

Computational Plenoptic Image Acquisition and Display

by

Gordon Wetzstein

Dipl., Bauhaus-Universität Weimar, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES
(Computer Science)

The University Of British Columbia
(Vancouver)

September 2011

© Gordon Wetzstein, 2011

Abstract

Recent advances in camera technology, computing hardware, and optical fabrication have led to the emergence of computational photography, a field exploring the joint design of optical light modulation and computational processing. While conventional cameras record two-dimensional images, the “ultimate” computational camera would capture all visual information with a single shot. The amount of control over such high-dimensional data is unprecedented: dynamic range, depth of field, focus, colour gamut, and time scale of a photograph can be interactively controlled in post-processing. Required visual properties include the colour spectrum as well as spatial, temporal, and directional light variation—the plenoptic function. Unfortunately, digital sensors optically integrate over the plenoptic dimensions; most of the desired information is irreversibly lost in the process.

We explore the multiplexed acquisition of the plenoptic function in this thesis. For this purpose, we introduce a mathematical framework that models plenoptic light modulation and corresponding computational processing. This framework not only allows us to evaluate and optimize the optical components of computational cameras, but also subsequent reconstructions.

The combined design of optical modulation and computational processing is not only useful for photography, displays benefit from similar ideas. Within this scope, we propose multi-layer architectures and corresponding optimization schemes for glasses-free 3D display. Compared to conventional automultiscopic displays, our devices optimize brightness, resolution, and depth of field while preserving thin form factors. In a different application, adaptive coded apertures are introduced to projection displays as next-generation auto-iris systems. Combined with computational processing that exploits limitations of human perception, these systems increase the depth of field and temporal contrast of conventional projectors. With computational optics, integrated into sunglasses or car windshields, the capabilities of the human visual system can be extended. By optically modulating perceived intensities and colours, we demonstrate applications to contrast manipulation, preattentive object detection, and visual aids for the colour blind.

Finally, we introduce computational probes as high-dimensional displays designed for computer vision applications, rather than for direct view. These probes optically encode refraction caused by transparent phenomena into observable changes in colour and intensity. Novel coding schemes enable single-shot reconstructions of transparent, refractive objects.

Preface

All publications, along with the relative contributions of collaborating authors, that have resulted from the research presented in this thesis are listed in the following.

On Plenoptic Multiplexing and Reconstruction *G. Wetzstein, I. Ihrke, W. Heidrich*, in submission.

This work is currently in submission [Wetzstein et al., 2011b] and discussed in Chapter 9. The manuscript is an extended and updated version of our CVPR 2010 conference paper [Ihrke et al., 2010b]. Dr. Ihrke had the initial idea of this project; he worked out the mathematical framework together with the author, co-wrote the manuscript, and generated some of the results and plots in the manuscript. Dr. Heidrich supervised the project and wrote parts of the paper. The author generated most of the results shown in the manuscript, wrote the software and the initial manuscript based on the conference version, revised the manuscript, and compiled the submission video.

Polarization Fields: Dynamic Light Field Display using Multi-Layer LCDs *D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, R. Raskar*, ACM Transactions on Graphics (Siggraph Asia) 2011.

Chapter 4 of this thesis is published as [Lanman et al., 2011]. Dr. Lanman wrote most of the manuscript and coordinated the project together with the author. Matthew Hirsch, together with Dr. Lanman, built and calibrated the multi-layer display prototype. Dr. Heidrich and Dr. Raskar supervised this project and contributed with insightful discussions. The author implemented off-line solvers in Matlab and a real-time solver in OpenGL, conducted most of the synthetic experiments, compiled the submission video and most figures in the manuscript, and wrote parts of the paper.

Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays *G. Wetzstein, D. Lanman, W. Heidrich, R. Raskar*, ACM Transactions on Graphics (Siggraph) 2011, Cover Feature.

This project is a collaboration with Dr. Lanman, who wrote most of the paper and coordinated the project together with the author, as well as Dr. Raskar and Dr. Heidrich,

who supervised the project and contributed with many insightful discussions. Dr. Heidrich also helped to write the paper. The author wrote parts of the paper, conducted most of the synthetic experiments, wrote the software, built the display prototypes, compiled the submission video, and co-presented the work with Dr. Lanman at Siggraph 2011. Chapter 3 contains an extended version of the paper.

Refractive Shape from Light Field Distortion *G. Wetzstein, D. Roodnick, R. Raskar, W. Heidrich*, IEEE International Conference on Computer Vision (ICCV) 2011.

Chapter 7 discusses this publication [Wetzstein et al., 2011e]. David Roodnick helped to build the hardware prototype and conducted some of the physical experiments. Dr. Heidrich and Dr. Raskar supervised this project and contributed with discussions; Dr. Heidrich also helped to edit the paper. The author conducted some of the physical experiments, coordinated the project, implemented the software, generated all synthetic results, wrote the paper, and compiled the submission video.

Hand-Held Schlieren Photography with Light Field Probes *G. Wetzstein, R. Raskar, W. Heidrich*, IEEE International Conference on Computational Photography (ICCP) 2011, Best Paper Award.

This work is discussed in Chapter 6 and was done in collaboration with Dr. Heidrich and Dr. Raskar, who contributed with discussions and comments. Dr. Heidrich also helped to write the paper. The author had the initial idea, implemented the software, built the hardware prototypes, conducted all experiments, wrote the paper, compiled the submission video, and presented the work at ICCP 2011.

State of the Art in Computational Plenoptic Imaging *G. Wetzstein, I. Ihrke, D. Lanman, W. Heidrich*, Computer Graphics Forum (Journal) 2011, presented at Eurographics 2011.

This survey paper [Wetzstein et al., 2011c,d] is part of Chapter 2. Dr. Ihrke and Dr. Lanman wrote parts of the paper. Dr. Heidrich supervised contributed with discussions. The author wrote most of the paper and co-presented the work at Eurographics 2011 together with Dr. Lanman.

Towards a Database of High-dimensional Plenoptic Images *G. Wetzstein, I. Ihrke, A. Gukov, W. Heidrich*, IEEE Conference on Computational Photography (ICCP) 2011,

Poster.

Parts of this work are included in Chapter 2. Alex Gukov built the multi-spectral camera that was used to record some of the datasets. Dr. Ihrke and Dr. Heidrich supervised the project. The author wrote all software, built the hardware setups, recorded the datasets, wrote the poster [Wetzstein et al., 2011a], and presented it at ICCP 2011.

Sensor Saturation in Fourier Multiplexed Imaging *G. Wetzstein, I. Ihrke, W. Heidrich*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.

Dr. Heidrich, Dr. Ihrke, and the author presented this work, as a poster, at CVPR 2010 [Wetzstein et al., 2010b]. All authors co-wrote the paper. Dr. Ihrke and Dr. Heidrich supervised this project. The author implemented the software, built the scanner-camera prototype, generated the results, and compiled the submission video. This project is discussed in Chapter 10.

A Theory of Plenoptic Multiplexing *I. Ihrke, G. Wetzstein, W. Heidrich*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010, oral presentation.

Dr. Ihrke and the author co-presented this work at CVPR 2010 [Ihrke et al., 2010b]. All authors co-wrote the paper. Dr. Ihrke and Dr. Heidrich supervised the project. Dr. Ihrke had the initial idea, implemented parts of the software, worked out the mathematical framework, and generated some of the results in the paper. The author helped to write the paper, revised the text and notation several times, conducted many of the experiments, implemented parts of the software, and generated some of the results. This work is discussed, in an extended form, in Chapter 9 of this thesis.

Coded Aperture Projection *M. Grosse, G. Wetzstein, A. Grundhöfer, O. Bimber*, ACM Transactions on Graphics (Journal) 2010, presented at ACM Siggraph 2010.

Chapter 5 presents a shortened version of this paper [Grosse et al., 2009]. Dr. Bimber supervised this project and wrote parts of the paper, Dr. Grundhöfer conducted some of the experiments, and Max Grosse built the projector prototype, implemented parts of the software, wrote parts of the paper, and co-presented the work at Siggraph 2010. The author also implemented parts of the software, performed all simulations, worked out the theory behind adaptive coded apertures, wrote most of the paper, helped to coordinate the project together with Dr. Bimber, co-presented the work at Siggraph 2010, and presented a preliminary version of the work as a talk at Siggraph 2009.

Optical Image Processing Using Light Modulation Displays *G. Wetzstein, D. Luebke, W. Heidrich*, Computer Graphics Forum (Journal) 2010.

The work published in this paper [Wetzstein et al., 2010a] was done in cooperation with Dr. Luebke and Dr. Heidrich. Both supervised the project and wrote most of the paper. The author implemented the software, built the hardware prototypes, conducted the experiments, and presented the work at Eurographics 2010. A discussion of this project is included in Chapter 8.

The Visual Computing of Projector-Camera Systems *O. Bimber, D. Iwai, G. Wetzstein, A. Grundhöfer*, Computer Graphics Forum (Journal) 2008.

Parts of this survey paper [Bimber et al., 2008] are included in Chapter 2. Dr Bimber supervised the project, all other authors wrote parts of the paper.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	vii
List of Tables	xiii
List of Figures	xiv
Acknowledgements	xviii
1 Introduction	1
1.1 Contributions	3
1.1.1 Contributions to Computational Displays	4
1.1.2 Contributions to Computational Illumination and Probes	5
1.1.3 Contributions to Computational Optics	6
1.1.4 Contributions to Computational Photography	6
1.2 Outline of Dissertation	7
2 Background and Related Work	9
2.1 The Plenoptic Function	9
2.2 Computational Plenoptic Image Acquisition	11
2.2.1 High Dynamic Range Imaging	12
2.2.2 Spectral Imaging	15
2.2.3 Light Field Acquisition	16
2.2.4 Multiplexing Space and Focal Surfaces	21
2.2.5 Multiplexing Time	23
2.2.6 Phase and Fluid Imaging	25
2.2.7 Acquiring Polarization	27
2.3 Plenoptic Image Display	28
2.3.1 High Dynamic Range Displays	28
2.3.2 Multi-Spectral Image Synthesis	29
2.3.3 Light Field and 3D Displays	30

2.3.4	Extended Depth of Field Projection	32
2.3.5	High-Speed Displays	33
2.3.6	Polarization-Rotating LCDs	34
2.4	Computational Optics for Direct View	35
2.4.1	Optical Image Processing	35
2.4.2	Night Vision	35
2.4.3	Augmented Reality	35
3	Tomographic Image Synthesis for Attenuation-based Multi-Layer Displays . .	37
3.1	Introduction and Motivation	37
3.1.1	Overview of Benefits and Limitations	39
3.2	Tomographic Image Generation	40
3.2.1	Modeling Volumetric Attenuation	40
3.2.2	Synthesizing Light Fields	41
3.2.3	Layered Attenuation-based Displays	44
3.3	Application to 3D Display	45
3.3.1	Assessing Performance for 3D Display	45
3.3.2	Characterizing Depth of Field	46
3.3.3	Optimizing Display Performance	49
3.4	Application to HDR Display	51
3.5	Implementation	55
3.5.1	Software	55
3.5.2	Hardware	56
3.6	Discussion	57
3.6.1	Benefits and Limitations	57
3.6.2	Future Work	59
4	Dynamic Light Field Display using Multi-Layer LCDs	61
4.1	Introduction and Motivation	61
4.1.1	Overview of Benefits and Limitations	63
4.2	Polarization Field Display	64
4.2.1	Overview of LCDs	65
4.2.2	Modeling Multi-Layer LCDs	66
4.2.3	Synthesizing Polarization Fields	67
4.3	Implementation	70
4.3.1	Hardware	71
4.3.2	Software	72
4.4	Assessment	74

4.5	Discussion	76
4.5.1	Benefits and Limitations	76
4.5.2	Future Work	78
4.5.3	Conclusion	80
5	Adaptive Coded Aperture Projection	81
5.1	Introduction	81
5.2	Coded Aperture Projection Principle	82
5.3	Prototype Designs	85
5.3.1	Static Broadband Aperture	85
5.3.2	Programmable Aperture	86
5.4	Defocus Compensation with Static Coded Apertures	86
5.4.1	Defocus Estimation and Data Preprocessing	87
5.4.2	Real-Time Compensation on the GPU	87
5.4.3	Static Aperture Results	87
5.5	Defocus Compensation with Adaptive Coded Apertures	88
5.5.1	Image Analysis	89
5.5.2	Dynamic Aperture Adaptation	91
5.5.3	Enforcing Temporal Consistency	93
5.5.4	Accounting for Different Amounts of Defocus	93
5.5.5	Incorporating Physical Constraints of the Prototype	94
5.5.6	Adaptive Coded Aperture Results	94
5.6	Evaluation and Comparison to Previous Work	95
5.7	Other Applications for Coded Apertures	98
5.8	Discussion	100
6	Hand-Held Schlieren Photography with Light Field Probes	102
6.1	Introduction and Overview	102
6.2	Theory	105
6.2.1	Image Formation	105
6.2.2	Designing Light Field Probes	106
6.3	Prototype and Experiments	107
6.3.1	Prototypes	107
6.3.2	Angular Filtering	108
6.3.3	Spatio-Angular Filtering	110
6.4	Comparison to Background Oriented Schlieren Photography	112
6.5	Limitations	112
6.6	Discussion	114

7	Refractive Shape from Light Field Distortion	116
7.1	Introduction and Motivation	116
7.2	Shape from Light Field Probes	117
7.2.1	Coding Light Field Illumination	117
7.2.2	Reconstructing Surface Normals	118
7.2.3	Point Cloud Estimation	119
7.2.4	Surface Estimation from Normals and Points	120
7.3	Experimental Results	121
7.4	Evaluation	122
7.5	Discussion	124
7.5.1	Limitations	125
7.5.2	Future Work	125
8	Optical Image Processing using Light Modulation Displays	126
8.1	Introduction	126
8.2	Prototype Designs	128
8.2.1	Monocular Scope Prototype	128
8.2.2	Window-style Prototype	129
8.3	Applications	130
8.3.1	Contrast Manipulations	130
8.3.2	Colour Manipulations	132
8.3.3	Object Highlighting Using Preattentive Cues	135
8.3.4	Defocused Light Modulation	135
8.4	Evaluation with User Study	136
8.5	Discussion	139
8.5.1	Feasibility and Limitations	139
8.5.2	Conclusions and Future Work	141
9	Plenoptic Multiplexing and Reconstruction	142
9.1	Introduction and Overview	142
9.1.1	Overview of Benefits and Limitations	143
9.2	Plenoptic Multiplexing	143
9.2.1	Plenoptic Image Formation	145
9.2.2	Basis Separation	146
9.2.3	Spatial Reconstruction	147
9.2.4	Fourier Reconstruction	149
9.2.5	Discussion	150
9.3	Application to Light Field Reconstruction	151

9.3.1	General Non-Refractive Modulators	151
9.3.2	Refractive Modulators	155
9.3.3	Plenoptic Dimension Transfer	157
9.4	Analyzing Aliasing	158
9.5	Analyzing Light Field Reconstruction Noise	159
9.6	Discussion	164
9.6.1	Benefits and Limitations	164
9.6.2	Future Work	165
9.6.3	Conclusion	165
10	Dynamic Range Boosting for Fourier Multiplexed Imaging	166
10.1	Introduction	166
10.2	Saturation Analysis in Fourier Space	168
10.3	Dynamic Range Boosting	170
10.4	Experimental Validation	173
10.4.1	Prototype	173
10.4.2	Optimization Results	174
10.4.3	Comparison to Spatial Reconstruction	175
10.4.4	Limitations	177
10.5	Combined Colour and HDR Optimization	178
10.6	Discussion	181
11	Discussion and Conclusion	182
11.1	Discussion	182
11.2	Future Work	185
11.3	Conclusion	186
	Bibliography	188
	Appendices	216
	Appendix A Additional Details on Multi-Layer Light Field Displays	216
A.1	Additional Photographs of Light Field Display Prototypes	216
A.2	Pseudo-Code for GPU-based SART Implementation	218
A.3	Synthetic Performance Evaluation of Multi-Layer Light Field Displays	219
A.4	Nonlinear Image Synthesis for Multi-Layer Displays	225
A.4.1	Derivation for Attenuation Layers	225
A.4.2	Derivation for Polarization-Rotating Layers	227

Appendix B Proofs of Plenoptic Multiplexing Theorems 229
 B.1 Proof of PSM Theorem 229
 B.2 Proof of PFM Theorem 230

List of Tables

Table 3.1	Benefits of Attenuation-based Multi-Layer Displays	40
Table 6.1	Overview of Light Field Background Oriented Schlieren Imaging	105
Table 6.2	Technical Specifications of Lenslet Arrays	107
Table 9.1	Plenoptic Multiplexing Notation	145

List of Figures

Figure 1.1	Illustration of the Scope of this Thesis	4
Figure 2.1	Multi-Spectral Light Fields	11
Figure 2.2	Taxonomy of Plenoptic Image Acquisition	12
Figure 2.3	Light Field Acquisition	18
Figure 3.1	Light Field Display using Volumetric Attenuators	38
Figure 3.2	Prototype Multi-Layer Display	39
Figure 3.3	Tomographic Analysis of Attenuation-based Displays	42
Figure 3.4	Multi-Layer 3D Display	45
Figure 3.5	Performance Evaluation with Respect to Number of Layers	47
Figure 3.6	Spectral Support for Multi-Layer Displays	48
Figure 3.7	Upper Bound on Multi-Layer Depth of Field	48
Figure 3.8	Display Performance Evaluation with Prototype	51
Figure 3.9	Display Performance Evaluation with Simulations	51
Figure 3.10	Multi-Layer, Parallax-free HDR Image Display	52
Figure 3.11	Heuristic vs. Tomographic HDR Image Synthesis	53
Figure 3.12	Contrast-Resolution Tradeoff in Multi-Layer HDR	54
Figure 3.13	Analysis of the Modulation Transfer Function	55
Figure 3.14	Benefits of Multi-Layer Automultiscopic Displays	58
Figure 3.15	Flip Animations with Automultiscopic Displays	59
Figure 3.16	Wide Field of View Light Field Display	60
Figure 4.1	Dynamic Light Field Display using Polarization Field Synthesis	62
Figure 4.2	Polarization-based vs. Attenuation-based Multi-Layer LCDs	64
Figure 4.3	Illustration of Polarization Field Setup	66
Figure 4.4	GPU-based SART vs. Off-line Solver	68
Figure 4.5	Constructing the Polarization Field Display Prototype	70
Figure 4.6	Polarization Field Display using the Multi-Layer Prototype	72
Figure 4.7	Simulated Reconstructions using Polarization Fields and Attenuators . . .	73
Figure 4.8	Radiometric Calibration of the Prototype	74
Figure 4.9	Performance of the GPU-based SART Implementation	74
Figure 4.10	Average PSNR for Attenuation Layers vs. Polarization Fields	76

Figure 4.11	Benefits of Multi-Layer Displays	77
Figure 4.12	Uncorrelated Views with Light Field Displays	77
Figure 4.13	Wide Field of View Light Field Display	79
Figure 5.1	Extended DOF Projection with Adaptive Coded Apertures	82
Figure 5.2	Different PSFs Produced by Aperture Patterns	83
Figure 5.3	Deconvolution with Different Kernels	84
Figure 5.4	Coded Aperture Projector Prototypes	86
Figure 5.5	Fourier Magnitudes	88
Figure 5.6	Static Coded Aperture Projection Results	88
Figure 5.7	Adaptive Thresholding	91
Figure 5.8	Performance Evaluation	95
Figure 5.9	Adaptive Coded Aperture Projection Results	96
Figure 5.10	Evaluation with Visual Difference Predictor	97
Figure 5.11	Depth of Field vs. Light Transmission	97
Figure 5.12	Projector Depixelation	99
Figure 5.13	Enhancing Temporal Contrast	99
Figure 6.1	Light Field Background Oriented Schlieren Photography	103
Figure 6.2	Optical Setups and Light Field Propagation	103
Figure 6.3	Refractive Index Gradients of Lens	108
Figure 6.4	Refractive Index Gradients of Plate	109
Figure 6.5	A Variety of Directional Filters	109
Figure 6.6	A Classic Rainbow Schlieren Filter	110
Figure 6.7	Gas Capture	110
Figure 6.8	Spatio-Angular Filtering	111
Figure 6.9	Comparison to BOS	113
Figure 6.10	Failure Case	114
Figure 7.1	Refraction of a Single Ray	117
Figure 7.2	Synthetic Results	120
Figure 7.3	Reconstructing Fluid Surfaces	122
Figure 7.4	Reconstructing Thin Solids	123
Figure 7.5	Qualitative Evaluation of Reconstruction	124
Figure 7.6	Quantitative Evaluation of Reconstruction	124
Figure 8.1	Concept Schematic	126
Figure 8.2	Prototypes	128
Figure 8.3	Optical Gamma Modulation	131

Figure 8.4	Contrast Reduction	131
Figure 8.5	Contrast Enhancement	132
Figure 8.6	Colour Demetamerization	134
Figure 8.7	Aiding the Colour Blind	134
Figure 8.8	Object Highlighting	136
Figure 8.9	User Study	137
Figure 8.10	User Study Evaluation	138
Figure 9.1	Overview of Multiplexed Image Reconstruction	144
Figure 9.2	Spatial and Fourier Reconstruction CFA Imaging	145
Figure 9.3	Illustration of Basis Separation	146
Figure 9.4	Light Field Parameterization	152
Figure 9.5	Illustration of Light Field Basis Separation	152
Figure 9.6	Quality Comparison of Light Field Reconstruction	154
Figure 9.7	Nonlinear Light Field Reconstruction	155
Figure 9.8	Plenoptic Manifold Reconstruction	158
Figure 9.9	Analyzing Aliasing Artifacts	159
Figure 9.10	SNR Comparison of CFAs and Light Field Masks	161
Figure 9.11	Covariance Matrices and Eigenvalues for Light Field Multiplexing	162
Figure 9.12	Light Field Attenuation Mask Noise Comparison	163
Figure 10.1	Illustration of Convolution Theorem	168
Figure 10.2	Simulated Fourier Saturation Artifacts 1D	169
Figure 10.3	Simulated Fourier Saturation Artifacts 2D	170
Figure 10.4	Scanner-Camera Prototype	173
Figure 10.5	High Dynamic Range Light Bulbs	174
Figure 10.6	Captured Fourier Saturation Artifacts	175
Figure 10.7	High Dynamic Range Clouds	176
Figure 10.8	Comparison to Assorted Pixels 1D	176
Figure 10.9	Comparison to Assorted Pixels 2D	177
Figure 10.10	Failure Case	178
Figure 10.11	Colour and HDR Capture	179
Figure 10.12	Comparison to Alternative CFAs	180
Figure A.1	Additional Photographs of Static Multi-Layer Display	216
Figure A.2	Additional Photographs of Dynamic Multi-Layer Display	217
Figure A.3	Additional Results for “Car” Scene	220
Figure A.4	Additional Results for “Dice” Scene	221

Figure A.5	Additional Results for “Dragon” Scene	222
Figure A.6	Additional Results for “Buddha” Scene	223
Figure A.7	Additional Results for “Red Dragon” Scene	224
Figure B.1	Illustration of Aliasing in Plenoptic Multiplexing	231

Acknowledgements

I wish to thank my research supervisor, Dr. Wolfgang Heidrich, for his guidance, inspiration, and support during my graduate studies. Furthermore, I would like to thank Ivo Ihrke, Douglas Lanman, Oliver Bimber, and Ramesh Raskar for inspiring discussions and fruitful collaborations. I have also benefited greatly from knowing and working with David Luebke, Anselm Grundhöfer, Matthew Hirsch, and Max Grosse. Thanks to my PhD supervisory committee members Robert Bridson and James Little as well as my external examiner Srinivasa Narasimhan. I have learned a great deal about research and life from all of you.

Without moral and financial support of my family, I would not be who or where I am today. Thank you!

Also thanks to all my friends and colleagues at UBC and in Vancouver: Stelian, Mike, Susie, Ernie, Heike, Rory, Trevor, Matt, Trent, Matt, Shaylee, Cheryl, Derek, Brad, James, Nasa, Allan, Abhi, Mike, Matthias, Anika, Emishaw, Tine, Jan, Thomas, Lukas, Lauren, Dom, Shawn, Cricket, Jonathan, Doris, Emi, Steve, Jason, Jean-Sebastian, everyone in imager, my Hapkido crew, and all my former roommates. You have made a difference!

This dissertation was supported by a Walter C Koerner Fellowship, a Theodore E Arnold Fellowship, a UBC Four Year Fellowship, and a UBC University Graduate Fellowship.

Chapter 1

Introduction

Evolution has resulted in the natural development of a variety of highly specialized visual systems among animals. The mantis shrimp retina, for instance, contains 16 different types of photoreceptors [Marshall and Oberwinkler, 1999]. The extraordinary anatomy of their eyes not only allows mantis shrimps to see 12 different colour channels, ranging from ultra-violet to infra-red, and distinguish between nuances of linear and circular polarization, but it also allows a mantis shrimp to perceive depth using trinocular vision with each eye. Other creatures of the sea, such as cephalopods [Mäthger et al., 2009], are also known to use their ability to perceive polarization for communication and unveiling transparency of their prey. Although the compound eyes found in flying insects have a lower spatial resolution compared to mammalian single-lens eyes, their temporal resolving power is far superior to the human visual system [Land and Nilsson, 2002].

Despite the tremendous advances in camera technology throughout the last decades, the basic principle of operation of modern cameras is still the same as that of Joseph Nicéphore Niépce’s camera, which he used to capture the first permanent photograph in 1826 [Gustavson and House, 2009]. Digital sensors, converting photons into electrons, have replaced light sensitive resins and films and on-board image processing using integrated computing hardware is now common practice, even for consumer-grade digital cameras. However, the acquired visual information has always been what a single human eye can perceive: a two-dimensional trichromatic image. Inspired by the natural diversity of perceptual systems and fueled by advances of digital camera technology, computational processing, and optical fabrication, image acquisition has begun to transcend limitations of film-based analog photography.

Computational photography, that is joint optical light modulation and computational processing of captured data, has emerged as an interdisciplinary field—spanning optics, sensor technology, image processing, and illumination—that is dedicated to the exploration of sophisticated approaches to capturing, analyzing, and processing visual information. Many of the proposed techniques aim at acquiring the dimensions of the plenoptic function [Adel-

son and Bergen, 1991] with combined optical modulation and computational reconstruction [Wetzstein et al., 2011d]. The plenoptic function provides a ray-based model of light encompassing most properties of interest for image acquisition, including the colour spectrum as well as spatial, temporal, and directional light variation. In addition to these more traditional plenoptic dimensions, we also consider dynamic range¹, polarization, and phase changes caused by refraction as desirable properties that can be associated with light rays.

Applications for the computerized acquisition of images with a high spatial, temporal, spectral, and directional resolution are manifold: medical imaging, remote sensing, shape reconstruction, surveillance, and automated fabrication are only a few examples. The amount of control over how such high-dimensional visual information can be converted into a presented image is unprecedented. Properties such as dynamic range, depth of field, focus, colour gamut, and time scale can be interactively controlled in post-processing after the data is captured.

Independent of the developments in the computational photography community, the film industry has been pushing stereoscopic image capture and presentation into the mass-market throughout the last few years with blockbuster movies such as ‘Avatar’. However, the underlying technology has not significantly changed since the invention of the stereoscope by Charles Wheatstone in 1838. Binocular disparity, captured in two views of the same scene, as observed from slightly different viewpoints, is presented to a human observer; most commonly requiring the viewer to wear specialized glasses. Automultiscopic displays that do not require special eyewear were invented at the beginning of the last century. Nevertheless, parallax barriers [Ives, 1903] and lenslet arrays [Lippmann, 1908] remain the two dominating technologies for glasses-free 3D displays. Only most recently, research has started to explore *computational displays* that exploit joint optical display design and computational processing in a similar fashion as computational photography (e.g., Lanman et al. [2010]; Seetzen et al. [2004]).

One of the key arguments throughout this thesis is that the combination of optical light modulation and computational processing is equally useful for photography and displays. We confirm this argument by designing and implementing novel approaches that push the boundaries of conventional light capture and display technology. The plenoptic function and the light field [Levoy and Hanrahan, 1996], which is the subset containing only spatio-angular plenoptic information, provide intuitive tools to characterize and analyze light, as it is displayed, interacts with matter, and is sensed either by a digital camera or a human observer. As outlined in Figure 1.1, two approaches to glasses-free 3D image presentation

¹Throughout this thesis, dynamic range is defined as the ratio of largest and smallest possible value in the range, as opposed to the domain, of the plenoptic function.

using computational displays are presented in Chapters 3 and 4. However, combined optical light modulation and computational processing is not only useful for 3D image synthesis, but also for extending the depth of field of projection displays, as discussed in Chapter 5. Instead of targeting human observers, light field displays also have applications in computer vision. We term displays that are specifically designed for computer vision applications, rather than for direct view, *computational probes*. Chapters 6 and 7 discuss applications of high-dimensional probes for the visualization of light refractions caused by transparent media and the reconstruction of certain classes of such media. *Computational optics*, that is optical image processing combined with automated, digital control, has the power to enhance the human visual system, as introduced in Chapter 8. Chapters 9 and 10 present a new theory and applications of optical modulation along with computerized reconstruction for acquiring high-dimensional plenoptic data.

The “ultimate” imaging system, toward which the approaches discussed in this thesis are steps, can be described as the development of technology that allows all dimensions of the plenoptic function to be displayed and captured with a single shot. Recording this kind of visual information, basically encompassing all measurable attributes of light rays, would also allow a maximum degree of freedom for post-processing captured data. The key requirement for any such approach is the optical control over the dimensions of the plenoptic function, which can also be used for directly modulating perceived imagery, thereby extending the capabilities of the human visual system.

1.1 Contributions

This dissertation contributes to a number of fields including computer graphics, computer vision, display design, phase imaging, augmented reality, optics, signal processing, and computerized tomography. The large variety is due to the fact that all of the presented work is inspired by ideas that blur the boundaries between traditional research areas by jointly designing computational processing and algorithms in unison with camera optics, display hardware, or illumination. For the purpose of this thesis, the following list of contributions, as well as the remaining chapters, are organized as *computational displays*, *computational illumination and probes*, *computational optics*, and *computational photography*.

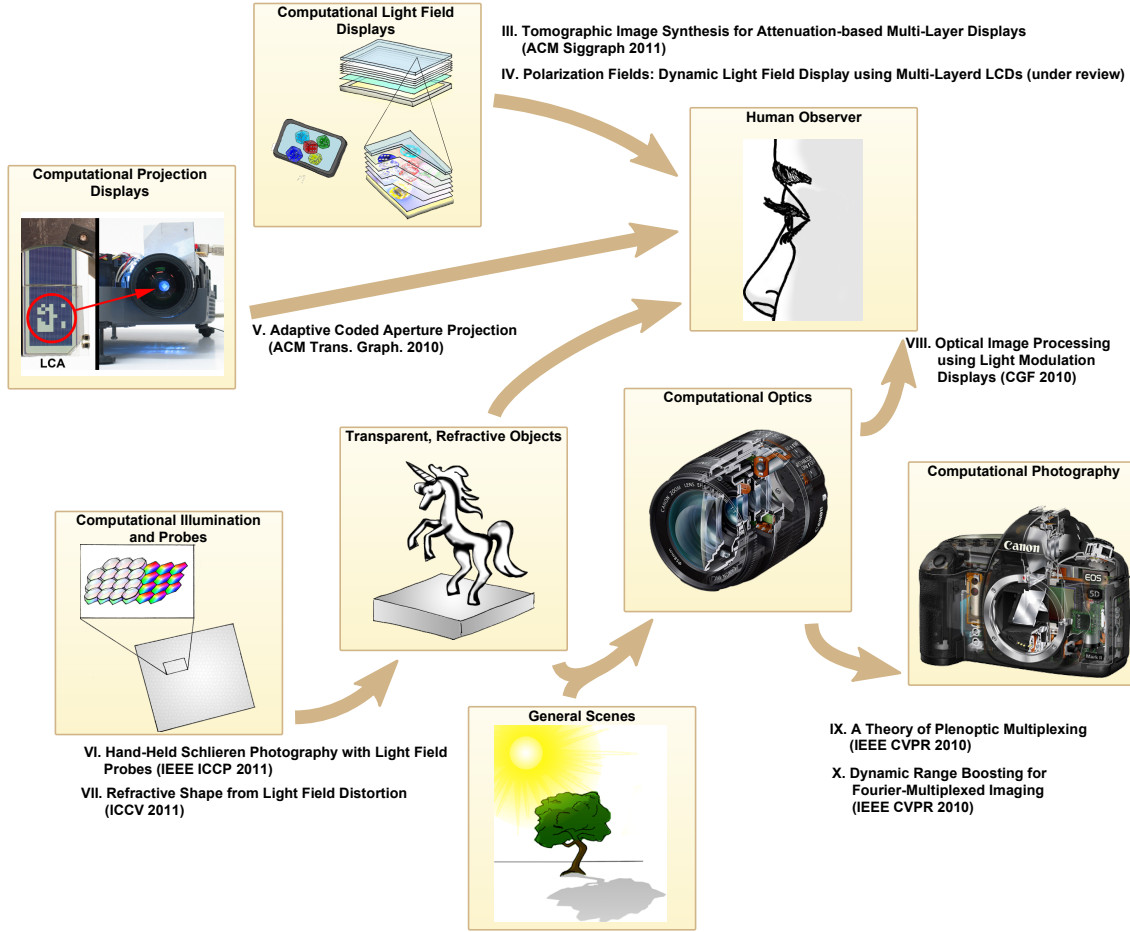


Figure 1.1: *Illustration of the scope of this thesis. A variety of applications for joint optical light modulation and computational processing is introduced: glasses-free 3D image synthesis using computational light field displays (Chapters 3 and 4), extended depth of field projection by means of adaptive coded apertures (Chapter 5), computational illumination and light field probes for visualizing and reconstructing refraction caused by transparent objects (Chapters 6 and 7), computational optics to enhance the power of the human visual system (Chapter 8), and approaches to overcome the limitations of standard image sensors with computational photography (Chapters 9 and 10).*

1.1.1 Contributions to Computational Displays

We demonstrate the practical benefits of multi-layered attenuators for light field display, compared to dual-layer devices, and establish theoretical limitations of all such displays. In Chapter 3, we present a tomographic approach that allows light fields to be optimally displayed by volumetric or layered attenuators. Compared to conventional parallax barriers, multi-layer displays achieve higher resolution, extended depth of field, and increased brightness. We show the proposed tomographic method also encompasses 2D high dynamic range

(HDR) display. With this method, we demonstrate the first HDR display using multiple, disjoint attenuators. For the case of dual-layer architectures, we confirm existing heuristic algorithms [Seetzen et al., 2004].

In Chapter 4, we introduce polarization field displays for optically-efficient light field display using multi-layer LCDs. Just like attenuation-based multi-layer displays, we demonstrate increased brightness, higher resolution, and extended depth of field, compared to existing dual-layer architectures. We propose to cast polarization-based light field display as a constrained linear least-squares problem and show that the simultaneous algebraic reconstruction technique (SART) allows dynamic light field display for both polarization-based and attenuation-based multi-layer LCD architectures. Interactive display is verified on a prototype multi-layer LCD using a GPU-based SART solver.

We introduce dynamic coded apertures to projection displays in Chapter 5. Compared to conventional projectors, we demonstrate improved depth of field and temporal contrast, at interactive framerates, using an integrated hardware solution and corresponding algorithms. We develop a content-dependent algorithm for computing dynamic adaptive projector apertures that exploit limitations of the human visual system.

1.1.2 Contributions to Computational Illumination and Probes

The computer vision community has been employing probes for decades. Most commonly, these are planar, diffuse checkerboard patterns used for camera calibration (e.g., Tsai [1987], Zhang [2000]). The diffuse nature of such probes makes them two-dimensional, because every point on the surface emits the same colour and intensity into all directions. In Chapter 6, we introduce the concept of computational light field probes for recording and processing new kinds of visual information with off-the-shelf cameras. The light emitted by these probes varies with respect to the position and direction on its surface, making them four-dimensional. We construct inexpensive light field probe prototypes using lenslet arrays and inkjet transparencies.

Schlieren photography is a non-intrusive method to visualize and capture transparent, refractive media. We present a new type of Background Oriented Schlieren imaging that is portable, alleviates the common problem of focus discrepancies, and allows spatial and angular information to be coded. We propose to capture transparent, refractive media with light field probes and develop optical coding schemes that are designed for efficient acquisition of such media. In Chapter 7, we demonstrate single image reconstruction of certain types of non-stationary transparent objects with the proposed techniques.

1.1.3 Contributions to Computational Optics

In Chapter 8, we introduce the concept of optical image processing, by means of light modulation, for enhancing the capabilities of the human visual system. Two hardware prototypes are built and serve as a platform to implement several perceptually-motivated applications of optical contrast manipulation, including the application of gamma curves, tone mapping-style contrast reduction, and contrast enhancement using counter-shading. We show the use of spatially modulated colour filters to perform de-metamerization and colour saturation enhancement or reduction, as well as providing visual aids for the colour blind. Object highlighting using the manipulation of colour, intensity, or contrast is discussed as another application. We analyze and demonstrate both in-focus and out-of-focus geometries for applications such as the above as well as both in-line and parallax geometries for the camera and observer. The effect of our contrast reduction approach on the perception of low-contrast details in high-contrast scenes is evaluated with a user study.

1.1.4 Contributions to Computational Photography

In Chapter 9, we show the capture and reconstruction of different dimensions of the plenoptic function to be very similar in nature. We demonstrate that sophisticated reconstruction methods developed for one dimension can be similarly applied to other dimensions. A mathematical framework for image formation in plenoptic multiplexing applications is developed. This model generalizes both spatial and Fourier multiplexing methods that have been proposed independently in the literature. We present, for the first time, spatial reconstructions of Fourier-multiplexed light fields and other plenoptic manifolds and show that the resulting image quality can be significantly increased. We also establish a metric for the quantitative evaluation of attenuation masks used in light field acquisition. This metric along with our framework allows us to compare a variety of attenuation patterns and analyze their performance with respect to sensor noise amplification.

The effects of sensor saturation for multiplexed image reconstruction in the Fourier domain are analyzed in Chapter 10. We propose a numerical optimization method for Fourier-based dynamic range boosting of multiplexed data. The approach is introduced for dynamic range multiplexing with neutral density filters. A prototype that allows novel colour filter arrays and other multiplexing masks to be developed and tested in a macroscopic scale is constructed. We introduce a novel colour filter array that allows our dynamic range boosting technique to be applied to colour photographs.

1.2 Outline of Dissertation

This section outlines the remaining chapters of the thesis.

Chapter 2. Background and Related Work. The basic concepts of the plenoptic function are reviewed in this chapter, along with the state of the art in computational plenoptic image acquisition and display as well as computational optics for direct view.

Chapter 3. Tomographic Image Synthesis for Attenuation-based Multi-Layer Displays. This chapter develops tomographic techniques for image synthesis on multi-layer attenuation-based displays. These displays are shown, both by theory and experiment, to exceed the performance of existing dual-layer architectures. For 3D display, spatial resolution, depth of field, and brightness are increased, compared to parallax barriers. For a plane at a fixed depth, the proposed optimization also allows optimal construction of high dynamic range displays, confirming existing heuristics and providing the first extension to multiple, disjoint layers.

Chapter 4. Dynamic Light Field Display using Multi-Layer LCDs. Polarization field displays are introduced as an optically-efficient construction allowing dynamic light field display using multi-layered LCDs. It is demonstrated that such displays can be controlled, at interactive refresh rates, by tomographic techniques that solve for the optimal spatially-varying polarization state rotations applied by each layer. The design is verified by constructing a prototype multi-layer LCD using modified off-the-shelf panels.

Chapter 5. Adaptive Coded Aperture Projection. This chapter demonstrates extended depth of field projection using adaptive coded apertures in combination with inverse filtering; two prototypes and corresponding algorithms for static and programmable projector apertures are presented.

Chapter 6. Hand-Held Schlieren Photography with Light Field Probes. A new approach to capturing refraction in transparent media is introduced. By optically coding the locations and directions of light rays emerging from a light field probe, changes of the refractive index field between the probe and a camera or an observer are captured.

Chapter 7. Refractive Shape from Light Field Distortion. Acquiring transparent, refractive objects is challenging as these kinds of objects can only be observed by analyzing the distortion of reference background patterns. This chapter presents a new, single image approach to reconstructing thin transparent surfaces, such as thin solids or surfaces of fluids, from the observed distortion of a light field probe.

Chapter 8. Optical Image Processing using Light Modulation Displays. This chapter proposes to enhance the capabilities of the human visual system by performing optical image processing directly on an observed scene. A number of perceptually-motivated algorithms including contrast manipulation, object highlighting for preattentive emphasis, colour saturation, de-saturation, and de-metamerization, as well as visual enhancement for the colour blind are demonstrated.

Chapter 9. On Plenoptic Multiplexing and Reconstruction. This chapter develops a mathematical framework that generalizes multiplexed imaging to all dimensions of the plenoptic function. This framework unifies a wide variety of existing approaches to analyze and reconstruct multiplexed data in either the spatial or the frequency domain. Many practical applications of this framework are demonstrated, including high-quality light field reconstruction, the first comparative noise analysis of light field attenuation masks, and an analysis of aliasing in multiplexing applications.

Chapter 10. Dynamic Range Boosting for Fourier Multiplexed Imaging. This chapter analyzes sensor saturation in Fourier-based image multiplexing approaches and proposes a computational photography approach to extending the dynamic range of captured images.

Chapter 11. Discussion and Conclusion. The contributions of this thesis are summarized and future directions of research are outlined.

Chapter 2

Background and Related Work

In this chapter, we review the state of the art in joint optical light modulation and computational processing for acquiring and displaying the dimensions of the plenoptic function. Furthermore, we discuss computational optics that are tailored for human perception. Specifically, we review the plenoptic function and related concepts in Section 2.1. A survey of computational plenoptic image acquisition follows in Section 2.2. We use an intuitive categorization based on plenoptic dimensions and hardware setups. Chapter 2.3 outlines approaches to plenoptic image display and Chapter 2.4 discusses computational optics approaches to enrich or enhance the human visual system.

2.1 The Plenoptic Function

The study of light has been pursued for centuries. Early work on the subject dates back to Ibn al-Haytham’s “Book of Optics” (11th century) and Leonardo da Vinci’s notebooks (15th century). Only recently, a comprehensive, ray-based model of light encompassing all perceivable properties has been proposed: the plenoptic function [Adelson and Bergen, 1991].

Consider a pinhole camera. Each sample on the sensor plane is a ray passing through the pinhole with a different direction of propagation, but at a fixed position in space. The recorded image is a 2D subset of all light rays in the scene, the full set being a 5D function that models the directional light distribution at all possible positions in 3D space. Radiance of light rays, however, does not change along their paths in free space. Hence, a 4D slice of the plenoptic function fully describes the spatio-angular light variation in a space free of occluders. In computer graphics and vision, this 4D set of rays is referred to as the *light field* [Gortler et al., 1996; Levoy and Hanrahan, 1996]. Light fields are used as a fundamental tool to analyze and characterize image acquisition and display systems throughout this thesis.

As opposed to light fields, the full 7D plenoptic function also considers the wavelength of

light as well as temporal variation. Polarization and phase are usually disregarded, because these properties are associated with wave-based light models. Figure 2.1 shows two different datasets that each contain a 5D subset of the plenoptic dimensions. These multi-spectral light fields are acquired by mounting a custom multi-spectral camera on a programmable X–Y translation stage [Wetzstein et al., 2011a]. Each of the different light field views is colour-coded with one of the acquired spectral channels. The larger dataset contains 15×15 perspectives of a scene, each containing 23 narrow-band colour channels ranging 460 nm to 680 nm. Each channel has a spatial resolution of 341×512 and is stored as a high dynamic range image composited from multiple exposures. Without image compression, this dataset alone requires 27 GB of storage and about 16 hours to be captured. The sheer size and the amount of required resources make it difficult to capture and process any such plenoptic data. Most acquisition setups (Sec. 2.2) therefore only consider subsets of the plenoptic function.

Limited resources for sensing and processing are not only a problem for machine vision, but also for the visual systems of animals. Natural evolution has resulted in a variety of different tradeoffs in these systems, each of which is optimized for survival in its natural environment. Animals with compound eyes, for example, often have an impressive spectral and temporal resolving power, but a limited spatial resolution [Land and Nilsson, 2002]. The human visual system (HVS), on the other hand, has its own limitations. The tristimulus nature of human colour perception severely limits our ability to resolve spectral distributions. Without adaptation, the maximum contrast that the HVS can resolve at any given time is about 10,000:1 [Reinhard et al., 2010]. Humans have a temporal resolving power of about 30 images per second [Goldstein, 2006] and sample the plenoptic function at the two different positions of the eyes. Unlike mantis shrimps, we are not, to any significant amount, sensitive to polarization.

Considering the limitations of the human visual system, the attempt to simultaneously acquire all plenoptic dimensions may seem redundant. However, captured visual information is not only important for displaying images to human observers but also for most computer vision applications. The degrees of freedom to post-process captured data, reconstruct additional information of the photographed scene, and potentially optimize the content for a specific display increase by magnitudes.

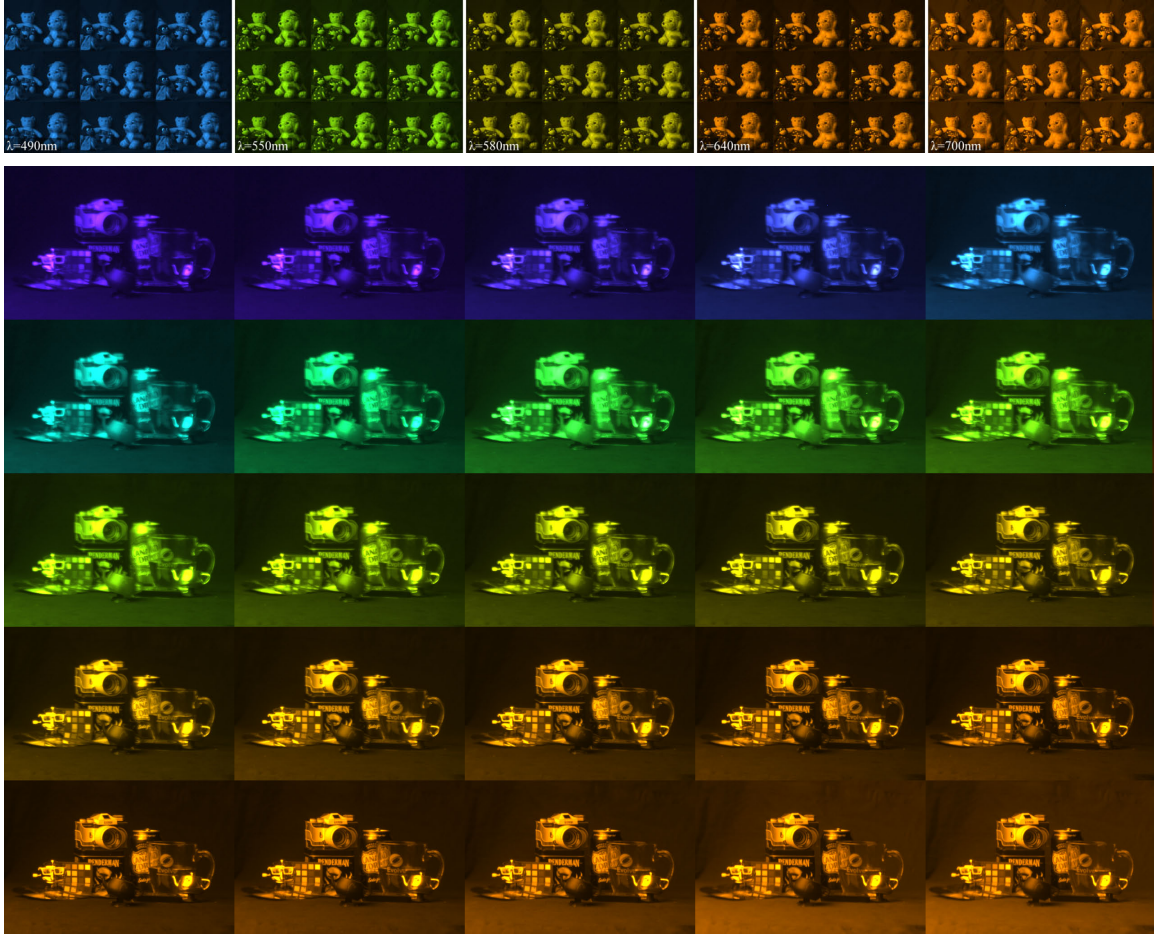


Figure 2.1: *Top: mosaics showing 3×3 viewpoints and five colour channels of a multi-spectral light field recorded from 5×5 viewpoints with 10 colour channels for each viewpoint. Bottom: another multi-spectral light field dataset with 15×15 viewpoints and 23 narrow-band colour channels for each viewpoint. The spectral channels range from 460 nm to 680 nm in 10 nm increments. Only 5×5 viewpoints are shown in this mosaic and each of those is colour-coded with one of the recorded colour channels for that viewpoint. This scene includes a variety of illumination effects including diffraction, refraction, inter-reflections, and specularities.*

2.2 Computational Plenoptic Image Acquisition

Standard sensors integrate over all of the dimensions of the plenoptic function. In order to record specific aspects, however, a lot of research has been done which will be outlined in the following. In the literature, these approaches have usually been categorized according to the imaged plenoptic dimension (see Figure 2.2). We use the same classification: we discuss high dynamic range imaging (Sec. 2.2.1), the acquisition of the colour spectrum (Sec. 2.2.2), light


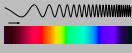
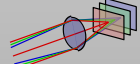
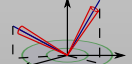



















Plenoptic Dimension					
Acquisition Approach	Dynamic Range	Color Spectrum	Space Focal Surfaces	Directions Light Fields	Time
 Single Shot Acquisition	 Assorted Pixels Gradient Camera Adaptive DR Imaging	 Color Filter Arrays Assorted Pixels Dispersive Optics	 Coded Apertures Focal Sweep Field Correction	 Plenoptic Cameras w/ Lenses, Masks, or Mirrors Compound Eye Cameras	 Assorted Pixels Flutter Shutter Reinterpretable Imager Sensor Motion
 Sequential Image Capture	 Exposure Brackets Generalized Mosaics HDR Video	 Narrow Band Filters Generalized Mosaicing Agile Spectrum Imaging	 Focal Stack Jitter Camera Super-Resolution	 Programmable Aperture Camera & Gantry	 High-Speed Imaging Temporal Dithering
 Multi-Device Setup	 Split Aperture Imaging Optical Splitting Trees	 Multi-Camera Arrays Optical Splitting Trees	 Multi-Camera Arrays	 Multi-Camera Arrays	 Multi-Camera Arrays Hybrid Cameras

Figure 2.2: *Taxonomy and overview of plenoptic image acquisition approaches.*

field capture (Sec. 2.2.3), spatial super-resolution and focal surfaces (Sec. 2.2.4), as well as high-speed imaging (Sec. 2.2.5). We also outline the acquisition of two light properties that are not directly included in the plenoptic function, but related: phase imaging (Sec. 2.2.6) and polarization (Sec. 2.2.7). Due to the fact that modern, digital acquisition approaches are often closely related to their analog predecessors, we outline these whenever applicable.

2.2.1 High Dynamic Range Imaging

High dynamic range (HDR) image acquisition has been a very active area of research for more than a decade. With the introduction of the HDR display prototype [Seetzen et al., 2004] and its successor models becoming consumer products today, the demand for high-contrast photographic material is ever increasing. Other applications for high dynamic range imagery include digital photography, physically-based rendering and lighting [Debevec, 2002], image editing, perceptual difference metrics based on absolute luminance [Mantiuk et al., 2005, 2011], virtual reality, and computer games. For a comprehensive overview of HDR imaging, including applications, radiometry, perception, data formats, tone reproduction, and display, the reader is referred to the textbook by Reinhard et al. [2010].

Single-Shot Acquisition

According to DxOMark (www.dxomark.com), the latest digital SLR cameras are equipped with CMOS sensors that have a measured dynamic range of up to 13.5 f-stops, which translates to a contrast of 11,000:1. This is comparable to that of colour negative films [Reinhard et al., 2010]. In the future, we can expect digital sensors to perform equally well as negative

film in terms of dynamic range, but this is not the case for most sensors today.

Specialized sensors that allow high dynamic range content to be captured have been commercially available for a few years. These include professional movie cameras, such as Grass Valley’s Viper [Valley, 2010] and Panavision’s Genesis [Panavision, 2010] as well as the SpheroCam HDR [SpheronVR, 2010], which is able to capture full spherical 360-degree images with 26 f-stops and 50 megapixels in a single scan. A technology that allows per-pixel exposure control on the sensor, thereby enabling adaptive high dynamic range capture, was introduced by Pixim [2010]. This level of control is achieved by including an analog-to-digital converter for each pixel on the sensor.

Capturing image gradients rather than actual pixel intensities was shown to increase the dynamic range of recorded content [Tumblin et al., 2005]. In order to reconstruct intensity values, a computationally expensive Poisson solver needs to be applied to the measured data. While a *Gradient Camera* is an interesting theoretical concept, to the knowledge of the author this camera has never actually been built.

The maximum intensity that can be resolved with standard neutral density (ND) filter arrays, as opposed to colour filter arrays, is limited by the lowest transmission of the employed ND filters. Large, completely saturated regions in the sensor image are usually filled with data interpolated from neighboring unsaturated regions [Nayar and Mitsunaga, 2000]. An analysis of sensor saturation in multiplexed imaging, along with a Fourier-based reconstruction technique that boosts the dynamic range of captured images beyond the previous limits, is presented in Chapter 10.

An alternative to mounting a fixed set of ND filters in front of the sensor is an aligned spatial light modulator, such as a digital micromirror device (DMD). This concept was explored as *Programmable Imaging* [Nayar et al., 2004, 2006] and allows for adaptive control over the exposure of each pixel. Unfortunately, it is rather difficult to align a DMD with a sensor on a pixel-precise basis, partly due to the required relay optics [Ri et al., 2006]. Although a transmissive spatial light modulator can, alternatively, be mounted near the aperture plane of the camera, as proposed by Nayar and Branzoi [2003], this *Adaptive Dynamic Range Imaging* approach only allows lower spatial frequencies in the image to be modulated. Whereas programmable imaging and adaptive dynamic range imaging manipulate the radiometric characteristics of the incoming light to aid photography and computer vision, we present similar manipulations intended for direct view in Chapter 8. Our target is the human visual system rather than a photographic or computational camera. This opens up a wide range of perceptually-motivated applications, as discussed in Chapter 8.

The most practical approach to adaptive camera exposures is a per-pixel control of the

readout in software, as implemented by the Pixim camera [Pixim, 2010]. This has also been simulated for the specific case of CMOS sensors with rolling shutters [Gu et al., 2010], but only on a per-scanline basis. The next version of the *Frankencamera* [Adams et al., 2010] is planned to provide non-destructive sensor readout for small image regions of interest [Levoy, 2010], which would be close to the desired per-pixel exposure control. Rouf et al. [2011] propose to encode both saturated highlights and low-dynamic range content in a single sensor image using cross-screen filters. Computerized tomographic reconstruction techniques are employed to estimate the saturated regions from glare created by the optical filters.

Multi-Sensor and Multi-Exposure Techniques

The most straightforward way of acquiring high dynamic range images is to sequentially capture multiple photographs with different exposure times and merge them into a single, high-contrast image [Debevec and Malik, 1997; Mann and Picard, 1995; Mitsunaga and Nayar, 1999; Robertson et al., 1999]. Some of these approaches simultaneously compute the nonlinear camera response function from the image sequence [Debevec and Malik, 1997; Mitsunaga and Nayar, 1999; Robertson et al., 1999]. Extensions to these techniques also allow HDR video [Kang et al., 2003]. Here, successive frames in the video are captured with varying exposure times and aligned using optical flow algorithms. Today, all of these methods are well established and discussed in the textbook by Reinhard et al. [2010].

In addition to capturing multiple exposures, a static filter with varying transmissivity, termed *Generalized Mosaicing* [Schechner and Nayar, 2003a], can be mounted in front of the camera but also requires multiple photographs to be captured. Alternatively, the optical path of an imaging device can be divided using prisms [Aggarwal and Ahuja, 2004] (*Split Aperture Imaging*) or beam-splitters [McGuire et al., 2007] (*Optical Splitting Trees*), so that multiple sensors capture the same scene with different exposure times. While these approaches allow dynamic content to be recorded, the additional optical elements and sensor hardware make them more expensive and increase the form factor of the device.

Analysis and Tradeoffs

Given a camera with known response function and dynamic range, Grossberg and Nayar [2003] and Hasinoff et al. [2010] analyze the best possible set of actual exposure values for a low dynamic range (LDR) image sequence used to compute an HDR photograph.

2.2.2 Spectral Imaging

Approaches to capturing colour information in a photograph date back almost as far as the first cameras themselves. James Clerk Maxwell [Maxwell, 1860] is usually credited with the discovery that any colour imaging system only needs to capture and display three different spectral bands in order to faithfully represent colour information for a human observer. One of the earliest collections of colour photographs was assembled by the Russian photographer Prokudin-Gorskii [1912]. *Time-sequential imaging* through different (programmable) colour filters remains one of the most important means to capture multi-spectral images with a large number of colour channels (e.g., CRI-INC [2009]).

The *instantaneous capture* of colour information can, alternatively, be achieved using *multiple sensors*. Three-sensor cameras, for instance, apply dichroic beam-splitter prisms (e.g., Optec [2011]) to optically separate different colour channels and image them with different sensors. Optical splitting trees [McGuire et al., 2007] generalize this concept to a variable number of sensors and corresponding optical information. A different principle, based on volumetric measurements, is employed by the Foveon sensor [Foveon, 2010], which captures tri-chromatic images at full spatial resolution.

The most popular approach to colour imaging in consumer cameras, however, is multiplexing with colour filter arrays (CFAs). Here, an interleaved array of colour filters, for instance a Bayer pattern [Bayer, 1976], is mounted directly on a monochromatic sensor, so that neighboring pixels record differently filtered light. A wide variety of CFAs exist [Compton, 2007; Hirakawa and Parks, 2006; Hirakawa and Wolfe, 2008; Lu and Vetterli, 2009], where different designs are optimized for different aspects of the imaging process. More recently, Sajadi et al. [2011] propose to use shiftable layers of CFAs; this design allows the colour primaries to be switched dynamically and provides an optimal SNR in different lighting conditions. *Assorted Pixels* [Narasimhan and Nayar, 2005; Yasuma et al., 2010] is a concept generalizing CFAs to arbitrary optical filter array configurations.

For all multiplexing approaches using filter arrays, a full-resolution image is usually reconstructed by interpolating all colour samples to every pixel. This is referred to as colour demosaicing. An overview of demosaicing techniques can be found in [Gunturk et al., 2005; Li et al., 2008b; Ramanath et al., 2002].

In Chapter 9, we generalize the multiplexed acquisition and reconstruction of colour information to all dimensions of the plenoptic function. We demonstrate that sophisticated reconstruction methods developed for one dimension, such as colour demosaicing, can be similarly applied to other plenoptic dimensions. In a different application, Chapter 10 ex-

plores CFA designs that allow colour information to be captured along with a high dynamic range. We present an optimized CFA pattern that, in conjunction with computational processing of the recorded data, can extend the dynamic range of a captured photograph.

2.2.3 Light Field Acquisition

A 5D subset, including 3D spatial and 2D angular variation, of the plenoptic function parameterizes all possible images of a general scene. Such a representation is a slice of constant time and wavelength of the full plenoptic function. Levoy and Hanrahan [1996] and Gortler et al. [1996] realized that, when the viewer is restricted to move outside the convex hull of an object, the 5D plenoptic function possesses one dimension of redundancy: the radiance of a given ray does not change in free space. Thus, in a region free of occluders, the 5D plenoptic function can be expressed as a 4D light field.

The concept of a light field predates its introduction in computer graphics. The term itself dates to the work of Gershun [1936], who derived closed-form expressions for illumination patterns projected by area light sources. Ashdown [1993] continued this line of research. Moon and Spencer [1981] introduced the equivalent concept of a *photic field* and applied it to topics spanning lighting design, photography, and solar heating. The concept of a light field is similar to epipolar volumes in computer vision [Bolles et al., 1987]. As demonstrated by Halle [1994], both epipolar volumes and holographic stereograms can be captured by uniform camera translations. The concept of capturing a 4D light field, for example by translating a single camera [Gortler et al., 1996; Levoy and Hanrahan, 1996] or by using an array of cameras [Wilburn et al., 2002], is predated by integral photography [Lippmann, 1908], parallax panoramagrams [Ives, 1903], and holography [Gabor, 1948].

This subsection catalogues existing devices and methods for light field capture, as well as applications enabled by such datasets. Note that a sensor pixel in a conventional camera averages the radiance of light rays impinging over the full hemisphere of incidence angles, producing a 2D projection of the 4D light field. In contrast, light field cameras prevent such averaging by introducing spatio-angular selectivity. Such cameras can be classified into those that primarily rely on multiple sensors or a single sensor augmented by temporal, spatial, or frequency-domain multiplexing.

Multiple Sensors

As described by Levoy and Hanrahan [1996], a light field can be measured by capturing a set of photographs taken by an array of cameras distributed on a planar surface. Each

camera measures the radiance of light rays incident on a single point, defined in the plane of the cameras, for a set of angles determined by the field of view of each camera. Thus, each camera records a 2D slice of the 4D light field. Concatenating these slices yields an estimate of the light field. Wilburn et al. [2005, 2002] achieve dynamic light field capture using an array of up to 125 digital video cameras (see Fig. 2.3, left). Yang et al. [2002] propose a similar system using 64 cameras. Nomura et al. [2007] create scene collages using up to 20 cameras attached to a flexible plastic sheet, combining the benefits of both multiple sensors and temporal multiplexing. Custom hardware allows accurate calibration and synchronization of the camera arrays. Such designs have several unique properties. Foremost, as demonstrated by Vaish et al. [2006], the captured light field can be considered as if it were captured using a single camera with a main lens aperture extending over the region occupied by the cameras. Such large-format cameras can not be practically constructed using refractive optics. Vaish et al. exploit this configuration by applying methods of synthetic aperture imaging to obtain sharp images of objects obscured by thick foliage.

Temporal Multiplexing

Camera arrays have several significant limitations; foremost, a sparse array of cameras may not provide sufficient light field resolution for certain applications. In addition, the cost and engineering complexity of such systems prohibit their use for many consumer applications. As an alternative, methods using a single image sensor have been developed. For example, Levoy and Hanrahan [1996] propose a direct solution; using a mechanical gantry, a single camera is translated over a spherical or planar surface, constantly reoriented to point towards the object of interest. Alternatively, the object can be mechanically rotated on a computer-controlled turntable. Ihrke et al. [2008] substitute mechanical translation of a camera with rotation of a planar mirror, effectively creating a time-multiplexed series of virtual cameras. Thus, by distributing the measurements over time, single-sensor light field capture is achieved. Taguchi et al. [2010a] show how capturing multiple images of rotationally-symmetric mirrors from different camera positions allow wide field of view light fields to be captured. Gortler et al. [1996] propose a similar solution; the camera is manually translated and computer vision algorithms are used to estimate the light field from such uncontrolled translations. These approaches trace their origins to the method introduced by Chen and Williams [1993], which is implemented by QuickTime VR.

The preceding systems capture the light field impinging on surfaces enveloping large regions (e.g., a sphere encompassing the convex hull of a sculpture). In contrast, hand-held light

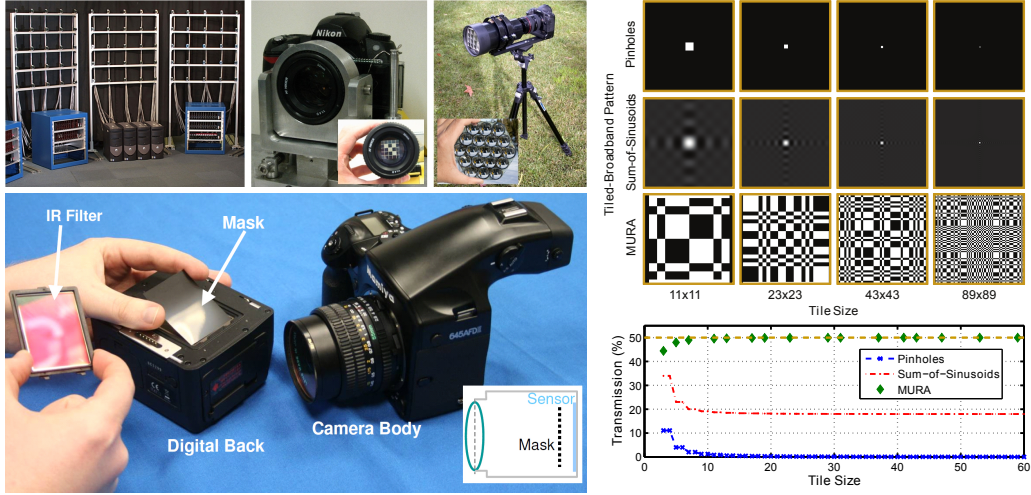


Figure 2.3: Left: light field cameras can be categorized by how a 4D light field is encoded in a set of 2D images. Methods include using multiple sensors or a single sensor with temporal, spatial, or frequency-domain multiplexing. (Top, Left) Wilburn et al. [2002] describe a camera array, Liang et al. [2008] achieve temporal multiplexing with a programmable aperture, Georgiev et al. [2008] capture spatially-multiplexed light fields using an array of lenses and prisms. (Bottom, Left) Raskar et al. [2008] capture frequency-multiplexed light fields by placing a heterodyne mask [Lanman et al., 2008; Veeraraghavan et al., 2008, 2007] close to the sensor. Right: Lanman et al. [2008] introduce tiled-broadband patterns for mask-based, frequency-multiplexed light field capture. (Top, Right) Each row, from left to right, shows broadband tiles of increasing spatial dimensions, including: pinholes [Ives, 1928], Sum-of-Sinusoids (SoS) [Veeraraghavan et al., 2007], and MURA [Gottesman and Fenimore, 1989; Lanman et al., 2008]. (Bottom, Right) The SoS tile converges to 18% transmission, whereas the MURA tile remains near 50%. Note that frequency multiplexing with either SoS or MURA tiles significantly outperforms conventional pinhole arrays in terms of total light transmission and exposure time. (Figures reproduced from [Wilburn et al., 2002], [Liang et al., 2008], [Georgiev et al., 2008], [Raskar et al., 2008], and [Lanman, 2010].)

field photography considers capturing the light field passing through the main lens aperture of a conventional camera. Adelson and Wang [1992], Okano et al. [1999], and Ng et al. [2005] extend integral photography to multiplex a 4D light field onto a 2D image sensor, as discussed in the following subsection. However, temporal multiplexing can also achieve this goal.

Liang et al. [2008] propose programmable aperture photography to achieve time-multiplexed light field capture. While Ives [1903] uses static parallax barriers placed close to the image sensor, Liang et al. use dynamic aperture masks (see Fig. 2.3, top left, center). For example, consider capturing a sequence of conventional photographs. Between each exposure a pinhole aperture is translated in raster scan order. Each photograph records a

pencil of rays passing through a pinhole located at a fixed position in the aperture plane for a range of sensor pixels. Similar to multiple sensor acquisition schemes, each image is a 2D slice of the 4D light field and the sequence can be concatenated to estimate the radiance for an arbitrary light ray passing through the aperture plane. To reduce the necessary exposure time, Liang et al. further apply Hadamard aperture patterns, originally proposed by Schechner et al. [2007], that are 50% transparent.

The preceding methods all consider conventional cameras with refractive lens elements. Zhang and Chen [2005] propose a lensless light field camera. In their design, a bare sensor is mechanically translated perpendicular to the scene. The values measured by each sensor pixel are recorded for each translation. By the Fourier projection-slice theorem [Ng, 2005], the 2D Fourier transform of a given image is equivalent to a 2D slice of the 4D Fourier transform of the light field; the angle of this slice is dependent on the sensor translation. Thus, tomographic reconstruction yields an estimate of the light field using a bare sensor, mechanical translation, and computational reconstruction methods.

Spatial and Frequency Multiplexing

Time-sequential acquisition reduces the cost and complexity of multiple sensor systems, however, it has one significant limitation: dynamic scenes cannot be readily captured. Thus, either a high-speed camera is necessary or alternative means of multiplexing the 4D light field into a 2D image are required. Ives [1903] and Lippmann [1908] provide two early examples of spatial multiplexing with the introduction of parallax barriers and integral photography, respectively. Such spatial multiplexing allows light field capture of dynamic scenes, but requires a trade-off between the spatial and angular sampling rates. Okano et al. [1999] and Ng et al. [2005] describe modern, digital implementations of integral photography, however numerous other spatial multiplexing schemes have emerged.

Instead of affixing an array of microlenses directly to an image sensor, Georgiev et al. [2006] add an external lens attachment with an array of lenses and prisms (see Fig. 2.3, center right). Ueda et al. [2008a,b] consider similar external lens arrays; however, in these works, an array of variable focus lenses, implemented using liquid lenses controlled by electrowetting, allow the spatial and angular resolution to be optimized depending on the observed scene.

Rather than using absorbing masks or refractive lens arrays, Unger et al. [2003], Levoy et al. [2004], Lanman et al. [2006], and Taguchi et al. [2010b] demonstrate that a single photograph of an array of tilted, planar mirrors or mirrored spheres produces a spatially-multiplexed estimate of the incident light field. Yang et al. [2000] demonstrate a large-format, lenslet-

based architecture by combining an array of lenses and a flatbed scanner. Related compound imaging systems, producing a spatially-multiplexed light field using arrays of lenses and a single sensor, were proposed by Ogata et al. [1994], Tanida et al. [2001, 2003], and Hiura et al. [2009].

Spatial multiplexing produces an interlaced array of elemental images within the image formed on the sensor. Veeraraghavan et al. [2007] introduce frequency multiplexing as an alternative method for achieving single-sensor light field capture. The optical heterodyning method proposed by Veeraraghavan et al. encodes the 4D Fourier transform of the light field into different spatio-angular bands of the Fourier transform of the 2D sensor image. Similar in concept to spatial multiplexing, the sensor spectrum contains a uniform array of 2D spectral slices of the 4D light field spectrum. Such frequency-domain multiplexing is achieved by placing non-refractive, light-attenuating masks slightly in front of a conventional sensor (see Fig. 2.3, bottom left).

As described by Veeraraghavan et al., masks allowing frequency-domain multiplexing (i.e., heterodyne detection) must have a Fourier transform consisting of an array of impulses (i.e., a 2D Dirac comb). In [Veeraraghavan et al., 2007], a *Sum-of-Sinusoids (SoS) pattern*, consisting of a weighted harmonic series of equal-phase sinusoids, is proposed. As shown in Figure 2.3 (right), such codes transmit significantly more light than traditional pinhole arrays [Ives, 1903]; however, these patterns are equivalent to a truncated Fourier series approximation of a pinhole array for high angular sampling rates. Lanman et al. [2008] propose *tilted-broadband patterns*, corresponding to periodic masks with individual tiles exhibiting a broadband Fourier transform. This family includes pinhole arrays, SoS patterns, and the tiled-MURA patterns proposed in that work (see Fig. 2.3, right). Such patterns produce masks with 50% transmission, enabling shorter exposures than existing methods.

In subsequent work, Veeraraghavan et al. [2008] propose adaptive mask patterns, consisting of aharmonic sinusoids, optimized for the spectral bandwidth of natural scenes. Georgiev et al. [2008] analyze such heterodyne cameras and further propose masks placed external to the camera body.

Chapter 9 of this thesis introduces a unifying mathematical framework that models the multiplexed acquisition of light fields and all other plenoptic dimensions. This framework allows us to analyze light field capture and reconstruction in either the spatial or the frequency domain. We show, for the first time, how Fourier-multiplexed light fields and other plenoptic manifolds can be reconstructed in the spatial domain, thereby significantly improving the resulting image quality. We also establish a metric for the quantitative evaluation of light field attenuation patterns in Chapter 9. This metric along with our framework allows us to

compare a variety of attenuation patterns, such as pinholes, sum-of-sinusoids, and MURA masks, and analyze their performance with respect to sensor noise amplification.

Light fields are not only a fundamental tool for computational photography but also for displays. Glasses-free 3D displays, as discussed in Chapters 3 and 4, require a light field as the desired input; reconstruction quality and display performance are evaluated in the spatial and frequency domain of light field space. Finally, the high-dimensional probes introduced in Chapters 6 and 7 are also designed, analyzed, and evaluated in 4D light field space.

Applications

Given the wide variety of light field capture devices, a similarly diverse set of applications is enabled by such high-dimensional representations of light transport. Light fields have proven useful for large variety of applications in computer graphics, digital photography, and 3D reconstruction. Specifically, these include image-based rendering [Gortler et al., 1996; Levoy and Hanrahan, 1996], image-based lighting [Debevec et al., 2000], 3D television [Carranza et al., 2003; Matusik et al., 2000; Matusik and Pfister, 2004; Starck and Hilton, 2008] and 3D display [Ives, 1928; Kanolt, 1918a; Lanman et al., 2010; Okano et al., 1999; Zwicker et al., 2006], shape reconstruction [Vlasic et al., 2009], gesture-based interaction [Hirsch et al., 2009], digital image refocusing [Ng, 2005], photographic glare reduction [Raskar et al., 2008; Talvala et al., 2007], extended depth of field microscopy [Levoy et al., 2004, 2006], and image stabilization [Smith et al., 2009].

2.2.4 Multiplexing Space and Focal Surfaces

The ability to resolve spatial light variation is an integral part of any imaging system. For the purpose of this chapter we differentiate between spatial variation on a plane perpendicular to the optical axis and variation along the optical axis inside a camera behind the main lens. The former quantity, transverse variation, is what all 2D sensors measure. Light variation along the optical axis can be described as the depth of field of an imaging system.

Although exotic camera systems can resolve structures in the order of 100 nm [van Putten et al., 2011], the resolution of standard photographs is usually limited by the physical layout and size of the photosensitive elements, the optical resolution of employed optical elements, and the diffraction limit. Attempts to break these limits, are referred to as super-resolution imaging [Agrawal and Raskar, 2007; Ashok and Neifeld, 2007; Baker and Kanade, 2002; Ben-Ezra et al., 2005; Borman and Stevenson, 1998; Landolt et al., 2001; Liu and Sun,

2011; Mohan et al., 2008; Shahar et al., 2011; Shechtman et al., 2005]. Gigapixel imaging is another field that, similar to super-resolution, aims at capturing very high-resolution imagery. The main difference is that gigapixel imaging approaches generally do not try to beat the limits of sensor resolution, but rather stitch a gigapixel panoramic image together from a set of megapixel images [Ben-Ezra, 2011; Cossairt et al., 2011; Kopf et al., 2007].

Depth of field (DOF), that is a depth-dependent (de)focus of a pictured scene, plays an important role in photography. While it is sometimes used as a photographic effect, for instance in portraits, ideally a photographer should be able to refocus or completely remove all defocus as a post processing step. Removing DOF blur from images is difficult, because the point spread function (PSF) is depth-dependent, resulting in a spatially-varying blur kernel. Furthermore, the PSF shape corresponds to that of the camera aperture, which is usually circular; due to the low Fourier magnitudes of these kinds of PSFs, high spatial frequencies are irreversibly filtered out in the image capture. The removal of DOF blur is therefore a deconvolution with an unknown, spatially-varying kernel that is not invertible. Applying natural image priors can improve reconstructions (see e.g., Levin et al. [2007a]), but does not change the ill-posedness of the problem.

In order to overcome the difficulties of defocus deblurring, the computational photography community has come up with two different approaches: optically changing the PSF to be depth-independent, resulting in a more tractable shift-invariant deconvolution, and modifying the PSF to be invertible. Point spread functions can be modified with *Focal Sweeps*, that is moving the object [Häusler, 1972] or sensor [Nagahara et al., 2008] during the exposure time, or by exploiting the wavelength-dependency of the PSF [Cossairt and Nayar, 2010]. Alternatively, the *apertures* of the imaging system can be *coded* with cubic phase plates [Dowski and Cathey, 1995] or other phase masks [Ben-Eliezer et al., 2005; Chi and George, 2001; Ojeda-Castaneda et al., 2005], diffusers [Cossairt et al., 2010; Garcia-Guerrero et al., 2007], attenuation patterns [Levin et al., 2007b; Veeraraghavan et al., 2007], polarization filters [Chi et al., 2006], or multi-focal elements [Levin et al., 2009]. All of these approaches optically modify the PSF of the optical system for an extended DOF. The captured images usually need to be post-processed, for instance by applying a shift-invariant deconvolution. An analysis of quality criteria of attenuation-based aperture masks for defocus deblurring was presented by Zhou and Nayar [2009]; this analysis was extended to also consider PSF invertibility [Baek, 2010].

Focal Stacks are image sequences, where the focal plane differs for each photograph in the stack. A single, focused image can be composited by selecting the best-focused match in the stack for each image region [Pieper and Korpel, 1983]. The optimal choice of parameters, including focus and aperture, for the images in a focal stack are well established [Hasinoff

and Kutulakos, 2008; Hasinoff et al., 2009]. Capturing a focal stack with a large-scale high-resolution camera was implemented by Ben-Ezra [2010]. Kutulakos and Hasinoff [2009] propose to multiplex a focal stack into a single sensor image in a similar fashion as colour filter arrays multiplex different colour channels into a RAW camera image. However, to the knowledge of the author, this camera has not yet been built.

Green et al. [2007] split the aperture of a camera using circular mirrors and multiplex the result into different regions of a single photograph. In principle, this approach captures multiple frames with varying aperture settings at a reduced spatial resolution in a single snapshot. Other applications for flexible focus imaging include 3D shape reconstruction with shape from (de)focus (e.g., Nayar and Nakagawa [1994]; Zhou et al. [2009]), *Confocal Stereo* [Hasinoff and Kutulakos, 2006, 2009], and video matting [McGuire et al., 2005].

In Chapter 5, we introduce dynamically coded apertures to extend the capabilities of projection displays. Inspired by coded apertures in computational photography [Levin et al., 2007b; Veeraraghavan et al., 2007; Zhou et al., 2009; Zhou and Nayar, 2009], we show how the depth of field of display devices can be significantly extended while preserving a high light transmission. Unlike in photography, the displayed images are known a priori, so corresponding aperture patterns can be optimized for a human observer in a content-dependent fashion.

2.2.5 Multiplexing Time

Capturing motion and other forms of movement in photographs has been pursued since the invention of the daguerreotype. Early pioneers in this field include Eadweard Muybridge (e.g., Muybridge [1957]) and Etienne-Jules Marey (e.g., Braun [1992]), who mostly built custom photographic apparatuses. More professional analog high-speed film cameras have been developed throughout the last century [Ray, 2002]. Today, high-speed digital cameras are commercially available. Examples are the Phantom Flex [Research, 2010], the FASTCAM SA5 [Photron, 2010], and the HyperVision HPV-2 [Shimadzu, 2010]. With the introduction of Casio’s Exilim camera series (exilim.casio.com), which records low resolution videos at up to 1,000 fps, high-speed cameras have entered the consumer market.

An alternative to high-speed sensors is provided by *Assorted Pixels* [Narasimhan and Nayar, 2005], where spatial resolution is traded for temporal resolution by measuring spatially interleaved, temporally staggered exposures on a sensor. This approach is very similar to what standard colour filter arrays do to acquire colour information (see Sec. 2.2.2). While this concept was initially only theoretical, it has recently been implemented by aligning

a digital micromirror device (DMD) with a CCD sensor [Bub et al., 2010]. Alternatively, the sensor readout could be controlled on a per-pixel basis, as for instance provided by non-destructive sensor readout (e.g., Semiconductor [2010]) or the Pixim camera [Pixim, 2010]. Reddy et al. [2011] build a liquid crystal on silicone (LCOS) based camera prototype that modulates the exposure of each pixel randomly throughout the exposure time. In combination with a nonlinear sparse reconstruction algorithm, the 25 Hz prototype has been shown to capture imagery with up to 200 frames per second without loss of spatial resolution by exploiting sparsity in the spatio-temporal volume. Coded rolling shutters [Gu et al., 2010] have the potential to implement this concept on a per-scanline basis. Chapter 9 generalizes many of these techniques in a common framework for *plenoptic multiplexing*.

Agrawal et al. [2010b] demonstrate how a pinhole in the aperture plane of a camera, which moves throughout the exposure time, allows the captured data to be adaptively re-interpreted. For this purpose, temporal light variation is directly encoded in the different views of the light field that is simultaneously acquired with a Sum-of-Sinusoids (SoS) attenuation-mask in a single shot (see Sec. 2.2.3). While this approach has exclusively been analyzed and reconstructed in the Fourier domain [Agrawal et al., 2010b], we show in Chapter 9 how a spatial reconstruction can be performed and results in higher quality data.

Rather than photographing a scene with a single high-speed camera, *multiple synchronized devices* can be used. The direct capture of high-speed events with camera arrays is discussed by Wilburn et al. [2004, 2005]. In this approach, the exposure windows of the cameras are slightly staggered so that a high-speed video can be composed by merging the data of the individual cameras. Shechtman et al. [2002, 2005] propose to combine the output of multiple low-resolution video cameras for space-time super-resolution. Coded exposures have been shown to optimize temporal super-resolution from multi-camera arrays [Agrawal et al., 2010a] by alleviating the ill-posedness of the reconstruction.

High-speed imagery can also be acquired by utilizing *high-speed illumination*. Harold ‘Doc’ Edgerton (e.g., [Project, 2009]) created this field by inventing electronic strobes in 1931. Stroboscopic illumination can be used to compensate for rolling shutter effects and synchronize an array of consumer cameras [Bradley et al., 2009]. Narasimhan et al. [2008] exploit the high-speed temporal dithering patterns of DLP-based projector illumination for a variety of vision problems, including photometric stereo and range imaging. Coded strobing, by either illumination or controlled sensor readout, in combination with reconstructions developed in the compressive sensing community, allows high-speed periodic events to be acquired [Veeraraghavan et al., 2011]. Another high-speed imaging approach that is inspired by compressive sensing is proposed by Gupta et al. [2010]. Here, a 3D spatio-temporal vol-

ume is adaptively encoded with a fixed voxel budget. This approach encodes fast motions with a high temporal, but lower spatial resolution, while the spatial resolution in static parts of the scene is maximized.

Motion Deblurring

Motion deblurring has been an active area of research over the last few decades. It is well known that deblurring is an ill-posed problem, which is why many algorithms apply regularizers [Lucy, 1974; Richardson, 1972] or natural image statistics (e.g., Fergus et al. [2006]; Levin et al. [2007a]) to solve the problem robustly.

Computational photography approaches to the problem of motion deblurring have been proposed. These include *coded single capture* approaches [Agrawal and Raskar, 2007; Agrawal and Xu, 2009; Raskar et al., 2006], *motion-invariant point spread functions* [Agrawal and Xu, 2009; Levin et al., 2008], techniques using *image sequences* [Agrawal et al., 2009; Bascle et al., 1996; Telleen et al., 2007], and *hybrid cameras* using multiple devices [Ben-Ezra and Nayar, 2004, 2003; Li et al., 2008a; Tai et al., 2010].

2.2.6 Phase and Fluid Imaging

Fluid imaging is a wide and active area of research. Generally, approaches to measure fluid flows can be categorized into optical and non-optical methods. Non-optical methods make velocity fields observable by inducing particles, dye, or smoke; alternatively, the surface shear, pressure forces, or thermal reactions can be measured on contact surfaces that are coated with special chemicals. Optical fluid imaging methods include Schlieren photography and holographic interferometry, which are based on ray and wave models of light, respectively. Light Field Background Oriented Schlieren imaging (LFBOS), as introduced in Chapters 6 and 7, is an optical fluid imaging approach that uses a ray-based light model. An extensive overview of fluid imaging techniques can be found in the book by Merzkirch [1987].

Traditional Schlieren Photography is a non-intrusive imaging method for dynamically changing refractive index fields. These techniques have been developed in the fluid imaging community over the past century, with substantial improvements in the 1940s by Schardin and his colleagues [Schardin, 1942]. An overview of different optical setups and the historic evolution of Schlieren and Shadowgraph imaging can be found in the book Settles [2001]. Unfortunately, the optical setups require precisely registered high-quality mirrors and lenses, which are expensive, usually bulky, and difficult to calibrate. LFBOS uses an inexpensive

light field probe that encodes refractions in variation of colour and intensity; our current prototype implements this concept with a lenslet array and a transparency.

In the last decade, *Background Oriented Schlieren Imaging* (BOS) [Dalziel et al., 2000; Elsinga et al., 2004; Hargather and Settles, 2009; Meier, 2002; Richard and Raffel, 2001] has been developed. Here, complicated optical apparatuses are replaced by optical flow calculations; an evaluation of these algorithms for BOS can be found in the work by Atcheson et al. [2009]. Because of the simplified setup, BOS makes it feasible to set up multi-view Schlieren imaging systems that can be used for tomographic reconstruction of 3D distortion volumes [Atcheson et al., 2008]. Just like BOS, the approach discussed in Chapter 6 uses a background probe, which is observed through the refractive medium. However, rather than estimating the distortion of a diffuse background with computationally expensive optical flow estimators, we optically encode spatial and angular light variation with light field probes. The computational part of our technique is a pre-processing step that determines the colours and layout of the probe.

A variety of techniques has been proposed to visualize and quantify phase retardation in transparent microscopic organisms [Murphy, 2001]. Many of these *phase-contrast imaging* approaches, such as Zernike phase contrast and differential interference contrast (DIC), require coherent illumination and are qualitative rather than quantitative. This implies that changes in phase or refractive events are encoded as intensity variations in captured images, but remain indistinguishable from the intensity variations caused by absorption in the medium. Quantitative approaches exist [Barone-Nugent et al., 2002], but require multiple images, are subject to a paraxial approximation, and are limited to orthographic cameras. The approach presented in Chapter 6 optically codes the refractions caused by solids or liquids in macroscopic environments in a single photograph; the method does not require coherent illumination.

Applications in Transparent, Refractive Object Reconstruction

Transparent object reconstruction has recently gained a lot of traction [Ihrke et al., 2010a]. Kutulakos and Steger [2005] analyze the space of these reconstructions based on acquisition setup and number of refractive events in the optical path of light rays. Generally, refractive object capture and reconstruction can be performed using a single camera but multiple images or, alternatively, using multiple cameras. Ben-Ezra and Nayar [2003] reconstruct smooth, parameterized refractive objects from the distortions of a diffuse background in an image sequence from a single view. Agarwal et al. [2004] extend optical flow with diffuse backgrounds to refractive environments. Miyazaki and Ikeuchi [2005] and Huynh et al. [2010]

exploit the polarization of refracted light to estimate transparent surfaces. A tomographic reconstruction of transparent solids from multiple images was proposed by Trifonov et al. [2006]. Ihrke et al. [2005] compute the shape of flowing water by dyeing it with fluorescent chemicals. Range scanning can be used for the acquisition of refractive solids, if they are immersed in a fluorescent liquid [Hullin et al., 2008]. Morris and Kutulakos [2007] show that the surface of complex refractive objects can be reconstructed from multiple photographs with changing illumination. Furthermore, specular objects can be acquired using shape from distortion [Bonfort et al., 2006; Tarini et al., 2005]. Multiple cameras have been used for dynamic refractive stereo [Morris and Kutulakos, 2005] and for the reconstruction of smooth gas flows [Atcheson et al., 2008]. In Chapter 7, we propose a new approach to reconstructing the shape of a class of non-stationary, refractive, transparent solids or liquids from a single image.

Single image reconstruction techniques include the seminal work by Murase [1990], where a wavy water surface is reconstructed by observing the distortions of a diffuse probe under water with an orthographic camera. Zhang and Cox [1994] also reconstruct a water surface with an orthographic camera by placing a big lens and a 2D screen at its focal length in the water. This allows the surface gradients to be measured, which can subsequently be integrated to compute the surface shape. For both approaches the mean water level needs to be known. Savarese and Perona [2002] present an analysis of single image reconstruction of smooth mirroring objects using shape from distortion. Compared to these techniques, the approach discussed in Chapter 7 also assumes that there is only a single refractive or reflective event; however, no constraints are placed on the camera setup. Furthermore, we show how to reconstruct both surface points and normals simultaneously from a single image.

2.2.7 Acquiring Polarization

Polarization is an inherent property of the wave nature of light [Collett, 2005], therefore not a dimension of the plenoptic function. Generally, polarization describes the oscillation of a wave traveling through space in the transverse plane, perpendicular to the direction of propagation. Linear polarization refers to transverse oscillation along a line, whereas spherical or elliptical polarization describe corresponding oscillation trajectories.

Although some animals, including mantis shrimp [Marshall and Oberwinkler, 1999], cephalopods (squid, octopus, cuttlefish) [Mäthger et al., 2009], and insects [Wehner, 1976], are reported to have photoreceptors that are sensitive to polarization, standard solid state sensors are not. The most straightforward way of capturing this information is by taking multiple

photographs of a scene with different polarizing filters mounted in front of the camera lens. These filters are standard practice in photography to reduce specular reflections, increase the contrast of outdoor images, and improve the appearance of vegetation. Alternatively, this kind of information can be captured using polarization filter arrays [Schechner and Nayar, 2003b] which, similar to generalized mosaics [Schechner and Nayar, 2005], require multiple photographs to be captured. Recently, polarized illumination [Ghosh et al., 2010] has been shown to have the potential to acquire all Stokes parameters necessary to describe polarization.

Computer graphics and vision applications for the acquisition of polarized light include image dehazing [Namer and Schechner, 2005; Schechner et al., 2001, 2003], improved underwater vision [Schechner and Karpel, 2004, 2005], specular highlight removal [Müller, 1996; Nayar et al., 1993; Umeyama and Godin, 2004; Wolff and Boulton, 1991], shape [Atkinson and Hancock, 2005; Miyazaki et al., 2004, 2003] and BRDF [Atkinson and Hancock, 2008] estimation, light source separation [Cula et al., 2007], surface normal acquisition [Ma et al., 2007], surface normal and refractive index estimation [Ghosh et al., 2010; Sadjadi, 2007], and the separation of transparent layers [Schechner et al., 1999].

2.3 Plenoptic Image Display

Similar to the dimension-based classification used in the last section, we outline approaches to plenoptic image display in this section.

2.3.1 High Dynamic Range Displays

The concept of high dynamic range image display was introduced by Seetzen et al. [2004]. For this purpose, two prototypes are proposed. Both are based on the concept of dual modulation, where the output of either a projector or a low-resolution light emitting diode (LED) array is modulated with a high-resolution liquid crystal display (LCD). The secondary modulation effectively reduces the blacklevel of the display, thereby increasing the contrast. Image processing [Trentacoste et al., 2007] is applied to the desired HDR image in order to decompose it into two components, one for each of the displays in the setup.

While Seetzen et al. [2004] propose to slightly defocus the projector on the secondary LCD, Pavlovych and Stuerzlinger [2005] use a similar setup with a focused projection. The advantages of the defocused version are reduced moiré and easier display registration. While high spatial frequencies are more difficult to be displayed with the highest possible contrast in a

defocused configuration, the human visual system cannot actually resolve these frequencies at a very high contrast due to masking effects [Seetzen et al., 2004].

Damberg et al. [2007] and Kusakabe et al. [2008] propose to add a secondary spatial light modulator (SLM), either an LCD or a digital micromirror device (DMD), into the optical path of a projector, inside the actual device. A practical solution to this problem is a setup composed of three low-resolution colour modulators inside the projector that are combined with a single device, which modulates the combined luminance at a high resolution [Kusakabe et al., 2008]. Hoskinson et al. [2010] proposed to reallocate the light within a projector in a content-dependent manner. For this purpose, an array of analog micromirrors is used to redirect the uniform backlight in the projector from darker parts of a displayed image into brighter parts before modulation by the actual SLM occurs. This concept allows the brightness in some parts of the projected image to be increased, whereas dual modulation approaches can only dim the backlight. Bimber and Iwai [2008] applied projector-based illumination to static prints, transparencies, and electronic paper to achieve an enhanced dynamic range of the content presented on these devices.

Due to its slim form factor and relatively low power consumption, the combined LED and LCD display design introduced by Seetzen et al. [2004] has been further developed and is commercially available today. Due to hardware constraints, a non-negligible gap between the LED and the LCD display is unavoidable. Previous approaches [Seetzen et al., 2004; Trentacoste et al., 2007] applied a heuristic decomposition, to the desired HDR image, taking this gap into account, when computing the contributions for the two different display layers. In Chapter 3, we confirm that these heuristics achieve near-optimal quality and we propose the first non-heuristic construction of HDR displays with two or more disjoint attenuation layers, such as stacked transparencies or LCDs.

2.3.2 Multi-Spectral Image Synthesis

Due to the tristimulus nature of the human visual system, almost all displays synthesize three colour channels that resemble the spectral sensitivities of the sensory cells responsible for colour perception. Multi-spectral image display, however, has the potential to increase the colour gamut of a device. This can, for instance, be achieved with multiple overlapping images that are each modulated with different optical filters [Michika and Brown, 1989]. Alternatively, the uniform backlight in an LCD panel can be replaced with coloured LEDs [Dolby, 2010], which increases the colour gamut of the device and also its dynamic range. Mohan et al. [2008] presented the idea of agile spectrum imaging, where the colour spectrum of projected imagery can be modulated with a grayscale attenuation mask.

While display colour gamuts beyond the resolving powers of the human visual system are rarely pursued in practice, these could achieve new degrees of freedom for the design of computational probes in computer vision applications (see Chapters 6 and 7).

2.3.3 Light Field and 3D Displays

This section reviews technologies for 3D image and light field display. A more detailed discussion and taxonomy of these approaches can be found in the courses by Lanman and Hirsch [2010] and Halle [2005].

Glasses-Bound Stereoscopic Displays

Glasses-bound displays are characterized by the user being required to wear additional glasses or other head-mounted accessories. *Head-Mounted Stereoscopic Displays* are often used in augmented reality applications. They can be further categorized as video see-through and optical see-through displays. A discussion of both can be found in Section 2.4. *Multiplexed Displays*, on the other hand, present a pair of stereo images on the same display surface to an observer. The two images are then demultiplexed by glasses worn by the user. Technologies to multiplex and demultiplex such imagery include passive colour filters or polarizers and active shutter glasses.

Automultiscopic Displays

Automultiscopic or autostereoscopic displays present three-dimensional imagery to a viewer without the need for special glasses. These types can be categorized as parallax-based, volumetric, and holographic displays.

The most popular approaches to glasses-free 3D display are parallax-based systems: *Parallax Barriers* [Ives, 1903; Kanolt, 1918b] and *Integral Imaging* [Lippmann, 1908]. Parallax-barrier approaches place light blocking elements, such as slits or pinholes, at a slight distance in front of a standard 2D screen. While reducing light efficiency and spatial resolution, this is a simple, yet effective approach to optically create distinct viewing zones. More recently, Isono et al. [1993] introduced dual-stack LCDs to achieve programmable parallax barriers. Kim et al. [2007] follow a similar principle, but enhance the spatial resolution of dual-stacked LCDs using time-shifted pinholes. Lanman et al. [2010] introduce content-adaptive parallax barriers, optimizing dual-layer displays with temporally-varying attenuation found with non-negative matrix factorization. This approach generalizes parallax-barrier systems

and improves light efficiency compared to prior approaches. One of the main advantages of LCD-based parallax barrier displays is the ability to dynamically switch between a high-resolution 2D mode and a lower-resolution 3D mode by switching one of the displays on or off [Jacobs et al., 2003].

Integral imaging methods use an array of lenslets on a 2D screen to synthesize parallax [Lippmann, 1908], thereby maximizing the light transmission of the display. While this approach makes it more difficult to switch between 2D and 3D display modes, recent lenticular designs have been shown to achieve this effect [Woodgate and Harrold, 2003]. The loss in spatial resolution is equivalent to the size of the employed lenslets and comparable to that of parallax barrier systems. Zwicker et al. [2006, 2007] analyze these resolution tradeoffs and discuss anti-aliasing for both parallax barrier and integral imaging architectures.

Blundell and Schwartz [1999] define a *Volumetric Display* as permitting “the generation, absorption, or scattering of visible radiation from a set of localized and specified regions within a physical volume”. Many volumetric displays exploit high-speed projection synchronized with mechanically-rotated screens. Such swept volume displays were proposed as early as 1912 [Favalora, 2005] and have been continuously improved [Cossairt et al., 2007]. While requiring similar mechanical motion, Jones et al. [2007] instead achieve light field display, preserving accurate perspective and occlusion cues, by introducing an anisotropic diffusing screen and user tracking. Related designs include the Seelinder by Yendo et al. [2005], exploiting a spinning cylindrical parallax barrier and LED arrays, and the work of Maeda et al. [2003], utilizing a spinning LCD panel with a directional privacy filter. Several designs have eliminated moving parts using electronic diffusers [Sullivan, 2003], projector arrays [Agocs et al., 2006], and beam-splitters [Akeley et al., 2004]. Others consider projection onto transparent substrates, including water drops [Barnum et al., 2010], passive optical scatterers [Nayar and Anand, 2007], and dust particles [Perlin and Han, 2006].

Multi-Layer Automultiscopic Displays with three or more attenuating layers were first considered for 3D display by Loukianitsa and Putilin [2002] and Putilin and Loukianitsa [2006]. The employed optimization uses neural networks. In closely-related works, Gotoda [2010, 2011] proposes optimizing the properties of layered LCDs with tomographic methods. As described by Bell et al. [2008], multi-layer LCDs exhibit decreased brightness, moiré, and colour crosstalk, with additional layers exacerbating problems. Similar limitations are expected with other spatial light modulators

Holographic Displays (e.g., Saxby [1994]) store wavefront information in microscopic scales on a holographic emulsion that is somewhat similar to photographic film. Holograms, when illuminated, can synthesize the recorded wavefront in great detail with directional variation

and high colour accuracy. Unfortunately, holographic recording technology can so far only produce static images with a high visual quality. Although there has been a lot of research on digitally synthesizing holograms [Slinger et al., 2005], the quality of these displays is far from that of recorded ones. Holograms can be augmented with computer generated content, as discussed by Bimber [2006] and Bimber et al. [2005c]. This combination allows the visual quality of holography and interactivity provided by real-time computer graphics to be combined.

In Chapters 3 and 4, we present multi-layered solution for glasses-free 3D display using attenuating and polarization-rotating layers, respectively. We demonstrate novel prototype designs and show that these can be controlled at interactive framerates. We establish theoretical limitations of all attenuation-based multi-layer displays and show how this design can, alternatively, be used for high contrast 2D image display.

2.3.4 Extended Depth of Field Projection

When projecting onto textured and geometrically complex surfaces, image distortion, colour modulation, and projector defocus are challenging problems. Recently developed radiometric compensation techniques allow seamless image projections onto everyday surfaces by pre-distorting the presented content and also pre-correcting the displayed colours [Bimber et al., 2005a,b; Grossberg et al., 2004; Wetzstein and Bimber, 2007].

Projections onto geometrically complex surfaces with a high depth variance generally do not allow the displayed content to be in focus everywhere. Common DLP or LCD projectors maximize their brightness with large apertures. Thus, they suffer from narrow depths of field and can only generate focused imagery on a single fronto-parallel screen. Laser projectors, which are commonly used in planetaria, are an exception. These emit almost parallel light beams, which make very large depths of field possible. However, the cost of a single professional laser projector can exceed the cost of several hundred conventional projectors. In order to increase the depth of field of conventional projectors, several approaches for deblurring unfocused projections with a single or with multiple projectors have been proposed.

Zhang and Nayar [2006] present an iterative, spatially-varying filtering algorithm that compensates for projector defocus. They employ a coaxial projector-camera system to measure the projections spatially-varying defocus. For this purpose, dot patterns are projected onto the screen and captured by the camera. The defocus kernels for each projector pixel can be recovered from the captured images. Given the environment light a desired input image,

a compensation image can be computed by minimizing the sum-of-squared pixel difference between the desired image and the expected projection. An alternative approach to defocus compensation for a single projector setup was presented by Brown et al. [2006]. Projector defocus is modeled as a convolution of a projected original image and Gaussian point spread functions (PSFs). The PSFs are estimated by projecting features on the canvas and capturing them with a camera. Assuming a spatially-invariant PSF, a compensation image can be synthesized by applying a Wiener deconvolution. Oyamada and Saito [2007] present a similar approach to single projector defocus compensation. Here, circular PSFs are used for the convolution and estimated by comparing the original image to various captured compensation images that were generated with different PSFs. The main drawback of these single projector defocus compensation approaches is that the image quality is highly dependent on the projected content. All of the discussed methods result in a pre-sharpened compensation image that is visually closer to the original image after being optically blurred by the defocused projection. While soft contours can be compensated, this is generally not the case for sharp features.

An alternative approach that is less dependent on the actual frequencies in the input image is discussed by Bimber and Emmerling [2006]. Multiple overlapping projectors with varying focal depths illuminate arbitrary surfaces with complex geometry and reflectance properties. A survey including all of the above listed approaches and other projector-camera systems is [Bimber et al., 2008].

In Chapter 5, we introduce coded apertures to projection displays. We demonstrate how static and dynamically coded projector apertures extend the depth of field of projection systems while maintaining a high light transmission. This work, along with the multi-layer light field displays discussed in Chapters 3 and 4, is another example of combined optical light modulation and computational processing for displays. The approach presented in Chapter 5 also considers the limitations of the visual system in order to optimize the presented content for an human observer.

2.3.5 High-Speed Displays

Although the human visual system requires about 30 frames per second (fps) to perceive smooth motion [Goldstein, 2006], the flicker fusion threshold is usually twice as high. For this reason, common displays are often designed to achieve at least 60 Hz. Recently, more and more LCD displays supporting up to 120 Hz have become available (e.g., www.viewsonic.com). Digital light processing (DLP) projectors are usually equipped with a spinning colour wheel and a high-speed digital micromirror device (DMD). DMDs can

achieve high framerates; the DLP LightCommander (www.ti.com/ww/en/analog/mems/-dplightcommander/), for instance, can project grayscale imagery with 500 Hz and binary content with 5000 Hz. The dithering patterns of DLP projectors, used to synthesize different intensity levels, can also be exploited for computer vision applications, such as 3D geometry reconstruction [Narasimhan et al., 2008].

Displays with more than 120 Hz, this framerate being required to present flicker-free stereoscopic content, are not of significant importance for image display to human observers. Just like multi-spectral displays, however, such devices have many applications in computer vision, for instance by extending the capabilities of the computational probes discussed in Chapters 6 and 7.

2.3.6 Polarization-Rotating LCDs

Liquid crystal displays are usually constructed by enclosing a layer of polarization-rotating liquid crystals with a pair of crossed linear polarizers [Yeh and Gu, 2009]. The light emitted by a uniform backlight is linearly polarized by the rear polarizer. The layer of programmable liquid crystal cells, each representing one pixel, has the ability to rotate the incident, linearly polarized light by a small, but controllable amount. The polarization angle between the rotated light and the front polarizer then determines the outgoing light intensity. Yeh and Gu [2009] formally characterize the polarization properties of LCDs, including twisted nematic (TN), vertical alignment (VA), and in-plane switching (IPS) panels.

Stripping an LC panel from its pair of crossed polarizers yields a pixel-precise, programmable polarization rotator. As discussed in more detail in Chapter 4, ideal polarization rotation, however, is only achieved for collimated light, passing through the LC panel at angles perpendicular to the display plane [Yeh and Gu, 2009]. In other cases, the outgoing light exhibits elliptical, rather than linear, polarization states. Davis et al. [2000] implement a two-dimensional polarization rotator using a custom parallel-aligned LCD covered by a pair of crossed quarter-wave plates. Moreno et al. [2007] construct a polarization rotator using a conventional TN panel. In both works, the liquid crystal is operated as a voltage-controlled wave plate to produce ideal polarization state rotations.

In Chapter 4, we introduce polarization field displays as an optically-efficient construction allowing dynamic light field display using multi-layered LCDs. Such displays are constructed by covering a stacked set of liquid crystal panels with a single pair of crossed linear polarizers.

2.4 Computational Optics for Direct View

This section outlines concepts for modulating light before it is perceived by a human observer.

2.4.1 Optical Image Processing

A vast literature exists on optical image processing methods using Fourier optics [Born and Wolf, 1999], including effects such as edge enhancement [Yelleswarapu et al., 2006] and image sharpening [Shih et al., 2001]. Fourier optics require coherent light and are thus ill-suited for direct view in natural environments.

2.4.2 Night Vision

Image amplification for low-light vision enhancement has been well studied, including wearable solutions for military personnel [Hradaynath, 2002]. Night-vision solutions using optical image amplification typically employ cascading opto-electrical effects. These approaches have a multiplicative effect on the incoming imagery, but they perform only uniform amplification rather than spatially selective filtering. In Chapter 8, we introduce an approach to direct manipulation of an observed scene. This is implemented using transmissive spatial light modulators that can attenuate light in a controlled, pixel-precise fashion.

2.4.3 Augmented Reality

Researchers in augmented reality (AR), also known as mediated reality, have been working on live manipulation of observed scenes for decades [Azuma et al., 2001; Feiner et al., 1993; Sutherland, 1968]. AR displays can be classified as *optical see-through* or *video see-through* [Rolland et al., 1994], with optical see-through AR displays being further categorized as either head-worn [Cakmakci and Rolland, 2006] or spatial [Bimber and Raskar, 2005].

Optical see-through AR displays usually split the light path so that an observer sees the real-world, but also a synthetic scene overlaid, which is shown on an additional display. Spatial light modulators (SLMs) in the optical paths of AR displays have so far been employed to achieve mutual occlusion [Cakmakci et al., 2004; Kiyokawa et al., 2001]. The prototypes discussed in Chapter 8 use designs resembling optical see-through displays long used in AR, but differs crucially in the way the display content is combined with the real

world. Traditional optical see-through AR uses combining optics, such as beam-splitters, to overlay synthetic content from a display, such as an organic LED (OLED) or LCD, onto real-world scenery in an additive fashion. In our approach, however, we use a partially transparent display as a spatial light modulator to filter, in real-time, the light arriving from the real-world environment in a multiplicative fashion. Rather than using the display to show artificial content, we program the SLM transparency in order to perform a variety of image processing operations to aid the user in understanding the real-world environment. Since the optical filtering acts immediately on the incident light, the processed real-world scene can be observed without latency in our approach.

Video see-through AR captures the incoming light with a camera, processes the resulting image to merge it with synthetic elements, and then displays the result. This design simplifies image processing and registration between synthetic and real imagery, and allows arbitrary manipulation of the real-world imagery, but at the cost of introducing the full system latency to the user's perception of the real world. Video see-through AR systems are also limited by the resolution and dynamic range of the display and camera. Optical see-through designs and the approach presented in Chapter 8 possess a crucial advantage over video-based AR: the real world is viewed directly with a quality only limited by the human eye, and without latency. As a result, such approaches avoid motion sickness, and with appropriate safeguards, could be used for safety-critical applications.

Recent advances in miniaturizing system components for head-worn displays make it possible to integrate these seamlessly into everyday clothing, a concept thoroughly explored by Mann et al. [2005] and Mann [1997]. Multipurpose contact lenses, as for instance introduced by Lingley and Parviz [2008], take the concept of integrating displays into clothing and accessories even further. The hardware components required by the technique described in Chapter 8 are very similar to the ones used in head-mounted AR applications.

Chapter 3

Tomographic Image Synthesis for Attenuation-based Multi-Layer Displays

In this chapter, we develop tomographic techniques for image synthesis on displays composed of compact volumes of light-attenuating material. Such volumetric attenuators recreate a 4D light field or high-contrast 2D image when illuminated by a uniform backlight. Since arbitrary oblique views may be inconsistent with any single attenuator, iterative tomographic reconstruction minimizes the difference between the emitted and target light fields, subject to physical constraints on attenuation. As multi-layer generalizations of conventional parallax barriers, such displays are shown, both by theory and experiment, to exceed the performance of existing dual-layer architectures. For 3D display, spatial resolution, depth of field, and brightness are increased, compared to parallax barriers (see Sec. 3.3). For a plane at a fixed depth, the optimization presented in this chapter also allows optimal construction of high dynamic range displays, confirming existing heuristics and providing the first extension to multiple, disjoint layers (see Sec. 3.4). This chapter concludes by demonstrating the benefits and limitations of attenuation-based light field displays using an inexpensive fabrication method: separating multiple printed transparencies with acrylic sheets. While this approach only allows for the display of static imagery, chapter 4 introduces a dynamic multi-layer display prototype along with real-time solutions to the tomographic reconstruction.

3.1 Introduction and Motivation

3D displays are designed to replicate as many perceptual depth cues as possible. As surveyed by Lipton [1982], these cues can be classified by those that require one eye (monocular) or both eyes (binocular). Artists have long exploited monocular cues, including perspective, shading, and occlusion, to obtain the illusion of depth with 2D media. Excluding motion parallax and accommodation, existing 2D displays provide the full set of monocular cues.

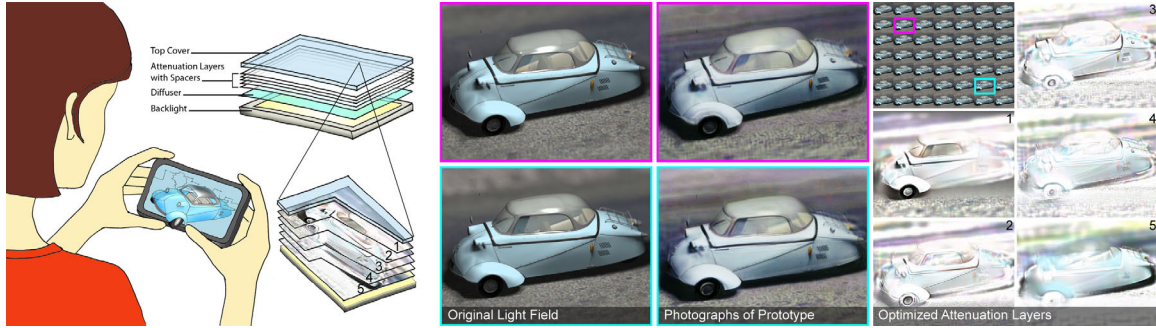


Figure 3.1: *Inexpensive, glasses-free light field display using volumetric attenuators. (Left) A stack of spatial light modulators (e.g., printed masks) recreates a target light field (here for a car) when illuminated by a backlight. (Right) The target light field is shown in the upper left, together with the optimal five-layer decomposition, obtained with iterative tomographic reconstruction. (Middle) Oblique projections for a viewer standing to the top left (magenta) and bottom right (cyan). Corresponding views of the target light field and five-layer prototype are shown on the left and right, respectively. Such attenuation-based 3D displays allow accurate, high-resolution depiction of motion parallax, occlusion, translucency, and specularity, being exhibited by the trunk, the fender, the window, and the roof of the car, respectively.*

As a result, 3D displays are designed to provide the lacking binocular cues of disparity and convergence, along with these missing monocular cues.

Most current 3D displays preserve disparity, but require special eyewear (e.g., LCD shutters, polarizers, or colour filters). In contrast, automultiscopic displays replicate disparity and motion parallax without encumbering the viewer. As categorized by Favalora [2005], such glasses-free displays include parallax barriers [Ives, 1903; Kanolt, 1918b] and integral imaging [Lippmann, 1908], volumetric displays [Blundell and Schwartz, 1999], and holograms [Slinger et al., 2005]. Holograms present all depth cues, but are expensive and primarily restricted to static scenes viewed under controlled illumination [Klug et al., 2001]. Research is addressing these issues [Blanche et al., 2010], yet parallax barriers and volumetric displays remain more practical, utilizing well-established, low-cost fabrication. Furthermore, volumetric displays can replicate similar depth cues with flicker-free refresh rates [Favalora, 2005].

This chapter considers automultiscopic displays comprised of compact volumes of light-attenuating material. Differing from volumetric displays with light-emitting layers, overlaid attenuation patterns allow objects to appear beyond the display enclosure and for the depiction of motion parallax, occlusion, and specularity.

The theoretical contributions of this chapter apply equally well to dynamic displays, such as

stacks of attenuating liquid crystal display (LCD) panels discussed in chapter 4. However, the prototype introduced in this chapter uses static printing to demonstrate the principles of tomographic image synthesis. Specifically, we produce multi-layer attenuators using 2D printed transparencies, separated by acrylic sheets (see Figures 3.1 and 3.2).



Figure 3.2: *Prototype multi-layer display. (Left) A multi-layer display is fabricated by separating transparencies with acrylic sheets and back-illuminating them with a light box (e.g., an LCD panel). (Right) Printed transparencies and acrylic layers.*

3.1.1 Overview of Benefits and Limitations

The relative benefits and limitations of the proposed approach are summarized in Table 3.1. Unlike many volumetric displays, attenuation-based multi-layer displays exploit multiplicative light absorption across multiple layers, rather than additive light emission. Such spatially-varying attenuation is inexpensively fabricated, without moving parts, using either *static* 2D or 3D printing, or by layering *dynamic* spatial light modulators (e.g., LCD, LCoS, or DMD). Modulation of light allows objects to appear beyond the display and for depiction of occlusion and specularity. Section 3.2 presents an optimal decomposition of light fields into two or more static layers, improving upon the method of Lanman et al. [2010]. Compared to parallax barriers, this multi-layer generalization enhances resolution, increases depth of field, and improves dynamic range. Finally, the first non-heuristic construction of HDR displays with two or more disjoint layers are given in Section 3.4.

Volumetric attenuators share the limitations of other multi-layer displays, particularly increased cost and complexity compared to monolithic or dual-layer designs. As described by Bell et al. [2008, 2010], multi-layer LCDs exhibit decreased brightness, moiré, and colour crosstalk, with additional layers exacerbating problems. Similar limitations are expected with other spatial light modulators. If fabricated with 2D/3D printing, fidelity is restricted by limited-contrast media and by scattering, misalignment, and interreflections. Similar to other automultiscopic displays, including parallax barriers and integral imaging, our design exhibits a finite depth of field. Most significantly, benefits of volumetric attenuators are

best realized by simultaneously increasing the number of layers and the display thickness.

	Volumetric	Integral Imaging	Parallax Barriers	Multi-Layer
2D Resolution	high	low	low	high
3D Resolution	high	moderate	moderate	high
Brightness	high	high	low	moderate-high
Contrast	moderate	moderate	moderate	high
Complexity	high	low	low	moderate
Flip Animations	no	yes	yes	low resolution*

Table 3.1: *Benefits of attenuation-based multi-layer displays. These displays enable sharper, brighter images than existing automultiscopic displays, but require additional layers and possibly thicker enclosures. *Flip animations, with uncorrelated multi-view imagery, are handled by enforcing parallax-barrier-like downsampling, described in Section 3.6.1.*

3.2 Tomographic Image Generation

This section describes how volumetric attenuators are optimally constructed to emit a target light field using tomographic principles. The analysis is presented in flatland, with a straightforward extension to 3D volumes and 4D light fields. First, the forward problem is considered: modeling the light field emitted by a backlit volumetric attenuator. The logarithm of the emitted light field is shown to equal to the negative Radon transform of the attenuation map. Second, the inverse problem is considered: synthesizing an attenuation map to approximate a target light field. The optimal solution is found, in the least-squares sense, using a series expansion method based on iterative tomographic reconstruction principles. Third, a description of how to apply these principles to the generation of images for *layered* attenuation displays is given.

3.2.1 Modeling Volumetric Attenuation

In flatland, a 2D volumetric attenuator is modeled by a continuously-varying attenuation map $\mu(x, y)$, such that the intensity I of a transmitted light ray \mathcal{C} is given by the Beer-Lambert law

$$I = I_0 e^{-\int_{\mathcal{C}} \mu(r) dr}, \quad (3.1)$$

where I_0 is the incident intensity [Hecht, 2001]. Additional scattering and reflection losses are assumed to be negligible. For convenience, the logarithm of the normalized intensity \bar{I} is defined as

$$\bar{I} = \ln \left(\frac{I}{I_0} \right) = - \int_{\mathcal{C}} \mu(r) dr. \quad (3.2)$$

This section considers a volumetric attenuator composed of a single slab, of width w and height h , such that $\mu(x, y)$ can be non-zero only within the interval $|x| < w/2$ and $|y| < h/2$. A relative two-plane light field parameterization $l(u, a)$ is adopted [Chai et al., 2000]. As shown in Figure 3.3, the u -axis is coincident with the x -axis, which bisects the slab horizontally. The orientation of ray (u, a) is defined by the slope $a = s - u = d_r \tan(\theta)$, where d_r is the distance of the s -axis from the u -axis.

In conventional parallel beam tomography [Kak and Slaney, 2001], the Radon transform $p(u, a)$ encodes all possible line integrals through the attenuation map, along each ray (u, a) , such that

$$p(u, a) = \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_{-\frac{w}{2}}^{\frac{w}{2}} \mu(x, y) \delta(d_r(x - u) - ay) dx dy, \quad (3.3)$$

where $\delta(\xi)$ denotes the Dirac delta function. Substituting into Equation 3.2 gives the following expression for the light field $l(u, a)$ emitted when a volumetric attenuator is illuminated by a backlight producing the incident light field $l_0(u, a)$:

$$\bar{l}(u, a) = \ln \left(\frac{l(u, a)}{l_0(u, a)} \right) = -p(u, a). \quad (3.4)$$

In practice, backlights produce uniform illumination so $l_0(u, a) = l_{\max}$ and the light field is normalized so $l(u, a) \in (0, l_{\max}]$. To summarize, tomographic analysis reveals a simple forward model: the logarithm of the emitted light field is equivalent to the negative Radon transform of the attenuation map. For a fixed linear angle $a = a_0$, a 1D slice of the Radon transform $p(u, a_0)$ corresponds to an oblique projection of the attenuation map and, correspondingly, to an emitted oblique view $\bar{l}(u, a_0)$, as shown in Figure 3.3.

3.2.2 Synthesizing Light Fields

With parallel beam tomography, an estimate of the attenuation map $\tilde{\mu}(x, y)$ is recovered from the projections $p(u, a)$ using the inverse Radon transform, conventionally implemented using the filtered backprojection algorithm [Kak and Slaney, 2001]. A direct application of this algorithm yields an estimate for a volumetric attenuator capable of emitting the target light field $\bar{l}(u, a)$:

$$\tilde{\mu}(x, y) = - \int_{-\infty}^{\infty} \bar{l}'(x - (a/d_r)y, a) da. \quad (3.5)$$

Geometrically, the spatially-filtered oblique views $\bar{l}'(u, a)$ are propagated through the attenuation volume, along the rays in Equation 3.3. However, a high-pass filter $\hat{h}(f_u)$ must be applied first to obtain a sharp estimate of $\mu(x, y)$, where the hat symbol denotes the

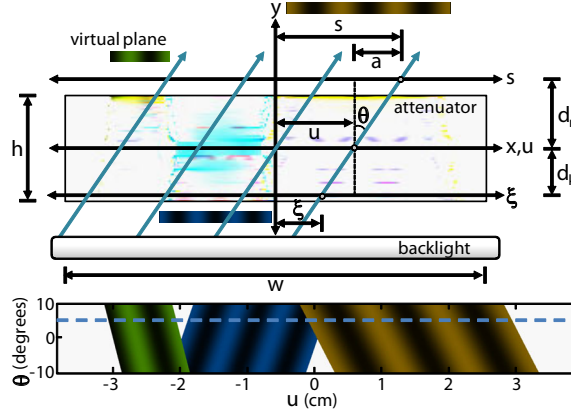


Figure 3.3: Tomographic analysis of attenuation-based displays. (Top) A volumetric attenuator $\mu(x, y)$ is optimized, using Equation 3.11, to emit a light field approximating the virtual planes. (Bottom) The target light field $l(u, a)$, with the dashed line denoting the oblique projection with rays passing through the attenuator above.

1D Fourier transform [Bracewell and Riddle, 1967]. This filter is implemented in the ray domain or frequency domain, as follows.

$$\bar{l}'(u, a) = \bar{l}(u, a) \otimes h(u) = \int_{-\infty}^{\infty} \hat{h}(f_u) \hat{l}(f_u, a) e^{2\pi j u f_u} df_u \quad (3.6)$$

For this application, the oblique views are known *a priori*, so an ideal ramp filter $\hat{h}(f_u) = |f_u|$ should be used.

Equations 3.5 and 3.6 fail to provide a practical method for constructing volumetric attenuators. First, high-pass filtering often leads to negative attenuations, prohibiting fabrication. Similar to the binary attenuation volumes considered by Mitra and Pauly [2009], the target set of oblique views may be inconsistent with any single attenuation map. While filtered backprojection can accommodate minor inconsistencies due to measurement artifacts, gross inconsistencies introduce negative attenuations. Second, filtered backprojection requires projections vary over a full hemisphere (i.e., $-\infty < a < \infty$). In practice, oblique views may be known only over a limited set of angles (e.g., if the light field is rendered or captured with cameras along a limited baseline). In this case, alternative methods are required to produce a sharp estimate of the attenuation map.

Iterative reconstruction algorithms present a flexible alternative to traditional transform methods [Herman, 1995]. While exhibiting greater computational complexity, such methods better account for inconsistent projections over limited angles. We consider a particular *series expansion method*, for which attenuation is modeled by a linear combination of N_b

non-negative basis functions $\bar{\phi}_k(x, y)$:

$$\mu(x, y) = \sum_{k=1}^{N_b} \alpha_k \bar{\phi}_k(x, y). \quad (3.7)$$

The basis can be conventional voxels used in volume rendering [Drebin et al., 1988], or general functions with compact or extended support. The choice of $\bar{\phi}_k(x, y)$ is discussed in Section 3.5.1.

Substituting Equation 3.7 into Equations 3.3 and 3.4 gives the following solution to the forward rendering problem.

$$\bar{l}(u, a) = - \sum_{k=1}^{N_b} \alpha_k \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_{-\frac{w}{2}}^{\frac{w}{2}} \bar{\phi}_k(x, y) \delta(d_r(x - u) - ay) dx dy \quad (3.8)$$

This expression leads to a linear system of equations, when considering a discrete light field \bar{l}_{ij} , such that

$$\bar{l}_{ij} = - \sum_{k=1}^{N_b} \alpha_k P_{ij}^{(k)}, \quad (3.9)$$

where (i, j) are the discrete indices corresponding to the continuous coordinates (u, a) . The projection matrix $P_{ij}^{(k)}$ is given by

$$P_{ij}^{(k)} = \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_{-\frac{w}{2}}^{\frac{w}{2}} \bar{\phi}_k(x, y) \delta(d_r(x - (i\Delta u)) - (j\Delta a)y) dx dy, \quad (3.10)$$

corresponding to line integrals through every basis function k along each ray (i, j) . This system is expressed in matrix-vector form as $\mathbf{P}\alpha = -\bar{\mathbf{l}} + \bar{\mathbf{e}}$, where $\bar{\mathbf{e}}$ is the approximation error. As surveyed by Herman [1995], a wide variety of iterative reconstruction algorithms exist to solve this system, primarily differing in computational complexity and in constraints placed on the error $\bar{\mathbf{e}}$. However, by the Weber-Fechner law, the human visual system responds approximately linearly to logarithmic changes in illumination [Reinhard et al., 2010]. As a result, the attenuation map synthesis is cast as the following non-negative linear least-squares problem.

$$\arg \min_{\alpha} \|\bar{\mathbf{l}} + \mathbf{P}\alpha\|^2, \text{ for } \alpha \geq 0 \quad (3.11)$$

Although requiring iterative optimization, this formulation as a convex optimization problem yields an optimal attenuation map, in the least-squares sense, that emits a target light field with consistent views. This problem is efficiently solved using optimization methods

described in detail in Section 3.5.1.

3.2.3 Layered Attenuation-based Displays

So far, an attenuating volume is considered with either a continuous spatially-varying absorption coefficient, or an expansion into a discrete set of basis functions that are uniformly distributed over the display volume. While such volumes could be fabricated with recent rapid prototyping hardware, other manufacturing processes, such as stacks of LCD panels (Sec. 4) or semi-transparent slides (Sec. 3.5.2), are better represented as a finite number of discrete attenuation layers.

The analysis naturally extends to such multi-layered attenuators. Rather than directly constructing attenuation maps, each mask controls spatially-varying transmittance in a single plane. Following Figure 3.3, ray (u, a) is modulated by N_l layers such that

$$l(u, a) = l_0(u, a) \prod_{k=1}^{N_l} t_k(u + (d_k/d_r)a), \quad (3.12)$$

where $t_k(\xi)$ is the transmittance of mask k (separated by a distance d_k). Taking the logarithm gives the forward model

$$\bar{l}(u, a) = \sum_{k=1}^{N_l} \ln t_k(u + (d_k/d_r)a) = - \sum_{k=1}^{N_l} a_k(u + (d_k/d_r)a), \quad (3.13)$$

where $a_k(\xi) = -\ln t_k(\xi)$ is the absorbance. Similar to Equation 3.9, the linear system $\bar{l}_{ij} = -\sum_{k=1}^{N_l} a_k P_{ij}^{(k)}$ is obtained for a discrete set of rays (i, j) . For multi-layered attenuators, the form of the projection matrix $P_{ij}^{(k)}$ is modified, now encoding the intersection of every ray with each mask. Thus, a similar optimization solves the inverse problem of constructing an optimal multi-layered attenuator. Practically, however, layers have a finite contrast (i.e., maximum transmissivity and opacity) and Equation 3.11 is solved as a *constrained* least-squares problem. As an additional benefit, our optimization encompasses additive decompositions, simply by interpreting \bar{l}_{ij} as the light field, rather than its logarithm, and a_k as the negative of the emittance. Section 3.5 describes additional issues that must be addressed for accurate fabrication (e.g., handling limited contrast and colour gamut, achieving accurate mechanical alignment, and mitigating scattering and reflection).



Figure 3.4: *Multi-layer 3D display.* The “dice” scene is rendered for vantage points to the right and left of the display, shown at the top and bottom of the left column, respectively. Corresponding views of the five-layer prototype (see Figure 3.2) are compared to the right. Inset figures denote the position of layers relative to the scene. Unlike conventional additive volumetric displays, both transparency and occlusion are accurately depicted, allowing for faithful representation of motion parallax for opaque objects. (Right) Counter-clockwise from the upper left are two, three, and five layer decompositions, ordered from the front to the back of the display. All layers are uniformly-spaced and span the same total display thickness.

3.3 Application to 3D Display

This section assesses multi-layer light field decompositions for automultiscopic 3D display. First, the qualitative performance of the tomographic algorithm presented in the previous section is documented, providing intuition into its behavior and how design parameters influence reconstruction accuracy. Second, the quantitative upper bound on depth of field for all multi-layer displays is established, informing system design and motivating display prefiltering [Zwicker et al., 2006]. Third, through experimental studies, rules for optimizing design parameters are developed, including the number of layers and display thickness, to minimize artifacts. This section concludes by evaluating scenes with varying degrees of disparity, occlusion, translucency, and specularly.

3.3.1 Assessing Performance for 3D Display

Consider the “dice” scene in Figure 3.4. Each die is approximately 1.5 cm on a side, with the scene extending 5 cm in depth. In this example and for all others in this chapter and

using the prototype, we assume a $5.7\text{ cm} \times 7.6\text{ cm}$ display with a thickness of 1.25 cm , with evenly-spaced layers. All layers have a resolution of 171 dots per inch (i.e., $149\text{ }\mu\text{m}$ pixels). The target scene is rendered as a light field with 7×7 oblique projections, spanning a field of view of ± 5 degrees from the display surface normal. Following Figure 3.3, the light field is parameterized with the u -axis bisecting the middle of the display and the s -axis coincident with the front layer. The scene is transformed so the red die is enclosed within the display, with other dice extending beyond the surface.

Multi-layer decompositions are obtained using the tomographic algorithm in Section 3.2.3. Figure 3.4 shows masks for two, three, and five layers. First, we observe that 3D objects can be displayed both inside and outside the enclosure. This illustrates the primary benefit of multiplicative displays over conventional additive volumetric displays: through modulation, spatio-angular frequencies are created corresponding to objects outside the display. Modulation also allows occlusion to be accurately depicted. Second, objects inside or near the display are rendered at full-resolution and with the same brightness as the target light field, representing the primary benefits compared to conventional automultiscopic displays (e.g., parallax barriers and integral imaging). Third, although intentionally rendered with a finite depth of field, halos appear around objects outside the enclosure, with additional layers mitigating these errors.

As shown in Figure 3.4, optimized masks exhibit predictable structure. Although not produced using filtered backprojection, the qualitative performance of iterative reconstruction can best be anticipated using this simple procedure; if applied, each view would first be sharpened, using Equation 3.6, and then smeared through the layers to assign absorbance by Equation 3.5. For a point on a virtual object within the display, a sharp image forms on the closest layer, since smeared views align there, and defocused images form on layers above or below. This is observed for the red die in Figure 3.4 and the wheel of the car in Figure 3.1, both appearing sectioned over the layers spanning their physical extent. For objects inside, our decomposition functions as the multiplicative equivalent to depth filtering in additive multi-layer displays [Akeley et al., 2004; Suyama et al., 2004]. However, iterative reconstruction enforces physical constraints on attenuation, resolves inconsistencies between views in a least-squares sense, and constructs attenuation patterns to illustrate objects beyond the display (e.g., the cyan and yellow dice).

3.3.2 Characterizing Depth of Field

The *depth of field* of an automultiscopic display characterizes the maximum spatial frequency that can be depicted, without aliasing, in a plane parallel to the display at a given

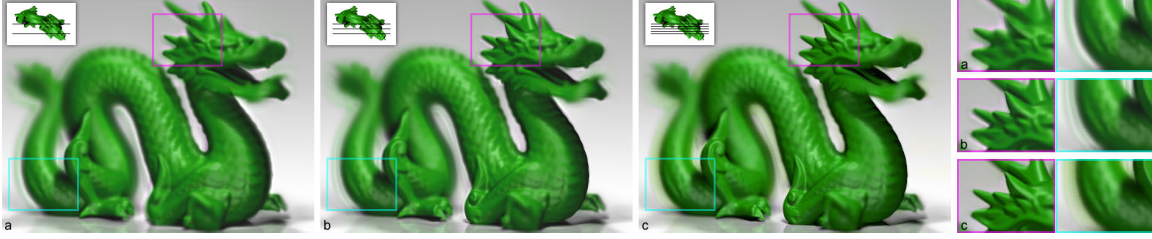


Figure 3.5: *Influence of multi-layer depth of field on reconstruction artifacts. Shown from left to right are reconstructions of the “dragon” scene, using two, three, and five layers, seen when viewing directly in front. Magnified regions are compared on the right. The magenta region, located on the head and inside the display, is rendered with increasing resolution as additional layers are incorporated. The cyan region, located on the tail and behind the display, exhibits noticeable halo artifacts, similar to the dice in Figure 3.4. As described in Section 3.3.2, the upper bound on depth of field for multi-layer displays indicates spatial resolution is inversely proportion to scene depth. Since high spatial frequencies are required to depict the tail edge, artifacts result without proper prefiltering. In this example, an approximation of the prefilter in Equation 3.17 is applied assuming a five-layer decomposition, allowing artifacts to persist in the two-layer and three-layer decompositions.*

distance. As described by Zwicker et al. [2006], depth of field is determined by the spectral properties of a display. For parallax barriers and integral imaging considered in flatland, discrete sampling of rays (u, a) produces a spectrum $\hat{l}(f_u, f_a)$ limited to a rectangle. Following Chai et al. [2000], the spectrum of a Lambertian surface, located a distance d_o from the middle of the display, is the line $f_a = (d_o/d_r)f_u$. Thus, the spatial cutoff frequency is found by intersecting this line with the spectral bandwidth. For parallax barriers and integral imaging, Zwicker et al. [2006] use this construction to show the spatial frequency f_ξ in a plane at d_o must satisfy

$$|f_\xi| \leq \begin{cases} \frac{f_0}{N_a}, & \text{for } |d_o| + (h/2) \leq N_a h \\ \left(\frac{h}{(h/2) + |d_o|} \right) f_0, & \text{otherwise} \end{cases}, \quad (3.14)$$

where N_a is the number of views, h is the display thickness, and $f_0 = 1/2p$ is the cutoff frequency for layers with pixels of width p .

As shown in Figure 3.5 and observed by Gotoda [2010], multi-layer displays exhibit finite depth of field, which, to date, has not been quantitatively described. We observe the upper bound on depth of field is found by similarly considering the maximum spectral bandwidth achievable with multiple layers. The Fourier transform of Equation 3.12 expresses the spectrum of any multi-layer display as

$$\hat{l}(f_u, f_a) = \bigotimes_{k=1}^{N_l} \hat{t}_k(f_u) \delta(f_a - (d_k/d_r)f_u), \quad (3.15)$$

where \bigotimes denotes repeated convolution (see Figure 3.6). Here, the backlight is uniform, such that $l_0(u, a) = 1$, and the light field is normalized such that $l(u, a) \in (0, 1]$. The upper bound on depth of field is found by intersecting the line $f_a = (d_o/d_r)f_u$ with the boundary of maximum spectral support given by Equation 3.15, using the fact that each mask spectrum $\hat{t}_k(f_\xi)$ has an extent of $\pm f_0$.

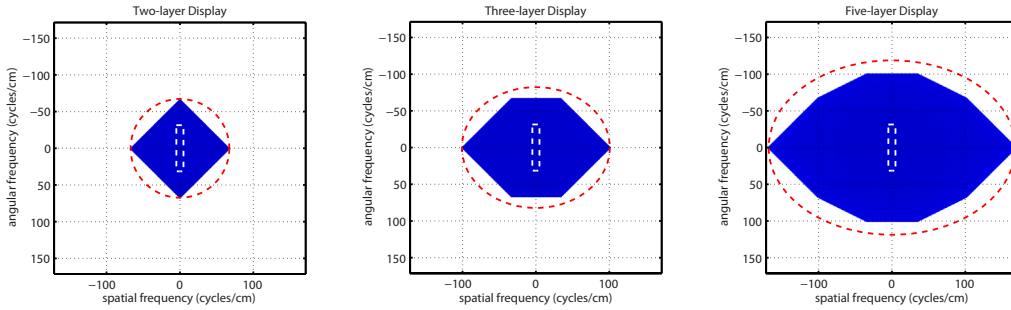


Figure 3.6: Spectral support for multi-layer displays. The spectral support (shaded blue) is illustrated for two-layer, (left), three-layer (middle), and five-layer (right) displays, evaluated using the geometric construction given by Equation 3.15. Note that the shaded area indicates the achievable region of non-zero spectral support. The system parameters correspond with the prototype in Section 3.3.1, with the variable $d_r = h/2$. The ellipse corresponding to the upper bound on achievable spatio-angular frequencies is denoted by a dashed red line. Note that the spectral support of a multi-layer display exceeds the bandwidth of a conventional, two-layer automultiscopic display with a similar physical extent, shown as a dashed white line and given by Equation 3.14.

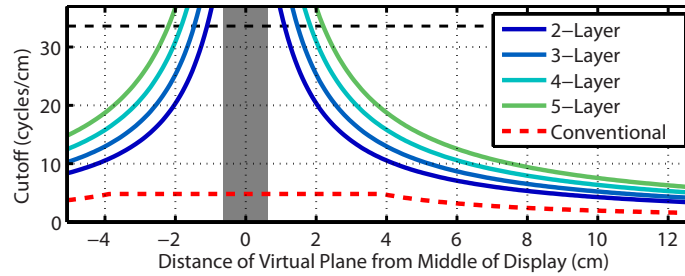


Figure 3.7: Upper bound on multi-layer depth of field. The spatial cutoff frequency is shown for conventional parallax barriers and integral imaging, using Equation 3.14, and for N_l -layer displays, using Equation 3.17. Parameters correspond with the prototype. Spatial resolution exceeds conventional architectures, particularly near or within the display enclosure (shaded region). The maximum spatial resolution of a single mask is denoted by a horizontal dashed line, indicating full-resolution display is possible within the enclosure.

For two layers, this construction yields the upper bound

$$|f_\xi| \leq \left(\frac{h}{(h/2) + |d_o|} \right) f_0. \quad (3.16)$$

Comparing Equations 3.14 and 3.16 indicates that parallax barriers and integral imaging, both of which employ fixed spatio-angular tradeoffs, achieve the optimal resolution for objects located far from the display, but reduce resolution for objects close to the display by up to a factor of N_a . As shown in this section, decompositions computed by the proposed tomographic reconstruction more fully realize the achievable spatio-angular bandwidth, obtaining higher-resolution images nearby.

The upper bound for multiple layers is assessed by similar methods, with the geometric construction providing the exact upper bound. However, we observe that repeated convolution of N_l mask spectra $\hat{t}_k(f_u, f_a)$, each with extent $|f_u| \leq f_0$ and constrained to the line $f_a = (d_k/d_r)f_u$, converges to a bivariate, zero-mean Gaussian distribution by the central limit theorem [Chaudhury et al., 2010]. The covariance matrix of this distribution is equal to the sum of the covariance matrices for each mask. Thus, contours of the light field spectrum $\hat{l}(f_u, f_a)$ will be ellipses. As before, intersecting the line $f_u = (d_o/d_r)f_a$ with the ellipse bounding the spectral bandwidth gives an approximate expression for the upper bound, as follows.

$$|f_\xi| \leq N_l f_0 \sqrt{\frac{(N_l + 1)h^2}{(N_l + 1)h^2 + 12(N_l - 1)d_o^2}} \quad (3.17)$$

Figure 3.7 compares the upper bound for multi-layer and conventional displays. As before, this upper bound indicates the potential to increase the resolution for objects close to the display; yet, even in the upper bound, multi-layer displays exhibit a finite depth of field similar to existing automultiscopic displays. For distant objects, resolution remains inversely proportional to object depth.

3.3.3 Optimizing Display Performance

To construct a practical multi-layer display, such as the prototype in Figure 3.2, one must select two key design parameters: the total thickness h and the number of layers N_l , where layers are assumed to be uniformly distributed such that $d_k \in [-h/2, h/2]$. The upper bound on the depth of field informs selection of h and N_l , yet, with the proposed optimization algorithm, further experimental assessment is required for clear design rules—since the upper bound may not be achievable for all scenes. As shown in Figure 3.8, optimization increases spatial resolution compared to conventional displays, but also introduces artifacts.

In the remainder of this section, display parameters are analyzed with the goal of minimizing artifacts. A more detailed, visual evaluation of reconstruction performance for multiple different light fields containing 3D scenes and varying display parameters is included in Appendix A.

For a conventional parallax barrier, with pixel width p , field of view α , and N_a views, the separation h_b between the layers is

$$h_b = \frac{N_a p}{2 \tan(\alpha/2)}, \quad (3.18)$$

with $h_b = 0.6$ cm for the display and light field parameters listed in Section 3.3.1. As shown by the red line in Figure 3.7, we expect parallax barriers to create lower-resolution images than multi-layer decompositions. This is confirmed in Figure 3.14. However, the upper bound does not indicate whether a high peak signal-to-noise ratio (PSNR) is obtained for a given display configuration.

As shown in Figure 3.9, a database of light fields facilitates display optimization. Views are rendered with a manually-selected depth of field, approximating combined light field anti-aliasing and display prefilters (see Section 3.5.1). Several observations can be made regarding general design principles. First, PSNR is maximized by enclosing the scene within the display (for a sufficiently large number of layers). Thus, multi-layer displays can be operated in a mode akin to additive volumetric displays, wherein high resolution is achieved for contained objects. However, particularly for mobile applications, displays must be thinner than the depicted volume. Second, addressing this, we find, for a fixed display thickness (e.g., that of a conventional parallax barrier), addition of layers increases PSNR. However, artifacts persist even with a large number of layers. Thus, the layered prototype closely approximates the performance of volumetric attenuators, despite relatively few layers. Third, for a fixed number of layers N_l , there is an optimal display thickness determined by the desired depth range. In summary, tomographic image synthesis obtains high-PSNR reconstructions with small numbers of static layers in compact enclosures. Through such simulations, the optimal design parameters h and N_l can be determined depending on form factor or system complexity, respectively, subject to image fidelity and depth range requirements.

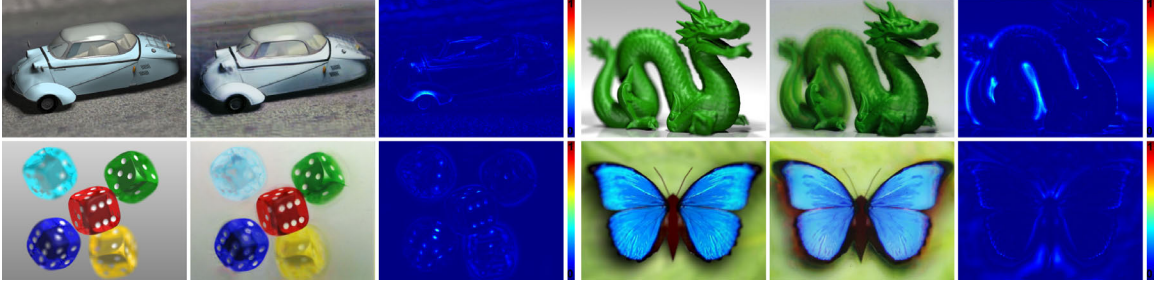


Figure 3.8: Multi-layer display performance. For each scene from left to right, a direct view of the target light field is compared to a prototype photograph and the absolute error between simulated and target views. Due to proper prefiltering, artifacts are evenly distributed in depth, occurring mostly near high-contrast edges (e.g., dragon silhouette and car fender). Note the specular highlight on the car roof is preserved, with reduced contrast, together with translucency of the windows.

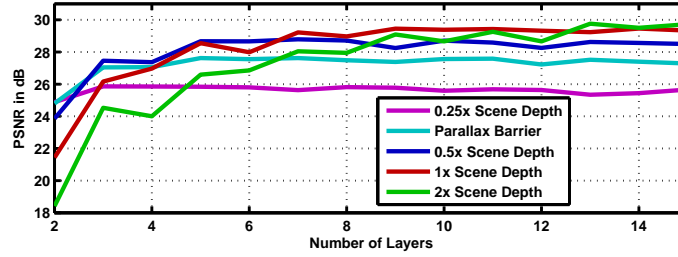


Figure 3.9: Minimizing artifacts with optimal display designs. To determine rules for optimizing the thickness h and number of layers N_l , the tomographic decomposition was applied to four scenes: “butterfly”, “dice”, “dragon”, and “car”. As shown above, the average PSNR is plotted as a function of N_l and h . For a fixed thickness, a finite number of masks closely approximates the maximum-achievable PSNR, obviating fabrication of continuously-varying attenuation.

3.4 Application to HDR Display

The previous sections show how displays composed of two or more layers can present 4D light fields. Multi-layer LCD panels are currently beginning to enter the consumer market, for example in the form of the Nintendo 3DS parallax barrier display [Jacobs et al., 2003]. Once such displays are available, they can not only be used for 3D display, but also for increasing the dynamic range of 2D images [Reinhard et al., 2010; Seetzen et al., 2004].

With non-negligible separations between attenuators, multi-layer HDR decomposition becomes a 3D display problem, since all viewpoints must produce an accurate rendition of the 2D image within the target field of view. A constrained tomographic solver inherently accounts for the limited contrast of each layer, thereby allowing simultaneous optimization

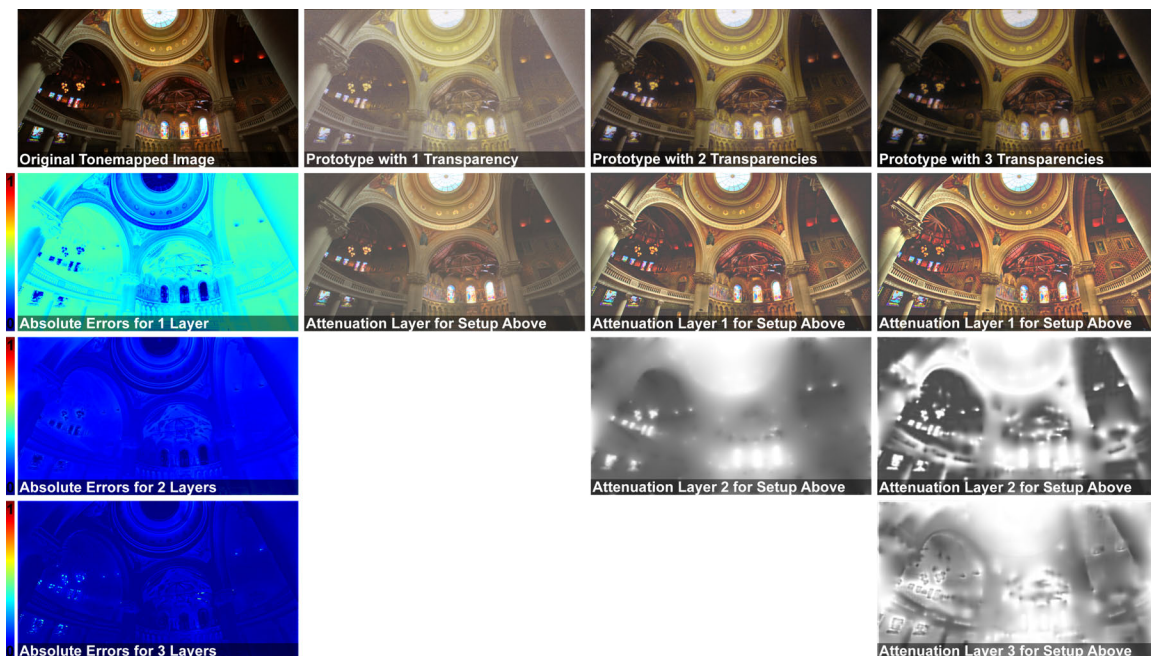


Figure 3.10: *Multi-layer, parallax-free HDR image display. The first row shows the desired high-contrast image, together with photographs of stacked printed transparencies placed on a light box. From left to right, the physical prototype configuration consists of a single transparency (second column), two transparencies separated by a 1/8" plexiglass spacer (third column), and three transparencies with 1/8" plexiglass spacers between each (fourth column). Although overall contrast increases, accurately displaying high-frequency details over a 10° field of view at high-contrast is problematic; for example, as seen around the left-most windows in the absolute error plots in the left column, artifacts occur even with multiple layers. The computed attenuation layers are shown underneath each photograph. Note that layers are optimized subject to the black level of the printer.*

of dynamic range and accurate multi-view imagery. In a 2D HDR display mode, the target light field encodes a single plane (e.g., coincident with the front layer), with a texture given by the desired HDR image. Figure 3.10 shows the result from a parallax-free 2D HDR display prototype. The optimized layers in the lower rows account for the non-zero black level of the printing process and are scaled appropriately before printing. Note the tomographic algorithm naturally handles decomposition into more than two disjoint layers.

In the prototype, the target light field has been optimized for a grid of 7×7 viewpoints, all showing the same HDR image. As described in Section 3.5, each layer is printed on a low-contrast inkjet transparency. For all experiments in this section, we follow a standard procedure for HDR image display, wherein the dynamic range is expanded for the luminance channel only [Reinhard et al., 2010; Seetzen et al., 2004], with chrominance assigned to the front layer after optimization. Although colour contrast is limited in this fashion, the human

visual system is more sensitive to luminance. In practice, such luminance-chrominance decompositions reduce system complexity, since monochromatic displays can be used for underlying layers, mitigating colour crosstalk and moiré.

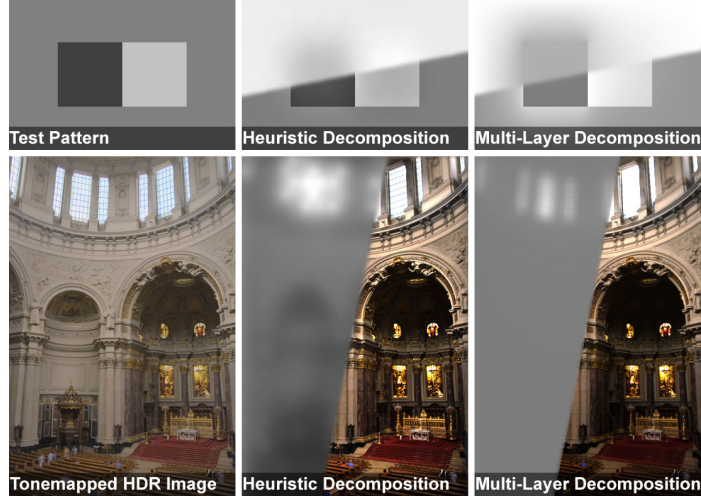


Figure 3.11: *Heuristic vs. tomographic HDR image synthesis. The tomographic method, assuming a high-resolution backlight, is compared to heuristic algorithms [Seetzen et al., 2004] designed to operate with a low-resolution backlight. The backlight and front panel images are divided into upper-left and lower-right halves in each example, respectively. Note the tomographic optimization produces similar patterns to prior heuristics. This confirms existing architectures achieve near-optimal results for HDR display, while providing a generalization to multiple layers.*

The optimized layers, as evaluated using Equation 3.11 and shown in Figure 3.10, suggest a trend for optimal constructions: the front layer is a sharpened target image and underlying layers appear blurred. This configuration is effective since a sharpened front layer preserves high spatial frequencies in a view-independent manner, with blurred underlying layers similarly enhancing dynamic range despite changes in viewpoint. The degree of blur is determined by the parallax, as defined by the field of view and layer spacing. Note that Seetzen et al. [2004] originally motivated blurring the back layer as a means to tolerate misalignment between layers; we observe that parallax due to spaced layers produces similar alignment issues. The solution in both cases is a low resolution or blurred back layer. Hence, our results are consistent with existing methods for dual-modulation HDR image display, pairing high-resolution LCD front panels with low-resolution LED backlights. As described by Seetzen et al. [2004], such displays use a “heuristic” image synthesis method motivated by both physiological and hardware constraints: the target HDR image is blurred to the resolution of the rear layer, while the front panel displays a compensated sharpened image.

Figure 3.11 compares this heuristic approach to the tomographic optimization, concluding that qualitatively-similar patterns are produced for dual-layer decompositions—despite significant differences between LED distributions and our assumed high-resolution backlight. This indicates existing HDR displays are near-optimal, in the least-squares sense, and further advocates for low-resolution, spatially-programmable backlighting.

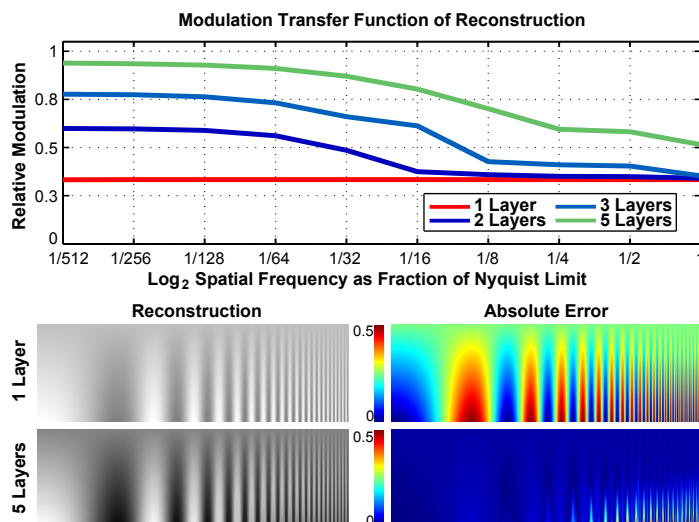


Figure 3.12: Contrast-resolution tradeoff in multi-layer HDR. (Top) MTF plots illustrating the reconstructed contrast, for sinusoids rendered with maximum contrast, as achieved using multiple disjoint layers with a fixed display thickness. (Bottom, Left) Reconstructed direct views of a light field for a single plane, coincident with the front layer, with a texture that increases in spatial frequency, from left to right, and contrast, from bottom to top. (Bottom, Right) Absolute error maps indicate multiple layers increase dynamic range, but high spatial frequencies with high contrast are difficult to display with disjoint, multi-layer architectures.

Additional layers increase dynamic range, yet not equally well for all spatial frequencies. Optimal performance occurs with no layer separation, but may not be practically achievable. As separation increases, depicting high spatial frequencies at high contrast becomes difficult over the full field of view. This is a fundamental limitation of multi-layered HDR displays, as well as existing dual-modulation architectures. In Figure 3.12, we characterize this effect by considering the light field of a plane, coincident with the front layer, containing maximum-contrast sinusoidal textures. This figure charts the modulation transfer function (MTF) [Hecht, 2001]: the achieved Michelson contrast, divided by maximum possible contrast, as a function of target spatial frequency. Contrast is averaged over all light field views. In addition to the MTF, which shows achievable contrast averaged over space and angle, Figure 3.13 illustrates spatial and angular variation in high contrast 2D target light fields that contain only spatial variation. The synthesized light fields in the second row illustrate how the achieved contrast decreases for increasing spatial frequencies. We conclude that

building multi-layer displays capable of both 3D and 2D HDR modes involves a careful interplay between design constraints; large depth of field for 3D applications necessitates larger gaps, limiting the field of view and maximum spatial frequency for which 2D HDR display is achieved.

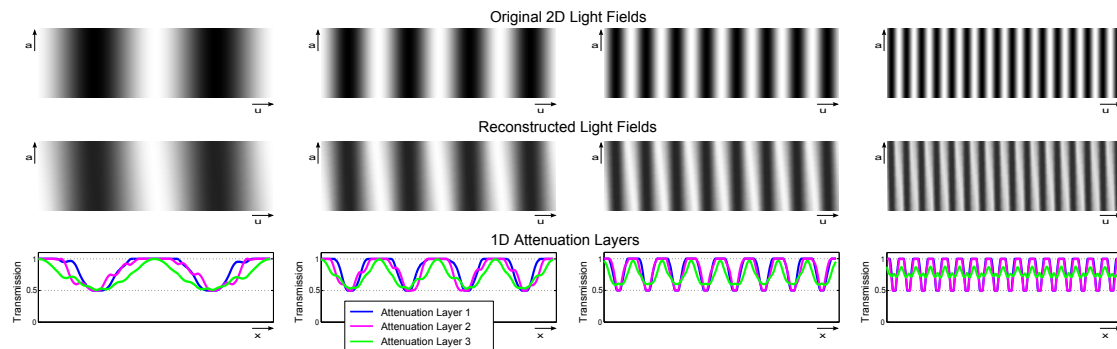


Figure 3.13: Analysis of the modulation transfer function. Multiple 2D light fields, containing different spatial frequencies at a high contrast, are shown in the top row. Light fields are parameterized with the spatial coordinate u on the front-most attenuation layer. The light fields are generated to contain only spatial variation along the u -axis and correspond to desired 1D image frequencies appearing on the front layer. A reconstruction for three disjoint layers, each having a limited contrast of 2:1, is shown in the centre row. The corresponding optimized 1D layers are shown along the bottom row. Note that, although high contrast can be reproduced for lower spatial frequencies, it becomes more challenging as the spatial frequency increases. The average difference in contrast between reconstructed and target light field for these examples provides the data for the modulation transfer function shown in Figure 3.12 and highlights that MTF computed for a fixed display thickness.

3.5 Implementation

This section describes our attenuation-based display prototypes, developing software for tomographic image synthesis and hardware for multi-layer displays comprising printed transparencies.

3.5.1 Software

Light fields are rendered with POV-Ray as 7×7 oblique projections within a field of view of ± 5 degrees. A depth-dependent 4D anti-aliasing filter is applied [Levoy and Hanrahan, 1996]; this is practically achieved by rendering each view with a limited depth of field (i.e., a finite aperture camera). As discussed in Section 3.3.2, volumetric and multi-layer displays also exhibit a limited depth of field, leading to a circle of confusion linearly proportional

to distance from the display. As a result, a combined anti-aliasing and display prefilter is approximated by a single selection of the camera aperture in POV-Ray, with the number of angular views determined by the user, based on the desired depth of field.

Following standard practice in tomographic reconstruction [Herman, 1995], we solve Equation 3.11 using a series expansion into a set of normalized linear basis functions, rather than with a discrete voxel representation. This yields smoother reconstructions and mitigates artifacts occurring due to discrete sampling of rays within the projection matrix. We note, however, that linear reconstruction filters applied to the logarithm of the light field do not correspond to the same filters applied to the light field directly.

We use a sparse, constrained, large-scale trust region method [Coleman and Li, 1996] to solve Equation 3.11. Limited layer contrast is incorporated as a constraint. For each light field, with $384 \times 512 \times 7 \times 7$ samples, the three colour channels are evaluated independently. While neglecting colour crosstalk between layers, such crosstalk could be incorporated into the optimization at the expense of greater memory requirements and longer computation time. A set of N_l masks, each with 384×512 pixels, are considered. For the light fields studied in this work, the solver typically converges in 8-14 iterations. On average, optimization takes 12 minutes, including projection matrix computation, for five layers using a 2.4 GHz 64-bit Intel Core 2 workstation with 8 GB of RAM. Although significant acceleration can be achieved with a GPU-based solver as described in Sections 4.2.3 and 4.3.2, the offline solver provides high quality reconstructions with sufficiently low compute times for static, layered prototypes.

3.5.2 Hardware

Our display prototypes (see Figure 3.2 and Appendix A) consist of five layers separated by clear acrylic sheets, each 0.3175 cm thick. Layers are printed on transparencies at 300 dots per inch using an Epson Stylus 2200 inkjet printer with six colour primaries. In practice, interreflections between layers do not lead to visual artifacts. Similarly, reflections on the display surface are minimized by dimming ambient lighting (an anti-reflection coating could be applied if needed). Moiré is not observed with our layered fabrication, likely due to ink naturally blending neighboring pixels. We ignore scattering and diffraction due to the ink pigments, although both likely place an upper limit on the achievable spatio-angular resolution. As shown in Figures 3.8 and 3.10, the resulting prototype demonstrates accurate 3D and 2D HDR display using tomographic image synthesis.

3.6 Discussion

3.6.1 Benefits and Limitations

Summarized earlier in Section 3.1.1 and Table 3.1, capabilities of multi-layer displays are comparable to volumetric and automultiscopic displays, particularly parallax barriers and integral imaging. While volumetric displays faithfully reproduce perceptual depth cues, most cannot represent objects beyond the display; many require moving assemblies and often cannot depict opaque scenes. Yet, the inexpensive alternatives of parallax barriers and integral imaging cannot reproduce high-resolution images, even in the display plane (see Figure 3.14). Attenuation-based multi-layer displays present a unique set of capabilities within this space. Foremost, depth of field can exceed that of conventional automultiscopic displays, allowing representation of objects floating beyond the enclosure. Within or near the display, full-resolution depiction is routinely achieved (see Figure 3.8). Multi-layer displays, implemented with printed transparencies, approach the brightness of integral imaging, with additional layers further enhancing contrast. Finally, similar to volumetric displays, accommodation is preserved for objects within the display, with sufficiently-dense layering.

These benefits come with increased mechanical and computational complexity, as well as the introduction of reconstruction artifacts. This complexity has proven manageable, involving additional alignment and separation of printed transparencies. As found in Section 3.3.3, performance is enhanced by simultaneously increasing the number of layers and the thickness of the display. Although the necessary “thick printing” processes for static displays may prove feasible for larger scales (e.g., by adapting general 3D printing to our multi-planar application), embodiments are currently limited by the capabilities of our iterative reconstruction. With our implementation, which stores the sparse projection matrix directly, the display dimensions, image resolution, and number of layers lie at the upper extent afforded by system memory. However, matrix-free optimization is routinely applied to resolve similar issues in computed tomography [Herman, 1995]; Section 4.3.2 shows how real-time framerates can be achieved with a GPU implementation of the simultaneous algebraic reconstruction technique (SART).

While the following chapter demonstrates dynamic light field display using multi-layer LCDs, the prototype introduced in this chapter appears best-suited for static 3D signage, being moderately more expensive and complex than parallax barriers, yet providing significantly enhanced resolution and brightness. Two market applications are seen within this scope, differing by the relation of interlaced views. First, for 3D display, multi-view imagery



Figure 3.14: *Benefits of multi-layer automultiscopic displays. Simulated views of the “dragon” are shown, from left to right, using integral imaging, parallax barriers, and the proposed multi-layer approach. Display parameters are selected to allow 7×7 views, leading to a similar reduction in resolution for conventional methods. Compared to parallax barriers, the multi-layer approach is 49 times brighter and exceeds the resolution of both methods by a similar factor.*

is correlated, for which Figure 3.8 demonstrates our method is well-suited. Second, primarily for advertising, multi-view imagery can be uncorrelated. Known colloquially as *flip animations*, different pictures are presented depending on viewpoint. Tomographic image synthesis does not produce compelling decompositions with such sequences (see Figure 3.15). In contrast, parallax barriers and integral imaging depict such sequences, although at reduced resolution. Yet, the addition of an equivalent spatial downsampling constraint to the tomographic solver yields similar results with enhanced contrast (Figure 3.15, right). As with other automultiscopic displays, limited depth of field is best exploited by transforming scene content to predominantly occupy the region near the display. The tomographic solver considers a finite field of view, outside of which artifacts may occur, although graceful degradation to a 2D image is observed. In contrast, conventional automultiscopic displays create periodic repetition of the central viewing zone, albeit with the cyclic appearance of pseudoscopic imagery.

Finally, a wide field of view is desirable. Figure 3.16 evaluates the reconstruction performance of attenuation-based multi-layer displays for an increasing field of view in the desired light field. Two different perspectives for fields of view of 10° , 20° , and 45° are shown. As seen in these simulations, the performance for wider fields of view drops due to the increased disparity between the different views. Even a solution to the nonlinear attenuation-based image formation (see Eq. 3.12), as derived in Appendix A.4, only produces moderately better results compared to the linear solution (Eq. 3.13).

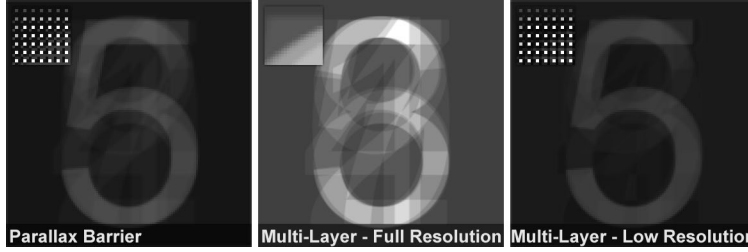


Figure 3.15: *Flip animations with automultiscopic displays. The “numbers” light field consists of a 3×3 mosaic of views, each depicting a different Arabic numeral. Central views with parallax barriers, the direct tomographic method, and the constrained form of the tomographic method are shown from left to right. Note that we assume a multi-layer display with three layers; each mask is assumed to have a contrast ratio of 4:1. While a direct application of Equation 3.11 leads to significant crosstalk, constraining the solver with similar downsampling yields an accurate flip animation, with enhanced contrast.*

3.6.2 Future Work

Any commercial implementation must address prototype limitations. Manual layer alignment is slow and error-prone; specialized multi-planar printing may be developed that, while similar to conventional rapid prototyping, can deliver higher throughput at lower cost. For example, existing 3D printers support 2D printing on fabricated, albeit opaque, 3D surfaces [Z Corporation, 2010]. Similarly, enhanced optimization methods are needed to allow larger, higher-resolution displays with greater numbers of layers.

The proposed multi-layer generalization of parallax barriers opens the door to similar modifications of existing display technologies. As documented in Figure 3.9, addition of layers alone cannot eliminate artifacts. Similar to Lanman et al. [2010], it may be possible to exploit temporal modulation to obtain more accurate light field reconstructions. While the upper bound on depth of field indicates a potentially significant gain in spatio-angular resolution, factorization methods must first be developed for such dynamically-modulated stacked displays. In contrast, combinations of additive and multiplicative layers may yield similar gains, while also enhancing brightness. Such displays are efficiently modeled with the emission-absorption volume rendering equation [Sabella, 1988]. Finally, the formulation introduced in this chapter facilitates the development of non-planar, volumetric displays with arbitrary curved surfaces.

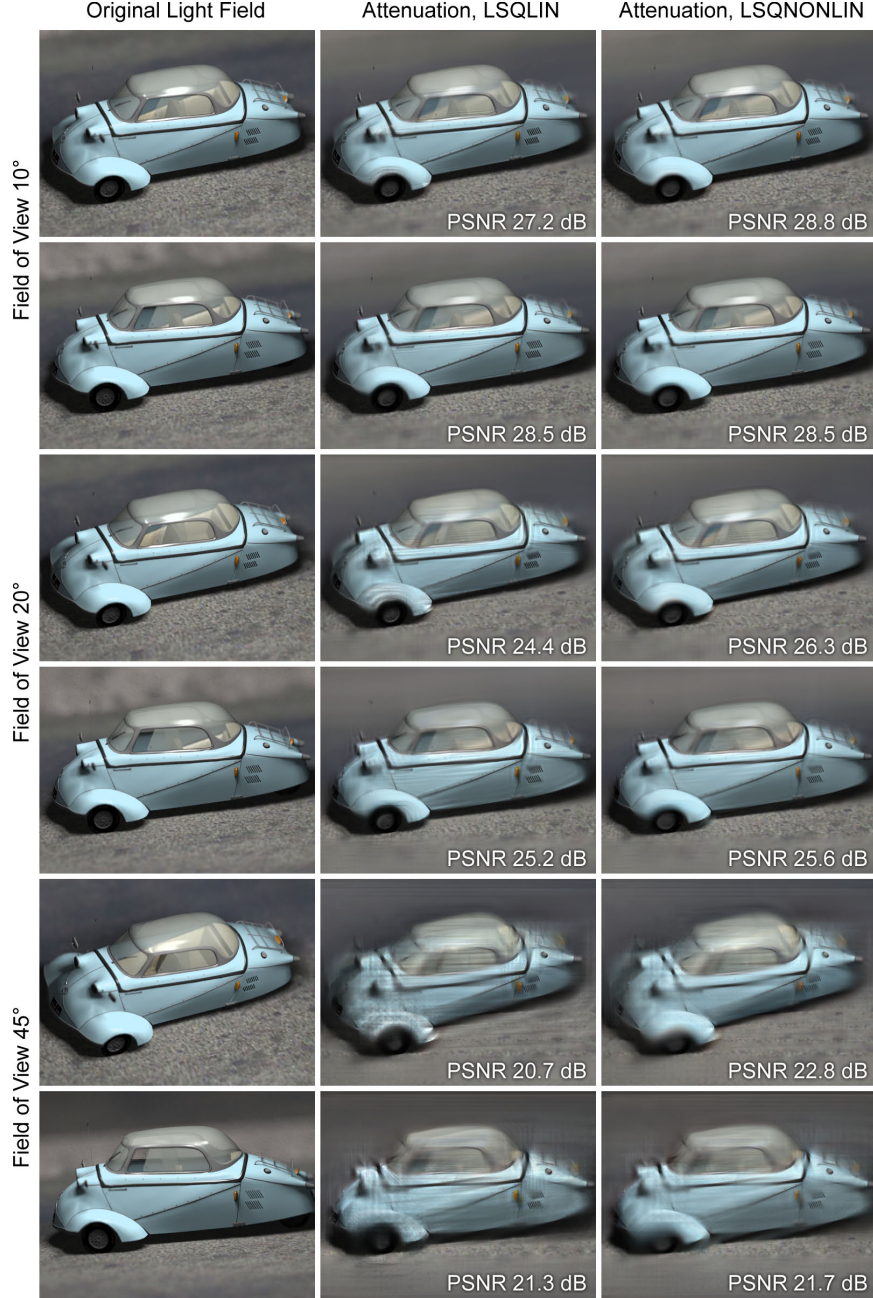


Figure 3.16: An increasing field of view of the desired light field decreases the correlation between different perspectives in a light field, thereby decreasing the reconstruction quality. The rows show two different perspectives of a light field for fields of view of 10°, 20°, and 45°. Column 1 shows the target light field, column 2 the synthesized light field using a linear solution in log-space, and column 3 shows the nonlinear solution that is detailed in Appendix A.4.1. Although the nonlinear solution mitigates some of the artifacts, wide field of view light field synthesis using multi-layer displays remains a challenging problem.

Chapter 4

Dynamic Light Field Display using Multi-Layer LCDs

This chapter introduces polarization field displays as an optically-efficient construction allowing dynamic light field display using multi-layered LCDs. Such displays are constructed by covering a stacked set of liquid crystal panels with a single pair of crossed linear polarizers. Compared to stacks of light attenuating layers, as discussed in the previous chapter, the individual layers in polarization field displays act as spatially-controllable polarization rotators. Section 4.2 demonstrates that multiple polarization rotating layers can be controlled by tomographic image generation techniques established for attenuation-based multi-layer displays in chapter 3.

The polarization field design is verified in Section 4.3.1 by constructing a prototype LCD stack using modified off-the-shelf panels. Furthermore, Section 4.3.2 introduces computational image synthesis techniques that achieve real-time framerates using a GPU-based SART implementation supporting both polarization-based and attenuation-based architectures. Experiments confirm the image formation model and verify polarization field displays achieve increased brightness, higher resolution, and extended depth of field, as compared to existing automultiscopic display methods for dual-layer and multi-layer LCDs.

4.1 Introduction and Motivation

The emergence of consumer, glasses-based stereoscopic displays has renewed interest in glasses-free automultiscopic alternatives. Manufacturers are beginning to offer such displays, supporting both binocular and motion parallax cues, primarily using two long-standing technologies: parallax barriers [Ives, 1903] and integral imaging [Lippmann, 1908]. Yet, these approaches have several well-documented limitations compared to existing stereoscopic displays, including: decreased resolution, potentially reduced brightness, and often narrow depths of field such that objects separated from the display appear blurred. Alter-

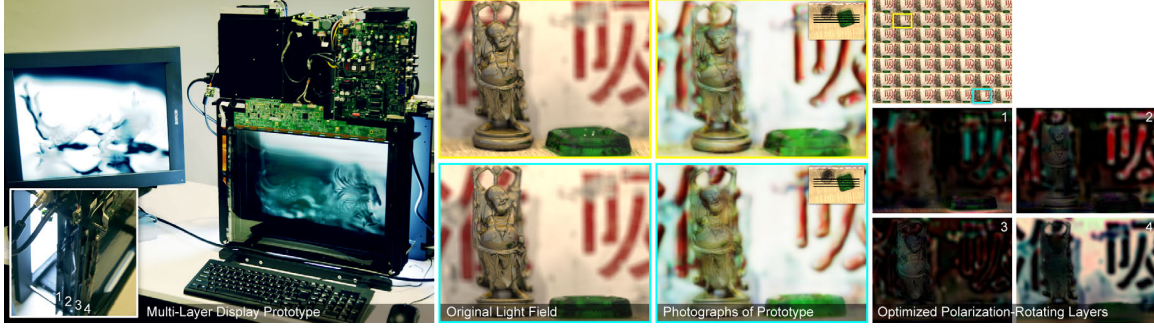


Figure 4.1: *Dynamic light field display using polarization field synthesis with multi-layered LCDs. (Left) An optically-efficient polarization field display is constructed by covering a stack of liquid crystal panels with crossed linear polarizers on either side of the stack. Each layer functions as a polarization rotator, rather than as a conventional optical attenuator. (Right, Top) A target light field. (Right, Bottom) Light fields are displayed, at interactive refresh rates, by tomographically solving for the optimal rotations to be applied at each layer. Layers are visualized by linearly mapping optimized rotations, restricted to $[0, \pi/2]$ radians, to image intensities. (Middle) A pair of simulated views is compared to corresponding photographs of the prototype on the left and right, respectively. Inset regions denote the relative position with respect to the display layers, shown as black lines, demonstrating objects can extend beyond the display surface.*

natives are being pursued, spanning volumetric to holographic systems; yet, particularly for mobile applications, a display is required that leverages existing or emerging spatial light modulation technologies compatible with thin form factors and requiring minimal power consumption.

The techniques presented in this chapter are inspired by systems that address these issues using well-developed LCD technology. As shown by Jacobs et al. [2003], dual-layered LCDs can be operated as parallax barriers, compatible with full-resolution 2D content and supporting 3D modes with reduced resolution and brightness. Lanman et al. [2010] increase the optical efficiency of dual-layer LCDs using content-adaptive parallax barriers, although at the cost of increased computational complexity. Several researchers have further described multi-layer LCDs for automultiscopic display. Early work on three-layer LCDs includes that of Loukianitsa and Putilin [2002] and Putilin and Loukianitsa [2006]. More recently, Gotoda [2010] and the techniques presented in chapter 3 of this thesis establish multi-layer light field displays in combination with tomographic image generation as an alternative automultiscopic display technology. Yet, these works share a common architecture: LCDs are directly stacked such that each layer implements a spatial light modulator that *attenuates* light.

In this chapter we describe optically-efficient architectures and computationally-efficient al-

gorithms for automultiscopic display using multi-layered LCDs. In contrast to prior work, we can operate these layered architectures as either attenuation-based light field displays or *polarization field displays*; the latter type is constructed by covering multiple liquid crystal panels with a pair of crossed linear polarizers. Each layer functions as a *polarization rotator*, rather than a light attenuator. In the approach described in this chapter, light throughput is increased by replacing colour filter arrays with field sequential colour illumination, as afforded by emerging high-speed LCDs. We also propose a computationally-efficient tomographic solver for interactive applications. The design is validated using a prototype constructed with modified off-the-shelf LCDs, confirming improvements in resolution, brightness, and depth of field relative to existing dual-layer and multi-layer LCDs. Thus, through polarization field displays, we endeavor to leverage existing and emerging LCD technology for practical dynamic light field display applications.

4.1.1 Overview of Benefits and Limitations

As shown in Figure 4.2, we construct a polarization field display and compare its performance to prior attenuation-based designs. A dynamic attenuation-based display is created by placing polarizers between every liquid crystal panel. While inspired by Gotoda [2010, 2011] and the techniques developed in chapter 3, the proposed construction is the first to extend the benefits established in those works to dynamic imagery, including increased spatial resolution, display brightness, and depth of field relative to parallax barriers and integral imaging. Brightness is further optimized by eliminating unnecessary polarizing films and colour filter arrays. Multi-layer polarization-based display offers similar benefits in resolution, brightness, and depth of field, while exhibiting fewer artifacts relative to attenuation layers. While attenuation-based displays were optimized off-line in chapter 3, the GPU-based SART solver presented in this chapter is the first to enable control of either attenuation-based or polarization-based displays at interactive refresh rates.

The proposed design shares the limitations of other multi-layer LCDs, including increased thickness, complexity, and cost. Layered constructions attenuate light and cause moiré. Attenuation is minimized using field sequential colour, requiring high-speed monochrome panels and strobed backlighting. While Bell et al. [2008] mitigate moiré with diffusers, our prototype does not incorporate such elements. Unlike the periodic viewing zones of parallax barriers and integral imaging [Dodgson, 2009], our design reproduces only the central zone and viewer tracking is necessary for wider fields of view. In contrast to attenuation layers, polarization field displays require detailed control of the polarization properties of LCDs since we assume panels can be operated as polarization rotators. Off-the-shelf LCDs deviate

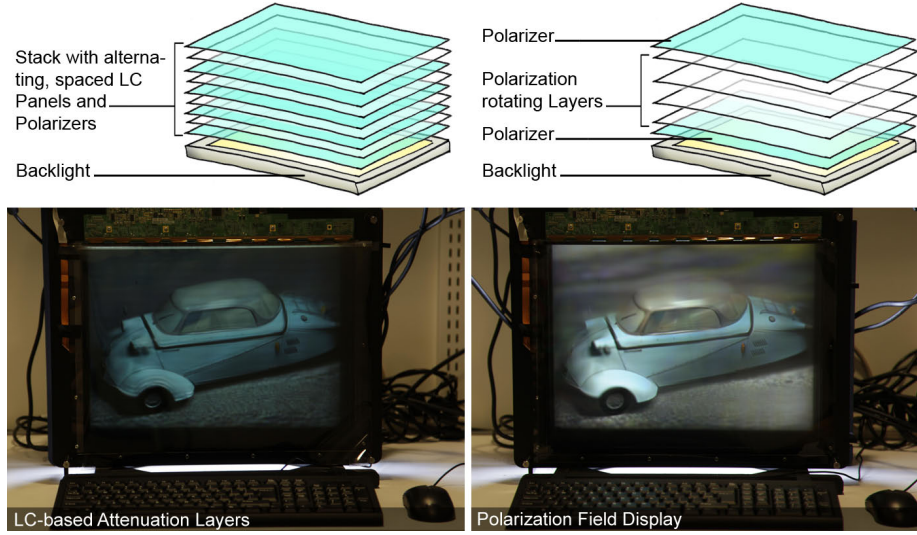


Figure 4.2: *Polarization-based vs. attenuation-based multi-layer LCDs. (Top, Left) An attenuation-based light field display requires stacking liquid crystal panels with polarizers between each layer. This construction effectively creates a programmable transparency stack. (Top, Right) Polarization-based light field displays improve optical efficiency using a single pair of crossed polarizers. (Bottom) Corresponding photographs of the prototype configured as an attenuation-based vs. polarization-based multi-layer LCD.*

from this model, particularly for oblique viewing, exhibiting partial attenuation for crossed linear polarizers and elliptical, rather than linear, polarization states [Yeh and Gu, 2009]. Commercial embodiments will require adopting optical models with increased fidelity, possibly complicating real-time control, or require engineering panels with the desired optical properties [Moreno et al., 2007].

4.2 Polarization Field Display

This section describes how to optimally construct polarization field displays using multi-layer LCDs to emit a target light field. First, we review conventional, single-layer LCD components and operation principles. Second, a general image formation model is described for polarization field displays, encompassing multi-layer LCDs as one embodiment. Under this model, each liquid crystal panel is considered as a spatially-controllable polarization rotator, and the entire set of panels is enclosed by a single pair of crossed linear polarizers. Third, we describe how to display dynamic light fields using polarization fields by adapting real-time tomographic algorithms to satisfy a least-squares optimality criterion.

4.2.1 Overview of LCDs

A liquid crystal display (LCD) consists of two primary components: a backlight and a spatial light modulator (SLM). The backlight is designed to produce uniform illumination, typically by conditioning the light produced by a cold cathode fluorescent lamp (CCFL) or light-emitting diode (LED) using a light guide and various diffusing and brightness-enhancing films. The spatial light modulator consists of a thin layer of liquid crystal, enclosed between glass sheets with embedded, two-dimensional electrode arrays. This stack is further enclosed by a pair of crossed linear polarizers.

When a voltage is applied across a given electrode pair (i.e., a single pixel), the orientation of the liquid crystal molecules reconfigures; this modifies the optical properties, inducing a rotation of the linear polarization state of light rays transversing this pixel [Yeh and Gu, 2009]. The transmitted intensity I is modeled by Malus' law:

$$I = I_0 \sin^2(\theta), \quad (4.1)$$

where I_0 is the intensity after passing through the first polarizer and θ is the angle of polarization after passing through the liquid crystal, defined relative to the axis of the first polarizer [Hecht, 2001]. Note that, conventionally, Malus' law is defined in terms of $\theta' = \theta - \pi/2$, the angle with respect to the axis of the second polarizer, such that $I = I_0 \cos^2(\theta')$. By controlling the voltages applied across the electrode array, an LCD renders two-dimensional images with varying shades of gray depending on the rotation induced within the liquid crystal. The rotation angle θ only must vary on the interval $[0, \pi/2]$ radians to reproduce all shades of gray—the range afforded by most commercial LCD panels, including widespread twisted nematic (TN) architectures. We note that this model only strictly applies for rays oriented perpendicular to the display surface. At oblique angles, light leakage occurs through crossed polarizers and birefringence of the liquid crystal produces elliptical, rather than linear, polarization states [Yeh and Gu, 2009]. However, as experimentally verified in Section 4.3, this model is a close approximation for the viewing angles considered in the prototype.

Two design alternatives enable colour LCDs: colour filter arrays and field sequential colour. In current LCDs, a colour filter array is deposited on the glass sheet closest to the viewer. Each pixel is divided into three subpixels by an array of filters with spectral transmittances corresponding to three colour primaries. This requires the resolution to be tripled along one display axis, increasing fabrication complexity and cost. Colour filter arrays also decrease brightness, typically to 30% of the backlight intensity. Rather than brightening the back-

light, which reduces power efficiency, field sequential colour (FSC) can be employed. With FSC, a strobed backlight successively illuminates a high-speed monochrome LCD with varying colour sources. If strobing occurs faster than the human flicker fusion threshold [Hart, 1987], a colour image is perceived. While yet to be widely commercially available, FSC LCDs are an active area of research [Chen et al., 2009; Stewart and Roach, 1994].

4.2.2 Modeling Multi-Layer LCDs

In this section we consider how multi-layer LCDs can be constructed to emit a four-dimensional light field, rather than a two-dimensional image. As shown in Figure 4.2, we consider the following architecture: a backlight covered by multiple, disjoint spatial light modulators. First, to maximize the optical efficiency, we assume field sequential colour illumination; this eliminates K layers of colour filters that would otherwise cause severe moiré [Bell et al., 2008] and brightness attenuation by a factor of approximately 0.3^K (e.g., 2.7% transmission for a three-layer LCD). Second, we observe that only two polarizing films are necessary, one on the top and bottom of the multi-layer stack. This creates a *polarization field display*, wherein each spatial light modulator consists of a liquid crystal layer functioning as a spatially-addressable, voltage-controlled polarization rotator.

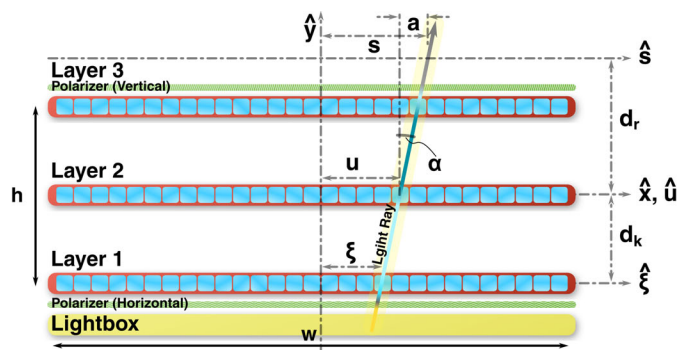


Figure 4.3: *Polarization field displays.* A K -layer display is constructed by separating multiple liquid crystal panels. The light field $l_0(u, a)$ emitted by the backlight is linearly polarized by the rear polarizer. The polarization state of ray (u, a) is rotated by $\phi_k(\xi)$ after passage through layer k , where $\xi = u + (d_k/d_r)a$. The emitted light field $\tilde{l}(u, a)$ is given by applying Equation 4.2 to the emitted polarization field $\tilde{\theta}(u, a)$ upon passage through the front polarizer.

Such displays must be controlled so the polarization field incident on the last polarizer accurately reproduces the target light field. In this section we present our analysis in flatland, considering 1D layers and 2D light fields, with a direct extension to 2D layers and 4D light fields. As shown in Figure 4.3, we consider a display of width w and height h , with

K layers distributed along the y -axis such that $d_k \in [-h/2, h/2]$. A two-plane light field parameterization $l(u, a)$ is used [Chai et al., 2000]. The u -axis is coincident with the x -axis and the slope of ray (u, a) is defined as $a = s - u = d_r \tan(\alpha)$, where the s -axis is a distance d_r from the u -axis.

The emitted light field $l(u, a)$ is given by applying Equation 4.1 to the polarization field $\theta(u, a)$ incident on the front polarizer:

$$l(u, a) = l_0(u, a) \sin^2(\theta(u, a)), \quad (4.2)$$

where $l_0(u, a)$ is the light field produced by the backlight after attenuation by the rear polarizer. The backlight is assumed to be uniform such that $l_0(u, a) = l_{max}$ and the light field is normalized such that $l(u, a) \in [0, l_{max}]$. This expression can be used to solve for the necessary target polarization field $\theta(u, a)$, as follows.

$$\theta(u, a) = \pm \sin^{-1} \left(\sqrt{\frac{l(u, a)}{l_0(u, a)}} \right) \bmod \pi \quad (4.3)$$

Under these assumptions, the principal value of the arcsine ranges over $[0, \pi/2]$. Note, with full generality, the target polarization field is multi-valued and periodic, since a rotation of $\pm\theta \bmod \pi$ radians will produce an identical intensity by application of Malus' law.

Each layer controls the spatially-varying polarization state rotation $\phi_k(\xi)$, as induced at point ξ along layer k . Ray (u, a) intersects the K layers, accumulating incremental rotations at each intersection, such that the emitted polarization field $\tilde{\theta}(u, a)$ is given by

$$\tilde{\theta}(u, a) = \sum_{k=1}^K \phi_k(u + (d_k/d_r)a). \quad (4.4)$$

Combining Equations 4.2 and 4.4 yields the following model for the light field $\tilde{l}(u, a)$ emitted by a K -layer polarization field display:

$$\tilde{l}(u, a) = l_0(u, a) \sin^2 \left(\sum_{k=1}^K \phi_k(u + (d_k/d_r)a) \right). \quad (4.5)$$

4.2.3 Synthesizing Polarization Fields

This section describes the optimization of multi-layer LCDs for polarization field display. We consider a discrete parameterization for which the emitted polarization field is represented



Figure 4.4: GPU-based SART allows real-time multi-layer optimization approaching the fidelity of the off-line solver. The first and second columns show different target views. Polarization-rotating layers are shown below each example. The off-line reference solver [Coleman and Li, 1996] produces sharp reconstructions (second row). A small number of SART iterations causes blurring (third row). Additional iterations converge to the reference (bottom row), with five iterations yielding similar quality (fourth row). Note that simulated views are shown, rather than prototype results.

as a column vector $\tilde{\theta}$ with M elements, each of which corresponds to the angle of polarization for a specific light field ray. Similarly, the polarization state rotations are represented as a column vector ϕ with N elements, each of which corresponds to a specific display pixel in a given layer. Under this parameterization, Equation 4.4 yields a linear model such that

$$\tilde{\theta}_m = \sum_{n=1}^N P_{mn} \phi_n, \quad (4.6)$$

where $\tilde{\theta}_m$ and ϕ_n denote ray m and pixel n of $\tilde{\theta}$ and ϕ , respectively. An element P_{mn} of the projection matrix \mathbf{P} is given by the normalized area of overlap between pixel n and ray m , occupying a finite region determined by the sample spacing.

An optimal set of polarization state rotations ϕ is found by solving the following constrained linear least-squares problem:

$$\arg \min_{\phi} \|\theta - \mathbf{P}\phi\|^2, \text{ for } \phi_{min} \leq \phi \leq \phi_{max}, \quad (4.7)$$

where each layer can apply a rotation ranging over $[\phi_{min}, \phi_{max}]$. Similar to Equation 3.11, Equation 4.7 can be solved using a sparse, constrained, large-scale trust region method [Coleman and Li, 1996]. However, we observe that this problem can be solved more efficiently by adapting the simultaneous algebraic reconstruction technique (SART). As proposed by Andersen and Kak [1984] and further described by Kak and Slaney [2001], SART provides an iterative solution wherein the estimate $\phi^{(q)}$ at iteration q is given by

$$\phi^{(q)} = \phi^{(q-1)} + \mathbf{v} \circ (\mathbf{P}^\top (\mathbf{w} \circ (\theta - \mathbf{P}\phi^{(q-1)}))), \quad (4.8)$$

where \circ denotes the Hadamard product for element-wise multiplication and elements of the \mathbf{w} and \mathbf{v} vectors are given by

$$w_m = \frac{1}{\sum_{n=1}^N P_{mn}} \quad \text{and} \quad v_n = \frac{1}{\sum_{m=1}^M P_{mn}}. \quad (4.9)$$

After each iteration, additional constraints on $\phi^{(q)}$ are enforced by clamping the result to the feasible rotation range. Building upon the Kaczmarz method for solving linear systems of equations [Kaczmarz, 1937], SART is shown to rapidly converge to a solution approaching the fidelity of that produced by alternative iterative methods, including trust region and conjugate gradient descent techniques [Kak and Slaney, 2001] (see Figure 4.4). Section 4.3 shows that SART allows for real-time optimization for interactive polarization field displays.

In summary, polarization fields present both an optically and computationally efficient

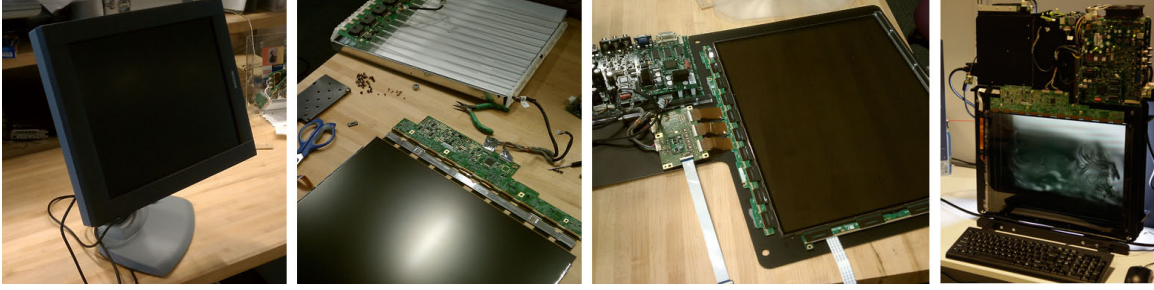


Figure 4.5: *Constructing the polarization field display prototype. Four monochrome LCDs were modified by Matt Hirsch at the MIT Media Lab to create a single multi-layer LCD. Photographs depict from left to right: an unmodified Barco E-2320 PA LCD, the liquid crystal panel and backlight after removing the case and power supply, a modified panel mounted on an aluminum frame, and the assembled prototype.*

architecture for dynamic light field display using multi-layer LCDs. We briefly contrast this architecture to that required for a direct extension of the attenuation-based method proposed in chapter 3. As shown in Figure 4.2, a multi-layered, attenuation-based display is fabricated by placing a polarizer on the backlight and additional polarizers after each liquid crystal layer, effectively creating a set of dynamically-programmable transparencies; however, such a design reduces the display brightness by a factor of 0.8^{K-2} compared to the proposed polarization field display, assuming a maximal transmission of 80% through each polarizer (as measured for those used in the prototype). Yet, we observe an adaptation of SART can similarly be applied to attenuation layers by substituting the logarithm of the emitted light field intensity \tilde{l}_m and the logarithm of the transmittance t_n for $\tilde{\theta}_m$ and ϕ_n in Equation 4.6, respectively; thus, we provide the first implementation for achieving interactive frame rates with such designs.

4.3 Implementation

This section describes the construction and performance of the prototype. First, we summarize the modifications made to commercial LCD panels to create a reconfigurable multi-layer display. Second, we review the off-line and real-time software for light field rendering, antialiasing, and optimizing layer patterns. Third, we assess the prototype, verifying the image formation model and illustrating the practical benefits and limitations of polarization field displays.

4.3.1 Hardware

Given that we require monochrome layers and field sequential colour, a custom prototype was necessary. This prototype was constructed by Matt Hirsch at the MIT Media Lab as part of a collaboration with the author at UBC.

PureDepth [Bell et al., 2008] offers dual-layer LCDs, but no supplier was found for multi-layer configurations. Each layer of the prototype consists of a modified Barco E-2320 PA LCD, supporting 1600×1200 8-bit grayscale display at 60 Hz, and an active area of 40.8×30.6 cm. As shown in Figure 4.5, the liquid crystal layer was separated from the case, backlight, and power supply. Polarizing films were removed and the adhesive was dissolved with acetone. By design, the driver board is folded behind the panel, blocking a portion of the display when used in a stacked configuration. An extended ribbon cable was constructed to allow the board to be folded above the display using a pair of 20-pin connectors and a flat flexible cable. The exposed panel, driver boards, and power supply were mounted to a waterjet-cut aluminum frame. Four such panels were constructed and stacked on a wooden stand. Arbitrary layer spacings are supported by translating the frames along rails. Acrylic spacers hold the layers at a fixed spacing of 1.7 cm for all experiments described in this chapter, yielding a total display thickness of 5.1 cm. The prototype is illuminated using an interleaved pair of backlights and controlled by a 3.4 GHz Intel Core i7 workstation with 4 GB of RAM. A four-head NVIDIA Quadro NVS 450 graphics card synchronizes the displays. Additional documentation of the prototype construction is included in Appendix A.

As shown in Figure 4.2, the display operates in either attenuation-based or polarization-based modes. The original polarizers were discarded and replaced with American Polarizers AP38-006T linear polarizers. By specification, a single polarizer has a transmission efficiency of 38% for unpolarized illumination. Transmission is reduced to 30% through a pair of aligned polarizers, yielding an efficiency of 80% for polarized light passing through a single, aligned polarizer. Five polarizers are required for attenuation-based display, with a pair of crossed polarizers on the rear layer followed by successively-crossed polarizers on each remaining layer. A polarization field display is implemented by enclosing the stack by a single pair of crossed polarizers. Field sequential colour is simulated, for still imagery, by combining three photographs taken while alternating the colour channel displayed on each layer. To assist registration, examples in this paper use the colour filters included in the Bayer mosaic of the camera.

Each panel must be radiometrically calibrated to allow an accurate mapping from optimized rotation angles to displayed image values. The Barco E-2320 PA is intended for medical diagnostic imaging and replicates the DICOM Grayscale Standard Display Function [DI-

COM, 2008]. The normalized displayed intensity $\bar{I} \in [0, 1]$ was measured as a function of the 8-bit image value $v \in [0, 255]$ using a photometer held against an unmodified panel. The resulting radiometric response curve is approximated by a gamma value of $\gamma = 3.5$ such that $\bar{I} = (v/255)^\gamma$. Thus, gamma compression is applied to map optimized pixel transmittances to image values when operating in the attenuation-based mode. When operated as a polarization field display, optimization yields the polarization state rotation ϕ for each pixel. For an unmodified panel we model this mapping using Equation 4.1 such that $\bar{I} = \sin^2(\phi)$. Equating this expression with the gamma curve yields the following mapping between rotations and image values.

$$v(\phi) = \lfloor 255 \sin^{2/\gamma}(\phi) + 0.5 \rfloor \quad (4.10)$$

Figures 4.1 and 4.6 compare target light field views to corresponding photographs of the prototype. Figure 4.2 compares the attenuation-based mode to the polarization-based mode. Videos confirming smooth motion parallax are included in the supplementary video.



Figure 4.6: Polarization field display using the multi-layer prototype. The central views (perpendicular to the display) for the “Buddha”, “dice”, “car”, and “red dragon” scenes. Target views are compared to photographs of the prototype on the top and bottom, respectively. Colour imagery is obtained by combining three photographs taken while varying the colour channel displayed by the monochrome panels. Section 4.4 quantitatively assesses performance and artifacts.

4.3.2 Software

The light fields in this chapter are rendered with a spatial resolution of 512×384 pixels and depict 3D scenes with both horizontal and vertical parallax from 7×7 viewpoints within a field of view of 10 degrees. POV-Ray is used to render the scenes shown in Figure 4.6.

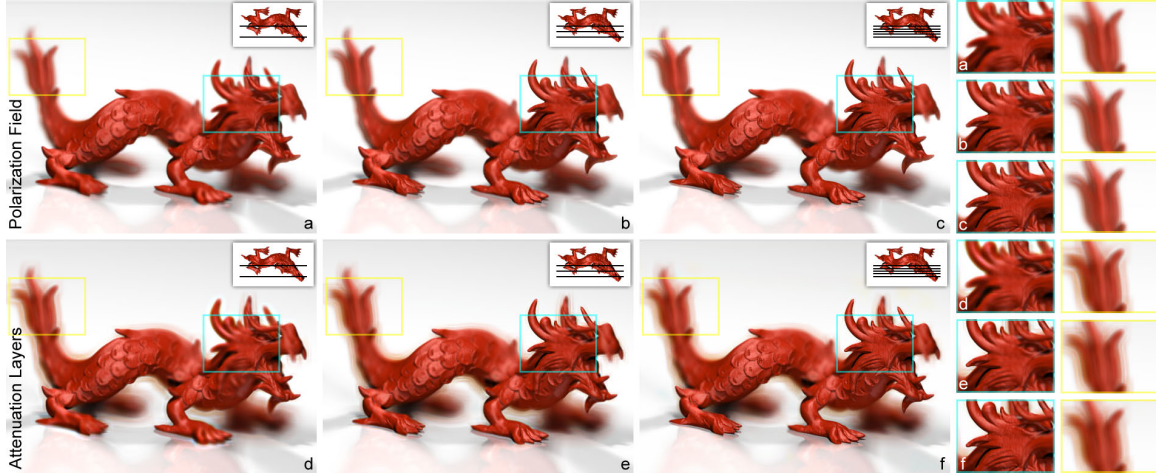


Figure 4.7: *Simulated light field reconstructions using polarization fields (top row) and attenuation layers (bottom row) are shown for two, three, and five layers from left to right. Layer positions with respect to the scene are illustrated in the insets. Note that the reconstruction fidelity of objects within and outside the physical display extent increases for a larger number of layers, as highlighted by the cyan and yellow regions, respectively. Due to bias in the least-squares solution for a log-domain objective, optimized tomographic reconstructions for attenuation-based displays suffer from halo artifacts around high-contrast edges, which is not the case for polarization field displays.*

Following Levoy and Hanrahan [1996] and Zwicker et al. [2006], we apply a 4D antialiasing filter to the light fields by rendering each view with a limited depth of field. As analyzed in Section 3.3, this antialiasing filter simultaneously approximates the limited depth of field established for multi-layer light field displays.

The Matlab LSQLIN solver serves as the reference solution to Equation 4.7, implementing a sparse, constrained, large-scale trust region method [Coleman and Li, 1996]. This solver converges in about 8 to 14 iterations for three to five attenuating or polarization-rotating layers. Solutions are found within approximately 10 minutes on the previously-described Intel Core i7 workstation.

The SART algorithm given by Equation 4.8 is implemented in Matlab and on the GPU. We observe SART is well-suited for parallel processing on programmable GPUs [Keck et al., 2009]. Our code is programmed in C++, OpenGL, and Cg. Light fields are rendered and antialiased in real-time using OpenGL, followed by several iterations of the GPU-based SART implementation. We achieve refresh rates of up to 24 frames per second using one iteration for four layers running on the NVIDIA Quadro NVS 450 (see Fig. 4.9). Figure 4.4 illustrates SART convergence, demonstrating that 2 to 5 iterations minimize reconstruction artifacts. Estimates for the previous frame seed the optimization for the current frame. For

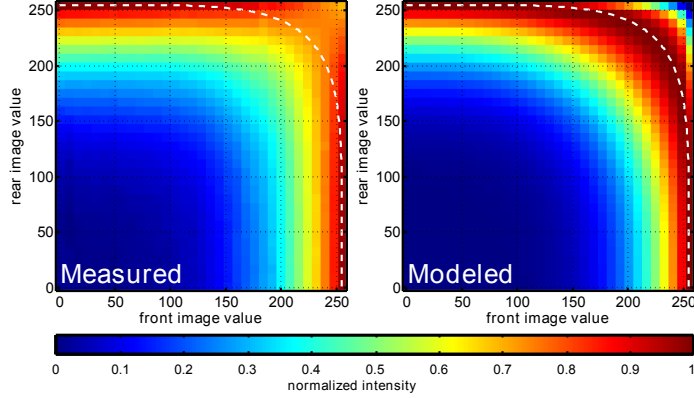


Figure 4.8: Radiometric calibration of the prototype. The measured (left) and modeled (right) normalized intensity \bar{I} is plotted as image values v_1 and v_2 are displayed on the rear and front layer, respectively. The model is a least-squares fit of Equation 4.11 to the measured intensities. Note that the prototype only uses rotations corresponding to values located on the lower left of the white lines.

static scenes, this effectively implements an increasing number of SART iterations over time, while providing a suitable initialization for successive frames in a dynamic environment. Pseudocode for the GPU implementation are included in Appendix A.

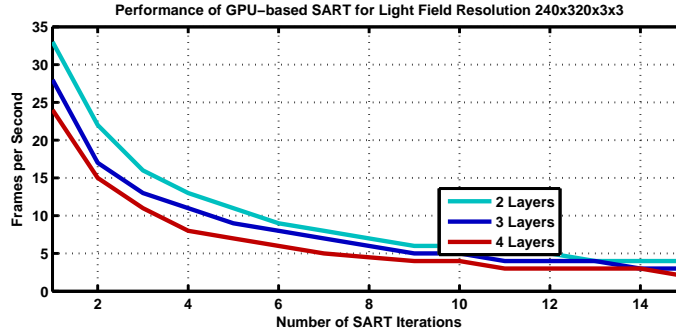


Figure 4.9: Performance of the GPU-based SART implementation as a function of varying numbers of SART iterations (per frame) and varying numbers of polarization-rotating or light-attenuating layers. The light field resolution is 320×240 spatial samples and 3×3 angular samples; layers have a similar spatial resolution. Timings are measured using the LCD-based prototype system described in the text.

4.4 Assessment

As shown in Figure 4.1, polarization fields accurately depict multiple perspectives of the “Buddha” scene. Note that variations in viewpoint capture highlights on the incense burner

and occlusions of the background characters. Figure 4.6 demonstrates faithful reproduction of translucency in the “dice” scene and through the windows in the “car” scene. Detailed analysis for each scene is included in Appendix A. Videos demonstrating motion parallax are reproduced in the supplemental video.

While confirming the prototype achieves automultiscopic display, photographs and videos exhibit artifacts not predicted by simulations. Moiré is present, although it could be mitigated using the method of Bell et al. [2008]. A colour cast also appears in the background of the “dice” and “dragon” scenes. We attribute this to the discrepancies between our prototype and the ideal construction composed of polarization-rotating layers. As characterized by Yeh and Gu [2009], most commercial panels do not operate as ideal two-dimensional polarization rotators, particularly at oblique angles. To this end, we used photometric measurements to assess the validity of our model. As shown in Figure 4.8, a photometer measured the normalized intensity \bar{I} as differing image values v_1 and v_2 were displayed on the rear and front layer, respectively. Substituting Equation 4.10 into Equation 4.5 yields the following prediction.

$$\bar{I}(v_1, v_2) = \sin^2 \left\{ \sin^{-1} \left[\left(\frac{v_1}{255} \right)^{\frac{\gamma}{2}} \right] + \sin^{-1} \left[\left(\frac{v_2}{255} \right)^{\frac{\gamma}{2}} \right] \right\} \quad (4.11)$$

We note that measured intensities are nearly identical upon interchanging v_1 and v_2 , validating the additive model in Equation 4.4—upon which the proposed tomographic optimization relies. However, we observe the measured contrast is limited when v_1 and v_2 are large. This is confirmed in the supplemental video; overlaying a pair of white images produces a darker image, but with reduced contrast. Thus, artifacts persist in the prototype due to the deviation of our off-the-shelf panels from ideal polarization rotators.

As shown in Figure 4.7, polarization fields perform comparably to attenuation layers in terms of reconstruction fidelity. Yet, halo artifacts are noticeably reduced using polarization fields. We attribute this primarily to different biases introduced by least-squares optimization of transformed objective functions. As proposed by Gotoda [2010] and in Chapter 3, attenuation-based displays optimize an object function, reminiscent of Equation 4.7, defined for the logarithm of the target intensities. This inherently penalizes artifacts in dark regions, leading to the observed halos. By comparison, polarization fields optimize an objective function defined for target intensities transformed by Equation 4.3; this transformation is more linear than for attenuation, thereby mitigating halos. This trend is confirmed by the average peak signal-to-noise ratio (PSNR) plots shown in Figure 4.10, in which polarization fields slightly outperform attenuation layers. Based on these trials, we conclude that polarization fields present an optically-efficient alternative to attenuation layers optimally-

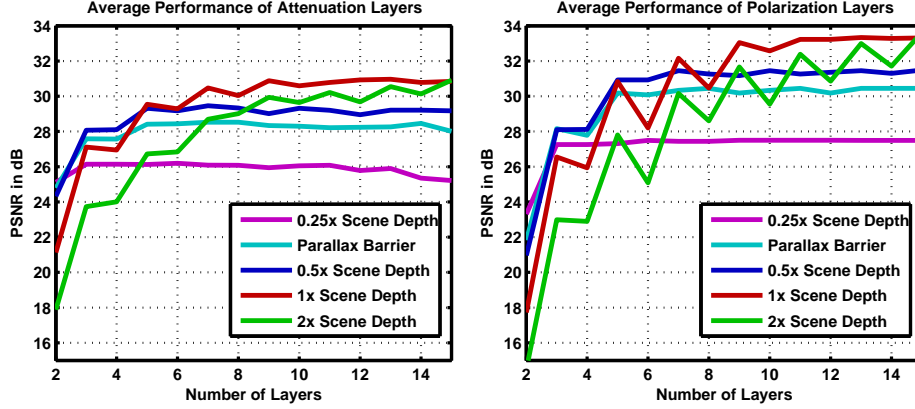


Figure 4.10: Average PSNR for attenuation layers vs. polarization fields. The PSNR was averaged for the four scenes in Figure 4.6 (and for two more in the video) depending on the number of layers and the relative display thickness. Note that polarization fields can accurately present objects beyond the display, but can also be operated in a volumetric mode enclosing the scene for reduced errors.

suiting to multi-layer LCDs, closely mirroring the PSNR trends and dependence on the layer numbers and display thickness previously established for attenuation-based displays.

4.5 Discussion

4.5.1 Benefits and Limitations

The benefits and limitations of polarization fields were previously outlined in Section 4.1.1. Here we briefly highlight additional insights drawn from the prototype. Polarization fields have several notable benefits over existing automultiscopic displays, particularly those supporting relatively thin form factors. Foremost, polarization fields are shown to compare favorably to closely-related attenuation layers. Yet, certain practical advantages are afforded by this construction, including increased optical efficiency and reduced reconstruction artifacts. Both methods significantly improve upon the dominant commercial alternatives, being parallax barriers and integral imaging, presenting imagery with greater spatial resolution, increased brightness, and extended depth of field.

Polarization fields inherit many of the limitations of other multi-layer displays. Foremost, performance improves by adding layers and increasing the display thickness, raising costs while restricting the scope for mobile applications. Layered constructions must introduce additional elements to mitigate moiré, scattering, and reflections. By relying on the polariza-

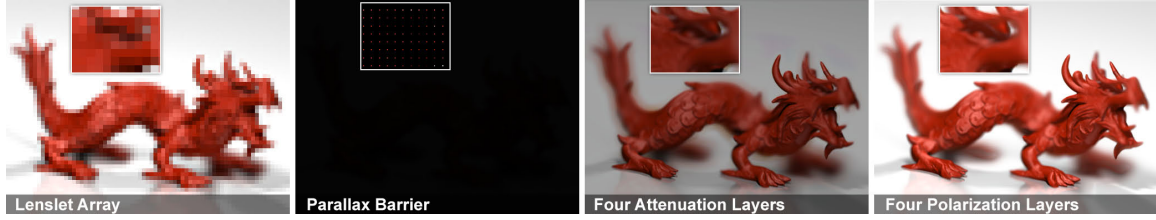


Figure 4.11: *Benefits of multi-layer displays. Simulated views of the “red dragon” are shown, from left to right, using integral imaging, parallax barriers as well as a multi-layer LCD display using attenuating layers, and polarization-rotating layers. The displayed light field contains 7×7 views, reducing the spatial resolution for integral imaging and parallax barriers by a similar amount. Compared to parallax barriers, polarization field displays are 48 times brighter. By decreasing the number of necessary polarizers, each reducing the total light transmission, in an LCD-based multi-layer display, polarization field displays have a better light efficiency than the same number of LCD-based attenuation layers.*

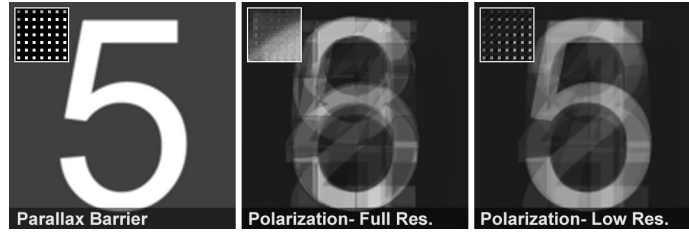


Figure 4.12: *Uncorrelated views with light field displays. (Left) Parallax barriers allow independent images to be projected in each direction, here corresponding to different Arabic numerals. (Middle) Similar to attenuation layers, polarization fields exploit correlated imagery and significant crosstalk is produced otherwise. (Right) Unlike attenuation layers, constraining only the subset of pixels used by a similar parallax barrier does not resolve this limitation.*

tion properties of LCDs, polarization fields require detailed optical models or, alternatively, engineering to produce two-dimensional polarization rotators. Perhaps most significantly, polarization fields converge to a moderate PSNR, even with the benefit of many layers and a large display enclosure. This indicates that such layered constructions have limited degrees of freedom. To this end, Figure 4.12 considers the performance for light fields for which neighboring views are independent. In contrast to parallax barriers and integral imaging, such imagery is not accurately rendered using polarization fields; thus, such displays share a key limitation with attenuation layers: requiring correlated views such as those that originate from natural scenes.

However, even the degree of correlation between different perspectives of a natural light field depends on many parameters. View-dependent global illumination effects, transparencies, and field of view (FOV) are only a few such variables. Figure 4.13 evaluates the

reconstruction performance of both attenuation-based and polarization-based multi-layer displays for an increasing field of view in the desired light field. Two different perspectives for fields of view of 10° , 20° , and 45° are shown. While the similarities between different viewpoints are fairly high in narrower fields of view, they drop, along with the reconstruction performance, for an increasing FOV. In addition to solutions to the constrained, linear formulations discussed in Sections 3.2.2 and 4.2.3, Figure 4.13 also shows a constrained, nonlinear solution to the periodic polarization field synthesis problem (Eq 4.2), which is derived in Appendix A.4.2. The nonlinear solution is computed with MATLAB’s iterative LSQNONLIN solver, with 30 iterations, where the linear solution serves as an initial guess. The display parameters for this experiment correspond to the multi-layer prototype introduced in Section 4.3.1.

4.5.2 Future Work

Additional engineering efforts are required to address prototype limitations. First, the proposed implementation only simulates field sequential colour; construction of a strobed LED backlight combined with high-speed monochrome LCDs would be necessary for dynamic colour imagery. Second, laboratory measurement of the Jones matrices characterizing the display panels [Ma et al., 2010; Moreno et al., 2003], together with a modified image formation model, would likely minimize artifacts. Third, inclusion of custom holographic diffusers could mitigate moiré and the addition of Fresnel lenses (as proposed by Gotoda [2011]) may extend the depth of field. Finally, the LCD panels can be replaced with displays that behave as polarization rotators or, alternatively, additional optical elements may be added to produce a similar result [Moreno et al., 2007].

Beyond these engineering efforts, several promising theoretical opportunities exist. A frequency-domain analysis of polarization fields, following Zwicker et al. [2006], may yield an analytic depth of field expression. A preliminary analysis, performed by evaluating the Fourier transform of Equation 4.5, indicates spatio-angular frequencies are produced well beyond the region supported by competing parallax barriers and integral imaging constructions. Given the dependence on correlated imagery and the limited degrees of freedom afforded by increasing layers, such performance assessment may be further supported by considering priors on natural light fields. Alternatively, to expand the degrees of freedom, time-multiplexed, multi-layer decompositions may be possible, although necessitating higher-speed panels and likely requiring alternate methods for real-time optimization.

Perhaps the most intriguing direction of future work is to re-examine the full potential afforded by the multi-valued, periodic target polarization field given by Equation 4.3. The



Figure 4.13: An increasing field of view of the desired light field decreases the correlation between different perspectives in a light field, thereby decreasing the reconstruction quality. Columns 1–4 show two views of the original light field, the simulated reconstruction of attenuating layers, the simulated reconstruction using linear polarization field synthesis (Sec. 4.2.3), and a nonlinear solution to the periodic polarization field problem (Eq. 4.2).

implementation presented in this chapter considers the principal value of this expression, limiting the target polarization field to $\theta(u, a) \in [0, \pi/2]$. If this restriction is lifted, additional degrees of freedom appear accessible; for example, larger rotations can *decrease* the intensity of an emitted ray via application of Malus' law. Similarly, incorporating panels that can apply positive and negative rotations over the full range such that $\phi_k(\xi) \in [-\pi, \pi]$ will likely increase reconstruction fidelity and potentially enable efficient, unconstrained optimization methods.

4.5.3 Conclusion

Polarization field displays are designed to maximize the optical and computational performance achieved when using multi-layered LCDs for automultiscopic display. Such displays eschew the long-standing trends of refractive, reflective, and attenuation-based architectures, instead focusing on the novel optical properties exhibited by stacked polarization rotators. In this chapter we establish the potential of such designs, as well as promising avenues of future research that may more fully exploit their potential. As LCDs have become the dominant spatial light modulator employed in consumer displays, it is our hope that polarization fields will inspire others to investigate their full potential for automultiscopic display.

Chapter 5

Adaptive Coded Aperture Projection

In this chapter, we present solutions for depth-of-field (DOF) enhanced projection through combined optical light modulation and computational processing. Coding a projector's aperture plane with dynamic, adaptive patterns together with inverse filtering allows the depth-of-field of projected imagery to be increased. We show how these patterns can be computed at interactive rates, by taking into account the image content and limitations of the human visual system. Applications such as projector defocus compensation, high quality projector de-pixelation, and increased temporal contrast of projected video sequences are supported.

5.1 Introduction

Modern video projectors are remarkable devices that can display large imagery with high resolution, brightness, and contrast. The latest high-end models even incorporate auto-focus and auto-iris objective lenses. These can enhance the temporal contrast of projected images by adjusting the aperture to the average brightness of the displayed content. Their flexibility and low cost make projectors irreplaceable for many applications including professional presentations, home entertainment, scientific visualization, as well as museum and art installations. We envision future generations of these displays as fully integrated systems with cameras, dynamically adjustable apertures and intelligent control mechanisms.

In this chapter, we present solutions for taking projectors to the next level. By placing coded masks at a projector's aperture plane, we show how the depth-of-field (DOF) of a projection can be enhanced. This allows focused imagery to be shown on complex screens with varying distances to the projector's focal plane, such as projection domes or cylindrical canvases. We demonstrate, for instance in Figure 5.1, that integrated broadband masks and adaptive dynamic apertures outperform previous methods of defocus compensation for circular aperture stops. In addition, our dynamic apertures can perform the type of contrast enhancement employed by common auto-iris projection lenses. Furthermore, they

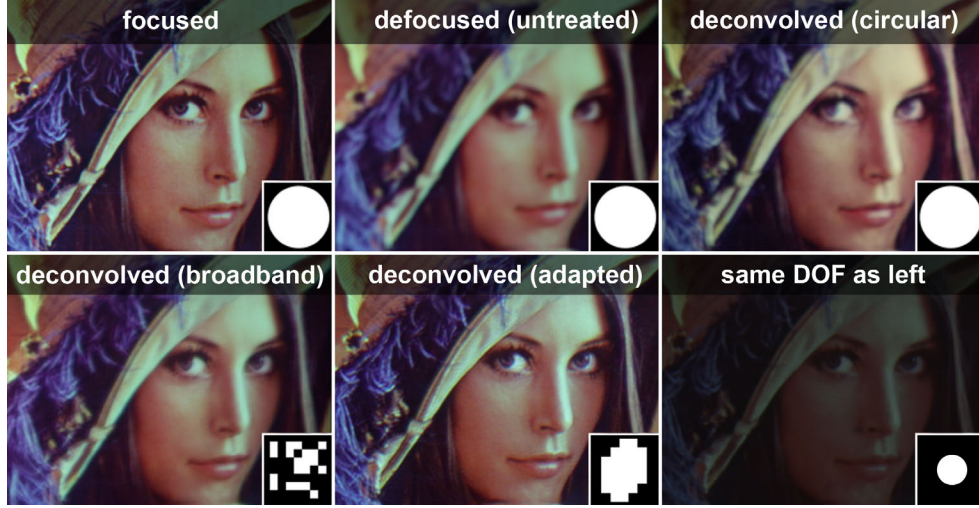


Figure 5.1: Displayed imagery can be pre-compensated to enhance a projector’s depth-of-field, but the resulting quality depends on the employed aperture shape. A projected image in focus (upper left), and with the same optical defocus (screen located at 2m distance to focal plane) modified in several different ways: with a circular aperture —untreated and deconvolved with corresponding PSF, compensated with static broadband coded aperture and with dynamic coded aperture. The result of the dynamic aperture is optimized for a viewing distance of at least 50cm and for an image diameter of 10cm where no artifacts should be visible. The deconvolved images are computed from the corresponding apertures (shown in the sub-figures). While all other images are adjusted to a similar brightness to enable a better comparison of focus, the lower right image is captured using the same exposure as the image to its left. A small circular aperture has been applied that achieves the same depth-of-field as with the adaptive coded aperture at the cost of a significant loss of brightness. Note that the first four and the last two images were displayed with two different projector prototypes (see Figure 5.4), which have slightly different colour gamuts and intensity ranges.

enable high-quality de-pixelated projections that are beneficial for rear-projection TV sets and other close-view displays.

5.2 Coded Aperture Projection Principle

Common projection displays have large aperture stops to optimize light transmission, thus narrow depths-of-field. When displaying imagery onto non-planar canvases or when a fronto-parallel display position is not feasible, projected content will be blurred due to optical defocus. This is often described as a convolution of the original image with a filter kernel that corresponds to the aperture shape of the imaging device (PSF, Figure 5.2 left). The scale of the kernel is directly proportional to the degree of defocus:

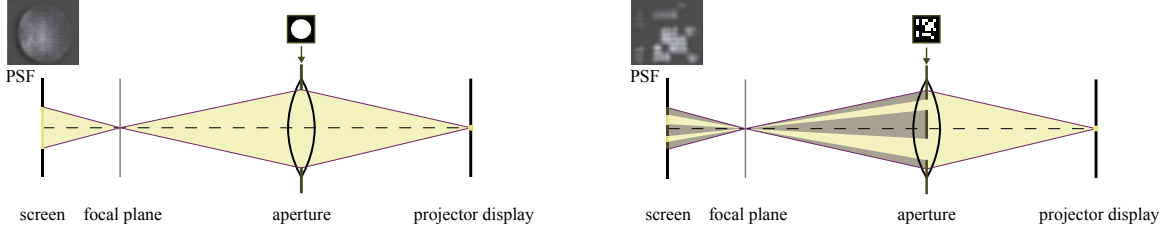


Figure 5.2: *Different aperture patterns lead to different point spread functions. A defocused projection of a point results in an image of the utilized aperture, as shown in case of a circular aperture (left) and a coded aperture (right).*

$$i_p = k_s \otimes i_d, \quad (5.1)$$

where i_d is the displayed image, k_s the aperture kernel at scale s , and i_p the optically blurred projection. Deconvolution, that is convolving an image with the inverse aperture kernel, will digitally sharpen an image and consequently compensate optical defocus:

$$i_d = k_s^{-1} \otimes i. \quad (5.2)$$

Here k_s^{-1} is the inverse aperture kernel, and i the original desired image. Convolution and deconvolution can be modeled easier in the frequency domain, rather than in the spatial domain, where a convolution corresponds to a multiplication $I_p = K_s \cdot I_d$ and deconvolution is equivalent to a division $I_d = I/K_s$. I , I_d , I_p , and K_s are the Fourier transforms of i , i_d , i_p , and k_s respectively. Deconvolution is an ill-posed problem and performing it directly through division in frequency space is generally avoided. Zeros in the Fourier representation of the convolution kernel and image noise often lead to ringing artifacts. For projector defocus compensation, however, the images are not measurements, but noiseless digital footage. This makes deconvolution by direct inverse filtering feasible if the kernel is broad-band, i.e. has no zeros in its Fourier transform. Instead of Wiener deconvolution, we apply a regularized inverse filter to compute a displayed compensation image I_d as

$$I_d = \frac{K_s^*(f_x, f_y)}{|K_s(f_x, f_y)|^2 + \alpha |L(f_x, f_y)|^2} I(f_x, f_y), \quad (5.3)$$

where $L(f_x, f_y)$ is the Fourier transform of the discrete Laplacian operator and $\alpha = 0.01$ is a user-defined smoothing parameter. The regularization term $\alpha |L(f_x, f_y)|^2$ prevents ringing artifacts by smoothing out any possible division by zero for higher frequencies. If chosen

too high, however, the regularizer prevents any significant sharpening. Intuitively, it can be understood as the signal-to-noise ratio term that is employed in a similar fashion for Wiener deconvolution of noisy camera images.

The deconvolution is computed in the frequency domain, for multiple kernel scales s , which are individually transformed back into the spatial domain. The resulting compensation images are fused into I_d so that each image region is deconvolved with the corresponding amount of measured defocus (see Bimber and Emmerling [2006] for more details). Thus, the larger the desired depth-of-field, the more scales are necessary.

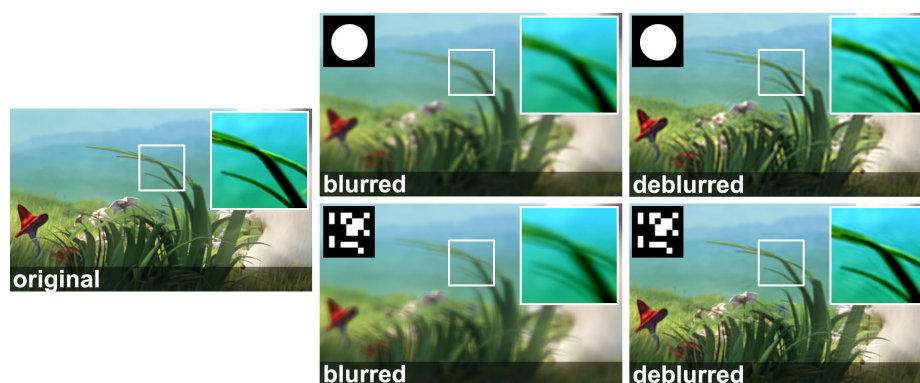


Figure 5.3: *Deconvolution with different kernels. Blurring an image (left) can, depending on the kernel, filter out important frequencies in the blurred result (center column), which often results in ringing artifacts after a deconvolution (upper right). A broad-band kernel can significantly improve the quality of the deblurred result (lower right). The close-up regions are contrast enhanced for improved visibility.*

The main limitation of previous projector defocus compensation approaches are ringing artifacts in the pre-filtered images. These are mainly due to the aforementioned divisions by zeros introduced by circular projector apertures. The resulting PSFs act as low pass filters on projected light by irreversibly canceling out high spatial image frequencies. The loss of these frequencies and resulting ringing in the compensation images is depicted in Figure 5.3 (upper row). Decreasing the size of the aperture opening reduces the number of low Fourier magnitudes, thus it effectively enhances the depth-of-field of a projection system. Using narrower aperture openings (up to pinhole size), however, will naturally decrease the light throughput significantly, which is unacceptable for most projection-based displays.

In order to reduce this problem, static coded apertures have been integrated into the objective compound of common projectors [Grosse and Bimber, 2008]. The aperture pattern influences how efficient high frequencies and light throughput can be retained. If this aperture is more broadband in the frequency domain and its Fourier transform has initially fewer

low magnitudes than a circular aperture with the same light throughput, more frequencies are retained and more image details can be reconstructed with less ringing artifacts.

However, one limitation of static coded apertures is that they reduce the light throughput by a constant factor which is independent of the actual image content. Adapting the aperture dynamically with respect to the input allows us to optimize light throughput and deconvolution quality for each projected frame individually. The quality difference to static aperture codes is compared in Figure 5.1.

In the remainder of this chapter, we explore the use of coded apertures that optimize a projector’s depth-of-field without the need for additional refractive lens elements. We show how dynamically adjusted attenuation masks in the projector’s aperture plane can preserve sharp image features with little ringing artifacts for defocused projections while preserving a high light transmission. The employed apertures are either designed to be broadband in the frequency domain or optimized to the frequency band of the current image. They comprise higher Fourier magnitudes than circular apertures with the same light throughput.

5.3 Prototype Designs

In order to demonstrate the feasibility of depth-of-field enhancement through coded projector apertures two prototypes were built by Max Grosse at the Bauhaus University in Weimar, Germany. The first prototype utilizes a static attenuation mask whereas a programmable liquid crystal display was integrated into a second prototype to experiment with adaptive aperture patterns.

5.3.1 Static Broadband Aperture

In the first prototype, a static attenuation mask is inserted into the aperture plane of the objective lens, as shown in Figure 5.4 (left). The projector is a Sony VPL-CX80 XGA 3-chip LCD projector and the applied aperture pattern was described by Veeraraghavan et al. [2007] as the 7x7 binary broadband mask that maximizes the Fourier-magnitudes of the zero padded convolution kernel.

Although photographic film is an obvious option for producing high-contrast aperture patterns, initial experiments show that it does not resist the heat generated in the projector. Hence, the codes are printed on transparencies, which prove more stable under varying temperatures. In a commercial display the masks could be manufactured from any heat re-



Figure 5.4: *Two prototypes fabricated by Max Grosse at the Bauhaus University Weimar: a static attenuation mask integrated into the aperture plane of a projector improves digital defocus compensation through inverse filtering (left). Replacing the aperture with a transparent liquid crystal array allows dynamic attenuation mask patterns to be encoded (right). We compute the adaptive patterns by optimizing for light throughput while preserving spatial frequencies that are perceivable by a human observer. Both approaches significantly improve depth-of-field through inverse light filtering when compared to circular aperture stops. While static coded apertures are easier to realize and less expensive, adaptive coded apertures are more efficient.*

sisting material, thus, providing an even higher contrast. They represent a low cost solution for coded aperture projection.

5.3.2 Programmable Aperture

Adaptive coded apertures are implemented by integrating a programmable liquid crystal array (LCA) into the projector’s aperture plane, as illustrated in Figure 5.4 (right). The LCA is an Electronic Assembly EA DOGM132W-5 display with a resolution of 132x32, 20Hz frame rate, and $0.4 \times 0.35 \text{ mm}^2$ pixel size. It is controlled via USB with an Atmel ATmega88 microcontroller. The projector is a BenQ 7765PA XGA (DLP).

5.4 Defocus Compensation with Static Coded Apertures

In this section we show results of depth-of-field enhancement through inverse light filtering with static coded apertures. All of the computations are performed on programmable graphics hardware and run at interactive framerates. As the author was not directly involved in the development of this particular part of the coded aperture projection project, implementation details are skipped; the interested reader is referred to [Grosse et al., 2009]. A basic understanding of defocus compensation with static apertures, however, is required

for the remainder of the chapter. Therefore, a general outline is given in the following.

5.4.1 Defocus Estimation and Data Preprocessing

The proposed defocus compensation approach relies on knowledge of the amount of defocus at each projector pixel, which directly corresponds to the scale of the convolution kernel or PSF at this point. We measure the spatially varying defocus by projecting structured light patterns and capturing them with a camera as discussed by Bimber and Emmerling [2006]. This is a one-time calibration step and takes less than a minute. Given the kernel scales for each pixel, we compute a non-uniform sub-division that partitions the projection images into regions with the same amount of defocus and, consequently, kernel scales. This image sub-division is constant for static settings and is therefore pre-computed along with the properly scaled inverse PSFs. As outlined in Section 5.2, a compensation can then be performed at runtime by image deconvolution with the appropriately scaled PSF.

5.4.2 Real-Time Compensation on the GPU

In each frame, the image needs to be transformed into the Fourier domain, where a compensation is performed according to Equation 5.3 for each defocus level. The result is transformed back into the spatial domain and displayed. Instead of performing the compensation on each colour channel separately, we apply the deconvolution only to the luminance channel. After compensation, the luminance channel and the original chrominance data are merged.

5.4.3 Static Aperture Results

Figure 5.5 compares the Fourier magnitudes of the chosen binary broad-band pattern [Veeraraghavan et al., 2007] and a circular mask with the same light transmission for different PSF scales. The de facto opening for both apertures is the same. In this plot, the scaled patterns are zero-padded to original image resolution, and show the amount of resulting Fourier magnitudes below 7% of the maximum value. The coded aperture outperforms the circular mask in all scales.

With our current prototype, we achieve 10-20 fps for XGA resolution and 1-14 scales (see Figure 5.5). Upcoming graphics hardware will clearly make real-time performance possible for a large number of scales. Figures 5.1 and 5.6 show examples of depth-of-field enhanced projection using the static mask. As seen in Figure 5.6, the optical defocus caused by

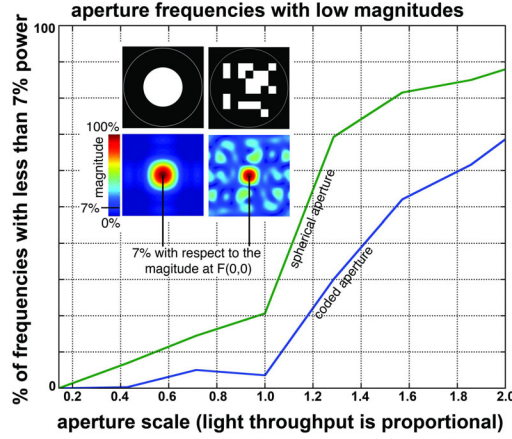


Figure 5.5: Low Fourier magnitudes cause ringing after inverse filtering. The power spectrum of a broadband code is always higher than that of a Gaussian PSF.

a projection onto a spherical screen can be compensated. Further evaluation of static apertures is presented in the course of a comparison with adaptive apertures in Section 5.6.



Figure 5.6: Projection onto a hemispherical dome (five different blur levels) focused at the outer frame (left). Defocus with spherical aperture at inner close-up section (center), and focus improvements with our static aperture mask (right).

One limitation of static coded apertures is that they reduce the light throughput by a constant factor which is independent of the actual image content. Adapting the aperture dynamically with respect to the input allows us to optimize light throughput and deconvolution quality for each projected frame as described in the following section.

5.5 Defocus Compensation with Adaptive Coded Apertures

As explained above, increasing the depth-of-field with a static coded broadband aperture comes at the cost of decreased light transmission, which is one of the most crucial aspects of all projector-based display systems.

In this section, we present a generalization of the techniques discussed in Section 5.4 to dynamic adaptive aperture patterns. Therefore, the data pre-processing including image segmentation and compensation via deconvolution can be performed as explained in the previous section. We now show how dynamic apertures can be computed based on an analysis of the projected image content. This analysis employs an intuitive model of the human visual system (HVS) and allows us to determine and filter out spatial frequencies of the input image that cannot be perceived by a human observer (Sec. 5.5.1). An adaptive aperture can then be computed by maximizing its light transmission while preserving the perceivable frequencies, rather than being restricted to support a constant but broad frequency band (Sec. 5.5.2). Sections 5.5.3 to 5.5.5 discuss how temporal consistency of the computed apertures is achieved and how physical constraints are incorporated. We will show that our adaptive dynamic apertures produce better results than previous methods with the same or even an increased amount of light transmission (Sec. 5.5.6).

5.5.1 Image Analysis

Although the HVS has enormous capacities such as dynamic range, spatial resolution and sensitivity to different wavelengths, it also has limitations. As all lenses in optical systems, the eyeball has a modulation transfer function (MTF) that acts as a low-pass filter on the incident light, thus attenuating or even removing high spatial frequencies. Furthermore, the sensory receptor cells in the retina are not infinitesimally small. Thus, the spatial frequencies that can be resolved on the retina are limited by the sampling theorem. We present a simple model of these optical limitations of the HVS that guide the computation of content-adapted dynamic apertures.

The sensitivity variations of the HVS according to spatial frequencies f_x, f_y are well studied and mathematically defined by the contrast sensitivity function (CSF) $S_{csf}(f_x, f_y)$. Various definitions of this function appear in the literature; we use the one described by Daly [1993]. The CSF depends on the viewing conditions only, not on the actual content. The sensitivity is defined as the inverse of the contrast required to produce a threshold response $S_{csf}(f_x, f_y) = 1/C_{thresh}(f_x, f_y)$, with C_{thresh} being the threshold contrast. Using the definition of Michelson contrast, this is given as $C_{thresh}(f_x, f_y) = \Delta L(f_x, f_y)/L_{mean}$, where ΔL is the necessary luminance difference given in cd/m^2 and L_{mean} is the mean image luminance. An absolute luminance threshold map can be computed as:

$$\Delta L(f_x, f_y) = \frac{L_{mean}}{S_{csf}(f_x, f_y)} \quad (5.4)$$

The sensitivity for frequencies in the range of 2 to 4 cycles per visual degree (cpd) is highest; it drops for lower and higher frequencies. As outlined by Ramasubramanian et al. [1999], an overestimation of the threshold ΔL for lower frequencies can be avoided by setting the frequency sensitivity below 4 cpd to its maximum. The threshold map is shown in Figure 5.7.

The data for Equation 5.4 was acquired by presenting human subjects gratings of various frequencies and intensities on a constant background luminance [Daly, 1993]. The equation is a mathematical model that allows a single spatial frequency to be altered below the allowed threshold without being perceived. The actual threshold represents a 50% probability for the change to be detected by a human observer.

For computing our dynamic apertures, we wish to eliminate all frequencies that do not contribute to perceivable image fidelity. An intuitive understanding of this is reached by regarding the CSF as the equivalent noise of the HVS. Similar approaches have been used to add noise with a spectral shape of the CSF to images [Ahumada and Watson, 1985]. Note that we do not aim at adding noise, but rather remove frequencies from the original footage that are within the range of the HVS noise. Image compression techniques such as the JPEG standard [ITU, 1993] employ similar ideas by quantizing the image frequencies. High frequencies with low magnitudes are usually quantized to zero.

The net-effect of modifying more than one spatial frequency within the threshold given by Equation 5.4 can be expressed using probability summation as $P_{net} = 1 - \prod_{i \in f_x, f_y} (1 - P_i)$ [Daly, 1993]. It is easy to see that modulating multiple frequencies, even within the threshold given by Equation 5.4, results in a noticeable effect. We compensate for the mutual interaction between modulated frequencies by lowering the threshold for each frequency to a detection probability that is very close to zero so that any combination of altered frequency has still a low probability of being noticeable¹.

The Fourier magnitudes of an image converted to absolute luminance values $L(f_x, f_y)$ correspond to the amount of spatial frequencies in the image. With this information, we can calculate a binary importance mask for the image frequencies as:

$$M(f_x, f_y) = \begin{cases} 1, & |L(f_x, f_y)| \geq s\Delta L(f_x, f_y) \\ 0, & otherwise \end{cases} \quad (5.5)$$

As seen in Figure 5.7, filtering the Fourier transform of an image with the binary importance mask M allows us to remove spatial frequencies that do not modify the perceived image

¹In practice we scale the threshold luminance in Equation 5.4 by a user defined parameter s . A scaling factor of 0.01 has proven successful for our experiments.

content for specific viewing conditions that include a fixed adaptation luminance, viewer position, and screen size.

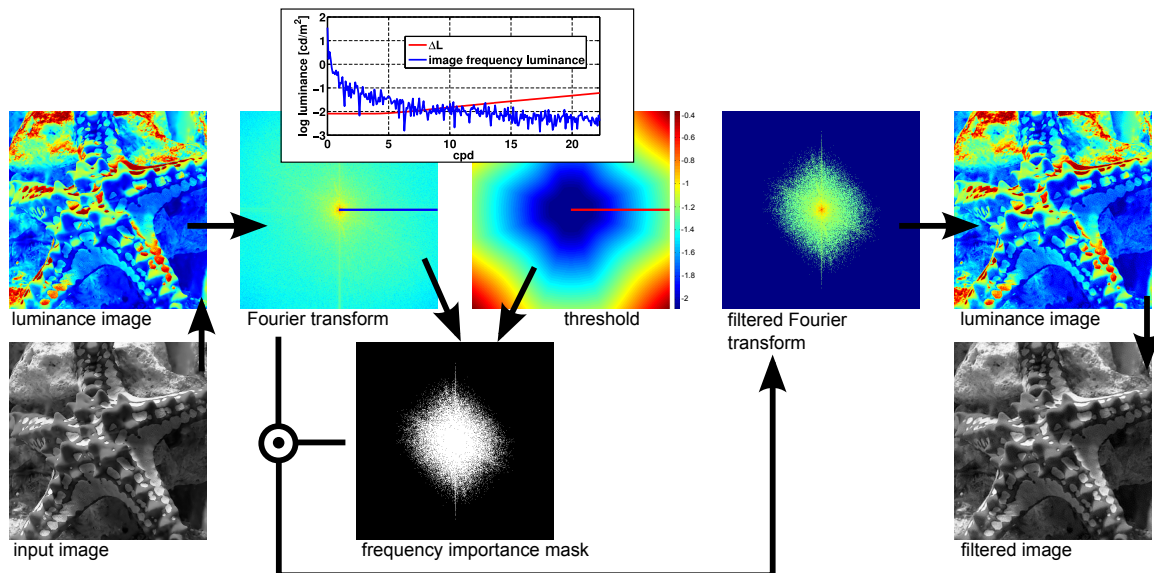


Figure 5.7: Adaptive thresholding: The original image is converted to absolute luminance values. A binary frequency importance mask can be computed by thresholding the image frequencies according to a model of the HVS. The difference between original and filtered image is not perceivable under specific viewing conditions. Scanlines of the image’s Fourier transform and the threshold map $\Delta L(f_x, f_y)$ are shown in the plot.

5.5.2 Dynamic Aperture Adaptation

Now let’s take a look at how to compute the dynamic aperture itself. We define the aperture as the sum of its individual pixels $a(x, y) = \sum_{i=1}^N a_i p_i$, where p_i is the pixel p at x_i and y_i (with a total of N pixels) and $a_i \in [0, 1]$ is its transmissivity. The Fourier transform of the aperture is $F\{a(x, y)\} = A(f_x, f_y) = \sum_{i=1}^N a_i P_i$. Our dynamic apertures should support all important frequencies in the input image with a minimal variance of their Fourier transform. In addition, they should maximize light throughput.

The variance of the aperture’s modulation transfer function (MTF) is a measure for how different frequencies are attenuated. Minimizing it for all important frequencies ensures that they are all supported. A similar criterion was employed by Raskar et al. [2006] for a one-dimensional binary temporal mask. The minimization can be mathematically expressed as an optimization problem:

$$\underset{a}{\text{minimize}} \quad \|\mathbf{M}\mathbf{B}a - b\|_2^2, \quad (5.6)$$

where b is a vector containing only 1s and $a_i > 0$ are the aperture pixel intensities. We do not enforce the pixel intensities to be below 1 in this formulation, but simply scale the resulting values so that the maximum is 1. This is equivalent with a scaling of the MTF and does not affect the variance criterion. \mathbf{M} is a diagonal matrix containing the binary frequency importance mask values described in Section 5.5.1. \mathbf{B} is a matrix with orthogonal basis functions in its columns which represent the optical transfer function (OTF) of the N individual aperture pixels P_i . This results in a linear system of the form $\mathbf{M}\mathbf{B}a = b$. Solving this heavily over-determined system in a least-squared error sense with the additional constraint to minimize $\|a\|_2^2$ will minimize the variance of the Fourier transform of the aperture for important frequencies. This formulation also intrinsically maximizes the light transmittance of the resulting aperture, because a small squared ℓ_2 -norm of a ($a_i \geq 0$) also minimizes the variance of the normalized pixel intensities in the spatial domain.

The linear system can certainly be solved with standard approaches, such as the conjugate gradient method for the normal equations or non-negative least squares solutions. However, this would not allow sufficiently high frame rates on commonly available computer hardware for standard image resolutions of 1024×768 and higher. Thus, we propose to solve the system using the pseudo-inverse matrix, an approach that has previously been employed in solving inverse problems for light projection in real-time [Wetzstein and Bimber, 2007]. Computing solutions of linear problems using the pseudo-inverse minimizes the least-squared error and the ℓ_2 -norm of the resulting vector, thus solving the variance and the light transmittance problem at the same time.

Reformulating our problem results in $a = \mathbf{B}^+ \mathbf{M}^+ b$, where $^+$ denotes the pseudo-inverse matrix. Since \mathbf{M} is a binary diagonal matrix its pseudo-inverse is the matrix itself, i.e. $\mathbf{M}^+ = \mathbf{M}$. \mathbf{B} comprises the set of orthogonal Fourier basis functions as its columns, thus $\mathbf{B}^+ = \mathbf{B}^*$. We need to employ the conjugate transpose \mathbf{B}^* , because \mathbf{B} is complex, hence:

$$a = \mathbf{B}^* \mathbf{M} b \quad (5.7)$$

In this formulation \mathbf{B}^* can be easily pre-computed. During run-time we solve the system with a matrix-vector multiplication. Since the solution a can contain negative values we clip these values and scale the result so that the maximum value is 1.

5.5.3 Enforcing Temporal Consistency

Altering the aperture of the projector during image display affects the average luminance from frame to frame. In order to avoid noticeable intensity variations in form of flickering during subsequent frames with a similar average luminance, we employ the model for temporal luminance adaptation of the HVS that was proposed by Durand and Dorsey [2000]. This was previously used for interactive tone mapping [Durand and Dorsey, 2000; Krawczyk et al., 2005] and adaptive radiometric compensation [Grundhöfer and Bimber, 2008].

The luminance adaptation process can be described as an exponential decay function [Durand and Dorsey, 2000]:

$$L_t = L_{t-1} + (L_t^* - L_{t-1}) \left(1 - e^{-\frac{T}{\tau}}\right), \quad (5.8)$$

where T is the time step between consecutive frames in seconds, τ is a time constant, we assume photopic vision, which is commonly modeled with $\tau = 0.1$ sec [Durand and Dorsey, 2000]. L_t is the smoothed average aperture luminance at time step t and L_t^* the unsmoothed average luminance of the normalized dynamic aperture.

Thus, instead of using the highest possible mean transmittance for a dynamic aperture computed at frame t , we use the same aperture scaled so that its mean transmittance does not, according to the model of luminance adaptation, result in perceivable flickering compared to the last frame.

5.5.4 Accounting for Different Amounts of Defocus

Although Equation 5.7 presents a valid solution for computing an adaptive, content-dependent aperture, this only accounts for a single aperture scale, which corresponds to exactly one fronto-parallel plane with a fixed distance from the optical focal plane. The distance is proportional to the scale of the aperture in the projected image.

In order to account for multiple scales we propose to choose the largest measured defocus and as a result the largest required scale for computing of the dynamic aperture. Recall that a decreased scale in the spatial domain is equivalent to an increased scale in the frequency domain. Consequently, if the Fourier transform of the dynamic aperture with the largest scale supports all important frequencies, so will all the smaller spatial scales. We assume that displayed images follow natural image statistics, which means that their spectrum somewhat resembles the shape $1/f$ in the frequency domain. It is important

to note that natural image statistics determine the general shape of our binary frequency importance mask M (Eq. 5.5). Assuming that lower frequencies have higher magnitudes leads to an M which is always completely filled. Since the dynamic aperture is optimized to support all frequencies within M for the largest spatial scale, all smaller scales (i.e., less-defocused areas) will contain M as a subset because they support a larger area in the frequency domain. Therefore, we compute an adaptive aperture using Equation 5.7 for the largest measured amount of defocus and resample the result to the resolution of the physical aperture display using integral sampling.

5.5.5 Incorporating Physical Constraints of the Prototype

The liquid crystal arrays in our prototypes are currently limited to binary mask patterns. The resulting values of Equations 5.6 and 5.7 are continuous, and cannot be displayed directly on binary LCAs. In order to solve Equation 5.6 for a binary LCA, discrete optimization approaches, such as integer programming, would have to be employed. However, these would not account for the limited contrast of LCAs. Furthermore, discrete optimization approaches are generally much slower than continuous methods. In order to reach interactive frame rates for dynamic image content, we first calculate a continuous adaptive aperture mask with Equation 5.7 on the GPU, and then apply a simple continuous-to-binary mapping scheme. Given the minimum and maximum aperture display transmittance (τ_{min} and τ_{max}) we compute the binary aperture a_{bin} as follows:

$$a_{bin}(x, y) = \begin{cases} 1, & a(x, y) \geq \tau_{min} + \frac{(\tau_{max} - \tau_{min})}{\delta} \\ 0, & otherwise \end{cases} \quad (5.9)$$

The parameter $\delta \geq 1$ represents a tradeoff between light transmissivity of the binarized aperture and accuracy. The minimum and maximum display transmittances are given by τ_{min} and τ_{max} respectively. Empirically, we found $\delta = 2$ to be appropriate.

5.5.6 Adaptive Coded Aperture Results

We have implemented the computation of our dynamic aperture patterns entirely on the GPU. Therefore, the B^* matrix is precomputed and uploaded onto the graphics hardware memory. We store the matrix as complex floating point values and limit the matrix size to $1024^2 \times 7^2$ entries, which requires 383 MB of memory storage. Hence, the compensation is performed on images with a resolution of 1024×1024 and an aperture resolution of 7×7 .

Using NVIDIA’s CUDA implementation of BLAS, we achieve about 13 ms for the entire computation of one aperture pattern given an original image on an NVIDIA GeForce 8800 Ultra.

Using the computed aperture kernel, the original image then needs to be pre-compensated to counteract the optical defocus. While a compensation of dynamic content with a static kernel allows the inverse kernel to be pre-computed in different scales, as described in Section 5.4.2, these need to be computed at each frame for dynamically adapted aperture patterns. The additional computation times for a varying number of kernel scales is summarized in Figure 5.8.

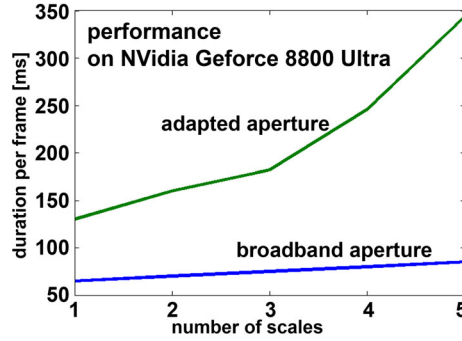


Figure 5.8: While all required data for a static aperture pattern can be precomputed, the dynamically computed adaptive apertures add additional computational load to each frame, thus lowering performance for a larger number of kernel scales, i.e. a larger number of different distances to the focal plane.

Figure 5.9 shows several different images that are compensated using our adaptive apertures. Each of the examples shows the in-focus and defocused projection, as well as the compensated defocused case. The sub-images show the corresponding continuous (Sec. 5.5.2) and binary (Sec. 5.5.5) aperture patterns. Since our prototype is limited to binary patterns, only the latter are displayed on the integrated LCA. Although natural images roughly follow a $1/f$ frequency distribution, Figure 5.9 shows that our adaptive apertures differ from a $1/f$ distribution depending on the image content and greatly enhanced the quality of a defocused projection.

5.6 Evaluation and Comparison to Previous Work

Due to the fact that our adaptive coded apertures are computed using a model for the limitations of the human visual system, we validate our theory using a visual difference predictor (VDP) [Mantiuk et al., 2005]. This operator can be applied to defocused and

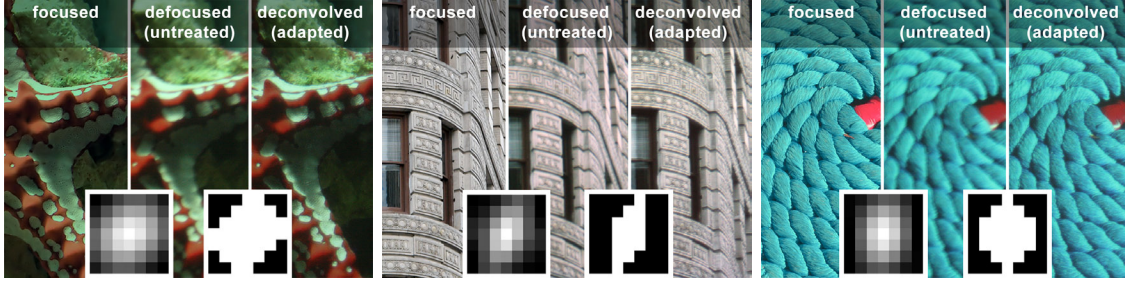


Figure 5.9: *Placing a transparent liquid crystal array at the aperture plane of a projector lens allows encoding the aperture’s mask pattern dynamically – depending on the perceivable frequencies of the displayed images. The inlays illustrate the computed intensity code and the applied binary masks. Each of the three image trippels are perception optimized and have been computed for viewing at a minimal distance of 50 cm when being displayed at a maximum diagonal of 21 cm on the screen. Possible artifacts could only be perceived when observing at closer distances or larger sizes.*

compensated images given in absolute luminance values under specific viewing parameters including adaptation luminance, distance to screen, and size of the canvas. The result is a map that contains the detection probabilities for each pixel compared to a reference image. We applied the VDP to our worst result (as explained in Fig. 5.11), the “house” scene in Figure 5.9. Figure 5.10 visualizes the probability of difference detection for the defocused image, the compensated defocused image projected with a circular aperture, and a compensation with our static and dynamic apertures, compared to the focused original image. All input photos for the VDP were taken with the same camera settings in a calibrated environment. The probabilities are overlaid on the original image showing a high detection probability in red, moderate probabilities in green, and low ones in the original grayscale image value. The adaptive coded aperture performs best.

All previous work that attempted to compensate defocus with single display devices and circular apertures [Brown et al., 2006; Oyamada and Saito, 2007; Zhang and Nayar, 2006] achieves results that are similar to the compensated case with circular apertures (Figures 5.1 and 5.10). Although all of them slightly differ in their exact solution, the best achievable quality is limited by the physical aperture itself and is thus represented by a compensation with circular apertures in this chapter.

For evaluation we also wish to quantitatively compare the depth-of-field that we can achieve with previous work and conventional projection, as well as the loss in light transmission. The DOF and the light throughput of a projection system depend mainly on the effective aperture diameter and on the focal length of the objective lens. In photography, the ratio of these two parameters is generally expressed by the objective’s f -stop number. A low

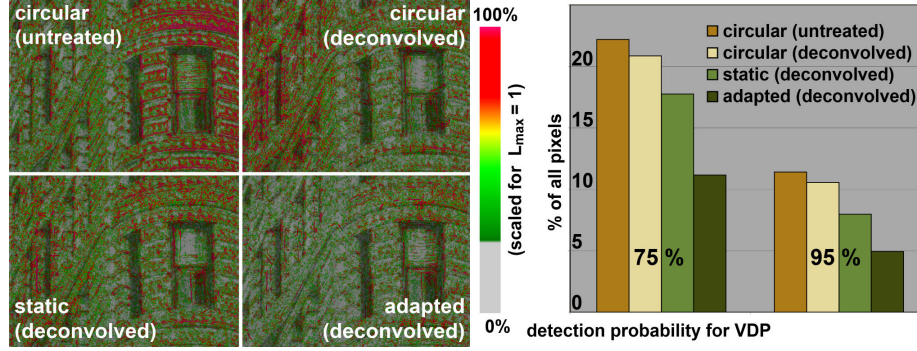


Figure 5.10: Results of visual difference prediction when compared with focused projection: colour coded detection probability (left) and bar charts for 75% and 95% difference detection probabilities (right) – all for house example in Fig. 5.9.

f -stop is equivalent to a large light throughput with a small depth-of-field and vice versa. A sufficiently wide depth-of-field is necessary for projections onto curved surfaces, such as domes or cylindrical screens. However, since bright images are desired for almost all applications, diaphragms that reduce the effective aperture diameter in favor to a wider depth-of-field are normally not incorporated into projectors. Exceptions are devices with automatic iris control (auto-iris) that can dynamically adjust the diameter of a circular diaphragms and, thus, control the DOF.

	defocused (untreated)	circular (deconvolved)	broadband (deconvolved)
$\hat{f}/\#$	3.1	3.7	4.3
$\tilde{f}/\#$	3.1	3.1	3.9

	adapted (deconvolved)			
	house	starfish	rope	lena
$\hat{f}/\#$	4.4	5.4	5.6	7.7
$\tilde{f}/\#$	4.0	3.5	3.7	3.7

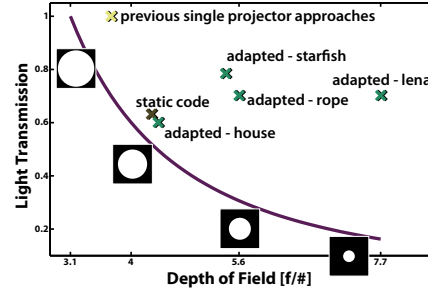


Figure 5.11: The table on the left shows a comparison of depth-of-field enhancement as well as light throughput of compensated projections with our coded and circular apertures. The values are given in relative f -numbers comparing DOF ($\hat{f}/\#$) and transmission ($\tilde{f}/\#$). Higher values are better for DOF enhancement, lower values better for light transmission. Both, DOF and light transmission increase from circular to broadband and even more to adapted apertures. The plot on the right illustrates the tradeoff between DOF and transmission for the examples in the tables and for an auto-iris system (curve).

Following Oyamada and Saito [2007], we determine the approximate scale of the PSF in a compensated projection by finding the best match of captured projection and the original image convolved with different scales of that PSF. Using the determined scale we compute

the f -number of an objective lens with a circular aperture (and constant focal length) that would lead to the same depth-of-field ($\hat{f}/\#$). For the “lenna” image in Figure 5.1, for instance, an $\hat{f}/7.7$ aperture stop is needed to achieve the same DOF with a standard projector, which corresponds to a tremendous amount of light loss (factor 0.16). As seen in Figure 5.11, previous approaches using deconvolution with circular apertures can slightly enhance the DOF, whereas our static broadband mask and especially the adaptive codes achieve a much better gain in DOF.

The enhanced DOF comes at the cost of reduced light transmission, which is the case for all the examples in Figure 5.11. We evaluate the light throughput also in f -numbers ($\tilde{f}/\#$) with respect to a comparable standard projector equipped with an auto-iris. The results show that our patterns can always achieve a larger DOF without losing as much light as an auto-iris projector, for which $\hat{f}/\# = \tilde{f}/\#$. Our adaptive codes also always achieve larger DOFs than the static coded aperture, while maintaining a higher light throughput.

5.7 Other Applications for Coded Apertures

So far, we have seen how static and dynamic coded apertures can improve the depth-of-field of a projection system, while maintaining a high light transmission. Our system is, however, more powerful and flexible than that. In the following, we outline two different applications of the proposed projection systems: smooth focused projections via de-pixelation and improved temporal image contrast.

Pixelation is an artifact created by all digital projectors. Due to spacing between the display pixels and their spatial discretization, focused projections often appear jaggy as shown in Figure 5.12. This effect becomes especially irritating for high resolution photographs of projected imagery, as modern digital cameras have an ever increasing spatial resolution. Following Zhang and Nayar [2006], we can introduce a slight optical defocus that lets the pixel structures optically vanish. Using our defocus compensation, we can invert the effect and thus display sharp images without visible pixel structures. Results for a static aperture de-pixelation (scene from Fig. 5.6) are shown in Figure 5.12 in the upper row, while the lower row presents a closeup of the adaptive “rope” scenario of Figure 5.9.

Another application of our system is improved temporal contrast. Modern auto-iris projection systems automatically adjust the iris opening to dynamically control the amount of projected light. For video sequences with frames that differ significantly in image brightness, this leads to an increase in inter-frame or temporal contrast. Specifically, the advantage for auto-iris displays is the option to either transmit all possible light for very bright images,

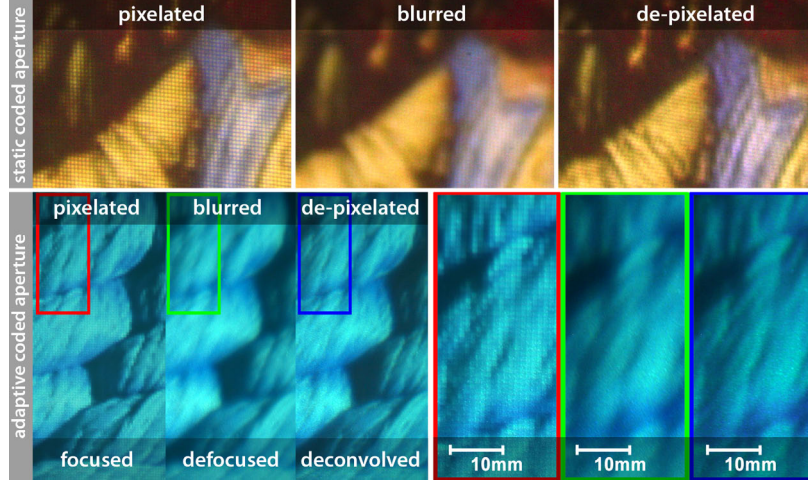


Figure 5.12: In some situations, the display pixel grid can become visible and disturbing for focused projections. A slight optical defocus diminishes the visibility of the grid structure, but also blurs image features. Using our coded apertures and corresponding defocus compensation algorithms, we can achieve smooth and sharp projections as shown for these examples. The upper row are magnifications for a de-pixelation performed with our static broad-band mask, while the lower row shows results from our adaptive aperture.

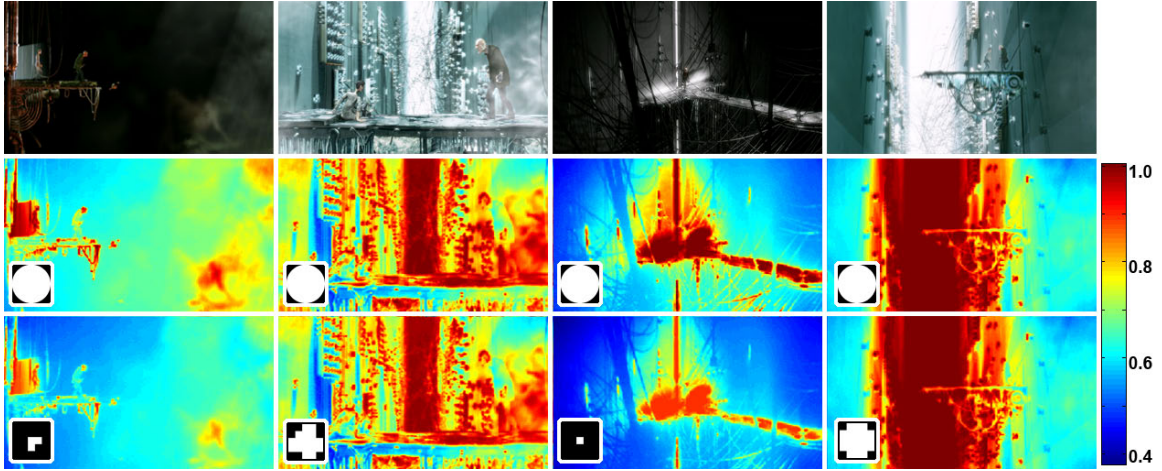


Figure 5.13: The temporal contrast of displayed video sequences can be increased using adaptive coded projector apertures. Therefore, the aperture pattern is adjusted in accordance with the presented image brightness. For very bright frames (columns 2 and 4) this yields an open aperture, while dark images result in small aperture openings (columns 1 and 3), thus effectively reducing the projector's blacklevel. Four frames of a video sequence are shown in the upper row, while rows 2 and 3 are logarithmically colour-coded luminance values with unmodified and scaled coded apertures, respectively. In contrast to standard auto-iris lenses, luminance-scaled adaptive coded apertures increase the depth-of-field additionally.

while reducing the aperture size for dark frames. The projector blacklevel can be reduced and the temporal contrast therefore increased. Using our adaptive aperture projector, we can adjust the attenuation pattern to control the light throughput, just like auto-iris systems. This is demonstrated in Figure 5.13, where the aperture opening is scaled according to the average image brightness for several frames of a video sequence (after being computed as explained in Section 5.5). The projected frames are colour-coded for better visualization and show how the blacklevel is decreased for dark images.

5.8 Discussion

In this chapter, we have presented algorithms and prototypes to compensate optical defocus of a single video projector by means of inverse light filtering and coded apertures. We have shown how integrated aperture patterns can increase the depth-of-field of ordinary video projectors. Adaptive coded apertures optimize depth-of-field versus light throughput based on the limitations of the human visual system, and can potentially lead to a new generation of auto-iris projector lenses. They outperform static broadband masks and circular aperture stops for digital projector defocus compensation and lead to a larger depth-of-field, higher quality projector de-pixelation, and increased temporal contrast of projected video sequences. Static coded apertures, however, are easy to manufacture and to integrate into existing projector designs, and are therefore more economical. In particular projector de-pixelation, where the discretized pixel structure is optically vanished through defocus while image details are recovered through inverse filtering (cf. Figure 5.12), will not only improve the image quality of close-view projection-based displays (such as rear-projected TV sets) but also of devices that utilize projection-based illumination techniques. High-dynamic range displays that apply spatially modulated projected backlight, such as [Seetzen et al., 2004], suffer from visible moiré patterns if the projection is focused on the screen plane. The hard- and software solutions we have presented in this paper could form an integral part of next generation projection devices.

The main limitations of our adaptive coded apertures are currently imposed by the employed LCA. Its low light transmittance (only 30% when completely transparent) results in a relatively high loss of light. Therefore, we trade light throughput for depth-of-field. As spatial light modulators (SLMs), such as a high contrast continuously valued LCAs with higher transmittance, reflective digital micro-mirror (DMD), and liquid crystal on silicon (LCoS) chips, become more widely available we expect better results with these displays. DMDs also do not suffer from the light loss and low contrast as LCDs do. We have shown that our approach is feasible and plan to experiment with alternative SLMs in the future.

We also believe that high contrast and brightness at low power consumption and heat development will become feasible with light engines that apply upcoming LED technology.

Another limitation of our current adaptive aperture prototype is the moderate performance for a large number of compensated defocus scales. This will improve with next generation graphics hardware, or with customized integrated image processors. With a higher performance, larger resolution aperture patterns can be computed. This would lead to less constraints for the discrete scaling operations that are required for inverse filtering and temporal aperture adaptations for a flicker-free increased temporal contrast. Currently, we are constrained to computations with an aperture resolution of 7×7 pixels.

Our adaptive coded aperture prototype is limited to binary patterns. Therefore, we have to round our adaptive masks. Although this rounding and other approximation make our eventually displayed aperture codes not quite optimal in practice, we believe that the quality of the presented results and the generality of our approaches, which also apply for alternative prototypes or commercial displays, demonstrate the feasibility of our algorithms and hardware setups.

Chapter 6

Hand-Held Schlieren Photography with Light Field Probes

While Chapters 3 and 4 introduce approaches to display light fields containing 3D scenes for glasses-free 3D display, in this chapter we establish methods for visualizing and capturing refraction in transparent media using light field probes. As compared to displays, which present visual information to a human observer, probes are engineered for computer vision applications. We call the presented approach Light Field Background Oriented Schlieren Photography (LFBOS). By optically coding the locations and directions of light rays emerging from a light field probe, we can capture changes of the refractive index field between the probe and a camera or an observer. Rather than using complicated and expensive optical setups as in traditional Schlieren photography LFBOS employs commodity hardware; our prototype consists of a camera and a lenslet array. By carefully encoding the colour and intensity variations of a 4D probe instead of a diffuse 2D background, expensive computational processing of the captured data is avoided, which is necessary for Background Oriented Schlieren imaging (BOS). This chapter also analyzes the benefits and limitations of LFBOS and discusses application scenarios.

6.1 Introduction and Overview

The acquisition of refractive phenomena caused by natural objects has been of great interest to the computer graphics and vision community. Joint optical light modulation and computational processing can be used to reconstruct refractive solids, fluids, and gas flows [Ihrke et al., 2010a], render complex objects with synthetic backgrounds [Zongker et al., 1999], or validate flow simulations with measured data. Unfortunately, standard optical systems are not capable of recording nonlinear trajectories that photons travel along in inhomogeneous media. In this chapter, we present a new approach to revealing refractive phenomena by coding the colours and intensities of a light field probe. As illustrated in Figure 6.1, the

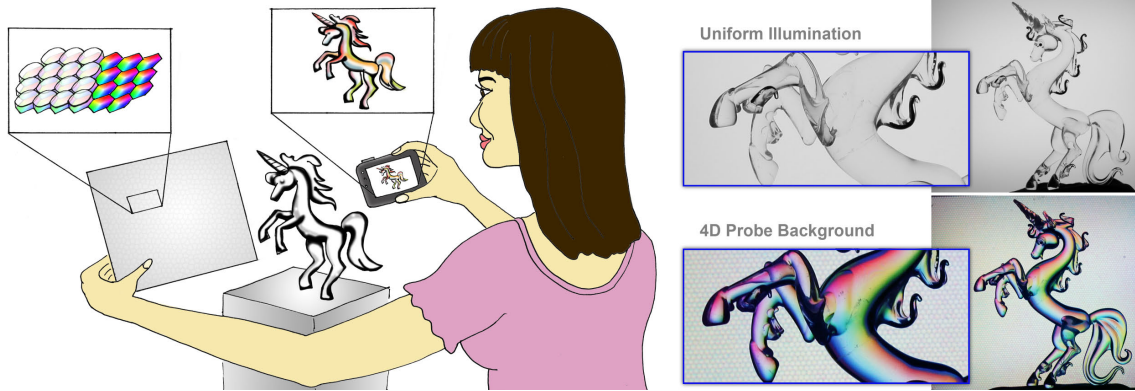


Figure 6.1: *Light field probes —when included into the background of a scene— allow otherwise invisible optical properties to be photographed. In this example, the probe contains a classic Rainbow Schlieren filter that codes the angles and magnitudes of complex refractive events in hue and saturation.*

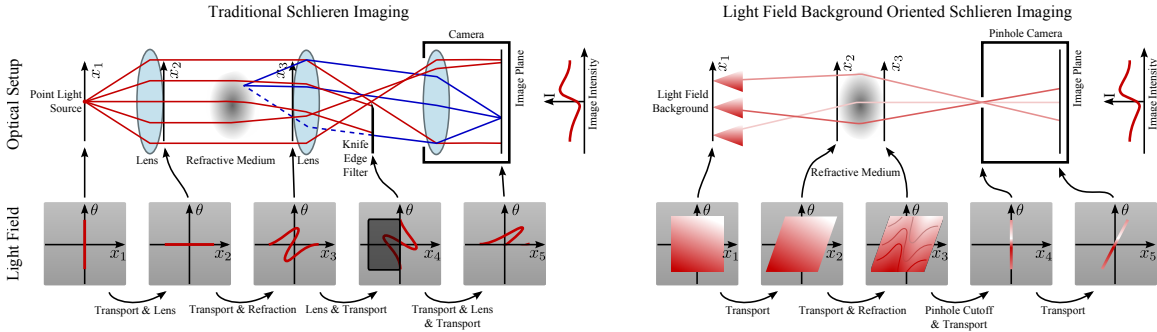


Figure 6.2: *Illustration of optical setups and light field propagation for traditional Schlieren imaging (left) and Light Field Background Oriented Schlieren photography (right). Optical paths for forward light propagation to the camera are red, whereas backward propagation paths from a camera pixel are blue. Please note that the 4D light field probe illustrated on the right only codes the two angular dimensions in this case.*

probe is positioned behind an object of interest and the object and probe are photographed by a camera. Due to refractions caused by the medium, apparent colours and intensities of the probe change with the physical properties of the medium, thereby revealing them to the camera or a human observer.

The idea of optically transforming otherwise invisible physical quantities into observed colours and changes in intensity is not new. In fact it occurs in nature in the form of caustics. These types of phenomena are generally referred to as Shadowgraphs and reveal only limited information of the underlying physical processes [Settles, 2001]. More sophisticated

techniques to visualizing and photographing gas and fluid flows, refractive solids, and shock waves were developed in the 1940s [Schardin, 1942]. Some of the phenomena that were depicted for the first time include the shock waves created by jets breaking the sound barrier and bullets flying through the air, or the heat emerging from our bodies. As illustrated in Figure 6.2 (left, red lines), traditional Schlieren setups require collimated illumination, which is then optically disturbed by changes in the refractive index of a medium. A lens deflects all light rays so that the ‘regular’ rays, those that were not refracted, intersect that plane at one specific point, usually the center. The ‘irregular’ or refracted rays intersect the plane at different points, which are determined by the angle and magnitude of the refraction. Optical filters such as knife edges or colour wheels can be mounted in that plane to encode these properties in colour or intensity. Further light propagation optically transforms the rays back to their ‘normal’ distribution and an image can be recorded with a camera. Each pixel on the sensor is focused on a specific point in the refractive medium (Figure 6.2, left, blue lines), which allows an image to be formed along with intensity or colour changes caused by the refraction. Although recent improvements have made traditional Schlieren setups more practical [Settles, 2010], fundamentally, these approaches require precise calibration and high-quality optical elements that are at least as big as the observed objects. Therefore, these systems are usually bulky, expensive, and mostly constrained to laboratory environments.

With the increase of computational power, Background Oriented Schlieren imaging (BOS, Dalziel et al. [2000]) was invented to overcome the difficulties of traditional Schlieren photography. In BOS, a digital camera observes a planar high-frequency background through a refractive medium. Optical flow algorithms are used to compute a per-pixel deflection vector with respect to an undistorted reference background. This type of optical flow estimation requires the background to be diffuse or photo-consistent.

Light Field Background Oriented Schlieren photography (LFBOS) also employs a background probe, but rather than coding only two dimensions, we can encode up to four dimensions of the light field: spatial and angular variation. Table 6.1 shows a comparison of the most important characteristics of LFBOS compared with BOS and traditional Schlieren imaging. The asterisk indicates that although BOS backgrounds have a high spatial resolution and no angular variation, these patterns are usually placed at a large distance to the object so that they effectively become angular-only probes. This increases the size of the setup and often results in focus discrepancies between background and object. Light field probes have a small form factor and do not suffer from focus mismatches.

	Trad. Schlieren	BOS	LFBOS
Coded Dimensions	2D	2D	4D
Setup Complexity	moderate-difficult	easy	easy
Angular Resolution	very high	N/A*	high
Spatial Resolution	N/A	very high*	moderate
Processing	low	very high	low
Cost	high	low	low
Form Factor	usually bulky	usually bulky*	small

Table 6.1: *Overview of Light Field Background Oriented Schlieren photography compared to traditional and Background Oriented Schlieren photography (*see text).*

6.2 Theory

6.2.1 Image Formation

The propagation of light in inhomogeneous refractive media is governed by the ray equation of geometric optics [Born and Wolf, 1999]:

$$\frac{\partial}{\partial s} \left(n \frac{\partial \mathbf{x}}{\partial s} \right) = \nabla n, \quad (6.1)$$

where \mathbf{x} is the position of a photon on a trajectory in space, ∂s is the differential path length along the trajectory, and n is the (spatially varying) refractive index field. The wavelength dependency of n is disregarded in this model. Equation 6.1 can be formulated as a coupled system of first-order ODEs [Atcheson et al., 2008; Ihrke et al., 2007]:

$$n \frac{\partial \mathbf{x}}{\partial s} = \mathbf{d}, \quad \frac{\partial \mathbf{d}}{\partial s} = \nabla n, \quad (6.2)$$

with \mathbf{d} being the local direction of propagation. Integrating Equation 6.2 leads to an expression for the global directional deformation within a refractive object [Atcheson et al., 2008]:

$$\mathbf{d}^{out} = \mathbf{d}^{in} + \int_c \nabla n \, ds. \quad (6.3)$$

In our setup, the refractive index field is observed against a known light field background, as shown in Figure 6.2 (right). In order to understand the distortion of this known 4D probe, Equation 6.2 can be solved numerically, for instance with forward Euler schemes

[Ihrke et al., 2007]. However, in our case it is more intuitive to trace the light field on the sensor plane back to the probe. Considering a pinhole camera, as shown in Figure 6.2 (right), the light field probe $l(x, y, \theta, \phi)$ is sampled in the following way:

$$i(\mathbf{x}_p) = l\left(\varsigma(\mathbf{x}_p, \mathbf{d}_p), \varphi\left(\mathbf{d}_p + \int_c \nabla n ds\right)\right). \quad (6.4)$$

Here, i is the sensor image and \mathbf{d}_p the normalized direction from a pixel $\mathbf{x}_p = (x_p^x, x_p^y)$ to the camera pinhole. The function $\varphi(\mathbf{d})$ maps a direction $\mathbf{d} = (d^x, d^y, d^z)$ to the angular parameterization of the light field, i.e. $\varphi_{\theta, \phi}(\mathbf{d}) = (\tan^{-1}(d^x/d^z), \tan^{-1}(d^y/d^z))$. The position and direction of a light ray incident on a pixel can be mapped to a position on the probe by the function ς ; this depends on the distances between camera and object, object and probe, and the ray displacement within the object. If the ray displacement is negligible, the refractive medium can be approximated as a thin element that causes a single refractive event, such as a thin lens. Reconstructions of such objects are presented in Chapter 7. With the above model we neglect intensity variations caused by the camera pinhole or the specific structure of the probe (e.g. lenslets).

Although pinhole cameras sample discrete rays of the refracted light field, in practice cameras usually have a finite aperture implying that each pixel integrates over a small range of light field directions on the probe. We assume that the distance between the light field probe and the camera is large compared to the size of the aperture and that the distance between the probe and the refractive object is small, which yields a good approximation of a pinhole camera.

6.2.2 Designing Light Field Probes

The goal of LFBOS is to encode the directions and locations of a 4D probe with colour and intensity variations so that the former parameters can be inferred from the colours in a photograph. The immediate question that arises is how one can design a light field probe that encodes positions and directions in a meaningful way. To answer this question, we need to consider a number of application-specific parameters that are discussed in the following.

Although a pinhole camera is a reasonable approximation under the above mentioned assumptions, finite pixel sizes and small apertures along with strong distortions of the wavefront caused by refraction often amplify the integration area of each camera pixel in the space of the light field probe. In order to compensate for this effect, the distribution of *colour tones and intensities* in the probe should be *smooth* in the 4D spatio-directional

domain. This also implies graceful degradation of the captured colours in case the pinhole assumption breaks down.

If the absorption of light within the refractive medium is not negligible, the colour distributions should ideally be independent of intensity changes. This can, for instance, be implemented by encoding the desired parameters only in hue and saturation, i.e. constant values in HSV colour space. Doing so also ensures resilience to vignetting and other possible intensity changes caused by the lenslets of the probe.

Positions and directions of a background light field can either be encoded in the *absolute* 4D reference frame of the probe or *relative* to a fixed camera position. This is not necessary for orthographic cameras, but to *compensate for the perspective* of a non-orthographic camera.

6.3 Prototype and Experiments

6.3.1 Prototypes

We have implemented prototypes of our light field probes using both lenticular sheets with cylindrical lenses and lenslet arrays with hexagonal grids of spherical lenses. These make a tradeoff between spatial and angular resolution. The specific lenslet arrays we have used in our experiments along with their spatial and angular resolution and fields-of-view are listed in Table 6.2. Alternative implementations of 4D probes include holograms (e.g., www.zebraimaging.com) or dynamic parallax barrier displays (e.g., Lanman et al. [2010]); these allow for probes with a very high spatial and angular resolution.

Lenslet Type	f [in]	d [in]	fov [°]	Angular Resolution [°]		
				600 dpi	2038 dpi	5080 dpi
MicroLens Animotion 10	0.11	0.1	48	0.80	0.24	0.09
MicroLens 3D 20	0.10	0.05	29	0.97	0.28	0.11
FresnelTech Hexagonal 300	0.12	0.09	42	0.77	0.23	0.09
FresnelTech Hexagonal 310	1.00	0.94	51	0.09	0.03	0.01

Table 6.2: *Technical specifications of the lenslet arrays used in our experiments.*

The lenslet arrays are mounted on a light box that provides a uniform background illumination. We print the probe codes at a resolution of 1200 dpi on transparencies that are manually aligned with the lenslet arrays before mounting them on the light box. For an increased contrast, multiple transparencies can be stacked. Care must be taken with the nonlinear colour transformations between specified digital images, the printer gamut, and

the colour gamut of the camera. We define our patterns within the device-specific CMYK colour gamut of the employed printer (RGB and CMYK ICC profiles of printers are usually available from the manufacturer), then print them with device-internal colour mappings disabled, and record camera images in sRGB space. This process allows us to transform the captured photographs to any desired colour space as a post-processing step.

6.3.2 Angular Filtering

Figures 6.3, 6.4, 6.5, and 6.6 demonstrate a variety of different intensity and colour filters encoded in the probes; these reveal angular variation caused by refraction. For the experiments discussed below, the encoded light field is pre-distorted to account for the perspective of the camera. An undistorted view of the probe from the camera position therefore shows the center of a specific angular probe code.

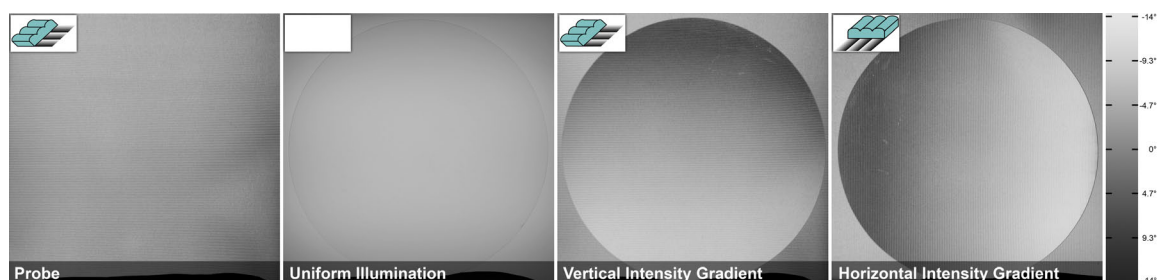


Figure 6.3: A lenticular light field probe encoding 1D directional light variation of the refraction created by a lens with an intensity gradient. This type of filter resembles the knife edge filter of traditional Schlieren photography in the specific direction. Left: probe encoding a horizontal intensity gradient without any object; center left: convex lens with uniform background illumination; center right: lens in front of the probe; right: lens in front of the rotated probe.

Intensity Gradients correspond to knife edge filters in traditional Schlieren imaging. These filters usually sacrifice half of the background illumination by coding undistorted rays in gray; the magnitudes of light deflection in a particular direction are coded in increasing or decreasing intensity. Examples of one-dimensional gradients are shown in Figures 6.3 and 6.4. Here, lenticular sheets are used as probes where the gradient under each cylindrical lens is equal except for a varying lens-to-lens pitch that corrects for the perspective of the camera. As an alternative to 1D gradients, a circular gradient under each circular lens of a hexagonal lenslet array can encode the magnitudes of deflected rays as seen in Figure 6.5 (center right).

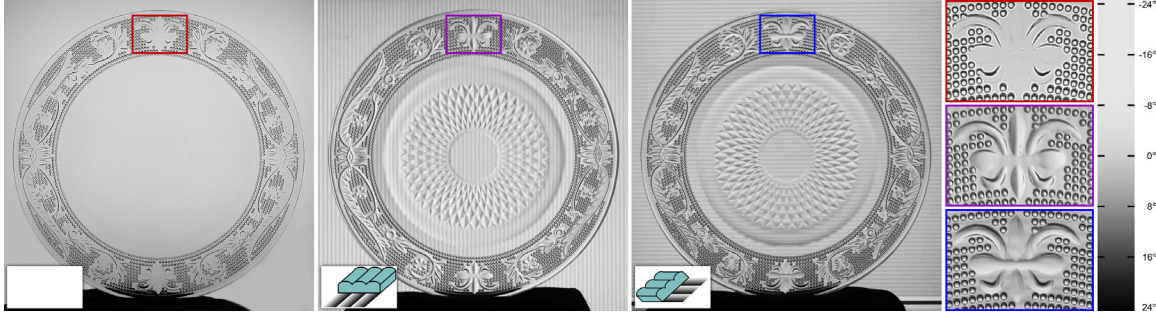


Figure 6.4: A plate in front of a uniform background (left), and a light field probe that encodes directional variation caused by refraction with a horizontal (center left) and a vertical (center right) intensity gradient. The magnifications (right) and the structure on the plate show how otherwise invisible information is revealed with our probes. The colour bar, mapping colours to magnitudes of refraction, is computed from the field of view of the lenticulars and a calibration gradient that is cropped from the photographs.

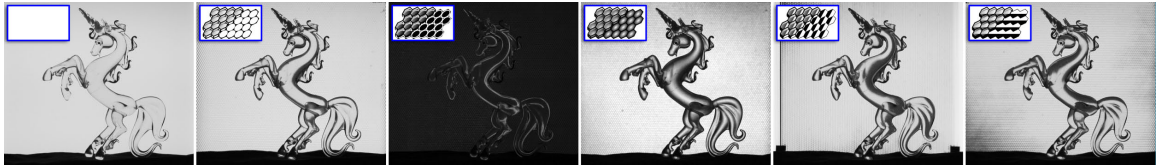


Figure 6.5: A variety of directional filters can be encoded in our probes. From left: uniform background, annular bright field, annular dark field, circular intensity gradient, horizontal cutoff, and vertical cutoff.

Annular Bright Field and Dark Field Filters are inspired by microscopy [Levoy et al., 2009; Murphy, 2001], where similar illumination patterns can be used to illuminate reflective or refractive specimen. The annular bright field probe code is a uniform circular pattern with a small black outline that does not effect undeflected rays or those that underwent only small amounts of refraction. Strong refractions, as seen on the object boundaries of the unicorn in Figure 6.5 (second from left), are completely blocked. The inverse of the annular bright field is the annular dark field, which only allows rays with strong refractions to reach the camera (see Figure 6.5, third from left).

Directional Cutoffs can be used to completely block ray deflections into a certain direction. An example of this filter type can be seen in the vertical-only and horizontal-only cutoff shown in Figure 6.5 (the two rightmost images).

Colour Filters are popular in traditional Schlieren setups. Usually, these approaches are called Rainbow Schlieren [Howes, 1984] and allow the magnitude and the angle of ray



Figure 6.6: *The classic Rainbow Schlieren filter encoded in our light field probe can visualize magnitudes and angles of refractions in complex media such as this mix of clear corn syrup and water.*

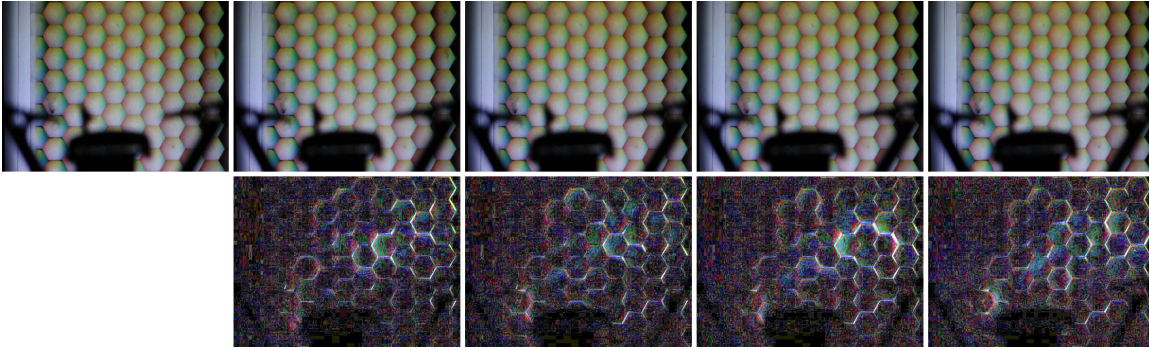


Figure 6.7: *Multiple frames of a video sequence showing a burning camping stove in front of a probe at 30 frames per second (upper row). The probe codes refraction with a hue-saturation colour wheel. Although the angular ray deflections are rather subtle in this case, contrast enhanced difference images (lower row) reveal colour variation due to changes in the refractive index field. For this experiment, the probe had to be placed at a distance to the stove so that it would not melt; this results in focus mismatches. The noise in the difference images is due to lossy mpeg compression; the darker boundaries between lenslets on the probe also cause artifacts.*

deflection to be encoded in different colour gradients. The HSV colour wheel is a popular choice for the filter. Undeflected rays are coded in white, whereas the deflection angle is coded in hue and the magnitude in saturation of the colour wheel. Examples of these filters are shown in Figures 6.1, 6.6, and 6.7.

6.3.3 Spatio-Angular Filtering

So far, we have only considered probes with purely angular filters. Light field probes, however, allow us to additionally encode the spatial locations on the probe. Following the probe design criteria discussed in Section 6.2.2, we experimented with light field probes that encode the 2D angular domain and one of the spatial dimensions in three colour primary gradients.

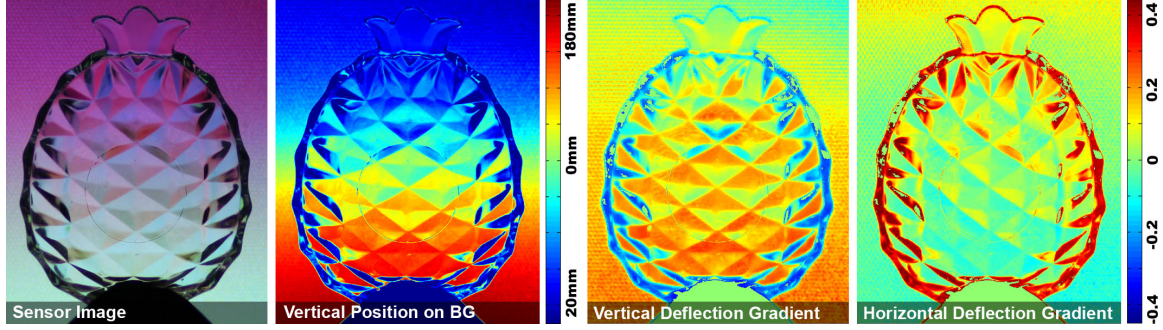


Figure 6.8: A refractive object in front of a light field probe (left) that encodes vertical ray displacement (center left) as well as vertical and horizontal ray deflection. The latter two quantities can be used to compute the gradient of the refractive index field as shown in the right columns.

The scene in Figure 6.8 (left) is photographed in front of a probe that codes horizontal and vertical angular ray deflections in red and blue colour gradients, respectively, and the vertical position on the probe in a green gradient. The absolute colour values are defined within the colour gamut of the printer.

In addition to the spatio-angular colour codes, we include fiducial markers on the probe (cropped from the image) that allow us to estimate the extrinsic camera parameters from a single photograph. Given the intrinsic and extrinsic camera parameters, we can easily compute the undistorted angle and position of each light ray emerging from a camera pixel on the probe background. The light field probe is registered with the markers, so that each ray that is emitted by the probe uniquely encodes its angle and location in colours.

Therefore, by taking a photograph of the probe without any refractive media in the optical path the expected ray locations and angles on the probe match the encoded colours. However, if variations in the refractive index field in between probe and camera cause changes in ray trajectories, these changes in ray angle and displacement are directly observable in the recorded colours as seen in Figure 6.8. By applying Equations 6.3 and 6.4 to the observed and expected light rays, we can compute a per-pixel refractive index gradient (Figure 6.8, right columns) and vertical ray displacement (Figure 6.8, center left). Reconstructing refractive index fields of fluid flows and the shape of transparent solids using these quantities is presented in the next chapter of this thesis.

6.4 Comparison to Background Oriented Schlieren Photography

Background Oriented Schlieren setups usually require a high-frequency background to be located at a large distance to the object. In this way, the per-pixel displacement vectors estimated by optical flow are proportional to the angular ray deflections, which is related to the refractive index gradient (Eq. 6.3). The form factor of these setups is therefore usually large. Furthermore, the camera needs to be focused on the background pattern so that its distortion can be tracked by the optical flow algorithm. The total amount of light in BOS setups is often limited, which is why cameras typically need to use a large aperture for capture. Unfortunately, this places the object of interest out-of-focus as seen in Figure 6.9 (upper left).

Strong refractions, for instance caused by fluids or solids, often lead to extreme distortions of the background pattern. These distortions may prevent a reliable optical flow estimation as shown in Figure 6.9 (upper right). Although an optical flow algorithm for refractive objects has been proposed [Agarwal et al., 2004], this requires many frames of a video sequence to be analyzed and is not practical for dynamic media such as fluids.

In comparison, our approach requires only a single image and the light field background can be placed at close proximity to the object, which alleviates the focus mismatch problem (Figure 6.9, lower left). Furthermore, if the light field probe encodes smooth gradients, as discussed in Section 6.2.2, even a defocused refractive object will reveal the mean of colours and intensities in the integration manifold of the 4D probe space to a camera pixel (Figure 6.9, lower right).

6.5 Limitations

As in most Schlieren approaches, the size of the refractive volume is limited by the size of the employed optical elements, in our case the light field probe. An exception are traditional Schlieren setups that use the sun or other distant point lights as the source of the required collimated illumination and Natural-Background Oriented Schlieren techniques [Hargather and Settles, 2009], which also work outdoors. Currently, both phase and amplitude of the medium caused by refraction and absorption, respectively, are captured. For faithful reconstruction of objects with non-negligible absorption, these effects need to be separated [Barone-Nugent et al., 2002]. Furthermore, we assume that there is no scattering or emission within the medium and neglect any wavelength-dependency of refractive events.

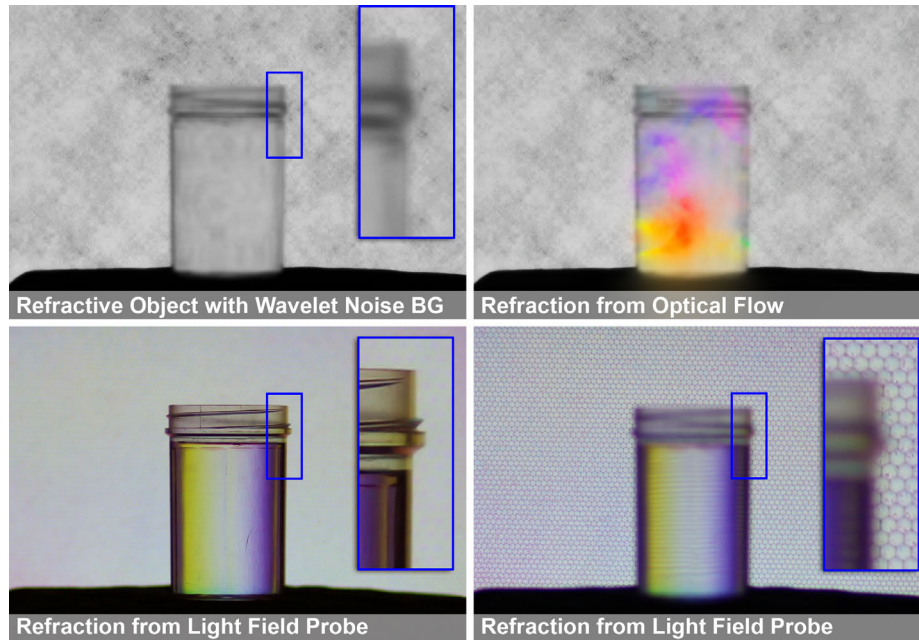


Figure 6.9: *Light Field Background Oriented Schlieren photography compared to a failure case of Background Oriented Schlieren imaging. Optical flow algorithms in BOS require the background to be focused, which places the object out-of-focus (upper left). In this case, the refractions are so strong that they blur out the background pattern and therefore prevent a reliable optical flow estimation (upper right). LFBOS works for in-focus (lower left) and out-of-focus settings (lower right).*

One of the major limitations of our prototypes is the limited resolution and colour gamut of the printer. All of the transparencies in our experiments were printed with off-the-shelf inkjet printers, usually with 1200dpi. Alternative processes are light valve technology, which exposes digital images with a high resolution, contrast, and colour gamut onto film (www.bowhaus.com), or professional offset printing.

All of our current prototypes are implemented with lenslet arrays or lenticulars, which trade spatial and angular resolution. A very high spatial and angular resolution can be achieved with alternative technologies, such as holograms (www.zebraimage.com) or dynamic parallax barrier displays [Lanman et al., 2010]. When the camera is focused on a lenslet-based probe, the space between individual lenses usually appears darker. This problem could also be overcome with alternative probe implementations.

As the fields-of-view of the lenslets in our current prototypes are defined by the manufacturing process, refractions that exceed the field-of-view cannot be coded reliably. An example of such a failure case is shown in Figure 6.10. The same problem often occurs in parallax-

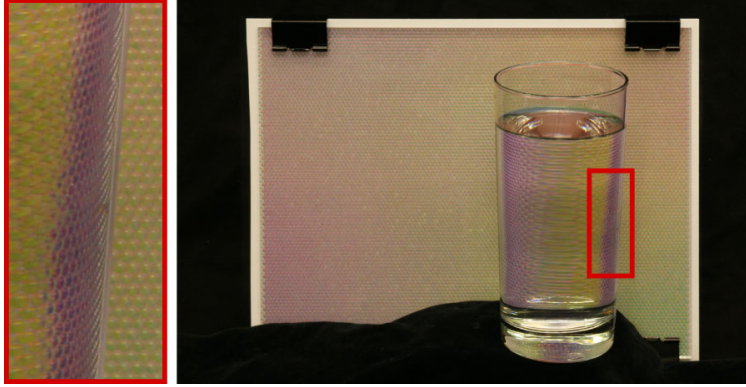


Figure 6.10: *Failure case: the field-of-view of the lenslet array is too narrow to properly encode the strong refractions near the sides of the glass. To overcome this, the lenslets should be chosen according to the amount of refraction in the scene.*

barrier or lenslet-based auto-stereoscopic displays. For our application, the lenslets for a specific experiment should be chosen in accordance with the expected amount of refraction. While the lenslets in our experiments (Table 6.2) are successful in capturing moderate to strong refractive events caused by liquids and solids, the precision and sensitivity of our current probes, which are made from off-the-shelf hardware, is currently too low to faithfully acquire the slight angular deflections within gas flows or shock waves.

6.6 Discussion

In summary, this chapter has presented a new approach to capturing refractive phenomena using light field probes. This approach presents a portable and inexpensive alternative to traditional and Background Oriented Schlieren imaging; it works well with strong refractions, which is often not the case for BOS, and also alleviates the focus mismatch between background and objects of interest. Inspired by Schlieren imaging and microscopy, we have shown how a variety of different filters for visualizing refractive events can be encoded in our light field probes and recorded with off-the-shelf cameras.

In the future, we would like to experiment with alternative technologies for 4D probes, which allow for a high spatial and angular resolution. We would also like to explore smaller probe designs using LED or OLED-based backlights instead of light boxes. Furthermore, it would be interesting to investigate more sophisticated colour coding schemes for the 4D light field probe space.

In addition to measuring refraction, 4D light field probes could be useful for a variety of other applications including BTDF and BRDF estimation, de-scattering, and separating local and global illumination.

Chapter 7

Refractive Shape from Light Field Distortion

The previous chapter has introduced high-dimensional light field probes and coding schemes for qualitative acquisition of transparent, refractive objects. In this chapter, we show that the distortion of these probes by refractive media can also be useful for quantitative measurements. Specifically, a new, single image approach to reconstructing thin transparent surfaces, such as thin solids or surfaces of fluids is presented. The proposed method is based on observing the distortion of 4D illumination emitted by a light field probe that contains both spatial and angular variation. Whereas commonly employed reference patterns are only two-dimensional by coding either position or angle on the probe, we show that the additional information can be used to reconstruct refractive surface normals and a sparse set of control points from a single photograph.

7.1 Introduction and Motivation

The reconstruction of transparent, refractive, and specular surfaces from photographs has been a target for active investigation in computer vision, but also other areas, including computer graphics and fluid imaging.

One strategy for dealing with such surfaces is to alter the reflectance or transmission characteristics of the surface under investigation to simplify the scanning. This can be achieved through coating with diffuse materials [Goesele et al., 2004] or immersion in special liquids [Hullin et al., 2008; Trifonov et al., 2006]. However, such intrusive methods are not always desirable or feasible, for example when the object under investigation is itself a liquid.

A popular alternative is to analyze the way in which the object distorts a diffuse background or illumination pattern [Agarwal et al., 2004; Ben-Ezra and Nayar, 2003; Bonfort et al., 2006; Kutulakos and Steger, 2005; Morris and Kutulakos, 2005, 2007; Murase, 1990; Savarese and Perona, 2002; Tarini et al., 2005]. Such approaches typically require multiple cameras,

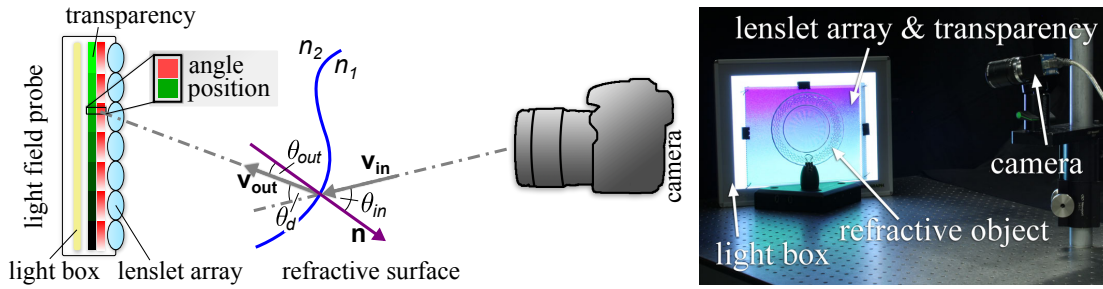


Figure 7.1: Schematic showing how both position and incident angle of a refracted ray are colour coded by a light field probe (left). Our probe prototypes consist of a light box, transparencies, and a lenslet array (right); these are positioned behind a refractive object when photographed.

or multiple images from the same camera taken with varying illumination or background patterns.

In the work presented in this chapter, we aim for a single camera, single image method more similar in spirit to photometric stereo [Woodham, 1980], and especially to single-image variants using coloured light sources [Hernandez et al., 2007]. We propose to reconstruct transparent surfaces from the observed distortion of higher-dimensional reference patterns, called light field probes. These probes are introduced in Chapter 6 and can encode the 2D spatial and the 2D angular domain on their surface; possible implementations include lenslet arrays, parallax-barriers, or holograms. The distortion of a light field emitted by such a probe allows us to simultaneously reconstruct the normals and a sparse set of absolute 3D points representing either a single refractive boundary surface or a thin refractive solid.

7.2 Shape from Light Field Probes

7.2.1 Coding Light Field Illumination

Light field probes are, as described in Chapter 6, capable of emitting 4D illumination by encoding the outgoing light ray positions and angles in varying intensities and colours. Standard displays only emit 2D illumination, because the light at each pixel is uniformly displayed to all directions. 4D probes can, for instance, be implemented by mounting high-resolution transparencies on a light box behind a lenslet array (see Fig. 7.1). This approach does not increase the total number of display pixels, but distributes them between spatial and angular resolution. The number of pixels under each lenslet corresponds to the angular

probe resolution, while the size of the lenslets determines the spatial resolution. Other hardware implementations, such as holograms, have the potential to overcome the resolution tradeoff of lenslet arrays.

For the purpose of single-shot transparent object reconstruction, the colour and intensity codes emitted by a light field probe need to satisfy two important criteria. First, the patterns are required to uniquely encode position and angle on the probe surface, so that a camera measures this information in a single image. Second, in order to account for slight miscalibrations of the probe prototype, the colour codes should be smooth in the 4D spatio-angular domain. We restrict our prototype to readily available hardware, as illustrated in Figure 7.1, and limit the feasible colours and intensities to the combined printer and camera gamut and dynamic range.

The most intuitive coding scheme satisfying the above requirements are colour gradients. In our implementation, we use red, blue, and green gradients to code the 2D directions and a 1D vertical position, respectively. As demonstrated in Section 7.2.3, the missing second spatial dimension can be recovered through geometric constraints in post-processing. This encoding is illustrated for a 1D case in Figure 7.1 (top). Here, the incident angle is coded in a shade of red and the position on the probe surface is coded in green.

This simple, yet effective coding scheme allows both angle and position of light rays to be encoded in observed colours and intensities. Without refraction in the optical path, the measured colours at each pixel of a calibrated camera correspond to the information predicted by the calibration, but in the presence of refraction these differ. In the following subsections we show how to reconstruct refractive surfaces from such measurements. The employed colour codes ignore the wavelength-dependency of refraction as well as attenuation and scattering caused by the medium.

7.2.2 Reconstructing Surface Normals

The normal of each surface point imaged by a camera pixel can be computed using Snell’s law: $n_1 \sin \theta_{in} = n_2 \sin \theta_{out}$. In our application, we seek the unknown normals given the incoming normalized rays \mathbf{v}_{in} , which are known from camera calibration, and the refracted ray directions \mathbf{v}_{out} , which are extracted from the imaged probe colour (cf. Fig. 7.1, left). The absolute angles θ_{in} and θ_{out} are unknown, but we can compute the difference between the two as $\cos \theta_d = \mathbf{v}_{in} \cdot \mathbf{v}_{out}$. For known refractive indices of the two media n_1 and n_2 , the angle between incoming ray and surface normal is then given as

$$\theta_{in} = \tan^{-1} \left(\frac{n_2 \sin \theta_d}{n_2 \cos \theta_d - n_1} \right). \quad (7.1)$$

Therefore, the surface normal \mathbf{n} can be computed independently for each camera pixel by rotating \mathbf{v}_{in} by the angle θ_{in} . The rotation is performed on the plane spanned by \mathbf{v}_{in} and \mathbf{v}_{out} , so

$$\mathbf{n} = \mathbf{R}(\theta_{in}, \mathbf{v}_{in} \times \mathbf{v}_{out}) (-\mathbf{v}_{in}), \quad (7.2)$$

where $\mathbf{R}(\theta, \mathbf{v})$ is a rotation matrix defined by angle θ around an axis \mathbf{v} .

7.2.3 Point Cloud Estimation

In order to triangulate absolute 3D surface points for each camera pixel, we need to determine the intersection of the lines $\mathbf{c} + t\mathbf{v}_{in}$ and $\mathbf{p} + s\mathbf{v}_{out}$. The camera position \mathbf{c} as well as the unrefracted ray directions \mathbf{v}_{in} are known from camera calibration and uniquely define a line in 3D space. The direction \mathbf{v}_{out} is estimated from the observed colours of the light field probe refracted by an object, however, only a single spatial coordinate is coded by the probe colour, i.e. p^y . Nevertheless, the intersection problem for the two lines results in a linear system with three equations and three unknowns p^x , s , and t because the origin of the coordinate system is defined on the plane of the probe, i.e. $p^z = 0$. Therefore, we can uniquely triangulate a 3D point per camera pixel as

$$t = \frac{1}{v_{in}^y - \frac{v_{in}^z v_{out}^y}{v_{out}^z}} \left(p^y + \frac{c^z v_{out}^y}{v_{out}^z} - c^y \right). \quad (7.3)$$

The triangulated positions are only numerically robust when significant refraction occurs along a ray; otherwise \mathbf{v}_{in} and \mathbf{v}_{out} are co-linear. At the same time, all measured ray directions \mathbf{v}_{out} will be noisy due to camera noise and possible colour nonlinearities of a fabricated probe. Therefore, we can only hope to robustly estimate a sparse set of 3D points from such measurements at camera pixels that observe a strong amount of refraction. The noise sensitivity of triangulated points is illustrated for a synthetic example in Figure 7.2.

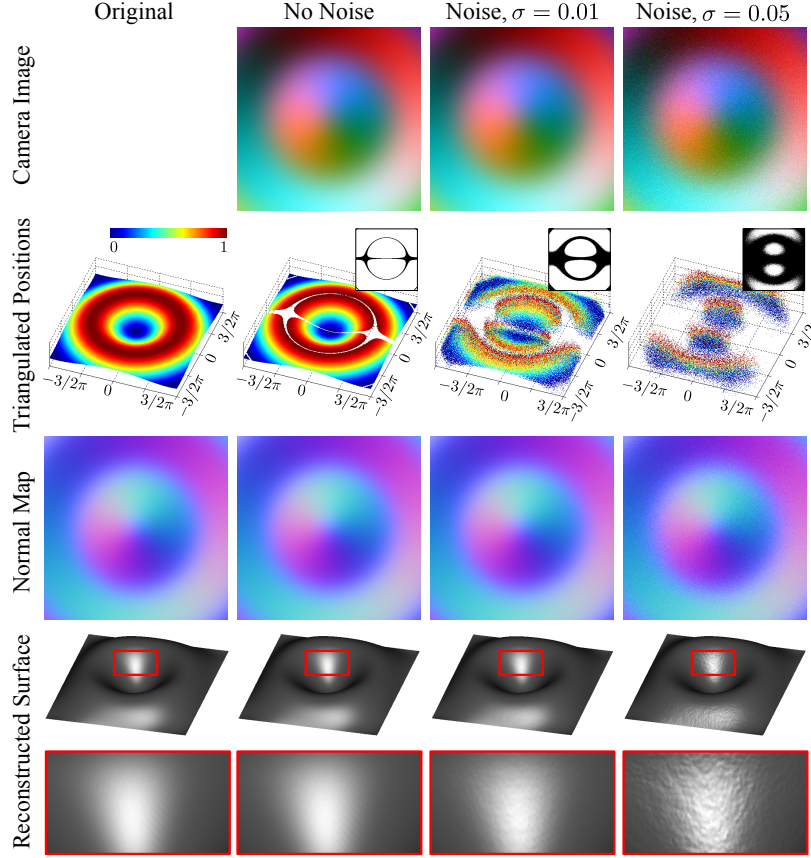


Figure 7.2: *Synthetic results for a refractive sinusoidal object. Normals and positions are shown for the original object (left column), and for reconstructions (other columns) from simulated camera images with an increasing amount of noise (top row).*

7.2.4 Surface Estimation from Normals and Points

While a normal field can be efficiently integrated to reconstruct surfaces (see e.g., Agrawal et al. [2006]), including an additional set of sparse 3D control points can remove ambiguities in these integration schemes [Horovitz and Kiryati, 2004]. For all of our reconstructions, we employ the integration method proposed by Ng et al. [2010], which uses an optimization with kernel basis functions.

We show synthetic results in Figure 7.2. Here, a sinusoidal function acts as the original surface with a refractive index corresponding to water; 3D positions and normals of the original surface are shown in the left column. We simulated photographs of an orthogonal camera that show the surface in front of a light field probe with the colour coding scheme discussed in Section 7.2.1 along with estimated normals, triangulated control points, and

final reconstructions. While the extracted normals are relatively resilient to an increasing amount of camera noise, the triangulated positions quickly become less reliable. Triangulated points that correspond to small angles between incoming and refracted rays for each pixel are masked out; the masks are shown in the insets of the second row.

7.3 Experimental Results

Our prototype (see Fig. 7.1, right) is composed of a light box, two stacked transparencies, a lenslet array, and a camera. The light box is LED-based, as opposed to fluorescent-based, in order to maintain consistent lighting throughout the capture process even when using a short exposure time, such as in video. The lenslet sheet is a FresnelTech hexagonal lenslet array with a focal length of 0.12" and a lenslet diameter of 0.09". The transparencies are printed with an Epson Stylus Photo 2200 printer at 1440 dpi, which, in combination with the lenslets, results in a theoretical angular resolution of 0.32° . This printer has six ink-based primaries; for improved contrast we stack two transparencies on top of each other. For still photographs we use a Canon D5 Mark II and for the videos a Prosilica EC1350C camera.

Intrinsic and extrinsic camera parameters are estimated in a pre-processing step using the method described by Atcheson et al. [2010]. The gamma curves of the printer are also estimated as a pre-processing step and compensated in the measurements.

Reconstructions of water surfaces are shown in Figure 7.3. Here, we positioned the probe underneath a rectangular water tank and filmed the scene from above (Fig. 7.3, rows 1 and 3). Secondary refractions from the glass tank bottom are negligible in this case. The results show a water drop falling into the tank in rows one and two; rows three and four depict water being poured into the tank. Some high-frequency noise is visible in the reconstruction, which is due to the printer half-toning patterns on the transparencies that become visible as noise on the probe when the camera is focused on it. Alternative printing technologies, such as light valve technology (www.bowhaus.com), could alleviate this problem.

Figure 7.4 shows reconstructions of three thin solid objects from a single photograph each. Although theoretically two refractive events occur for each camera ray, one at the air-glass interface toward the camera and another one at the glass-air boundary on the other side, the objects are thin enough that ray displacements within the glass are negligible. This is a common assumption for thin lenses. The reconstructed normals (Fig. 7.4, column 3) for these examples therefore show the difference between front and back normal of the surface; for the plate and the pineapple, the front side is flat and parallel to the fine details on the

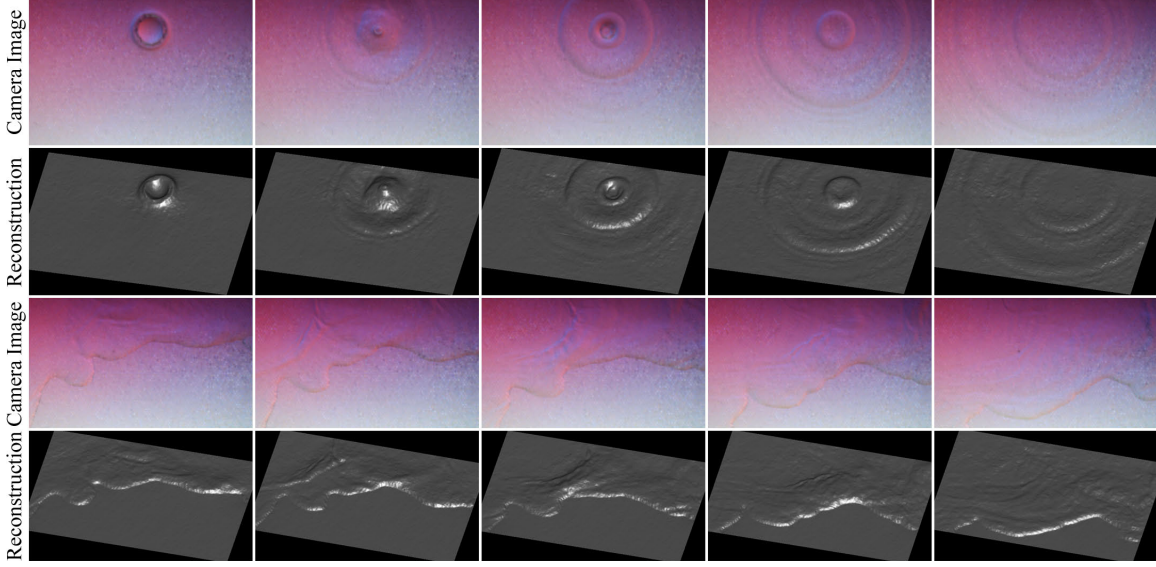


Figure 7.3: *Camera images and reconstructed surfaces of dynamic water surfaces. The upper rows shows a drop falling into the water, whereas the lower rows depict water being poured into the tank.*

rear side. The reconstructed surfaces (Fig. 7.4, right) only contain a flat triangle mesh and corresponding normals.

7.4 Evaluation

The acquisition of ground truth data for these objects is difficult. We qualitatively evaluate reconstructions of our prototype by comparing a rendering of the three lenses (see Fig. 7.4) with analytic descriptions of the same lenses in Figure 7.5. The diameters and focal lengths of these lenses are known and used to simulate them as bi-convex refractive surfaces in front of a textured background with POV-Ray (www.povray.org). The same procedure is used to simulate the reconstructed lens surfaces in front of the background. Slight differences in the lower left lens are mainly due to a violation of the thin lens model.

A quantitative evaluation of the proposed reconstruction algorithm with respect to camera noise and refractive index mismatches is shown in Figure 7.6. In this experiment, we simulate the acquisition and reconstruction of a 1D parabolic surface. An orthographic camera observes the scene from above with a light field probe illuminating it from the bottom. The surface represents the boundary between two media, the upper one is air and the lower one has a refractive index of $n = 1.5$. We add zero-mean Gaussian noise to the

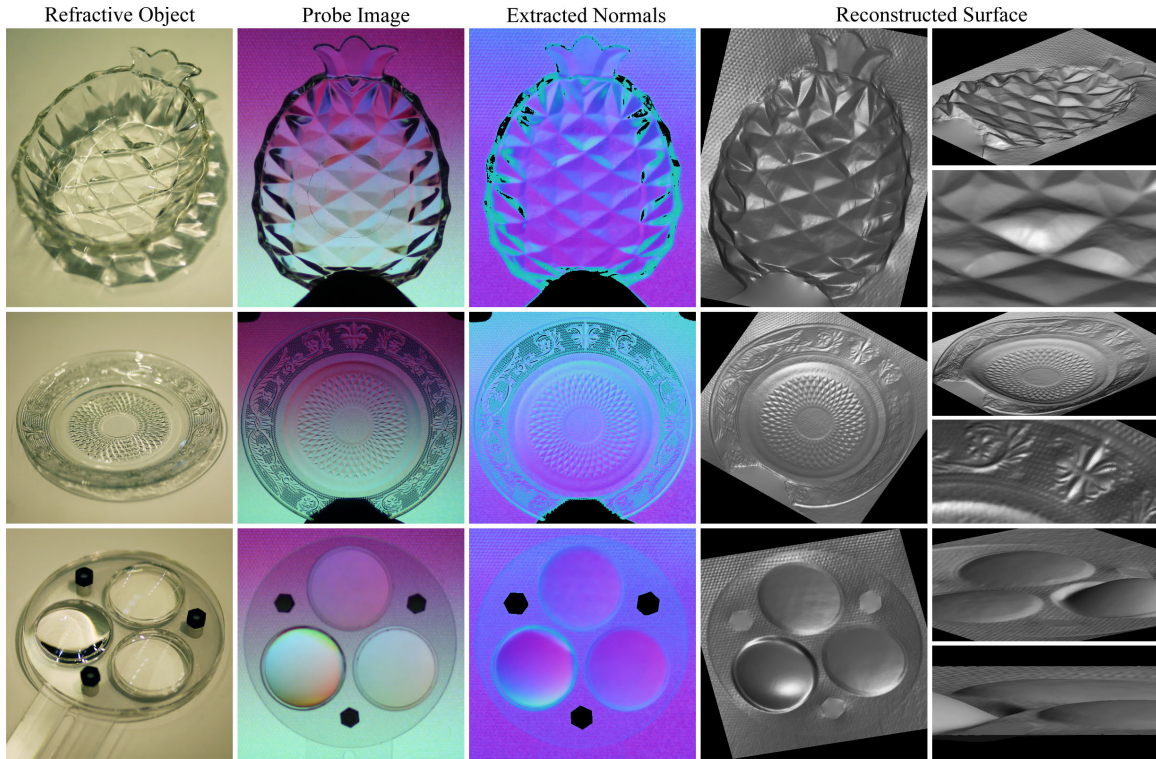


Figure 7.4: *Three thin refractive objects under room illumination (left column) and in front of a light field probe (center left column). The distorted colours of the probe allow us to estimate refractive surface normals from a single image (center row), which can be integrated to reconstruct thin shapes that approximate the geometry of transparent, refractive solids (right).*

simulated sensor measurements and evaluate reconstruction quality for different refractive index mismatches. Surface gradients (Fig. 7.6, center) are directly computed from the noisy sensor measurements and subsequently integrated to yield the actual surfaces (Fig. 7.6, left).

Based on these experiments, we can see that a mismatch in the refractive index results in a vertical shear of the gradients (Fig. 7.6, center, purple line), which corresponds to low frequency distortions of the actual surface (Fig. 7.6, left, purple line). The mean squared error (MSE) between original surface and reconstruction is particularly high when the assumed refractive index is lower than that of the medium (Fig. 7.6, top right, purple line). Furthermore, there is an approximately linear relationship between sensor noise and the noise observed in both reconstructed gradients and surfaces (Fig. 7.6, right). The mean squared error plots on the right of Figure 7.6 are averaged over 500 experiments, each exhibiting random noise.



Figure 7.5: Reconstructed and synthetic lenses from Figure 7.4 rendered as a refractive mesh in front of an image.

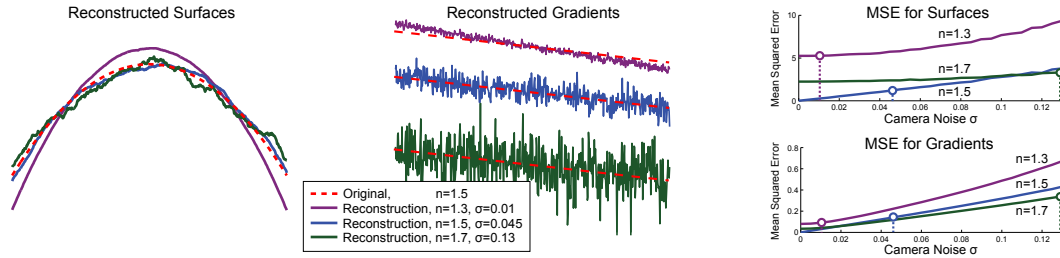


Figure 7.6: Evaluation of reconstruction with respect to noise and refractive index mismatch. A 1D parabolic surface (left, dotted red) is simulated to be captured with a light field probe and reconstructed with different amounts of camera noise and mismatches in the refractive index of the medium (left). While noise results in high frequency artifacts, a mismatch in the refractive index causes low frequency distortions. We show the mean squared error of surfaces (top right) and gradients (bottom right) for an increasing amount of sensor noise.

7.5 Discussion

In summary, this chapter has presented the first single image approach to thin refractive surface acquisition. Instead of analyzing the distortion of purely diffuse or purely angular reference background patterns, as done in previous work, we encode the angular and spatial dimensions of a light field probe simultaneously. The observed distortion of high-dimensional light fields allows us to reconstruct surface normals and triangulate a sparse set of control points from a single photograph. While the normals are relatively resilient to sensor noise and allow high-quality reconstructions, the triangulated control points are very sensitive to noise, but allow low-frequency ambiguities of the surface normals to be corrected.

7.5.1 Limitations

Our approach is currently mostly limited by the employed off-the-shelf hardware. Instead of using lenslet arrays and printed transparencies as light field probes, we expect much better results with alternative light field display technologies, such as holograms (www.zebraimaging.com). Furthermore, the lenslets have a limited field of view and introduce intensity variations over the probe surface, which become visible in the reconstructions; holograms could resolve this problem as well. Colour nonlinearities and cross-talk introduced by the printing process also affect the accuracy of the reconstructions.

Disregarding the prototype, our approach is fundamentally limited by the light field coding scheme and the reconstruction algorithm. Although the employed colour codes are optimized for single image reconstructions, attenuation and scattering within the medium as well as wavelength-dependency of refraction are assumed to be negligible. Alternative, dynamic codes can overcome these limitations at the cost of requiring multiple photographs. The proposed reconstruction algorithm requires the refractive index of the medium to be known and restricts light rays to refract only once in the scene. In combination with advanced coding schemes, novel algorithms could overcome these limitations as well.

7.5.2 Future Work

In the future, we would like to experiment with alternative technologies for fabricating light field probes, such as holograms, and test more sophisticated light field coding schemes. Applying temporal multiplexing with dynamic probes could lift current limitations; multi-spectral displays and cameras could improve the amount of coded information as well. We would like to explicitly separate attenuation and refraction caused by the medium and test our approach with multi-camera, multi-probe configurations.

Chapter 8

Optical Image Processing using Light Modulation Displays

In this chapter, we explore the potential to enhance the power of the human visual system by applying on-the-fly optical image processing using a spatial light modulation display. We introduce the concept of see-through optical processing for image enhancement (SOPhIE) by means of a transparent display that modulates the colour and intensity of a real-world observation. The modulation patterns are determined dynamically by processing a video stream from a camera observing the same scene.

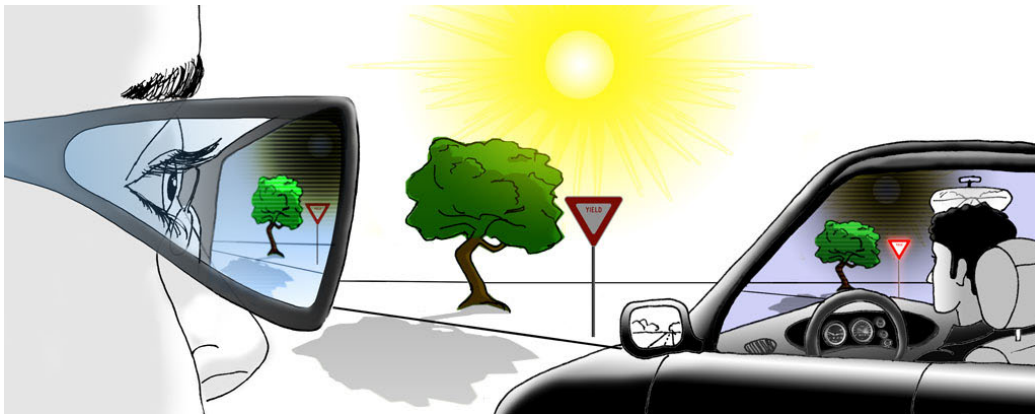


Figure 8.1: A conceptual illustration of our approach. A light modulation display locally filters a real-world scene to enhance the visual performance of a human observer, in this case by reducing the contrast of the sun and boosting the saturation of the traffic sign for a driver. Our approach would apply to see-through scenarios such as car windshields and eye glasses as depicted, as well as to binoculars, visors, and similar devices.

8.1 Introduction

The human visual system (HVS) is a remarkable optical device possessing tremendous resolving ability, dynamic range, and adaptivity. The HVS also performs an impressive

amount of processing in early (preattentive) stages to identify salient features and objects. However, the HVS also has some properties that limit its performance under certain conditions. For example, veiling glare due to extremely high contrast can dangerously limit object detection in situations such as driving at night or driving into direct sunlight. On the other hand, conditions such as fog or haze can reduce contrast to a point that significantly limits visibility. The tri-stimulus nature of human colour perception also limits our ability to resolve spectral distributions, so that quite different spectra may be perceived as the same colour (*metamers*). Any form of colour blindness exacerbates the problem.

We propose to enhance the power of the human visual system by applying on-the-fly optical image processing using a spatial light modulation display. To this end, we introduce the concept of see-through optical processing for image enhancement (SOPhIE) by means of a transparent display that modulates the colour and intensity of a real-world observation. The modulation patterns are determined dynamically by processing a video stream from a camera observing the same scene. Our approach resembles and builds on work in computational photography and computer vision, but we target a human observer rather than a camera sensor, and our goals are thus perceptual rather than photographic. The work presented in this chapter also resembles traditional ‘see-through’ augmented reality (AR) displays, which present the observer with an image that shows synthetic imagery overlaid (i.e. added to) the real world scene. Unlike optical AR displays, we use the display to spatially *filter* the incoming light at the observer’s position, allowing us to perform image processing operations such as contrast reduction, contrast enhancement, or colour manipulation without the latency introduced by video-based AR systems.

Potential applications for such an approach, once perfected, range from tone-mapping sunglasses and ski goggles to ‘smart’ automobile windshields that reduce dangerous contrast, enhance low contrast, or subtly draw attention to important features such as street signs or an erratic vehicle. We take the first steps toward such ambitious applications by demonstrating several examples of optical processing in a prototype setup consisting of a monocular scope-like device (see Section 8.2.1). Our primary prototype enables the display and the observed scene to be in focus, and enables the camera and observer to share an optical axis. We also envision and analyze different scenarios such as the aforementioned glasses or car windshields that place the display out of focus and impose a parallax between the camera and observer.

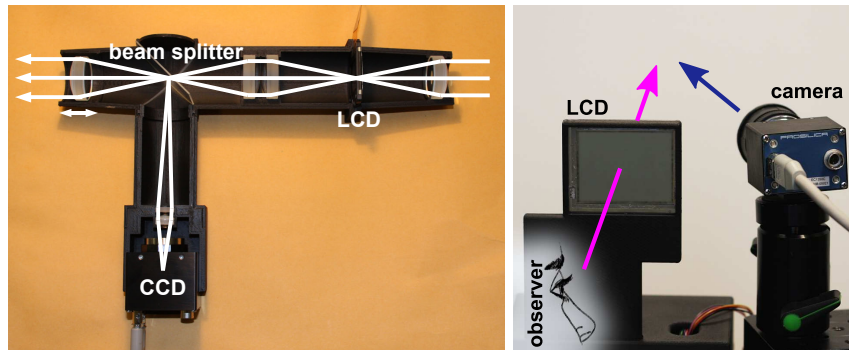


Figure 8.2: *Two prototypes. Left: the opened scope system with optical paths indicated. Right: the window-style setup is simply an LCD panel stripped of its backlight in a custom housing with the camera located next to it.*

8.2 Prototype Designs

To demonstrate the feasibility of the SOPhIE approach, we have implemented two physical prototypes (Figure 8.2) that let us experiment with different algorithms. The first one is a scope, and the second one a small see-through window-like configuration that can simulate, for example, one side of a pair of glasses. We do not deal with binocular stereo parallax that occurs when both eyes of the user look through the same display. For windshields or visors, this issue will eventually have to be addressed, but we believe this will be possible given HVS characteristics such as the dominant eye effect. We leave the investigation of such solutions for future work.

8.2.1 Monocular Scope Prototype

Our first prototype setup is a scope-like system in which a lens assembly brings both the scene and a see-through LCD panel into focus simultaneously. A beam splitter ensures that the camera shares the optical axis of the system. Both the camera and the LCD panel are greyscale-only in this setup.

LCD Display

The LCD panel was taken from a Sharp PG-D100U video projector, which uses three such panels, each with a resolution of 800x600, to project a colour image. The control electronics were left in the projector housing, which was connected to the scope via a ribbon cable. Focusing optics for the display were assembled from four achromatic doublet lenses.

Camera

The camera, a 1.5 megapixel C-mount camera from Prosilica, observes the scene through a beam-splitter. We successfully used both a standard half-silvered mirror and a reflective polarizer for this purpose. The reflective polarizer has the advantage of minimizing light absorption through the LCD panel, as it pre-polarizes the light transmitted through it. The camera image is formed by two additional achromatic doublets that re-image the image projected by the front-most, i.e. object-side, lens of the assembly.

Assembly

We used a rapid prototyping machine to create a custom housing for the components of the scope. The white ABS plastic was subsequently spray-painted black to minimize scatter. The housing allows the front-most lens to be moved along the optical axis for focusing the scope at different depths.

Calibration

The response curves of both the camera and LCD panel are measured and compensated for using standard techniques [Reinhard et al., 2010]. Geometric calibration and alignment procedures are performed by replacing the human observer with a second camera, and manually aligning the captured images using a model of affine transformation plus radial distortion.

8.2.2 Window-style Prototype

Our second prototype allows a user to directly observe a scene through a colour LCD panel without refocusing optics. The camera, a colour version of the same Prosilica model used in the monocular scope, is located off-axis in this setup and the system corrects for the resulting parallax. This second prototype allows us to analyze the issues associated with target applications such as car windshields, sunglasses, or helmet visors where the relative positions of eyes and display are fixed.

Such setups can suffer from colour fringing and other diffraction artifacts caused by small pixel structures. This problem can be avoided by using displays with large pixels. The resulting lower resolution can be tolerated for our application, since the display will generally be much closer to the human observer than the scene under investigation. This causes the

display to be strongly blurred due to defocus, so that the display resolution is not a primary concern.

For our prototype, we chose a 2.5” active matrix TFT panel with a resolution of 480×234 pixels from Marshall Electronics (V-LCD2.5-P). In order to turn this display into a see-through spatial light modulator, we removed the backlight. Like most LCD displays, the Marshall panel is coated on the front surface with a film that acts as a polarizer, but also shapes the directional distribution of emitted light. Since this film prevents see-through applications by blurring the scene, we replaced it with a generic polarizer.

Geometric and radiometric calibration of this second setup proceed as for the first setup. A radial distortion term is not necessary due to the lack of refocusing lenses.

8.3 Applications

The SOPhIE framework enables a variety of applications, which are explored in this section.

8.3.1 Contrast Manipulations

Direct View Gamma Curves

A straightforward use of SOPhIE is the application of a nonlinear ‘response’ function to the view. For example, a gamma curve with $\gamma < 1$ can be used to reduce contrast in environments with harsh lighting, while $\gamma > 1$ could boost contrast, for example on an overcast day. Gamma adjustments are frequently used to boost contrast on conventional displays. For example, the inverse gamma curve used by video standards deliberately differs from the gamma of physical output devices, in order to improve contrast [ITU, 1990]. SOPhIE enables the same method for use in direct observations by human viewers. To apply a response curve such as a gamma curve, we record a normalized image I of the scene using the SOPhIE camera. The ratio image I^γ/I is the desired transmission of the LCD panel.

Contrast Reduction

A gamma curve results in a relatively subtle compression or expansion of contrast. For more dramatic contrast reduction, we can apply strategies similar to tone mapping (see for example Reinhard et al. [2010]). One simple approach is to display the inverse of the image

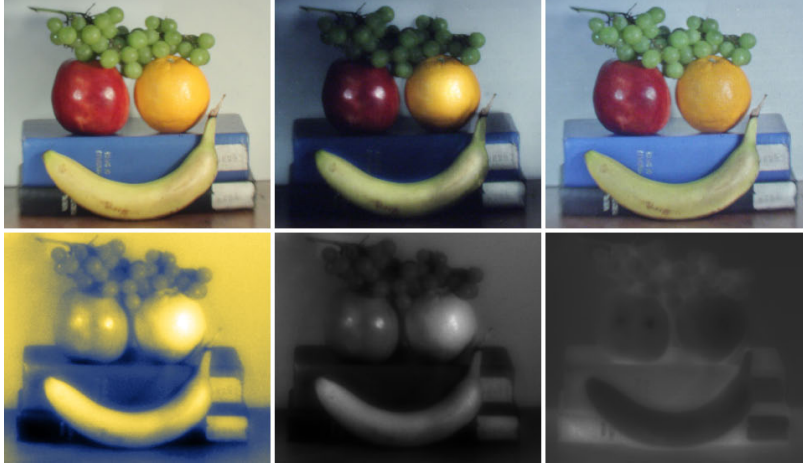


Figure 8.3: *Optical gamma modulation. The top row shows photographs taken from the point of view of a human observer using SOPhIE. Left: unmodified image. Center and right: a gamma of 2.5 and 0.625, respectively. Bottom left: false-colour rendition of the image observed by SOPhIE’s camera. Bottom center/right: modulation images shown on the SLM to produce the gamma curve applications above.*

seen by the camera on the LCD panel, so that brighter regions in the camera image become darker (i.e. more heavily attenuating) regions on the see-through display. Alternatively, we can leave most of the scene untouched, but dim the regions with very bright light sources or specular reflections. Both approaches are shown in Figure 8.4.



Figure 8.4: *Contrast reduction. Three photographs taken from the point of view of a human observer. From left to right: no correction; correction with an inverted and blurred camera image; and darkening a blurred region around saturated pixels. The insets show the corresponding modulation patterns.*

Contrast Enhancement with Countershading

Since we cannot amplify light with our setup, we can alternatively exploit the special characteristics of the HVS to achieve images that appear to have increased contrast even though

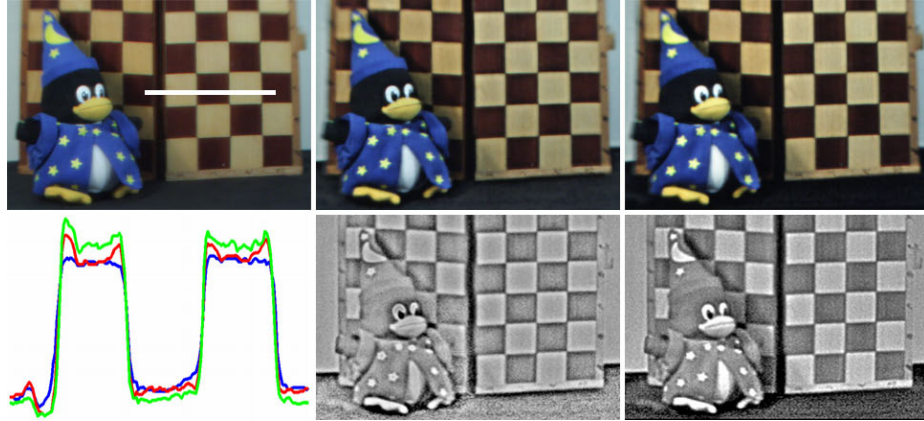


Figure 8.5: *Contrast enhancement. Top row: photographs taken from the viewpoint of a user with (from left to right) no contrast enhancement, unsharp-masking, and countershading. Bottom left: a cross-section of the intensity profile for the scanline marked in white (blue: original image, red: unsharp-masking, green: countershading). Bottom center/right: modulation images shown on the SOPhIE display.*

the actual dynamic range is unchanged. Two examples of such methods are the familiar *unsharp-mask* filter, and the more sophisticated *countershading* approach introduced by Krawczyk et al. [2007]. Both methods rely on the *Cornsweet effect* [Kingdom and Moulden, 1988], in which the perceived difference in intensity between adjacent image regions can be amplified by exaggerating the difference in the boundary region only. The unsharp-mask filter provided by many image manipulation packages is an ad-hoc realization of this effect, while adaptive countershading is based on a principled analysis of the spatial frequencies in the image, and how the human visual system perceives them.

In the SOPhIE system, we experimented with both unsharp-masking and countershading. As seen in the intensity profiles in Figure 8.5, unsharp masking amplifies the contrast by modifying intensities in one of the higher frequency bands, while countershading alters multiple spatial frequency bands at the same time. The unsharp-mask implementation is similar to the method discussed before; we compute the ratio image of the processed and unprocessed camera image, and show the result on the LCD display. In order to ensure that the scanlines can be compared with the same exposure settings, the unmodified case was captured with the LCD transmission set to a constant value of 0.8.

8.3.2 Colour Manipulations

The applications discussed thus far only require modulating intensity. Next, we discuss applications that become possible with the addition of colour displays and cameras.

Manipulations of Colour Saturation

An intuitive application of a colour system is the spatial manipulation of colour saturation. For example, we can reduce the colour saturation in the following way. For each pixel, the camera observes an RGB colour. We set the transmission of the colour channel with the smallest value to one, but reduce the transmission of the other two channels such that they match the smallest channel value. Likewise, we can boost saturation by choosing a transmission value of one for the dominant, i.e. largest, channel, but reducing the transmission for the other channels by a certain factor.

Note that this approach enables us to drastically reduce colour saturation, but not, in general, to completely eliminate all colour in the scene. The SOPhIE camera senses colour with RGB sensor arrays and filters it with RGB displays. Like all tri-stimulus systems (including the HVS) such an approach permits metamers, i.e. different spectral distributions that result in an identical sensor response. The RGB filters in the SOPhIE system have different spectral distributions from the S, M, L photoreceptor types in the human eye, so different metamers with a given sensor response would require slightly different RGB filter settings to completely remove all perceived colour from the image.

Since our monocular scope prototype does not have a colour display, we demonstrate the method with the second prototype, which has a parallax between camera and observer. Results are shown in Figure 8.8 (right, top row). They demonstrate that it is possible to drastically reduce colour saturation with a coloured SOPhIE system, although some residual colour remains due to the effects discussed above.

Colour De-metamerization

De-metamerization addresses the related problem of making metamers of a single perceived colour visually distinct for a human observer. We can achieve this in a straightforward manner by using the colour display as a spatially uniform, but programmable colour filter. Figure 8.6 shows a result of this approach, again using the second prototype with a parallax between camera and observer. The exposure times of the right photograph was adjusted, simulating intensity adaptations performed by the HVS.

Colour Deficiency

A significant portion of the population suffers from some kind of colour vision deficiency. The most widespread such deficiency is deuteranopia (red-green colour blindness), which



Figure 8.6: *De-metamerization: Coloured flowers that appear similar under orange illumination (a). Modifying the colour transmission of the LCD reveals the visual differences (b).*

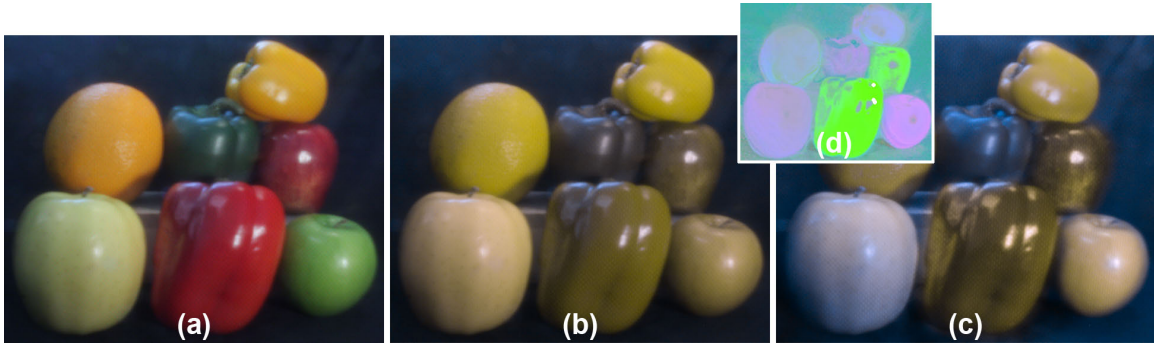


Figure 8.7: *Using SOPhIE for aiding colour-deficient viewers. (a) shows the original scene and (b) the scene as perceived by a deuteranopic colour-deficient viewer. Modulating the colours of the SOPhIE display as shown in (d) preserves visual differences for colour blind persons (c). All images except (d) are photographed through the SOPhIE prototype assembly and images (b) and (c) are post-processed to simulate the perception of a deuteranopic viewer.*

results from a lack of medium-wavelength sensitive cones. Brettel et al. [1997] showed how to simulate images as perceived by colour-deficient viewers. For a given RGB image (Figure 8.7 (a)), the view of a deuteranopic dichromat is simulated in Figure 8.7 (b). Note how the colours of various objects in the scene become indistinguishable.

We can optically modulate the colours of a real-world scene so that visible differences are preserved for people suffering from colour blindness. To this end, we employ the algorithm introduced by Rasche et al. [2005] for re-colourization to calculate a desired colour-modified image from a captured photograph. Dividing the former by the latter and accounting for the camera and display response curves enables us to compute a compensation image that is displayed on the SOPhIE LCD as seen in Figure 8.7 (d). A deuteranopic person would not be able to perceive the original colours, but could discriminate individual objects

which would appear similar otherwise (Figure 8.7 (c)). Again, the exposure time of (c) was increased.

8.3.3 Object Highlighting Using Preattentive Cues

Most of the above examples apply the same image processing operation uniformly across the image. These methods can be made spatially varying by making use of any external object recognition or tracking method. For example, de-metamerization trivially extends to spatially selective de-metamerization, that is if one wished to de-metamerize camouflage from foliage but leave the colour of the sky perceptually unmodified. More generally, spatially varying processing enables us to highlight objects of interest and to shift the user's attention. The effect ranges from subtle to quite dramatic.

For example, by increasing contrast in regions of interest we obtain an effect much like depth of field-driven attention in cinematography. On the other hand, reducing brightness or colour saturation outside the regions of interest is a very strong way of directing attention (see Figure 8.8). Significant brightness [Beck et al., 1983; Healey and Enns, 1999] and colour [Healey and Enns, 1999; Nagy and Sanchez, 1990] differences trigger preattentive processing mechanisms in the human visual system, so that in effect visual search tasks are performed in parallel rather than sequentially [Beck et al., 1983; Nagy and Sanchez, 1990]. As a result, the regions of interest stand out strongly even over very cluttered surroundings.

In our demonstration system, we highlight regions determined by either manual selection or simple background subtraction to identify regions of change. However, one could easily use dedicated tracking or recognition systems to highlight, for example, people, faces, or street signs. In multi-use collaborative systems, a user could use an interactive interface to draw the attention of other users to specific features.

8.3.4 Defocused Light Modulation

Conceptually, response function manipulations are operations that require a one-to-one correspondence between display pixels and scene points, and thus an in-focus display. However, in practice most algorithms also work very well with blurred ratio images, and thus out-of-focus displays. One exception is the countershading approach, which requires precise alignment between display and world, and thus does not work with out-of-focus settings. All other algorithms presented here, however, can be used with window-like or near-eye setups without refocusing optics, as required for glasses, helmets, or windshields. The ability to use defocused light modulation also makes the SOPhIE approach quite robust under

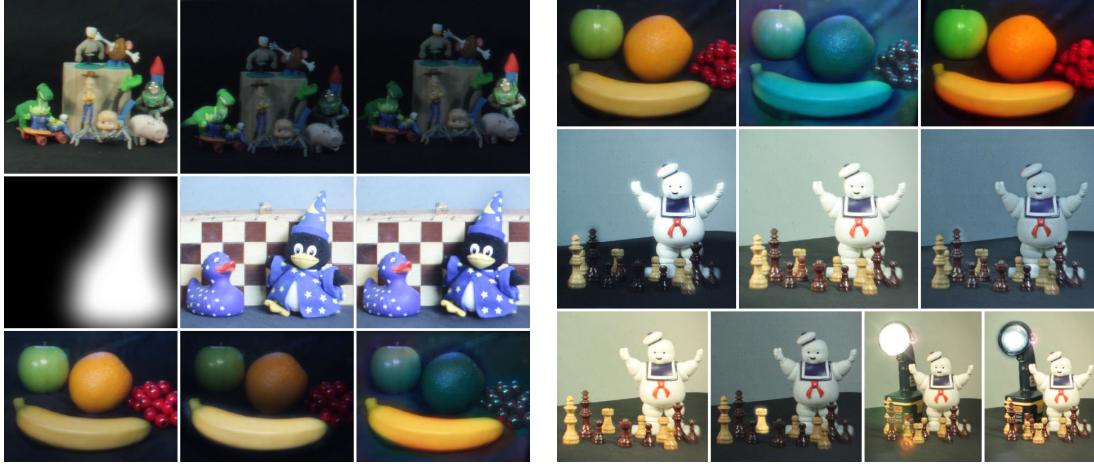


Figure 8.8: *Left side: example of object highlighting. Top row: object highlighting by darkening all but the region of interest. Middle row: more subtle direction of attention (right) by selectively sharpening some regions (left) of the image (center). Bottom row: highlighting by brightness (center) and colour (right) manipulations using our second setup. Right side: results from parallax experiments: colour saturation changes using our second prototype (top), synthetic gamma (center row), object highlighting (lower left), and contrast reduction (lower right).*

misalignment between the see-through display and the real world. Consequently, we can handle setups in which a parallax exists between camera and observer, a situation unlike most results described so far, in which the camera shared the optical axis with the human observer. This feature is again important for systems such as glasses or windshields. Figure 8.8 (right) shows several examples of out-of-focus photographs with a 5° parallax between viewer and camera.

Finally, and crucially, the robustness under misalignment produces a resulting robustness under motion. Even though our configuration has a system latency comparable to optical see-through augmented and mixed reality systems – or worse, since we have invested little effort in reducing latency – this does not translate into a noticeable misalignment if the modulation image is defocused or synthetically blurred. As a result, SOPhIE successfully processes dynamic scenes with moderately fast-moving objects without apparent latency, a key advantage over watching a processed video stream on a standard display.

8.4 Evaluation with User Study

To validate the effectiveness of the SOPhIE framework, we picked one of the proposed applications for evaluation with a user study. The application we chose is contrast reduction

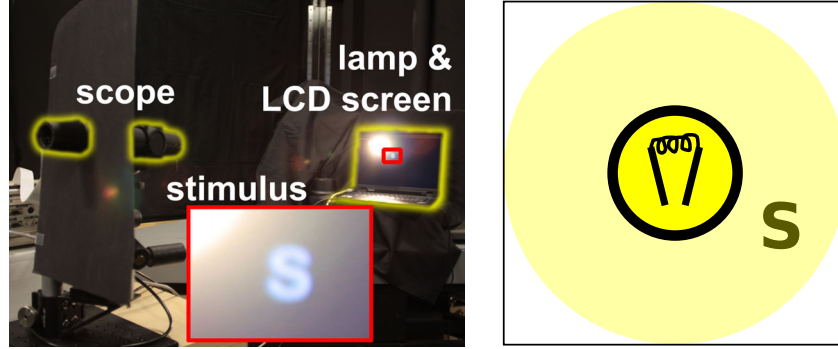


Figure 8.9: A photograph of the physical setup of our user study with its different components highlighted (left). The schematic on the right shows what a human observer sees through the scope: a character being displayed on the screen and veiled by glare created in the eye of the observer by the bright light source.

according to Section 8.3.1. We selected this application due to its wide applicability in all possible hardware embodiments of SOPhIE, including both binoculars and scopes, as well as sunglasses or windshields. We leave a full evaluation of the other application scenarios for future work.

For our study, we set up a bright light source and an LCD screen that subjects had to observe through our scope prototype (Figure 8.9). The visual path was blocked such that the only means to see the screen was through the scope. One hundred random characters, that were closely located to the light source, were presented to each of 12 subjects with normal or corrected-to-normal vision. Each character flashed for a fixed duration of 0.5 seconds. The lamp had a size of 2 visual degrees, the characters 0.7 degrees, and the distance from the center of the light source to the center of the stimuli was 2 degrees.

For every character, we randomly enabled or disabled our contrast reduction. The characters were shown on a black background in five different intensity levels, randomly selected for each character, with Weber contrasts of 5.25, 8.55, 14.2, 45.6, and 102 respectively. We measured the Weber contrast of the light source with respect to the background as 14900. Weber contrast is given as $(L - L_b) / L_b$, where L is the luminance of the stimulus and L_b that of the background. Weber contrast is the preferred unit for non-grating stimuli presented on uniform backgrounds.

Results of our study are shown in Figure 8.10. The detection probability is the number of correctly recognized characters divided by the number of displayed characters for each of the 10 different settings. The plot shows average detection probability, and according variances across the subjects (left). Psychometric functions (as seen on the right) were fitted using the *psignifit* toolbox for Matlab (see bootstrap-software.org/psignifit/) which implements

the maximum-likelihood method described by Wichmann and Hill [2001]. The expected detection probability of 0.5 could be shifted from a Weber contrast of 42 without modification to 5 using our contrast modulation technique, which is a significant improvement.

As expected, the light source creates glare that is caused by light scattering in the human eye. Effectively, this increases the perceived intensity around the lamp and masks details in the veiling glare. Instead of uniformly darkening the entire scene like regular sunglasses, our contrast reduction decreases the light intensity for only the brightest parts. This unveils imperceivable details that would otherwise be hidden in the halo of the light source created by scattering in the human visual system.

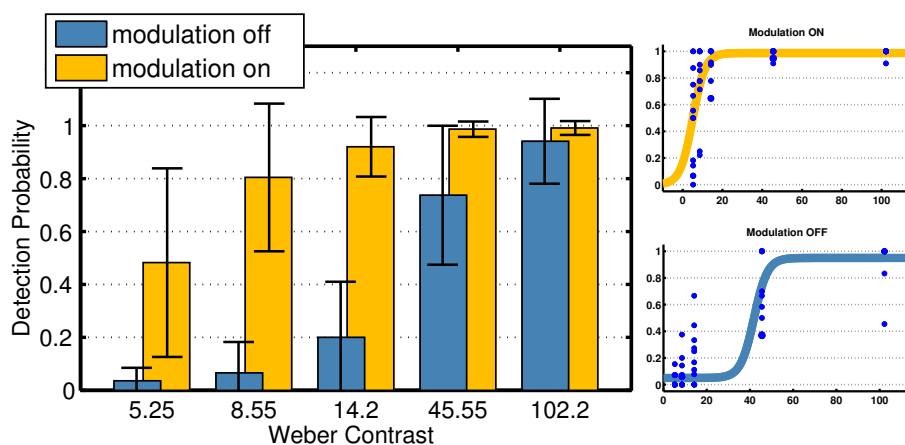


Figure 8.10: *The results of our user study confirm that details in high contrast scenes that are imperceivable due to veiling glare can be unveiled with our framework. These plots shows average stimulus detection probabilities and variances (left) as well as psychometric functions fitted to the data (right).*

The acquired data supports our arguments; stimuli of lower contrast can be veiled by objects of higher contrast. In situations such as driving, the sun or headlights of other cars can limit a driver’s visual performance. Traffic signs or other crucial information may not be perceived as intended. By selectively reducing the contrast of a scene as proposed in this chapter, we can significantly reduce glare created in the eye, and enhance the visual performance of a human observer.

8.5 Discussion

8.5.1 Feasibility and Limitations

The sunglasses concept presented in Figure 8.1 may seem futuristic, but all the necessary ingredients are already present in an ubiquitous computing platform: modern mobile phones contain small liquid crystal displays, very small cameras, and low-power processors suited for image processing calculations. As more processing power becomes available, sunglasses seem quite feasible as a SOPhIE platform in the near future. Work in augmented reality has already demonstrated that these components can be integrated into wearable systems [Mann, 1997]. Other scenarios, such as automobile windshields, still face some limitations of the underlying technology that must be addressed. Here we discuss the challenges in display, tracking, and processing technology that face a practical implementation of the SOPhIE approach.

Display Technology

Liquid crystal displays are a mature, mass-produced technology, but have several disadvantages for our purposes. First, they rely on polarization of the incoming light and thus immediately cut the incoming intensity by half. This may prove acceptable for sunglasses, or related scenarios like ski goggles, but not for windshields. Our prototypes have a light throughput of about 10%. For context, modern greyscale LCD panels have maximum transparencies of up to 45%, considerably more transmissive than most sunglasses (5%-20%).

Mass-produced LCDs have increasingly high resolution. However, for strongly defocused settings such as sunglasses this resolution is unnecessary, as fine details are blurred away, and even undesirable, since a small dot pitch causes colour banding due to diffraction. We viewed scenes through pixel grids printed onto transparencies to empirically determine that a minimum pixel size of about 0.02 mm avoids visible diffraction artifacts in such out-of-focus settings. LCD panels designed for SOPhIE applications would require fast response, low resolutions, high contrast, and for some applications greyscale rather than colour—an unusual combination not currently mass-produced. Finally, the flat nature of current LCDs could hamper the ergonomic and fashion design of products like sunglasses; however, companies are beginning to offer curved and flexible LCD panels [Crawford, 2005].

Other technologies exist but possess their own limitations; for example, commercial electrochromic products offer an excellent attenuation range but are slow and power hungry.

The reflective micromirror arrays used by Nayar et al. [2006] would be ergonomically awkward and suffer diffraction problems in defocused situations.

Real-Time Image Processing

Clearly any image processing performed by the SOPhIE system must be done in real time. This limits the algorithms we can consider employing. However, modern mobile processors are extremely powerful, particularly for algorithms with dedicated fixed-function support. For example, the NVIDIA Tegra processors perform H.264 720p video decode at 30 frames per second, using well under 200 milliwatts full-chip [NVIDIA Tegra, 2008]. Furthermore, many of our algorithms operate at relatively low resolutions because of the eventual defocus blurring of the already low resolution display. In our prototypes almost all of the presented algorithms run at several hundred frames per second on current GPUs and thus have no measurable impact on latency. The exceptions are countershading and Rasche’s re-colourization algorithm [Rasche et al., 2005], for which we currently use unoptimized CPU-based implementations.

The latency of our prototypes is fairly high due to the loose coupling of cameras and displays with a normal desktop PC. Specifically, the scope and the window-style prototype have system latencies of 72 ms and 61 ms respectively. In a commercial handheld system these latencies would be drastically reduced, because the display and camera would be custom-designed and much more tightly integrated with the processor. Nevertheless, even our current prototypes proved sufficiently fast to handle speeds up to 1 m/s at a distance of 3 m without noticeable visual artifacts.

Head and Feature Tracking

We assume in all of our applications that the relative positions of the camera and display are fixed and calibrated, and that the eye position relative to the display is either fixed, or known approximately through other means. These assumptions hold in any setup in which the optical system is either held to the head, such as binoculars and monocular scopes, or attached to the head in a fixed fashion as with eye glasses or helmet visors. Larger systems like windows or car windshields could use head tracking, although multi-user scenarios require further research.

Our object highlighting applications may require more sophisticated real-time recognition and tracking of important features, which imply more daunting computational requirements. However, advances in graphics hardware and computer vision algorithms appear promising.

For example, real-time face tracking is available commercially [Seeing Machines, 2008] and robust detection and tracking of pedestrians is a ‘hot topic’ in computer vision [Sabzmejdani and Mori, 2007]. As processors continue to evolve, it seems reasonable to imagine such applications running in real-time on mobile hardware.

8.5.2 Conclusions and Future Work

In summary, we have presented a novel approach for improving the performance of the human visual system by spatially modulating the light incident at a human observer. Applications such as contrast enhancement and reduction, colour manipulation, and object highlighting could help humans process visual information more effectively, as demonstrated by our user study. This approach could someday help reduce risk in safety critical applications.

We have evaluated the feasibility of both a scope-like setup, and a thin device, such as eye glasses. The basic technologies for lightweight, mobile SOPhIE systems exist today; no significant technical hurdles prevent implementations of such devices. Advances in display technology and processing power will increase the reach and application of the SOPhIE approach. We plan to investigate the feasibility of more complex optical setups and algorithms including spatial convolution and nonlinear operations.

In the future we will further customize specific algorithms and applications for SOPhIE, and extensively test and evaluate these with more user studies. A detailed analysis of binocular parallax effects is also left for future work. We believe that the dominant eye effect and other properties of the human visual system can be effectively exploited to produce high quality results in setups where both eyes share the same modulation display.

We also plan to investigate applications of the SOPhIE approach to future AR display systems. Combining traditional additive AR features with our multiplicative attenuation technique to achieve effects such as consistent illumination[Bimber et al., 2003] appears natural; many such tasks, while difficult to realize with either additive or multiplicative display in isolation, become feasible when using both at once.

Chapter 9

Plenoptic Multiplexing and Reconstruction

Photography has been striving to capture an ever increasing amount of visual information in a single image. Digital sensors, however, are limited to recording a small subset of the desired information at each pixel. The most common approach to overcoming the limitations of sensing hardware is the optical multiplexing of high-dimensional data into a photograph. While this is a well-studied topic for imaging with colour filter arrays, we develop a mathematical framework that generalizes multiplexed imaging to all dimensions of the plenoptic function. This framework (Sec. 9.2) unifies a wide variety of existing approaches to analyze and reconstruct multiplexed data in either the spatial or the frequency domain. We demonstrate many practical applications of our framework including high-quality light field reconstruction (Sec. 9.3), the first comparative noise analysis of light field attenuation masks (Sec. 9.5), and an analysis of aliasing in multiplexing applications (Sec. 9.4).

9.1 Introduction and Overview

As outlined in Chapter 1, the “ultimate” camera would capture all plenoptic dimensions in a single image using *plenoptic multiplexing*. This can be achieved with something as simple as a colour filter array or, more generally, consider additional plenoptic quantities [Narasimhan and Nayar, 2005]. In either case, a full-resolution image is computed from an interleaved sensor image by interpolating the captured data. Alternatively, an encoding of the spatio-angular plenoptic dimensions, commonly referred to as light fields [Levoy and Hanrahan, 1996], can be achieved by multiplexing directional light variation into spatial frequency bands using optical heterodyning [Lanman et al., 2008; Veeraraghavan et al., 2008, 2007].

In this Chapter, we introduce a mathematical framework for describing and analyzing plenoptic multiplexing systems. This allows us to cast a large variety of existing multiplexed imaging approaches into a common framework for analysis, reconstruction, and

performance evaluation.

9.1.1 Overview of Benefits and Limitations

The framework introduced in Section 9.2 shares limitations of other image processing methods, such as colour demosaicing: the captured sensor images are assumed to be composed of repeating *super-pixels*. Each of these super-pixels contains different samples of the plenoptic function, but the sampling layout within the super-pixels is spatially invariant. While standard colour filter arrays (CFAs, e.g. Bayer [1976]) only perform an interpolation of the spatially interleaved samples, our framework targets more sophisticated multiplexing schemes that require additional data processing after the interpolation. A general assumption for all such approaches is that the sampled signal is band-limited. In practice, this is achieved using optical anti-aliasing filters (e.g., Greivenkamp [1990]).

As illustrated in Figure 9.1, we demonstrate that our image formation allows for reconstructions, that is interpolation and subsequent processing, of multiplexed data in both the spatial and Fourier domain. Although this may seem straightforward for some applications, such as colour demosaicing, a variety of mask-based light field acquisition approaches have recently been proposed with corresponding analyses and reconstructions being exclusively performed in the Fourier domain [Agrawal et al., 2010b; Georgiev et al., 2008; Lanman et al., 2008; Veeraraghavan et al., 2008, 2007]. Our framework is the first to generalize optical multiplexing to all plenoptic dimensions and to demonstrate a unified reconstruction approach in either domain.

Finally, the proposed formulation allows, for the first time, a quantitative evaluation of attenuation masks for light field acquisition. We compare different designs and demonstrate that the optimal choice, in terms of signal-to-noise ratio, is dependent on camera noise characteristics. We do not propose new optical light modulation techniques to capture any of the plenoptic dimension, but analyze and unify a variety of existing methods; we outline important criteria for the design of optimal light field attenuation masks.

9.2 Plenoptic Multiplexing

The most popular approach to capturing high-dimensional visual information with a single photograph is multiplexing. For this purpose, a modulator optically separates this information so that a sensor records an image mosaic containing the desired data. Computational processing is then applied to reconstruct the final full-resolution image.

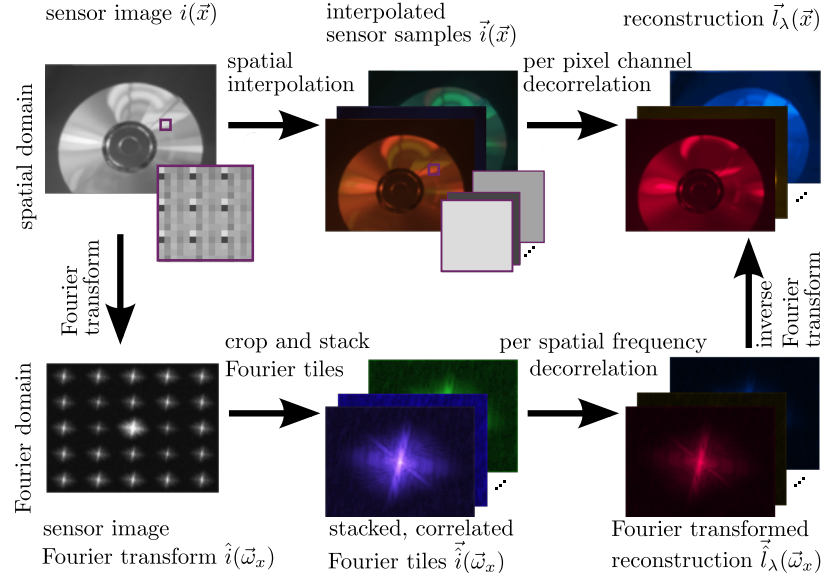


Figure 9.1: Overview of multiplexed image reconstruction. The plenoptic function can be reconstructed by interpolating the sensor samples and performing a local decorrelation in the spatial domain (upper row). Alternatively, it can be reconstructed in the Fourier domain by cropping and locally decorrelating Fourier tiles that are created by the periodic structure of the employed optical filters (lower row).

Consider the example shown in Figure 9.2. A CFA, in this case a Bayer pattern, optically filters the light before it reaches a sensor so that the captured RAW photograph consists of repetitive super-pixels, each encoding four colour samples (Fig. 9.2, upper left). A standard reconstruction or demosaicing interpolates all colour channels to every pixel (Fig. 9.2, lower left). Alternatively, the RAW image can be analyzed in the Fourier domain (Alleyson et al. [2005], Fig. 9.2, upper right), where four different tiles are created that each contain contributions from all colour channels. These tiles can be cropped and decorrelated before being transformed back into the spatial domain (Fig. 9.2, lower right). Although more sophisticated Fourier reconstructions [Li et al., 2008b] may mitigate visible artifacts, a spatial reconstruction produces much better results in this case, because it is usually more resilient to aliasing artifacts (see Sec. 9.4).

Imaging with a Bayer pattern is a well-known problem and only serves as an intuitive, motivating example. An additional processing step after interpolating the sensor samples is, in this particular case, not necessary. Here, a Fourier reconstruction is practically not very useful; understanding the process, however, is essential for later parts of the Chapter, where we consider light field multiplexing approaches that have previously been analyzed exclusively in the Fourier domain.

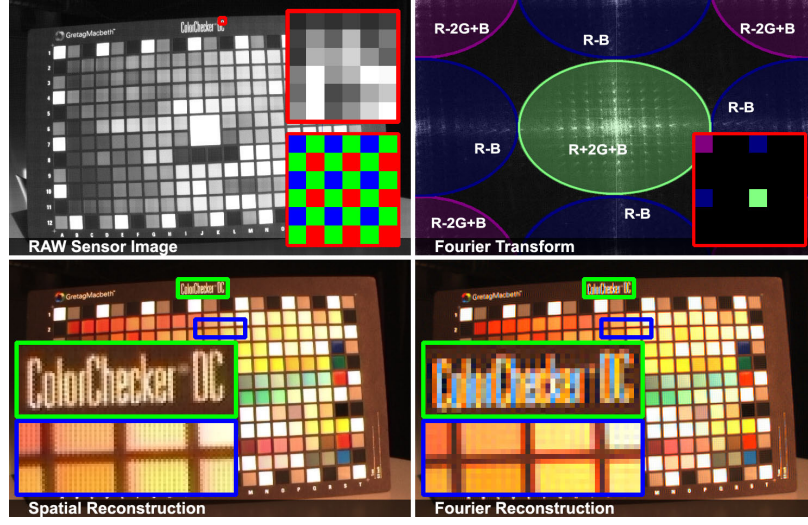


Figure 9.2: Upper left: RAW sensor image with close-up and corresponding CFA. Upper right: Fourier transform with channel correlations illustrated for the entire image and the magnified CFA. Reconstructions of the non-perfectly band-limited signal in the spatial (lower left) and Fourier (lower right) domain reveal different aliasing artifacts.

Symbol	Definition	Physical interpretation	
		Imaging with CFAs	Light field capture
$I_\lambda(\vec{x}, \vec{p})$	The plenoptic function	colour photograph	Light field
$\vec{I}_\lambda(\vec{x}), \vec{I}_\lambda(\vec{\omega}_x)$	Vector of plenoptic quantities at \vec{x} & FT	Same as above	
$\vec{i}(\vec{x})$	Monochromatic sensor image	RAW sensor mosaic with one sample per pixel	
$\vec{\tilde{i}}(\vec{x})$	Vector of interpolated sensor samples	Sensor samples interpolated to all positions	
$\vec{\tilde{i}}(\vec{\omega}_x)$	Cropped & stacked Fourier tiles of sensor image	Correlated, high-dimensional FT of $\vec{I}_\lambda(\vec{x})$	
$\sigma_j(\vec{x}), \hat{\sigma}_j(\vec{\omega}_x)$	Spatial basis functions & FT, $j = 1 \dots N$	CFA Layout	Layout / spatial frequencies
$\pi_j(\vec{p})$	Plenoptic basis functions	Spectral transm. of CFA	All angular frequencies
$\rho_j(\vec{x}), \hat{\rho}_j(\vec{\omega}_x)$	Plenoptic coefficients & FT	colour channels	Sampled angular frequencies
$\Sigma, \hat{\Sigma}$	Spatial correlation matrix & FT	Constant weights, defined by $\rho_j(\vec{x})$, in matrix form	
$\Pi\{\cdot\}$	Projection operator into the plenoptic basis	Projection into colour channels or directions	
$\mathcal{F}\{\cdot\}$	Projection operator into the Fourier basis		
$\vec{i}, \vec{\tilde{i}}, \vec{\tilde{i}}_\lambda, \vec{I}_\lambda, \Pi, \mathcal{F}$	Discrete versions of above quantities		

Table 9.1: A summary of the notation used in this chapter with references to the most important equations.

In the following, we introduce an image formation model for plenoptic multiplexing (Secs. 9.2.1, 9.2.2) and demonstrate how this can be used to derive a generic spatial reconstruction algorithm (Sec. 9.2.3) as well as a corresponding Fourier interpretation (Sec. 9.2.4). The notation introduced in this section, along with physical interpretations, is summarized in Table 9.1. All formulations are continuous unless stated otherwise.

9.2.1 Plenoptic Image Formation

We consider the acquisition of the plenoptic function on the sensor plane, behind the main lens of a camera. A sensor image $\vec{i}(\vec{x})$ is formed by integrating the plenoptic function

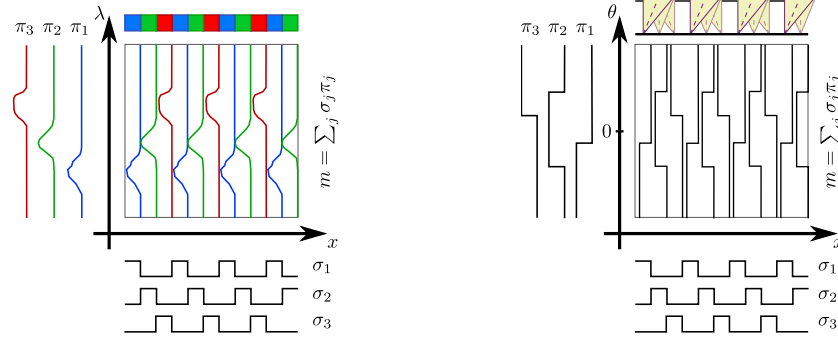


Figure 9.3: 1D illustration of the plenoptic modulator being separated into a spatial and a plenoptic basis. Left: imaging with CFAs; right: light field capture with an array of pinholes.

$l_\lambda(\vec{x}, \vec{p})$ over the plenoptic domain \mathcal{P}

$$i(\vec{x}) = \int_{\mathcal{P}} m(\vec{x}, \vec{p}) l_\lambda(\vec{x}, \vec{p}) d\vec{p}. \quad (9.1)$$

In this formulation, the plenoptic dimensions \vec{p} , including directional $\vec{\theta}$ and temporal t variation as well as the colour spectrum λ , are separated from the spatial location on the sensor \vec{x} . A *plenoptic modulator* $m(\vec{x}, \vec{p})$, which is capable of selectively attenuating each dimension at every location, models a generic optical filter. In the case of colour imaging this modulator is the CFA, but we show in the following sections that this general formulation includes a wide variety of optical elements. Equation 9.3 not only accounts for multiplexing different slices of one plenoptic dimension, such as different colour channels, onto a sensor but also for the combined acquisition of multiple dimensions.

9.2.2 Basis Separation

Our framework is based on the separation of the plenoptic modulator into a sum of mutually independent spatial and plenoptic basis functions:

$$m(\vec{x}, \vec{p}) = \sum_{j=1}^N \sigma_j(\vec{x}) \times \pi_j(\vec{p}). \quad (9.2)$$

This separation is similar in spirit to other basis decompositions, such as the singular value decomposition, and a convenient mathematical tool to interpret plenoptic multiplexing in

either the spatial or the Fourier domain, as demonstrated in the following two subsections. The plenoptic basis $\boldsymbol{\pi}$ is most often defined by the optical properties of a specific modulator (see Table 9.1), but can be chosen as any spatially-invariant basis. As illustrated in Figure 9.3 (left), the plenoptic basis functions $\boldsymbol{\pi}_j(\vec{p})$ in colour imaging, for instance, can model the spectral transmissions of the employed colour filters, whereas the spatial basis functions $\boldsymbol{\sigma}_j(\vec{x})$ describe the layout and mixing of colour samples on the sensor. Figure 9.3 (right) illustrates a choice of these bases for light field cameras with a pinhole array mounted at a slight distance to the sensor.

In the remainder of this Chapter, we assume the spatial basis functions to be periodic, thereby implementing the super-pixel concept. All digital imaging systems are designed to acquire a discrete set of $j = 1 \dots N$ plenoptic samples, such as colours or directions, at every pixel. These samples represent projections of the plenoptic function into the set of plenoptic basis functions

$$\boldsymbol{\rho}_j(\vec{x}) = \int_{\mathcal{P}} \boldsymbol{\pi}_j(\vec{p}) \boldsymbol{l}_\lambda(\vec{x}, \vec{p}) d\vec{p}. \quad (9.3)$$

We term these projections $\boldsymbol{\rho}_j(\vec{x})$ *plenoptic coefficients* (see Table 9.1). Their number N often corresponds to the number M of sensor pixels in each super-pixel, but may be lower as in many CFAs where $N = 3$, $M = 4$. Combining Equations 9.1–9.3 as

$$\begin{aligned} \boldsymbol{i}(\vec{x}) &= \sum_{j=1}^N \boldsymbol{\sigma}_j(\vec{x}) \int_{\mathcal{P}} \boldsymbol{\pi}_j(\vec{p}) \boldsymbol{l}_\lambda(\vec{x}, \vec{p}) d\vec{p} \\ &= \sum_{j=1}^N \boldsymbol{\sigma}_j(\vec{x}) \boldsymbol{\rho}_j(\vec{x}), \end{aligned} \quad (9.4)$$

allows us to model a sensor image at each position \vec{x} as a linear combination of all plenoptic coefficients $\boldsymbol{\rho}_j(\vec{x})$. Spatial multiplexing approaches are designed to directly sample one plenoptic coefficient per sensor pixel. However, Equation 9.4 similarly models the acquisition of differently weighted linear combinations of all plenoptic coefficients at each pixel.

9.2.3 Spatial Reconstruction

The goal of a spatial reconstruction is the recovery of all plenoptic quantities at every pixel of the full-resolution sensor image; initially, only one sample is recorded at each pixel.

Under the assumptions of an underlying band-limited signal and a super-pixel-periodic spatial basis, this is a standard interpolation or demosaicing of the M interleaved sensor sub-images in $\mathbf{i}(\vec{x})$, resulting in the vector-valued image $\vec{\mathbf{i}}(\vec{x})$. In order to compute the desired plenoptic samples from $\vec{\mathbf{i}}(\vec{x})$, the spatial and plenoptic bases subsequently need to be inverted in a per-pixel manner.

Before interpolation, every sensor pixel in $\mathbf{i}(\vec{x})$ is associated with a single value of all N spatial basis functions $\sigma_j(\vec{x})$ at \vec{x} . Interpolating the $k = 1 \dots M$ sub-pixels within the super-pixels to all locations also interpolates the spatial basis. Therefore, each position in $\vec{\mathbf{i}}(\vec{x})$ is associated with an array of constants $\Sigma \in \mathcal{R}^{M \times N}$, as defined by the interpolated spatial basis functions. With this notation, we can define

Theorem 1 (*Plenoptic Spatial Multiplexing, PSM*). *The interpolated sensor samples $\vec{\mathbf{i}}(\vec{x})$, representing weighted combinations of all plenoptic coefficients at a particular location \vec{x} , are **locally** related to the plenoptic function as*

$$\vec{\mathbf{i}}(\vec{x}) = \Sigma \vec{\rho}(\vec{x}) = \Sigma \Pi \left\{ \vec{\mathbf{l}}_\lambda(\vec{x}) \right\}. \quad (9.5)$$

The operator $\Pi \{ \cdot \}$ projects the plenoptic function into the plenoptic basis, $\vec{\mathbf{l}}_\lambda(\vec{x})$ is the continuous plenoptic function in vector form, and $\vec{\rho}(\vec{x})$ consists of the N corresponding plenoptic coefficients at each position. This reconstruction is illustrated in the upper row of Figure 9.1. The proof for Theorem 1 is included in Appendix B.

The PSM theorem shows that we can reconstruct the plenoptic function $\vec{\mathbf{l}}_\lambda(\vec{x})$ from sensor samples $\mathbf{i}(\vec{x})$ by performing a *local* decorrelation¹ on the interpolated measurement samples $\vec{\mathbf{i}}(\vec{x})$. The decorrelation inverts the spatial basis Σ and, depending on the application, also the projection into the plenoptic basis. However, Theorem 1 not only shows that the correlation between the measured samples is spatially local, but also that the correlation is in fact a linear operator, yielding

Corollary 1 *Any linear interpolation filter can be applied to the measured sensor samples $\mathbf{i}(\vec{x})$ prior to decorrelation while yielding equivalent results to application after the decorrelation.*

Image processing operations such as upsampling, edge detection, blurring, sharpening, etc. can thus be performed on the correlated image without affecting the end result. Although this only applies to linear filters in theory, we show in Section 9.3 that nonlinear filters

¹In the context of this thesis, decorrelation is defined as the inversion of a linear operator.

can achieve high-quality reconstruction results in practice. Nonlinear filters are already the preferred choice for colour demosaicing; we show that these can also be applied to light field reconstruction under certain conditions.

9.2.4 Fourier Reconstruction

In recent literature, multiplexing strategies have often been analyzed exclusively in the Fourier domain [Agrawal et al., 2010b; Georgiev et al., 2008; Lanman et al., 2008; Veeraraghavan et al., 2008, 2007]. For this reason we provide the dual Fourier view of plenoptic multiplexing and reconstruction in the following.

By applying the convolution theorem, the Fourier transform of an acquired image (Eq. 9.4) is given as

$$\mathcal{F}_x \{ \mathbf{i}(\vec{x}) \} = \mathcal{F}_x \left\{ \sum_{j=1}^N \boldsymbol{\sigma}_j(\vec{x}) \boldsymbol{\rho}_j(\vec{x}) \right\} = \sum_{j=1}^N \hat{\boldsymbol{\sigma}}_j(\vec{\omega}_x) \otimes \hat{\boldsymbol{\rho}}_j(\vec{\omega}_x), \quad (9.6)$$

where $\hat{\cdot}$ denotes the Fourier transformed version of a quantity and $\vec{\omega}_x$ are the spatial frequencies. The Poisson summation formula dictates that the Fourier transform of a periodic function is a weighted set of Dirac peaks. Thus, the Fourier transform of the super-pixel-periodic spatial basis functions is given by

$$\hat{\boldsymbol{\sigma}}_j(\vec{\omega}_x) = \sum_{k=1}^M \hat{\boldsymbol{\sigma}}_j^k \delta(\vec{\omega}_x - k\Delta\vec{\omega}_x), \quad (9.7)$$

where $\Delta\vec{\omega}_x$ is the frequency offset or distance between successive Dirac peaks² and the values $\hat{\boldsymbol{\sigma}}_j^k$ are complex weighting factors for basis function j . These weights correspond to the Fourier transform of a single period of that specific basis function. Combining Equations 9.6 and 9.7 as

$$\mathcal{F}_x \{ \mathbf{i}(\vec{x}) \} = \sum_{k=1}^M \left(\delta(\vec{\omega}_x - k\Delta\vec{\omega}_x) \otimes \sum_{j=1}^N \hat{\boldsymbol{\sigma}}_j^k \hat{\boldsymbol{\rho}}_j(\vec{\omega}_x) \right) \quad (9.8)$$

shows that M different tiles containing linear combinations of Fourier transformed plenoptic

²For a 2D sensor image the bases are periodic in both spatial dimensions, but we will omit the second one in our notation of k for clarity.

coefficients $\hat{\rho}_j(\vec{\omega}_x)$ are created in the frequency domain. As illustrated in Figure 9.2 (upper right), these tiles can be cropped from the Fourier transformed sensor image and arranged in a stack $\vec{\mathbf{i}}(\vec{\omega}_x)$ (see Fig. 9.1, bottom center). The correlation of these stacked tiles is *local in the spatial frequencies* $\vec{\omega}_x$. In analogy to the PSM theorem, Equation 9.5, we can therefore state the following

Theorem 2 (*Plenoptic Fourier Multiplexing, PFM*). *Cropping and stacking the individual Fourier tiles of a multiplexed sensor image allows the plenoptic function to be expressed as a correlation in the Fourier domain that is **local** for each spatial frequency:*

$$\vec{\mathbf{i}}(\vec{\omega}_x) = \hat{\Sigma} \vec{\rho}(\vec{\omega}_x) = \hat{\Sigma} \Pi \left\{ \vec{\mathbf{i}}_\lambda(\vec{\omega}_x) \right\}. \quad (9.9)$$

The correlation matrix $\hat{\Sigma}_{jk} = \hat{\sigma}_j^k$ is determined by the Fourier weights of the spatial basis. The projection into the plenoptic basis $\Pi \{ \cdot \}$ remains unchanged because of its independence of \vec{x} (see Eq. 9.2), which allows us to change the order of operation, i.e. $\mathcal{F}_x \left\{ \Pi \left\{ \vec{\mathbf{i}}_\lambda(\vec{x}) \right\} \right\} = \Pi \left\{ \mathcal{F}_x \left\{ \vec{\mathbf{i}}_\lambda(\vec{x}) \right\} \right\} = \Pi \left\{ \vec{\mathbf{i}}_\lambda(\vec{\omega}_x) \right\}$. This theorem is illustrated in the lower row of Figure 9.1; the proof is included in Appendix B.

9.2.5 Discussion

Previously proposed Fourier multiplexing approaches have analyzed the image formation and reconstruction exclusively in the frequency domain. The mathematical framework presented in this section, however, formulates plenoptic multiplexing and demultiplexing in very general terms. Not only does our framework model the acquisition and reconstruction of arbitrary combinations of plenoptic dimensions, but Theorems 1 and 2 also allow us to analyze and process multiplexed data in either the spatial or the Fourier domain. The notation introduced in this section makes it easy to understand the close connections between the two different interpretations, which is important because each has its own advantages. A spatial reconstruction, that is an interpolation of the sensor samples followed by a per-pixel decorrelation, is generally the preferred method for processing captured data (Sec. 9.3) and analyzing reconstruction noise (Sec. 9.5). A Fourier perspective of the same problem, on the other hand, provides a powerful tool for analyzing many important properties such as aliasing (Sec. 9.4).

Our analysis demonstrates that both spatial and Fourier multiplexing schemes are closely related. The cropping operation in Fourier space is a multiplication with a rect function, which is equivalent to a spatial sinc filter (see App. B). Therefore, all previously proposed

Fourier reconstruction methods use a fixed spatial reconstruction filter: the sinc. We demonstrate in the next section that this choice negatively affects the quality of demultiplexed data. More sophisticated apodization approaches (i.e. using a soft roll-off rather than hard cropping) can potentially improve the quality of Fourier-based reconstructions, and are in fact equivalent to using non-sinc linear filters in a spatial reconstruction. Note, however, that nonlinear reconstruction filters, including those commonly used for demosaicing, cannot easily be interpreted as Fourier-domain operations.

A consequence of our analysis is that multiplexing schemes have, independent of the reconstruction domain, nominally the same band-limitation requirements. Due to the large choice of linear and nonlinear filters, however, a spatial reconstruction can be made more resilient to residual high frequencies, thereby mitigating aliasing artifacts (see Secs. 9.3, 9.4).

9.3 Application to Light Field Reconstruction

In the following, we demonstrate how the general framework introduced in the last section applies to the reconstruction of light fields. We show, for the first time, how Fourier multiplexed light fields captured with non-refractive attenuation masks can be reconstructed with a superior quality in the spatial domain (Sec. 9.3.1). Although the general plenoptic modulator introduced in Section 9.2.1 only models selective *attenuation* for each plenoptic dimension, we show in Section 9.3.2 how similar concepts apply to a variety of acquisition systems with *refractive* optical elements. In Section 9.3.3, we demonstrate how our framework allows the tempo-directional plenoptic manifolds proposed by Agrawal et al. [2010b] to be reconstructed with a higher quality than the originally proposed Fourier processing.

Throughout this section, we employ a two-plane parameterization for light fields. As illustrated for a 1D case in Figure 9.4, this includes a position \vec{x} on the sensor and the relative distance on a plane at unit distance $\vec{v} = \tan(\vec{\theta})$, which replaces an actual angles $\vec{\theta}$.

9.3.1 General Non-Refractive Modulators

Attenuation masks that do not include refractive optical elements have recently been popularized for light field acquisition [Georgiev et al., 2008; Lanman et al., 2008; Veeraraghavan et al., 2008, 2007]. All of these approaches have been analyzed and reconstructed exclusively in the Fourier domain. Here, we show how the employed periodic attenuation masks can be separated into a spatial and a plenoptic basis. This separation allows the aforementioned techniques to be expressed in the framework introduced in Section 9.2.

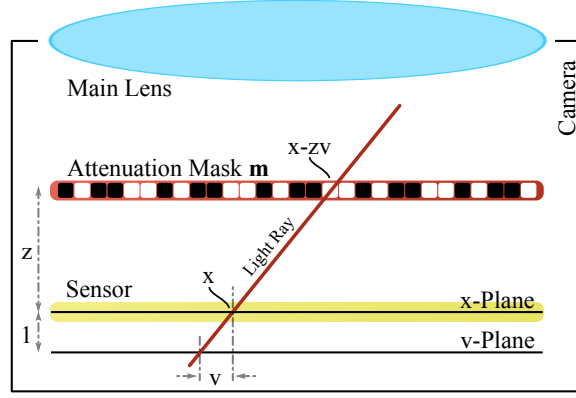


Figure 9.4: A light field can be parameterized by a spatial position x on the sensor plane and a relative distance v on a plane at unit distance.

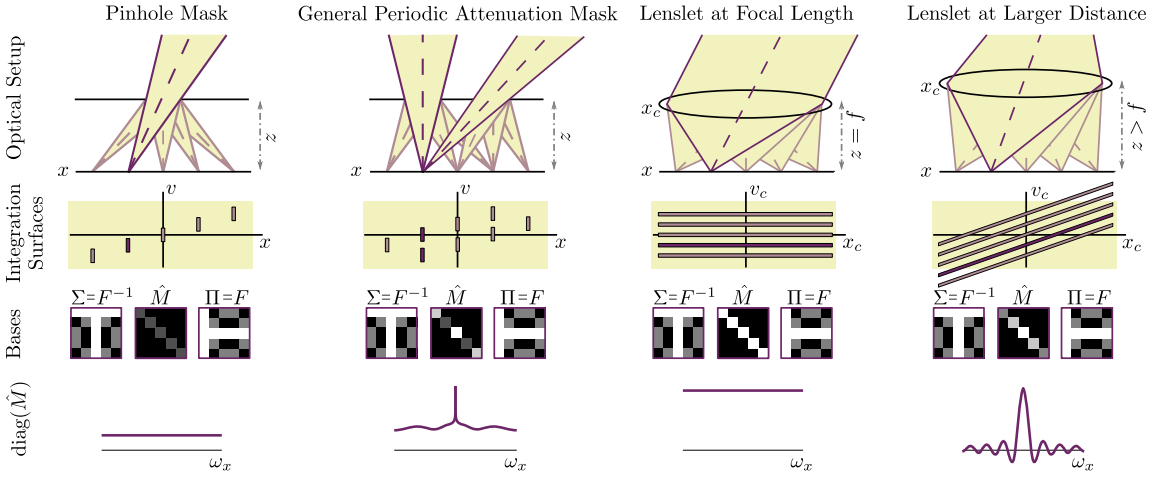


Figure 9.5: An illustration of the optical setups, integration surfaces, spatial and plenoptic bases as well as the weighting factors $\hat{\mathbf{m}}$ for a variety of super-pixel based light field cameras. The convolution of a light field and a periodic attenuation mask or refractive optical element, resulting in the captured sensor image, can be separated into a spatial and a plenoptic part using the Fourier basis (row 3). This allows us to perform a light field reconstruction in either the spatial or Fourier domain by applying the PSM or PFM theorem, respectively. Please note that the camera and sensor coordinates for non-refractive elements are identical; the integration surfaces for refractive optical elements already include the mapping from sensor space to world space on the microlens plane.

As illustrated in Figure 9.4, the plenoptic modulator (Eq. 9.1) for attenuation masks at a distance z to a sensor is $\mathbf{m}(\vec{x}, \vec{v}) = \mathbf{m}(\vec{x} - z\vec{v})$. This formulation models the light transport in free space from sensor to mask as well as the attenuation caused by the latter. A separation of this modulator into a purely spatial and a plenoptic, in this case directional, part can be achieved by substituting the modulator $\mathbf{m}(\vec{x} - z\vec{v})$ with the inverse of its Fourier transform

$$\begin{aligned}
\mathbf{i}(\vec{x}) &= \int_{\vec{v}} \mathbf{l}_{\lambda}(\vec{x}, \vec{v}) \mathbf{m}(\vec{x} - z\vec{v}) d\vec{v} \\
&= \int_{\vec{v}} \mathbf{l}_{\lambda}(\vec{x}, \vec{v}) \int_{\vec{\omega}_x} \hat{\mathbf{m}}(\vec{\omega}_x) e^{2\pi i(\vec{x} - z\vec{v}) \cdot \vec{\omega}_x} d\vec{\omega}_x d\vec{v} \\
&= \int_{\vec{\omega}_x} \hat{\mathbf{m}}(\vec{\omega}_x) e^{2\pi i\vec{x} \cdot \vec{\omega}_x} \int_{\vec{v}} \mathbf{l}_{\lambda}(\vec{x}, \vec{v}) e^{-2\pi i z \vec{v} \cdot \vec{\omega}_x} d\vec{v} d\vec{\omega}_x.
\end{aligned} \tag{9.10}$$

Equation 9.10 shows that the plenoptic basis projects the light field into its angular frequencies, which are then multiplexed, with mask-dependent weights $\hat{\mathbf{m}}(\vec{\omega}_x)$, into the spatial frequencies of a sensor image.

For a practical processing of digital images, the integrals can be discretized as $\mathbf{i} = \mathbf{F}^{-1} \hat{\mathbf{M}} \mathbf{F} \mathbf{l}_{\lambda}$, with $\hat{\mathbf{M}} = \text{diag}(\hat{\mathbf{m}})$. As illustrated in Figure 9.5 (center left), this formulation allows us to choose the inverse discrete Fourier transform (DFT) as the spatial basis $\mathbf{\Sigma} = \mathbf{F}^{-1}$, and the DFT as the discretized plenoptic projection $\mathbf{\Pi} = \mathbf{F}$. A spatial reconstruction can therefore be performed by solving Equation 9.5. Again, a discrete spatial reconstruction is performed by interpolating the measured samples in the sensor image \mathbf{i} to all positions and then decorrelating the resulting, vector-valued, discrete image $\vec{\mathbf{i}}$ in a per-pixel manner as

$$\vec{\mathbf{l}}_{\lambda} = \mathbf{\Pi}^{-1} \mathbf{\Sigma}^{-1} \vec{\mathbf{i}} = \mathbf{F}^{-1} \hat{\mathbf{M}}^{-1} \mathbf{F} \vec{\mathbf{i}}. \tag{9.11}$$

Equation 9.11 is a *per-image-pixel deconvolution* with a kernel that is defined by the attenuation pattern of the mask. As shown in Section 9.5, a deconvolution with sum-of-sinusoids patterns [Agrawal et al., 2010b; Veeraraghavan et al., 2007], for instance, represents a high-pass filter. Unfortunately, this type of filter decreases the signal-to-noise ratio (SNR) of the reconstructed light field significantly by amplifying sensor noise.

Alternatively, the previously proposed discrete Fourier reconstruction can be performed by directly solving Equation 9.9 as:

$$\vec{\mathbf{l}}_{\lambda} = \mathbf{\Pi}^{-1} \hat{\mathbf{\Sigma}}^{-1} \vec{\mathbf{i}} = \mathbf{F}^{-1} \hat{\mathbf{M}}^{-1} \vec{\mathbf{i}}. \tag{9.12}$$

Equation 9.12 shows that the stacked discrete Fourier tiles $\vec{\mathbf{i}}$ need to be re-weighted on a *per-spatial-frequency* basis; the weighting factors $\hat{\mathbf{m}}$ depend on the applied mask. An inverse DFT is required to invert the plenoptic basis $\mathbf{\Pi}$. An additional inverse DFT is applied to compute the desired plenoptic samples $\vec{\mathbf{l}}_{\lambda}$ from $\vec{\mathbf{l}}_{\lambda}$, which is equivalent to an inverse 4D

Fourier transform of cropped and stacked light field Fourier tiles [Georgiev et al., 2008; Lanman et al., 2008; Veeraraghavan et al., 2008, 2007].

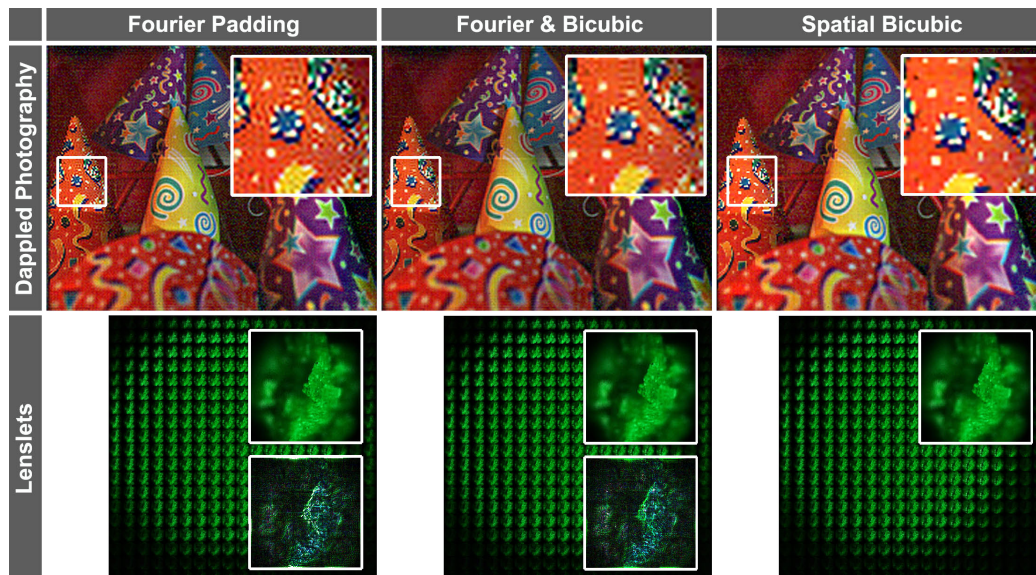


Figure 9.6: Comparison of reconstruction quality for Cones data set [Veeraraghavan et al., 2007] captured with a non-refractive sum-of-sinusoids mask and lenslet-based Fluorescent Crayon Wax data set [Levoy et al., 2006]. All results are three-times upsampled during reconstruction. Left column: upsampling by zero-padding the 4D inverse DFT. Center column: low resolution 4D inverse DFT followed by bicubic upsampling. Right column: bicubic up-sampling followed by local decorrelation. For the right column, we show one of the light field views and a contrast enhanced difference image to the spatial reconstruction in the magnifications. Ringing artifacts are clearly visible.

Figure 9.6 (left) shows comparisons of light field reconstructions in the Fourier domain, as previously proposed, and in the spatial domain. Noise and ringing artifacts are significantly reduced in the spatial reconstruction, which is enabled by our framework. Even with a simple linear spatial interpolation scheme such as cubic interpolation, common ringing artifacts associated with Fourier-based techniques can be avoided. A more detailed discussion on aliasing artifacts can be found in Section 9.4.

Although Corollary 1 is theoretically only valid for linear filters, Figure 9.7 demonstrates that a practical reconstruction works well with nonlinear filters. The presented example shows two different viewpoints of the reconstructed Mannequin dataset [Lanman et al., 2008] and, more importantly, a spatial reconstruction with a non-linear joint bilateral filter. This filter is just one of many possible choices for sophisticated spatial reconstruction filters facilitated by Theorem 1. The joint bilateral filter can, in this case, reconstruct a slightly sharper image than a bicubic filter.

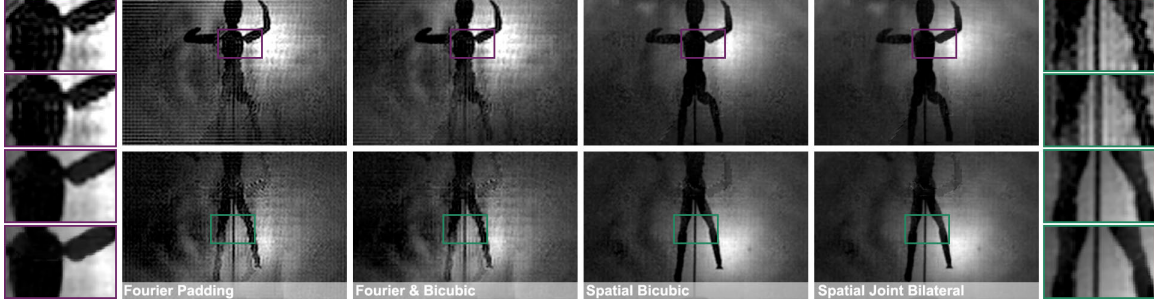


Figure 9.7: *Reconstruction results for Mannequin data set [Lanman et al., 2008]. The results are three-times upsampled and show two different views of the reconstruction. The fifth column shows a reconstruction that was computed with a nonlinear joint bilateral filter. Close-ups are shown for the top row on the left and for the bottom row on the right.*

9.3.2 Refractive Modulators

The general plenoptic modulator introduced in Section 9.2.1 does not directly model ray deflections caused by refractive optical elements. However, the image formation can be modeled as a convolution of the plenoptic function and a refractive plenoptic kernel $\mathbf{k}(\vec{x}, \vec{x}_c, \vec{v}_c)$ [Levin et al., 2009]:

$$\mathbf{i}(\vec{x}) = \int_{\vec{x}_c} \int_{\vec{v}_c} \mathbf{k}(\vec{x}, \vec{x}_c, \vec{v}_c) \mathbf{l}_\lambda(\vec{x}_c, \vec{v}_c) d\vec{x}_c d\vec{v}_c, \quad (9.13)$$

where \vec{x} is the spatial coordinate on the sensor surface and \vec{x}_c is the spatial coordinate defined on the plane of the refractive elements (see Fig. 9.5, right). Here, $\mathbf{l}_\lambda(\vec{x}_c, \vec{v}_c)$ is the light field on the plane of the refractive elements; the directional coordinate \vec{v}_c describes the directional light variation before modulation by the kernel inside the camera behind the main lens. Plenoptic modulation kernels for different optical elements are well known. Under paraxial approximations and disregarding the element's aperture and wavelength of light, the kernel for most refractive elements is of the form

$$\mathbf{k}(\vec{x}, \vec{x}_c, \vec{v}_c) = \delta(\vec{x} - z\vec{v}_c - \phi(\vec{x}_c)). \quad (9.14)$$

The specific kernel for a lens at focal distance to the sensor is given by $\phi_{z=f}(\vec{x}_c) = 0$. The term becomes non-zero when the lens is moved away from the focal distance, as proposed by Lumsdaine and Georgiev [2009]: $\phi_{z \neq f}(\vec{x}_c) = s\vec{x}_c$, with $s = 1 - z/f$ being the slope of the integration surface. Optical setups and integration surfaces for all of these cases are illustrated in Figure 9.5.

In the following, we demonstrate how a separation of the refractive plenoptic kernel into a spatial and a plenoptic basis can be performed. This is shown for a single element of an array of optical elements such as lenslets. The plenoptic function is assumed to be spatially band-limited, which in this case requires it to be constant over a *single optical element*, i.e. $\mathbf{l}_\lambda(\vec{x}_c, \vec{v}_c) = \mathbf{l}_\lambda(\vec{v}_c)$. Combining Equations 9.13 and 9.14 and substituting the plenoptic kernel with the inverse of its Fourier transform yields

$$\begin{aligned} \mathbf{i}(\vec{x}) &= \int_{\vec{x}_c} \int_{\vec{v}_c} \mathbf{l}_\lambda(\vec{v}_c) \delta(\vec{x} - z\vec{v}_c - \phi(\vec{x}_c)) d\vec{x}_c d\vec{v}_c \\ &= \int_{\vec{x}_c} \int_{\vec{v}_c} \mathbf{l}_\lambda(\vec{v}_c) \int_{\vec{\omega}_x} e^{2\pi i(\vec{x} - z\vec{v}_c - \phi(\vec{x}_c)) \cdot \vec{\omega}_x} d\vec{\omega}_x d\vec{x}_c d\vec{v}_c \\ &= \int_{\vec{\omega}_x} \hat{\mathbf{m}}(\vec{\omega}_x) e^{2\pi i \vec{x} \cdot \vec{\omega}_x} \int_{\vec{v}_c} \mathbf{l}_\lambda(\vec{v}_c) e^{-2\pi i z \vec{v}_c \cdot \vec{\omega}_x} d\vec{v}_c d\vec{\omega}_x, \end{aligned} \quad (9.15)$$

where the extra term $\hat{\mathbf{m}}(\vec{\omega}_x) = \int_{\vec{x}_c} e^{-2\pi i \phi(\vec{x}_c) \cdot \vec{\omega}_x} d\vec{x}_c$ varies for different refractive elements and represents their optical transfer function (OTF, see Figure 9.5). Lenses at focal distance to the sensor are not affected by the extra term, i.e. $\hat{\mathbf{m}}(\vec{\omega}_x) = 1, \forall \vec{\omega}_x$.

Figure 9.6 (right) shows a comparison of spatial and Fourier reconstructions of a light field that was captured inside a microscope using a lenslet array at focal distance to the sensor [Levoy et al., 2006]. A decorrelation of the interpolated sensor image $\vec{\mathbf{i}}(\vec{x})$ is in this case redundant, as every sensor pixel measures uncorrelated directional samples of the light field. Although artifacts are visible in the magnifications of Figure 9.6 (right), they are more subtle than those in mask-based light field reconstructions (Fig. 9.6, left). This can be attributed to pixels under each microlens integrating spatial light variation over the entire lenslet area, which provides a proper spatial band-limit and therefore minimizes aliasing [Levoy and Hanrahan, 1996] (see Sec. 9.4).

Employing lenslets at a distance to the sensor that is different to their focal lengths, i.e. $z \neq f$, was explored by Lumsdaine and Georgiev [2009]. It was shown that such a setup in combination with a custom resorting algorithm is capable of reconstructing light fields with a higher spatial but reduced angular resolution. In this particular setup, the corresponding OTF is a sinc (Fig. 9.5, right). Applying our theory would aim at reconstructing the full spatial and directional resolution of the light field, which is ill-conditioned due to the zero-crossings of the OTF, unless additional statistical priors are incorporated.

9.3.3 Plenoptic Dimension Transfer

Agrawal et al. [2010b] propose to equip a camera aperture with a pinhole mask that can be moved throughout the exposure time of a single photograph. This motion encodes temporal light variation in the directions of a light field. The light field itself is acquired by mounting an additional sum-of-sinusoids attenuation mask at a small distance to the camera sensor. Such a setup allows the captured photograph to be reinterpreted as either a high spatial resolution image, a light field, or a video for different parts of the scene in post-processing. Mathematically, the image formation can be formulated as

$$\mathbf{i}(\vec{x}) = \int_{\vec{v}} \int_t \mathbf{l}_{\lambda}(\vec{x}, \vec{v}, t) \mathbf{m}_t(\vec{x}, \vec{v}, t) \mathbf{m}_v(\vec{x}, \vec{v}, t) dt d\vec{v}, \quad (9.16)$$

where $\mathbf{m}_v(\vec{x}, \vec{v}, t) = \mathbf{m}(\vec{x} - z\vec{v})$ is the non-refractive attenuation mask, as introduced in Section 9.3.1, and $\mathbf{m}_t(\vec{x}, \vec{v}, t) = \delta(\vec{v} - \psi_{\vec{v}}(t))$ is the moving pinhole aperture. The pinhole motion is described by $\psi_{\vec{v}}(t)$. In our framework, this can be expressed as

$$\begin{aligned} \mathbf{i}(\vec{x}) &= \int_{\vec{v}} \int_t \mathbf{l}_{\lambda}(\vec{x}, \vec{v}, t) \delta(\vec{v} - \psi_{\vec{v}}(t)) \mathbf{m}(\vec{x} - z\vec{v}) dt d\vec{v} \\ &= \int_t \mathbf{l}_{\lambda}(\vec{x}, \psi_{\vec{v}}(t), t) \mathbf{m}(\vec{x} - z\psi_{\vec{v}}(t)) dt \\ &= \int_{\vec{\omega}_x} \hat{\mathbf{m}}(\vec{\omega}_x) e^{2\pi i \vec{x} \cdot \vec{\omega}_x} \\ &\quad \int_t \mathbf{l}_{\lambda}(\vec{x}, \psi_{\vec{v}}(t), t) e^{-2\pi i z \psi_{\vec{v}}(t) \cdot \vec{\omega}_x} dt d\vec{\omega}_x. \end{aligned} \quad (9.17)$$

The pinhole motion $\psi_{\vec{v}}(t)$ introduces a manifold in the tempo-directional domain of the plenoptic function over which the sensor integrates. A reconstruction can only recover this manifold; temporal and directional light variation are coupled. In our framework, the plenoptic basis for this example is the Fourier transform of the plenoptic manifold, whereas the spatial basis is, just as in the case of attenuation masks and refractive optical elements, the inverse Fourier transform.

Figure 9.8 shows spatial and Fourier-based reconstructions of one of the datasets used by Agrawal et al. [2010b]. Columns 1 and 2 show two different views of the reconstructed light field. Each of these views additionally encodes a different time in the animation: the Rubik’s cube is moved away from the camera. As seen in the close-ups, a spatial reconstruction, enabled by Theorem 1, can reduce ringing artifacts as compared to previously employed

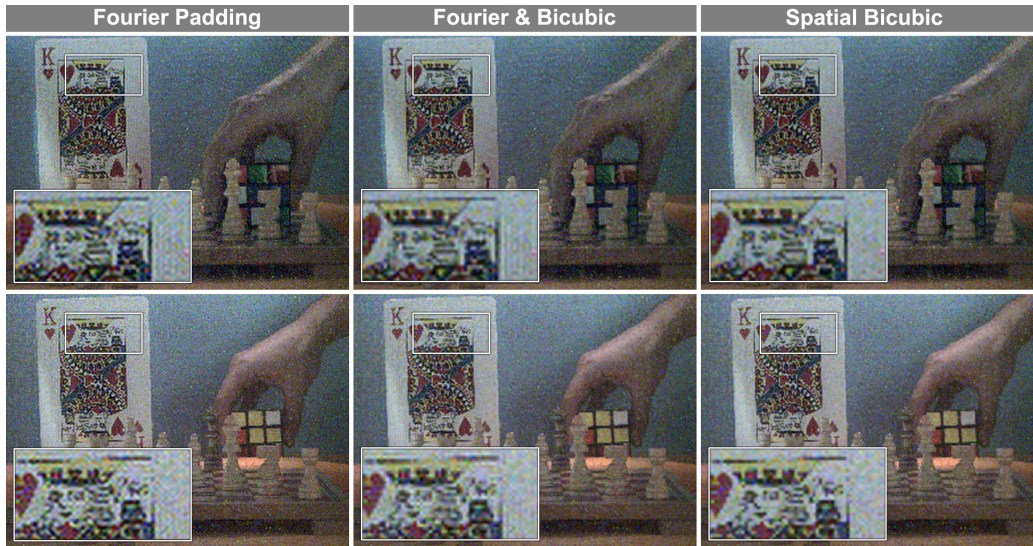


Figure 9.8: Two different views of a light field reconstructed with an upsampling factor of three. Each view encodes a different temporal slice within the exposure time of a single photograph, as proposed by Agrawal et al. [2010b]. The spatial reconstruction (bottom row) increases the reconstruction quality by reducing ringing artifacts that are visible in previously proposed Fourier reconstructions (top and center row).

Fourier reconstructions.

9.4 Analyzing Aliasing

One of the main arguments throughout the last section is that a spatial reconstruction of multiplexed data can improve the image quality. We demonstrate in this section, that the difference in quality is mainly due to aliasing, that is violations of the band-limit assumption. While spatial processing with sophisticated reconstruction filters offers the benefit of improved image quality, a Fourier perspective allows aliasing to be analyzed in a convenient manner.

Consider the experiment in Figure 9.9. A light field is multiplexed onto a single sensor with a MURA attenuation mask. The mask consists of a repetitive pattern of 5×5 pixels, as introduced by Gottesman and Fenimore [1989]. Figure 9.9 (upper left) shows the Fourier transform of a simulated sensor image without any anti-aliasing applied. The grid of 5×5 different Fourier tiles is clearly visible. Without optically filtering the captured signal, the copies slightly overlap, thereby causing pre-aliasing in the reconstructions (Fig. 9.9, columns 2, 3). As expected from the results presented in Section 9.3, aliasing artifacts in the

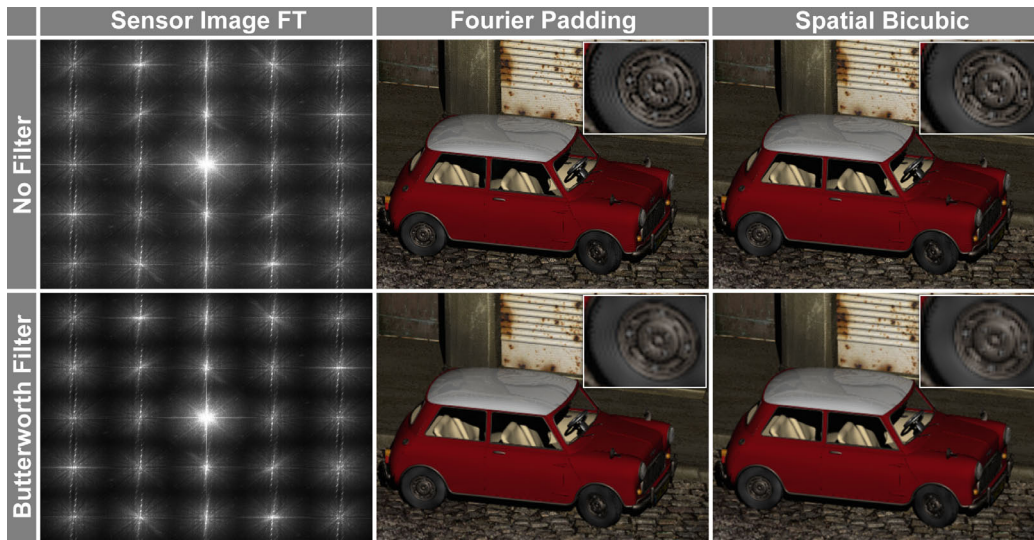


Figure 9.9: *Analyzing aliasing artifacts. Multiplexing a light field with 5×5 views onto a sensor is simulated without (left column) and with (right column) a synthetic optical anti-aliasing filter applied. The Fourier tiles of the sensor image slightly overlap without any filtering (upper left). A proper anti-aliasing filter mitigates this overlap and thereby artifacts in the reconstructions (right). A spatial processing of multiplexed data (lower left) is usually more robust to signal aliasing than a corresponding Fourier approach (center left).*

Fourier reconstruction are much stronger than in the corresponding spatial reconstruction. With a proper anti-aliasing filter applied to the signal before capture, however, the quality differences become more subtle (Fig. 9.9, bottom row). For this experiment we applied a first-order Butterworth low-pass filter to each input image, before multiplexing them on the sensor. This filter significantly reduces pre-aliasing in the recorded signal by eliminating high spatial frequencies from the light field. In the Fourier domain, reduced aliasing is visible by less overlap of the frequency tiles (Fig. 9.9, lower left).

Optical anti-aliasing filters are common practice for imaging with CFAs [Greivenkamp, 1990]. While attenuation masks have been optimized for the 4D frequency distribution of natural scenes [Veeraraghavan et al., 2008], anti-aliasing mechanisms have yet to be implemented for mask-based light field cameras.

9.5 Analyzing Light Field Reconstruction Noise

The last sections have discussed a theoretical framework for modeling plenoptic multiplexing with practical applications in high-quality reconstructions. In many multiplexing tasks, however, there is a wide choice of optical filters or modulators that can be employed to ac-

quire the same visual information. Several different CFA designs are, for instance, available for colour imaging. While RGB-based colour filters are usually preferred in bright lighting conditions, CMY filters provide better noise characteristics in low-light conditions [Sajadi et al., 2011]. In this section, we analyze and compare a variety of optical multiplexing filters for two different applications: imaging with colour filter arrays and light field acquisition. While colour imaging serves as an intuitive example to validate the noise model introduced in this section, we present the first comparative analysis of non-refractive light field multiplexing masks. Most of the employed attenuation patterns [Lanman et al., 2008; Veeraraghavan et al., 2008, 2007] have been exclusively analyzed in the Fourier domain, which makes it difficult to evaluate the noise characteristics of the actual mask pattern. The spatial analysis enabled by Theorem 1 allows us to compare the performance of alternative attenuation masks with respect to the signal-to-noise ratio (SNR) in reconstructed imagery.

We employ a noise model that is commonly applied in computer vision [Schechner et al., 2007; Wuttig, 2005]. The total noise variance ς^2 of a camera image is modeled as the combination of a signal-independent additive term ς_c^2 , which includes dark current and amplifier noise, as well as a signal-dependent photon shot noise term $\mathbf{i}(\vec{x})\varsigma_p^2$. Following standard practice [Schechner et al., 2007], we approximate the image intensity term in the photon noise by the mean light transmission τ of the plenoptic modulator, yielding the following noise variance in the captured image:

$$\varsigma^2 = \varsigma_c^2 + \tau\varsigma_p^2. \quad (9.18)$$

In order to compare alternative plenoptic modulators, we need to propagate the sensor noise ς of a specific setup to the demultiplexed reconstruction. For this purpose, we introduce a noise amplification term α that is based on our discretized image formation (Sec. 9.3, see Schechner et al. [2007] for more details):

$$\alpha = \sqrt{\frac{1}{N} \text{trace} \left((\mathbf{\Pi}\mathbf{\Sigma})^T \mathbf{\Sigma}\mathbf{\Pi} \right)^{-1}}. \quad (9.19)$$

The signal-to-noise ratio in the demultiplexed plenoptic function requires expressions for the signal and for the noise term. Assuming a normalized plenoptic function and orthogonal plenoptic basis functions, the signal term in the demultiplexed signal can be approximated by $1/N$, where N is the number of sampled plenoptic coefficients (e.g, colour channels or light field views). The reconstruction noise term is the standard deviation of the propagated sensor noise $\alpha\varsigma$, resulting in an SNR of

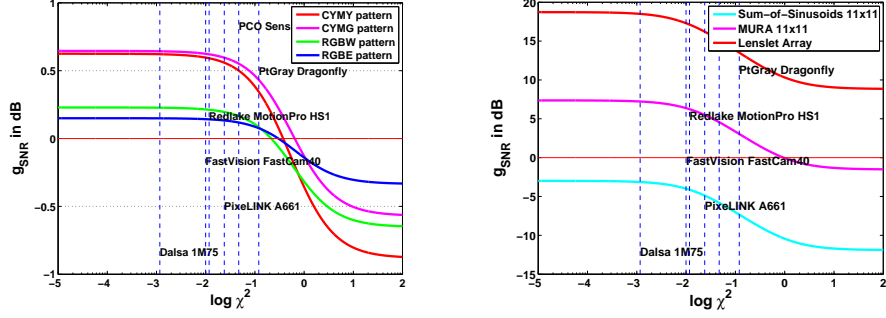


Figure 9.10: SNR comparison of various alternative CFA patterns to the Bayer pattern (left). SNR comparison of different light field attenuation masks (right). The vertical lines indicate χ^2 values for several machine vision cameras tested by Schechner et al. [2007]. All cameras are operating in the gain region of the filters, i.e. $g_{\text{SNR}} > 0$. Note, however, that the lines can be moved left and right along the χ^2 -axis by increasing and decreasing the gain setting of a camera, respectively.

$$\text{SNR} = 10 \log_{10} \left(\frac{1}{N \alpha \sqrt{\varsigma_c^2 + \tau \varsigma_p^2}} \right), \quad (9.20)$$

where the SNR is defined in dB. The gain of SNR for this demultiplexed signal compared to some demultiplexed reference signal with a noise term of ς_{ref} is then

$$g_{\text{SNR}} = \text{SNR} - \text{SNR}_{ref} = 10 \log_{10} \left(\frac{\alpha_{ref} \varsigma_{ref}}{\alpha \varsigma} \right). \quad (9.21)$$

A positive gain indicates an improved SNR, relative to the reference signal, while a negative gain indicates noise amplification. A plot of SNR gain for a standard Bayer CFA filter, which serves as the reference signal, compared to the more transmissive cyan-yellow-magenta-yellow (CYMY), cyan-yellow-magenta-green (CYMG), and red-green-blue-white (RGBW) CFAs is shown in Figure 9.10 (left). The plots demonstrate that all alternative filters produce a slightly better SNR than the Bayer pattern when the additive noise term dominates (left part of the plot). However, performance drops below that of the Bayer pattern once photon shot noise becomes dominant. We employ the notation introduced by Wuttig [2005], where a parameter $\chi = \varsigma_p / \varsigma_c$ describes the ratio of signal-dependent and signal-independent noise terms. This makes it more convenient to plot the performance of a multiplexing scheme with different camera noise parameters up to a global scale, as seen in Figure 9.10. Our CFA noise analysis helps to determine the exact gain regions of a particular setup, which is especially important for dynamically switchable implementations

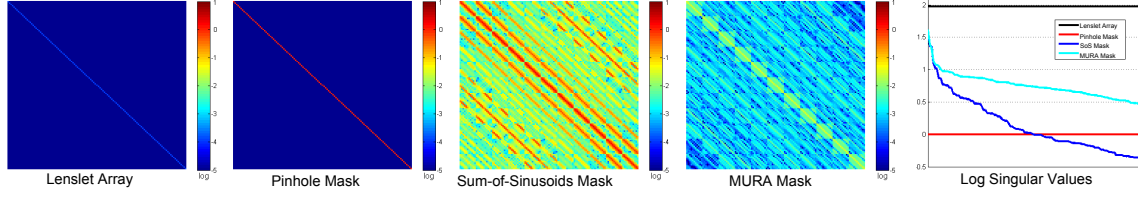


Figure 9.11: *Covariance matrices and eigenvalues for different light field multiplexing schemes. Large values and especially off-diagonal entries in the covariance matrices indicate amplification of additive noise in the sensor images. The sum-of-sinusoids mask is thus expected to perform worse than a pinhole mask for dominating dark current noise in the sensor image, which can also be inferred from the plots showing the multiplexing matrix’s singular values (right).*

[Sajadi et al., 2011].

Similarly, we compare the noise performance of various light field acquisition approaches, where the pinhole attenuation mask serves as the reference. The size of the simulated pinhole matches that of a sensor pixel. Sensor quantization and other nonlinearities are disregarded in this experiment. The plot in Figure 9.10 (right) shows that lenslets at focal distance to the sensor always perform best in terms of SNR. Among the non-refractive multiplexing methods, MURA-based attenuation masks [Lanman et al., 2008] perform very well for a dominating additive noise term, i.e. at high camera gain settings in low-light conditions. However, their SNR gain drops below that of a pinhole for an increasingly dominating photon noise term. Sum-of-sinusoids masks [Agrawal et al., 2010b; Veeraraghavan et al., 2007] always perform worse than a pinhole.

When considering only additive, signal-independent noise in the captured sensor images, which is most often the case in low-light environments, the noise of the reconstructed plenoptic slices can be quantified by the covariance matrix C :

$$C = \varsigma^2 \left((\Sigma \Pi)^T (\Sigma \Pi) \right)^{-1}, \quad (9.22)$$

where ς^2 is the variance of an additive, zero-mean Gaussian noise distribution in the sensor image $\mathbf{i}(\vec{x})$. Figure 9.11 shows the magnitudes of C for several light field acquisition schemes assuming $\varsigma^2 = 1$. Values larger than 1 amplify noise in the camera image and off-diagonal entries accumulate noise from different regions of the captured images. The covariance matrix of the sum-of-sinusoids (SoS) mask has many large-valued off-diagonal entries, which indicates noise amplification in the reconstruction. The matrix for MURA masks does have off-diagonal entries, but with much smaller magnitudes than SoS masks. Similar interpretations can be inferred from the plots of the singular values of the multiplexing

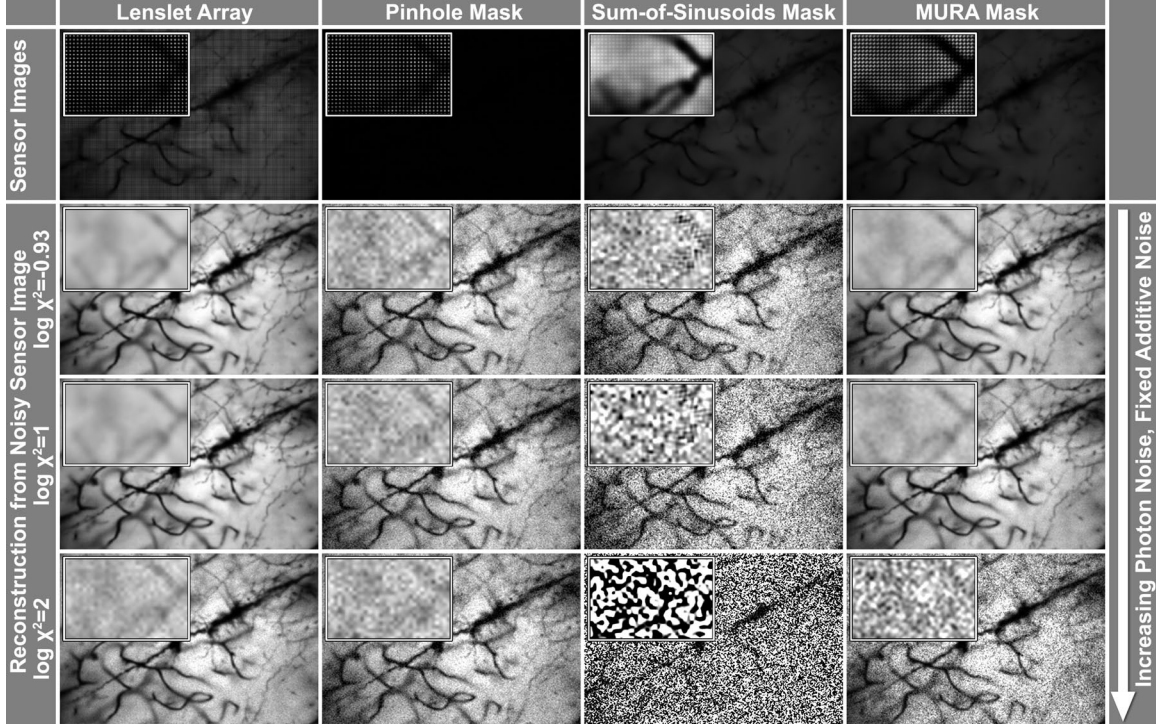


Figure 9.12: Comparison of noise amplification for different light field acquisition schemes on the golgi stained neuron dataset (lightfield.stanford.edu). Row 1 shows simulated sensor images with contrast enhanced close-ups. The other rows show a single view of the reconstructed light field from a noisy sensor image. The ratio χ^2 of signal-dependent photon noise and signal-independent dark current noise varies for the different reconstructions. Row 2 simulates a reconstruction with a dominating additive noise term, while rows 3 and 4 show the effect of an increasingly dominating photon noise term in the sensor images.

matrix $\Sigma\Pi$ in Figure 9.11 (right).

Based on the covariance analysis and predicted SNR gain (Fig. 9.10, right), we expect SoS masks to amplify sensor noise more than both pinhole and MURA attenuation masks for a dominating additive noise term. In order to validate this prediction, we simulate the acquisition of a light field with a variety of different methods and camera noise parameters (Fig. 9.12). In this experiment, we use attenuation masks with a resolution of 11×11 for each super-pixel and a similar sensor resolution. Each lenslet in the first column of Figure 9.12 covers the same area as a corresponding mask super-pixel. As expected, for low-light conditions (Fig. 9.12, row 2) lenslet arrays and MURA masks have a better noise performance than pinhole masks, whereas SoS masks perform worse (Fig. 9.12, column 3). The value $\log(\chi^2) = -0.93$ corresponds to a PointGrey Dragonfly camera [Schechner et al., 2007] where the additive noise term dominates. The lower two rows in Figure 9.12 show how

the noise increases for an increasingly dominating photon noise term up to a point where a pinhole mask performs better than even the MURA mask. The same effect was described by Wenger et al. [2005] for Hadamard codes in an illumination multiplexing application. Please note that for our analysis the exposure times of the simulated sensors were equal for each method, resulting in visible intensity differences between the sensor images (Fig. 9.12, top row).

9.6 Discussion

In this Chapter, we have introduced a framework that unifies a variety of plenoptic multiplexing approaches. Previously, these techniques have been analyzed with respect to a specific plenoptic dimension. In most cases, reconstructions have been performed exclusively in either the spatial or the Fourier domain. We have demonstrated the importance of such a unified view: certain properties, such as aliasing, can be theoretically analyzed more conveniently in the frequency domain. Other characteristics, such as noise amplification of the employed optical modulators, are easier to be evaluated in the spatial domain. We show, for the first time, how the quality of practical reconstruction mechanisms for some of the discussed techniques can be increased with spatial processing, rather than previously proposed Fourier-based algorithms. The latter, however, may require fewer computational resources.

9.6.1 Benefits and Limitations

The proposed framework generalizes multiplexing systems where the underlying signal is sampled in a super-pixel-periodic fashion. While this is the most common approach for colour imaging and light field acquisition, several methods that sample the plenoptic function in a completely random manner have been proposed [Reddy et al., 2011; Veeraraghavan et al., 2011]. Due to the lack of a regular sampling structure, these specific approaches are not supported by our framework. However, we envision multiplexing approaches that combine random plenoptic projections with super-pixel-periodic spatial sampling patterns to be an exciting avenue of future research. Our image formation unifies a wide range of previously proposed multiplexing schemes and paves the way for novel multiplexing techniques. We generalize the analysis and reconstruction to either the spatial or the Fourier domain. Practically, this allows for higher quality reconstructions of Fourier multiplexed data and the formulation of optimality criteria of employed optical modulators.

We do not propose new optical multiplexing methods, but evaluate and unify a variety of existing approaches. The theory presented in this Chapter allows knowledge of colour demosaicing, which has been built up for decades within the computer vision community, to be transferred to the reconstruction of light fields and other dimensions of the plenoptic function.

9.6.2 Future Work

As mentioned above, we would like to explore strategies that sample the plenoptic function in a random but super-pixel-periodic fashion. A combination of compressive sensing paradigms and traditional, periodic sampling approaches could prove essential in the quest for plenoptic resolution improvements beyond the Nyquist limit. Application-specific reconstruction filters, exploiting natural image statistics, could further push the boundaries of conventional image acquisition. The exploitation of natural image statistics is common practice for imaging with colour filter arrays and subsequent demosaicing. However, there is significant potential to develop similar techniques for demosaicing other multiplexed plenoptic information, for instance light fields [Levin and Durand, 2010].

9.6.3 Conclusion

The unifying theory presented in this Chapter is a crucial step toward the “ultimate” camera capturing all visual information with a single shot. Only within the last few years has the research community started to investigate approaches to acquire the plenoptic function with joint optical modulation and computational processing. Our work ties many of these new techniques to more traditional ways of sampling visual information. The proposed framework is essential for the evaluation and optimization of plenoptic multiplexing schemes of the future.

Chapter 10

Dynamic Range Boosting for Fourier Multiplexed Imaging

Optically multiplexed image acquisition techniques have become increasingly popular for encoding colour, light fields, and other properties of light onto two-dimensional image sensors. Recently, a new category of Fourier-multiplexed approaches has been proposed, with the goal of superior light transmission and signal-to-noise characteristics.

In this chapter, we show that such Fourier space encodings suffer from severe artifacts in the case of sensor saturation, i.e. when the scene dynamic range exceeds the capabilities of the image sensor. We analyze the problem, and propose new multiplexing masks and optimization methods that not only suppress such artifacts, but also allow us to recover a wider dynamic range than existing image space multiplexing approaches.

10.1 Introduction

Photography has evolved as one of the primary means by which we represent and communicate the three-dimensional world around us. Over the past decade, we have seen a push to digital photo-sensors, such as charge coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS) sensors, which have largely replaced traditional film photography.

Digital sensors, however, have a relatively limited dynamic range compared to both the human visual system and photographic film. The recent development of consumer displays that support a high dynamic range [Seetzen et al., 2004] has increased the demand for high contrast content beyond the scope of movie theaters.

In order to capture a high dynamic range image or other visual information with standard digital sensors, a variety of multiplexing techniques have been proposed. One of the most popular approaches is to successively capture different exposures of the same scene with a

single camera [Debevec and Malik, 1997; Mitsunaga and Nayar, 1999; Schechner and Nayar, 2003a]. Alternatively, multiple aligned image sensors can be employed to simultaneously capture images of a scene that are either differently exposed or filtered [Aggarwal and Ahuja, 2004; McGuire et al., 2007]. The latter approach is costly, while the former usually does not allow for acquisition of dynamic environments.

Different optical information can also be encoded into a single photograph. Examples of this approach include high dynamic range imaging [Nayar and Mitsunaga, 2000], photography with colour filter arrays (CFAs) [Bayer, 1976; Compton, 2007], and light field acquisition [Adelson and Wang, 1992; Levoy et al., 2006; Lippmann, 1908; Ng, 2005]. These techniques can be generalized as *Assorted Pixels* [Narasimhan and Nayar, 2005], where each sensor pixel captures an exposure or some other part of the plenoptic function [Adelson and Bergen, 1991]. Micro-lens arrays can be equipped with different apertures to capture high dynamic range light fields [Georgiev et al., 2009].

Recently, more sophisticated multiplexing schemes have been explored. Modulators that optically encode light properties in different spatial frequency bands have been proposed for the acquisition of light fields [Veeraraghavan et al., 2008, 2007] and occluder information [Lanman et al., 2008]. The corresponding reconstruction has usually been performed in the *Fourier domain*, although a recent analysis [Ihrke et al., 2010b] shows that a spatial reconstruction is possible. Compared to standard spatial multiplexing techniques, Fourier multiplexing methods allow for a superior light transmission and potentially increased signal-to-noise ratio (SNR) of the demultiplexed signal.

Although the effect of sensor saturation on spatial reconstructions of multiplexed data is well understood, it has so far been ignored for Fourier-based reconstruction methods. We analyze the problem and show that saturation results in severe artifacts for these approaches as the global frequency content is altered by the saturation.

Inspired by the idea of Fourier-based image reconstruction, we present a joint optical light modulation and computational reconstruction approach to boosting the dynamic range of multiplexed photographs. Previously, plausible dynamic range values were estimated from demosaicked images using heuristics [Banterle et al., 2006; Masia et al., 2009; Rempel et al., 2007] or priors on the colour distributions for the case of a single saturated colour channel in a photograph [Zhang and Brainard, 2004]. In contrast to this, we develop a numerical optimization method for Fourier-based reconstruction and dynamic range boosting of multiplexed data that precedes the demosaicking step and recovers the raw sensor data.

Restoration of clipped general [Abel and Smith, 1991] and ultrasonic [Olofsson, 2005] signals based on inequality constrained optimization has been successful, but it does not consider

the artifacts introduced by reconstructing individual channels from a saturated multiplexed signal.

10.2 Saturation Analysis in Fourier Space

The principle of Fourier multiplexed image reconstruction is illustrated in Figure 10.1. A band-limited optical signal is filtered with a periodic light modulator that is mounted directly on the sensor, such as a colour filter array (CFA) or an array of different neutral density (ND) filters. The optical filtering, or multiplication, of incident light and modulator is, according to the convolution theorem, equivalent to a convolution in the Fourier domain. The convolution enables a copy mechanism that allows Fourier multiplexing approaches to directly encode desired visual information in different frequency bands of the image. As shown in Figure 10.1 (lower right), these can then be cropped in the Fourier transform of the sensor image and individually transformed back to the spatial domain.

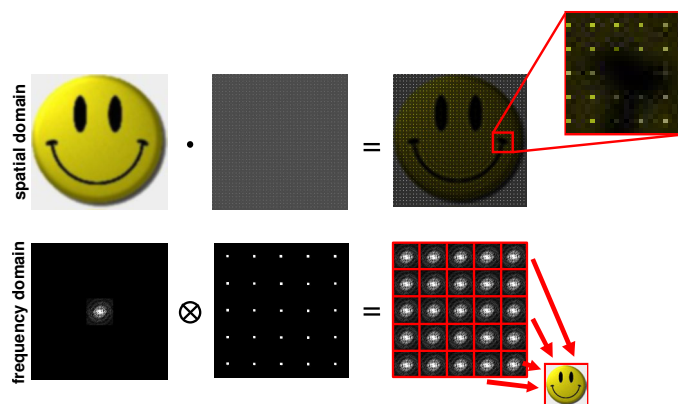


Figure 10.1: A simulated, band-limited optical signal is filtered with a semi-transparent mask placed directly in front of the image sensor. This multiplication (upper row) corresponds to a convolution in the frequency domain (lower row). Fourier multiplexing exploits the resulting copy mechanism to capture several differently filtered versions of a scene in different frequency bands of a photograph.

Although specialized Fourier multiplexing masks have so far only been used to acquire light fields [Lanman et al., 2008; Veeraraghavan et al., 2008, 2007] it is straightforward to extend this concept to colour filter arrays (see Section 10.5). The potential of Fourier-based reconstruction methods for colour demosaicking of raw CFA imagery, even with standard Bayer filters, has only recently been discovered [Alleyson et al., 2005].

In order to understand the effects of sensor saturation on the individual Fourier tiles, let us consider a 1D unmodulated, band-limited scene as shown in Figure 10.2 (left). Due to

the band-limited nature of the signal, only low image frequencies are present that occupy a single frequency band (Fig. 10.2, lower left), while all other frequency bands are empty and reserved for additional information to be optically encoded with a modulator. Sensor saturation, as illustrated in Figure 10.2 (right), destroys the band-limited nature of the signal by modifying the frequency content of the signal. The frequency bands that were originally reserved for additional data are corrupted with high frequency components.

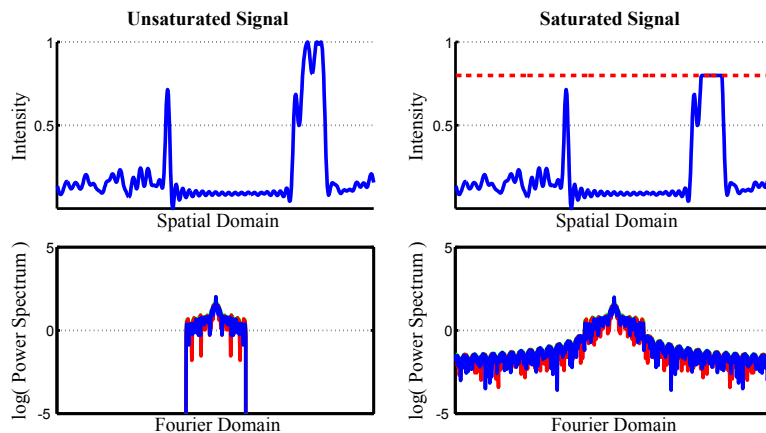


Figure 10.2: A band-limited signal (upper left), consisting of a single scanline taken from a high dynamic range image, and the same signal clipped at 0.8 maximum intensity level (upper right). The Fourier transform of the original scanline is band-limited (lower left), while the Fourier transform of the clipped version of the same scanline is corrupted by high frequency components.

In order to understand the effect of saturation, let us now consider a simple example scene, shown in Figure 10.3, where a constant white signal is captured through an attenuation mask consisting of different neutral density (ND) filters. The Fourier transform of the underlying signal is a single value for the DC term. According to the Poisson summation formula, the Fourier transform of the unsaturated periodic sensor image is a series of differently weighted Fourier peaks as seen in Figure 10.3 (b). If the dynamic range of this mask-modulated scene exceeds that of the sensor, some of the ND filters are saturated (c), and the global scale of the Fourier copies is altered (d). In this case only parts of each super-pixel, that is one spatial period of the modulator, are saturated; therefore, we refer to these regions as being *partially* saturated.

Another case of saturation occurs when the scene regionally exhibits a very high dynamic range. Here, one or more spatially neighboring super-pixels are *fully* saturated. Due to the spatial structure of the super-pixels being completely removed in such regions, the local information cannot be copied into the high frequency bands of the image. Instead, other frequency bands are corrupted.

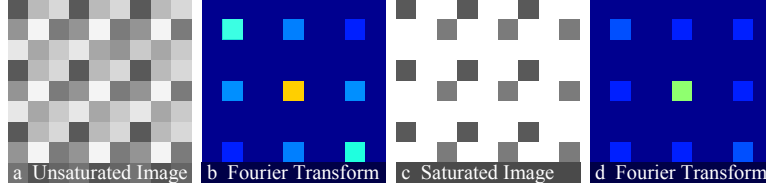


Figure 10.3: A constant signal captured through a 3×3 pattern of neutral density filters (a) and its Fourier transform (b). Partial saturation of this sensor image (c) causes changes in the magnitudes of the signal’s Fourier copies (d). Peak magnitudes are colour-coded.

In summary, saturation has a significant impact on the performance of Fourier multiplexing techniques. Saturation, if not dealt with properly, can introduce severe artifacts in multiplexed information. In the next section, we demonstrate how to recover lost information and extend the dynamic range of the captured content.

10.3 Dynamic Range Boosting

Based on the analysis presented in the last section, we now introduce a Fourier-space optimization approach to recover dynamic range information in saturated image regions. The optimization is based on the idea that monochromatic neutral density filter masks can be used to create differently scaled copies in Fourier space. In the absence of saturation, these neutral density filters produce multiple Fourier copies of the image with different intensities. The reconstructed image can represent the same dynamic range as that of the corresponding spatial reconstruction [Nayar and Narasimhan, 2002]. In the presence of saturation, however, we have seen that these copies are corrupted, even if only the most transmissive among the neutral density filters saturates. The optimization method derived in the following not only suppresses these artifacts, but results in an increased dynamic range compared to spatial methods.

To avoid the resulting artifacts, we make use of two pieces of information about the signal we are recording. First, the original signal before modulation is band-limited. Second, the filter mask should create Fourier tiles with different (known) amplitudes, corresponding to the coefficients of the spatial Fourier basis. We are thus able to formulate an error measure ϵ in Fourier space, which incorporates these two constraints:

$$\epsilon = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \|T_i(f_x, f_y) - T_j(f_x, f_y)\|_2^2 \quad (10.1)$$

Here, $T_i(f_x, f_y) = s_i \cdot (F_i\{L(x, y)\} + \eta_i)$ is a tile describing a single copy of the sensor image $L(x, y)$ in the Fourier domain. η_i is the Fourier-transformed sensor noise. F_i is the Fourier transform that maps a full-resolution image from the spatial domain into the subset of the frequency space that is spanned by tile T_i . The scaling factors $1/s_i$ describe the relative amplitudes of the individual tiles, as introduced by the modulation mask. For notational simplicity we assume in the following that all tiles have been normalized by dividing through the corresponding factor after capture.

We are now splitting the desired, mask-modulated spatial target image L into a part where the sensor pixels saturate, and a second part where they do not. That is, $L = L_{\text{unsat}} + L_{\text{sat}}$, where

$$\begin{aligned} L_{\text{unsat}}(x, y) &= \begin{cases} L(x, y) & ; L(x, y) < L_{\text{max}} \\ 0 & ; \text{else} \end{cases} \\ L_{\text{sat}}(x, y) &= \begin{cases} 0 & ; L(x, y) < L_{\text{max}} \\ L(x, y) & ; \text{else} \end{cases} . \end{aligned}$$

The corresponding relationship in Fourier space is: $F\{L\} = F\{L_{\text{unsat}}\} + F\{L_{\text{sat}}\}$. The individual tiles are now given as

$$T_i = F_i\{L_{\text{unsat}}\} + F_i\{L_{\text{sat}}\} + \eta_i. \quad (10.2)$$

The term $F_i\{L_{\text{unsat}}\}$ can readily be computed from the captured image, and represents measured data. $F_i\{L_{\text{sat}}\}$ includes the unknown variables (the non-zero subset of L_{sat}) that, in the presence of saturation, will cause the saturation noise in the Fourier domain. Combining Equations 10.1 and 10.2 yields

$$\begin{aligned} \varepsilon = \sum_{i=1}^{N-1} \sum_{j=i+1}^N & \| F_i\{L_{\text{unsat}}\} - F_j\{L_{\text{unsat}}\} \\ & + F_i\{L_{\text{sat}}\} - F_j\{L_{\text{sat}}\} + \eta_i + \eta_j \|_2^2, \end{aligned} \quad (10.3)$$

We assume that the sensor noise is independently distributed in the spatial domain, and observes a Gaussian distribution in the per-pixel image intensities. Thus, η_i has a uniform power spectrum with a Gaussian characteristic in each Fourier coefficient. This simple noise

model allows us to use a quadratic error norm for optimization in Fourier space.

We encode ε in a linear system of equations, where we optimize the *spatial* pixel intensities L_{sat} using an error measure defined *in Fourier space*. We show a simplified example assuming one copy at the *DC* peak and one copy of equal scale in a higher frequency band:

$$\min \left\| \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} F_1 \\ F_{DC} \\ F_1^* \end{bmatrix} (L_{\text{unsat}} + L_{\text{sat}}) \right\|_2^2. \quad (10.4)$$

In matrix notation, Equation 10.4 becomes $\min \|RF(L_{\text{unsat}} + L_{\text{sat}})\|_2^2$, with R encoding the relationship between the different Fourier copies, and F performing the transformation from the spatial domain into the single frequency tiles. Since we are minimizing differences, matrix R is not of full rank, which we compensate for by adding a regularization term S on the spatial reconstructions of the individual Fourier tiles. In this regularizer, we encode a spatial smoothness constraint (i.e. a curvature minimizing term), which is justified by the assumption of a band-limited signal:

$$\min \|(RF + \alpha S)(L_{\text{unsat}} + L_{\text{sat}})\|_2^2. \quad (10.5)$$

Differentiating Equation 10.5 with respect to L_{sat} and setting the gradient to zero, we finally obtain a least squares description of our error measure:

$$\begin{aligned} (F^* R^* R F + \alpha S^* S) L_{\text{sat}} = \\ - (F^* R^* R F + \alpha S^* S) L_{\text{unsat}}. \end{aligned} \quad (10.6)$$

Note that the right hand side of this system is constant and represents our image space measurements.

Equation 10.6 describes the final optimization process that we use to recover lost dynamic range. The equation can be solved using any linear solver. In our work, we employ conjugate gradients for least-squares (CGLS), combined with image-space operations, which allows us to represent the system without explicitly forming the matrix.

10.4 Experimental Validation

10.4.1 Prototype

To validate this optimization procedure on real data, we built a digital camera from a flatbed scanner [Wang and Heidrich, 2004]. This camera is easy to construct, provides a very high-resolution, and is large-scale, which makes it easy to use simple transparencies as masks. Rather than focusing the camera directly on the sensor and applying colour filters outside the camera as proposed by Wang and Heidrich [2004], we attach our modulation masks directly onto the glass plate, where the incoming light as well as the underlying sensor elements are focused. A holographic diffuser, which allows the incident light to form an image, is also mounted over the filter on the scanner glass plate as seen in Figure 10.4. It simultaneously serves as a band-limiter for the light incident on the sensor.

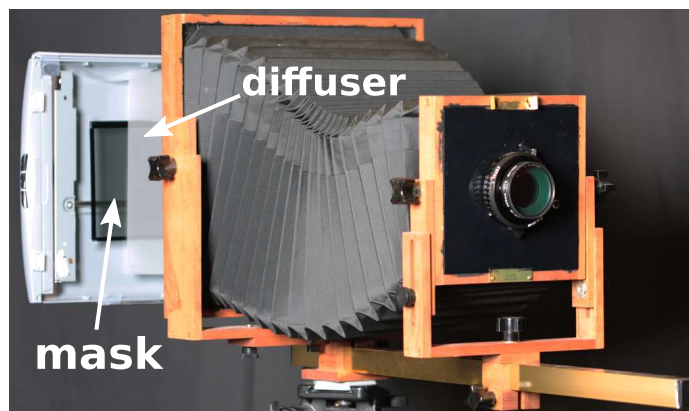


Figure 10.4: Our prototype is a large format high-resolution scanner camera. The filters are mounted under a diffuser on the scanner’s glass plate.

The masks are high resolution RGB digital images, exposed onto photographic film using light-valve technology (LVT). Colour transparencies with a resolution of up to 2032 dpi and a high contrast can be ordered at professional print service providers such as Bowhaus (www.bowhaus.com). We use 4”x5” transparencies and perform scans with 2400 dpi.

Due to slight mis-registrations (rotation and shift) of the filter in front of the sensor in our prototype camera, as well as dust and scanner sensor artifacts, the point spread functions (PSFs) of the filter tiles in the Fourier domain do not exactly correspond to the filter specification before the print. In order to calibrate for these effects, we estimate the PSFs of the individual filter tiles in the Fourier transform of a calibration image.

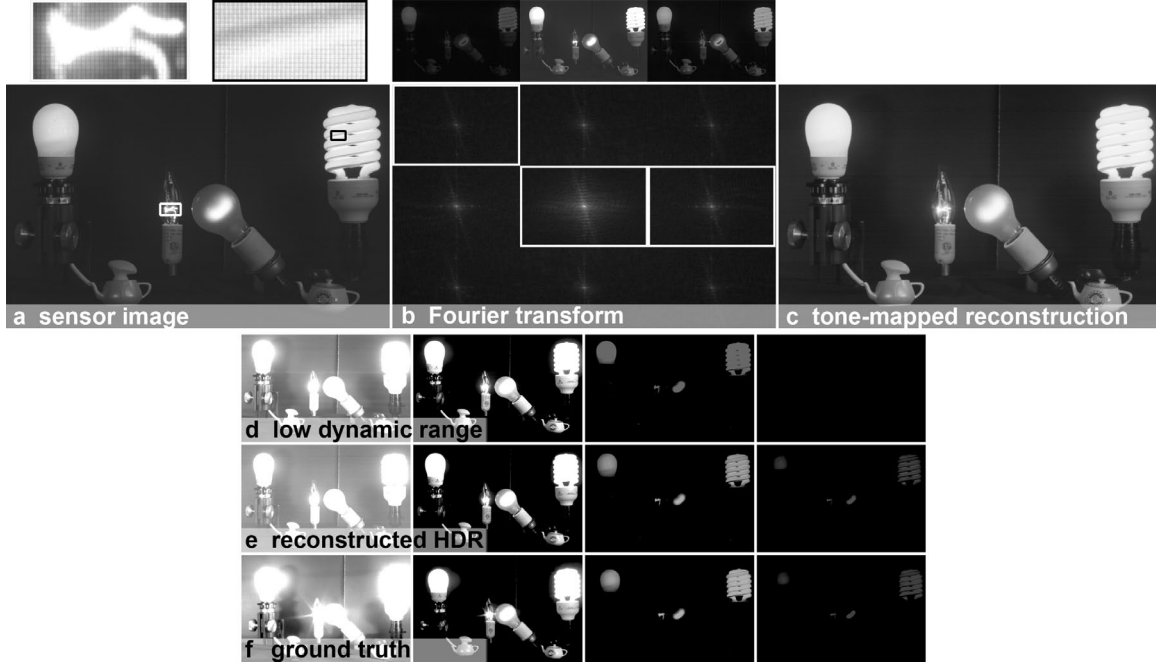


Figure 10.5: *A mask-modulated LDR image captured with a prototype scan camera (a). The pattern introduced by the mask and saturated regions are enlarged on the upper left. The Fourier transform of the captured image (b) contains nine copies, of which three (b, top) are individually transformed into the spatial domain and reveal saturation artifacts (see Figure 10.6 for enlargement). The tone-mapped result (c) does not contain the mask pattern or Fourier copies. The right part shows a linearly mapped exposure sequence of an unmodulated LDR image (d) and our reconstruction (e) compared to ground truth (f).*

10.4.2 Optimization Results

An example scene containing saturated regions is shown in Figure 10.5. Two magnifications of the multiplexed image captured with our scan camera prototype (2628×1671 px) are shown on the upper left (a, top). Three of the Fourier tiles are individually transformed back into the spatial domain and presented above the Fourier transform of the image (b). One of them is also enlarged in Figure 10.6. Note that they are differently affected by saturation and sensor noise, as well as scanner artifacts. Figure 10.5 (c) shows the tone-mapped result of our HDR reconstruction. In order to compute it, we performed our optimization on a grid of 3×3 tiles in the Fourier domain, each with a resolution of 876×557 pixels. The data was in the range of 0-1 and our optimization algorithm converged in about 750 iterations to a residual of 10^{-6} . The large number of iterations can be explained by the very high noise level of our camera prototype. The dynamic range of the captured scene is extended by a factor of 1.58 in this case. Note that this factor is obtained *after* the tiles have already been

divided by their relative intensity scale factors s_i , as described in Section 10.3. Therefore, the factor of 1.58 is an additional improvement on top of the one that would be obtained by using the same neutral density filter array in combination with Assorted Pixel spatial reconstruction [Nayar and Mitsunaga, 2000]. The total gain in dynamic range compared to an unmodulated image is 1.58 times the contrast of the used filters.

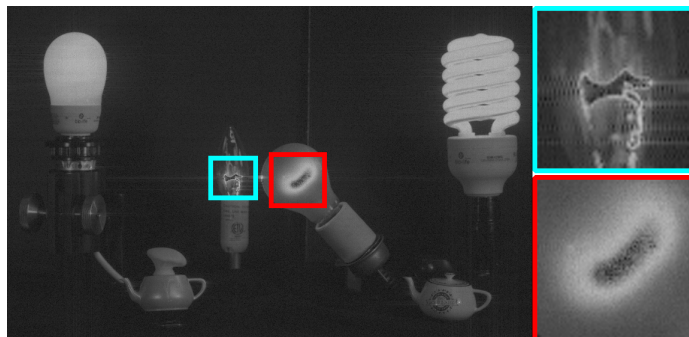


Figure 10.6: *Sensor saturation yields ringing and other artifacts in Fourier multiplexed data. This image is the spatial version of one of the higher frequency tiles in Figure 10.5.*

For validation, we compare our result to a high dynamic range ground truth image that was generated by combining 12 exposures of an SLR camera located next to the scan camera. As seen in the multi-exposure sequence in the bottom row of Figure 10.5 (e), the dynamic range can be faithfully recovered. Note in particular the structure recovered in the cold fluorescent light bulb. The depicted LDR image (d) was photographed using our scan camera without the attenuation mask, and lacks details in bright image regions. The SLR image is shown at the bottom (f).

Another example scene is presented in Figure 10.7. Here, we also show results of our optimization for saturation in dark regions (highlighted in green), as well as in bright image parts (red highlights). The magnifications in the lower row illustrate how saturation (c,d) is recovered (e,f) using our approach. Due to the noise floor in the sensor image, our approach can, in this case, not successfully push the intensity values below the blacklevel in dark image regions. Although saturated pixels in the close-ups (c,d) should be entirely flat, errors are introduced by camera noise and the calibration image of the mask which distort the final reconstruction slightly.

10.4.3 Comparison to Spatial Reconstruction

We show comparisons of our image reconstruction and the Assorted Pixels approach [Nayar and Mitsunaga, 2000] for a 1D and a 2D scene in Figures 10.8 and 10.9 respectively. For

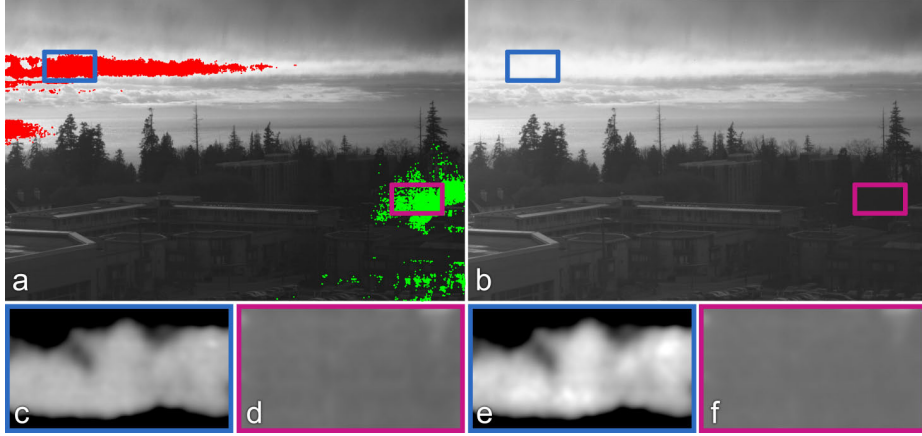


Figure 10.7: Outdoor scene captured with our scan camera prototype. The sensor image (a) has saturated parts indicated in red (bright) and green (dark); (b) is the reconstructed HDR image. The magnified parts show linearly mapped intensities before (c,d) and after (e,f) reconstruction for bright and dark saturation respectively.

the former case we simulated a 1D sensor image by multiplying a repetitive array of neutral density (ND) filters with a test signal taken from a real HDR image and saturating it at 8% of the dynamic range of the scene. The Assorted Pixels approach performs the reconstruction by dividing by the ND mask values followed by a bi-cubic interpolation to estimate the saturated pixels (blue line in Figure 10.8, right). Alternatively, we can apply our Fourier-based reconstruction approach to get a much better estimate of the original function (cyan-coloured line in Figure 10.8).

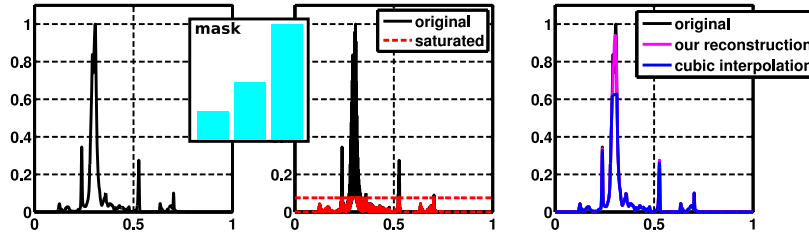


Figure 10.8: A band-limited 1D signal (left) is modulated with an attenuation mask pattern (one of the repeating tiles shown in inset), and captured by a simulated sensor with a limited dynamic range (red dotted line, center). Our reconstruction (right, magenta) performs better than previously developed interpolation methods (right, blue).

The 2D comparison presented in Figure 10.9 shows three different exposures of an HDR image (left column), reconstructions of a simulated sensor image with the Assorted Pixels approach (center column), as well as our method (right column). The sensor image was saturated at 7% of the dynamic range of the scene. The mask was in this case a repeating pattern of 2×2 ND filters with transmission values of 1, 0.5, 0.25, and 0.125. As expected,

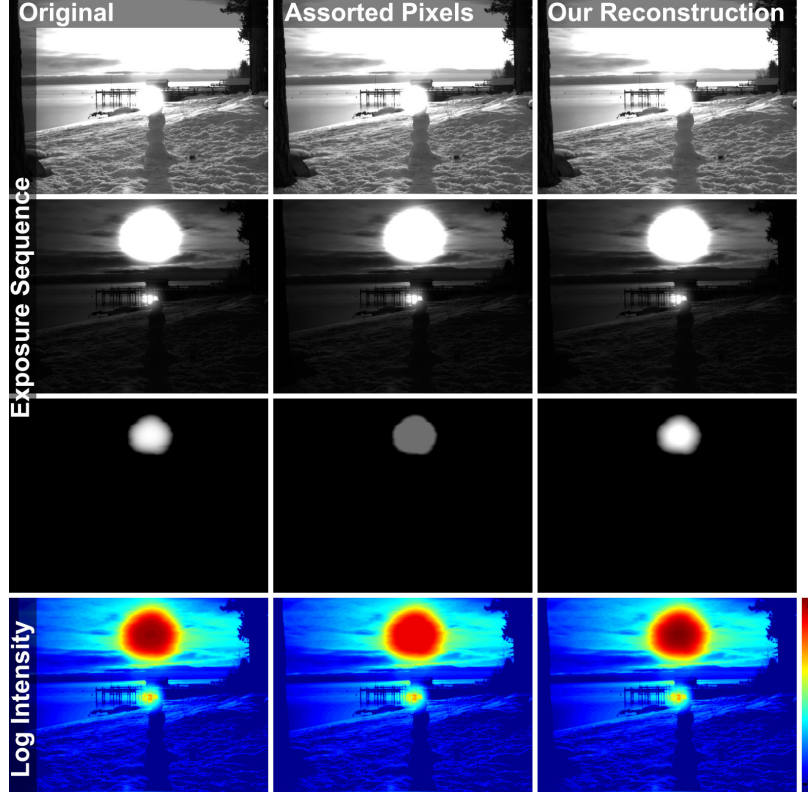


Figure 10.9: A 2D comparison of previously proposed interpolation and our reconstruction for a repeating 2×2 pattern of neutral density filters. The simulated sensor image is saturated at 7% of the dynamic range of the scene. Three different exposures of the HDR images are shown in rows 1-3 and colour coded intensities are visualized in the lower row.

dividing by the mask and performing a bi-cubic interpolation can only recover a maximum image intensity that is defined by the ND filter with the lowest light transmission. Our approach can boost the recovered intensity beyond that limit.

10.4.4 Limitations

For very large areas of saturation, we expect our method to eventually produce unsatisfying images, because the regularization term results in an over-smoothing of such regions. In order to test the performance of our algorithm under extreme situations, we photographed a scene that exhibits a very high dynamic range with a number of different aperture settings using our scanner camera prototype (see Figure 10.10). As the size of fully saturated regions grows, an unregularized solution of Equation 10.6 would become less stable, thus the regularization term counteracts this by filling in smooth image data. The gain in dynamic range is highest for an aperture of 32 (c), because the saturated region is not too large, and

the lost dynamic range of the original image is higher than that of the smaller apertures. The gain, however, gets lower for aperture 22 (d), as the equation system becomes more ill-conditioned; the regularization term starts to dominate and over-smooth the solution. The effect is even stronger for aperture 11 (e), where we stopped the optimization after 1000 iterations without convergence. We expect a smooth and plausible solution with more iterations, this example just demonstrates slower convergence behavior for larger saturated regions and the effect of over-smoothed solutions. Artifacts in the captured images with aperture settings 22 (d) and 11 (e) are caused by sensor blooming in our prototype camera.

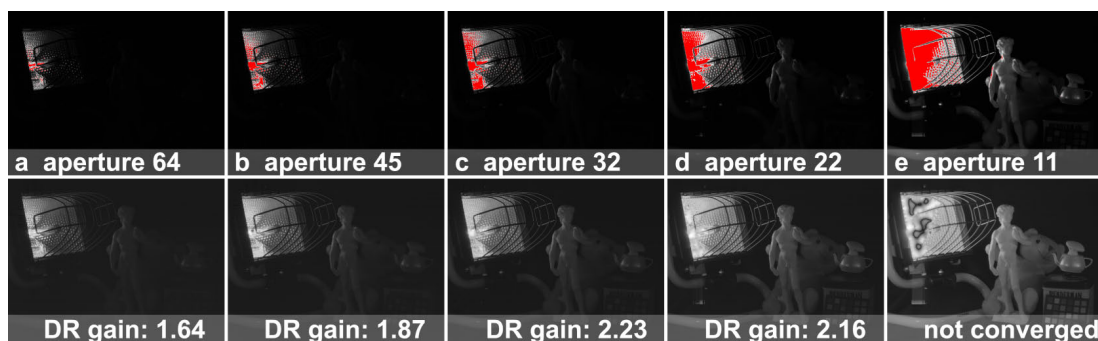


Figure 10.10: A set of images captured with different camera apertures. For very large fully saturated regions, the optimization over-smoothes the solution. Saturated sensor images are shown in the upper row, tonemapped HDR reconstructions in the lower. The right case did not yet converge for the maximum of 1000 iterations.

10.5 Combined Colour and HDR Optimization

So far we have only considered the grayscale case. In this section, we show how to incorporate our reconstruction approach into a Fourier multiplexed capture of colour images. This is practically useful, as we introduce a novel colour filter array that, in combination with our dynamic range boosting technique, allows us to simultaneously capture RGB imagery and a high dynamic range.

Our colour filter array is inspired by a recent analysis of standard Bayer CFAs in the Fourier domain. As discovered by Alleyson et al. [2005], a raw sensor image captured through a Bayer pattern inherently contains four differently filtered copies of the image in the Fourier domain. Specifically, these copies are one luminance tile ($R + 2G + B$), two *similar* chrominance tiles ($R - B$), and a fourth tile ($R - 2G + B$). Instead of redundantly encoding the same chrominance tile twice ($R - B$), we propose a CFA design that contains two *different* chrominance tiles ($R - G$) and ($B - G$) as well as two differently scaled luminance tiles ($R + 2G + B$) in the Fourier transform. Our CFA, just like a Bayer pattern,

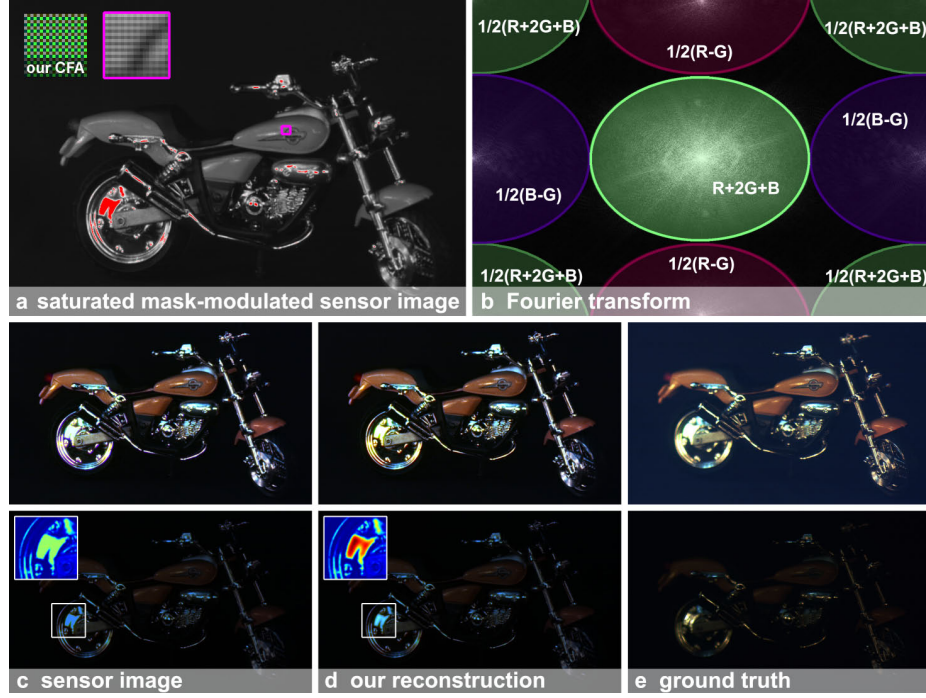


Figure 10.11: A sensor image captured with our prototype through a novel CFA with saturated regions in red (a). The mask creates two differently scaled luma and two chroma copies in the Fourier transform (b). The dynamic range of the captured image can be significantly extended using our optimization approach as seen in the recoloured magnifications in the lower images.

comprises a repeating pattern of coloured super-pixels with a resolution of 2×2 pixels. Rather than sampling each colour channel directly, the colour distribution for each of our CFA's super-pixels in the spatial domain is

$$R = \begin{pmatrix} 0.5 & 0.25 \\ 0 & 0.25 \end{pmatrix}, G = \begin{pmatrix} 0.5 & 0.25 \\ 0.25 & 1 \end{pmatrix}, B = \begin{pmatrix} 0.5 & 0 \\ 0.25 & 0.25 \end{pmatrix}. \quad (10.7)$$

Figure 10.11 shows a raw sensor image captured with our scanner camera prototype through the proposed colour filter array. The photograph contains saturation (a); the spatial pattern of our CFA and a corresponding region of the sensor image are magnified (a, upper left). We illustrate the chrominance and differently scaled luminance tiles in the Fourier domain (b).

Our optimization can, in this case, be applied to the two different luminance tiles. This implies that no colour information can be recovered in saturated regions, as the dynamic range is only boosted in the luminance channel. The lower part of Figure 10.11 shows

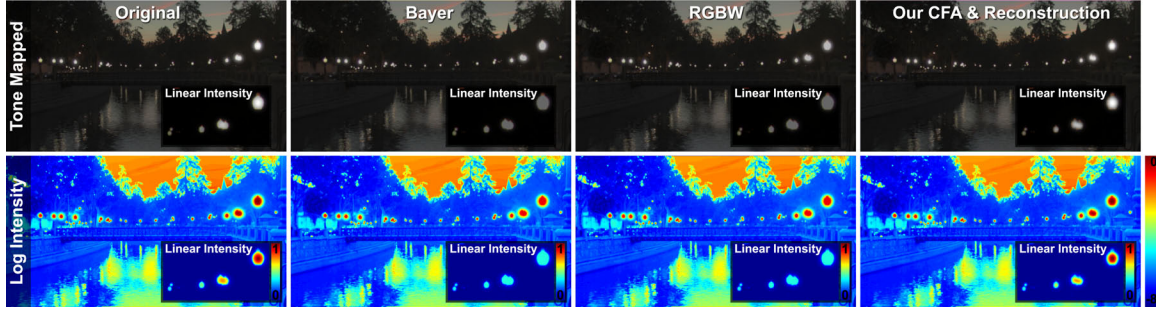


Figure 10.12: A comparison of imaging through a standard Bayer colour filter array (center left column), an RGBW CFA with the same light transmission (center right column), and our CFA with the proposed reconstruction (right column). A sensor image was simulated for all three cases with a dynamic range of 4% of that of the photographed scene. The upper row shows the tone mapped original image and the three reconstructions with linearized magnifications. The lower row shows colour coded logarithmic intensities and linear intensities in the magnifications.

two exposures for different reconstructions of the example scene. The left column (c) is directly reconstructed from the saturated sensor image. It therefore represents an image that is comparable to one captured through a Bayer pattern. Our reconstruction (column d) can significantly increase the dynamic range and reconstruct saturated image regions. The right column (e) shows an HDR image assembled from photographs of the scene without our mask, but with pure red, green, and blue filters, and 3 different aperture settings each. A gain factor of 1.9 in dynamic range could be achieved for our reconstruction, as compared to the standard reconstruction from the sensor image. The high noise level of our camera prototype results in a different black level for the ground truth and mask modulated images. Colour differences can be explained by imperfect PSF calibration.

A synthetic result is shown in Figure 10.12. For this example, the original HDR image (left column) was modulated by a Bayer pattern and reconstructed with standard colour demosaicking (center left column). Additionally, we simulated a sensor image by applying an RGBW CFA (center right column) with the same mean light transmission as a Bayer pattern and our CFA. The RGBW CFA consists of one red, green, and blue filter each, and an additional white ND filter. A sensor image with our CFA was simulated and saturated at the same intensity as the Bayer and the RGBW sensor images. The result of our reconstruction is shown in the right column of Figure 10.12. We can see that the dynamic range of the reconstruction can be significantly boosted by our approach.

Many alternative CFA designs are possible, the one presented in this section is just one example for a design that is compliant with our optimization approach. The average light transmission of our CFA is similar to that of a Bayer pattern and the same band-limitation

requirements apply, as both CFAs consist of repeating 2×2 super-pixels.

10.6 Discussion

In summary, we have presented an analysis of saturation artifacts for Fourier multiplexed imaging in this chapter. Based on this analysis, we have proposed an optimization framework that uses optically encoded information in the Fourier transform and facilitates the suppression of such artifacts, as well as an expansion of the dynamic range. We have shown limitations of the employed optimization and presented an approach to incorporate it into the capture of Fourier multiplexed colour images.

In the future we wish to experiment with alternative colour masks, and test our optimization in the context of further optical Fourier multiplexing applications, such as light fields. We believe that the loss of resolution in all Fourier reconstruction methods could be compensated by incorporating priors on channel correlations, which could also lead to better results of our optimization. An interesting avenue of future research is the possibility of spatial reconstruction of Fourier multiplexed data. In this case our optimization could be performed as a pre-processing step before the actual image reconstruction, e.g. colour demosaicking in the case of multiplexed colour channels.

Chapter 11

Discussion and Conclusion

In this thesis, we have discussed a variety of approaches to plenoptic image acquisition and display using the joint design of optical systems and computational processing. We have introduced technologies for glasses-free 3D display using stacks of attenuating and polarization-rotating layers. While the degrees of freedom of these displays are limited, convincing 3D effects can be achieved. Computational probes, requiring precise control over the emitted light, have been introduced as high-dimensional displays that facilitate visualization and reconstruction of transparent, refractive media. By using adaptive coded apertures, we have shown that the depth of field of projection systems can be extended while maintaining a high light efficiency. Computational optics, as introduced in this thesis, can enhance the capabilities of the human visual system with combined optical modulation and real-time processing. Finally, we have presented new colour filter arrays and corresponding optimization schemes to enhance the dynamic range of colour photographs and we have developed a mathematical framework that unifies plenoptic image acquisition and reconstruction.

11.1 Discussion

We briefly re-iterate over the contributions of all proposed methods in this section. A discussion of each approach within the scope of computational image acquisition and display is included as well. More specific discussions on details of individual techniques can be found in the respective chapters of this thesis.

Multi-layer light field displays, as introduced in Chapters 3 and 4, are practical alternatives to conventional glasses-free 3D displays. Compared to parallax barriers and integral imaging they provide a higher resolution and extended depth of field without compromising light efficiency. We present two prototype implementations: inexpensive, static displays composed of layered transparencies and dynamic stacks of LCDs. We establish the theoretical limits of all attenuation-based multi-layer displays and demonstrate how these displays can,

alternatively, achieve 2D high dynamic range display. Both attenuating and polarization-rotating layers support the display of 3D objects outside the device enclosure as well as view-dependent and global illumination effects within the 3D scenes. Optimal layer decompositions are computed with tomographic principles; real-time framerates can be achieved with a GPU-based implementation of the SART algorithm.

The degrees of freedom of multi-layer displays are limited because arbitrary, uncorrelated 4D light fields cannot be displayed. Yet these devices create convincing 3D effects. This indicates that the human visual system does not necessarily require physically correct solutions but rather perceptually plausible results. *Perceptually-driven optimization*, however, is an active area of research (e.g., Mantiuk et al. [2011]) and often far from being *computationally efficient*. Decomposing high-resolution light fields into multiple layers, for example, requires a significant amount of computational resources, even for linearized tomographic formulations. *Globally-optimal* solutions to the underlying nonlinear formulations have the potential to optimize numerical errors, but the quality of all light field displays should ideally be evaluated and optimized with perceptual error metrics.

With adaptive coded apertures, we introduce next-generation auto-iris systems to projection displays (Chapter 5). Programmable projector apertures allow the depth of field of a projector to be increased or, alternatively, the temporal contrast of presented imagery to be optimized. The employed algorithm for computing adaptive attenuation patterns represents a tradeoff between computational efficiency and perceptual optimality. Limitations of human perception with respect to contrast resolving capabilities are exploited to increase the light efficiency and depth of field of the optical system while allowing real-time framerates.

The concept of computational optics for direct view by a human observer is introduced in Chapter 8. By combining precisely-controlled optical modulation of an observed scene with perceptually-driven computation, we show how some of the capabilities of the human visual system can be enhanced. Specific applications include contrast and colour saturation manipulation, colour de-metamerization, preattentive object highlighting, and visual aids for the colour blind.

For all displays designed for human perception, the complex interplay between perceptually-driven error metrics, computational efficiency, and constraints imposed by the employed hardware is of crucial importance. Ideally, the perceptual error of presented content is minimized, taking hardware constraints into account. The human visual system, however, is highly nonlinear, resulting in inefficient computations; real-time framerates or other requirements may prevent sophisticated perceptual error metrics from being practical. Application-specific tradeoffs between perceptual optimality, computational efficiency,

and hardware requirements are, in practice, the only viable solution for any computational display.

In Chapters 6 and 7, we present a new approach to capturing refractive media using light field probes. We demonstrate that a 4D encoding of the positions and outgoing directions on a probe surface has many advantages compared to only coding two dimensions, as in conventional approaches. The proposed probe prototypes have a thin form factor, are portable, and can inexpensively be constructed with lenslet arrays and inkjet transparencies. Furthermore, we show how a sparse set of 3D control points and dense gradient information of refractive media can be reconstructed from a single photograph.

The high-dimensional patterns emitted by computational probes are designed to optically encode properties of the observed physical medium. As opposed to standard displays, these probes and corresponding codes do not need to be optimized for human perception but with respect to the employed hardware and computational processing. Numerical methods to determine a globally-optimal encoding strategy should, in this case, maximize the code entropy while considering the hardware constraints imposed by display and camera hardware.

The mathematical framework introduced in Chapter 9 generalizes multiplexed photography to all dimensions of the plenoptic function. We show that this framework not only allows us to evaluate and optimize the optical components of computational cameras, but also subsequent reconstructions. With the novel colour filter arrays and a corresponding optimization approach discussed in Chapter 10, we present a new computational camera design that extends the dynamic range of captured photographs.

In comparison, the algorithms driving computational displays have a major advantage over those in computational cameras: the presented image content is known a priori. This knowledge allows the data to be pre-processed in a content-adaptive fashion while optimizing the result for a human observer. Computational processing in cameras, on the other hand, often relies on statistical priors for natural scenes, which only describe common properties, such as frequency distributions, of a class of images rather than the recorded photograph itself. Furthermore, photographs, as opposed to displayed images, usually include camera noise which has a significant impact on the algorithms applied in post-processing. Deconvolution, for instance, is highly sensitive to noise; while coded aperture projection allows simple deconvolution algorithms to be applied for image decomposition, similar approaches usually fail in the presence of noise, as is the case in computational photography.

11.2 Future Work

The wide variety of techniques, advancing computational plenoptic image acquisition and display discussed in this thesis enables a similarly diverse body of possible future work.

While the multi-layer displays discussed in Chapters 3 and 4 have proven successful in light field synthesis at interactive framerates, the degrees of freedom of these approaches are limited. The underlying formulations have the potential to be solved with a higher quality, at the cost of increased computational complexity, using nonlinear optimization techniques. Such solutions could improve the field of view, depth of field, and support for global illumination effects for all multi-layer light field displays. A temporally multiplexed decomposition of the light fields, as proposed for dual-layer architectures by Lanman et al. [2010], could further enhance the capabilities of dynamic multi-layer designs. For interactive applications, requiring real-time framerates, the linear solutions may be the most practical, but for off-line content, such as 3D movies, the increased computational complexity of nonlinear approaches may be acceptable for optimized image quality.

Adaptive coded aperture projection systems, as introduced in Chapter 5, optimize the trade-off between light transmission, depth of field extension, and image degradation perceived by a human observer. While the employed attenuating apertures have the advantage of being easily switched on or off, enabling an auto-iris mode and a standard projection mode of the display, dynamic spatial light modulators that only modulate the phase of the projected light could achieve similar depth of field extension without compromising light efficiency at all.

Considering recent advances in display technology and wearable computing, the optical see-through prototypes presented in Chapter 8 could be miniaturized or even integrated into computational contact lenses, car windshields, ski goggles, or motorcycle visors. With practical implementations of such displays, new application-specific algorithms can be explored. A combination of the light-modulating approach discussed in this thesis with traditional optical see-through augmented reality, that is additively overlaid synthetic content, would further extend the impact of our concepts.

Computational displays and optics synthesize or modulate light for observation by the human visual system. The perceived image quality of all of these approaches could be significantly improved by employing perceptually-motivated error metrics (e.g., Mantiuk et al. [2011]). However, there is much potential in adapting these metrics to be computationally efficient.

As all displays presented in this thesis, the lenslet-based light field probes in Chapters 6 and 7 are only prototypes. More sophisticated display technology, such as holograms, could maximize the resolution of these probes. Extending the control over emitted light beyond the spatio-angular dimensions, for instance by including the colour spectrum and temporal variation, will help traditional structured illumination and reconstruction approaches to transcend the limitations that conventional displays impose on these methods. The reconstruction of transparent, refractive objects, as discussed in this thesis, requires the absorption of the medium to be negligible, because refraction is directly encoded in visible changes of colour and intensity. For most natural objects, however, the absorption is not completely negligible. Novel coding schemes and optical setups could improve the practical application of high-dimensional probes by explicitly separating absorption from refraction. Extending the presented reconstruction methods to a wider class of refractive media would further improve their impact.

Finally, the multiplexed acquisition of the plenoptic function with standard sensors would greatly benefit from sophisticated mathematical priors on the recorded scenes. While natural image statistics are commonly used for two-dimensional imagery, for instance by exploiting common frequency distributions of natural scenes, such statistics have yet to be developed for other plenoptic dimensions and especially for the correlations between them. In conjunction with these priors, compressive sensing approaches could maximize the amount of visual information that can be reconstructed from a limited number of samples.

11.3 Conclusion

The “ultimate” camera, capturing all visual information with a single shot, is a concept that would allow unprecedented control over processing captured data. Dynamic range, depth of field, focus, colour gamut, and time scale are only a few parameters that can be modified in post-processing. Applications that would greatly benefit from high-dimensional plenoptic image data include the adaptation of recorded content to a specific display, varying in size, resolution, gamut, and contrast, or reconstructions of additional, scene-specific information using computer vision techniques. One crucial requirement for the “ultimate” camera is the ability to optically modulate all plenoptic dimensions with a high precision. This level of control, however, is not only essential for maximizing the flexibility of camera design, but also beneficial for the design of computational optics that enhance the capabilities of the human visual system. Integrated into contact lenses and sunglasses, these systems could have a profound impact on our daily lives.

Plenoptic displays are yet another technology requiring high-dimensional light modulation. One incarnation of such displays are computational probes with applications to computer vision. While we have started to explore their potential for reconstructing transparent, refractive phenomena, there is a wide variety of different applications where high-dimensional probes could also make a significant impact. In any probe implementation, precise control over all aspects of the emitted light is necessary. Displays intended for observation by the human visual system, on the other hand, may not require all possible degrees of freedom. As long as the presented content is perceptually plausible, rather than physically correct, such devices have the potential to advance the state of the art of display technology. The multi-layer displays proposed in this thesis, for instance, are constrained in their ability to emit arbitrary 4D light fields, yet they create convincing 3D imagery that can be observed without glasses. With these devices, and also adaptive coded aperture projectors, desired effects, such as high-resolution light field display or light-efficient depth of field enhancement, are achieved with constrained optical control by exploiting imperfections of the human visual system.

The key factor to advance technology beyond the limits of conventional approaches to image acquisition and display is the combination of computational processing and optical design. This combination has been pursued in computational photography for a few years but has only begun to affect the design of displays and optical see-through devices. The approaches presented in this thesis are essential steps toward next-generation technology for computational cameras, displays, and probes as well as computational optics for enhancing the capabilities of the human visual system.

Bibliography

- Abel, J. S. and Smith, J. O. (1991). Restoring a Clipped Signal. In *Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1745–1748. IEEE.
- Adams, A., Talvala, E.-V., Park, S. H., Jacobs, D. E., Ajdin, B., Gelfand, N., Dolson, J., Vaquero, D., Baek, J., Tico, M., Lensch, H. P. A., Matusik, W., Pulli, K., Horowitz, M., and Levoy, M. (2010). The Frankencamera: an Experimental Platform for Computational Photography. *ACM Trans. Graph. (SIGGRAPH)*, 29:29:1–29:12.
- Adelson, E. H. and Bergen, J. R. (1991). The Plenoptic Function and the Elements of Early Vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press.
- Adelson, E. H. and Wang, J. Y. A. (1992). Single Lens Stereo with a Plenoptic Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106.
- Agarwal, S., Mallick, S. P., Kriegman, D., and Belongie, S. (2004). On Refractive Optical Flow. In *Proc. ECCV*, pages 483–494.
- Aggarwal, M. and Ahuja, N. (2004). Split Aperture Imaging for High Dynamic Range. *Int. Journal of Computer Vision*, 58(1):7–17.
- Agocs et al. (2006). A large scale interactive holographic display. In *IEEE Virtual Reality*, pages 311–312.
- Agrawal, A., Gupta, M., Veeraraghavan, A., and Narasimhan, S. (2010a). Optimal Coded Sampling for Temporal Super-Resolution. In *Proc. IEEE CVPR*, pages 374–380.
- Agrawal, A. and Raskar, R. (2007). Resolving Objects at Higher Resolution from a Single Motion-Blurred Image. In *Proc. IEEE CVPR*, pages 1–8.
- Agrawal, A., Raskar, R., and Chellappa, R. (2006). What is the Range of Surface Reconstructions from a Gradient Field? In *Proc. ECCV*, pages 483–494.
- Agrawal, A., Veeraraghavan, A., and Raskar, R. (2010b). Reinterpretable Imager: Towards Variable Post-Capture Space, Angle and Time Resolution in Photography. In *Proc. Eurographics*, pages 1–10.

- Agrawal, A. and Xu, Y. (2009). Coded Exposure Deblurring: Optimized Codes for PSF Estimation and Invertibility. In *Proc. IEEE CVPR*, pages 1–8.
- Agrawal, A., Xu, Y., and Raskar, R. (2009). Invertible Motion Blur in Video. *ACM Trans. Graph. (SIGGRAPH)*, 28(3).
- Ahumada, A. J. and Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *J. Opt. Soc. Am.*, A(2):1133–1139.
- Akeley, K., Watt, S. J., Girshick, A. R., and Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Trans. Graph.*, 23:804–813.
- Alleyson, D., Süssstrunk, S., and Héroult, J. (2005). Linear Demosaicing inspired by the Human Visual System. *IEEE Trans. Im. Proc.*, 14(4):439–449.
- Andersen, A. and Kak, A. (1984). Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94.
- Ashdown, I. (1993). Near-field photometry: A new approach. *Journal of the Illuminating Engineering Society*, 22(1):163–180.
- Ashok, A. and Neifeld, M. A. (2007). Pseudorandom Phase Masks for Superresolution Imaging from Subpixel Shifting. *Applied Optics*, 46(12):2256–2268.
- Atcheson, B., Heide, F., and Heidrich, W. (2010). CALTag: High Precision Fiducial Markers for Camera Calibration. In *Proc. VMV*.
- Atcheson, B., Heidrich, W., and Ihrke, I. (2009). An Evaluation of Optical Flow Algorithms for Background Oriented Schlieren Imaging. *Experiments in Fluids*, 46(3):467–476.
- Atcheson, B., Ihrke, I., Heidrich, W., Tevs, A., Bradley, D., Magnor, M., and Seidel, H.-P. (2008). Time-resolved 3D Capture of Non-stationary Gas Flows. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 27(5):132.
- Atkinson, G. and Hancock, E. (2005). Multi-view surface reconstruction using polarization. In *Proc. ICCV*, volume 1, pages 309–316.
- Atkinson, G. A. and Hancock, E. R. (2008). Two-dimensional BRDF Estimation from Polarisation. *Comput. Vis. Image Underst.*, 111(2):126–141.
- Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Comput. Graph. Appl.*, 21(6):34–47.

- Baek, J. (2010). Transfer Efficiency and Depth Invariance in Computational Cameras. In *Proc. ICCP*, pages 1–8.
- Baker, S. and Kanade, T. (2002). Limits on Super-Resolution and How to Break Them. *IEEE Trans. PAMI*, 24:1167–1183.
- Banterle, F., Ledda, P., Debattista, K., and Chalmers, A. (2006). Inverse Tone Mapping. In *Proc. GRAPHITE*, pages 349–356.
- Barnum, P. C., Narasimhan, S. G., and Kanade, T. (2010). A multi-layered display with water drops. *ACM Trans. Graph.*, 29:76:1–76:7.
- Barone-Nugent, E. D., Barty, A., and Nugent, K. A. (2002). Quantitative Phase-Amplitude Microscopy I: Optical Microscopy. *Journal of Microscopy*, 206(3):194–203.
- Bascle, B., Blake, A., and Zisserman, A. (1996). Motion Deblurring and Super-resolution from an Image Sequence. In *Proc. ECCV*, pages 573–582.
- Bayer, B. E. (1976). Color imaging array. US Patent 3,971,065.
- Beck, J., Prazdny, K., and Rosenfeld, A. (1983). *Human and Machine Vision*, chapter A theory of textural segmentation. Academic Press.
- Bell, G. P., Craig, R., Paxton, R., Wong, G., and Galbraith, D. (2008). Beyond flat panels: Multi-layered displays with real depth. *SID Digest*, 39(1):352–355.
- Bell, G. P., Engel, G. D., Searle, M. J., and Evanicky, D. (2010). Method to control point spread function of an image. U.S. Patent 7,742,239.
- Ben-Eliezer, E., Marom, E., Konforti, N., and Zalevsky, Z. (2005). Experimental Realization of an Imaging System with an Extended Depth of Field. *Appl. Opt.*, 44(11):2792–2798.
- Ben-Ezra, M. (2010). High Resolution Large Format Tile-Scan Camera. In *Proc. IEEE ICCP*, pages 1–8.
- Ben-Ezra, M. (2011). A Digital Gigapixel Large-Format Tile-Scan Camera. *IEEE Computer Graphics and Applications*, 31:49–61.
- Ben-Ezra, M. and Nayar, S. (2004). Motion-based Motion Deblurring. *IEEE Trans. PAMI*, 26(6):689–698.
- Ben-Ezra, M. and Nayar, S. K. (2003). What Does Motion Reveal About Transparency? In *Proc. ICCV*, pages 1025–1032.

- Ben-Ezra, M., Zomet, A., and Nayar, S. (2005). Video Superresolution using Controlled Subpixel Detector Shifts. *IEEE Trans. PAMI*, 27(6):977–987.
- Bimber, O. (2006). Augmenting Holograms. *IEEE Computer Graphics and Applications*, 26(5):12–17.
- Bimber, O. and Emmerling, A. (2006). Multifocal Projection: A Multiprojector Technique for Increasing Focal Depth. *IEEE TVCG*, 12(4):658–667.
- Bimber, O., Emmerling, A., and Klemmer, T. (2005a). Embedded Entertainment with Smart Projectors. *IEEE Computer*, 38(1):56–63.
- Bimber, O., Grundhöfer, A., Wetzstein, G., and Knödel, S. (2003). Consistent Illumination within Optical See-Through Augmented Environments. In *Proc. ISMAR*, pages 198–207.
- Bimber, O. and Iwai, D. (2008). Superimposing Dynamic Range. In *ACM Trans. Graph. (Siggraph Asia)*, pages 1–8.
- Bimber, O., Iwai, D., Wetzstein, G., and Grundhöfer, A. (2008). The Visual Computing of Projector-Camera Systems. *Computer Graphics Forum*, 27(8):2219–2245.
- Bimber, O. and Raskar, R. (2005). *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters, Ltd.
- Bimber, O., Wetzstein, G., Emmerling, A., and Nitschke, C. (2005b). Enabling View-Dependent Stereoscopic Projection in Real Environments. In *Proc. IEEE/ACM ISMAR*, pages 14–23.
- Bimber, O., Zeidler, T., Grundhoefer, A., Wetzstein, G., Moehring, M., Knoedel, S., and Hahne, U. (2005c). Interacting with Augmented Holograms. In *Proc. SPIE Conference on Practical Holography XIX: Materials and Applications*, pages 1–8.
- Blanche, P.-A. et al. (2010). Holographic 3-d telepresence using large-area photorefractive polymer. *Nature*, 468:80–83.
- Blundell, B. and Schwartz, A. (1999). *Volumetric Three-Dimensional Display Systems*. Wiley-IEEE Press.
- Bolles, R. C., Baker, H. H., and Marimont, D. H. (1987). Epipolar-plane image analysis: An approach to determining structure from motion. *IJCV*, 1(1):7–55.
- Bonfort, T., Sturm, P., and Gargallo, P. (2006). General Specular Surface Triangulation. In *Proc. ACCV*, pages 872–881.

- Borman, S. and Stevenson, R. (1998). Super-resolution from image sequences - A review. In *Proc. Symposium on Circuits and Systems*, pages 374–378.
- Born, M. and Wolf, E. (1999). *Principles of Optics*. Cambridge University Press, 7 edition.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Bracewell, R. N. and Riddle, A. C. (1967). Inversion of fan-beam scans in radio astronomy. *Astrophysical Journal*, 150:427–434.
- Bradley, D., Atcheson, B., Ihrke, I., and Heidrich, W. (2009). Synchronization and Rolling Shutter Compensation for Consumer Video Camera Arrays. In *Proc. ProCams*, pages 1–8.
- Braun, M. (1992). *Picturing Time: The Work of Etienne-Jules Marey (1830-1904)*. The University of Chicago Press.
- Brettel, H., Vinot, F., and Mollon, J. D. (1997). Computerized simulation of color appearance for dichromats. *JOSA A*, 14(10):2647–2655.
- Brown, M. S., Song, P., and Cham, T.-J. (2006). Image Pre-Conditioning for Out-of-Focus Projector Blur. In *Proc. IEEE CVPR*, volume II, pages 1956–1963.
- Bub, G., Tecza, M., Helmes, M., Lee, P., and Kohl, P. (2010). Temporal Pixel Multiplexing for Simultaneous High-Speed, High-Resolution Imaging. *Nature Methods*, 7:209–211.
- Cakmakci, O., Ha, Y., and Rolland, J. P. (2004). A Compact Optical See-Through Head-Worn Display with Occlusion Support. In *Proc. ISMAR*, pages 16–25.
- Cakmakci, O. and Rolland, J. (2006). Head-Worn Displays: A Review. *IEEE Journal of Display Technology*, 2(3):199–216.
- Carranza, J., Theobalt, C., Magnor, M. A., and Seidel, H.-P. (2003). Free-viewpoint video of human actors. *ACM Transactions on Graphics (TOG)*, 22(3):569–577.
- Chai, J.-X., Tong, X., Chan, S.-C., and Shum, H.-Y. (2000). Plenoptic sampling. In *ACM SIGGRAPH*, pages 307–318.
- Chaudhury, K. N., Muñoz-Barrutia, A., and Unser, M. (2010). Fast space-variant elliptical filtering using box splines. *IEEE Trans. Image*, 19(9):2290–2306.
- Chen, C.-H., Lin, F.-C., Hsu, Y.-T., Huang, Y.-P., and Shieh, H.-P. D. (2009). A field sequential color LCD based on color fields arrangement for color breakup and flicker reduction. *Display Technology*, 5(1):34–39.

- Chen, S. E. and Williams, L. (1993). View interpolation for image synthesis. In *Proc. ACM SIGGRAPH*, pages 279–288.
- Chi, W., Chu, K., and George, N. (2006). Polarization Coded Aperture. *Optics Express*, 14(15):6634–6642.
- Chi, W. and George, N. (2001). Electronic Imaging using a Logarithmic Asphere. *Optics Letters*, 26(12):875–877.
- Coleman, T. and Li, Y. (1996). A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6(4):1040–1058.
- Collett, E. (2005). *Field Guide to Polarization*. SPIE Press.
- Compton, J. (2007). Color filter array 2.0. <http://johncompton.pluggedin.kodak.com>.
- Cossairt, O., Miao, D., and Nayar, S. K. (2011). Gigapixel Computational Imaging. In *Proc. ICCP*.
- Cossairt, O. and Nayar, S. K. (2010). Spectral Focal Sweep: Extended Depth of Field from Chromatic Aberrations. In *Proc. ICCP*, pages 1–8.
- Cossairt, O., Zhou, C., and Nayar, S. K. (2010). Diffusion Coded Photography for Extended Depth of Field. *ACM Trans. Graph. (Siggraph)*, 29(3):31.
- Cossairt, O. S., Napoli, J., Hill, S. L., Dorval, R. K., and Favallora, G. E. (2007). Occlusion-capable multiview volumetric three-dimensional display. *Applied Optics*, 46(8):1244–1250.
- Crawford, G. (2005). *Flexible Flat Panel Displays*. John Wiley and Sons.
- CRI-INC (2009). VariSpec Liquid Crystal Tunable Filters. www.cri-inc.com/varispec.
- Cula, O. G., Dana, K. J., Pai, D. K., and Wang, D. (2007). Polarization Multiplexing and Demultiplexing for Appearance-Based Modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):362–367.
- Daly, S. (1993). The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity. In Watson, A., editor, *Digital Image and Human Vision*, pages 179–206. Cambridge, MA: MIT Press.
- Dalziel, S., Hughes, G., and Sutherland, B. (2000). Whole-Field Density Measurements by Synthetic Schlieren. *Experiments in Fluids*, 28(4):322–335.

- Damberg, G., Seetzen, H., Ward, G., Heidrich, W., and Whitehead, L. (2007). High Dynamic Range Projection Systems. In *Society of Information Displays Symposium Digest*, pages 4–7.
- Davis, J. A., McNamara, D. E., Cottrell, D. M., and Sonehara, T. (2000). Two-dimensional polarization encoding with a phase-only liquid-crystal spatial light modulator. *Applied Optics*, 39(10):1549–1554.
- Debevec, P. (2002). Image-Based Lighting. *IEEE Computer Graphics and Applications*, pages 26–34.
- Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M. (2000). Acquiring the Reflectance Field of a Human Face. In *Proc. ACM SIGGRAPH*, pages 145–156.
- Debevec, P. E. and Malik, J. (1997). Recovering High Dynamic Range Radiance Maps from Photographs. In *Proc. ACM SIGGRAPH*, pages 369–378.
- DICOM (2008). Part 14: Grayscale standard display function. In *Digital Imaging and Communications in Medicine*, pages 1–55. National Electrical Manufacturers Association.
- Dodgson, N. A. (2009). Analysis of the viewing zone of multi-view autostereoscopic displays. In *SPIE Stereoscopic Displays and Applications XIII*, pages 254–265.
- Dolby (2010). Dolby PRM-4200. http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/ProMonitor_OverviewSpecsheet_24374_1010_final.pdf.
- Dowski, Jr, E. and Cathey, W. (1995). Extended Depth of Field Through Wave-Front Coding. *Applied Optics*, 34(11):1859–1866.
- Drebin, R. A., Carpenter, L., and Hanrahan, P. (1988). Volume rendering. *ACM SIGGRAPH*, 22:65–74.
- Durand, F. and Dorsey, J. (2000). Interactive Tone Mapping. In *Proc. Eurographics Workshop on Rendering*, pages 219–230.
- Elsinga, G., Oudheusen, B., Scarano, F., and Watt, D. (2004). Assessment and Application of Quantitative Schlieren Methods: Calibrated Color Schlieren and Background Oriented Schlieren. *Experiments in Fluids*, 36(2):309–325.
- Favalora, G. E. (2005). Volumetric 3D displays and application infrastructure. *IEEE Computer*, 38:37–44.

- Feiner, S., Macintyre, B., and Seligmann, D. (1993). Knowledge-Based Augmented Reality. *Comm. of the ACM*, 36(7):53–62.
- Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., and Freeman, W. T. (2006). Removing Camera Shake from a Single Photograph. *ACM Trans. Graph.*, 25:787–794.
- Foveon (2010). X3 Technology. <http://www.foveon.com>.
- Gabor, D. (1948). A new microscopic principle. *Nature*, pages 777–778.
- Garcia-Guerrero, E. E., Mendez, E. R., and Leskova, H. M. (2007). Design and Fabrication of Random Phase Diffusers for Extending the Depth of Focus. *Optics Express*, 15(3):910–923.
- Georgiev, T., Intwala, C., Babacan, S., and Lumsdaine, A. (2008). Unified frequency domain analysis of lightfield cameras. In *Proc. ECCV*, pages 224–237.
- Georgiev, T., Lumsdaine, A., and Goma, S. (2009). High Dynamic Range Image Capture with Plenoptic 2.0 Camera. In *Proc. of Signal Recovery and Synthesis*, pages 1–3.
- Georgiev, T., Zheng, C., Nayar, S., Curless, B., Salesin, D., and Intwala, C. (2006). Spatio-angular resolution trade-offs in integral photography. In *Proc. EGSR*, pages 263–272.
- Gershun, A. (1936). The light field. *Journal of Mathematics and Physics*, XVIII:51–151. Translated by P. Moon and G. Timoshenko.
- Ghosh, A., Chen, T., Peers, P., Wilson, C. A., and Debevec, P. (2010). Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph. (Siggraph Asia)*, 27(5):134.
- Goesele, M., Lensch, H. P. A., Lang, J., Fuchs, C., and Seidel, H.-P. (2004). DISCO: Acquisition of Translucent Objects. *ACM Trans. Graph. (Siggraph)*, 23:835–844.
- Goldstein, E. B. (2006). *Sensation and Perception*. Wadsworth Publishing.
- Gortler, S., Grzeszczuk, R., Szelinski, R., and Cohen, M. (1996). The Lumigraph. *ACM Trans. Graph. (SIGGRAPH)*, pages 43–54.
- Gotoda, H. (2010). A multilayer liquid crystal display for autostereoscopic 3D viewing. In *SPIE-IS&T Stereoscopic Displays and Applications XXI*, volume 7524, pages 1–8.
- Gotoda, H. (2011). Reduction of image blurring in an autostereoscopic multilayer liquid crystal display. In *SPIE-IS&T Stereoscopic Displays and Applications XXII*, volume 7863, pages 1–7.

- Gottesman, S. R. and Fenimore, E. E. (1989). New Family of Binary Arrays for Coded Aperture Imaging. *Applied Optics*, 28(20):4344–4352.
- Green, P., Sun, W., Matusik, W., and Durand, F. (2007). Multi-Aperture Photography. *ACM Trans. Graph. (Siggraph)*, page 68.
- Greivenkamp, J. (1990). Color Dependant Optical Prefilter for the Suppression of Aliasing Artifacts. *Applied Optics*, 29(5):676–684.
- Grossberg, M., Peri, H., Nayar, S., and Belhumeur, P. (2004). Making one object look like another: Controlling appearance using a projector-camera system. In *Proc. IEEE CVPR*, volume I, pages 452–459.
- Grossberg, M. D. and Nayar, S. K. (2003). High Dynamic Range from Multiple Images: Which Exposures to Combine. In *Proc. ICCV Workshop CPMCV*.
- Grosse, M. and Bimber, O. (2008). Coded Aperture Projection. In *Proc. EDT*, pages 1–4.
- Grosse, M., Wetzstein, G., Grundhöfer, A., and Bimber, O. (2009). Adaptive Coded Aperture Projection. *Submitted to ACM Trans. Graph.*, pages 1–18.
- Grundhöfer, A. and Bimber, O. (2008). Real-Time Adaptive Radiometric Compensation. *IEEE TVCG*, 14(1):97–108.
- Gu, J., Hitomi, Y., Mitsunaga, T., and Nayar, S. K. (2010). Coded Rolling Shutter Photography: Flexible Space-Time Sampling. In *Proc. IEEE ICCP*, pages 1–8.
- Gunturk, B., Glotzbach, J., Altunbasak, Y., Schafer, R., and Mersereau, R. (2005). Demosaicking: Color Filter Array Interpolation in Single-Chip Digital Cameras. *IEEE Signal Processing*, 22(1):44–54.
- Gupta, M., Agrawal, A., Veeraraghavan, A., and Narasimhan, S. G. (2010). Flexible Voxels for Motion-Aware Videography. In *Proc. ECCV*, pages 100–114.
- Gustavson, T. and House, G. E. (2009). *Camera: A History of Photography from Daguerreotype to Digital*. Sterling Innovation.
- Halle, M. (2005). Autostereoscopic Displays and Computer Graphics. In *ACM Siggraph Courses, Session: Spatial Displays and Computer Graphics*, pages 1–104.
- Halle, M. W. (1994). Holographic stereograms as discrete imaging systems. In *SPIE Practical Holography*, pages 73–84.

- Hargather, M. J. and Settles, G. S. (2009). Natural-Background-Oriented Schlieren Imaging. *Experiments in Fluids*, 48(1):59–68.
- Hart, W. M. (1987). The temporal responsiveness of vision. In Moses, R. A. and Hart, W. M., editors, *Adler’s Physiology of the Eye*. C.V. Moseby Company.
- Hasinoff, S. W., Durand, F., and Freeman, W. T. (2010). Noise-Optimal Capture for High Dynamic Range Photography. In *Proc. IEEE CVPR*, pages 1–8.
- Hasinoff, S. W. and Kutulakos, K. N. (2006). Confocal Stereo. In *Proc. ECCV*, pages 620–634.
- Hasinoff, S. W. and Kutulakos, K. N. (2008). Light-Efficient Photography. In *Proc. of ECCV*, pages 45–59.
- Hasinoff, S. W. and Kutulakos, K. N. (2009). Confocal Stereo. *Int. J. Comp. Vis.*, 81(1):82–104.
- Hasinoff, S. W., Kutulakos, K. N., Durand, F., and Freeman, W. T. (2009). Time-Constrained Photography. In *Proc. of ICCV*, pages 333–340.
- Häusler, G. (1972). A Method to Increase the Depth of Focus by Two Step Image Processing. *Optics Communications*, 6(1):38–42.
- Healey, C. and Enns, J. (1999). Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE TVCG*, 5(2):145–167.
- Hecht, E. (2001). *Optics*. Addison Wesley.
- Herman, G. T. (1995). Image reconstruction from projections. *Real-Time Imaging*, 1(1):3–18.
- Hernandez, C., Vogiatzis, G., Brostow, G., Stenger, B., and Cipolla, R. (2007). Non-Rigid Photometric Stereo with Colored Lights. In *Proc. ICCV*.
- Hirakawa, K. and Parks, T. W. (2006). Joint Demosaicing and Denoising. *IEEE Trans. Im. Proc.*, 15(8):2146–2157.
- Hirakawa, K. and Wolfe, P. (2008). Spatio-Spectral Color Filter Array Design for Optimal Image Recovery. *IEEE Trans. Im. Proc.*, 17(10):1876 –1890.
- Hirsch, M., Lanman, D., Holtzman, H., and Raskar, R. (2009). BiDi screen: A thin, depth-sensing LCD for 3D interaction using lights fields. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 28(5).

- Hiura, S., Mohan, A., and Raskar, R. (2009). Krill-eye : Superposition Compound Eye for Wide-Angle Imaging via GRIN Lenses. In *Proc. OMNIVIS*, pages 1–8.
- Horovitz, I. and Kiryati, N. (2004). Depth from Gradient Fields and Control Points: Bias Correction in Photometric Stereo. *Image and Vision Computing*, 22:681–694.
- Hoskinson, R., Stoeber, B., Heidrich, W., and Fels, S. (2010). Light Reallocation for High Contrast Projection Using an Analog Micromirror Array. *ACM Trans. Graph. (Siggraph Asia)*, 29:165.
- Howes, W. L. (1984). Rainbow Schlieren and its Applications. *Applied Optics*, 23(4):2449–2460.
- Hradaynath, R. (2002). *An Introduction to Night Vision Technology*. Defence Research & Development Organization (India).
- Hullin, M. B., Fuchs, M., Ihrke, I., Seidel, H.-P., and Lensch, H. P. A. (2008). Fluorescent Immersion Range Scanning. In *ACM Trans. Graph. (Siggraph)*, pages 87:1–87:10.
- Huynh, C. P., Robles-Kelly, A., and Hancock, E. (2010). Shape and Refractive Index Recovery from Single-View Polarisation Images. In *Proc. CVPR*, pages 1229–1236.
- Ihrke, I., Goldluecke, B., and Magnor, M. (2005). Reconstructing the Geometry of Flowing Water. In *Proc. ICCV*, pages 1055–1060.
- Ihrke, I., Kutulakos, K. N., Lensch, H. P. A., Magnor, M., and Heidrich, W. (2010a). Transparent and Specular Object Reconstruction. *Computer Graphics Forum*, 29(8):2400–2426.
- Ihrke, I., Stich, T., Gottschlich, H., Magnor, M., and Seidel, H. (2008). Fast incident light field acquisition and rendering. In *Proc. of WSCG*, pages 177–184.
- Ihrke, I., Wetzstein, G., and Heidrich, W. (2010b). A Theory of Plenoptic Multiplexing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. oral.
- Ihrke, I., Ziegler, G., Tevs, A., Theobalt, C., Magnor, M., and Seidel, H.-P. (2007). Eikonal Rendering: Efficient Light Transport in Refractive Objects. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 26(3):59.
- Isono, H., Yasuda, M., and Sasazawa, H. (1993). Autostereoscopic 3-D Display using LCD-Generated Parallax Barrier. *Electronics and Communication in Japan*, 76(7):77–84.
- ITU (1990). ITU-R BT.709, basic parameter values for the HDTV standard for the studio and for international programme exchange. Standard Recommendation 709, International Telecommunication Union.

- ITU (1993). Iso/iec 10918-1: 1993(e) ccit recommendation t.81.
- Ives, H. (1903). Parallax Stereogram and Process of Making Same. US patent 725,567.
- Ives, H. (1928). Camera for Making Parallax Panoramagrams. *J. Opt. Soc. Amer.*, 17:435–439.
- Jacobs, A., Mather, J., Winlow, R., Montgomery, D., Jones, G., Willis, M., Tillin, M., Hill, L., Khazova, M., Stevenson, H., and Bourhill, G. (2003). 2D/3D Switchable Displays. Sharp Technical Journal.
- Jones, A., McDowall, I., Yamada, H., Bolas, M., and Debevec, P. (2007). Rendering for an interactive 360° light field display. *ACM Trans. Graph.*, 26:40:1–40:10.
- Kaczmarz, S. (1937). Angenäherte auflösung von systemen linearer gleichungen. *Bull. Acad. Pol. Sci. Lett. A*, 35:335–357.
- Kak, A. C. and Slaney, M. (2001). *Principles of Computerized Tomographic Imaging*. Society for Industrial Mathematics.
- Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2003). High Dynamic Range Video. In *Proc. ACM SIGGRAPH*, pages 319–325.
- Kanolt, C. W. (1918a). Parallax Panoramagrams. US patent 1,260,682.
- Kanolt, C. W. (1918b). Photographic method and apparatus. U.S. Patent 1,260,682.
- Keck, B., Hofmann, H., Scherl, H., Kowarschik, M., and Hornegger, J. (2009). GPU-accelerated SART reconstruction using the CUDA programming environment. In *SPIE*, volume 7258.
- Kim, S.-J., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. (2007). An Interior-Point Method for Large-Scale ℓ_1 -Regularized Least Squares. *IEEE Journal on Selected Topics in Signal Processing*, 1(4):606–617.
- Kingdom, F. and Moulden, B. (1988). Border effects on brightness: a review of findings, models and issues. *Spatial Vision*, 3(4):255–262.
- Kiyokawa, K., Kurata, Y., and Ohno, H. (2001). ELMO: An Enhanced Optical See-Through Display Using an LCD Panel for Mutual Occlusion. In *Proc. ISMAR*, pages 186–187.
- Klug, M., Holzbach, M., and Ferdman, A. (2001). Method and apparatus for recording one-step, full-color, full-parallax, holographic stereograms. U.S. Patent 6,330,088.

- Kopf, J., Uyttendaele, M., Deussen, O., and Cohen, M. F. (2007). Capturing and Viewing Gigapixel Images. *ACM Trans. on Graph. (SIGGRAPH)*, 26(3).
- Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2005). Perceptual Effects in Real-Time Tone Mapping. In *Spring Conference on Computer Graphics 2005*. ACM.
- Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2007). Contrast restoration by adaptive countershading. *Computer Graphics Forum (Proc. Eurographics)*, 26:581–590.
- Kusakabe, Y., Kanazawa, M., Nojiri, Y., Furuya, M., and Yoshimura, M. (2008). A YC-separation-type projector: High dynamic range with double modulation. *Journal of The Society for Information Display*, 16(2):1–9.
- Kutulakos, K. N. and Hasinoff, S. W. (2009). Focal Stack Photography: High-Performance Photography with a Conventional Camera. In *Proc. IAPR Conference on Machine Vision Applications*, pages 332–337.
- Kutulakos, K. N. and Steger, E. (2005). A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation. In *Proc. ICCV*, pages 1448–1455.
- Land, M. and Nilsson, D. (2002). *Animal Eyes*. Oxford University Press.
- Landolt, O., Mitros, A., and Koch, C. (2001). Visual Sensor with Resolution Enhancement by Mechanical Vibrations. In *Proc. Advanced Research in VLSI*, pages 249–264.
- Lanman, D. (2010). *Mask-based Light Field Capture and Display*. PhD thesis, Brown University, School of Engineering.
- Lanman, D. and Hirsch, M. (2010). Build your own 3D Display. *ACM Siggraph Course Notes*, pages 1–171.
- Lanman, D., Hirsch, M., Kim, Y., and Raskar, R. (2010). Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Trans. Graph.*, 29:163:1–163:10.
- Lanman, D., Raskar, R., Agrawal, A., and Taubin, G. (2008). Shield Fields: Modeling and Capturing 3D Occluders. *ACM Trans. Graph. (SIGGRAPH Asia)*, 27(5):131:1–131:10.
- Lanman, D., Wachs, M., Taubin, G., and Cukierman, F. (2006). Spherical Catadioptric Arrays: Construction, Multi-View Geometry, and Calibration. In *Proc. 3DPVT*, pages 81–88.

- Lanman, D., Wetzstein, G., Hirsch, M., Heidrich, W., and Raskar, R. (2011). Polarization Fields: Dynamic Light Field Display using Multi-Layer LCDs. *ACM Trans. Graph. (Siggraph Asia)*, 3:1–9.
- Levin, A. and Durand, F. (2010). Linear View Synthesis Using a Dimensionality Gap Light Field Prior. In *Proc. IEEE CVPR*, pages 1–8.
- Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007a). Deconvolution using Natural Image Priors. groups.csail.mit.edu/graphics/CodedAperture/SparseDeconv-LevinEtAl07.pdf.
- Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007b). Image and Depth from a Conventional Camera with a Coded Aperture. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):70.
- Levin, A., Hasinoff, S. W., Green, P., Durand, F., and Freeman, W. T. (2009). 4D Frequency Analysis of Computational Cameras for Depth of Field Extension. *ACM Trans. Graph. (SIGGRAPH)*, 28(3):1–14.
- Levin, A., Sand, P., Cho, T. S., Durand, F., and Freeman, W. T. (2008). Motion-invariant photography. *ACM Trans. Graph. (Siggraph)*, 27(3):71.
- Levoy, M. (2010). Computational Photography and the Stanford Frankencamera. Technical Talk. <http://www.graphics.stanford.edu/talks/>.
- Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., and Bolas, M. (2004). Synthetic Aperture Confocal Imaging. *ACM Trans. Graph. (Siggraph)*, 23(3):825–834.
- Levoy, M. and Hanrahan, P. (1996). Light Field Rendering. *ACM Trans. Graph. (SIGGRAPH)*, pages 31–42.
- Levoy, M., Ng, R., Adams, A., Footer, M., and Horowitz, M. (2006). Light field microscopy. *ACM Trans. Graph. (SIGGRAPH)*, 25(3):924–934.
- Levoy, M., Zhang, Z., and McDowall, I. (2009). Recording and Controlling the 4D Light Field in a Microscope. *To appear in Journal of Microscopy*.
- Li, F., Yu, J., and Chai, J. (2008a). A Hybrid Camera for Motion Deblurring and Depth Map Super-Resolution. In *Proc. IEEE CVPR*, pages 1–8.
- Li, X., Gunturk, B., and Zhang, L. (2008b). Image Demosaicing: a Systematic Survey. In *SPIE Conf. on Visual Comm. and Image Proc.*, pages 68221J–68221J–15.

- Liang, C.-K., Lin, T.-H., Wong, B.-Y., Liu, C., and Chen, H. H. (2008). Programmable Aperture Photography: Multiplexed Light Field Acquisition. *ACM Trans. Graph. (SIGGRAPH)*, 27(3):1–10.
- Lingley, A. and Parviz, B. (2008). Multipurpose Integrated Active Contact Lenses. *The Neuromorphic Engineer*.
- Lippmann, G. (1908). La Photographie Intégrale. *Academie des Sciences*, 146:446–451.
- Lipton, L. (1982). *Foundations of the Stereoscopic Cinema*. Van Nostrand Reinhold.
- Liu, C. and Sun, D. (2011). A Bayesian Approach to Adaptive Video Super Resolution. In *Proc. IEEE CVPR*, pages 1–8.
- Loukianitsa, A. and Putilin, A. N. (2002). Stereodisplay with neural network image processing. In *SPIE-IT&T Stereoscopic Displays and Virtual Reality Systems IX*, volume 4660, pages 207–211.
- Lu, Y. M. and Vetterli, M. (2009). Optimal Color Filter Array Design: Quantitative Conditions and an Efficient Search Procedure. In *Proc. SPIE 7250*, pages 1–8.
- Lucy, L. B. (1974). An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745–754.
- Lumsdaine, A. and Georgiev, T. (2009). The Focused Plenoptic Camera. In *Proc. ICCP*, pages 1–8.
- Ma, B., Yao, B., Ye, T., and Lei, M. (2010). Prediction of optical modulation properties of twisted-nematic liquid-crystal display by improved measurement of Jones matrix. *Appl. Physics*, 107.
- Ma, W.-C., Hawkins, T., Peers, P., Chabert, C.-F., Weiss, M., and Debevec, P. (2007). Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proc. EGSR*, pages 183–194.
- Maeda, H., Hirose, K., Yamashita, J., Hirota, K., and Hirose, M. (2003). All-around display for video avatar in real world. In *IEEE/ACM ISMAR*, pages 288–289.
- Mann, S. (1997). Wearable Computing: A First Step Toward Personal Imaging. *IEEE Computer*, 30(2):25–32.
- Mann, S., Fung, J., Aimone, C., Sehgal, A., and Chen, D. (2005). Designing EyeTap Digital Eyeglasses for Continuous Lifelong Capture and Sharing of Personal Experiences. In *CHI extended abstracts*.

- Mann, S. and Picard, R. W. (1995). Being 'Undigital' with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures. In *Proc. IS&T*, pages 442–448.
- Mantiuk, R., Daly, S., Myszkowski, K., and Seidel, H.-P. (2005). Predicting Visible Differences in High Dynamic Range Images - Model and its Calibration. In Rogowitz, B. E., Pappas, T. N., and Daly, S. J., editors, *Electronic Imaging*, volume 5666, pages 204–214.
- Mantiuk, R., Kim, K. J., Rempel, A., and Heidrich, W. (2011). HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph. (Siggraph)*, 30(3):1–12.
- Marshall, J. and Oberwinkler, J. (1999). Ultraviolet Vision: the Colourful World of the Mantis Shrimp. *Nature*, 401(6756):873–874.
- Masia, B., Agustin, S., Fleming, R., Sorkine, O., and Gutierrez, D. (2009). Evaluation of Reverse Tone Mapping through Varying Exposure Conditions. *ACM Trans. Graph. (Siggraph Asia)*, 28(5):1–8.
- Mäthger, L. M., Shashar, N., and Hanlon, R. T. (2009). Do Cephalopods Communicate using Polarized Light Reflections from their Skin? *Journal of Experimental Biology*, 212:2133–2140.
- Matusik, W., Buehler, C., Raskar, R., Gortler, S. J., and McMillan, L. (2000). Image-based visual hulls. In *ACM SIGGRAPH*, pages 369–374.
- Matusik, W. and Pfister, H. (2004). 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Transactions on Graphics*, 23:814–824.
- Maxwell, J. C. (1860). On the Theory of Compound Colours, and the Relations of the Colours of the Spectrum. *Phil. Trans. R. Soc. Lond.*, 150:57–84.
- McGuire, M., Matusik, W., Pfister, H., Chen, B., Hughes, J. F., and Nayar, S. K. (2007). Optical Splitting Trees for High-Precision Monocular Imaging. *IEEE Comput. Graph. & Appl.*, 27(2):32–42.
- McGuire, M., Matusik, W., Pfister, H., Hughes, J. F., and Durand, F. (2005). Defocus Video Matting. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):567–576.
- Meier, G. (2002). Computerized Background-Oriented Schlieren. *Experiments in Fluids*, 33(1):181–187.

- Merzkirch, W. (1987). *Flow Visualization*. Academic Press.
- Michika, D. and Brown, R. J. (1989). Single-aperture multi-spectral reticle projector. US Patent 4881796.
- Mitra, N. J. and Pauly, M. (2009). Shadow art. *ACM Trans. Graph.*, 28:156:1–156:7.
- Mitsunaga, T. and Nayar, S. K. (1999). Radiometric Self Calibration. In *IEEE CVPR*, pages 374–380.
- Miyazaki, D. and Ikeuchi, K. (2005). Inverse Polarization Raytracing: Estimating Surface Shapes of Transparent Objects. In *Proc. CVPR*, pages 910–917.
- Miyazaki, D., Kagesawa, M., and Ikeuchi, K. (2004). Transparent Surface Modelling from a Pair of Polarization Images. *IEEE Trans. PAMI*, 26(1):73–82.
- Miyazaki, D., Tan, R. T., Hara, K., and Ikeuchi, K. (2003). Polarization-based inverse rendering from a single view. In *Proc. ICCV*, pages 982–998.
- Mohan, A., Raskar, R., and Tumblin, J. (2008). Agile Spectrum Imaging: Programmable Wavelength Modulation for Cameras and Projectors. *Computer Graphics Forum (Eurographics)*, 27(2):709–717.
- Moon, P. and Spencer, D. E. (1981). *The Photic Field*. MIT Press.
- Moreno, I., Martínez, J. L., and Davis, J. A. (2007). Two-dimensional polarization rotator using a twisted-nematic liquid-crystal display. *Appl. Optics*, 46(6):881–887.
- Moreno, I., Velásquez, P., Fernández-Pousa, C. R., Sánchez-López, M. M., and Mateos, F. (2003). Jones matrix method for predicting and optimizing the optical modulation properties of a liquid-crystal display. *Appl. Physics*, 94.
- Morris, N. J. W. and Kutulakos, K. N. (2005). Dynamic Refraction Stereo. In *Proc. ICCV*, pages 1573–1580.
- Morris, N. J. W. and Kutulakos, K. N. (2007). Reconstructing the Surface of Inhomogeneous Transparent Scenes by Scatter Trace Photography. In *Proc. ICCV*, pages 1–8.
- Müller, V. (1996). Elimination of Specular Surface-Reflectance Using Polarized and Unpolarized Light. In *Proc. IEEE ECCV*, pages 625–635.
- Murase, H. (1990). Surface Shape Reconstruction of an Undulating Transparent Object. In *Proc. ICCV*, pages 313–317.

- Murphy, D. B. (2001). *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley-Liss.
- Muybridge, E. (1957). *Animals in Motion*. first ed. Dover Publications, Chapman and Hall 1899.
- Nagahara, H., Kuthirummal, S., Zhou, C., and Nayar, S. (2008). Flexible Depth of Field Photography. In *European Conference on Computer Vision (ECCV)*.
- Nagy, A. and Sanchez, R. (1990). Critical color differences determined with a visual search task. *JOSA A*, 7(7):1209–1217.
- Namer, E. and Schechner, Y. Y. (2005). Advanced Visibility Improvement Based on Polarization Filtered Images. In *Proc. SPIE 5888*, pages 36–45.
- Narasimhan, S. and Nayar, S. (2005). Enhancing Resolution along Multiple Imaging Dimensions using Assorted Pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530.
- Narasimhan, S. G., Koppal, S. J., and Yamazaki, S. (2008). Temporal Dithering of Illumination for Fast Active Vision. In *Proc. ECCV*, pages 830–844.
- Nayar, S. and Anand, V. (2007). 3D display using passive optical scatterers. *IEEE Computer Magazine*, 40(7):54–63.
- Nayar, S. and Branzoi, V. (2003). Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time. In *Proc. IEEE ICCV*, volume 2, pages 1168–1175.
- Nayar, S., Branzoi, V., and Boulton, T. (2004). Programmable Imaging using a Digital Micromirror Array. In *Proc. IEEE CVPR*, volume I, pages 436–443.
- Nayar, S., Fang, X.-S., and Boulton, T. (1993). Removal of Specularities using Color and Polarization. In *Proc. IEEE CVPR*, pages 583–590.
- Nayar, S. and Mitsunaga, T. (2000). High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proc IEEE CVPR*, volume 1, pages 472–479.
- Nayar, S. and Narasimhan, S. (2002). Assorted Pixels: Multi-Sampled Imaging With Structural Models. In *European Conference on Computer Vision (ECCV)*, volume IV, pages 636–652.
- Nayar, S. K., Branzoi, V., and Boulton, T. E. (2006). Programmable Imaging: Towards a Flexible Camera. *International Journal on Computer Vision*, 70(1):7–22.

- Nayar, S. K. and Nakagawa, Y. (1994). Shape from Focus. *IEEE Trans. PAMI*, 16(8):824–831.
- Ng, H.-S., Wu, T.-P., and Tang, C.-K. (2010). Surface-from-Gradients without Discrete Integrability Enforcement: A Gaussian Kernel Approach. *IEEE Trans. PAMI*, 32:2085–2099.
- Ng, R. (2005). Fourier Slice Photography. *ACM Trans. Graph. (Siggraph)*, 24(3):735–744.
- Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. (2005). Light field photography with a hand-held plenoptic camera. Technical Report Computer Science CSTR 2005-02, Stanford University.
- Nomura, Y., Zhang, L., and Nayar, S. (2007). Scene Collages and Flexible Camera Arrays. In *Proc. EGSR*, pages 1–12.
- NVIDIA Tegra (2008). www.nvidia.com/page/handheld.
- Ogata, S., Ishida, J., and Sasano, T. (1994). Optical Sensor Array in an Artificial Compound Eye. *Optical Engineering*, 33(11):3649–3655.
- Ojeda-Castaneda, J., Landgrave, J. E. A., and Escamilla, H. M. (2005). Annular Phase-Only Mask for High Focal Depth. *Optics Letters*, 30(13):1647–1649.
- Okano, F., Arai, J., Hoshino, H., and Yuyama, I. (1999). Three-Dimensional Video System Based on Integral Photography. *Optical Engineering*, 38(6):1072–1077.
- Olofsson, T. (2005). Deconvolution and Model-Based Restoration of Clipped Ultrasonic Signals. *IEEE Trans. Instrumentation and Measurements*, 54(3):1235–1240.
- Optec (2011). Separation prism technical data. http://www.alt-vision.com/color_prisms_tech_data.htm.
- Oyamada, Y. and Saito, H. (2007). Focal Pre-Correction of Projected Image for Deblurring Screen Image. In *Proc. IEEE ProCams*.
- Panavision (2010). Genesis. <http://www.panavision.com>.
- Pavlovych, A. and Stuerzlinger, W. (2005). A High-Dynamic Range Projection System. *Proc. SPIE annual meeting*, 5969.
- Perlin, K. and Han, J. Y. (2006). Volumetric display with dust as the participating medium. U.S. Patent 6,997,558.

- Photron (2010). FASTCAM SA5. http://www.photron.com/datasheet/FASTCAM_SA5.pdf.
- Pieper, R. J. and Korpel, A. (1983). Image Processing for Extended Depth of Field. *Applied Optics*, 22(10):1449–1453.
- Pixim (2010). Digital Pixel System. <http://www.pixim.com>.
- Project, E. D. C. (2009). Harold ‘Doc’ Edgerton. <http://www.edgerton-digital-collections.org/techniques/high-speed-photography>.
- Prokudin-Gorskii, S. M. (1912). The Prokudin-Gorskii Photographic Records Recreated. <http://www.loc.gov/exhibits/empire/>.
- Putilin, A. N. and Loukianitsa, A. (2006). Visualization of three dimensional images and multi aspect imaging. U.S. Patent 6,985,290.
- Ramanath, R., Snyder, W., Bilbro, G., and Sander, W. (2002). Demosaicking Methods for Bayer Color Arrays. *Journal of Electronic Imaging*, 11(3):306–315.
- Ramasubramanian, M., Pattanaik, S. N., and Greenberg, D. P. (1999). A Perceptually Based Physical Error Metric for Realistic Image Synthesis. *ACM Trans. Graph.*, 25(3):73–82.
- Rasche, K., Geist, R., and Westall, J. (2005). Re-coloring Images for Gamuts of Lower Dimension. *Computer Graphics Forum (Proc. Eurographics)*, 24(3):423–432.
- Raskar, R., Agrawal, A., and Tumblin, J. (2006). Coded Exposure Photography: Motion Deblurring using Fluttered Shutter. *ACM Trans. Graph. (SIGGRAPH)*, 25(3):795–804.
- Raskar, R., Agrawal, A., Wilson, C. A., and Veeraraghavan, A. (2008). Glare Aware Photography: 4D Ray Sampling for Reducing Glare Effects of Camera Lenses. *ACM Trans. Graph. (SIGGRAPH)*, 27(3):1–10.
- Ray, S. F. (2002). *High Speed Photography and Photonics*. SPIE Press.
- Reddy, D., Veeraraghavan, A., and Chellappa, R. (2011). P2C2: Programmable Pixel Compressive Camera for High Speed Imaging. In *Proc. IEEE CVPR*, pages 1–8.
- Reinhard, E., Ward, G., Debevec, P., Pattanaik, S., Heidrich, W., and Myszkowski, K. (2010). *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann.
- Rempel, A. G., Trentacoste, M., Seetzen, H., Young, H. D., Heidrich, W., Whitehead, L., and Ward, G. (2007). Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. In *ACM Trans. Graph. (Proc. SIGGRAPH)*, page 39.

- Research, V. (2010). Phantom Flex. <http://www.visionresearch.com/Products/High-Speed-Cameras/Phantom-Flex/>.
- Ri, S., Fujigaki, M., Matui, T., and Morimoto, Y. (2006). Accurate pixel-to-pixel correspondence adjustment in a digital micromirror device camera by using the phase-shifting moiré method. *Applied optics*, 45(27):6940–6946.
- Richard, H. and Raffel, M. (2001). Principle and Applications of the Background Oriented Schlieren (BOS) Method. *Measurement Science and Technology*, 12(9):1576–1585.
- Richardson, H. W. (1972). Bayesian-Based Iterative Method of Image Restoration. *JOSA*, 62(1):55–59.
- Robertson, M. A., Borman, S., and Stevenson, R. L. (1999). Estimation-Theoretic Approach to Dynamic Range Enhancement Using Multiple Exposures. *Journal of Electronic Imaging*, 12:2003.
- Rolland, J., Holloway, R. L., and Fuchs, H. (1994). A comparison of optical and video see-through head-mounted displays. In *SPIE 2351*, pages 293–307.
- Rouf, M., Mantiuk, R., Heidrich, W., Trentacoste, M., and Lau, C. (2011). Glare Encoding of High Dynamic Range Images. In *Proc. IEEE CVPR*, pages 1–8.
- Sabella, P. (1988). A rendering algorithm for visualizing 3D scalar fields. *ACM SIGGRAPH*, 22:51–58.
- Sabzmeydani, P. and Mori, G. (2007). Detecting Pedestrians by Learning Shapelet Features. In *Proc. CVPR*, pages 1–8.
- Sadjadi, F. (2007). Extraction of Surface Normal and Index of Refraction using a Pair of Passive Infrared Polarimetric Sensors. In *Proc. IEEE CVPR*, pages 1–5.
- Sajadi, B., Majumder, A., Hiwada, K., Maki, A., and Raskar, R. (2011). Switchable Primaries Using Shiftable Layers of Color Filter Arrays. *ACM Trans. Graph. (Siggraph)*, 30(3):1–10.
- Savarese, S. and Perona, P. (2002). Local Analysis for 3D Reconstruction of Specular Surfaces - Part II. In *Proc. ECCV*, pages 759–774.
- Saxby, G. (1994). *Practical Holography*. Prentice Hall.
- Schardin, H. (1942). Die Schlierenverfahren und ihre Anwendungen. *Ergebnisse der Exakten Naturwissenschaften*. English translation available as NASA TT F-12731, April 1970 (N70-25586), 20:303–439.

- Schechner, Y., Narasimhan, S. G., and Nayar, S. K. (2001). Instant Dehazing of Images using Polarization. In *Proc. IEEE CVPR*, pages 325–332.
- Schechner, Y. and Nayar, S. (2003a). Generalized Mosaicing: High Dynamic Range in a Wide Field of View. *International Journal on Computer Vision*, 53(3):245–267.
- Schechner, Y. and Nayar, S. (2005). Generalized Mosaicing: Polarization Panorama. *IEEE Trans. PAMI*, 27(4):631–636.
- Schechner, Y., Nayar, S., and Belhumeur, P. (2003). A Theory of Multiplexed Illumination. In *Proc. IEEE ICCV*, volume 2, pages 808–815.
- Schechner, Y., Nayar, S., and Belhumeur, P. (2007). Multiplexing for Optimal Lighting. *IEEE Trans. PAMI*, 29(8):1339–1354.
- Schechner, Y. and Nayar, S. K. (2003b). Polarization Mosaicking: High dynamic Range and Polarization Imaging in a Wide Field of View. In *Proc. SPIE 5158*, pages 93–102.
- Schechner, Y., Shamir, J., and Kiryati, N. (1999). Polarization-Based Decorrelation of Transparent Layers: The Inclination Angle of an Invisible Surface. In *Proc. ICCV*, pages 814–819.
- Schechner, Y. Y. and Karpel, N. (2004). Clear Underwater Vision. In *Proc. IEEE CVPR*, pages 536–543.
- Schechner, Y. Y. and Karpel, N. (2005). Recovery of Underwater Visibility and Structure by Polarization Analysis. *IEEE Journal of Oceanic Engineering*, 30(3):570–587.
- Seeing Machines (2008). www.seeingmachines.com.
- Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., and Vorozcovs, A. (2004). High dynamic range display systems. *ACM Trans. Graph.*, 23(3):760–768.
- Semiconductor, C. (2010). LUPA Image Sensors. <http://www.cypress.com/>.
- Settles, G. S. (2001). *Schlieren and Shadowgraph Techniques*. Cambridge University Press.
- Settles, G. S. (2010). Important Developments in Schlieren and Shadowgraph Visualization during the Last Decade. In *Proc. Int. Symp. on Flow Vis.*, page 267.
- Shahar, O., Faktor, A., and Irani, M. (2011). Space-Time Super-Resolution from a Single Video. In *Proc. IEEE CVPR*, pages 1–8.

- Shechtman, E., Caspi, Y., and Irani, M. (2002). Increasing Space-Time Resolution in Video. In *Proc. ECCV*, pages 753–768.
- Shechtman, E., Caspi, Y., and Irani, M. (2005). Space-Time Super-Resolution. *IEEE Trans. PAMI*, 27(4):531–545.
- Shih, M., Shishido, A., and Khoo, I. (2001). All-optical image processing by means of a photosensitive nonlinear liquid-crystal film: edge enhancement and image addition&subtraction. *Optics Letters*, 26:1140–1142.
- Shimadzu (2010). HyperVision HPV-2. <http://www.shimadzu.com/an/test/hpv-2/hpv1.html>.
- Slinger, C., Cameron, C., and Stanley, M. (2005). Computer-generated holography as a generic display technology. *Computer*, 38(8):46–53.
- Smith, B. M., Zhang, L., Jin, H., and Agarwala, A. (2009). Light Field Video Stabilization. In *Proc. of ICCV*, pages 1–8.
- SpheronVR (2010). SpheroCam HDR. <http://www.spheron.com>.
- Starck, J. and Hilton, A. (2008). Model-based human shape reconstruction from multiple views. *Computer Vision and Image Understanding (CVIU)*, 111(2):179–194.
- Stewart, R. G. and Roach, W. R. (1994). Field-sequential display system utilizing a backlit LCD pixel array and method for forming an image. U.S. Patent 5,337,068.
- Sullivan, A. (2003). A solid-state multi-planar volumetric display. In *SID Digest*, volume 32, pages 207–211.
- Sutherland, I. E. (1968). A head-mounted three dimensional display. *Fall Joint Computer Conference*, pages 757–764.
- Suyama, S., Ohtsuka, S., Takada, H., Uehira, K., and Sakai, S. (2004). Apparent 3-D image perceived from luminance-modulated two 2-D images displayed at different depths. *Vision Research*, 44(8):785–793.
- Taguchi, Y., Agrawal, A., Ramalingam, S., and Veeraraghavan, A. (2010a). Axial Light Fields for Curved Mirrors: Reflect Your Perspective, Widen Your View. In *Proc. IEEE CVPR*, pages 1–8.
- Taguchi, Y., Agrawal, A., Veeraraghavan, A., Ramalingam, S., and Raskar, R. (2010b). Axial-Cones: Modeling Spherical Catadioptric Cameras for Wide-Angle Light Field Rendering. *ACM Trans. Graph.*, 29:172:1–172:8.

- Tai, Y., Hao, D., Brown, M. S., and Lin, S. (2010). Correction of Spatially Varying Image and Video Motion Blur using a Hybrid Camera. *IEEE Trans. PAMI*, 32(6):1012–1028.
- Talvala, E.-V., Adams, A., Horowitz, M., and Levoy, M. (2007). Veiling Glare in High Dynamic Range Imaging. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):37:1–37:9.
- Tanida, J., Kumagai, T., Yamada, K., Miyatake, S., Ishida, K., Morimoto, T., Kondou, N., Miyazaki, D., and Ichioka, Y. (2001). Thin Observation Module by Bound Optics (TOMBO): Concept and Experimental Verification. *Applied Optics*, 40(11):1806–1813.
- Tanida, J., Shogenji, R., Kitamura, Y., Yamada, K., Miyamoto, M., and Miyatake, S. (2003). Color Imaging with an Integrated Compound Imaging System. *Optics Express*, 11(18):2109–2117.
- Tarini, M., Lensch, H. P., Goesele, M., and Seidel, H.-P. (2005). 3D Acquisition of Mirroring Objects using Striped Patterns. *Graphical Models*, 67(4):233–259.
- Telleen, J., Sullivan, A., Yee, J., Wang, O., Gunawardane, P., Collins, I., and Davis, J. (2007). Synthetic Shutter Speed Imaging. *Computer Graphics Forum (Eurographics)*, 26(3):591–598.
- Trentacoste, M., Heidrich, W., Whitehead, L., Seetzen, H., and Ward, G. (2007). Photometric Image Processing for High Dynamic Range Displays. *Journal of Visual Communication and Image Representation*, 18(5):439–451.
- Trifonov, B., Bradley, D., and Heidrich, W. (2006). Tomographic Reconstruction of Transparent Objects. In *Proc. EGSR*, pages 51–60.
- Tsai, R. Y. (1987). A Flexible new Technique for Camera Calibration. *IEEE Journal of Robotics and Automation*, 3(4):323–344.
- Tumblin, J., Agrawal, A., and Raskar, R. (2005). Why I want a Gradient Camera. In *Proc. IEEE CVPR*, pages 103–110.
- Ueda, K., Koike, T., Takahashi, K., and Naemura, T. (2008a). Adaptive IP Imaging with Variable-Focus Lens Array. In *Proc SPIE: Stereoscopic Displays and Applications XIX*, pages 68031A–9.
- Ueda, K., Lee, D., Koike, T., Takahashi, K., and Naemura, T. (2008b). Multi-Focal Compound Eye: Liquid Lens Array for Computational Photography. ACM SIGGRAPH New Tech Demo.

- Umeyama, S. and Godin, G. (2004). Separation of Diffuse and Specular Components of Surface Reflection by Use of Polarization and Statistical Analysis of Images. *IEEE Trans. PAMI*, 26(5):639–647.
- Unger, J., Wenger, A., Hawkins, T., Gardner, A., and Debevec, P. (2003). Capturing and Rendering with Incident Light Fields. In *Proc. EGSR*, pages 141–149.
- Vaish, V., Szeliski, R., Zitnick, C. L., Kang, S. B., and Levoy, M. (2006). Reconstructing Occluded Surfaces using Synthetic Apertures: Stereo, Focus and Robust Measures. In *Proc. IEEE CVPR*, pages 23–31.
- Valley, G. (2010). Viper FilmStream Camera. <http://www.grassvalley.com>.
- van Putten, E., Akbulut, D., Bertolotti, J., Vos, W., Lagendijk, A., and Mosk, A. (2011). Scattering Lens Resolves Sub-100 nm Structures with Visible Light. *Physical Review Letters*, 106(19):1–4.
- Veeraraghavan, A., Raskar, R., Agrawal, A., Chellappa, R., Mohan, A., and Tumblin, J. (2008). Non-Refractive Modulators for Encoding and Capturing Scene Appearance and Depth. In *Proc IEEE CVPR*.
- Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., and Tumblin, J. (2007). Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocussing. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):69.
- Veeraraghavan, A., Reddy, D., and Raskar, R. (2011). Coded Stroboscopic Photography: Compressive Sensing of High Speed Periodic Videos. *IEEE Trans. PAMI*, page to appear.
- Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., and Matusik, W. (2009). Dynamic Shape Capture using Multi-View Photometric Stereo. In *ACM Trans. Graph. (SIGGRAPH Asia)*, pages 1–11.
- Wang, S. and Heidrich, W. (2004). The Design of an Inexpensive Very High Resolution Scan Camera System. *Computer Graphics Forum (Eurographics)*, 23(10):441–450.
- Wehner, R. (1976). Polarized-Light Navigation by Insects. *Scientific American*, 235:106115.
- Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., and Debevec, P. (2005). Performance Telighting and Reflectance Transformation with Time-Multiplexed Illumination. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):756–764.
- Wetzstein, G. and Bimber, O. (2007). Radiometric Compensation through Inverse Light Transport. In *Proc. Pacific Graphics*, pages 391–399.

- Wetzstein, G., Heidrich, W., and Luebke, D. (2010a). Optical Image Processing Using Light Modulation Displays. *Computer Graphics Forum*, 29:1934–1944.
- Wetzstein, G., Ihrke, I., Gukov, A., and Heidrich, W. (2011a). Towards a Database of High-dimensional Plenoptic Images. In *Proc. ICCP (Poster)*.
- Wetzstein, G., Ihrke, I., and Heidrich, W. (2010b). Sensor Saturation in Fourier Multiplexed Imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Wetzstein, G., Ihrke, I., and Heidrich, W. (2011b). On Plenoptic Multiplexing and Reconstruction. *in submission*, pages 1–17.
- Wetzstein, G., Ihrke, I., Lanman, D., and Heidrich, W. (2011c). Computational Plenoptic Imaging. *Computer Graphics Forum (to appear)*, pages 1–25.
- Wetzstein, G., Ihrke, I., Lanman, D., and Heidrich, W. (2011d). State of the Art in Computational Plenoptic Imaging. In *Proc. Eurographics (STAR)*, pages 1–24.
- Wetzstein, G., Raskar, R., and Heidrich, W. (2011e). Refractive Shape from Light Field Distortion. In *Proc. ICCV*, pages 1–7.
- Wichmann, F. A. and Hill, N. J. (2001). The Psychometric Function: Fitting, Sampling and Goodness-of-Fit. *Perception and Psychophysics*, 63(8):1293–1313.
- Wilburn, B., Joshi, N., Vaish, V., Levoy, M., and Horowitz, M. (2004). High Speed Video Using a Dense Camera Array. In *Proc IEEE CVPR*, volume 1, pages 472–479.
- Wilburn, B., Joshi, N., Vaish, V., Talvala, E.-V., Antunez, E., Barth, A., Adams, A., Horowitz, M., and Levoy, M. (2005). High Performance Imaging Using Large Camera Arrays. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):765–776.
- Wilburn, B., Smulski, M., Lee, K., and Horowitz, M. A. (2002). The Light Field Video Camera. In *SPIE Electronic Imaging*, pages 29–36.
- Wolff, L. B. and Boult, T. E. (1991). Constraining Object Features using a Polarization Reflectance Model. *IEEE Trans. PAMI*, 13(7):635–657.
- Woodgate, G. and Harrold, J. (2003). High Efficiency Reconfigurable 2D/3D Autostereoscopic Display. *SID DIGEST*.
- Woodham, R. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*.

- Wuttig, A. (2005). Optimal Transformations for Optical Multiplex Measurements in the Presence of Photon Noise. *Appl. Opt.*, 44(14):2710–2719.
- Yang, J., Lee, C., Isaksen, A., and McMillan, L. (2000). A Low-Cost Portable Light Field Capture Device. ACM SIGGRAPH Technical Sketch.
- Yang, J. C., Everett, M., Buehler, C., and McMillan, L. (2002). A Real-Time Distributed Light Field Camera. In *Proc EGSR*, pages 77–86.
- Yasuma, F., Mitsunaga, T., Iso, D., and Nayar, S. K. (2010). Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. *IEEE Trans. Im. Proc.*, 99.
- Yeh, P. and Gu, C. (2009). *Optics of Liquid Crystal Displays*. John Wiley and Sons.
- Yelleswarapu, C., Wu, P., Kothapalli, S., Rao, D., Kimball, B., Sai, S., Gowrishankar, R., and Sivaramakrishnan, S. (2006). All-optical spatial filtering with power limiting materials. *Optics Express*, 14:1451–1457.
- Yendo, T., Kawakami, N., and Tachi, S. (2005). Seelinder: the cylindrical lightfield display. In *ACM SIGGRAPH Emerging Technologies*.
- Z Corporation (2010). ZPrinter 650. <http://www.zcorp.com>.
- Zhang, C. and Chen, T. (2005). Light Field Capturing with Lensless Cameras. In *Proc. ICIP*, pages III – 792–5.
- Zhang, L. and Nayar, S. K. (2006). Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. Graph. (Siggraph)*, 25(3):907–915.
- Zhang, X. and Brainard, D. H. (2004). Estimation of Saturated Pixel Values in Digital Color Imaging. *JOSA A*, 21(12):2301–2310.
- Zhang, X. and Cox, C. S. (1994). Measuring the Two-dimensional Structure of a Wavy Water Surface Optically: A Surface Gradient Detector. *Experiments in Fluids*, 17:225–237.
- Zhang, Z. (2000). A Flexible new Technique for Camera Calibration. *IEEE Trans. PAMI*, 22(11):1330–1334.
- Zhou, C., Lin, S., and Nayar, S. K. (2009). Coded Aperture Pairs for Depth from Defocus. In *Proc. ICCV*.

- Zhou, C. and Nayar, S. (2009). What are Good Apertures for Defocus Deblurring? In *Proc. ICCP*, pages 1–8.
- Zongker, D. E., Werner, D. M., Curless, B., and Salesin, D. H. (1999). Environment Matting and Compositing. In *ACM Trans. Graph. (Proc. SIGGRAPH)*, pages 205–214.
- Zwicker, M., Matusik, W., Durand, F., and Pfister, H. (2006). Antialiasing for automultiscopic 3D displays. In *Eurographics Symposium on Rendering*.
- Zwicker, M., Vetro, A., Yea, S., Matusik, W., Pfister, H., and Durand, F. (2007). Resampling, antialiasing, and compression in multiview 3-D displays. *IEEE Signal Processing Magazine*, 24(6):88–96.

Appendix A

Additional Details on Multi-Layer Light Field Displays

This appendix contains additional images of the multi-layer light field display prototype constructions, pseudo-code for the GPU-based SART implementation, and further documentation of simulated display performance for different solvers and parameters with polarization-rotating layers and light-attenuating layers. Furthermore, approaches for nonlinear layer decomposition of light field for multi-layer displays are formulated.

A.1 Additional Photographs of Light Field Display Prototypes

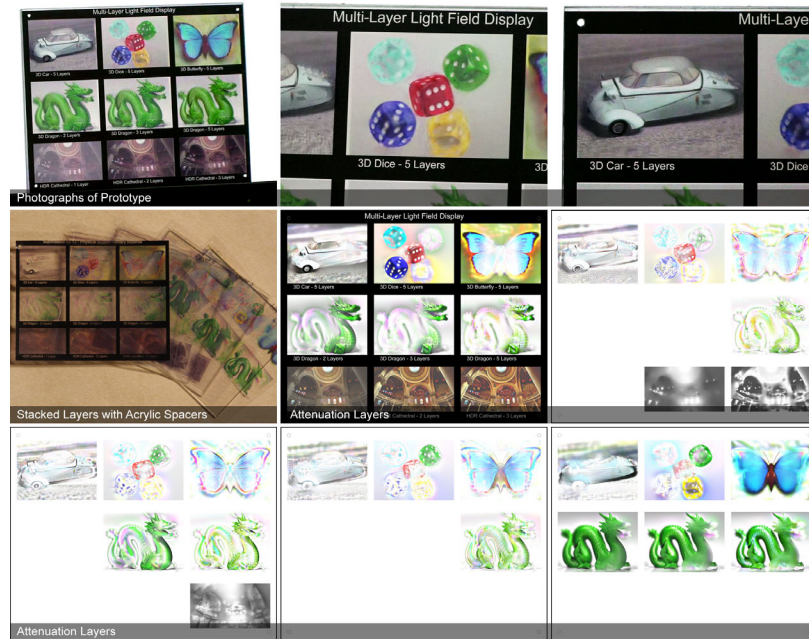


Figure A.1: Additional photographs of a multi-layer transparency-based light field display prototype (top row), along with the individual attenuation layers (center and bottom rows).



Figure A.2: *Photographic documentation of the multi-layer LCD prototype constructed by Matthew Hirsch at the MIT Media Lab. Four monochrome, off-the-shelf medical LCDs were modified. Polarizing films were removed and electronics repositioned so the panels could be mounted on custom-fabricated frames. The layers are separated by acrylic spacers and illuminated by a uniform backlight.*

A.2 Pseudo-Code for GPU-based SART Implementation

In the following, pseudo-code for the GPU-based SART implementation is provided. The implementation, tested on both Windows and Linux, is programmed in C++ and uses OpenGL and Cg shaders. The algorithm assumes a light field with N distinct views and K layers positioned at user-defined depths along the optical axis. Intermediate quantities, including the target light field views and temporary variables storing weights and layer patterns, are internally rendered into 16-bit off-line framebuffers (i.e., Framebuffer Objects, FBOs) before the optimized patterns are displayed on the individual polarization-rotating layers. Only three separate Cg fragment programs are required, each performing the action implied by their names.

Algorithm SART

variables FBO_LF[N], FBO_LF_TMP[N], FBO_LAYERS[K], FBO_W[N], FBO_V[K], DEPTH[K]

function displayLoop()

if not initialized

 initialize all FBO_W, FBO_V

end

 drawLightField();

 runSART();

 drawReconstructedLayers();

end

function drawLightField()

for all light field views i

 activate FBO_LF[i]

 set perspective i

 drawScene(); // render desired 3D scene (e.g., a teapot)

end

end

function runSART()

for all iterations k

 // 1. compute $Ax^{(k)}$

for all light field views i

 activate FBO_LF_TMP[i]

 enable BLEND_MODE

 set perspective i

for all layers l

 draw 2D plane at DEPTH[l] textured with FBO_LAYERS[l]

end

end

 // 2. given $Ax^{(k)}$, compute $W(b - Ax^{(k)})$

for all light field views i

 activate FBO_LF_TMP[i]


```

        activate CG_SHADER_MULTIPLY_SUBTRACT(FBO_W[i],FBO_LF[i],FBO_LF_TMP[i])
        draw orthographic 2D plane with normalized texcoords
    end
    // 3. given  $W(b - Ax^{(k)})$ , compute  $VA^T(W(b - Ax^{(k)}))$ 
    for all layers  $l$ 
        activate FBO_LAYER[l]
        activate BLEND_MODE
        for all light field views  $i$ 
            set projective texture to perspective  $i$ 
            activate CG_SHADER_MULTIPLY(FBO_V[l],FBO_LAYER[l])
            draw orthographic 2D plane with automatic texcoord generation
        end
    end
    // 4. enforce constraints by clamping values outside feasible range
    for all layers  $l$ 
        activate FBO_LAYER[l]
        activate CG_SHADER_CLAMP(FBO_LAYER[l])
        draw orthographic 2D plane textured with FBO_LAYER[l]
    end
end
end

function drawReconstructedLayers()
    for all layers  $l$ 
        set viewport for display  $l$ 
        draw orthographic 2D plane textured with FBO_LAYERS[l]
    end
end
end

```

A.3 Synthetic Performance Evaluation of Multi-Layer Light Field Displays



Figure A.3: Additional results for the “car” scene, taken from the archive available at <http://www.layered3d.com>. Simulated views are compared for polarization-rotating layers (first and second columns) and attenuating layers (fourth and fifth columns). Rows illustrate, from top to bottom: target views, reconstructions using the off-line solver for two, three, and four layers (for the same depth range), and SART reconstructions with two, five, and fifty iterations with four layers. Columns three and six present decomposed layers.

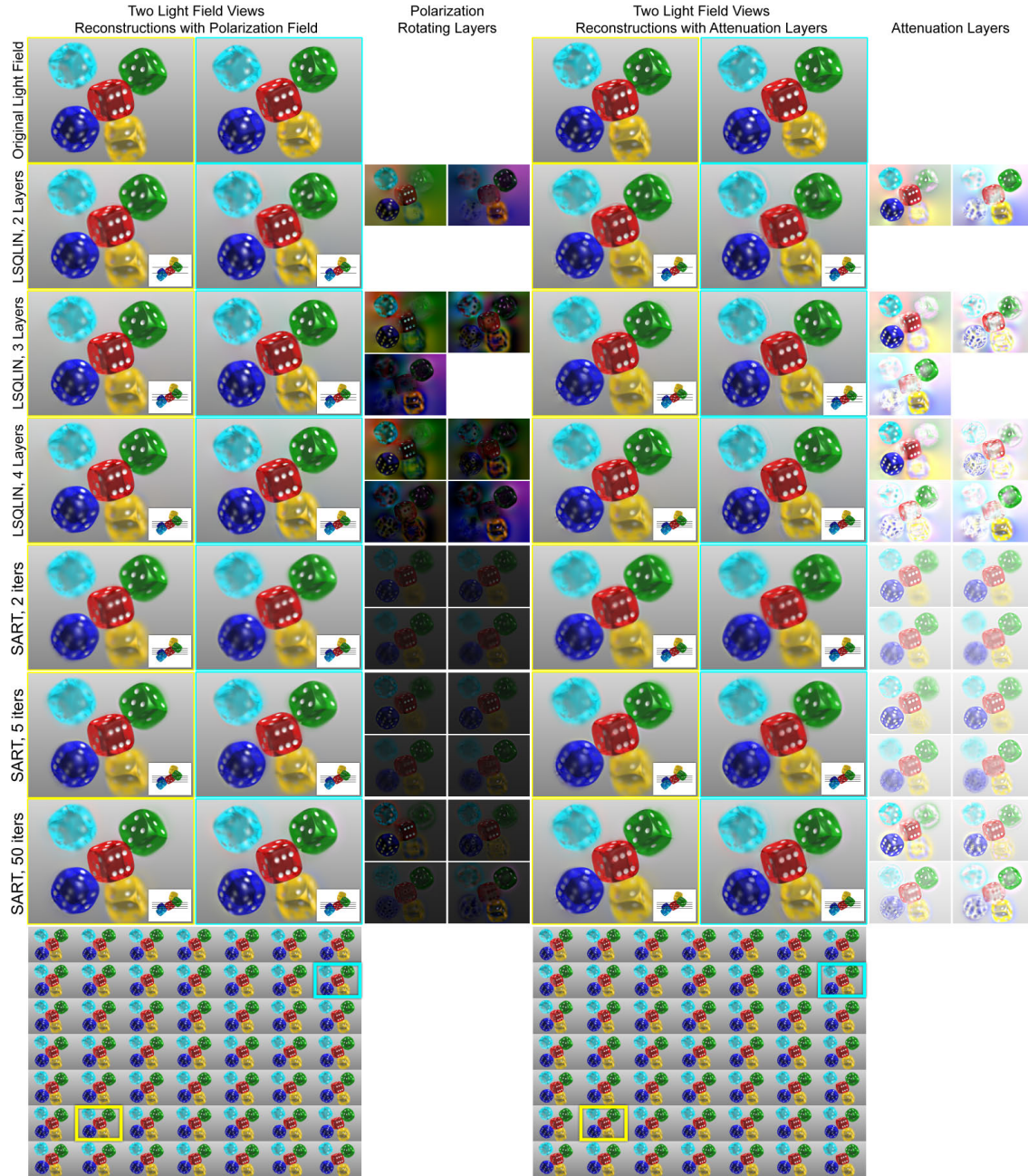


Figure A.4: Additional results for the “dice” scene, taken from the archive available at <http://www.layered3d.com>. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure A.3. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.

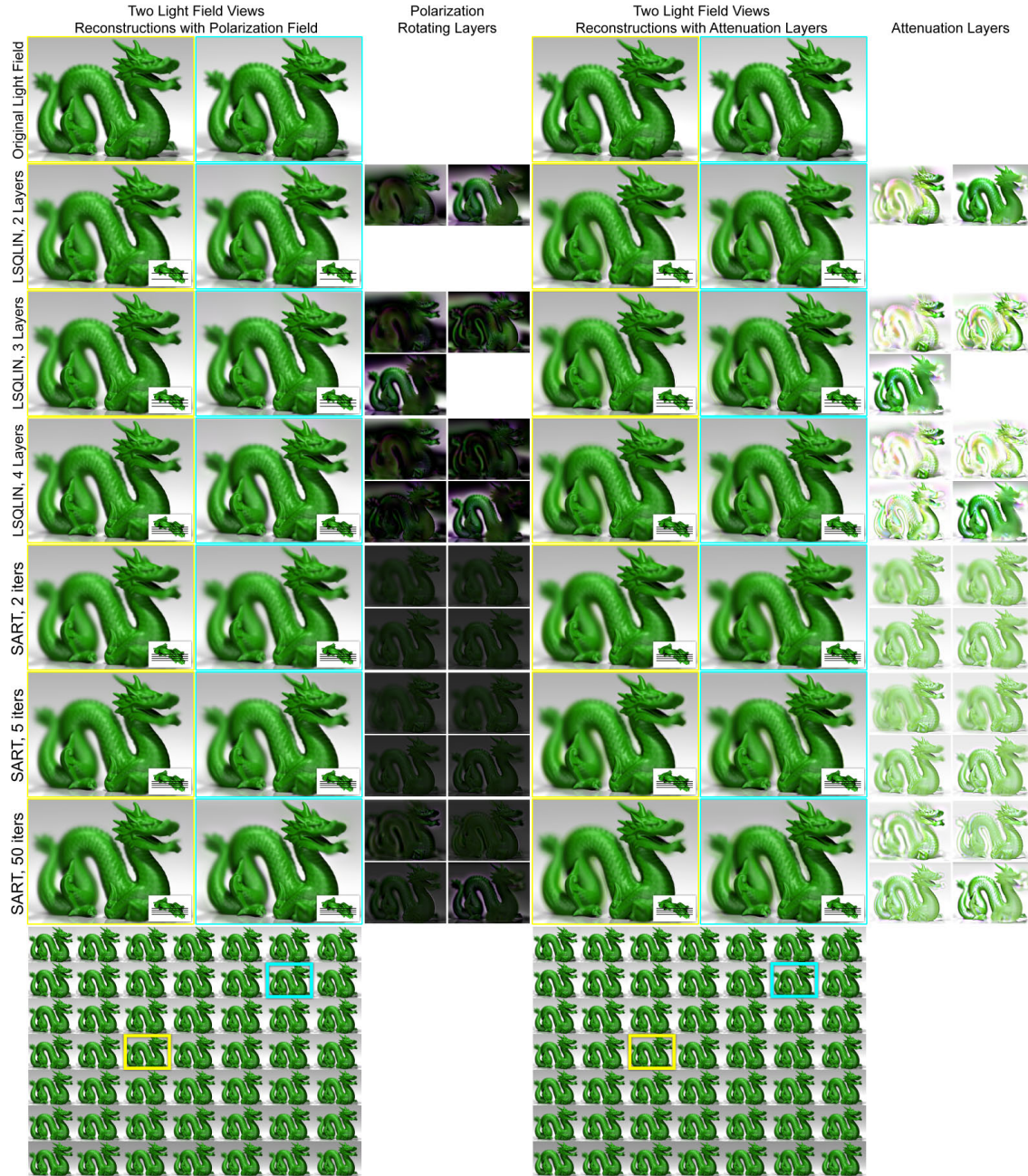


Figure A.5: Additional results for the “dragon” scene, , using the “dragon” model from <http://graphics.stanford.edu/data/3Dscanrep/>. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure A.3. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.



Figure A.6: Additional results for the “Buddha” scene, using the “Buddha” model from <http://graphics.stanford.edu/data/3Dscanrep/>. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure A.3. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.

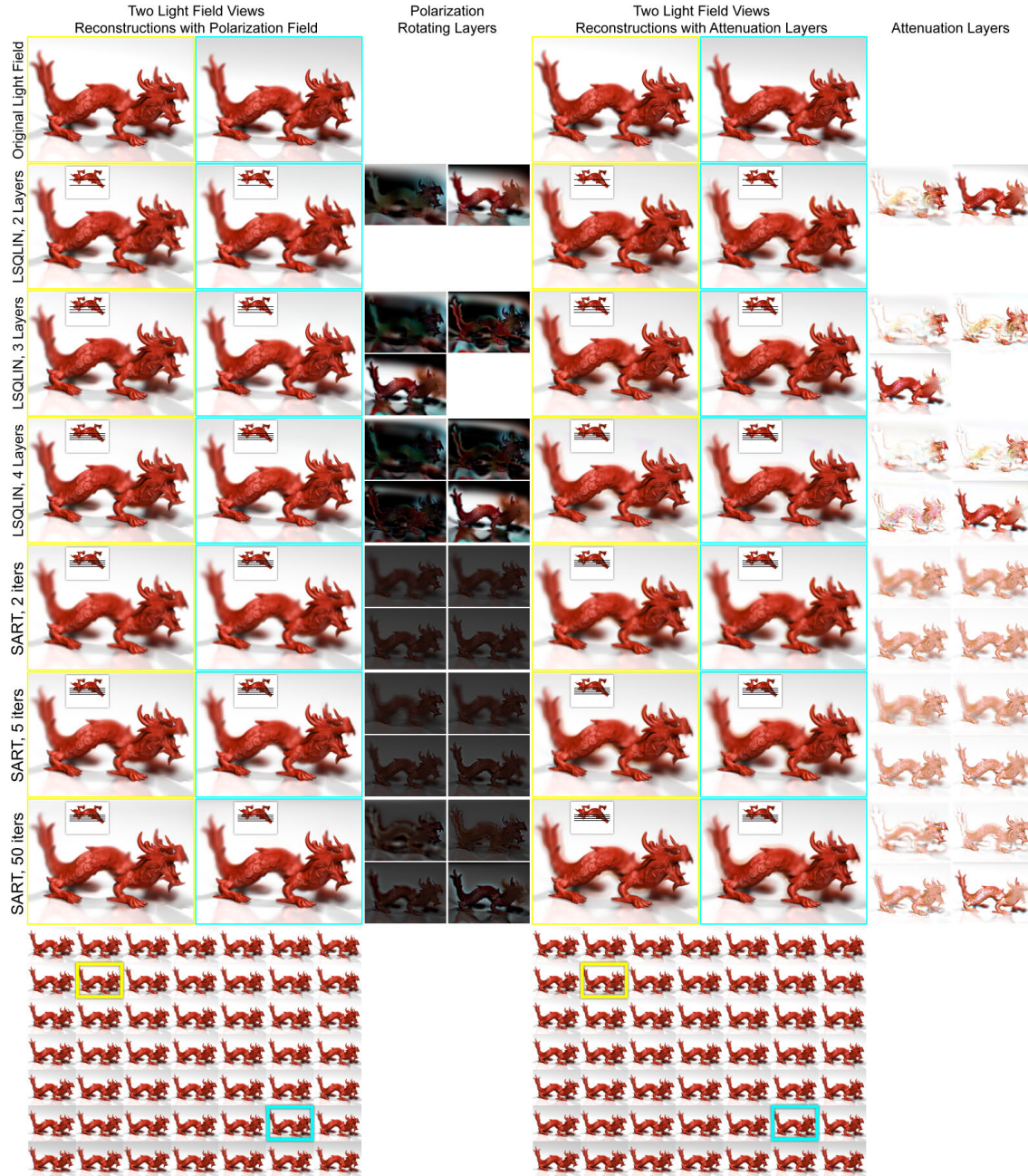


Figure A.7: Additional results for the “red dragon” scene, using the “dragon” model from <http://graphics.stanford.edu/data/3Dscanrep/>. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure A.3. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.

A.4 Nonlinear Image Synthesis for Multi-Layer Displays

This section documents solutions to nonlinear image synthesis for both attenuating and polarization-rotating multi-layer light field displays. While Chapters 3 and 4 explore computationally efficient approaches to solving the image formation (Eqs. 3.12 and 4.5) using linearized formulations, we derive expressions for solving the nonlinear equations in the following subsections. Note that we derive these expressions for the layered image formation, with extensions for volumetric attenuators being straightforward.

A.4.1 Derivation for Attenuation Layers

It is shown in Section 3.2.3 that the light field $\tilde{l}(u, a)$ emitted by an attenuation-based multi-layer display is given by

$$\tilde{l}(u, a) = \prod_{k=1}^{N_l} t_k(u + (d_k/d_r)a), \quad (\text{A.1})$$

where $t_k(\xi)$ is the transmittance of mask k . The backlight is assumed to be uniform and unit-valued¹. We demonstrate in Equation 3.13 that Equation A.1 becomes linear in log-space. Tomographic approaches usually minimize a least-squared error function based on the linearized formulation (i.e., Eq. 3.11), because of improved computational efficiency. This approach, however, slightly changes the problem that is being solved in favor of a more tractable solution; a least-squared error formulation in log-space will have a different optimal solution than a least-squared error objective in linear space. Specifically, the differences of small values, close to zero, are weighted more heavily than those of larger values close to one. The squared difference between 0.9 and 0.8, for instance, is 0.01 and that between 0.2 and 0.1 is also 0.01. In log-space, however, these values become 0.01 and 0.48, respectively. In practice, this error bias becomes visible as dark halos around high contrast edges, which is particularly problematic for light fields with a wide field of view (see Fig. 3.16).

A nonlinear objective function based on Equation A.1 can be formulated as

$$\mathcal{O}(u, a) = l(u, a) - \prod_{k=1}^{N_l} t_k(u + (d_k/d_r)a), \quad (\text{A.2})$$

¹In practice the light distribution of the backlight can be estimated as a one-time pre-processing step and divided out of the light field.

where $l(u, a)$ is the target light field.

Least-squared error solutions to general, nonlinear objective functions can be found using iterative optimization schemes as, for instance, the Gauss-Seidel or the Levenberg-Marquardt methods. In any iterative approach, the unknown variables are initialized with a guess, for example random values, which are subsequently updated in each iteration. The update for the estimated solution at some iteration can be computed by optimizing a first-order approximation of the objective function around the current estimate of the solution [Boyd and Vandenberghe, 2004]. This approximation is defined as the matrix of partial derivatives or Jacobian matrix \mathbf{J} of a nonlinear function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. The function f is defined as $y_1(x_1 \dots x_n), \dots, y_m(x_1 \dots x_n)$, and its Jacobian matrix $\mathbf{J} \in \mathbb{R}^m \times \mathbb{R}^n$ as

$$\mathbf{J} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}. \quad (\text{A.3})$$

The Jacobian matrix of Equation A.2 is then

$$\mathbf{J}_{\text{attenuation}} = - \left[\text{diag} \left(\prod_{k=1 \dots N_l}^{k \neq 1} t_k(u + (d_k/d_r)a) \right) \cdots \text{diag} \left(\prod_{k=1 \dots N_l}^{k \neq N_l} t_k(u + (d_k/d_r)a) \right) \right], \quad (\text{A.4})$$

where each of the sparse block matrices $\text{diag} \left(\prod_{k=1 \dots N_l}^{k \neq i} t_k(u + (d_k/d_r)a) \right)$ is of size $m \times n/N_l$ and corresponds to a specific layer i . In this case, \mathbf{J} is large-scale and sparse, which allows for memory-efficient representations. An analytic computation of the Jacobian matrix is generally preferred to a finite difference approximation, which is provided by many commercial optimization packages, because it is more precise and computationally more efficient.

Another important property of Equation A.2 is that it is convex. By definition, the linear formulation in log-space is convex; the nonlinear objective function (Eq. A.2) simply adds an exponential function around the log-linear objective. As derived by Boyd and Vandenberghe [2004] (Chapter 3.2.4), this operation preserves convexity.

Convexity of the objective function ensures that each successive update in an iterative optimization scheme, if computed with the proper step length, is closer to the optimal solution than the previous estimate. Convergence toward the unique optimal solution is, therefore, guaranteed.

We use the Matlab LSQLIN solver, implementing a sparse, constrained, large-scale trust region method [Coleman and Li, 1996] for linear and nonlinear equation systems. The nonlinear solver usually converges in the same number (about 8–14) of iterations as the linear solution outlined in Section 3.5.1. The processing time, however, increases by an order of magnitude to about 10 hours for similar resolutions as discussed in Section 3.5.1, seeding it with the result of the linear approximation improves the performance.

A.4.2 Derivation for Polarization-Rotating Layers

A similar strategy can be pursued for solving the nonlinear problem of estimating polarization-rotating layers given a light field. As discussed in Section 4.5, the light field $\tilde{l}(u, a)$ emitted by a polarization field display is

$$\tilde{l}(u, a) = \sin^2 \left(\sum_{k=1}^K \phi_k(u + (d_k/d_r)a) \right), \quad (\text{A.5})$$

where the backlight is assumed to be uniform. This results in the following nonlinear objective function:

$$\mathcal{O}(u, a) = l(u, a) - \sin^2 \left(\sum_{k=1}^K \phi_k(u + (d_k/d_r)a) \right). \quad (\text{A.6})$$

This periodic objective function has a lot more degrees of freedom as the linearized version, outlined in Equation 4.7, which only uses the principle values of the sine ranges. Following Section A.4.1, we can derive an analytic expression for the Jacobian matrix of the nonlinear formulation in Equation A.6 as

$$\mathbf{J}_{\text{polarization}} = \text{diag} \left(-2 \sin \left(\sum_{k=1}^K \phi_k(u_1 + (d_k/d_r)a_1) \right) \cos \left(\sum_{k=1}^K \phi_k(u_1 + (d_k/d_r)a_1) \right) \right) \mathbf{P}, \quad (\text{A.7})$$

where $\mathbf{P} \in \mathbb{R}^m \times \mathbb{R}^n$ is the projection matrix defined in Section 4.2.3 and the $\text{diag}(\cdot)$ operation results in a diagonal matrix with a size $m \times m$.

We have implemented a nonlinear solution to Equation A.6 using Matlab’s LSQNONLIN solver, with solutions requiring about six hours to be computed. Results are shown in Figure 4.13.

Unfortunately, the nonlinear polarization field image formation (Eq. A.6) is, as opposed to attenuating layers, not convex. The sine term introduces periodicity; experiments with iterative nonlinear optimization approaches have shown that these solution, when seeded with the output of the linearized formulation, converge to the same local minima as the linear approximations (see Fig. 4.13), whereas random initial guesses have not led to any meaningful results. Global optimization schemes could exploit the additional degrees of freedom, overcoming limitations of field of view and depth of field. Practical implementations for high-resolution polarization field displays, however, are to be implemented in the future.

Appendix B

Proofs of Plenoptic Multiplexing Theorems

B.1 Proof of PSM Theorem

Throughout Chapter 9, we assume that the spatial basis functions $\sigma_j(\vec{x})$ are super-pixel-periodic¹, i.e. $\sigma_j(\vec{x}) = \sigma_j(\vec{x} + t\Delta\vec{x}_s)$, $\forall t \in \mathbb{Z}$. The offset between successive super-pixels is denoted as $\Delta\vec{x}_s$. This notation allows us to define a sampling operator as

$$\text{III}_k(\vec{x}) = \sum_{t \in \mathbb{Z}} \delta(\vec{x} + t\Delta\vec{x}_s + \Delta\vec{x}_k), \quad (\text{B.1})$$

where $\Delta\vec{x}_k$ is the offset of individual samples within each super-pixel. This sampling operator basically extracts a sub-image or channel $k = 1 \dots M$ from an interleaved sensor image, where only corresponding sub-pixels within the super-pixels are included in each channel. Sampling such a channel $\tilde{\mathbf{i}}_k(\vec{x})$ from a sensor image $\mathbf{i}(\vec{x})$, in combination with Equation 9.4, results in the following expression

$$\begin{aligned} \tilde{\mathbf{i}}_k(\vec{x}) &= \text{III}_k(\vec{x})\mathbf{i}(\vec{x}) = \text{III}_k(\vec{x}) \left(\sum_{j=1}^N \sigma_j(\vec{x})\rho_j(\vec{x}) \right) \\ &= \sum_{j=1}^N \sigma_j^k \text{III}_k(\vec{x})\rho_j(\vec{x}). \end{aligned} \quad (\text{B.2})$$

As discussed in Section 9.2.3, the spatial basis functions $\sigma_j(\vec{x})$ become spatially-invariant constants σ_j^k in the sampled channels. Each channel $\tilde{\mathbf{i}}_k(\vec{x})$ is thus associated with N constants σ_j^k defined by the spatial basis. Reconstructing a channel $\mathbf{i}_k(\vec{x})$ from its sampled representation $\tilde{\mathbf{i}}_k(\vec{x})$ is performed by convolving with a reconstruction filter kernel $\mathbf{f}(\vec{x})$:

¹Following Sec. 9.2.4, we omit periodicity in the second spatial dimension in our notation of t for clarity.

$$\mathbf{i}_k(\vec{x}) = \widetilde{\mathbf{i}}_k(\vec{x}) \otimes \mathbf{f}(\vec{x}). \quad (\text{B.3})$$

According to the sampling theorem, the original signal must be spatially band-limited for this reconstruction to be a faithful representation.

Due to the spatial basis being a set of constants at every position in the interpolated channels $\mathbf{i}_k(\vec{x})$, and the plenoptic basis being spatially-invariant, a convolution with the filter kernel can be formulated as

$$\begin{aligned} \mathbf{i}_k(\vec{x}) &= \left(\sum_{j=1}^N \sigma_j^k \mathbb{I}\mathbb{I}_k(\vec{x}) \boldsymbol{\rho}_j(\vec{x}) \right) \otimes \mathbf{f}(\vec{x}) \\ &= \sum_{j=1}^N \sigma_j^k \int_{\mathcal{P}} \pi_j(\vec{p}) \left(\mathbb{I}\mathbb{I}_k(\vec{x}) \mathbf{l}_\lambda(\vec{x}, \vec{p}) \otimes \mathbf{f}(\vec{x}) \right) d\vec{p} \end{aligned} \quad (\text{B.4})$$

Equation B.4 shows that all channels $\mathbf{i}_k(\vec{x})$ are locally related to the plenoptic function via a linear combination. It also shows that applying a reconstruction filter to the measured channels, i.e. before decorrelation, is equivalent to applying the same filter to the plenoptic function itself, i.e. after decorrelation.

B.2 Proof of PFM Theorem

The proof of Theorem 2 follows Equations 9.6–9.8. To provide additional detail, we start with Equation 9.8

$$\mathcal{F}_x \{ \mathbf{i}(\vec{x}) \} = \sum_{k=1}^M \left(\delta \left(\vec{\omega}_x - \Delta \vec{\omega}_x^k \right) \otimes \sum_{j=1}^N \hat{\sigma}_j^k \hat{\boldsymbol{\rho}}_j(\vec{\omega}_x) \right), \quad (\text{B.5})$$

where $k\Delta\vec{\omega}_x$ has been replaced by $\Delta\vec{\omega}_x^k$, a vector to the center frequency of a Fourier tile (Fig. B.1, right), to allow for generalized sampling patterns. If the signal is properly band-limited the image's Fourier transform separates into disjoint sets, each encoding one Fourier tile $\hat{\mathbf{i}}_k(\vec{\omega}_x)$ (Fig. B.1, right). To separate the notation for spatial frequencies in the sensor image and Fourier tiles we use the substitution $\vec{\omega}_x' = \vec{\omega}_x - \Delta\vec{\omega}_x^k$.

The Fourier tiles can be cropped from the Fourier transformed sensor image by applying a

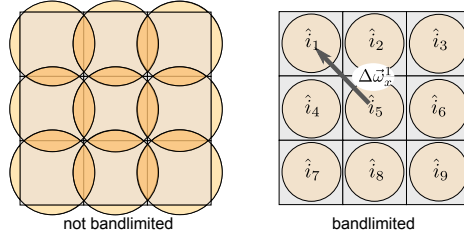


Figure B.1: Information overlaps in the Fourier domain if the signal is not suitably band-limited (left). With the appropriate band-limitation, the Fourier representation decomposes into distinct correlated Fourier tiles $\hat{\mathbf{i}}_k$. The arrow in the right figure indicates the position $\Delta\vec{\omega}_x^k$ of the Dirac peaks, i.e. the center frequencies, of a Fourier tile.

rect filter²:

$$\begin{aligned}\hat{\mathbf{i}}_k(\vec{\omega}'_x) &= \text{rect}^k(\vec{\omega}'_x) \delta(\vec{\omega}'_x) \otimes \sum_{j=1}^N \hat{\boldsymbol{\sigma}}_j^k \hat{\boldsymbol{\rho}}_j(\vec{\omega}'_x) \\ &= \text{rect}^k(\vec{\omega}'_x) \sum_{j=1}^N \hat{\boldsymbol{\sigma}}_j^k \int_{\mathcal{P}} \boldsymbol{\pi}(\vec{p}) \hat{\mathbf{l}}_\lambda(\vec{\omega}'_x, \vec{p}) d\vec{p}.\end{aligned}\tag{B.6}$$

The convolution with a Dirac train in Equation 9.8 reduces to a convolution with a single Dirac peak $\delta(\vec{\omega}'_x)$ because of band-limitation. This is a unit operation and thus removed from the equations. \square

In addition, by inverse Fourier transforming and sampling the Fourier tile $\hat{\mathbf{i}}_k(\vec{\omega}'_x)$ we see that the reconstruction filter is indeed a sinc:

$$\mathbf{i}_k(\vec{x}) = \sum_{j=1}^N \boldsymbol{\sigma}_j^k \int_{\mathcal{P}} \boldsymbol{\pi}_j(\vec{p}) \left(\mathbb{I}_k(\vec{x}) \mathbf{l}_\lambda(\vec{x}, \vec{p}) \otimes \text{sinc}(\vec{x}) \right) d\vec{p}\tag{B.7}$$

²For other sampling patterns, corresponding, e.g. hexagonal, masking functions would be used. In addition, apodization functions can be used to reduce ringing artifacts in the reconstruction at the expense of decreasing the effective resolution, see e.g. [Veeraraghavan et al., 2007]. This does not affect the proof.