial/Angular Trade-off.

The effect of changing the microlens size Δs is of particular interest because it induces a tradeoff between angular and spatial sampling density, as seen in the third scheme of Fig. 27.

Fig. 53 compares DSI images generated from light fields within the tradespace at $N_u = 3$ and $N_u = 9$, subjected to Gaussian noise. Though the images appear to be impacted differently by the light field noise, it isn't clear that one better highlights the minima associated with correct slope. The second and fourth images show banding

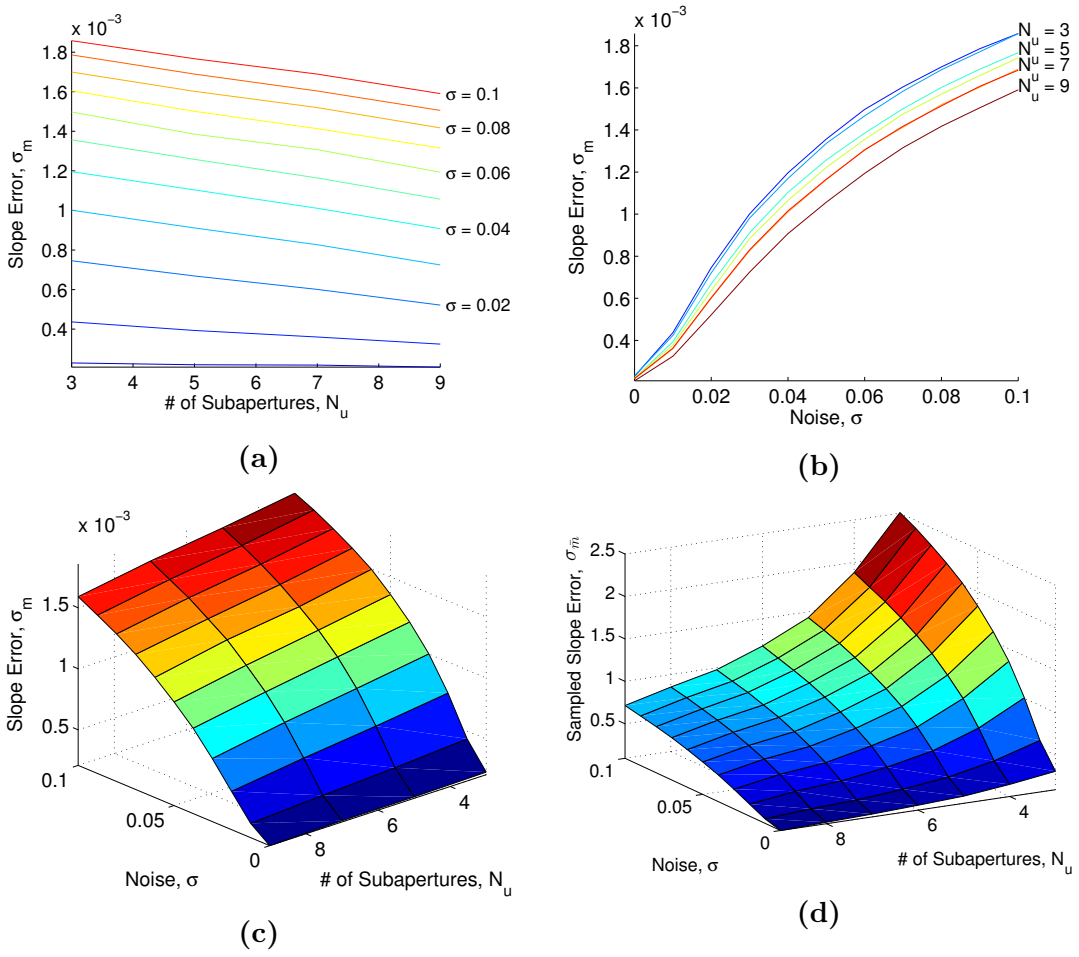
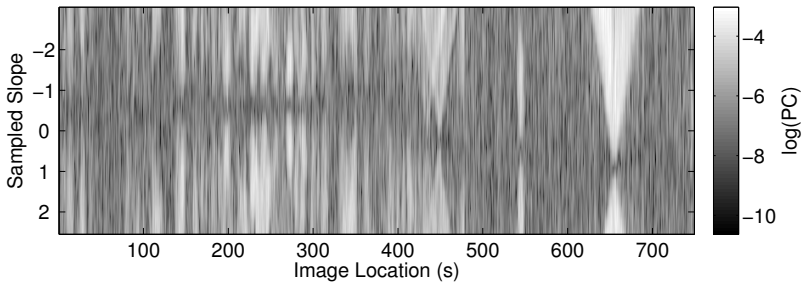
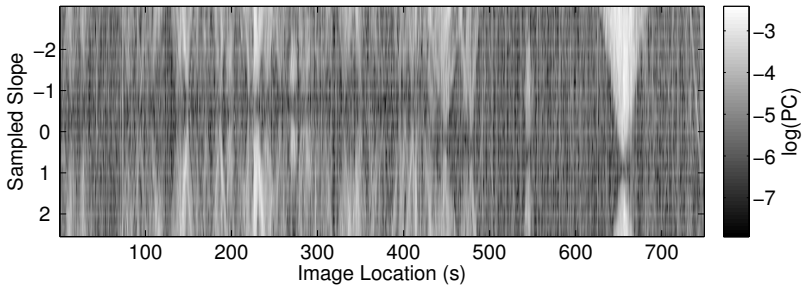


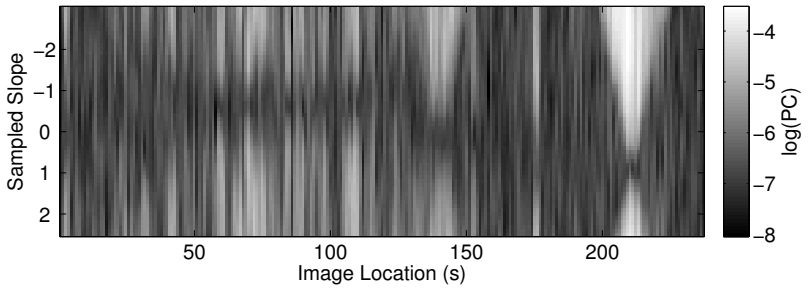
Figure 52. Experimental Slope Uncertainty, Varying Detector Size: Effects of increasing the number of angular samples N_u by decreasing detector size. (a) through (c) show continuous light field slope error, while (d) shows sampled light field slope error. The best fit for (c) follows $N_u^{-1.15}$, while the fit to (d) follows $N_u^{-0.17}$, in good agreement with the theory.



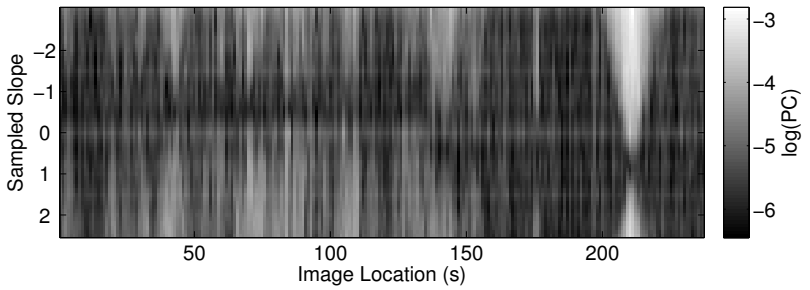
(a) $\sigma^2 = 0.0025$



(b) $\sigma^2 = 0.0025$



(c) $\sigma^2 = 0.0025$



(d) $\sigma^2 = 0.0025$

Figure 53. Comparison of DSIs Generated from Noisy EPIs. (a) and (b) use a light field having three angular samples, while (c) and (d), have nine angular samples, and correspondingly lower spatial resolution. (a) and (c) use a Gaussian interpolation kernel with a 1 pixel standard deviation and having a constant size of three pixels, while (b) and (d) use linear interpolation. Where linear interpolation is used, cusp artifacts are apparent as horizontal stripes.

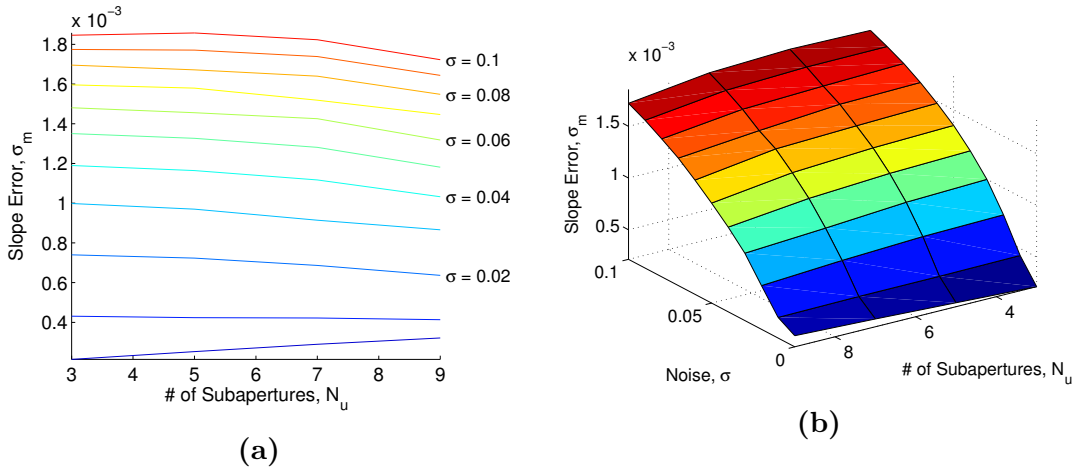


Figure 54. Experimental Slope Uncertainty, Varying Microlens Size: Effects of angular/spatial resolution trade-off achieved by varying microlens size on slope uncertainty under presence of noise. The increase in performance with angular resolution is minimal.

artifacts associated with linear interpolation, which are suppressed by the use of a Gaussian interpolation kernel rather than linear interpolation, as discussed above.

The estimation performance is quantified in Fig. 54, which shows the error under different noise conditions as Δs is altered (effecting a change in N_u). The improvement in performance with N_u in this case is hardly noteworthy. To understand why this is the case, it is necessary to start by assessing the agreement between the sampled light field slope error and Eq. 108. Next, the transformation between continuous light field quantities and sampled quantities must be examined.

Fig. 55 shows the sampled light field slope error as a function of image gradient, $\bar{g}$, and number of angular samples, N_u. Viewing uncertainty in terms of image gradient is necessary in this case, since image gradient is not expected to remain constant under the spatial resampling involved in altering microlens size. The best fit to the data is noteworthy because the dependence on N_u is a more dramatic N_u^{-2} than the expected $N_u^{-5/4}$ falloff. This is somewhat surprising considering the fact that the continuous light field slope error shows almost no dependence on N_u. The stronger dependence than expected is slightly disconcerting and hints at some form of systematic error.

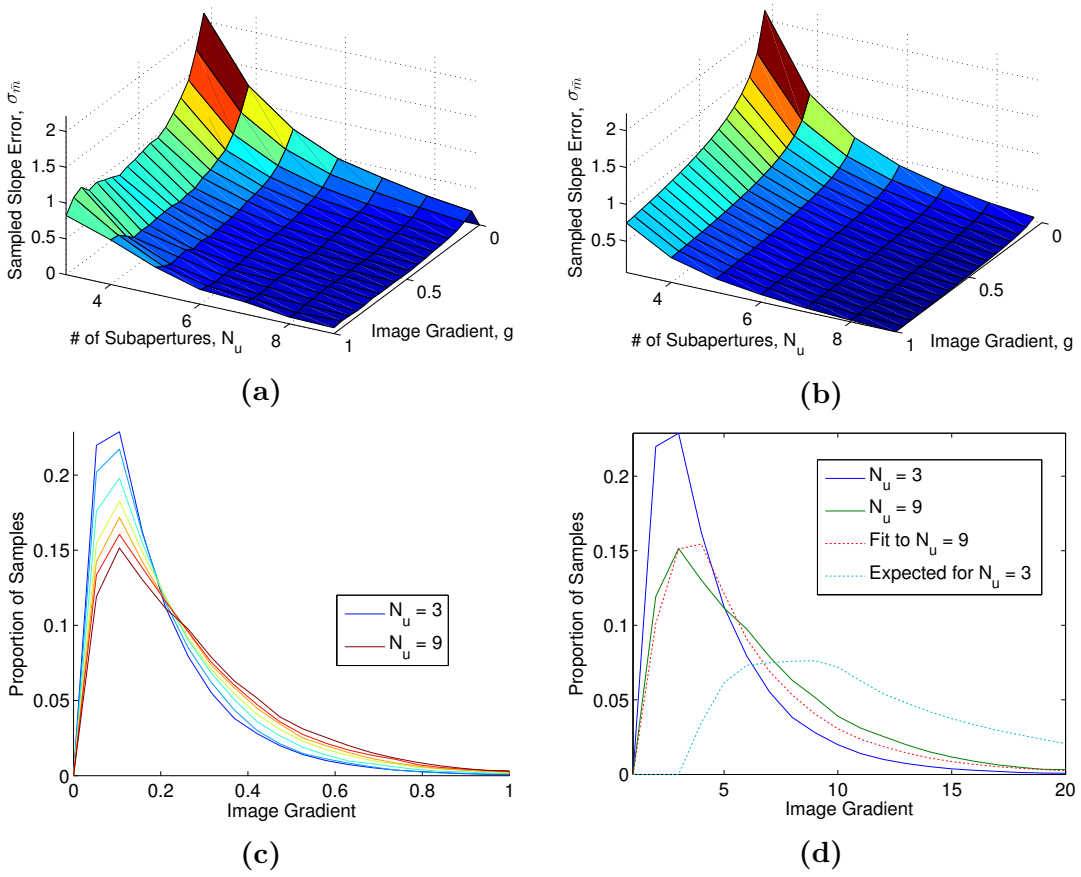


Figure 55. Experimental Slope Uncertainty, Varying Microlens Size (cont.): (a) shows the sampled light field slope estimation error as a function of image gradient and number of angular samples for noise level $\sigma = 0.4$. The exponential fit, shown in (b) follows $N_u^{-2}\sigma^{-0.37}$. Though the dependence on N_u is greater than expected, it does not result in a strong dependence for the continuous slope error in Fig. 54. (c) shows the distribution of the image gradient as Δs increases. The distribution shifts slowly to the right, but not nearly to the extent expected according to the simple scaling assumption that $\bar{g} = g\Delta s$, as shown in (d).

The explanation possibly involves the uniform estimation error behavior discussed above, wherein the average error becomes coupled to the separation of the bounds of the shearing slope used to generate the DSI. If the same continuous slope range is used, this separation does diminish with a $1/N_u^2$ dependence for the sampled light field slope.

That the strong dependence on N_u seen here does not result in a stronger dependence in Fig. 54 is due to two discrepancies between the gradient-related behavior assumed in the development of Eq. 112 and the behavior observed in Fig. 55. Most importantly, the falloff of error with image gradient displays a relationship closer to $\bar{g}^{-1/3}$ than the expected $\bar{g}^{-1}$.

A second discrepancy worth noting involves the scaling of image gradients under the impact of spatial resampling. In the derivation of Eq. 112, it was assumed that continuous gradients would relate to sampled gradients according to $\bar{g} = g\Delta s$. Fig. 55 shows the distribution of gradients as resampling takes place to simulate altering the microlens size. As Δs increases, the distribution gradually shifts to the right. However, under the assumption that $\bar{g}_1/\bar{g}_2 = \Delta s_1/\Delta s_2 = \alpha$, we note that

$$f(\bar{g}_2)d\bar{g}_2 = \frac{1}{\alpha} f\left(\frac{\bar{g}_1}{\alpha}\right) d\bar{g}_1. \tag{113}$$

Fig. 55 shows a fit of the distribution at $N_u = 3$ to the distribution at $N_u = 9$ using the transformation shown here. For the best fit, $\alpha = 1.3$. However, the expected value for α over this range is 3. The figure also shows what the distribution would look like if the assumption about gradient scaling had been correct. In theory, the updated relation $\bar{g} \approx 0.3g\Delta s$ should lead to an increased uncertainty for any type of sampling, without affecting the dependence on N_u.

However, the altered fall-off in $\bar{g}$ has greater ramifications. To show this, we start with an empirical model which reflects the altered dependencies seen if Fig. 55, given

by

$$\Delta\bar{m} = \frac{4.5\sigma}{N_u^2 \bar{g}^{1/3}}. \tag{114}$$

Upon substituting $\Delta m = \Delta\bar{m}/\gamma$ and $\bar{g} = g\Delta s/3$, the equation for continuous quantities is

$$\Delta m \approx \frac{6.5\sigma\Delta q^{2/3}}{D N_u^{1/3} g^{1/3}}. \tag{115}$$

Due to the altered gradient scaling, the N_u^{-2} dependence for the sampled case is reduced to a very mild $N_u^{1/3}$ dependence for the continuous case.

4.5 Range Finding via Refocusing

Depth through Refocusing.

The ability to produce refocused imagery is one of the most striking capabilities latent in the light field captured by the plenoptic camera. When properly focused on an object within a scene, an image will be characterized by sharp edges, steep gradients, and comparatively large amounts of energy in high spatial frequencies. Thus, a natural approach to range finding is to search for refocused images containing these characteristics.

In the spatial domain, this means producing a stack of refocused images and determining in which frame the image gradient reaches a maximum at each pixel. This approach bears strong similarity to the photo-consistency approach discussed in the previous section. Indeed, refocusing involves the same shearing operation used there to identify the slope of an EPI. While the photo-consistency approach looks for low variance in the samples that will be summed together to make up the final image pixel, the depth-through-refocusing technique first performs this summation, and then looks for the high spatial gradients that are made possible when accurate refocusing minimizes the effective point spread function of an object point. When

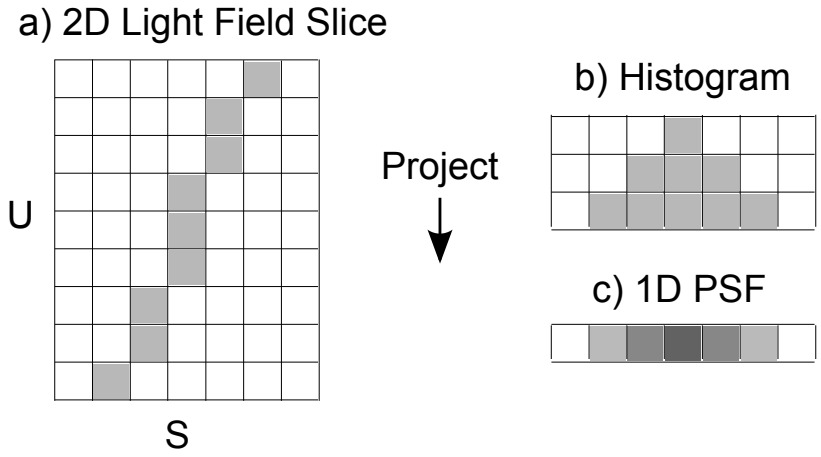


Figure 56. Point Spread Function, 1D. When image formation is performed via projection of the light field, sloped lines within the light field (a) will be spread across a range of image pixels. This distribution, which can be envisioned as the histogram in (b), determines point spread functions for objects at this depth, visualized in (c).

the photo-consistency (variance) is low, this is because the samples associated with a single object point are gathered together under a single image pixel, rather than spread across neighboring pixels where they would reduce spatial image gradients.

In order to consider the effects of defocus on imagery, it is necessary to know the defocus-induced point spread function. Fig. 56 illustrates the formation of the PSF for a two dimensional slice of the light field. A defocused point is represented by a sloped line within the light field, and the image formation operation projects the line down into one dimension. The projection of the line is then equal to the point spread function for an object at the distance giving a line of that slope. The projection operation consists of counting up the number of u samples associated with each s sample. This same approach can be utilized for the case of the 2D PSF generated from the 4D light field, as illustrated in Fig. 57. The light field slope dictates the range of (u, v) samples over which each (s,t) sample is spread. The resulting PSF is a disk with radius $r = \bar{m}N_u\Delta u/2\Delta s = \bar{m}D/2\Delta s$.

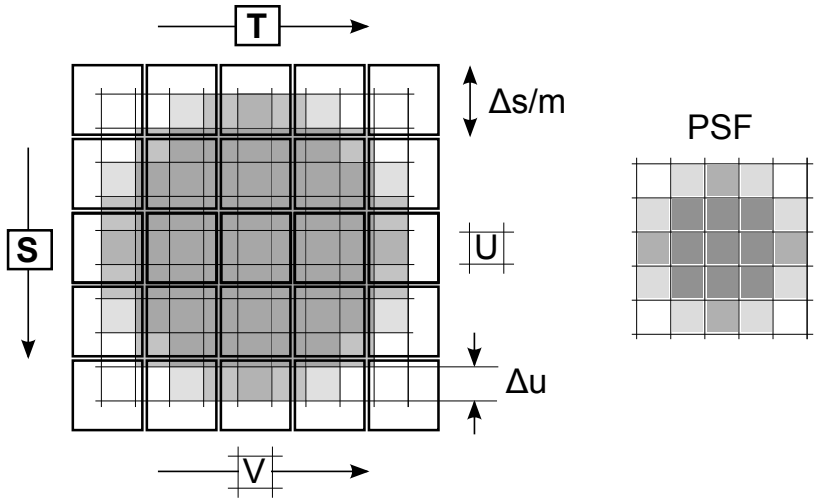


Figure 57. Point Spread Function. The projection required to form the 2D PSF is visualized more easily as counting up the number of angular samples associated with each spatial sample.

Fourier Domain Ranging.

Ranging in the Fourier domain attempts to determine the planes in object space which contain objects by searching for slices of the Fourier transformed light field which contain large amounts of high spatial frequency content. In general, this involves weighting spectral intensities by some increasing function of spatial frequency, and then summing to provide an image sharpness metric. A simple linear weighting was found to yield good results.

As demonstrated in the next section, this method can provide good results when there is only one object in the scene. However, when multiple objects having different depths exist in the scene, the method can have difficulty distinguishing them. This is because the sharpness metric curve associated with a single object can be very broad, with energy slowly receding into lower spatial frequency regions as an object becomes out of focus. These curves can blend together or overwhelm one another, such that planes containing objects do not appear as local maxima in the curve.

Fourier domain ranging is highly sensitive to aliasing and other Fourier reconstruction artifacts. When simple cubic interpolation is used to slice an image spectrum for

the transformed light field, these artifacts manifest as high spatial frequency content peaking at $m = 0$, and dying away as $|m|$ increases. This can cause the approach to fail in the same way that it is liable to fail for multiple objects at different depths. Reconstruction using a Kaiser-Bessel filter, as discussed in [11], was found to ameliorate this effect. However, the recommended zero-padding of the light field prior to Fourier transforming was observed to introduce a harmful 'ringing' effect.

Fourier Ranging Resolution.

In this section, we develop a simple model for estimating the depth resolving capability attainable via ranging in the Fourier domain. Because Fourier domain ranging is a global operation, it is easy for small objects to be overwhelmed by more dominant objects in the scene. In this section, we address the cause of this phenomenon, and provide a depth resolution expression for two objects in a scene having about the same size and characteristics.

To start out generally, we assume that the scene consists of N_1 delta functions at slope m_1 and N_2 delta functions at m_2. Based on the discussion of the previous section, the refocused image will be given by a depth dependent blurring of the scene. For simplicity, we replace the disc-shaped Point Spread Function of the previous section with a Gaussian kernel having $\sigma = r/2$, given by

$$k(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{x^2 + y^2}{2\sigma^2}\right). \tag{116}$$

The refocused image is given by the convolution of the delta functions with the appropriate Gaussian kernel.

$$im(x, y) = \sum_{i=1}^{N_1} \delta(x - x_i^1, y - y_i^1) * k_1(x, y) + \sum_{i=1}^{N_2} \delta(x - x_i^2, y - y_i^2) * k_2(x, y). \tag{117}$$

We define $g(k_x, k_y) = FT^2[im(x, y)]$ as the Fourier transform of the refocused image. Via the convolution, shift, and linearity properties of the Fourier transform, $g(k_x, k_y)$ is given by

$$g(k_x, k_y) = \hat{k}_1(k_x, k_y) \sum_{i=1}^{N_1} e^{-2\pi i\left(x_i^1 k_x + y_i^1 k_y\right)} + \hat{k}_2(k_x, k_y) \sum_{i=1}^{N_2} e^{-2\pi i\left(x_i^2 k_x + y_i^2 k_y\right)}. \quad (118)$$

The summands are plane waves whose frequency and direction of propagation are determined by the delta function locations (x_i, y_i). These plane waves determine the fabric of Fourier space, the intensity of which is then modulated by the Fourier transformed Gaussian kernels. It is difficult to simplify further without making further assumptions about the structure of the image. We make a large simplification by assuming that for our purposes the distribution in Fourier space can be adequately represented by a uniform distribution multiplied by a weighting factor which corresponds to the overall 'prominence' of the objects at each slope within the scene. In reality, this means reducing the initial distribution of delta functions to one weighted delta function at $(x = 0, y = 0)$ for each depth:

$$g(k_x, k_y) = W_1 \hat{k}_1(k_x, k_y) + W_2 \hat{k}_2(k_x, k_y). \quad (119)$$

The Fourier transform of the Gaussian convolution kernel is a second Gaussian with inverted variance:

$$\hat{k} = \sqrt{\frac{2}{\pi}} \exp\left(-2\pi^2 \sigma^2 (k_x^2 + k_y^2)\right). \quad (120)$$

Image sharpness is quantified by a metric which measures the amount of energy at high spatial frequencies in the image. Here, the $g(k_x, k_y)$ is multiplied by a parabolic

weighting function and then summed:

$$\text{metric} = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}(k_x^2 + k_y^2)|g(k_x, k_y)|dk_x dk_y. \tag{121}$$

For a single delta function with a weight of unity, this integral may be solved approximately in polar coordinates, as in

$$\text{metric} = \int_0^{2\pi}\int_0^{1/2} r^2|g(r)|rdrd\theta \tag{122}$$

Via integration by parts, this is solved by

$$\text{metric} = \frac{1}{a^2}\left[1 - \exp\left(-\frac{a}{4}\right)\left(1 + \frac{a}{4}\right)\right] \tag{123}$$

where $a = 2\pi^2\sigma^2$. This expression approaches a limit of $1/32$ at $a = 0$, and a second derivative of $1/1024$. It is fairly well approximated by the Gaussian having these same properties,

$$\text{metric} \approx \frac{1}{32}\exp\left(-\frac{a}{4\sqrt{2}}\right). \tag{124}$$

Upon substituting, consecutively, for a, σ, and r, we get the metric in terms of slope, m, which is given by

$$\text{metric} = \exp\left(-\frac{\pi^2 D^2(m - m_1)^2}{32\sqrt{2}\Delta s^2}\right) = \exp\left(-\frac{(m - m_1)^2}{2\alpha^2}\right) \tag{125}$$

where $\alpha = 4(2^{1/4})\Delta s/\pi D$, and m is the slope at which the light field is sheared to produce the image. The multiplicative factor has been dropped, since only relative scale is important.

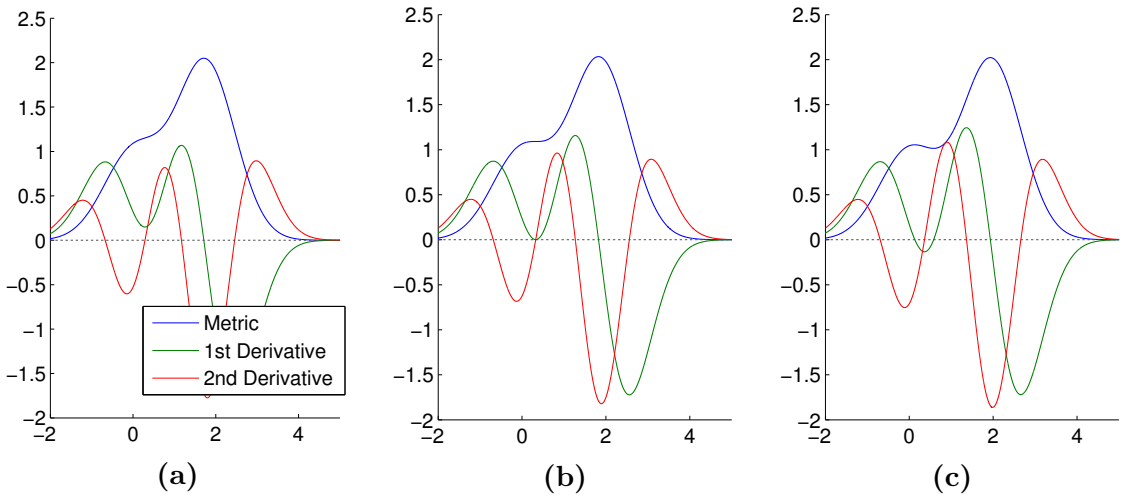


Figure 58. Sparrow Resolvability Criterion. Under the sparrow resolvability criterion, two peaks are considered resolved at the separation which produces a point where the first and second derivatives of the combined curve jointly go to zero.

Without loss of generality, we let one object be located at $m_1 = 0$, and the other be located some interval Δm away. The total metric is then given by

$$\text{metric} = W_1 \exp\left(-\frac{m^2}{2\alpha^2}\right) + W_2 \exp\left(-\frac{(m-\Delta m)^2}{2\alpha^2}\right). \tag{126}$$

The question we seek to investigate is how close the two objects can be in slope space before the two peaks can no longer be resolved. The Sparrow resolvability criterion specifies the point beyond which the two objects will no longer produce two distinct maxima (the point at which the intervening minimum disappears). This occurs where both the first and second derivative of the combined signal are simultaneously zero. Fig. 58 gives a graphical illustration.

Unfortunately, for two Gaussian functions of unequal size, the separation that satisfies this criterion cannot be solved analytically. For the case where $W_1 = W_2$, the criterion is satisfied at

$$\Delta m = 2\alpha = \frac{8(2^{1/4})\Delta s}{\pi D}. \tag{127}$$

Or, in terms of the sampled light field slope, $\bar{m} = m\Delta u/\Delta s$,

$$\Delta\bar{m} = \frac{8(2^{1/4})}{\pi N_u}. \tag{128}$$

Fig. 59 provides an experimental assessment of this expression, using the Lytro Light Field Camera. The Lytro camera has approximately 11x11 subpixels per microlens, and thus the minimum slope separation evaluates to $\Delta\bar{m} = 0.28$. In the experimental tests, the minimum separation for two objects was seen to be two or three times this value. A piece of the explanation may involve the fact that only two orthogonal strips of the Fourier transformed light field were used in forming the metric due to speed considerations. The theory and experimental results do agree that Fourier domain ranging is *not* highly effective at distinguishing objects in a scene, compared to other methods. Some possible approaches to more effective Fourier domain ranging would be a) to split up the light field into spatial segments and apply the method separately to each segment, or b) to fit a Gaussian to the most prominent peak in the metric and then extract its contribution in order to see if any other peaks are made visible.

Camera Calibration.

The relationship between light field slope and object distance is given by

$$\bar{m} = \left(1 - \frac{l_m}{f} + \frac{l_m}{z_o}\right)\gamma. \tag{129}$$

This relationship involves the main lens focal length, the distance from the main lens to the microlens plane, and implicitly parameters such as the main lens diameter and microlens size. This research was performed partly with a camera for which these parameters were not known. Thus, the relationship between light field slope and

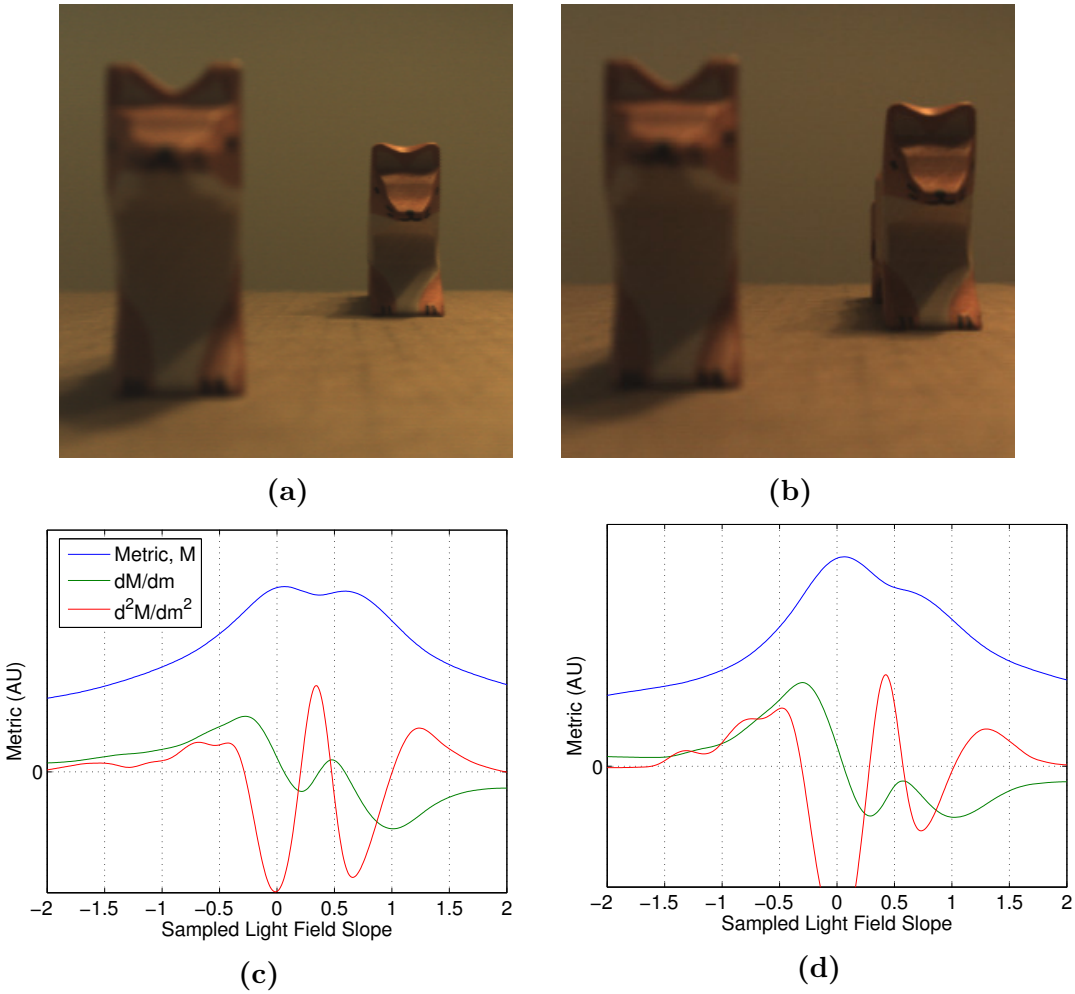


Figure 59. Fourier Ranging Test: An experimental assessment of the ability to resolve objects using Fourier domain ranging. The minimum resolvable separation is several times larger than predicted by the simplified model. However, both results indicate that Fourier domain ranging is not well suited to resolving details about a scene when applied globally.

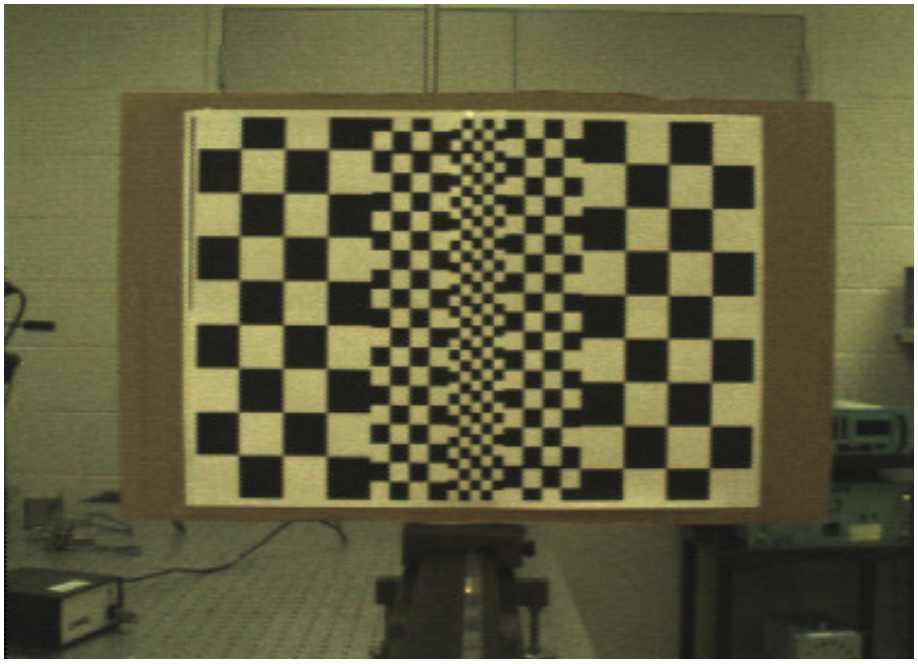


Figure 60. Camera Calibration Target. Checkered grids of varying size were used so that regions of the pattern would remain sharp under camera MTF effects as the target grew more distant.

object distance was determined empirically by imaging a target (See Fig. 60) at a set of known distances.

The Fourier domain ranging method provides a convenient approach for performing this calibration since it naturally gives the distance of the most prominent object in a scene, and there is no need to attempt the removal of noisy depth estimates occurring in previously discussed methods where image gradient is low.

Fig. 61 shows a comparison of the Fourier domain image sharpness metric with a number of alternative metrics. Metric #2 was formed by spatially refocusing the image and then taking its Fourier transform. Metric # 3 is simply a plot of the maximum gradient magnitude contained in a spatially refocused image. The methods show good agreement concerning the slope corresponding to maximum image sharpness. The Fourier domain metric is preferable since it avoids the expensive spatial refocusing step required by the other methods.

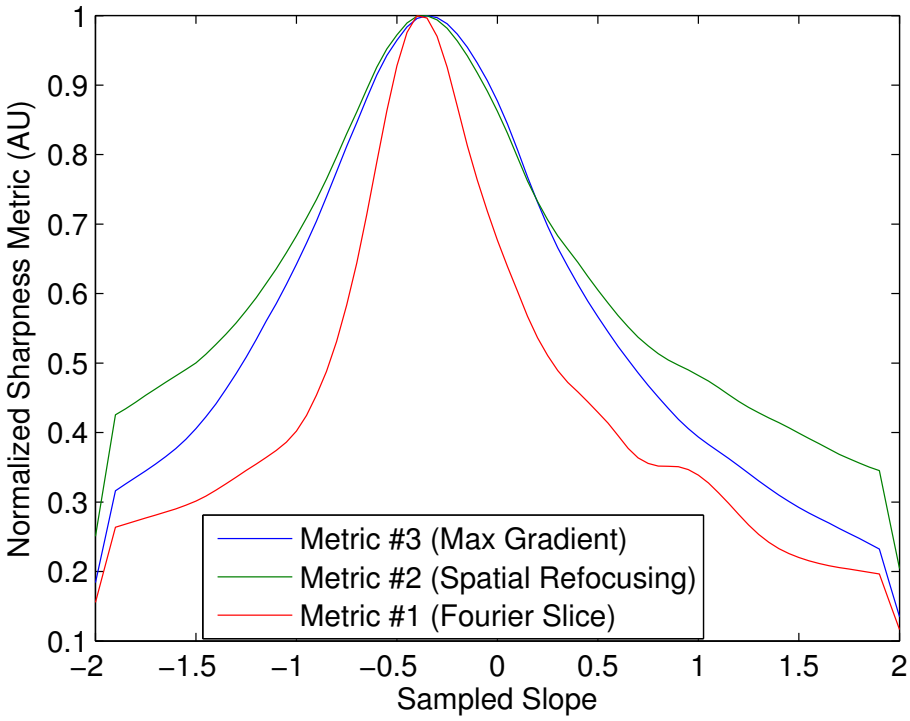


Figure 61. Slope Estimation Results. The metric employed by method 2 is the maximum gradient magnitude contained in the region of interest. Method 3 uses a weighted sum of spectral intensities.

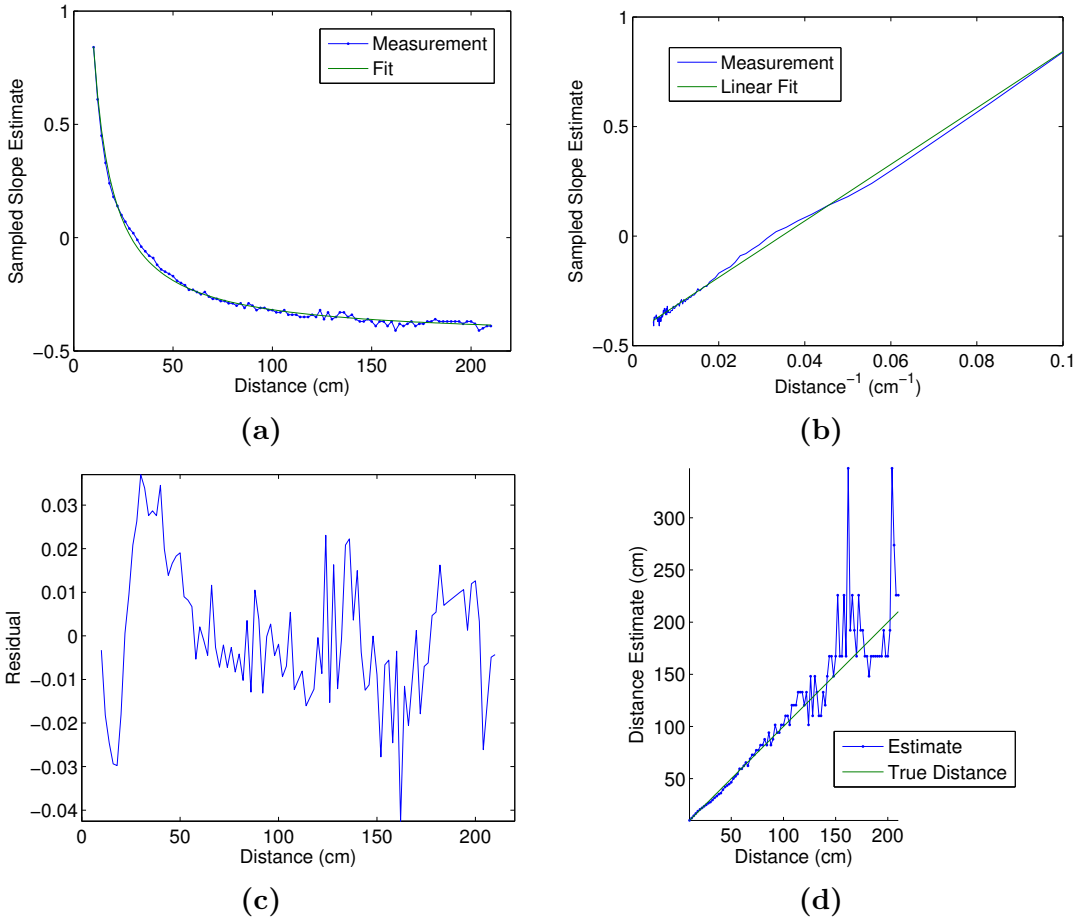


Figure 62. Camera Calibration Plots. Slope measurements obtained for a target at increasing distances from the camera are shown in (a). Plot (b) demonstrates the linear relationship between z_o^{-1} and $\bar{m}$. The residuals of the slope estimation, shown in (c), maintain a regular magnitude as the slope changes. However, since $\Delta s_o \propto s_o^2 \Delta m$, uncertainty in distance estimation does not remain constant, as demonstrated in (d).

Fig. 62 shows the results of a 100-point calibration using a constant 90px by 30px spatial region of the light field, cropped prior to Fourier transformation. Fig. 62b illustrates the linear relationship between z_o^{-1} and the quantized slope, $\bar{m}$, defined in Eq. 129. The fit to this line provides the information needed to use the camera for absolute ranging. The magnitude of the residual between this fit and the slope estimate (See Fig. 62c) spans a similar range as the slope changes. However, as object distance increases, slope estimation errors are amplified, such that the depth estimation error increases quadratically with distance, as seen in Fig. 62d.

4.6 Summary

A major goal of this chapter was to develop models for describing the behavior of the sampled light field slope uncertainty, $\sigma_{\bar{m}}$, in terms of attributes of the sampled light field, such as the number of subapertures N_u, the image gradient $\bar{g}$, and the degree of noise, σ. Such a model is the key piece of a generalized uncertainty model, since the sampled light field slope uncertainty can be directly translated to a distance uncertainty as long as various camera parameters are known. Though the chapter derives several such analytic models, the models do not consistently line up with empirical uncertainties calculated using synthetic light fields.

In the case of feature matching, theoretical modeling indicates that σ_{qm} should diminish as N_u is increased due to the additional samples provided to the simple linear regression. However, the observed fall-off is weaker than expected. Further investigation is needed to determine if the discrepancy results from faulty assumptions made about the nature of the localization error of the feature detector (namely, that it is normally distributed).

For the photo-consistency method, theoretical modeling indicates that σ_{qm} should decrease as N_u increases due to what is effectively an improvement in the signal to noise ratio of the photo-consistency curve. This type of behavior is observed, but not in a manner that consistently follows the analytic model. The discrepancy is likely due to the existence of a behavior not accounted for by the analytic model, in which the mean square error for an entire light field becomes coupled to the size of the range of slopes used to perform the initial shearing of the light field. To achieve more conclusive demonstration of agreement between the theoretical model and empirical results, it will be necessary to avoid this coupling or to find a meaningful uncertainty metric that avoids this problem.

Finally, for the case of Fourier domain ranging, both theoretical and empirical results indicate that the performance of this method is much inferior to that of the spatial domain methods. Therefore, this method is not of further interest for our purposes.

In the absence of a theoretical model that consistently describes the behavior of the sampled slope uncertainty, $\sigma_{\bar{m}}$, the empirical results provided in the preceding sections, and summarized in Table 5, can be scaled to provide range uncertainties for an arbitrary camera. The table gives the average empirical sampled slope uncertainties yielded by the feature matching and photo-consistency methods for the case of zero noise added to the light field. In viewing the table, it is good to keep in mind that the number of depth estimations provided by the SIFT method is much less than that generated by the photo-consistency method.

Table 5. Empirical Values of Sampled Slope Uncertainty, $\sigma_{\bar{m}}$, for Zero Added Noise.

Method	$N_u = 3$	$N_u = 5$	$N_u = 7$	$N_u = 9$
SIFT (Fig. 37a)	0.100	0.082	0.077	0.075
SIFT (Fig. 37b, Changing Δs)	0.228	0.118	0.074	0.069
SIFT (Fig. 37b, Changing Δq)	0.229	0.133	0.100	0.074
SIFT (Avg)	**0.186**	**0.111**	**0.084**	**0.073**
Photo-Consistency (Fig. 50)	0.122	0.104	0.099	0.094
Photo-Consistency (Fig. 52)	0.310	0.178	0.126	0.094
Photo-Consistency (Avg)	**0.216**	**0.141**	**0.113**	**0.094**

As discussed in the chapter introduction, the sampled light field slope uncertainties provided in the table can be scaled to the continuous light field slope uncertainties by

$$\sigma_m = \sigma_{\bar{m}}/\gamma = \sigma_{\bar{m}}\frac{\Delta s}{\Delta u}. \tag{130}$$

Applying the substitutions $\Delta s = N_u \Delta q$ and $\Delta u = D/N_u$, we see that

$$\sigma_m = \frac{\Delta q}{D} N_u^2 \sigma_{\bar{m}}. \tag{131}$$

Via Eq. 64, which relates the continuous slope uncertainty to distance uncertainty, the distance uncertainty is then given by

$$\sigma_z = \frac{z_o^2}{l_m}\frac{\Delta q}{D}N_u^2\sigma_{\bar{m}}. \tag{132}$$

Typically, the value of $\sigma_{\bar{m}}$ in Table 5 does not fall with N_u rapidly enough to counter the N_u^2 factor in this equation. For this reason, a choice of $N_u = 3$ is optimal because it minimizes uncertainty while allowing for application of the photo-consistency method. Since lowering N_u while maintaining a constant detector size Δq is achieved by reducing the microlens size Δs, this also has the advantage of a higher spatial sampling rate at the microlens plane. There may be a practical limit to how small N_u may become under the traditional plenoptic camera framework. If this is the case, the focused plenoptic camera provides a viable alternative, since it can mimic the performance of a traditional plenoptic camera while using larger microlenses (see the equivalences in Table 3).

Choosing $N_u = 3$ and picking a value of $\sigma_{\bar{m}} \sim 0.1$ from the table gives an empirical range uncertainty formula, which will be explored in the next chapter:

$$\sigma_z \approx \frac{z_o^2}{l_m}\frac{\Delta q}{D}. \tag{133}$$

V. Plenoptic Camera Utility

The previous chapter provides the following empirical expression for the range uncertainty of a camera with $N_u = 3$ subapertures.

$$\sigma_z \approx \frac{z_o^2}{l_m}\frac{\Delta q}{D}. \tag{134}$$

This chapter will use this equation to provide an assessment of the applications for which plenoptic camera is well suited. Strictly speaking, the equation relates uncertainty to l_m, the distance from the main lens plane to the microlens plane. However, in order to avoid the light field spreading effects discussed in section 3.4, l_m must be on the same order as f, the main lens focal length. Thus, we represent f rather than l_m as the limiting factor in depth uncertainty. Uncertainty is then reduced by minimizing pixel size, Δq, or maximizing the focal length, f, and lens diameter, D.

To make Eq. 134 easy to grasp, its results are visualized in Fig. 63 for three sample cameras. The camera specifications were selected to nominally represent cameras from three different application regimes. The space-based camera has dimensions similar to that of the Hubble space telescope. The airborne camera is sized such that it might feasibly be mounted on some type of manned aircraft or UAV. The ground-based camera might be wielded by a small scale robot interacting with nearby objects. The space-based camera experiences uncertainty on the order of 10 meters at a distance of 50km. The airborne camera yields comparable uncertainty at a distance of 500 meters. The robotic camera has an uncertainty of 1 meter for a point around 25 meters away. The logarithmic scale in Fig. 64 is useful for assessing the performance of each camera over a broader range of distances. For more general purposes, Fig. 65 provides a nomograph usable for determining the uncertainty of an arbitrary camera.

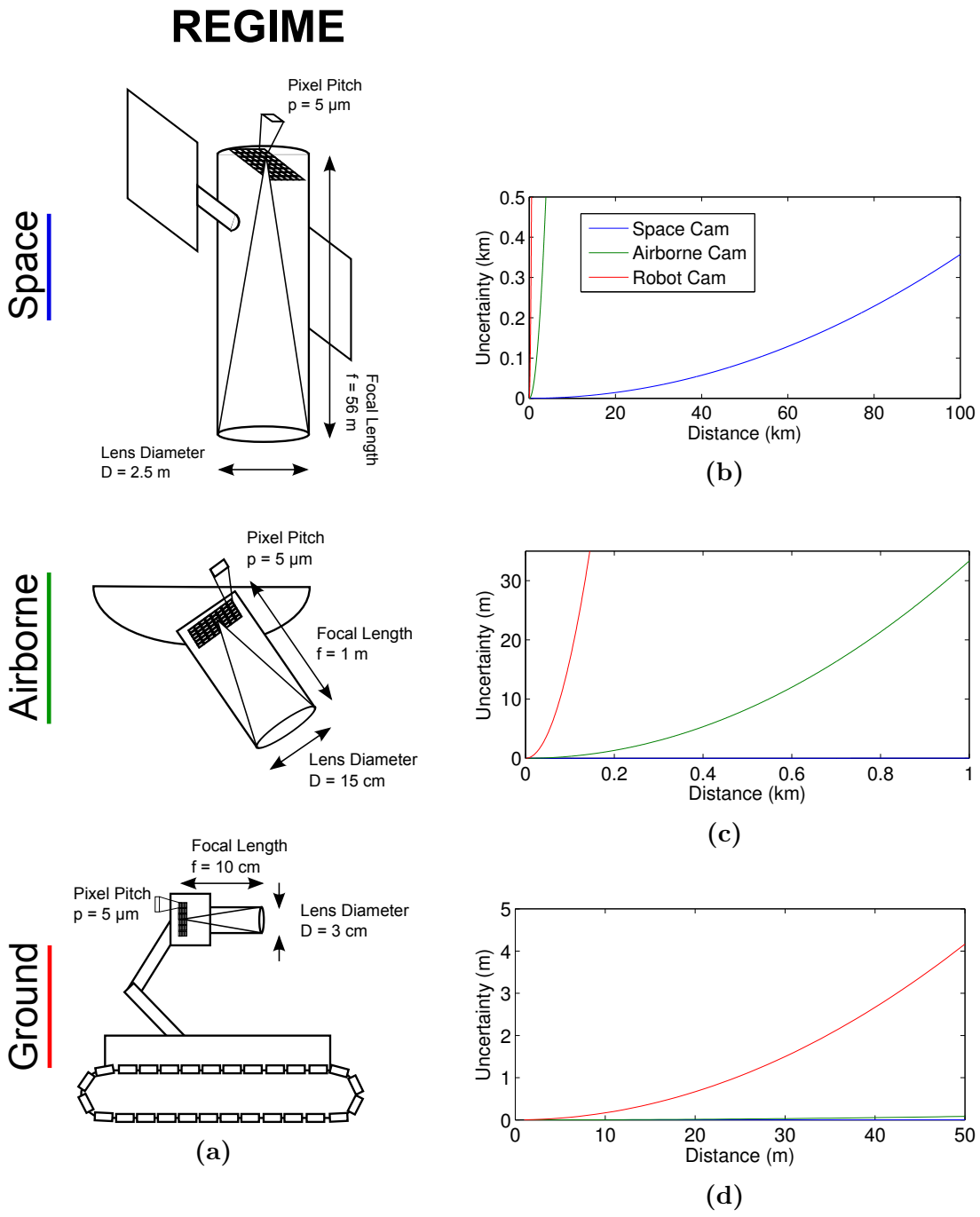


Figure 63. Plenoptic Camera Performance Regimes. The space-based camera experiences uncertainty on the order of 10 meters at a distance of 50km. The airborne camera yields comparable uncertainty at a distance of 500 meters. The robotic camera has an uncertainty of 1 meter for a point around 25 meters away.

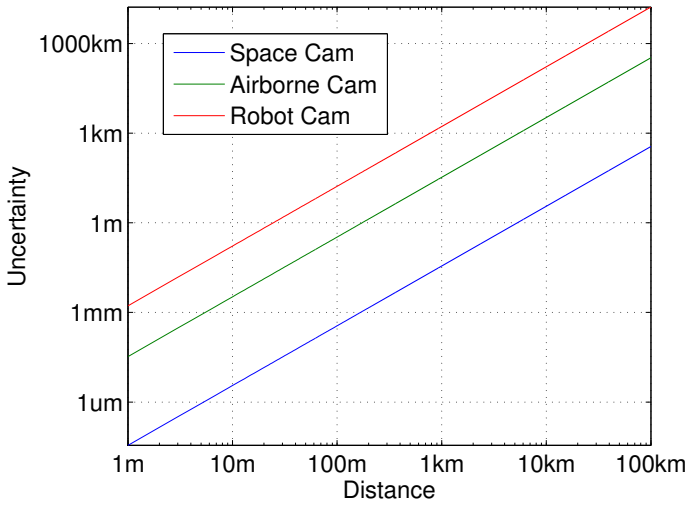


Figure 64. Logarithmic Scale Uncertainties.

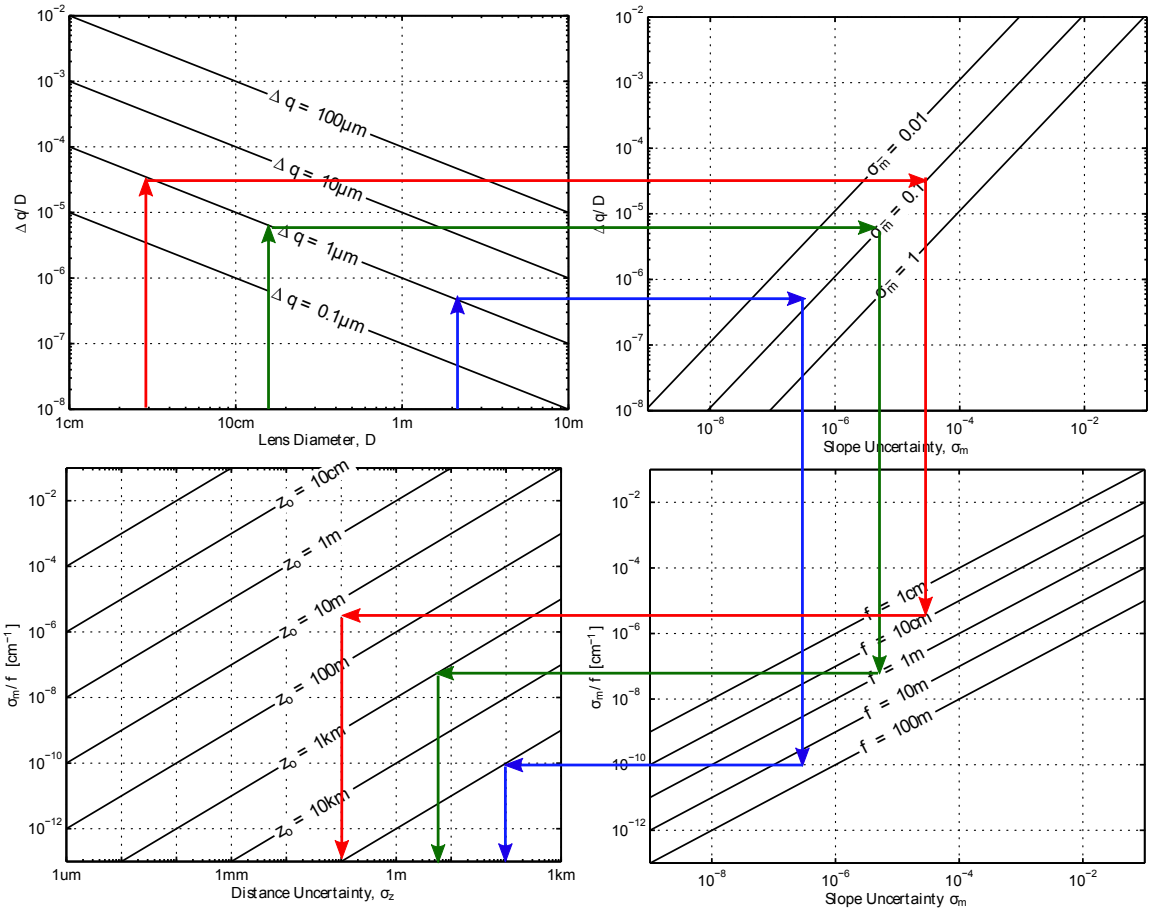


Figure 65. Uncertainty Nomograph. The nomograph can be used to determine the ranging uncertainty for a camera with arbitrary parameters. The paths plotted correspond roughly to the notional cameras presented in Fig. 63.

5.1 Remote Sensing Application

The modern remote sensing landscape is outfitted with a variety of solutions for obtaining depth information. Two such methods include aerial photogrammetry and lidar. Photogrammetry is the process of extracting real-world position information from images of objects. In the era of film photography, stereo-plotters enabled cartographers to identify correspondences between overlapping images taken by an airborne camera. The development of Lidar, a technology which analyzes the reflected response from an active light signal to determine distance, provided advantages over the photogrammetric process both in terms of precision and automated workflow. These initial advantages appear to have given the technology a large user base within the remote sensing community [33].

Lidar systems achieve accurate depth estimations typically by measuring the time of flight of a laser pulse reflected from a surface back toward the receiver. Since Lidar does not depend on parallax in obtaining depth, its depth resolution performance is largely independent of distance, as long as the laser source is powerful enough and the detector sensitive enough for the signal to be registered after propagating the full distance.

In spite of the lead taken by Lidar within the remote sensing field, a combination of factors including the proliferation of high resolution digital imagers, the advancement of the graphics processing unit, and the fruition of automated feature matching techniques developed within the computer vision community have brought new vitality to an era of digital photogrammetry, in which 3D maps can be generated with comparable accuracy and greater efficiency than afforded by laser scanning techniques [33]. Due to the greater complexity of lidar systems, photogrammetry also tends to offer a significant cost advantage.

Lidar and Photogrammetry are held to standards of accuracy set within the remote sensing community. The United States National Map Accuracy Standards (NMAS) specify tolerances for contour maps generated from 3D geographic data, [34]. This standard specifies that "not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval." More recent standards specify stricter tolerances [35], and typical contour intervals range from 0.5 feet to 10 feet [36]. Users of Lidar systems are typically able to certify their results to the 1' contour interval NMAS standard [37].

Fig. 63 illustrates that, for practical airborne and orbital altitudes, this level of accuracy is not attainable. Simply put, since the plenoptic camera does not significantly improve upon the ranging performance afforded by a stereo system, it is not surprising that it is not a suitable candidate for 3D terrestrial mapping from airborne and orbital platforms.

5.2 Autonomous Navigation

Though the accuracy afforded by plenoptic camera ranging is most likely not well-suited to terrestrial mapping applications, the camera does afford the sort of accuracy appropriate for the task of autonomous navigation. This is not surprising, since the performance of the plenoptic camera is demonstrated within this thesis to be very comparable to that of stereo ranging systems, which are commonly employed for robotic applications.

As a monocular system, the plenoptic camera provides a passive ranging option that requires a minimal amount of hardware and space. Light field ranging techniques such as the photo-consistency method discussed in this thesis may offer advantages over standard approaches in terms of ease of implementation and computational load.

These advantages make the plenoptic camera a likely choice for incorporation on a small, autonomous robotic system

Other technologies employed within this regime include structured light scanning and time of flight cameras. Though these technologies are often able to provide better accuracy than the plenoptic camera, as active techniques they involve increased complexity and expense. Unlike passive, image-correspondence based systems, these technologies do not simultaneously provide depth information and imagery. The need for a separate camera to provide imagery further increases the complexity, bulk, and expense of such a system.

5.3 3D Video

The ability to simultaneously collect imagery and depth information, with potential for real time processing, all with a single aperture camera, makes it difficult not to imagine easily recording 3D videos with a plenoptic camera. Outside of the realm of entertainment, this technology has applications that will probably only be fully understood as it matures and proliferates.

As technologies for displaying 3D video reach greater maturity, it seems likely that 3D video will come to play a stronger role in allowing intelligence analysts or central command headquarters to receive a better understanding of a tactical situation via 3D cameras deployed to the site of operations. This will lead to an increased demand for devices capable of capturing 3D video at minimal cost and operational difficulty. The plenoptic camera stands alongside other technologies, like stereoscopic cameras, in a position to fulfill this need.

Fig. 63 indicates that a plenoptic camera mounted on board a UAV would likely yield only coarse landscape depth information when operating at typical altitudes. In contrast, a hand-held or helmet-mounted camera might more easily succeed at

providing meaningful 3D interaction with objects within 10 to 20 meters of the observer. Security cameras of this scale might also provide additional detail sufficient to improve facial identification.

5.4 Future Development

In the near future, the plenoptic camera appears likely to find its most comfortable applications in the domains of small scale autonomous robots, hand-held 3D video recording, and any other systems requiring passively-obtained depth information for close ranges. Various advancements stand to extend this application space.

Advancements in the availability of cheap, light-weight, large-diameter optics may play a part in making larger scale plenoptic range cameras practical. Eq. 134 indicates that increasing the diameter of the camera's aperture leads to significant improvement in depth resolution ability. The plenoptic camera may assist in this effort by allowing for computational correction of optical aberrations. Aberrations tend to affect how different regions of a collecting lens map object points to the imaging plane, resulting in a broadened point spread function for the system. By separately collecting light from different aperture regions, the plenoptic camera allows for the mapping from each subaperture to be separately modified, in order to tighten the overall point spread function. The demonstration of this capability is presented in [4].

Future iterations of the plenoptic camera may no longer use microlenses to achieve angular sensitivity. Angularly sensitive pixels have been demonstrated which use stacked gratings to selectively transmit light [38]. Other sensor designs integrate optical elements directly into the detector [39]. The focused plenoptic camera model allows for gaps between the apertures to exist without resulting in spatial sampling gaps. These gaps in turn allow for a sensor design which enables smaller pixel sizes than is otherwise achievable, though at the cost of a lower SNR [39]. Any of these

technologies, by yielding an angularly-sensitive pixel smaller than the combination of pixels and lenses employed within lenslet based plenoptic cameras, could feasibly provide a significant improvement to range performance.

VI. Conclusion

6.1 Contributions

This thesis contains a number of contributions to the literature relating to plenoptic cameras. At the level of general plenoptic imaging, the thesis improves upon previous descriptions of the role of diffraction within a plenoptic camera, as in [4]. In particular, the effects of diffraction are described comprehensively and quantitatively, and it is shown that it is impossible to achieve critical sampling for all four dimension of the light field. Again, with respect to plenoptic imaging, the thesis provides a link between the traditional plenoptic camera [4], and the 'focused' plenoptic camera [24], eliminating any notion of a fundamental disparity in the capabilities of the two approaches.

Finally, the thesis provides novel analytic models for describing the uncertainty in the plenoptic camera's ranging uncertainty for a number of estimation frameworks. For example, though depth estimation using the photo-consistency technique described here has been previously demonstrated in [7], no characterization of the estimation uncertainty is provided. Again, plenoptic rangefinding within the Fourier domain is described in [12] absent of any model for range uncertainty. This thesis supplies these techniques with models which describe the scale of the uncertainty to be expected, as well as the behavior of the uncertainty as various camera parameters of varied. The outcome of these models is a recommendation concerning the camera construction yielding the best overall performance for the purposes of range estimation.

6.2 Future Work

Within the realm of plenoptic ranging itself, much can be done and has been done beyond the techniques described within this thesis. However, the comments here will be limited to work that might be done to improve upon the uncertainty modeling which constituted the primary task of this thesis. Improvements in this regard might conceivably take two forms.

First, none of the models presented in this thesis deal adequately with the impact of sampling on the image characteristics which ultimately relate to ranging accuracy. The model dealing with accuracy in the context of feature matching makes no attempt to describe the behavior of the localization error of the feature detecting algorithm. For the photo-consistency method, uncertainty is modeled in terms of the local gradient strength of the light field. However, there is no straightforward relationship between a gradient within a continuous image and the gradient within a sampled or blurred image. Rather, the effects of sampling and blurring are best understood with respect to image spatial frequencies. For a more robust uncertainty characterization, it would be necessary either to find a better description of the behavior of image gradients under resampling, or to formulate uncertainty in terms of the spatial frequency content of an image.

A different possibility would be to take a more empirical approach to uncertainty by using a large sample set with a variety of estimation methods to create a database describing the performance of different approaches in various scenarios. Both efforts would be aided by an improved plenoptic camera simulation framework. The synthetic light fields used within this research are suspect because they arise from an approach which generates subaperture images using simulated 'pinhole' cameras. Section 3.3 illustrates that subaperture images will not always have the depth of field characteristics of such a 'pinhole image.' An improved simulation framework might

also include diffraction and optional lens aberrations. Finally, a simulation employing accurate radiometry would assist correctly modeling signal to noise ratio, which is demonstrated in this thesis to strongly effect ranging performance. A simulation incorporating all of these elements would be crucial in any effort to create a meaningful database of range estimation performance.

6.3 Final Remarks

The plenoptic camera is a milestone device which promises to help usher in a new era of computational photography. The plenoptic camera's striking ability to render refocused images stems from the fact that it samples the 4D radiance distribution at its detector plane, rather than the 2D intensity distribution sampled by conventional cameras. When this distribution is formulated as a light field, a sequence of shearing and projecting the light field imitates the image formation process of a conventional camera with adjustable focal length.

The light field is distinguished from data sets collected by stereoscopic systems because it contains images obtained by an N by N grid of apertures, rather than just the two apertures of the stereoscopic system. Though these additional views enable the camera to perform novel functions like the generation of refocused imagery, it is not clear that they provide a significant advantage in terms of depth resolution.

Though theoretical considerations within this paper indicate that increasing the angular sampling density of the camera, all other things fixed, should result in a better rangefinding accuracy, experimental results indicate that the improvement may be fairly minimal. This means that, when there is a choice between spatial and angular resolution, as in the tradeoff induced by varying the microlens size in a conventional plenoptic camera, it is typically desirable to maximize spatial resolution.

Though the effectiveness of light field ranging techniques compares closely to that of techniques employed in stereoscopic computer vision, the plenoptic camera may still offer practical advantages in terms of its small footprint, low cost, and minimal need for calibration. At its present level of development, the plenoptic camera fits nicely into an application space that includes robotic navigation, 3D video recording, and security monitoring. This application space may continue to expand as developing technologies allow the camera to achieve acceptable accuracy at greater ranges.

Appendix A. Projection Slice Theorem

In this appendix, we show that the sequence of shearing, projecting, and Fourier transforming the light field is equivalent to the sequence of Fourier transforming, shearing, and slicing. We confine the math to the two dimensions s and u. However, extension to the full 4D case is straightforward. For this appendix, we drop the convention of representing normalized coordinates with over-bars. The goal is to show the following equality:

$$(\mathcal{FT} \circ \mathcal{P} \circ \mathcal{B}_{\bar{m}})[L(\mathbf{x})] = (\mathcal{S} \circ \bar{\mathcal{B}}_{\bar{m}}^{-T} \circ \mathcal{FT}^2)[L(\mathbf{x})]. \tag{135}$$

The various operators employed are defined in Table 6.

Eq. 135 is demonstrated by simply following the two sequences of operations and manipulating the result to show equivalence. The first consists of shearing, projecting, and taking the Fourier transform. The shearing operator is applied first:

$$\mathcal{B}_{\bar{m}}[L(\mathbf{x})](\mathbf{x}) = L(\mathcal{B}_{\bar{m}}^{-1}\mathbf{x}) = L\left(s + \bar{m}u, u\right). \tag{136}$$

This is followed by projection,

$$(\mathcal{P} \circ \mathcal{B}_{\bar{m}})[L(\mathbf{x})](s) = \sum_{u=0}^{N_u-1} L\left(s + \bar{m}u, u\right), \tag{137}$$

and finally, a one-dimensional Fourier transform,

$$(\mathcal{FT} \circ \mathcal{P} \circ \mathcal{B}_{\bar{m}})[L(\mathbf{x})](k_s) = \sum_{s'=0}^{N_s-1} \sum_{u=0}^{N_u-1} L(s' + \bar{m}u, u) \exp\left(-2\pi i \frac{k_s s'}{N_s}\right). \tag{138}$$

Table 6. Operator Definitions

Description	Symbol	Definition
1D Fourier Transform	$\mathcal{FT}[f(s)](k_s)$	$\sum_{s=0}^{N_s-1} f(s) \exp\left[-2\pi i \left(k_s \frac{s}{N_s}\right)\right]$
Projection	$\mathcal{P}[f(\mathbf{x})](s)$	$\sum_{u=0}^{N_u-1} f(s,u)$
Shear	$\mathcal{B}[f(\mathbf{x})](\mathbf{x})$	$f(\mathcal{B}^{-1}\mathbf{x}) \qquad \mathcal{B}_{\bar{m}} = \begin{bmatrix} 1 & -\bar{m} \\ 0 & 1 \end{bmatrix}$
Slice	$\mathcal{S}[f(\mathbf{k})](k_s)$	$f(k_s, 0)$
Modified Shear	$\bar{\mathcal{B}}[f(\mathbf{x})](\mathbf{x})$	$f(\bar{\mathcal{B}}^{-1}\mathbf{x}) \qquad \bar{\mathcal{B}}_{\bar{m}} = \begin{bmatrix} 1 & -\bar{m}N_u/N_s \\ 0 & 1 \end{bmatrix}$
2D Fourier Transform	$\mathcal{FT}^2[K(\mathbf{x})](\mathbf{k})$	$\sum_{s=0}^{N_s-1}\sum_{u=0}^{N_u-1} K(s,u) \exp\left[-2\pi i \left(k_s \frac{s}{N_s} + k_u \frac{u}{N_u}\right)\right]$

We use the substitution $s = s' + \bar{m}u$ to slightly alter the form of the equation:

$$(\mathcal{FT} \circ \mathcal{P} \circ \mathcal{B}_{\bar{m}})[L(\mathbf{x})](k_s) = \sum_{u=0}^{N_s-1} \sum_{s=\bar{m}u}^{N_u-1+\bar{m}u} L(s,u) \exp\left(-2\pi i \frac{k_s}{N_s}(s - \bar{m}u)\right). \quad (139)$$

The second route is to take the 2D Fourier transform of the light field, and then to extract a slice at the correct angle. Here, angled slicing is represented by applying the modified shear operator ($\bar{\mathcal{B}}$), and then taking the central angular slice. First, we apply to 2D Fourier transform:

$$G(\mathbf{k}) = \mathcal{FT}^2[L(\mathbf{x})](\mathbf{k}) = \sum_{s=0}^{N_s-1} \sum_{u=0}^{N_u-1} L(s,u) \exp\left(-2\pi i \left(k_s \frac{s}{N_s} + k_u \frac{u}{N_u}\right)\right). \quad (140)$$

Next, the modified shearing operator:

$$(\bar{\mathcal{B}}_{\bar{m}}^{-T} \circ \mathcal{FT}^2)[L(\mathbf{x})](\mathbf{k}) = \bar{\mathcal{B}}_{\bar{m}}^{-T}[G(\mathbf{k})](\mathbf{k}) = G\left(\bar{\mathcal{B}}_{\bar{m}}^{T}\mathbf{k}\right) = G\left(k_s, -\bar{m}\frac{N_u}{N_s}k_s + k_u\right). \quad (141)$$

Finally, the slicing operator sets k_u to zero:

$$(\mathcal{S} \circ \bar{\mathcal{B}}_{\bar{m}}^{-T} \circ \mathcal{F}\mathcal{T}^2)[L(\mathbf{x})](k_s) = G\left(k_s, -\bar{m}\frac{N_u}{N_s}k_s\right)$$
$$= \sum_{s=0}^{N_s-1}\sum_{u=0}^{N_u-1} L(s,u)\exp\left(-2\pi i\frac{k_s}{N_s}(s-\bar{m}u)\right). \quad (142)$$

Comparison reveals that Eqs. 139 and 142 are identical, save a minor difference in the limits of the summation in s.

References

1. C. Chang and S. Chatterjee, "Quantization error analysis in stereo vision," in *Conference Record of The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers*. IEEE, 1992, pp. 1037–1041.

2. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, ser. Cambridge Books Online. Cambridge University Press, 2003.

3. E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

4. R. Ng, "Digital light field photography," Ph.D. dissertation, Stanford University, 2006.

5. M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 1996, pp. 31–42.

6. R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

7. A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 51–85, 2005.

8. S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4d light fields," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 41–48.

9. A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 297–306.

10. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, 2005.

11. R. Ng, "Fourier slice photography," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 735–744.

12. Y.-H. Kao, C.-K. Liang, L.-W. Chang, and H. H. Chen, "Depth detection of light field," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1. IEEE, 2007, pp. I–893.

13. J. Bigun and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *Proceedings of the First International Conference on Computer Vision*, 1987, pp. 433–438.

14. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 1996, pp. 43–54.

15. E. L. Dereniak and G. Boreman, *Infrared detectors and systems*, ser. Wiley Series in Pure and Applied Optics. Wiley, 1996.

16. R. Fiete, *Modeling the Imaging Chain of Digital Cameras*, ser. Tutorial Text Series. SPIE Press, 2010.

17. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE CVPR, 2013, pp. 1027–1034.

18. J. Goodman, *Introduction to Fourier Optics*, ser. McGraw-Hill Physical and Quantum Electronics Series. Roberts & Company, 2005.

19. M. Eismann, *Hyperspectral Remote Sensing*. SPIE Press, 2012.

20. R. D. Fiete, "Image quality and λfn/p for remote sensing systems," *Optical Engineering*, vol. 38, no. 7, pp. 1229–1240, 1999.

21. T. Malzbender, "Fourier volume rendering," *ACM Transactions on Graphics (TOG)*, vol. 12, no. 3, pp. 233–250, 1993.

22. J. I. Jackson, C. H. Meyer, D. G. Nishimura, and A. Macovski, "Selection of a convolution function for fourier inversion using gridding [computerised tomography application]," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 473–478, 1991.

23. A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *IEEE International Conference on Computational Photography*. IEEE, 2009, pp. 1–8.

24. T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, 2010.

25. S. Wanner, J. Fehr, and B. Jähne, "Generating epi representations of 4d light fields with a single lens focused plenoptic camera," in *Advances in Visual Computing*. Springer, 2011, pp. 90–101.

26. D. Wackerly, W. Mendenhall, and R. Scheaffer, *Mathematical Statistics with Applications*. Cengage Learning, 2007.

27. M. Loeve, *Probability Theory I*, ser. Graduate Texts in Mathematics. Springer, 1978.

28. S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields," in *Vision, Modelling and Visualization (VMV)*, 2013.

29. H. Robbins and C.-H. Zhang, "Maximum likelihood estimation in regression with uniform errors," *Lecture Notes-Monograph Series*, pp. 365–385, 1986.

30. J. Conway and R. Guy, *The Book of Numbers*, ser. Copernicus Series. Springer, 1996.

31. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

32. A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

33. F. Leberl, A. Irschara, T. Pock, P. Meixner, M. Gruber, S. Scholz, and A. Wiechert, "Point clouds: Lidar versus 3d vision," *Photogrammetric Engineering and Remote Sensing*, vol. 76, no. 10, pp. 1123–1134, 2010.

34. "United states national maps accuracy standards." U.S. Bureau of the Budget, 1947.

35. "ASPRS accuracy standards for large-scale maps." American Society for Photogrammetry and Remote Sensing, 1990.

36. M. Flood, "ASPRS guidelines: Vertical accuracy reporting for lidar data." American Society for Photogrammetry and Remote Sensing, 2004.

37. "Lidar accuracy, an airborne 1 perspective." Airborne1. [Online]. Available: http://www.airborne1.com/LiDARAccuracy.pdf

38. A. Wang, P. R. Gill, and A. Molnar, "An angle-sensitive cmos imager for single-sensor 3d photography," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. IEEE, 2011, pp. 412–414.

39. K. Fife, A. El Gamal, and H.-S. Wong, "A 3d multi-aperture image sensor architecture," in *IEEE Custom Integrated Circuits Conference*. IEEE, 2006, pp. 281–284.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
27–03–2014	Master's Thesis	Oct 2012 - Mar 2014

4. TITLE AND SUBTITLE

Range Finding with a Plenoptic Camera

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Robert A. Raynor, 2LT, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENP-14-M-29

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

The plenoptic camera enables simultaneous collection of imagery and depth information by sampling the 4D light field. The light field is distinguished from data sets collected by stereoscopic systems because it contains images obtained by an N by N grid of apertures, rather than just the two apertures of the stereoscopic system. By adjusting parameters of the camera construction, it is possible to alter the number of these 'subaperture images,' often at the cost of spatial resolution within each. This research examines a variety of methods of estimating depth by determining correspondences between subaperture images. A major finding is that the additional 'apertures' provided by the plenoptic camera do not greatly improve the accuracy of depth estimation. Thus, the best overall performance will be achieved by a design which maximizes spatial resolution at the cost of angular samples. For this reason, it is not surprising that the performance of the plenoptic camera should be comparable to that of a stereoscopic system of similar scale and specifications. As with stereoscopic systems, the plenoptic camera has its most immediate, realistic applications in the domains of robotic navigation and 3D video collection.

15. SUBJECT TERMS

Plenoptic Camera, Range Finding, Light Field, Passive, Monocular, Depth, Uncertainty

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Col Karl C. Walli
U	U	U	UU	152	19b. TELEPHONE NUMBER *(include area code)* (703) 808-4932/14D00A; wallikar@nro.mil