BELEN MASIA

COMPUTATIONAL IMAGING: COMBINING OPTICS, COMPUTATION AND PERCEPTION

delivered by
EUROGRAPHICS DIGITAL LIBRARY
www.eg.org diglib.eg.org

COMPUTATIONAL IMAGING: COMBINING OPTICS, COMPUTATION AND PERCEPTION

BELEN MASIA

Supervisor: DIEGO GUTIERREZ

Tesis Doctoral · Ingeniería Informática
Departamento de Informática e Ingeniería de Sistemas
Escuela de Ingeniería y Arquitectura
Universidad de Zaragoza

October 2013

Belen Masia: *Computational Imaging: Combining Optics, Computation and Perception,* , © October 2013

To the memory of Enrique, Tere and Mirentxu.

ABSTRACT

This thesis presents contributions on the different stages of the imaging pipeline, from capture to display, and including interaction as well; we embrace all of them under the concept of Computational Imaging. The addressed topics are diverse, but the driving force and common thread has been the conviction that a *combination* of improved optics and hardware (*optics*), computation and signal processing (*computation*), and insights from how the human visual system works (*perception*) are needed for –and will lead to– significant advances in the imaging pipeline. In particular, we present contributions in the areas of: coded apertures for defocus deblurring, reverse tone mapping, disparity remapping for automultiscopic and stereoscopic displays, visual comfort when viewing stereo content, interaction paradigms for light field editing, and femto-photography and transient imaging.

MEASURABLE CONTRIBUTIONS AND MERITS

The realization of this thesis has yielded the following results; a more detailed list can be found in Section 1.5:

- 7 JCR-indexed journal publications (3 of them ACM Transactions on Graphics) [305, 302, 308, 307, 112, 459, 490]
- 5 peer-reviewed conference publications (one of them a SIGGRAPH Talk) [303, 304, 458, 199, 135]
- 1 peer-reviewed tutorial course [151]
- 2 research stays (totalling seven months) at MIT Media Lab
- 1 research visit (ten days) to Tsinghua University
- 2 PhD grants and an NVIDIA graduate fellowship
- 3 supervised PFCs and 1 more in progress
- 2 best papers
- 4 invited talks
- Participation in 5 research projects
- Reviewer for 4 journals and 8 international conferences, and program committee member for 3 international conferences

RESUMEN

Esta tesis presenta contribuciones en distintas partes del *pipeline* de imagen, desde la captura de imágenes, hasta la presentación de las mismas en un monitor u otro dispositivo, pasando por el procesamiento que se produce en los pasos intermedios. Englobamos las distintas técnicas y algoritmos utilizados en las diferentes etapas bajo el concepto de Imagen Computacional (*Computational Imaging* en inglés). Los temas son diversos, pero el motor e hilo conductor ha sido la idea de que una combinación de óptica avanzada, computación y procesamiento de señal, y conocimiento del funcionamiento de la percepción y el sistema visual humano son necesarias y conducirán a mejoras significativas en cómo capturamos y mostramos el mundo.

Las primeras cámaras fotográficas comerciales datan de 1839. Hoy en día, tras más de 150 años, y con la aparición de la fotografía digital, el concepto de cámara fotográfica es muy similar al de esas primeras cámaras. La luz que viaja a través de una escena se puede ver como una función multidimensional denominada función plenóptica [6]; una fotografía convencional muestrea sólo dos dimensiones de dicha función, integrando sobre un rango de las otras dimensiones. Así, una gran cantidad de información de la escena se pierde.

La fotografía computacional, combinando dos pilares, óptica y computación, parte de la idea de codificar la información que llega al sensor, de forma que podamos muestrear otras dimensiones de esa función plenóptica, para a posteriori decodificarla, obteniendo una imagen que no hubiera sido posible capturar con técnicas tradicionales.

El tercer pilar de esta tesis es la percepción, el funcionamiento del sistema visual humano. Argumentamos y mostramos que conocer el sistema visual humano y explotar sus características ayuda a superar las limitaciones del hardware y los algoritmos existentes, y puede contribuir a mejorar la experiencia del espectador o usuario.

Esta combinación de óptica, computación y percepción también puede dar y ha dado sus frutos en el campo de los *displays* (monitores, dispositivos de visualización), a los que se dedica la segunda parte de esta tesis. Los displays son limitados en cuanto a su capacidad de representar el mundo real, y conocer cómo procesa nuestro sistema visual la información puede ayudar a superar limitaciones existentes. Esta idea no es nueva, pero todavía hay un gran número de problemas sin resolver que se pueden beneficiar de esta manera de abordarlos, de este enfoque multidisciplinar. A lo largo de esta tesis hemos ahondado en esta idea, proponiendo soluciones a un número de problemas existentes en el *pipeline* de imagen. En particular, presentamos contribuciones en las siguientes áreas: correción de desenfoque mediante aperturas codificadas, reproducción de tono inversa, remapeo de disparidades (*disparity remapping*) para monitores automultiescópicos y estereoscópicos, comfort y fatiga durante la visualización de contenido estéreo, paradigmas de interacción para edición de *light fields*, y femtofotografía.

Le vrai luxe, c'est le temps.
— J. J. Goldman, 1951–
(via G. Drettakis)

ACKNOWLEDGMENTS

So many people have helped me throughout this thesis, or have made these four years better, either by teaching me, encouraging me, or simply being at my side when I needed it. I particularly would like to thank a series of people. The list is long. It has to be.

Diego, for so many things. When I decided to accept your offer to do a PhD, I did it under the conviction that you were a great professional from whom I could learn and be a better researcher; you have more than fulfilled that expectation. For your help, your advice, your support, your patience, and your example, thank you.

Ramesh, for creating and maintaining an amazing combination of people together, and giving me the opportunity to be within that group of people, twice. Also for being an inspiration, for that incredible combination of knowledge and vision.

Gordon, because of how much you have taught me throughout this PhD. For sharing your knowledge, both technical and non-technical; for pushing when you had to push; for understanding when I needed it; for setting an example to follow; for teaching me how to play. And for the beers too.

The co-authors of all my publications, and of the work in progress which still has not materialized in a paper. Thanks for all I have learned from you.

The people from the *Graphics and Imaging Lab*, for being such a great team, always willing to contribute, to help out, to create good stuff, for the discussions, and of course for all the good times, beers, coffees, lunch times, dinners, and nights ended up at the Crapula or the Zeta.

From them, I cannot help a special mention to: *Adrián Jarabo*, for his relevance in the work here presented, and for his hard work and deep knowledge; plus, deadlines and early mornings are always easier when there is someone sitting next to you. *Carlos Aliaga*, who was working two days after major surgery, and not only for his work but also for the good times in the lab. *Paz Hernando*, for her invaluable help in some of the projects, and for her constant smile, enthusiasm and positive attitude. And *Elisa Amorós*, for being so good at what she does.

The students I supervised, *Lara Presa*, *Adrián Corrales*, *Luis García* and *Sara Álvarez*, for the trust; I hope you have learned from me, I definitely learned from

you.

The people from the Camera Culture group, for accepting me as one more, and for being a source of inspiration.

Again, specially: *Chris Barsi*, for a couple of enlightening conversations; *Andreas Velten*, amazing at explaining things; *Di Wu*, for how nice, enthusiastic, willing to learn and contribute, and hard-working she is; and *Taya Leary*, for all the good times together, and for sharing her vision of life.

Rafa, for making my first incursion in the world of serious teaching much easier and much more entertaining.

Óscar, because you already deserved an acknowledgement for the PFC and it ended up not being there, at least in black on white, and because you did not stop teaching me stuff back then and you still haven't.

Matt, for being a source of balance, good conversation, and fun; for the *quieres un poco de té*, the skating, the way back to Central, the beers at the Miracle, and on could the list go.

L'uboš and *Ayush*, for the pier, and the invaluable conversations.

Kshitij, for the explanations, the conversation, the trust, and the fun; and *Amna*, *Rahul*, *Dhairya*, *Prashant*, and *Harshit* for a bunch of good moments, you are an amazing lot.

Carmelo, *Clara*, *Héctor*, *María*, *Marta*, *Noelia*, *Rosa*, *Samuel*, *et al.*, for making both undergrad and grad school a great time; for understanding all my "nos" when I had to work, and for separating me from work so many other times, for the nights out, the weekend trips, the Pirineos Sur, the random beers, the coffees... for all the good times.

Sara, for being like a part of me outside myself; always.

and my family, for whom I do not have enough words; for the love, the support, and above all, for the education they have given me.

In terms of funding, I have to thank the following, for their generous support, which made this work possible:

- Ministerio de Educación (beca de Formación de Profesorado Universitario)
- NVIDIA's Graduate Fellowship program
- Diputación General de Aragón (beca de doctorado)
- Projects TIN2010-21543, Fundación ARAID OTRI 2011/0180, TIN2007-63025, UZ2007-TEC06, ASI/B7-301/98/679-051, and EU 7th FP 251415.

CONTENTS

i INTRODUCTION & OVERVIEW 1

1 INTRODUCTION 3
- 1.1 Photography and Computational Photography 3
- 1.2 Displays and Computational Displays 5
- 1.3 Introducing Perception in the Pipeline 6
- 1.4 Goal and Overview of this Thesis 7
- 1.5 Contributions and Measurable Results 9
  - 1.5.1 Publications 9
  - 1.5.2 Awards 10
  - 1.5.3 Research Stays and Visits 11
  - 1.5.4 Supervised Students (PFCs) 11
  - 1.5.5 Research Projects 12
  - 1.5.6 Other merits 12

ii CAPTURE AND PROCESSING 15

2 CODED APERTURES FOR DEFOCUS DEBLURRING 17
- 2.1 Introduction 17
- 2.2 Previous Work 18
- 2.3 The Imaging Process 19
- 2.4 Perceptual Quality Metrics 20
- 2.5 Perceptually-Optimized Apertures 22
  - 2.5.1 Optimization 22
- 2.6 Performance of the Apertures in Simulation 25
  - 2.6.1 Influence of the Perceptual Metrics 26
  - 2.6.2 Influence of Noise 27
  - 2.6.3 Comparison with other metrics 27
- 2.7 Performance of the Apertures with Real Data 29
- 2.8 Exploring Coded Apertures for Defocus Deblurring of HDR Images 36
  - 2.8.1 Processing Models 36
  - 2.8.2 Simulation of Processing Models 37
  - 2.8.3 Performance Comparison 39
  - 2.8.4 Validation in Real Scenarios 40
- 2.9 Conclusions and Future Work 45

3 REVERSE TONE MAPPING 47
- 3.1 Introduction 47
- 3.2 Previous Work 48
  - 3.2.1 Reverse tone mapping 48
  - 3.2.2 User studies 50
- 3.3 Experiment One: rTMO Evaluation 50
  - 3.3.1 Stimuli 52
  - 3.3.2 Subjects 53

3.3.3 Procedure 53
3.3.4 Results 53
3.4 Experiment Two: HDR vs. LDR Monitor 56
3.5 Expanding over-exposed content 58
3.5.1 Determining the value of γ 58
3.5.2 Validation 61
3.6 Discussion 63
3.7 Selective Reverse Tone Mapping 63
3.7.1 Using the Zone System for rTM 64
3.7.2 Content-aware rTM 66
3.7.3 Results and Discussion 70
3.8 Conclusions 71

III COMPUTATIONAL DISPLAYS 73

4 A SURVEY ON COMPUTATIONAL DISPLAYS 75
4.1 Introduction 75
4.2 Improving Contrast and Luminance Range 77
4.2.1 Perceptual Considerations 77
4.2.2 Display Architectures 79
4.2.3 HDR Content Generation and Processing 80
4.3 Improving Color Gamut 84
4.3.1 Perceptual Considerations 85
4.3.2 Display Architectures 85
4.3.3 Achieving Faithful Color Reproduction 86
4.4 Improving Spatial Resolution 89
4.4.1 Perceptual Considerations 89
4.4.2 Display Architectures 90
4.4.3 Generation of Content 92
4.5 Improving Temporal Resolution 94
4.5.1 Perceptual Considerations 94
4.5.2 Temporal Upsampling Techniques 95
4.6 Improving Angular Resolution I: Stereoscopic Displays 98
4.6.1 Perceptual Considerations 98
4.6.2 Display Architectures 101
4.6.3 Software Solutions for Improving Depth Reproduction 103
4.7 Improving Angular Resolution II: Automultiscopic Displays 105
4.7.1 Perceptual Considerations 106
4.7.2 Display Architectures 106
4.7.3 Image Synthesis for Automultiscopic Displays 108
4.7.4 Applications 111
4.8 Conclusion and Outreach 112

5 DISPLAY-ADAPTIVE 3D CONTENT REMAPPING 115
5.1 Introduction 115
5.2 Related Work 118
5.3 Display-specific Depth of Field Limitations 119
5.4 Optimization Framework 120
5.5 Implementation Details 123

5.6 Retargeting for Stereoscopic Displays 124
5.7 Results 125
5.8 Comparison to Other Methods 127
5.9 Conclusions and Future Work 133
6 VISUAL COMFORT IN STEREOSCOPIC MOTION 137
6.1 Introduction 137
6.2 Related Work 139
6.3 Methodology 140
6.3.1 Parameter Space 140
6.3.2 Stimuli 141
6.3.3 Procedure 142
6.4 Analysis 143
6.4.1 Discussion 144
6.5 Metric of Visual Comfort 144
6.6 Validation 148
6.7 Applications 150
6.8 Conclusion and Future Work 152

IV INTERACTION 155
7 EVALUATION OF INTERACTION PARADIGMS FOR LIGHT FIELD EDITING 157
7.1 Introduction 157
7.2 Related Work 159
7.3 Interfaces 160
7.4 Experiments 162
7.5 Analysis 164
7.6 Discussion and Conclusions 168
7.7 Future Work 171

V FEMTO-PHOTOGRAPHY AND TRANSIENT IMAGING 173
8 FEMTO-PHOTOGRAPHY: ACQUISITION AND VISUALIZATION 175
8.1 Introduction 175
8.2 Related Work 177
8.3 Capturing Space-Time Planes 179
8.4 Capturing Space-Time Volumes 180
8.5 Depicting Ultrafast Videos in 2D 181
8.6 Time Unwarping 182
8.7 Captured Scenes 184
8.8 Conclusions and Future Work 187
9 RELATIVISTIC RENDERING FOR TRANSIENT IMAGING 189
9.1 Introduction 189
9.2 Related Work 190
9.3 Relativistic Rendering 190
9.3.1 Frames of Reference 190
9.3.2 Relativistic Effects 192
9.3.3 Relativistic Rotation 193
9.4 Implementation 196

9.5 Conclusions and Future Work 196

VI CONCLUSION 199

10 CONCLUSIONS AND FUTURE WORK 201

VII APPENDICES 205

A REVERSE TONE MAPPING: IMAGE STATISTICS 207

B F-TESTS FOR ASSESSING THE APPROPRIATENESS OF ADDING NEW PREDICTORS TO A MODEL 209

C GOODNESS OF FIT IN MULTILINEAR REGRESSIONS 211

D REVERSE TONE MAPPING: RESULTS OF THE OBJECTIVE EVALUATION 213

E DISPLAY ADAPTIVE 3D CONTENT REMAPPING: OBJECTIVE FUNCTION AND ANALYTICAL DERIVATIVES IN THE OPTIMIZATION 219

E.1 Term 1: Optimizing Luminance and Contrast 219

E.2 Term 2: Preserving Perceived Depth 220

F DISPLAY ADAPTIVE 3D CONTENT REMAPPING: A DICHOTOMOUS ZONE OF COMFORT 223

G VISUAL COMFORT IN STEREO MOTION: ADDITIONAL DATA 225

G.1 Slices of the Comfort Function 225

G.2 Comfort Zones 225

G.3 Ratings for the Stimuli 226

H LIGHT FIELD EDITING INTERFACES: INTERFACE IMPLEMENTATION DETAILS 227

I LIGHT FIELD EDITING INTERFACES: INSTRUCTIONS FOR INTERFACE EVALUATION TASKS 229

J LIGHT FIELD EDITING INTERFACES: HYBRID INTERFACE AND ADVANCED TOOLS 231

K LIGHT FIELD EDITING INTERFACES: ADDITIONAL DATA FROM THE ANALYSIS OF INTERACTION PARADIGMS 235

K.1 Error in Depth 235

K.2 Time to Completion 236

K.3 Ratings 237

K.4 Rankings 239

K.5 Workflow in open tasks 239

L LIGHT FIELD EDITING INTERFACES: LIGHT FIELDS USED IN THE EVALUATION 247

BIBLIOGRAPHY 251

LIST OF FIGURES

Figure 1.1 Illustration of a camera obscura 4
Figure 1.2 Our eye as a color or light meter 7
Figure 1.3 Overview of the structure of the thesis 8
Figure 2.1 Camera used in our tests 20
Figure 2.2 Power spectra of different apertures 20
Figure 2.3 Image pattern used in the optimization 23
Figure 2.4 Apertures obtained for the four variations of the evaluation function 25
Figure 2.5 Samples of the image database used for evaluation 26
Figure 2.6 Performance of the chosen binary aperture across a range of noise values 29
Figure 2.7 Performance comparison of different apertures 30
Figure 2.8 Calibrated point spread functions 31
Figure 2.9 Recovered images for different apertures 33
Figure 2.10 Defocused and recovered images obtained with the chosen aperture 34
Figure 2.11 Correlation between real and simulated results 35
Figure 2.12 Comparison of real recovered images 35
Figure 2.13 Processing models for deblurring of HDR images 38
Figure 2.14 Performance of the different models and priors according to HDR-VDP-2 (simulation) 39
Figure 2.15 Recovered images with different priors for the *one-shot* model (simulation) 40
Figure 2.16 Recovered images with different priors for the *HDR* model (simulation) 41
Figure 2.17 Calibrated PSFs for the aperture by Zhou and Nayar 41
Figure 2.18 Performance of the different models and priors according to HDR-VDP-2 for scene one (real capture) 42
Figure 2.19 Comparison of results for two selected processing models for scene one 43
Figure 2.20 Performance of the different models and priors according to HDR-VDP-2 for scene two (real capture) 43
Figure 2.21 Comparison of results for two selected processing models for scene two 44
Figure 2.22 Recovered images with different priors 45
Figure 2.23 Effect of noise in the recovered image 45
Figure 3.1 The reverse tone mapping problem 51
Figure 3.2 Representative samples of the stimuli used in our tests 52
Figure 3.3 Bracketed sequence for two stimuli 52
Figure 3.4 Ratings for the bright and dark series 55
Figure 3.5 Distribution of outlier indices for all four rTMOs 56

Figure 3.6 Ratings on an HDR monitor vs ratings on an LDR monitor 57
Figure 3.7 Predictive accuracy of the regression shown in Equation 15 61
Figure 3.8 Predictive accuracy of robust regression vs. OLS 61
Figure 3.9 Comparison of several rTMOs with an objective metric 62
Figure 3.10 Images containing large saturated areas 64
Figure 3.11 Division of luminance in zones according to Ansel Adams' System 65
Figure 3.12 Luminance decomposition for zone-based reverse tone mapping 65
Figure 3.13 Zone-based reverse tone mapping 66
Figure 3.14 Saliency detection with different methods 68
Figure 3.15 Complete pipeline of our selective rTM approach 70
Figure 3.16 rTM using zone-based expansion functions 71
Figure 4.1 Early implementation of the concept of automultiscopic displays 76
Figure 4.2 Overview of the field of computational displays 78
Figure 4.3 Low dynamic range depictions of a high dynamic range scene 79
Figure 4.4 Luminance levels for scotopic, mesopic and photopic vision 79
Figure 4.5 Dual modulation HDR displays 81
Figure 4.6 Superimposing dynamic range for medical applications 82
Figure 4.7 Examples of tone mapping 82
Figure 4.8 Afterimage simulation of a traffic light 84
Figure 4.9 Binocular tone mapping 84
Figure 4.10 Color gamut expansion methods 86
Figure 4.11 Color appearance of a high dynamic range image 87
Figure 4.12 Pipeline for tone reproduction algorithms and color appearance models 88
Figure 4.13 Color reproduction taking into account display and viewing conditions 88
Figure 4.14 Spatial resolution enhancement by temporal superposition in a wobbling display 90
Figure 4.15 Spatial resolution enhancement by optical pixel sharing 91
Figure 4.16 Spatial resolution enhancement by temporal superposition in a conventional display 93
Figure 4.17 Simulation of hold-type blur 95
Figure 4.18 An example of temporal upsampling 97
Figure 4.19 Sensitivity of the HVS to nine different depth cues as a function of distance to the observer 99
Figure 4.20 Accommodation-vergence conflict in stereoscopic displays 100
Figure 4.21 Zone of comfort as a function of disparity and velocity of motion 100
Figure 4.22 Perceived disparity as predicted by a luminance-contrast aware disparity metric 101
Figure 4.23 An example of 3D+2D TV 102

Figure 4.24 Microstereopsis and backward compatible stereo 105
Figure 4.25 Two examples of volumetric displays 106
Figure 4.26 Tensor displays: prototype and computed layered patterns 108
Figure 4.27 Progressive reconstruction of a light field by adaptive image synthesis 109
Figure 4.28 Tailored displays for enhanced visual acuity 111
Figure 5.1 Our 3D content retargeting for a glasses-free lenticular display 116
Figure 5.2 Simulated views of a scene for three different displays 116
Figure 5.3 Depth of field for different display architectures and target displays 120
Figure 5.4 Sensitivity values and thresholds for contrast and disparity 122
Figure 5.5 Multiscale weights for a sample scene 123
Figure 5.6 Magnitude notation and convention definition 124
Figure 5.7 Dichotomous and non-dichotomous zones of comfort for different devices 125
Figure 5.8 Results on a commercial lenticular display 126
Figure 5.9 Our 3D content retargeting for a multilayer light field display 127
Figure 5.10 Simulated results on a multilayer light field display 128
Figure 5.11 Results of our retargeting algorithm on complex scenes 129
Figure 5.12 Sample non-central views of retargeted light fields 130
Figure 5.13 Result illustrating a limitation of the algorithm 130
Figure 5.14 Retargeting results for stereo content 131
Figure 5.15 Comparison against other methods (I) 134
Figure 5.16 Comparison against other methods (II) 135
Figure 6.1 Comfort zone computed from our measurements and comfort metric example 138
Figure 6.2 Defintions and notation: disparity and velocities 141
Figure 6.3 Sample stimulus for our experiment on visual comfort 142
Figure 6.4 Different combinations of disparity and motion in depth for the stimuli 142
Figure 6.5 Slices of our measurement function C 145
Figure 6.6 Projections of our measurement function C 146
Figure 6.7 Representative frames and their computed comfort map M_p 148
Figure 6.8 Stimuli for the validation of the effect of luminance contrast spatial frequency 148
Figure 6.9 Stimuli for the validation of the effect of saliency 149
Figure 6.10 Stimuli for the validation on real footage 150
Figure 6.11 Results from the metric validation study 151
Figure 6.12 Discomfort distribution for an example video 153
Figure 7.1 Example results of two novice subjects editing light fields using our two interaction paradigms 158
Figure 7.2 User interfaces used in our tests 160
Figure 7.3 Workflow when drawing a stroke in each paradigm 161

Figure 7.4 Target images given to users and sample results of user edits 163
Figure 7.5 Analysis of error per interface per task 165
Figure 7.6 Analysis of mean time to completion per interface per task 166
Figure 7.7 Analysis of rankings from final questionnaire (I) 168
Figure 7.8 Analysis of ratings from final questionnaire (I) 169
Figure 7.9 Analysis of rankings and ratings from final questionnaire (II) 169
Figure 7.10 Distribution of times for Tasks 6 and 7 170
Figure 8.1 What does the world look like at the speed of light? 176
Figure 8.2 Our setup for capturing a 1D space-time photo 178
Figure 8.3 Our ultrafast imaging setup 179
Figure 8.4 Performance validation of our system 181
Figure 8.5 Going from 2D to 3D 181
Figure 8.6 Three visualization methods for a captured scene 183
Figure 8.7 Understanding reversal of events in captured videos 183
Figure 8.8 Time unwarping in 1D for a streak image 184
Figure 8.9 Time unwarping for the *bottle* scene 185
Figure 8.10 More scenes captured with our setup 185
Figure 9.1 Time unwarping between camera time and world time for synthesized new views of a scene 191
Figure 9.2 Relativistic effects shown separately for the *cube* scene 194
Figure 9.3 Results for the *cube* scene including all effects 195
Figure 9.4 Results for the *bunny* scene including all effects 195
Figure 9.5 Relativistic rotation 195
Figure F.1 Dichotomous and non-dichotomous zones of comfort for different devices 223
Figure G.1 Slices of our comfort function 225
Figure G.2 Comfort zones derived from our comfort function 225
Figure G.3 Ratings for the stimuli in our experiments 226
Figure H.1 Screenshot of the interfaces based on the focus paradigm 228
Figure I.1 Target images given to users in *directed* tasks. 230
Figure I.2 Sample images given to users in *open* tasks. 230
Figure J.1 Screen-shot of our hybrid interface 231
Figure J.2 Sample views of an edited *San Miguel* light field. 232
Figure J.3 Sample views of an edited *Vase* light field. 233
Figure J.4 Sample views of another edited *Vase* light field 234
Figure K.1 Confidence intervals at 95% for mean difference of error in depth between interfaces for Tasks 1 to 5. 236
Figure K.2 Confidence intervals at 95% for mean difference in time to completion between interfaces for Tasks 1 to 5. 238
Figure K.3 Mean ratings for each interface for questions on general aspects asked in final questionnaire. 238
Figure K.4 Rankings for each interface for questions on general aspects asked in final questionnaire. 239
Figure K.5 Workflow for Task 7, subjects 1-8. 244

Figure K.6 Workflow for Task 7, subjects 9-16. 245
Figure K.7 Workflow for Task 7, subjects 17-20. 246
Figure L.1 Sample views of the *San Miguel* light field. 247
Figure L.2 Sample views of the *Vase* light field. 248
Figure L.3 Sample views of the *Head* light field. 249
Figure L.4 Sample views of the *Teapots* light field. 250

LIST OF TABLES

Table 2.1 Performance evaluation of binary apertures 28
Table 2.2 Performance evaluation of non-binary apertures 32
Table 3.1 Results of the Wilcoxon rank sum tests 54
Table 3.2 γ values for the stimuli 59
Table 7.1 Results of the repeated measures ANOVA for the interface factor for the error in depth 166
Table A.1 Statistics for the images in our dataset 208
Table K.1 Significance of pairwise comparisons for error in depth in directed tasks. 235
Table K.2 ANOVA results for time to completion in directed tasks. 236
Table K.3 Significance of pairwise comparisons for time to completion in directed tasks. 237
Table K.4 ANOVA results for ratings in final questionnaire (I). 238
Table K.5 ANOVA results for ratings in final questionnaire (and II). 239
Table K.6 Significance of pairwise comparisons for ratings in final questionnaire. 241
Table K.7 Kruskal-Wallis results for rankings in final questionnaire (I). 241
Table K.8 Kruskal-Wallis results for rankings in final questionnaire (and II). 242
Table K.9 Significance of pairwise comparisons for rankings in final questionnaire. 242
Table K.10 Rank products per interface for rank scores on final questionnaire. 243
Table K.11 Switching between interfaces in Task 7. 243

Part I

INTRODUCTION & OVERVIEW

1 INTRODUCTION

"If I have seen further it is by standing on the shoulders of giants."
Sir Isaac Newton, 1642–1727.

This thesis presents contributions on the different stages of the imaging pipeline, from capture to display, and including interaction as well; we embrace all of them under the concept of Computational Imaging. The addressed topics are diverse, but the driving force and common thread has been the conviction that a *combination* of improved optics and hardware (*optics*), computation and signal processing (*computation*), and insights from how the human visual system works (*perception*) are needed for –and will lead to– significant advances in the imaging pipeline. This chapter tries to provide a brief overview of the relevant areas and the contributions of this thesis.

1.1 PHOTOGRAPHY AND COMPUTATIONAL PHOTOGRAPHY

Ibn Al-Haytham (also known as Alhazen), Muslim astronomer and mathematician, was the first to make an early analysis of a *camera obscura*, the predecessor of modern photographic cameras [460]. He demonstrated that, through an aperture small enough, the image of an outdoors scene could be reproduced inside a dark room, as depicted conceptually in Figure 1.1[1]. The addition of a lens to the aperture to increase light throughput, and the use of a photosensitive material to record the image of the scene being formed inside the camera obscura essentially gave birth to the photographic camera. The first permanent photograph dates back to 1826, taken with materials that required hours of exposure. After that, the first commercial cameras appeared in 1839, based on the photographic process called daguerrotype, after Louis Daguerre. Today, more than 150 years later, and with digital photography, the concept behind a photographic camera is hardly different from that of those early cameras. Perhaps most surprising is the fact that digital cameras, with their processing power and potential, have, since their birth, been trying to mimic their analog counterparts, let aside that there is no longer need for development. Of course there is a great amount of processing typically done in the camera: A/D conversion, demosaicing, denoising, color space conversion, white balancing, color enhancement, gamma encoding and final quantization before saving the image as a jpeg file [151]. But still, the idea of an image being formed by rays of light converging through a lens and impinging on a photosensitive material to form the image persists.

Computational photography emerged as a field trying to change this. Light traveling through a scene can be seen as a multidimensional function, as noted by Adel-

1 The idea of the camera obscura, and the fact that light would travel through small apertures and illuminate the other side, was already known in ancient China and by the Greek philosophers such as Aristotle, or Euclid, who used it as a proof that light travels in straight lines.



Figure 1.1: Illustration of camera obscura from *Sketchbook on military art, including geometry, fortifications, artillery, mechanics, and pyrotechnics*. 17th century. Library of Congress Online Catalog.

son and Bergen [6]. They define the plenoptic function as a complete representation of the visual world, and this is formalized as light being a function of position (x, y, z), direction (θ, ϕ), wavelength λ, and time t: $L = L(x, y, z, \theta, \phi, \lambda, t)$.

A conventional photograph samples just two dimensions of the plenoptic function (x, y), integrating over a certain range of all the other dimensions: The shutter is responsible for the integration over t, the lens and aperture determine the angular integration over (θ, ϕ), and the sensor and color filter array are responsible for the sampling and integration in λ. An enormous amount of information from the scene is lost when taking a conventional photograph. The reason for this is probably that cameras have aimed at mimicking what can be captured by our own eyes, which essentially works the same way as a camera does. Computational photography, by combining optics, specialized hardware, and computation, has shown us that this need not be the case, that much more can be done.

Examples of this include cameras that can refocus after an image has been taken by avoiding the integration in the angular dimension [332]; cameras that can capture a sharp (and properly exposed) image from a moving scene by coding the shutter [360]; cameras that can recover depth from a single image [259], or a pair of images [508], and/or correct for defocus blur [506, 305] by coding the aperture; we can now capture images whose spatial resolution is not bounded by sensor resolution [84]; and of course (it was one of the first problems addressed by computational photography) capture images of a larger dynamic range than that possible with a 10- or 12-bit sensor [289, 98]. Recently, we have demonstrated that, by using computational photography techniques, we can even capture light at a temporal resolution of picoseconds, effectively allowing us to see light *as it propagates* through a macroscopic scene [459] (part of this work is described in

Part V of this thesis). Such is the power of this computation and optics combination.

Some of the aforementioned techniques have even, despite the youth of the field, made it to the consumer market and are present in most new cameras, such as the bracketing function for HDR creation. Others have given rise to new commercial cameras with enhanced capabilites over a traditional digital camera: This is the case of *light field cameras* such as those sold by Lytro™ or Raytrix™.

Thus, as we explain in [151], "when we speak about computational photography, we commonly refer to how its goal is the enhancement of the abilities of conventional digital (or analog) photography". This is done by sampling the plenoptic function along more dimensions in non-trivial ways, which we refer to as plenoptic imaging (see the work by Wetzstein and colleagues [481] for a comprehensive survey on the topic). Typically, the gist is to somehow code the information arriving to the sensor in ways in which more information from the scene is acquired, that a posterior decoding step allows to recover. Ramesh Raskar, leader of the Camera Culture Group at MIT Media Lab, summarized it by saying: "Photographs will no longer be taken; they will be computed".

1.2 DISPLAYS AND COMPUTATIONAL DISPLAYS

Displays, understood as any means capable of showing a previously captured (or generated) image (or representation) of a scene, can be seen as the next stage of the imaging pipeline after the capture and processing has been done. Somehow, and as was introduced in the seminal work dual-photography [396], they can be seen as the counterparts of cameras, as interchangeable, and thus a lot of concepts applied in computational photography can be taken advantage of in computational displays.

We have argued that cameras have changed relatively little since invention, and that it is in the recent years, with computational photography and plenoptic imaging, when efforts are systematically devoted to capturing richer representations of the scene. With displays, a similar analogy can be made: Computational Displays have aimed at enhancing the content along the different dimensions of the plenoptic function explained in Section 1.1. Again, it has been the combination of hardware and processing that has allowed for this. Chapter 4 of this thesis contains a survey of computational displays (published in [308]); we will thus not extend ourselves here in the topic, and will just provide an illustrative example. Automultiscopic displays are displays capable of showing stereo content to multiple viewers or different viewpoints without the need to wear glasses or other additional equipment. Currently, consoles, desktop monitors, tablets, and even cell phones exist that have an automultiscopic display. All these displays are based on a technology which was presented more than a century ago, and that is parallax barriers, patented by Frederic Ives in 1903 [198], and integral imaging, introduced by Gabriel Lippmann in 1908 [266]. Apart from those, volumetric displays have been developed, but that can only reproduce scenes within the enclosure of the display [130], and holographic imaging has been described and developed (see e.g. [410]), but as of today the cost –together with other issues such as the need for controlled illumination [232]– restrict or prevent its com-

mercial or widespread use. In 2011, a group of researchers presented a new type of displays which could reproduce three-dimensional (3D) images outside the enclosure of the display without the need to wear glasses [482]. These are multi-layer displays with a number of discrete layers in which patterns are shown and rear-illuminated to produce the final scene; these patterns are obtained based on principles from computed tomography, and individually do not resemble the original scene, yet together they succeed at creating a 3D scene with higher spatio-angular resolution than displays based on parallax barriers or integral imaging. Later, directional backlighting was added to these displays [478]. This example shows how non-trivial processing of the data, together with improved hardware, yielded new display architectures with enhanced capabilities.

1.3 INTRODUCING PERCEPTION IN THE PIPELINE

Plato, in his book *The Republic*, includes the well known "Allegory of the cave". In this allegory, Plato, by means of a dialogue between Glaucon and Socrates, presents a situation in which a set of prisoners are kept inside a cave all their lives, facing its wall, with their backs to the entrance, and without the ability to turn their heads around or see anything other than the wall in front of them. Behind the prisoners, a fire is set, and objects are moved between the fire and the prisoners, projecting shadows on the wall that the prisoners see. Those shadows are all that the prisoners have seen of reality, and they thus believe that that is how the world looks like. While the allegory then goes on to describe what happens when one of the prisoners is set free of his chains and sees the world outside the cave, this first part already conveys a key idea, still valid today: the world that we perceive through our senses is, or can be, just a "bad" copy, or a fragment, of the real world, and is often influenced by our previous knowledge and assumptions. Plato would use this idea to defend his rationalist view of epistemology, here, we want to convey the much more modest idea that it is important to take into account the way we perceive the world, the way we perceive images, when designing capture and display devices.

Optical illusions allow us to reveal and better understand the functioning of the human visual system (HVS) and the assumptions that are made by it. In Figure 1.2, left, the right eye (left for the observer) of the girl is seen as cyan, despite it being gray; the reason is *color constancy*, a feature of our color perception system which is essentially what allows us to recognize the same material under different illumination conditions, by "subtracting" the light. Our brain assumes the left half of the picture is illuminated by red light, and subtracts it from the image; as a consequence, the gray eye, once red is subtracted by our visual system, appears as being cyan. Figure 1.2, right, depicts an illusion known as Adelson's checkerboard: Squares A and B are the shade of gray, but square A appears much darker. One of the reasons is *local contrast*[2]. The intensity of its neighboring squares influences our perception of brightness for a particular square.

2 A complete explanation of all the factors involved can be found in `http://persci.mit.edu/gallery/checkershadow`

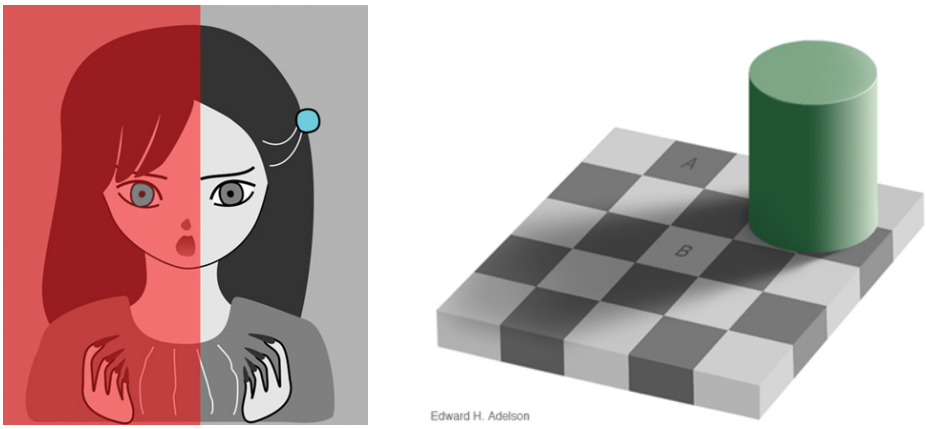


Figure 1.2: Left: Drawing by artist Akiyoshi Kitaoka. The girl's right eye, despite being gray, appears cyan due to color constancy (image adapted from [297]). Image Right: Adelson's checkerboard. Squares A and B are the same shade of gray, yet square A appears darker; one of the reasons is local contrast [5].

These two illusions clearly show that our visual system does not work like an absolute light meter, or a color meter (while still, many image processing algorithms work with absolute intensity values, or RGB values), and illustrate the importance of understanding and taking into account visual perception in the imaging pipeline. Knowing, and exploiting can help overcome limitations of current hardware

1.4 GOAL AND OVERVIEW OF THIS THESIS

The overall goal of this thesis is to improve the imaging pipeline by delving in the idea that a combination of enhanced hardware and optics, computation and processing, *and* knowledge of visual perception can help us overcome inherent limitations of current hardware and algorithms and overall improve the viewing experience. As outlined above, we are not the first to dwell on this idea of combining optics and computation, which has already yielded a number of successful techniques over the last years and the idea of applying perception to the imaging pipeline, is of course not new either (see, e.g., [100]). Still, there is a long way to go, and numerous unsolved problems in imaging that can benefit from this joint and holistic approach. Throughout this thesis we push further in this direction, providing solutions to a number of varied existing problems along the imaging pipeline.

Figure 1.3 provides an overview of the structure of this thesis. The different parts of the thesis correspond roughly to different stages of the imaging pipeline. Within each part, different computational imaging problems are tackled, as described below. This division in parts should not be seen as a "hard" division; due to the inherent interrelations between stages of the pipeline there are certain contributions which could be placed in more than one part. In particular, we have devoted a final part to femto-photography; this project could have been placed under "capture and processing" (Part II). However, because of the context in which it was done (I participated in this project as a consequence of my

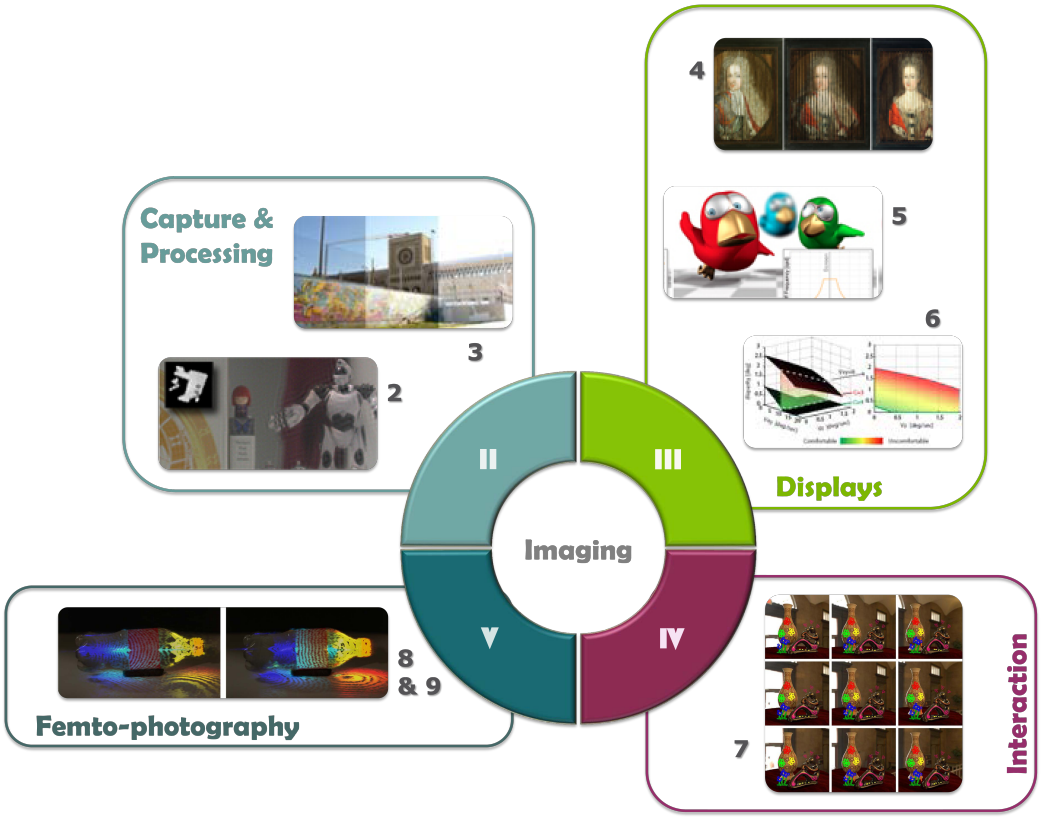


Figure 1.3: Overview of the structure of the thesis

visit to the Camera Culture Group at MIT, inventors of this technique), and because there is no perceptual component to it, we have decided to include it as a separate part (Part V).

- Part II deals with the capture and processing stages. In Chapter 2 we tackle the problem of defocus deblurring of images, with the use of *coded apertures*, and incorporate perceptual metrics to the design of the aperture. The topic of Chapter 3 is *reverse tone mapping*, or how to expand the dynamic range of a conventional image to be shown on a high dynamic range display. We perform a study that allows us to identify a limitation of existing algorithms, and we provide and validate a solution that improves over existing approaches for that limitation. Further, we also propose a more artistic semi-automatic technique for range expansion.

- Part III is devoted to displays. Chapter 4 is a *survey of computational displays*. In the survey we categorize existing displays and display-related algorithms along the dimensions of the plenoptic function, and for each of them we outline the main relevant perceptual aspects, and the work in display architectures and in generation of content or software solutions. In Chapter 5 we address a limitation of automultiscopic displays, namely their limited depth of field. We leverage computational models of perception to propose a *disparity remapping* method that strives to minimize *visible* errors. The final chapter dealing with displays, Chapter 6, investigates the problem of discomfort associated to stereoscopic viewing, and in particular, the

influence of motion in this visual comfort. We perform the most comprehensive measurements of the *influence of stereo motion in comfort* up to date, and also propose a comfort metric derived from this measurements.

- PART IV (Chapter 7) describes a project which belongs to the field of interaction: We study *interaction paradigms for light field editing*, that is, what is the best way to edit a high dimensional representation of a scene that is a light field [261, 143].
- PART V is devoted to *femto-photography*. The ability to capture light at picosecond resolution opens up a whole new world of possibilities. In Chapter 8 we describe the acquisition system and the processing the data undergoes for correct visualization. Visualization of the data is further explored in Chapter 9, in which we deal with the relativistic effects that arise if the camera is moved through the scene.

This thesis and its contributions would not have been possible had I worked all by myself, or only with my supervisor. Multiple projects here presented have involved collaborations with other researchers, and oftentimes, the project is described in its whole to provide context and smooth readability. Thus, at the beginning of each chapter, a section named "About this chapter" contextualizes the work presented in it and describes, when necessary, which parts have been done by myself.

1.5 CONTRIBUTIONS AND MEASURABLE RESULTS

1.5.1 *Publications*

A large part of the work presented in this thesis has already been published (in seven journals indexed in JCR, including three papers in ACM Transactions on Graphics presented at SIGGRAPH or SIGGRAPH Asia, and five peer-reviewed international conferences):

- Coded Apertures for Defocus Deblurring (Chapter 2, Part II):
  - The work on coded apertures for defocus deblurring (Chapter 2) has been published in Computer Graphics Forum [305]. This journal has an impact factor of 1.63, and its position in the JCR index is 15th out of 103 (Q1) in the category Computer Science, Software Engineering (data from 2012).
  - Previous results were published in the international conference SIACG 2011 [304].
  - Follow up related work has been published in the Spanish Conference on Computer Graphics (CEIG) 2012 [135].
- Reverse Tone Mapping (Chapter 3, Part II):
  - The main work on reverse tone mapping was accepted to SIGGRAPH Asia 2009 and published in ACM Transactions on Graphics [302], and in a technical report [301]. This journal has an impact factor of 3.62, and its position in the JCR index is 2nd out of 93 (Q1) in the category Computer Science, Software Engineering (data from 2009).

  - Follow up related work has been published in the Spanish Conference on Computer Graphics (CEIG) 2010 [303].

- Survey on Computational Displays (Chapter 4, Part III):
  - This survey has been accepted for publication in Computers & Graphics [308]. This journal has an impact factor of 1.00, and its position in the JCR index is 40th out of 103 (Q2) in the category Computer Science, Software Engineering (data from 2011).

- Display Adaptive Disparity Remapping (Chapter 5, Part III):
  - This work has been accepted for publication in Computers & Graphics [307]. This journal has an impact factor of 1.00, and its position in the JCR index is 40th out of 103 (Q2) in the category Computer Science, Software Engineering (data from 2011).

- Comfort in Stereoscopic Motion (Chapter 6, Part III):
  - This work has been accepted to SIGGRAPH Asia 2013 and will be published in ACM Transactions on Graphics [112]. This journal has an impact factor of 3.49, and its position in the JCR index is 2nd out of 103 (Q1) in the category Computer Science, Software Engineering (data from 2012).

- Femto-photography and Transient Imaging (Chapters 8 and 9, Part V):
  - The acquisition system and data visualization techniques (Chapter 8) have been accepted to SIGGRAPH 2013 and published in ACM Transactions on Graphics [459]. This journal has an impact factor of 3.49, and its position in the JCR index is 2nd out of 103 (Q1) in the category Computer Science, Software Engineering (data from 2012).
  - Previously, it was accepted as a talk to SIGGRAPH 2012 [458].
  - The relativistic rendering framework (Chapter 9) has been accepted to CEIG 2013 [199].
  - A minor collaboration on a project dealing with analysis of light transport using time-resolved data (included at the beginning of Chapter 8) which has been accepted for publication in the International Journal on Computer Vision (IJCV) 2013 [490]. This journal has an impact factor of 5.35, and its position in the JCR index is 5th out of 92 (Q1) in the category Artificial Intelligence (data from 2010).

1.5.2 *Awards*

We include here a list of awards and fellowships received throughout this thesis. Their generous support allowed the realization of the work here presented:

- FPU grant from the Spanish Ministry of Science and Education (4-year PhD grant)

- NVIDIA Graduate Fellowship Program grant[3] (includes the donation of $25,000 for the development of the awarded project)

3 https://research.nvidia.com/content/nvidia-graduate-fellowship-results-2012

- PhD grant from Diputacion General de Aragon (4-year PhD grant, had to give it up to accept the FPU grant, since they are incompatible)
- NVIDIA Academic Program: Tegra prototype gift (*Mobile Computational Photography: Appearance Capture and Editing*; with A. Jarabo and D. Gutierrez)

Additionally, some of the projects of this thesis were awarded a special recognition:

- Best paper (1 in 2) at CEIG 2013 for the work *Rendering Relativistic Effects in Transient Imaging* (proposed for extension and submission to the journal Computer Graphics Forum; the extension is currently work in progress)
- Best paper (1 in 3) at SIACG 2011 for the work *Coded Apertures for Defocus Deblurring* (proposed for extension and submission to the journal Computer Graphics Forum; the extension got accepted to the journal)

1.5.3 *Research Stays and Visits*

Two research stays were carried out during this PhD; both funded mainly by the Spanish Ministry of Science and Education (tuition and administrative fees were covered by the receiving institution). Based on them, a fruitful collaboration started between the Graphics & Imaging Lab (GIGA) in Universidad de Zaragoza and the Camera Culture group at MIT, which continues today and from which a number of publications have spawned.

- August 2011 – December 2011 (four months): Visiting student at Camera Culture Group, MIT Media Lab. Supervisor: Prof. Dr. Ramesh Raskar.
- March 2013 – June 2013 (three months): Visiting student at Camera Culture Group, MIT Media Lab. Supervisor: Prof. Dr. Ramesh Raskar.

Additionally, a visit of 10 days to Tsinghua University (Beijing, China) took place in November 2012. The collaboration on the project on visual comfort in stereo motion [112] emerged as a result of that visit.

1.5.4 *Supervised Students (PFCs)*

The students supervised throughout the course of this thesis are final year students of the Spanish 5-year engineering degrees. Supervision is during their final degree project, termed *Proyecto Fin de Carrera* or *PFC*.

- In progress: Sara Álvarez. *Low-Cost Recovery of Spectral Power Distributions of Light Sources*. Expected graduation date: December 2013.
- 2011 – 2012: Luis García. *Refocusing High Dynamic Range Images: Coded and Multiple Apertures*. Co-supervised with Lara Presa. Graduated June 2012. Grade: 9.8/10.

- 2011: Lara Presa. *Coded Apertures for Depth Estimation and Image Recovery*. Graduated June 2011. Grade: 9.7/10.
- 2010 – 2011: Adrián Corrales. *Design of Coded Apertures for Defocus Deblurring*. Graduated June 2011. Grade: 9.0/10.

1.5.5 *Research Projects*

The following are research projects in which I have been involved, at different levels, during my Ph.D. studies.

- GOLEM: Realistic Virtual Humans. *European Commission Marie Curie Industry-Academia Program, Seventh Framework. Grant agreement no.: 251415*. PI: Diego Gutierrez.
- MIMESIS: Técnicas de bajo coste para la adquisición de modelos de apariencia de materiales. *Spanish Ministry of Science and Education (TIN2010-21543)*. PI: Diego Gutierrez.
- Aumento del rendimiento gráfico, para sistemas de simulación y visualización en tiempo real, a través de técnicas de antialiasing morfológico. *Fundación ARAID (OTRI 2011/0180)*. PI: Diego Gutierrez.
- TANGIBLE: Humanos realistas e interacción natural y tangible. *Spanish Ministry of Science and Education (TIN2007-63025)*. PI: Francisco J. Seron.
- Fotografía computacional: nuevos algoritmos de procesamiento de imágenes en alto rango dinámico. *Universidad de Zaragoza (UZ2007- TEC06)*. PI: Diego Gutierrez.

1.5.6 *Other merits*

We include here, in somewhat random order, some additional merits:

a. Practical Morphological Anti-Aliasing (MLAA)

   During the development of my PhD I had the chance to participate in another project, not directly related to the topic of my thesis. This project, led by Jorge Jimenez, was a real-time anti-aliasing algorithm [203]. It was published in a peer-reviewed book on GPU techniques, *GPU Pro 2*, highly relevant and well-regarded by professionals in the field. The technique had an enormous impact in the industry and the media:

   - It has been used in the *Torque 3D* engine and in games, e.g. *Rabbids Alive and Kicking*.
   - It has been featured in relevant media such as *Game Developer Magazine*, *Eurogamer* or *Games Industry*.
   - It has drawn the attention of game and hardware companies including Activision, Microsoft, Ubisoft, ZeniMax, Criterion and Intel.

- It has generated a broad interest and discussion over the internet, our project page being linked by over 5400 pages.
- It was featured on the cover of GPU Pro 2.

B. IMPACT OF FEMTO-PHOTOGRAPHY

This project has drawn a large amount of attention from both the research community and the media. The work has appeared in numerous media including the New York Times, the BBC, or MIT News. There is also a TED Talk on the topic by Prof. Ramesh Raskar.

C. INVITED TALKS

The work presented in this thesis has also been presented (partially or as a whole) in a number of invited talks at Tsinghua University (November 2012), Microsoft Research Asia (November 2012), the Max Planck Institut für Informatik (September 2013), and REVES-INRIA Sophia Antipolis (September 2013).

D. PROFESSIONAL SERVICE

I have been given the chance to give something back to the research community by reviewing and serving on the program committee of several journals and conferences. Over the years I have reviewed papers for ACM SIGGRAPH, ACM SIGGRAPH Asia, Eurographics, Pacific Graphics, IEEE Transactions on Image Processing, Pattern Recognition Letters, Computers & Graphics, IEEE Computer Graphics & Applications, APGV and SIACG; and I have served on the committee of the International Conference in Central Europe on Graphics, Visualization and Vision (WSCG), the International Conference on Computer Graphics and Visualization, and will serve next year for the Spring Conference on Computer Graphics. Finally, I was on the local organizing committee for the Eurographics Symposium on Rendering (EGSR) 2013, held in Zaragoza and hosted by our group.

Part II

CAPTURE AND PROCESSING

The first half of this part is devoted to coded apertures for defocus deblurring; in particular the main contribution lies in the introduction of perceptual metrics in the design of the aperture. The topic of the second half is reverse tone mapping, or how to expand the dynamic range of a conventional image to be shown on a high dynamic range display; the main contribution here is that we perform a study that allows us to identify a limitation of existing algorithms, and we provide a new solution for it that improves over existing approaches.

2

CODED APERTURES FOR DEFOCUS DEBLURRING

ABOUT THIS CHAPTER

The work here presented has been published in three papers: an initial paper was presented at an international conference, the Iberoamerican Symposium on Computer Graphics (SIACG) 2011, and selected as one of the three papers invited to submit an extended version to the journal Computer Graphics Forum. The paper was subsequently extended and published in the journal. This first part explores the use of perceptual metrics in the design of coded apertures for defocus deblurring. While I led the line of work (under the supervision of Diego Gutierrez), Lara Presa and Adrián Corrales participated in the work, helping mainly in capturing the databases, and generating the results. Lara Presa further helped with the analysis. The second part was a follow-up of this work, carried out as the final degree project of Luis García, co-supervised by Lara Presa and myself. This part deals with the use of coded apertures for deblurring of high dynamic range images; and was published in the Spanish Conference in Computer Graphics (CEIG) 2012.

B. Masia, A. Corrales, L. Presa and D. Gutierrez.
Coded Apertures for Defocus Deblurring.
In Proc. of SIACG 2011.

B. Masia, L. Presa, A. Corrales and D. Gutierrez.
Perceptually-Optimized Coded Apertures for Defocus Deblurring.
Computer Graphics Forum 2012.

L. Garcia, L. Presa, D. Gutierrez and B. Masia.
Analysis of Coded Apertures for Defocus Deblurring of HDR Images. In Proc. of CEIG 2012.

2.1 INTRODUCTION

In the past few years, the field of computational photography has yielded spectacular advances in the imaging process. One strategy is to code the light information in novel ways before it reaches the sensor, in order to decode it later and obtain an enhanced or extended representation of the scene being captured. This can be accomplished for instance by using structured lighting, new optical devices or modulated apertures or shutters. Here we focus on *coded apertures*. These are masks obtained by means of computational algorithms which, placed at the camera lens, encode the defocus blur in order to better preserve high frequencies in the original image. They can be seen as an array of multiple ideal pinhole apertures (with infinite depth and no chromatic aberration), whose location on

the 2D mask is determined computationally. Decoding the overlap of all pinhole images yields the final image.

Some existing works interpret the resulting coded blur attempting to recover *depth from defocus*. Given the nature of the blur as explained by simple geometrical optics, this approach imposes a multi-layered representation of the scene being depicted. While there is plenty of interesting on-going research in that direction, in this work we limit ourselves to the problem of *defocus deblurring*: we aim to obtain good coded apertures that allow us to recover a sharp image from its blurred original version. We follow standard approaches and pose the imaging process as a convolution between the original scene being captured and the blur kernel (plus a noise function). In principle, this would lead to a blind deconvolution problem, given that the such blur kernel is usually not known. Assuming no motion blur nor camera shake, this kernel is reduced to the point spread function of the optical system. Traditional circular apertures, however, have a very poor response in the frequency domain: not only do they lose energy at high frequencies, but they exhibit multiple zero-crossings as well; it is thus impossible to recover information at such frequencies during deconvolution.

Inspired by previous works [506], we rely on the average power spectra of natural images to guide an optimization problem, solved by means of genetic algorithms. Our main contribution is the use of two existing image quality perceptual metrics during the computation of the apertures; this leads to a new evaluation function that minimizes errors in the deconvolved images that are predicted to be perceived by a human observer. Our results show better performance compared to similar approaches that only make use of the L2 metric in the evaluation function. Additionally, we explore the possibility of computing non-binary masks, and find a trade-off between ringing artifacts and sharpness in the deconvolved images. Our work demonstrates a novel example of applying perceptual metrics in different contexts; as these perceptual metrics evolve and become more sophisticated, some existing algorithms may be revisited and benefit from them.

2.2 PREVIOUS WORK

Coded apertures have been traditionally used in astronomy, coding the direction of incoming rays as an alternative to focusing imaging techniques which rely on lenses [192]. Possibly the most popular patterns were the MURA patterns (Modified Uniformly Redundant Array) [144]. In the more recent field of computational photography, Veeraraghavan et al. [455] showed how a 4D light field can be reconstructed from 2D sensor information by means of a coded mask. Placed at the lens, the authors achieve refocusing of images at full resolution, provided the scene being captured contains only Lambertian objects. Nayar and Mitsunaga [330], extended the dynamic range capabilities of an imaging system by placing a mask of spatially varying transmittance next to the sensor, and then mapping the captured information to high dynamic range.

Other works have proposed different coded apertures for defocus deblurring or depth approximation. To restore a blurred image, the apertures are designed to have a broadband frequency response, along with none (or distinguishable)

zero-crossings in the Fourier domain. Hiura and Matsuyama [172] proposed a four-pinhole coded aperture to approximate the depth of the scene, along with a deblurred version of it, although their system required multiple images. Liang et al. [264] use a similar approach, combining tens of images captured with Hadamard-based coded patterns. Levin et al. [259] attempted to achieve all-focus and depth recovery simultaneously, relying on image statistics to design an optimal aperture and for the subsequent deconvolution. Depth recovery is limited to a multi-layered representation of the scene. Last, the idea of encoding the information before it reached the sensor has not only been limited to the spatial domain but also transferred to the temporal domain by applying a coded exposure aimed at motion deblurring [360].

Another approach to recovering both a depth map of the scene and in-focus images was that of Zhou et al. [507], in this case obtaining a pair of coded apertures using both genetic algorithms and gradient descent search. The same year, a framework for evaluating coded apertures was presented, based on the quality of the resulting deblurred image and taking into account natural image statistics [506]. Near-optimal apertures are obtained by means of a genetic algorithm. In this work we extend previous approaches by introducing two existing perceptual metrics in the optimization process leading to an aperture design. Further, we explore the potential benefits of non-binary masks.

2.3 THE IMAGING PROCESS

Image blur due to defocus is caused by the loss of high frequency content when capturing the image. The capture process can be modeled as a convolution between the scene being captured and the point spread function (PSF) of the camera, plus some noise:

$$f = k_d * f_0 + \eta \tag{1}$$

where f_0 represents the real scene being photographed, f is the captured image, k_d is the PSF and η accounts for the noise introduced in the imaging process. Subscript d accounts for the dependency of the PSF with the defocus depth d (distance of the scene to the in-focus plane). Additionally, the PSF varies spatially across the image and depends on the absolute position of the in-focus plane as well. We will assume that the noise follows a Gaussian distribution of zero mean and a standard deviation denoted by σ, $N(0, \sigma^2)$. By means of deconvolution, an approximation $\hat{f}_0$ of the original sharp image can be obtained.

As Figure 2.1 shows, the PSF is also characterized by the pattern and size of the aperture. Since, as mentioned, blur is caused by the loss of information at certain frequencies, the response of an aperture is better analyzed in the frequency domain, where Equation 1 can be written as:

$$F = K_d \cdot F_0 + \zeta \tag{2}$$

Figure 2.2 shows two plots of the power spectra of different apertures: the traditional circular pattern, an optimal aperture from related previous work [506], and three of the perceptually-optimized apertures presented in this paper. Note that the y-axis, showing the square of the amplitude of the response for different frequencies, is log-scale. Circular apertures exhibit zero crossings at several

frequencies, and thus information at those frequencies is lost during the imaging process and cannot be recovered. Optimal apertures for deblurring therefore seek a smooth power spectrum, while keeping the transmitted energy as high as possible.

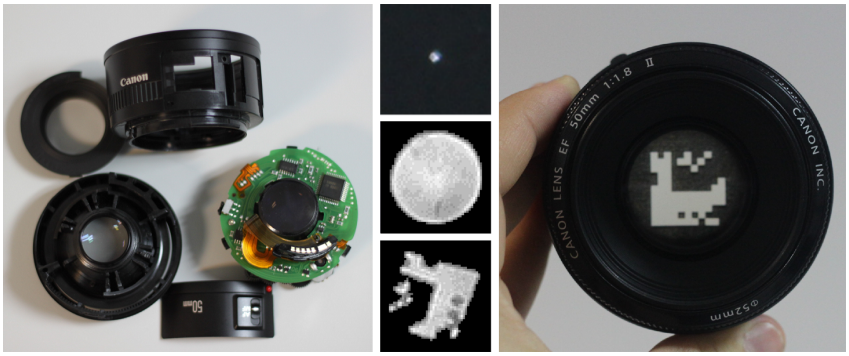


Figure 2.1: *Left:* Disassembled Canon EOS 50mm f/1.8 used in our tests. *Middle:* Point spread function for different apertures and degrees of defocus (*from top to bottom:* circular aperture, focused; circular aperture, defocus depth = 90cm; and one of our coded apertures, defocus depth = 80cm). *Right:* The lens with one of our coded apertures inserted.

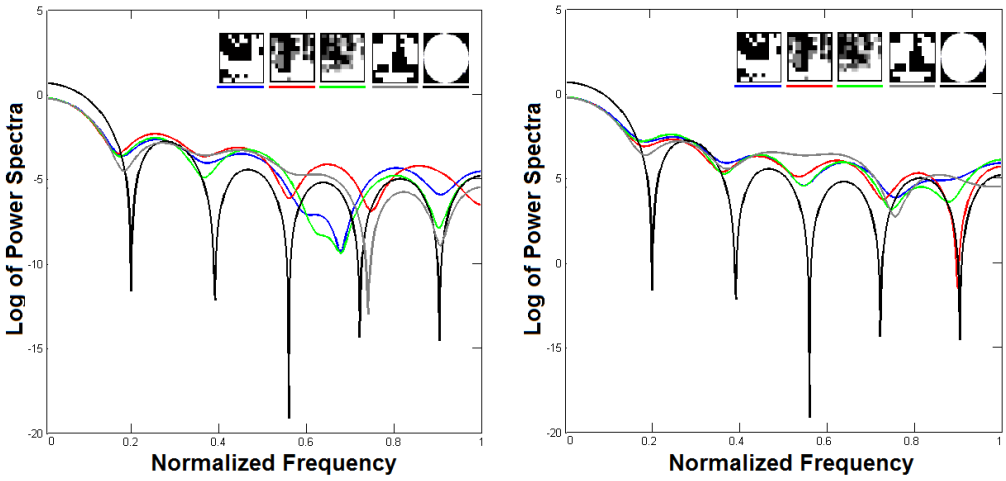


Figure 2.2: Power spectra of different apertures. Spectra for a conventional circular aperture and for an aperture specifically designed for defocus deblurring [506] are shown in black and gray, respectively. Blue, red and green curves show the spectra of some of our perceptually-optimized apertures (please refer to the text for details).

2.4 PERCEPTUAL QUALITY METRICS

Devising an aperture pattern whose frequency response is optimal can be done in different manners. In this paper we build on the approach of Zhou and Nayar [506]; in their work, the authors define their quality metric, i.e. the objective function, as the expectation of the L_2 distance between the deconvolved image $\hat{F}_0$ and the ground truth image F_0 with respect to ζ.

However, objective metrics working at pixel level (such as the L2 norm) are not necessarily correlated with human perception: images with completely different per-pixel information may share a visual quality that will be easily identified by humans [4]. Inspired by this observation, we introduce two additional perceptually-based metrics to guide the design of the apertures, by minimizing errors in the deconvolved images that are predicted to be perceived by a human observer. Furthermore, we include a more reliable prior based on the statistics of a large number of natural images from a recently published database [353]. The perceptual metrics that we use are SSIM (Structural Similarity)[467] and the recent HDR-VDP-2 [294], which we briefly describe in the following subsections.

SSIM: The Structural Similarity Index Measure (SSIM) was introduced by Wang et al.[467], to compute the similarity between two images. It is based on a measure of structural similarity between corresponding local windows in both images. It assumes that the human visual system is very well adapted to extract structural information from a scene, and therefore evaluates the similarity between a distorted image and a reference image based on the degradation of such structural information.

Assuming x and y to be non-negative image signals, belonging to the two images to be compared, SSIM compares luminance $l(\mathbf{x},\mathbf{y})$, contrast $c(\mathbf{x},\mathbf{y})$,and the structure $s(\mathbf{x},\mathbf{y})$ between the images. The latter, $s(\mathbf{x},\mathbf{y})$, is termed structural similarity and defined as the correlation between the two image signals after normalization. The three components are multiplied to obtain the final similarity measure (please refer to the original publication for details):

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + A_1)(2\upsilon_{xy} + A_2)}{(\mu_x^2 + \mu_y^2 + A_1)(\sigma_x^2 + \sigma_y^2 + A_2)} \tag{3}$$

where μ represents mean luminance, and σ is the standard deviation, used as an estimate of the image contrast. υ is the correlation coefficient between the images, obtained as the inner product of the unit vectors $(x-\mu_x)/\sigma_x$ and $(y-\mu_y)/\sigma_y$. In our case, the local window to compute the needed statistics has been set to a 8×8 pixels square window weighted by a rotationally symmetric Gaussian function with a standard deviation $\sigma = 1.5$. The constants A_i avoid instabilities when either $(\mu_x^2+\mu_y^2)$ or $(\sigma_x^2+\sigma_y^2)$ are very close to zero; we set their values to $A_1 = (B_1L)^2$ and $A_2 = (B_2L)^2$ where L is the dynamic range of the pixel values (255 for 8-bit grayscale images), $B_1 = 0.01$, and $B_2 = 0.03$.

HDR-VDP-2: HDR-VDP-2 is a very recent metric that uses a fairly advanced model of human perception to predict both visibility of artifacts and overall quality in images [294]. The visual model used is based on existing experimental data, and accounts for all visible luminance conditions. The results of this metric show a significant improvement over its predecessor, HDR-VDP. This metric makes use of a detailed model of the optical and retinal pathway (including intraocular light scatter, photoreceptor spectral sensitivities and luminance masking) and takes into account contrast sensitivity for a wide range of luminances, as well as inter- and intra-channel contrast masking. We again refer the reader to the original publication for the details.

HDR-VDP-2 can yield different outputs: an estimation of the probability of detecting differences between the two images compared, or an estimation of the quality of the test image with respect to the reference image. In this work we have used the latter, a prediction of the quality degradation with respect to the reference image, expressed as a *mean-opinion-score* (from 0 to 100). We set the *color encoding* parameter of the metric to *luma-display* in order to work with the luminance channel of LDR images; the *pixels-per-degree* parameter, related to the viewing distance and the spatial resolution of the image, is set to a standard value of 30.

2.5 PERCEPTUALLY-OPTIMIZED APERTURES

The Fourier transform of the recovered image $\hat{F_0}$ can be obtained using Wiener deconvolution as follows [506]:

$$\hat{F_0} = \frac{F \cdot \bar{K}}{|K|^2 + |C|^2} \tag{4}$$

where $\bar{K}$ is the complex conjugate of K, and $|K|^2 = K \cdot \bar{K}$. $|C|^2 = C \cdot \bar{C}$ is the matrix of noise-to-signal power ratios (NSR) of the additive noise. We precompute this matrix as $|C|^2 = \sigma^2/S$, where S is the estimated power spectra of a natural image and σ^2 is the noise variance. To estimate S, we rely on recent work on statistics of natural images by Pouli et al. [353], and select from their database 180 images from an extensive collection of two different categories: half of the images belong to the *manmade-outdoors* category, while the other half belongs to the *natural* category. The estimated power spectra is obtained as the average of the power spectra over small windows of each of the 180 images and will be used as our prior in the deconvolution process.

The quality of the recovered image $\hat{f_0}$ with respect to the real image f_0 is measured using a combination of the L2 norm, the *SSIM* index and the *HDR-VDP-2* score (VDP2). The aperture quality metric Q is then given by:

$$Q = \lambda_1(1 - L2) + \lambda_2(SSIM) + \lambda_3(VDP2/100) \tag{5}$$

For the normalized L2 norm, 0 represents perfect quality, while 1 means worst quality. The SSIM index can yield values in the range [-1, 1], but we observe that for the specific case of blurred images the structural information does not change enough for the index to reach negative values. Therefore, values for the SSIM index range from 0 (worst quality) to 1 (best quality). The values for VDP2 range from 0 (worst quality) to 100 (best quality). Last, the vector $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ represents the weights assigned to each metric (discussed in Subsection 2.5.1).

2.5.1 *Optimization*

Our goal is to obtain apertures with the largest possible Q value according to our quality metric. Once we have introduced a way of evaluating a certain aperture with Equation 5, an optimization method can be used to obtain the maximum value of Q over the space of all possible apertures. This space is infinite, limited

only by physical restrictions (i.e. apertures with negative values are not realizable in practice and resolution is limited by the printing process). Resolution is additionally limited by diffraction effects, which appear as the size of the pixels in the aperture gets smaller, and hinder its performance. In our case, we fix the resolution of the apertures to 11×11.

Transmissivity is an additional issue to be taken into account when designing an aperture. Coded apertures typically have lower transmission rates than their circular counterparts, and the use of a longer exposure time to obtain an equivalent brightness to that of the circular aperture can cause other problems such as motion blur. We fix the transmission rate in our apertures to 0.578. We have chosen this value empirically since it yields adequate exposure times, while being similar to other coded apertures proposed for defocused deblurring.

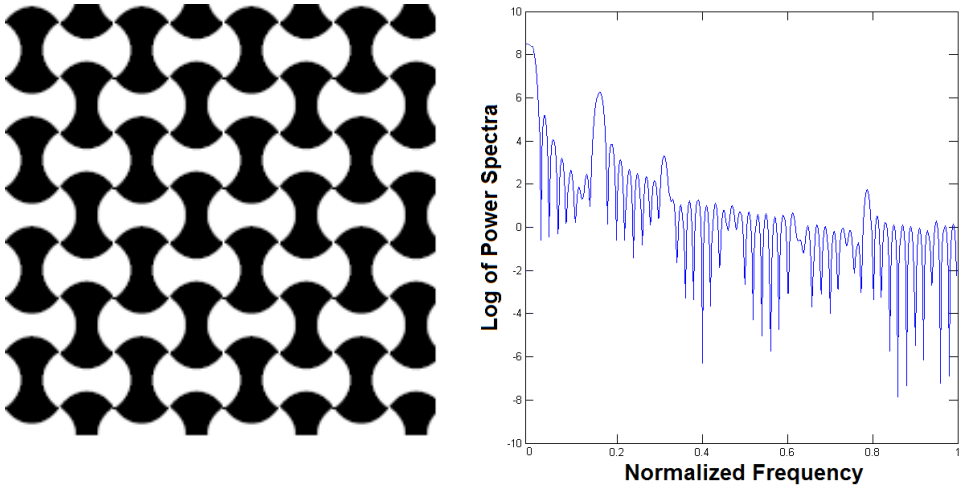


Figure 2.3: *Left:* Image pattern, after [210], used in the evaluation function of the genetic algorithm. *Right:* Wide bandwidth power spectra of the selected pattern.

In order to search for the best aperture pattern we have implemented a genetic algorithm (similar to [506, 304]), which uses our novel quality metric as evaluation function (i.e. objective function). The algorithm has the following scheme:

- *Initialization.* An initial population of $N = 1500$ apertures is randomly generated. An aperture is defined by a vector of $P = 121$ elements, each element corresponding to an aperture pixel.

- *Selection.* We evaluate each aperture by simulating the capture process, multiplying the Fourier transform of a sharp image F_0 by the OTF (response of the aperture in the frequency domain) and adding the Fourier transform of the gaussian noise (Equation 2). We then perform Wiener deconvolution with our prior $|C|^2$ of natural images (Equation 4). The quality of the recovered image is measured using our quality metric Q (Equation 5), and the $M = 150$ apertures with best quality result are selected. The image used to perform this step, which is 200×200 pixels in size, is similar to the pattern used by Joshi et al. [210] (see Figure 2.3), since this pattern has a wide bandwidth spectrum in the frequency domain.

- *Reproduction.* The M selected apertures are used to populate the next generation by means of crossover and mutation. Crossover implies randomly selecting two apertures, duplicating them, and exchanging corresponding bits between them with probability $c_1 = 0.2$, obtaining two new apertures. Mutation ensures diversity by modifying each bit of the aperture with probability $c_2 = 0.05$.

- *Termination.* The reproduction and selection steps are repeated until the termination condition is met. In our case, the algorithm stops when the increment in the quality factor is less than 0.1%, which generally occurs before $G = 80$ generations.

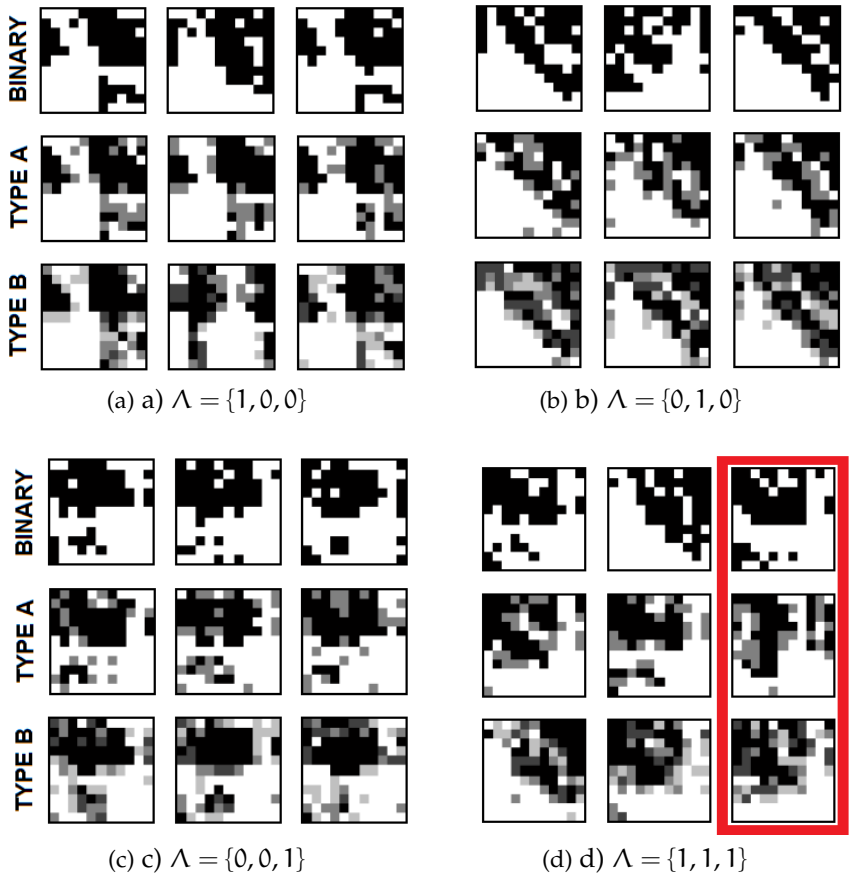
The standard deviation of the noise applied in the *selection* process is set in principle to $\sigma = 0.005$ (we later explore this parameter in Section 2.6.2). This is based on previous findings where apertures designed for σ values of 0.001 and 0.005 proved to work best for a wide variety of images [304]. Following Equation 5, we consider four variations of our evaluation function, characterized by the weight assigned to each metric:

- $\Lambda = \{1, 0, 0\}$: just using the *L2* norm
- $\Lambda = \{0, 1, 0\}$: just *SSIM*
- $\Lambda = \{0, 0, 1\}$: just *HDR-VDP-2*
- $\Lambda = \{1, 1, 1\}$: combining *L2*, *SSIM*, and *HDR-VDP-2*

We have run the genetic algorithm three times for each variation of the evaluation function, yielding three executions to which we will refer as $I = \{1, 2, 3\}$. The top row for each weight vector Λ in Figure 2.4 shows the twelve binary apertures obtained. The other two rows show the results for non-binary apertures, explained next.

NON-BINARY APERTURES Binary codes have the initial advantage of reducing the search space, and are usually preferred in the existing literature. However, there is no principled motivation to restrict the aperture pixel values to either black or white, other than apparent simplicity. A notable exception in this regard is the work by Veeraraghavan and colleagues [455], where the authors report the advantages of continuous-valued apertures, found by gradient descent optimization: reduced computational times and less noise in the recovered (deblurred) images.

In order to analyze if our perceptual metrics also improve the performance of non-binary apertures, we repeat our optimization process, but allowing the solutions of the genetic algorithm to include values between 0 and 1. In order to limit the search space, in practice we restrict the set of possible values to i) one level of gray (the allowed pixel values thus being $\{0, 0.5, 1\}$) and ii) three levels of gray ($\{0, 0.25, 0.5, 0.75, 1\}$). We call the results of both options non-binary type A and non-binary type B, respectively. The middle and bottom rows in Figure 2.4 show the apertures obtained for both types (again, we obtain three different apertures for each weight vector Λ).



(a) a) $\Lambda = \{1, 0, 0\}$ (b) b) $\Lambda = \{0, 1, 0\}$

(c) c) $\Lambda = \{0, 0, 1\}$ (d) d) $\Lambda = \{1, 1, 1\}$

Figure 2.4: Apertures obtained for the four variations of the evaluation function. For each weight vector Λ, the top row shows the results of the binary apertures; while second and third rows show the non-binary type A and non-binary type B results. Columns correspond to the different executions $I = \{1, 2, 3\}$. The apertures which exhibit the best performance (Section 2.6) are highlighted in red.

2.6 PERFORMANCE OF THE APERTURES IN SIMULATION

In this section, we analyze first the performance of binary apertures; then discuss their non-binary counterparts. We simulate the capture process by first convolving a sharp image f_0 with the aperture k_d and adding noise η as described by Equation 1. To recover the deblurred image $\hat{f}_0$, we perform Wiener deconvolution using our prior $|C|^2$ derived from natural images (Equation 4). Note that in practice we work in the frequency domain.

The quality of each recovered image is measured using the L2 norm, the SSIM index and the HDR-VDP-2 score. In order to take in account the results of all three metrics together we calculate the *aggregate* quality factor Q_a as:

$$Q_a = (1 - L2) + (SSIM) + (VDP2/100) \tag{6}$$

where larger values of Q_a correspond to better quality in the recovered images ($Q_a \in [0, 3]$).



Figure 2.5: Some of the images used for evaluating the obtained apertures. Image licensed under Creative Commons copyright of freemages and flickr users (in reading order) Christophe Eyquem, Stig Nygaard, Paola Farrera and Juampe Lopez.

We repeat this process using 30 images of different types of scenes (nature, people, buildings), in order to include a large and varied enough selection. A few examples of the images used are shown in Figure 2.5. For each aperture, we calculate the values for the three different metrics plus the aggregate quality factor Q_a for the 30 recovered images. We therefore have, for each type of aperture (binary, type A or type B) and each weight vector Λ, a total of 90 Q_a values. We denote each of these values as $Q_{a(i,j)}$, where i refers to the execution number ($I = \{1, 2, 3\}$) and j to the image number ($J = [1..30]$)[1]. In the following we analyze separately the influence of the perceptual metrics and the noise level in the performance of the obtained apertures.

2.6.1 *Influence of the Perceptual Metrics*

We compute the aggregate quality factor of the *best binary* aperture obtained for each Λ averaged along the 30 images $Q_{a(i_{best},J)}$ (together with the corresponding standard deviation); we also compute the mean along the 30 test images of the individual scores of the three metrics *L2*, *SSIM* and *HDR-VDP-2*. These serve as an indicative of the performance of a particular aperture. Additionally, we obtain the mean aggregate quality factor of the three executions, $Q_{a(I,J)}$, together with its standard deviation $\sigma(Q_{a(I,J)})$. These values will illustrate the appropriateness of including each of the perceptual metrics in the evaluation function.

1 Note that Q_a values conform a four-dimensional set of data. One dimension corresponds to the type of aperture (binary, type A, or type B), another dimension is the weight vector Λ, and the third and fourth dimensions are the number of executions $i \in I$ and the number of test images $j \in J$.

Table 2.1 compiles these results for binary apertures. The first five columns refer to individual data for the best aperture of the three executions, whereas the last two refer to the averaged values for that particular evaluation function:

$$Q_{a(I,J)} = \frac{1}{|I|} \sum_{i} \left(\frac{1}{|J|} \sum_{j} Q_{a(i,j)} \right), \tag{7}$$

with $|I| = 3$ and $|J| = 30$. It can be seen how the combination of the three metrics ($\Lambda = \{1, 1, 1\}$) yields the highest Q_a scores, which translates into better apertures for defocus deblurring. Although we have limited ourselves in this work to equal weights when combining the three metrics, leaving further exploration of other possibilities for future work, these results clearly suggest the benefits of using perceptual metrics when deriving the apertures.

2.6.2 *Influence of Noise*

The apertures analyzed so far have all been computed assuming an image noise level of $\sigma = 0.005$. We now explore performance of our apertures over a wider range of noise levels, to ensure that our findings generalize to different image conditions. Figure 2.6 shows L2, SSIM, HDR-VDP-2 and Q_a for images captured and deblurred using our best perceptually-optimized binary aperture. The images used are the same 30 test images described before, but after synthetically adding to them noise of increasing standard deviation: σ= 0.0001, 0.0005, 0.001, 0.002, 0.005, 0.008, 0.01 and 0.02. It can be seen how our optimized patterns perform well across all noise levels, in contrast to standard circular apertures which have been proved to be very sensitive to high noise levels [506].

2.6.3 *Comparison with other metrics*

We now compare the performance of our best binary aperture (marked in red in Figure 2.4) with a conventional circular aperture and with the best aperture described by Zhou et al. [506] for a noise level of $\sigma = 0.005$. Note that Zhou's aperture has been optimized using only a L2 norm quality metric.

Figure 2.7 shows the results for both comparisons (top: against a circular aperture; bottom: against Zhou's aperture). We have used each of the three metrics to compare the quality of corresponding recovered images. Each dot in the diagrams represents the values obtained for a given image in the 30-image data set used in this work. Thus, values on the diagonal would indicate equal performance of the two apertures being compared. For the case of the L2 norm, values above the diagonal favor our binary aperture (plotted in the x-axis), whereas for the other two metrics, values below the diagonal are preferred. It is clear from these data that our binary aperture consistently outperforms not only the conventional circular aperture, but Zhou's aperture as well (although obviously by a lesser margin). This translates into recovered images of better quality according to all the metrics, as will be shown in Section 2.7.

We perform the same simulated validation explained above for the non-binary apertures. Our results confirm that again the combination of the three metrics

Binary							
	$L2_{(i_{best},J)}$	$SSIM_{(i_{best},J)}$	$VDP2_{(i_{best},J)}$	$Q_{a(i_{best},J)}$	$\sigma(Q_{a(i_{best},J)})$	$Q_{a(I,J)}$	$\sigma(Q_{a(I,J)})$
$\Lambda = \{1,0,0\}$	1.0893	0.8870	68.6038	2.5622	0.1405	2.5564	0.0042
$\Lambda = \{0,1,0\}$	1.1716	0.8994	71.7686	2.6054	0.1164	2.5479	0.0407
$\Lambda = \{0,0,1\}$	1.0359	0.8883	70.9459	2.5874	0.1190	2.5619	0.0183
$\Lambda = \{1,1,1\}$	1.0261	0.8990	74.4169	**2.6329**	0.1026	**2.5921**	0.0315

Table 2.1: Performance evaluation of binary apertures obtained with the different objective functions (i.e. different weight vector Λ). The first five columns of each table show values of the different metrics and aggregate quality factor for the best binary apertures of each evaluation function averaged across the 30 test images, plus the standard deviation of the latter. The two rightmost columns show, for each evaluation function, the mean aggregate quality factor of the three executions and its standard deviation. Note that the L2 norm is shown as a percentage with respect to the maximum error.

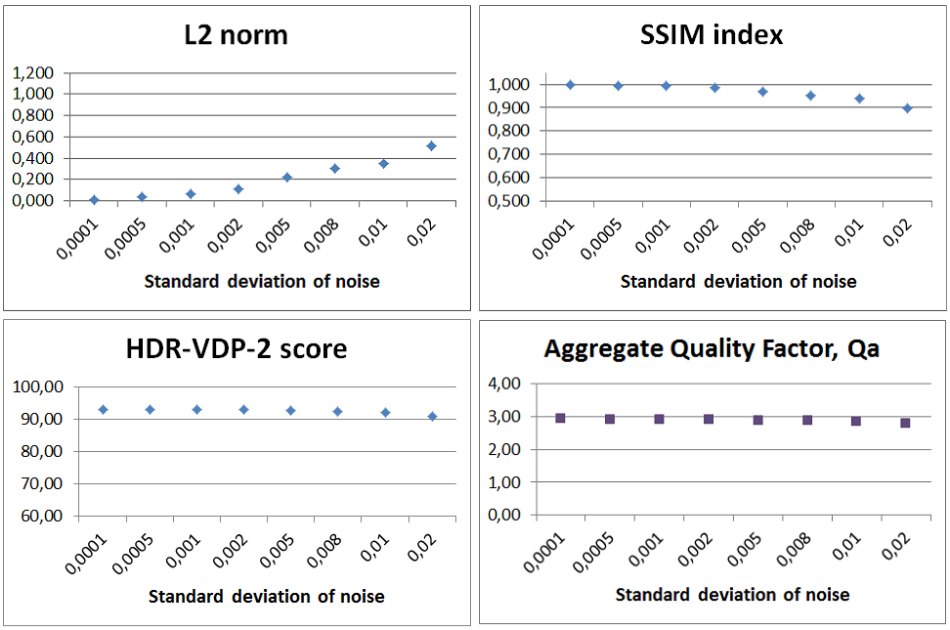


Figure 2.6: Performance of the best perceptually-optimized binary coded aperture across eight different levels of noise, measured with the L2, SSIM, HDR-VDP-2 and Q_a metrics. The L2 norm shows percentages with respect to the maximum error.

with equal weights $\Lambda = \{1, 1, 1\}$ yields apertures with better overall performance. Table 2.2 summarizes the results. In an analogous manner to the analysis for binary apertures, the first five columns show data for the best non-binary aperture in each case, averaged across the 30 test images. The last two columns show averaged values across the 30 images and the three executions computed for each evaluation function.

2.7 PERFORMANCE OF THE APERTURES WITH REAL DATA

While in the previous sections we have evaluated the performance of the apertures by simulating the capture process, in this section we test our apertures on a real scenario; we print and insert the masks into a camera, calibrate the system, and capture real scenes. We have used a Canon EOS 500D with a EF 50mm f/1.8 II lens, shown (disassembled) in Figure 2.1. To calibrate the response of the camera (PSF) at different depths, we used a LED which we made as close as possible to a point light source with the aid of a pierced thick black cardboard. We locked the focus at 1,20 m and took an initial focused image, followed by images of the LED at 20, 40, 60 and 80 cm with respect to the in-focus plane. For each depth, the actual cropped image of the LED served us as PSF, after appropriate thresholding of surrounding values which contain residual light, and subsequent normalization for energy conservation purposes. The resulting PSFs for one of our binary apertures are shown in Figure 2.8, next to the PSFs of a conventional, circular aperture for comparison purposes.

Once calibration has been performed, images of three scenes at the four defocus depths (20, 40, 60 and 80 cm) were taken with each of the selected apertures.

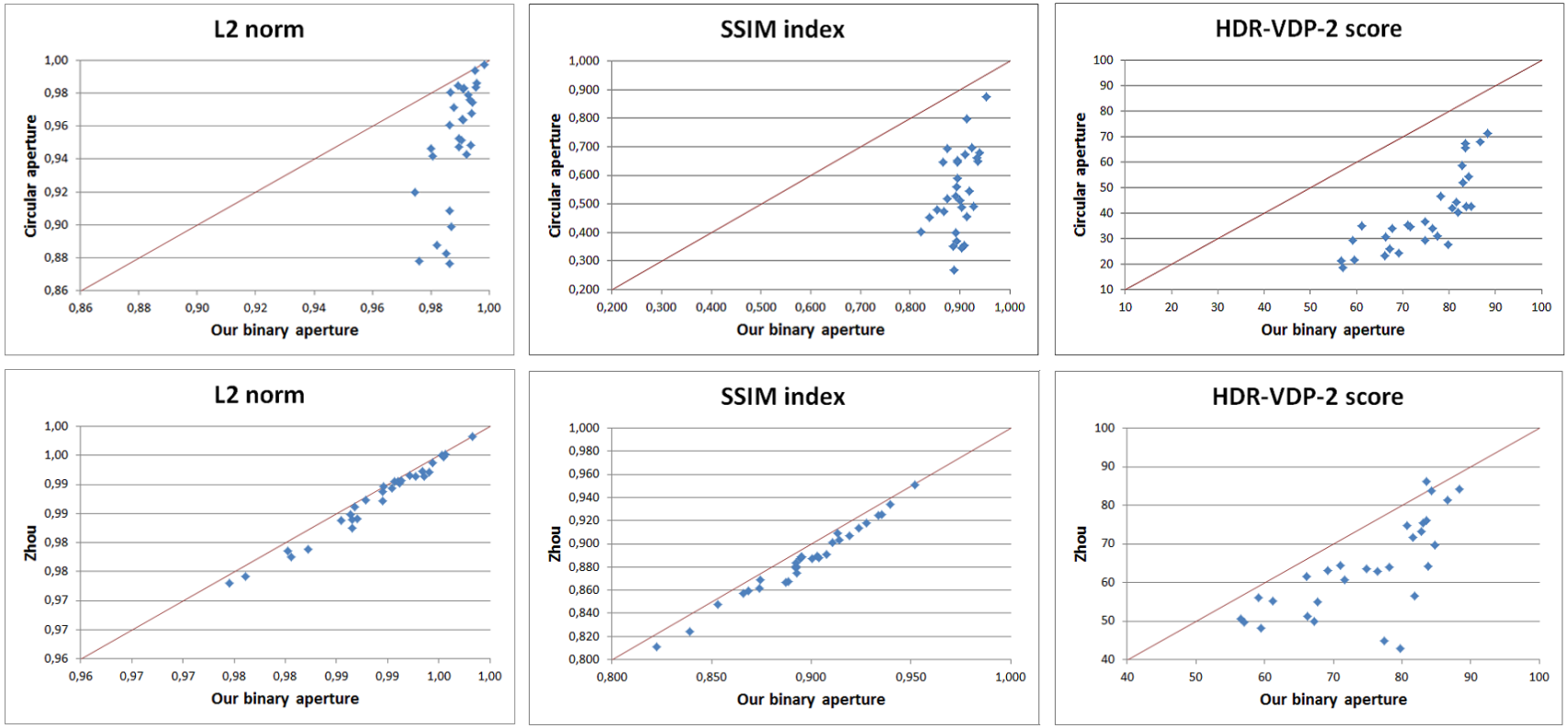


Figure 2.7: Scatter plots showing the performance of our best binary coded aperture against that of a circular aperture (*top row*) and against the coded aperture proposed by Zhou et al. [506] for an image noise of $\sigma = 0.005$ (*bottom row*). For the sake of consistency, the L2 norm is depicted as $(1 - \text{L2}/100)$, L2 being the percentage with respect to the maximum error. It can be seen how our proposed aperture outperforms the other two.

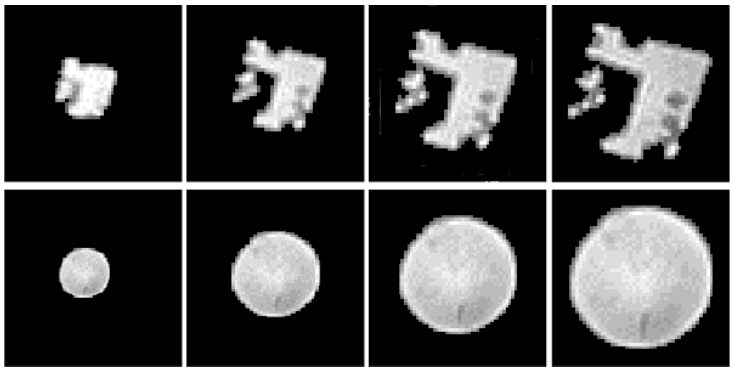


Figure 2.8: PSFs at four different defocus depths (20, 40, 60 and 80 cm). *Top row*: For our binary coded aperture. *Bottom row*: For a circular aperture.

During the capture process, the aperture was set to F2.0, and the exposure time to 1/20 for all scenes and apertures, to ensure a fair comparison. The captured defocused images are then deblurred using the corresponding calibrated PSF by means of Wiener deconvolution. We used Wiener deconvolution with a NSR of 0.005 instead of the prior of natural images, since in real experiments it gave better results. This may be caused by the fact that our prior $|C|^2$ is calculated with the power spectra of images from *manmade day* and *natural day* scenes, which have similar spectral slopes, while the spectral slope for images from *manmade indoors* scenes (similar to the scenes we capture) is slightly different [353]. The same exposure and aperture settings were used for all our coded apertures. Figure 2.9 depicts the recovered images for three different apertures: a circular aperture, our best binary coded aperture and the best aperture obtained by Zhou et al. [506] for a noise value of $\sigma = 0.005$, to which we have also compared in Section 2.6. Defocus depths are 60 cm for recovered images (b), (c) and (d) and 80 cm for (e) and (f). Insets depict the corresponding PSF.

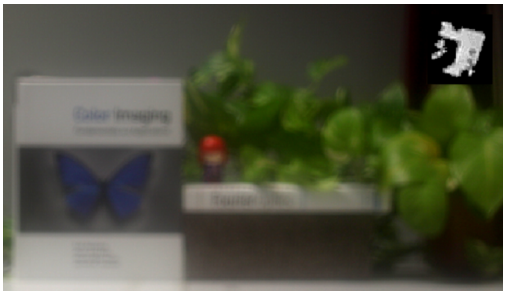
Our aperture clearly outperforms the circular one, which was to be expected from the existing body of literature about coded apertures. More interesting is the comparison with a current state-of-the-art coded aperture; when compared to the aperture described by Zhou et al., our perceptually-optimized approach yields less ringing artifacts, exhibiting, qualitatively, a better overall performance. Additional results for two other scenes at four defocus depths (20, 40, 60 and 80 cm) can be seen in Figure 2.10. Please note that the slight changes in brightness in the images are due to different illumination conditions, and not to the light transmitted by the aperture.

Minor artifacts that appear in our recovered images are probably due to errors in the calibrated PSF. Another possible cause of error may be inaccurately modeled image noise [398]. Additionally, although the PSF actually varies spatially across the image [259], we consider here one single PSF, measured at the center of the image, for the entire image plane.

The non-binary apertures obtained in Section 2.5.1 were also evaluated in a real scenario. Figure 2.12 shows the recovered images obtained with the best binary aperture (left), the best non-binary aperture of type A (middle) and the best non-binary aperture of type B (right). Although non-binary apertures seem to yield images with lower background noise, evidence is not strong enough to derive

Non-binary type A							
	$L2_{(i_{best},J)}$	$SSIM_{(i_{best},J)}$	$VDP2_{(i_{best},J)}$	$Q_{a(i_{best},J)}$	$\sigma(Q_{a(i_{best},J)})$	$Q_{a(I,J)}$	$\sigma(Q_{a(I,J)})$
$\Lambda = \{1,0,0\}$	1.0671	0.8887	69.3253	2.5812	0.1265	2.5645	0.0053
$\Lambda = \{0,1,0\}$	1.1840	0.9012	63.4331	2.5245	0.1435	2.5258	0.0026
$\Lambda = \{0,0,1\}$	1.0707	0.8867	70.0342	2.5763	0.1352	2.5584	0.0127
$\Lambda = \{1,1,1\}$	1.1368	0.8928	72.3652	**2.6050**	0.1186	**2.5963**	0.0069
Non-binary type B							
	$L2_{(i_{best},J)}$	$SSIM_{(i_{best},J)}$	$VDP2_{(i_{best},J)}$	$Q_{a(i_{best},J)}$	$\sigma(Q_{a(i_{best},J)})$	$Q_{a(I,J)}$	$\sigma(Q_{a(I,J)})$
$\Lambda = \{1,0,0\}$	1.1436	0.8880	68.9516	2.5660	0.1460	2.5460	0.0168
$\Lambda = \{0,1,0\}$	1.1979	0.9022	63.1623	2.5218	0.1399	2.5153	0.0054
$\Lambda = \{0,0,1\}$	1.1392	0.8869	69.4941	2.5705	0.1203	2.5460	0.0019
$\Lambda = \{1,1,1\}$	1.1542	0.8998	70.8276	**2.5965**	0.1224	**2.5692**	0.0234

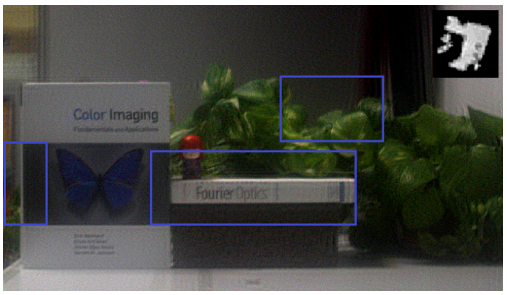
Table 2.2: Performance evaluation of non-binary apertures obtained with the different objective functions (i.e. different weight vector Λ). The first five columns show values of the different metrics and aggregate quality factor for the best non-binary apertures of each evaluation function averaged across the 30 test images, plus the standard deviation of the latter. The two rightmost columns show, for each evaluation function, the mean aggregate quality factor of the three executions and its standard deviation. Note that the L2 norm is shown as a percentage with respect to the maximum error.



(a) a) Defocused image captured with our best binary aperture



(b) b) Result obtained for a circular aperture (d = 60cm)



(c) c) Result obtained for our best binary aperture (d = 60cm)



(d) d) Result for the aperture by Zhou et al. for σ = 0.005 (d = 60cm)



(e) Close-ups of c)



(f) Close-ups of d)



(g) e) Result obtained for our best binary aperture (d = 80cm)



(h) f) Result for the aperture by Zhou et al. for σ = 0.005 (d = 80cm)

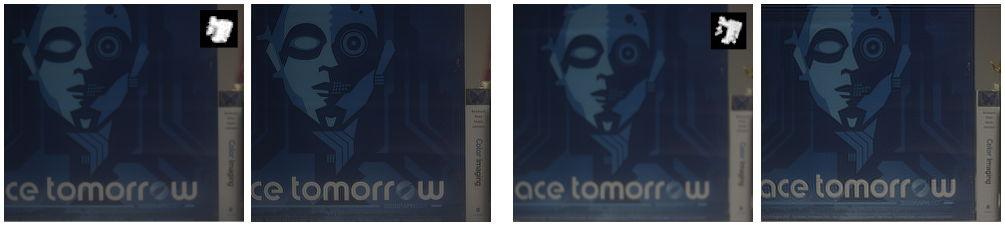


(i) Close-ups of e)



(j) Close-ups of f)

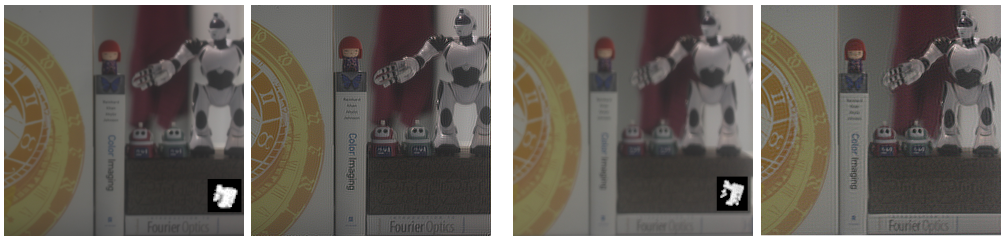
Figure 2.9: Recovered images for different apertures (circular, Zhou's for σ = 0.005 and our best perceptually-optimized binary aperture) and different defocus depths d. Close-ups of this images show the improved quality and fewer ringing artifacts of images recovered with the perceptually-optimized aperture. Insets depict the PSF of the aperture used in each case. Note that results for the circular aperture are significantly brighter because of its higher transmission rate.



(a) a) d = 20cm (b) b) d = 40cm



(c) c) d = 60cm (d) d) d = 80cm



(e) a) d = 20cm (f) b) d = 40cm



(g) c) d = 60cm (h) d) d = 80cm

Figure 2.10: Defocused and recovered images at four different defocus depths d obtained with the perceptually-optimized binary coded aperture for two different scenes.

any definite conclusion. It is worth noting that metrics based on simulations of the capture process yield similar quality values for binary apertures and their non-binary counterparts (see Tables 2.1 and 2.2). This may suggest the need for a more complex image formation model, essentially in what regards to the additive noise, a need which has already been observed by other authors in the field [455].



Figure 2.11: Correlation between real-capture and simulated-capture results. Average quality of the recovered images for both cases (real and simulated) according to each metric for the four defocus depths tested (20, 40, 60 and 80 cm) and to the *aggregate* quality factor Q_a calculated according to Equation 6.

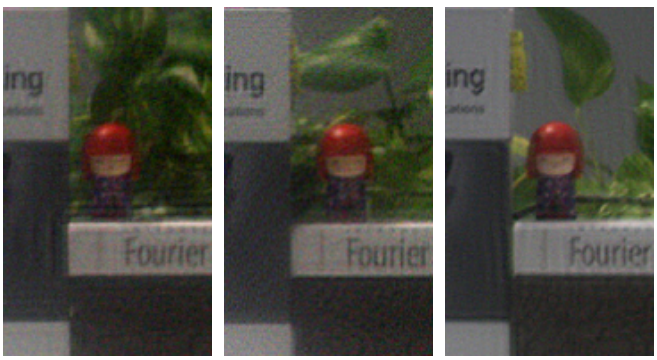


Figure 2.12: Comparison between deblurred images captured using perceptually-optimized binary (*left*), non-binary type A (*middle*), and non-binary type B (*right*) apertures.

Observations from real-world images are consistent with the power spectra shown in Figure 2.2, where our perceptually-optimized apertures exhibit larger amplitudes for the majority of the spectrum compared to Zhou's and the circular aperture. Additionally, in order to assess how well real results correlate with sim-

ulated ones we have compared results from a real setup with results simulated for the same conditions. We have done this for our best binary coded aperture selected in red in Figure 2.4. To do this we compute the size of the blur for the different defocus depths used in the real scenario (20, 40, 60 and 80 cm) and scale the PSF accordingly when computing the simulated blurred images. Althought this scaling is only an approximation to what the real PSF would be, it does give information on how well simulated results extrapolate to real results. Figure 2.11 shows the results obtained by the different quality metrics (plus the aggregate factor Q_a) for real and simulated results. We can clearly see how both exhibit the same behavior and trends, thus showing the validity of the use of simulated capture processes for the evaluation of the different apertures.

Finally, the time until convergence when running the algorithm on an Intel core i7 930 @2.80GHz is 13,72 hours for the evaluation function using ($\Lambda = 1, 1, 1$), which is obviously the most expensive scenario. As expected, computing the HDR-VDP 2 metric consumes the largest amount of time (62% of the total execution time when $\Lambda = 1, 1, 1$), followed by SSIM; there is clearly a trade-off between complexity of the metrics included and performance of the resulting apertures.

2.8 EXPLORING CODED APERTURES FOR DEFOCUS DEBLURRING OF HDR IMAGES

While it was well known that the use of coded apertures for defocus deblurring offers good performance with LDR images [506], to our knowledge it had not been tested in the context of HDR imaging, so we set out to explore how they would perform in this context. For this purpose, we rely on a coded aperture specifically designed for defocus deblurring of LDR images by Zhou et al. [506] and use it to analyze this problem in HDR images. The pattern of this aperture can be seen in Figure 2.2 (gray line) together with its power spectrum compared to that of a circular aperture. Note that this aperture offers a better frequency response for defocus deblurring than the circular aperture, as seen before.

In this section we propose and analyze three different processing models for recovering focused HDR images, one from a single blurred HDR radiance and two from an input of blurred LDR exposures, and analyze them first in a simulation environment and finally in real scenarios. We also analyze the use of deconvolution statistical priors, made both from HDR and from LDR images, taking into account the work of Pouli et al. [353] and following the idea that, to solve HDR problems, the use of HDR priors instead of LDR ones would lead to better results due to the existing statistical differences between both types of images.

2.8.1 *Processing Models*

The capture process of an image can be modeled with a convolution, as explained in Section 2.3. In order to study the viability of the employment of coded apertures for defocus deblurring in HDR images, we simulate the capture process and attempt to recover a sharp image from the simulated blurred image.

Being f_0^{HDR} an HDR scene, we can use the approximation given by Equation 8 to simulate the capture of a High Dynamic Range radiance f^{HDR} only if we are able to capture it in one single shot.

$$f^{HDR} = k * f_0^{HDR} + \eta \tag{8}$$

Some existing cameras allow the capture of extended dynamic range, but in most cases HDR images are obtained by capturing series of LDR exposures and merging them later.

Then, being $f_{0n}^{LDR}, (n = 1, ..., N)$ a set of LDR exposures of the same focused HDR scene f_0^{HDR}, we can simulate the capture of the defocused HDR radiance by first simulating the capture of each exposure following Equation 9, and second merging them into a single HDR defocused radiance as expressed in Equation 10, g being the HDR merging operator.

$$f_n^{LDR} = f_{0n}^{LDR} * k + \eta \tag{9}$$

$$f^{HDR} = g(f_1^{LDR}, f_2^{LDR}, ..., f_N^{LDR}) \tag{10}$$

Once f^{HDR} is obtained, we can recover the focused HDR radiance $\hat{f}_0^{HDR}$ by performing a single deconvolution. However, since we have the LDR defocused exposures, it is possible to deblur them separately with a set of N deconvolutions and merge them later to obtain $\hat{f}_0^{HDR}$, following Equation 11.

$$\hat{f}_0^{HDR} = g(\hat{f}_{01}^{LDR}, \hat{f}_{02}^{LDR}, ..., \hat{f}_{0N}^{LDR}) \tag{11}$$

According to this, we present three different models for recovering focused HDR radiances:

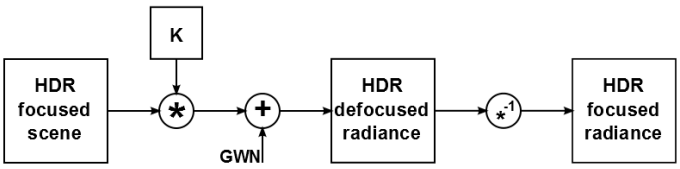
1. **One-shot model:** Processing HDR radiance obtained with a single shot. Equation 8 is used to model the capture process and the focused radiance is recovered with a single deconvolution, as seen in Figure 2.13a.

2. **HDR model:** Processing HDR radiance obtained by merging LDR exposures. Equations 9 and 10 are used and the focused HDR image is recovered with a single deconvolution. The pipeline of this processing is shown in Figure 2.13b.

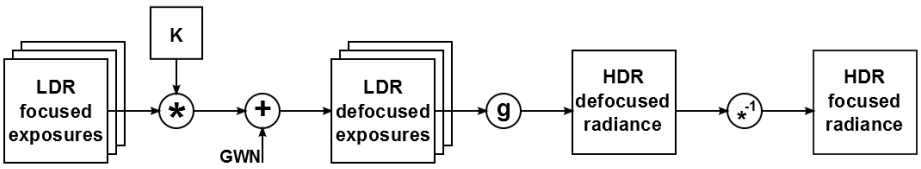
3. **LDR model:** Processing LDR exposures separately before merging. We follow Equation 9 to model the capture process of the N input images, and recover the focused LDR exposures with N deconvolutions, then merging them as in Equation 11 to obtain the HDR focused radiance. This pipeline can be seen in Figure 2.13c.

2.8.2 *Simulation of Processing Models*

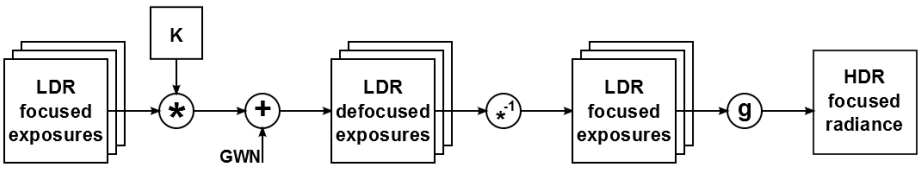
First we analyze these three models by performing simulations in order to study their viability before proceeding to real experiments. To carry them out, we use one of the coded apertures developed by Zhou et al. [506], which is shown in



(a) Pipeline for the one-shot model



(b) Pipeline for the HDR model



(c) Pipeline for the LDR model

Figure 2.13: Pipelines for all different processing models, where **k** is the convolution kernel, *GWN* is Gaussian White Noise, g is the HDR merging operator and $*$ is the convolution operator.

Figure 2.2 (gray line). This aperture is known to work well for defocus deblurring LDR images.

For the simulations we use a set of seven HDR photographs with different dynamic ranges for the first model, and their three corresponding LDR exposures for the other two. The main goal is to recover the focused HDR images with all three processing models. We use the perceptual metric *HDR-VDP2* [294] in order to assess the quality of the results. This metric works on luminance, comparing a reference HDR image with its distorted version, providing quality and visibility (probability of detection) measures based on a calibrated model of the human visual system. In this work we focus in obtaining the quality factor Q, a prediction of the quality degradation of the recovered HDR image with respect to the reference HDR image, expressed as a mean-opinion-score (with values between 0 and 100). This metric can not only work with HDR images, but also with their LDR counterparts.

We test four different noise levels ($\sigma = 0.0005$, 0.001, 0.005 and 0.05), and three different deconvolution variations based on Wiener deconvolution, whose formulation in frequency is given by Equation 4.

From this deconvolution, we study these three different variations:

- **Wiener deconvolution without prior, with a constant NSR matrix.** Replacing $|C|^2$ in Equation 4 by a constant NSR matrix. We tested several values and found that there is a trade-off between noise and ringing in resulting

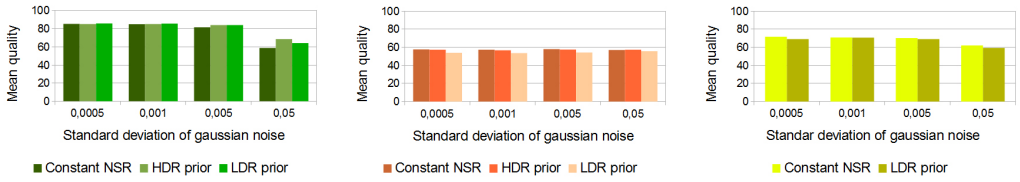
images. We finally decided to set the NSR to 0.005, achieving good balance between both artifacts.

- **Wiener deconvolution using an HDR image prior.** Approximating $|F_0|^2$ in Equation 4 by a statistical prior matrix averaging power spectra of a series of 198 HDR images. We construct the prior employing *manmade* (*day* and *indoors*) HDR images from the database of Tania Pouli[2]).
- **Wiener deconvolution using an LDR image prior.** Replacing $|F_0|^2$ as in the previous, using a prior of 198 *manmade* (*day* and *indoors*) LDR images instead, extracted from the database of Tania Pouli.

We explore the use of HDR priors in the one-shot and HDR models, given that we are deconvolving an HDR radiance, inspired by Pouli et al. [353]. Note that we do not test the LDR model with an HDR prior since we are deconvolving LDR images in it. Since the aperture we are using is optimized for a noise level of $\sigma = 0.005$, we set this value as standard deviation of the Gaussian noise in our deconvolutions with priors.

2.8.3 *Performance Comparison*

Once all the simulations are finished, we compute the mean quality factor Q, given by the *HDR-VDP2* metric, of the seven images obtained with the three proposed processing models shown in Figure 2.13. For each model we analyze four different noise levels and the three different deconvolution variations explained in Section 2.8.2 (except for the LDR model, as explained). This information is collected in Figure 2.14.



(a) One-shot model (b) HDR model (c) LDR model

Figure 2.14: Mean Q obtained with the *HDR-VDP2* metric for each processing model, with all different combinations of noise level and deconvolution prior.

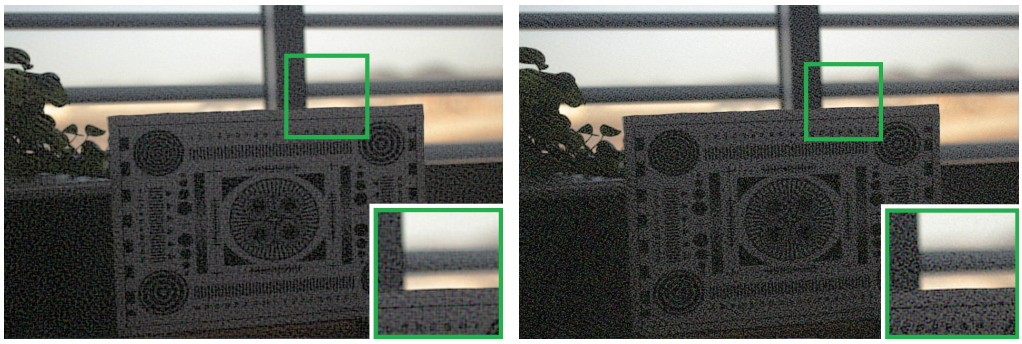
We can see how the use of priors is strongly recommended for the one-shot model when image noise is very high. In this noisy scenario, an HDR prior offers better results than an LDR prior. However, when image noise decreases, all three different deconvolutions produce similar behaviours. As expected, using an HDR prior outperforms using an LDR prior in the HDR model, but we can see how the use of Wiener deconvolution with a constant NSR matrix seems to offer similar or even better quality all along the noise range. For the LDR model, the use of a constant NSR matrix in the deconvolution seems to offer better results than the LDR prior, although differences are not significant.

2 http://taniapouli.co.uk/research/statistics/

With regard to the comparison between all three processing models, we can see how the one-shot model clearly derives in better results than the other two, and would be the ideal method, if the appropriate hardware becomes widely available. Meanwhile, the HDR model seems to perform worst. Note that the merging operation is a non-linear process, and therefore the deconvolution is performed over content which has been non-linearly transformed. Also, the added GWN can be amplified during this process. It must be noted, however, that function g is approximately linear for a wide range of luminances. In the LDR model, three deconvolutions are performed, and it is well-known that deconvolution is a noisy process. However, in HDR images the relative difference between neighbour pixels is bigger than in LDR ones. This increases ringing significantly, and along with the amplified GWN and non-linearity may be what causes the HDR model results to be the worst of all.

In terms of computational cost, the lowest is offered by the one-shot model, as it only requires one deconvolution, while the HDR model requires one deconvolution and one exposure fusion, and the LDR model requires one deconvolution for each exposure and one exposure fusion.

In Figure 2.15 we show the result of one of the noisy simulations ($\sigma = 0.05$) using the one-shot model, with both priors. We can see how the use of an HDR prior slightly reduces the recovered image noise. In Figure 2.16 we show an example of the same HDR scene recovered with the HDR model, with both priors, this time with $\sigma = 0.0005$. In this low noise scenario we can appreciate how the use of an HDR prior instead of an LDR one results in a reduction of ringing artifacts.

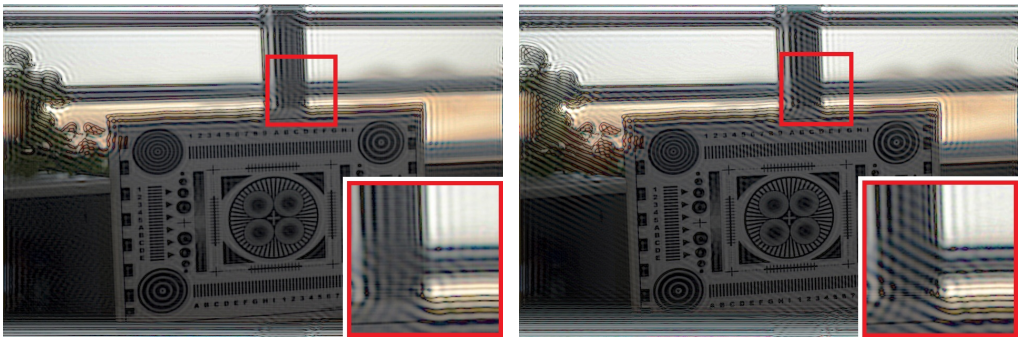


(a) HDR prior (b) LDR prior

Figure 2.15: Comparison between images recovered after simulation of the one-shot model, with HDR and LDR priors and $\sigma = 0.05$. Note how the use of the HDR prior instead of the LDR one slightly reduces image noise.

2.8.4 *Validation in Real Scenarios*

After performing the simulations we proceed to validate the same processes in real scenarios. We cannot validate the one-shot model in real scenes because of the lack of the required equipment: an HDR camera that allows to capture an HDR image with a single shot. For this reason, physical validation is restricted to the HDR and LDR models. For these, the image capture process is analogous



(a) HDR prior (b) LDR prior

Figure 2.16: Comparison between images recovered after simulation of the HDR model, with HDR and LDR priors and $\sigma = 0.0005$. Note how using the HDR prior instead of the LDR one seems to reduce image ringing.

to that described in Section 2.7. We construct a scene with a large luminance range and capture three images using the multi-bracketing camera option set to relative exposures of +2, 0 and -2 stops. For these captures we fix the ISO setting value at 100 and aperture size at F2.0, leading to exposure times of 1/5, 1/20 and 1/80 seconds. We place the scene 180 cm away from the camera, and set the focus plane at 120 cm, leading to a defocus distance of 60 cm. We also take three exposures of the well focused scene to obtain a ground truth HDR image that allows comparison, using the same capture parameters described above. All images are taken in RAW format, with a size of 4752x3168 pixels. To reduce computational time and cost we resize images by a factor of 0.2, reducing them to 951x634 pixels.

In terms of system calibration, the process is again like that described in Section 2.7. The only clarification to make is that in this cas, in order to be coherent with image capture, we obtain three images, one for each exposure value, with the same capture parameters used to capture the scene. We also obtain an HDR image of the montage to obtain the PSF that we will use in the deconvolution in the HDR model. As in Section 2.7, the cropped grayscaled image of the LED serves us as PSF, after thresholding it in order to eliminate residual light, and normalizing it to preserve energy in the deconvolution process. Note that the threshold value changes for each PSF, increasing with the exposure value: 0.39 for underexposed, 0.5 for well-exposed and 0.8 for overexposed. For the PSF used in the HDR model the threshold value is 0.2. The resulting PSFs are shown in Figure 2.17. After resizing the kernel, its size is 14×14 pixels.

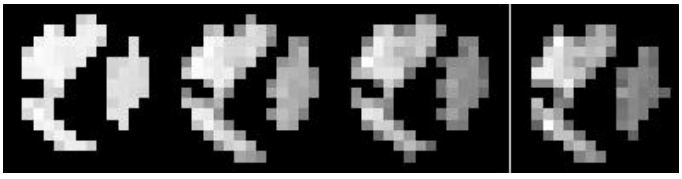
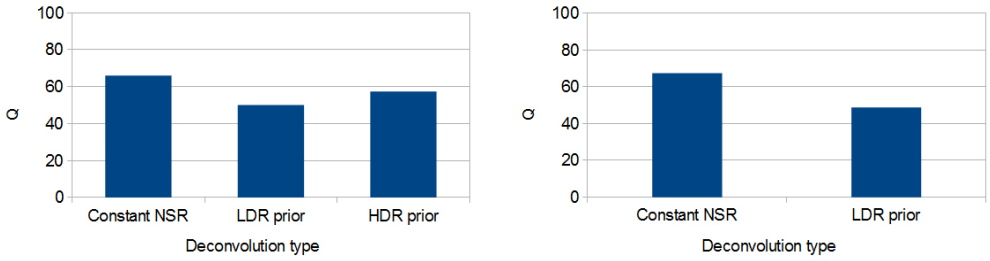


Figure 2.17: PSFs obtained for deconvolution. From left to right: PSF for the high, central and low exposure, used in the LDR model, and PSF obtained by merging the three exposures used in the HDR model.

Once we obtain the PSFs we recover the sharp images following the HDR and LDR models. For the HDR one we merge the defocused exposures into a defocused HDR radiance and obtain the deblurred HDR image performing a single deconvolution using the HDR kernel, as in Figure 2.13b. For the LDR model we perform one deconvolution for each defocused exposure, using the corresponding PSF for each one, and then we merge the resulting recovered exposures into the focused HDR image, as in Figure 2.13c. In each case, we carry out the same Wiener deconvolution variations described in Section 2.8.2, excluding again the use of an HDR prior for the LDR model. Once we perform all the experiments, we compare the results of both models. We compute the quality factor Q given by the *HDR-VDP2* metric of the HDR recovered images and show the results in two different scenes. We also check the effect of the use of the different deconvolution variations, specially those which employ deconvolution priors.

2.8.4.1 *Model comparison*

For our first scene, in Figure 2.18 we show the quality factor Q, given by the *HDR-VDP2* metric, of the HDR images recovered with each processing model. These results indicate that, while simulation results suggested that the LDR model offered better results than the HDR model (see Figure 2.14), real experiments point out that both models offer very similar qualities. Note also that, according to the metrics, the use of priors results in worse performance. We explore this fact further in Subsection 2.8.4.2.

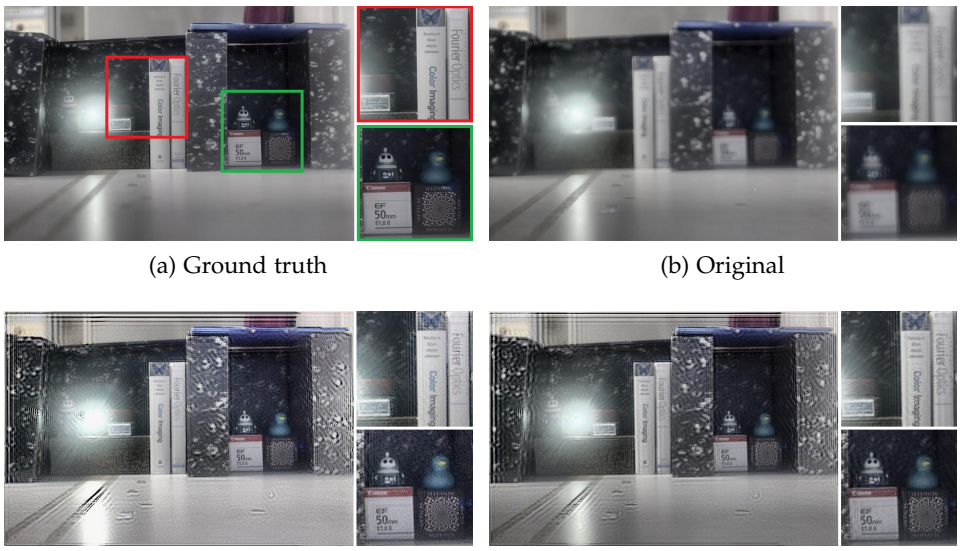


(a) HDR model (b) LDR model

Figure 2.18: Quality factor Q obtained with the *HDR-VDP2* metric for our first real scene, for each processing model and deconvolution prior. We can observe how in the HDR model the HDR prior outperforms the LDR one, and how both LDR and HDR models using constant NSR offer similar quality.

We show the result of both models, using constant NSR, in Figure 2.19, in order to offer a visual comparison of how both models perform. We also show the original (blurred) HDR radiance and the ground truth ideal HDR radiance. We can see how visual appearance is consistent with the results yielded by the metric. The image recovered with the HDR model shows more ringing due possibly to the biggest relative difference between neighbour pixels (see also Section 2.8.3). Furthermore, attending to the highlighted details and comparing recovered and original images we see how both models are able to recover the well-focused

HDR radiance (see e.g. book titles or text in the lens box in the images). These images prove that the employment of coded apertures for defocus deblurring of HDR images is viable and presents a good performance.

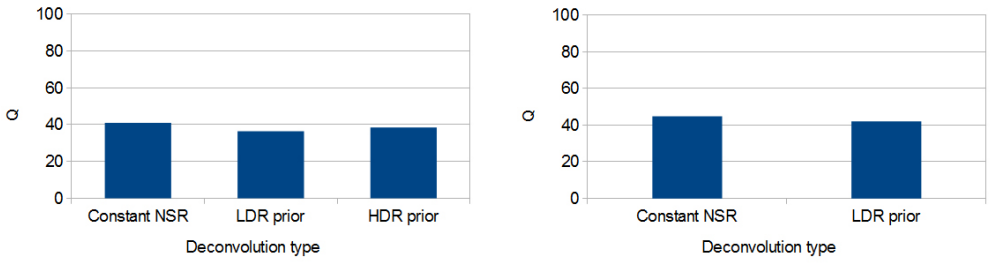


(a) Ground truth (b) Original

(c) HDR model with constant NSR (d) LDR model with constant NSR

Figure 2.19: HDR results obtained for our first real scene with the best processing models in terms of Q (c,d), compared to the ground truth and original images, all of them tonemapped. Here we see how both models offer good and similar results.

We test again our approximations performing the experiments in a new scene, in order to check if results correlate with the first ones. In Figure 2.20 we show the quality factor Q given by the *HDR-VDP2* metric for this second scene.

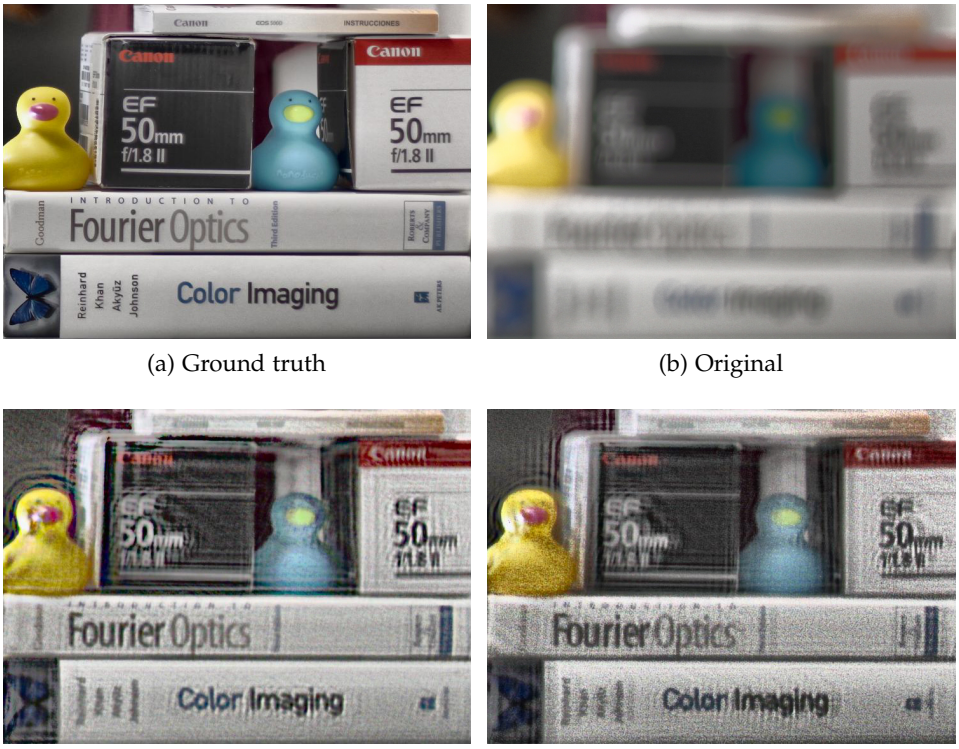


(a) HDR model (b) LDR model

Figure 2.20: Quality factor Q obtained with the *HDR-VDP2* metric for our second real scene, for each processing model and deconvolution prior. Note that in the HDR model the HDR prior outperforms the LDR one, and that in both models the use of a constant NSR offers the best results.

Again, the use of priors derives in worse results than the use of a constant NSR, for both processing models. In Figure 2.21 we show the HDR images of this scene recovered with the HDR and LDR models with constant NSR. As we

can see, both models offer good results when recovering the focused image, and again the HDR model exhibits slightly more ringing than its LDR counterpart.



(a) Ground truth (b) Original

(c) HDR model (d) LDR model

Figure 2.21: HDR results obtained for our second real scene with the best processing models in terms of Q, compared to the ground truth and original images, all of them tonemapped. We can see how both models are able to recover sharp details such as the book titles.

In Section 2.8.3 we have already pointed out possible causes for one model performing better than the other in simulation. When incorporating results in real scenarios, the Q metric seems to indicate similar results for both models, although it would be advisable to perform more tests with more data. Also, the *HDR-VDP2* metric works only with luminance values, not taking into account color, and while it has been specifically tested for some types of distortions, such as white noise or Gaussian blur, it has not been designed nor tested for e.g. ringing artifacts. Finally, modelling noise as GWN is another source of inaccuracy, an approximation, since image noise does not follow a Gaussian distribution.

2.8.4.2 *Effects of using a prior*

As shown in Figures 2.18 and 2.20, in real experiments we see that both HDR and LDR models perform much better when no deconvolution prior is used. We inspect the images recovered with both priors in order to know why this happens. If we carefully observe these images we can appreciate a grid shaped distortion, as seen in Figure 2.22. This distortion clearly reduces the visual quality of the images recovered with deconvolution prior. Further, we notice again that



(a) HDR model with HDR prior (b) HDR model with LDR prior (c) LDR model with LDR prior

Figure 2.22: Detail of our recovered images of the first real scene using priors, where we can appreciate a clear grid shape distortion. Note that, in the HDR model, using an HDR prior instead of an LDR one reduces this effect. All the images are tonemapped.

HDR prior outperforms LDR prior in the HDR model, as it minimizes, but not completely removes, this distortion.

We explore the variation of σ in the deconvolution process and see the impact of this alteration in the described distortion. This variation corresponds to a higher weight for the deconvolution prior. In Figure 2.23 we see some of the images obtained with different σ in the deconvolution process for the LDR model. We see how increasing this value we obtain a better reduction of prior distortion and ringing. In exchange, we find that this increase leads to less sharp results, resulting in a trade-off between both effects.



(a) $\sigma = 0.0005$ (b) $\sigma = 0.005$ (c) $\sigma = 0.05$

Figure 2.23: Effect of the variation of σ in the deconvolution for the LDR model. We can see a trade-off between the grid shape distortion and image sharpness. All the images are tonemapped.

2.9 CONCLUSIONS AND FUTURE WORK

In this chapter we have presented a method to obtain coded apertures for defocus deblurring, which takes into account human perception for the computation of the optimal aperture pattern. Following previous approaches, we pose the problem as an optimization, and, to our knowledge, propose the first algorithm

that makes use of perceptual quality metrics in the objective function. We explore the performance of different quality metrics for the design of coded apertures, including the well-established SSIM, and the state-of-the-art HDR-VDP-2, which features a comprehensive model of the HVS, as well as the L2 norm, previously used in related works. The results obtained show that the best apertures are obtained when a combination of the three metrics is used in the objective function, clearly outperforms existing apertures, both in simulated and real scenarios, results obtained by conventional circular apertures and by an existing aperture pattern specifically designed for defocus deblurring.

Additionally, we have explored non-binary aperture patterns, often neglected in the literature. Even though results with real images seem to indicate a better performance (i.e. less ringing artifacts) of non-binary apertures with respect to their binary counterparts, sharpness appears somewhat hindered by non-binary masks in comparison to binary patterns, resulting in a trade-off between both.

One of the challenges for the future is in noise modeling: Employing a better model for the noise inherent to the capture process would allow a better modeling of the process and thus a better design of coded aperture patterns. Although we show that simulated and real results correlate fairly well, differences remain, which may be overcome with a better model.

Additionally, we explore for the first time, to our knowledge, the use of coded apertures for defocus deblurring of HDR images, showing that these techniques, which used to be employed in LDR images, can be extended for HDR imaging. We also see that the use of deconvolution priors made of HDR images instead of conventional LDR priors leads to better performance. However, maybe due to the fact that the prior we are employing is far from optimal, the best results come when no prior is employed in the process. From this, and relying on the work of Pouli et al. [353], we believe that more research related to HDR priors is needed. Since many optimization problems benefit from the use of statistical regularities of the images, and taking into account the advances on HDR imaging, the construction of good HDR priors is another avenue of future work.

3

REVERSE TONE MAPPING

ABOUT THIS CHAPTER

The work here presented has been published in two papers and one technical report. The first paper, accepted to SIGGRAPH Asia and published in Transactions on Graphics, studies rTMOs across exposures and proposes an expansion method for over-exposed content, and was in part inspired by the work of Martin and colleagues [296]. The technical report further explores and improves upon the results of that paper, and the third paper, presented at CEIG, seeks a more artistic, and interactive, approach and proposes a semi-automatic method for range expansion.

B. Masia, S. Agustin, R. Fleming, O. Sorkine and D. Gutierrez.
Evaluation of Reverse Tone Mapping Through Varying Exposure Conditions.
ACM Transactions on Graphics, 28(5) (Proc. of SIGGRAPH Asia 2009).

B. Masia, R. Fleming, O. Sorkine and D. Gutierrez.
Selective Reverse Tone Mapping.
In Proc. of CEIG 2010.

B. Masia and D. Gutierrez.
Multilinear Regression for Gamma Expansion of Overexposed Content.
Technical Report RR-03-11, Universidad de Zaragoza. July 2011.

3.1 INTRODUCTION

High dynamic range display devices are becoming increasingly common [389], yet very large amount of existing low dynamic range legacy content and prevalence of 8-bit photography persist. This presents us with the problem of reverse tone mapping. The aim of reverse tone mapping operators (rTMOs) is to endow low dynamic range (LDR) imagery with the appearance of a higher dynamic range without introducing objectionable artifacts. Ideally, an rTMO should take a standard LDR image as input and reconstruct as accurately as possible the true luminance values of the original scene. As depicted in Figure 3.1, this is an ill-posed problem. For most scenes and imaging devices, the image data is irreversibly distorted by unknown nonlinearities, sensor noise, lens flare, blooming, and perhaps most importantly, sensor saturation, which clips high intensities to a constant value. Reverse tone mappers must somehow reconstruct the missing data, or boost the contrast in a way that does not cause the clipped regions to appear visually unpleasant.

Existing rTMOs tackle this ill-posed problem in different ways, leading them to succeed and fail in different conditions. For example, some reverse tone mapping

strategies may handle small clipped highlights well, but cause large saturated regions to appear unnatural. Conversely, other rTMOs may avoid introducing artifacts in over-exposed conditions, but fail to enhance under-exposed images sufficiently. The key is to understand which strategies produce the best possible visual experience, for which a number of user studies have recently been conducted [503, 390, 12, 34]. These experiments have yielded many valuable insights which may guide future rTMO and even HDR display design. However, they have been applied only to subjectively *correctly exposed* images, usually with knowledge of the dynamic range of the original, real-world scene. A key challenge in rTMO design is how to handle non-optimal LDR content, particularly images that are incorrectly exposed.

Our research here is dedicated to finding non-intrusive ways to take advantage of the higher dynamic range of the display medium, irrespective of the dynamic range of the original image. Reverse tone mapping also sheds light on a general problem in signal processing: taking partial, distorted or corrupted data and reconstructing the original as faithfully as possible. Here our quality criterion is perceptual faithfulness rather than physical accuracy.

The vast amount of LDR legacy content spans a large range of exposures. Under- or over-exposure may be due to different reasons, including bad choices by the photographer or pure artistic intentions. Legacy professional material may have been shot to make the most appropriate use of the dynamic range available at the time, very different from what is currently available. Additionally, the information about the dynamic range of the real scene is typically not recorded. It is therefore crucial to extend previous studies by taking into consideration varying exposure conditions for a set of images without additional information.

We have performed a series of psychophysical studies assessing how rTMOs handle images across a wide range of exposure levels. We have found that, while existing rTMOs perform sufficiently well for dimmer (under-exposed) images, their performance systematically decreases for brighter (over-exposed) input images. This suggests that there is a need for an rTM method that effectively deals with over-exposed content. We show that simply boosting the dynamic range by means of an adaptive γ curve achieves good results that outperform the current rTMOs, and propose a simple method to obtain a suitable value of γ for each image.

We additionally observe that artifacts produced by some rTMOs are also visible in low dynamic range renditions of the images. This is because many artifacts are not simply due to inappropriate intensity levels, but also have a spatial component. We perform a second user study to shed light on which type of inaccuracies introduced by reverse tone mapping most hamper our perception of the final image. This information can further help future rTMO design.

3.2 PREVIOUS WORK

3.2.1 *Reverse tone mapping*

Dynamic range expansion, along with related subsequent problems such as contour artifacts, has been initially addressed by bit-depth extension techniques [90]

and decontouring methods [91]. However, these techniques are designed for extension to bit-depths much lower than that of HDR displays. More recently, a few works have looked at the problem of reverse tone mapping for the display of LDR images and videos on HDR displays. The general approach of these reverse tone mapping techniques has been to identify the bright areas within the image, and in particular areas that have been clamped due to sensor saturation, such as light sources. Those areas are typically significantly expanded, while the rest is left unchanged or mildly expanded, to prevent noise amplification. We offer here a brief discussion on reverse tone mapping techniques, and refer the reader to the work by Banterle and colleagues [33] for a comprehensive review on the topic.

Banterle et al. [30, 31] apply the inverse of Reinhard's tone mapping operator [364] to the LDR image and detect areas of high luminance in the resultant HDR image. They then produce a so-called *expand-map* by density estimation of the bright areas, and use this map to interpolate between the LDR image and the initial inverse tone mapped HDR image, thus modulating the expansion range. This framework has been extended to video by designing a temporally-coherent version of the expand-map [32]. The LDR2HDR framework of Rempel et al. [368] is similar in spirit, but their expand-map (which they term *brightness enhancement function*) can be computed in real time using the GPU. The image intensity is first linearized, and a binary mask is computed by thresholding the saturated pixels; the brightness enhancement map is computed as a blurred version of the binary mask, combined with an edge stopping function to retain contrast of prominent edges. The contrast of the LDR image is then scaled according to the enhancement map. Note that the expansion is affected by the size of the bright objects: larger objects may receive more brightness boost. Recently, Kovaleski and Oliveira [240] presented a reverse tone mapping technique which is also based on real-time computation of a brightness enhancement function, but substitutes a bilateral filter for the combination of a Gaussian blur and an edge stopping function used by Rempel et al. [368].

Meylan et al. [316, 317] explicitly focus on specular highlight detection and apply a steep linear tone mapping curve to the presumably clamped areas, whereas the rest of the image is expanded by a mild linear curve. A more sophisticated segmentation and classification of bright areas in the image is done in the work of Didyk and colleagues [101]: they segment the bright image areas and label them as diffuse surfaces, light sources, specular highlights and reflections using a trained classifier. Different expansion functions are designed for each class to reproduce the dynamic range more accurately (in particular, the luminance of light sources and highlights is expanded more than that of reflections, while bright diffuse surfaces are not expanded). The method is suitable for high-quality video enhancement thanks to the temporal coherence of the segmentation and the expansion function. Finally, Wang et al. [465] propose to fill in the texture information of the clamped bright areas by transferring texture from other (well exposed) areas, although the method may not be viable if a suitable region for transferring detail is not found elsewhere. Both methods [101, 465] rely on user assistance to guide the process, whereas we are interested in more automatic approaches.

3.2.2 *User studies*

It is now generally accepted that HDR displays provide a richer visual experience than their LDR counterparts. However, different parameters such as luminance, contrast or spatial resolution influence our visual experience, which makes it difficult to come up with an ideal combination. Additionally, image content probably also affects our preferences. In computer graphics, several researchers have performed a series of user studies, the findings of which may even influence future hardware development.

Yoshida et al. [503] judged subjective preference (without a reference image) and fidelity (by comparing to a real world scene) for a series of tone mapped images. Users could adjust brightness, contrast and saturation for each individual image. Although their work was geared towards the design of a forward tone mapping operator, their conclusions are also useful for rTMO development: they found that, in general, brighter images were preferred over dimmer ones. Interestingly, however, in certain cases users would break this tendency and keep a significant portion of the image dark, reducing overall brightness and giving more importance to contrast.

Seetzen et al. [390] analyzed the influence of luminance, contrast and amplitude resolution of HDR displays, to guide future display designs. Their studies show that the preferred luminance and contrast levels are related: for a given contrast, perceived image quality increases with peak luminance, reaches a maximum and then slowly decreases.

Akyüz and colleagues [12] performed a series of psychophysical studies which revealed that a linear range expansion of the LDR image could surpass the appearance of a true HDR image, suggesting that simple solutions may suffice for reverse tone mapping. Recently, Banterle et al. [34] have presented a psychophysical evaluation of existing reverse tone mapping techniques, the results of which indicate that nonlinear contrast enhancement may yield better results overall.

These previous studies provide useful insight into the desirable behavior of tone mapping operators. A key difference with our work is that they were performed on correctly exposed images, whereas we are interested in analyzing reverse tone mapping across varying exposure conditions. In this work, we define over-exposed pixels as those with values $\geqslant 254$, and under-exposed pixels as those with null values [368, 296].

3.3 EXPERIMENT ONE: RTMO EVALUATION

To assess the overall performance of an rTMO, it is important to evaluate it across a range of different imaging conditions. To this end, we have performed a user study in which subjects directly compared the output of three reverse tone mapping schemes (plus standard LDR visualization) across a range of exposures, from clearly under-exposed to clearly over-exposed images. We asked subjects to rate the appearance of the reverse tone mapped images on a calibrated Brightside DR37-P monitor (32.26″ wide and 18.15″ high), with a black level of 0.015 cd/m^2 and a peak luminance of over 3000 cd/m^2. Calibration of the Brightside monitor was performed to confirm linearity and stable performance during

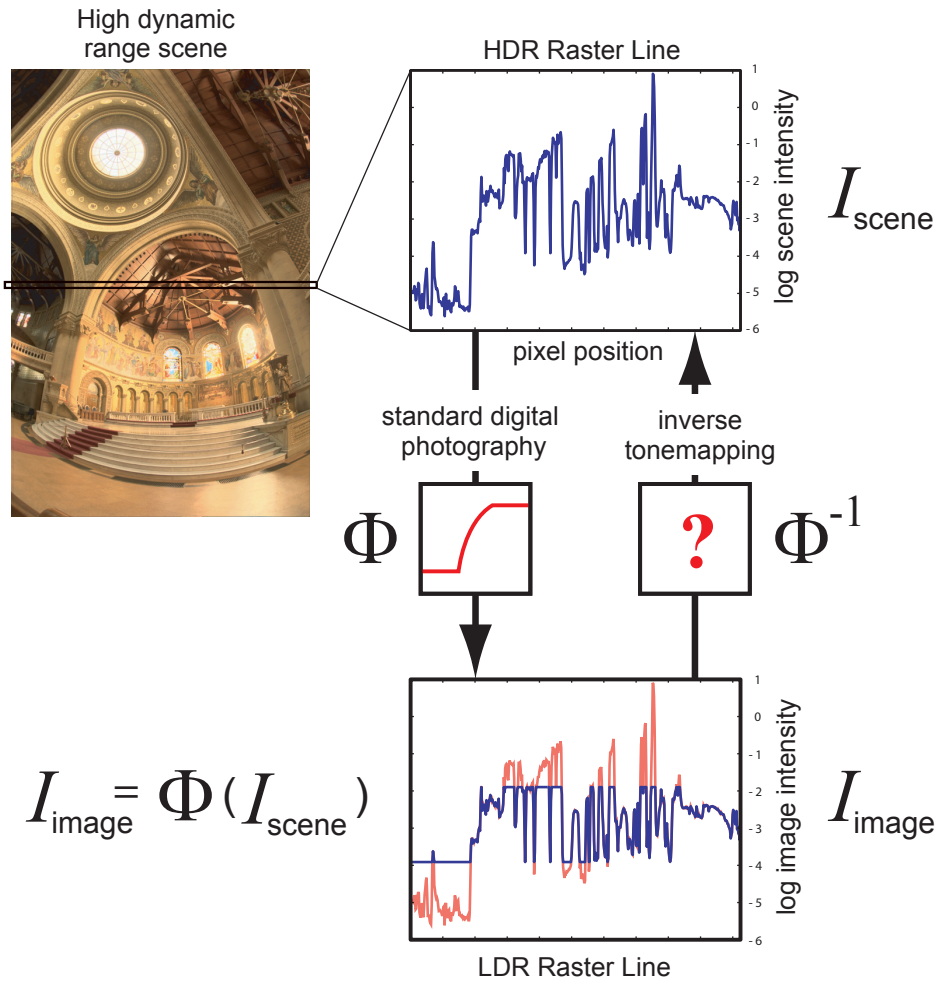


Figure 3.1: The reverse tone mapping problem. Standard imaging loses data by transforming the raw scene intensities I_{scene} through some unknown function Φ, which clips and distorts the original scene values to create the I_{image}, shown in the bottom panel (values clipped from the original are shown in red). The goal of an rTMO is to invert Φ to reconstruct the original scene data, or to convincingly "fake" it.

the experiment and to enable comparison to specific intensities in cd/m^2 should the need have arisen in the analysis, as per standard practice in psychophysics. Temperature compensation was turned off to avoid changes in intensity (this was possible thanks to the air conditioning in the room). The LDR versions of the images were displayed by approximately matching the contrast to a typical desktop TFT (Dell).

Ambient luminance was kept at about 20 cd/m^2, and the participants were seated approximately one meter away from the monitor. Based on the subjects' ratings, we can infer which rTMOs are most effective at recreating the experience of an HDR scene without visually objectionable side-effects. As opposed to other studies, we do not provide a ground truth HDR image for direct comparison, since it is almost always unavailable in the case of legacy content.



Figure 3.2: Representative samples of the stimuli used in our tests. Top: bright images (*Building, Lake, Graffiti, Strawberries, Sunset*), each showing a certain degree of over-exposure. Bottom: dark images (*Car, Flowers, Crayons, Pencils*), with varying degrees of under-exposure.



Figure 3.3: The complete bracketed sequence for the *Building* and *Flowers* scenes.

3.3.1 *Stimuli*

The stimuli consist of photographs of nine scenes with different lighting conditions, captured with a Nikon D200 at an original resolution of 3872 by 2592 (down-sampled for visualization purposes on the Brightside monitor, which has a 1920 by 1080 pixel resolution). Each scene was captured with four different exposure times. Five scenes were made up of bright images (from approximately correct exposure to clearly over-exposed), and the remaining four were made up of dark images (from clearly under-exposed to approximately correct). Figure 3.2 shows a representative image of each scene, while Figure 3.3 shows the four exposures for two example scenes. The stimuli (please refer to `http://webdiis.unizar.es/~bmasia/downloads/thesis/rTM_Stimuli.zip` for the complete series of all the scenes) have been obtained from a previous study on exposure perception [296], where the authors analyze basic image data to try to obtain a correlation between image statistics and the perception of under- and over-exposure. s From each exposure in the bracketed sequence, we obtained three candidate renditions for display on the HDR monitor using a representative subset of reverse tone mapping algorithms: LDR2HDR [368], Banterle's operator [30] and linear contrast scaling [12]. Except for the straightforward linear scaling (in Yxy color space, and thus performed on linearized values) we obtained the images from the authors of the original algorithms, in order to ensure accuracy in the implementation. For the LDR2HDR algorithm the parameters used were 150 pixels for the standard deviation of the large Gaussian blur applied to the mask, a brightness amplification factor $\alpha = 4$ and a gradient image baseline width for divided differences of 5 pixels, plus a 9×9-pixel kernel for the antialiasing blur and a 4-pixel radius for the *open* operator used to clean up the final edge stopping function (please refer to the original paper for a detailed explanation of these parameters).

In the case of Banterle's operator, when generating the expand-map, the parameters of the density estimation were a radius ranging from 16 to 42 pixels (smaller radius for lower exposures) and a threshold of 1 to 4 light sources (lower threshold for higher exposures), being 2048 the number of generated light sources for Median Cut sampling. In both cases, Banterle's operator and LDR2HDR, images were linearized using gamma correction ($\gamma = 2.2$). We also added a fourth LDR rendition in which the original images are presented within a luminance range matched to a typical desktop TFT monitor. The goal of this fourth image is to study whether the established assumption that visual preference is given to HDR holds over a range of exposures.

3.3.2 *Subjects*

A gender-balanced set of twelve subjects with normal or corrected-to-normal acuity and normal color vision were recruited to participate in the experiment. All subjects were unaware of the purpose of the study, and were unfamiliar with HDR imaging.

3.3.3 *Procedure*

Participants viewed the stimuli on the Brightside HDR display in a dark room. On each trial, subjects were presented with all four renditions of a given exposure of a given scene in a 2×2 array (a *stimulus quadruple*). The positions of the four renditions within the array were random across trials, and the order of the trials was random with the constraint that consecutive trials did not present the same scene. The subjects' task was to rate the quality of the four renditions on a scale from 1 to 7, according to how accurately the images depicted how the scene would appear to the subject if they were actually present in the scene. Thus the key criterion for comparison was the subjective fidelity of the renditions. Subjects were given unlimited time for each trial and could modify their rating of any of the renditions on a given trial before proceeding to the next trial. Additionally, they were free to assign the same values to all four renditions on a given trial, although they were instructed to try to use as much of the 1-7 scale as possible within the experiment as a whole. To aid them in setting their scale, and to accustom them to the experimental procedure, the subjects were presented with a number of practice trials before the start of the experiment.

3.3.4 *Results*

Several conclusions can be drawn from this test. First, for our images, there was a clear difference in perceived quality between the bright and the dark series: subjects clearly preferred the reverse tone mapped depictions of darker images over brighter ones. This can be seen in Figure 3.4: not only is the overall mean value significantly higher in the former case, but it is relatively stable across exposure as well. In contrast, for the bright images, there is a general downward trend in ratings across the four exposure levels.

(i-j)	$p_b(i,j)$	$p_d(i,j)$
LDR2HDR - Banterle's	2.0532e-21	2.8633e-7
LDR2HDR - Linear	0.5734	0.0283
LDR2HDR - LDR	1.7762e-6	1.4976e-11
Banterle's - Linear	1.1739e-22	0,0013
Banterle's - LDR	4.4489e-11	0.1938
Linear - LDR	1.4697e-7	2.0538e-6

Table 3.1: Results of the Wilcoxon rank sum tests for the bright and dark series (denoted by subindices b and d respectively). Values of $p < 0.05$ are considered to indicate statistically significant differences between rTMOs. Thus, all differences were significant except for LDR2HDR vs. Linear in the bright series and Banterle vs. LDR in the dark series.

Note that this gradual decrease in performance does not correlate with the subjective perception of quality of the original LDR image: in a previous pilot study, users picked different exposures for each series as the subjective best, not necessarily the same as the objective best (defined as the one with the smallest proportion of under- and over-exposed pixels [12]). The trend instead correlates with the proportion of over-exposed pixels and the mean luminance, which do increase with exposure.

Secondly, we can observe *systematic* differences between the rTMOs. On average, subjects rated the LDR2HDR and the Linear rTMOs best (the difference between the two failed to reach statistical significance), followed by the LDR images, and finally the output of the Banterle's rTMO (see Figure 3.4). Pairwise Wilcoxon rank sum tests (similar to a non-parametric version of the t-test) reveal that these differences were significant to $p < 0.05$, except for LDR2HDR vs. Linear in the bright series and Banterle's operator vs. the LDR depiction in the dark series (see Table 3.1 for the complete results).

It is important to note, however, that this ordering does not hold for all conditions. For instance, the LDR depiction was systematically ranked lower than two of the rTMOs, suggesting that indeed HDR visualization is still preferred over LDR, even for under- and over-exposed images. Surprisingly, though, it ranked higher on average than Banterle's rTMO for bright images. The poor overall performance of Banterle's rTMO with this data set is probably due to the fact that it often exaggerates the errors in poorly exposed images, resulting in intrusive artifacts. This becomes clear when we measure the extent to which each rTMO yields outlier rating values for each image. We calculate the median rating for each image across rTMOs. We then obtain the *outlier index* as the difference in rating for each rTMO relative to this median value. When an rTMO is *neutral*, simply reflecting the overall quality of the exposure of the image, then the outlier index tends to be close to zero. However, when an rTMO stands out relative to the others (for example due to the introduction of artifacts), then the outlier index tends to deviate from zero. In Figure 3.5, we plot the histogram of the outlier index values for the three rTMOs and the LDR depiction. It is notable that for LDR2HDR, Linear and LDR, the distribution tends to be relatively tightly

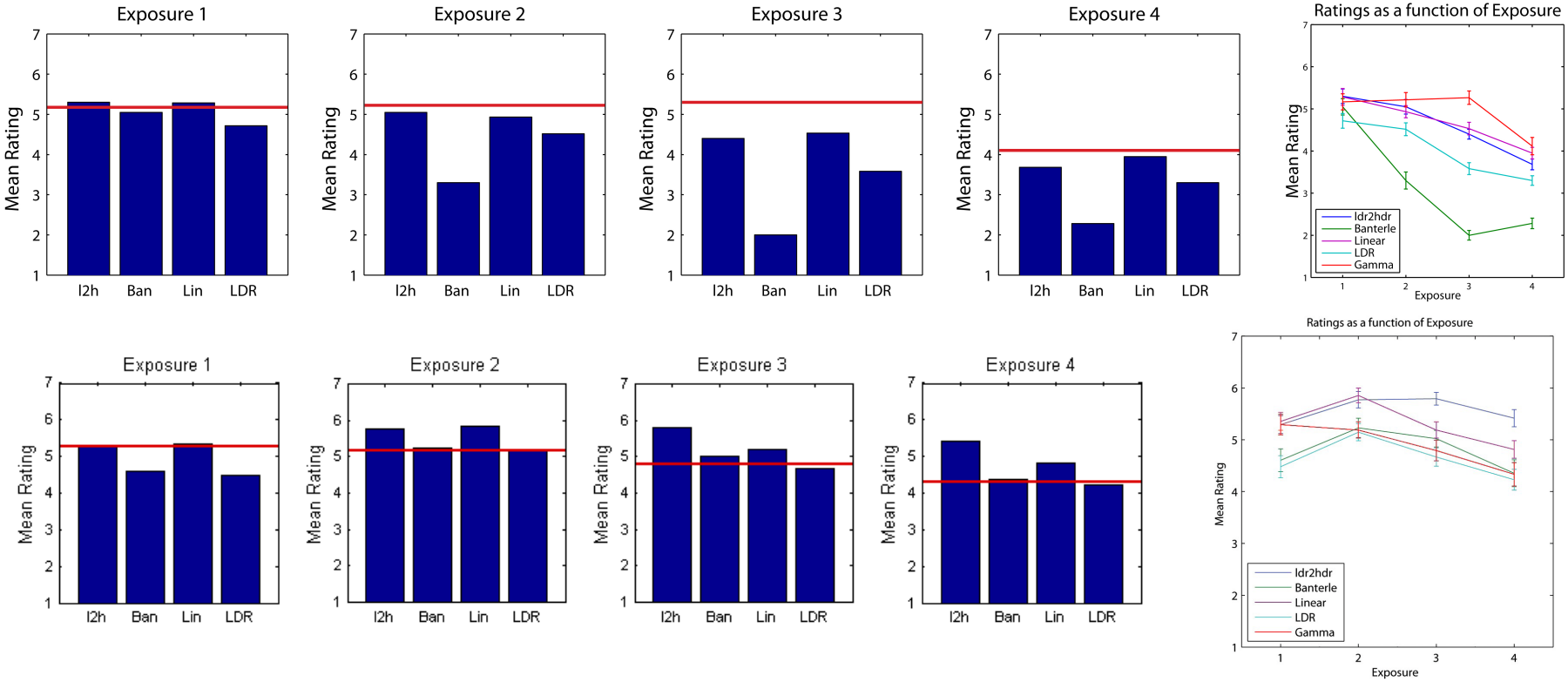


Figure 3.4: Top: bright images series. The blue bars represent the mean ratings across subjects for the four rTMOs (LDR2HDR, Banterle's, Linear and LDR) with increasing exposure levels (see Section 3.3). The last chart clearly shows the downward trend in perceived image quality. Error bars represent standard errors on the mean. The red line in the first four charts represents the mean ratings for our proposed γ-curve expansion (see Section 3.5). It can be seen that it rates generally higher and is more stable. Bottom: same information for the dark images series, showing higher overall means and a more stable perceived quality across exposures; the absence of significant improvement derives from the fact that our expansion method is designed for bright images.

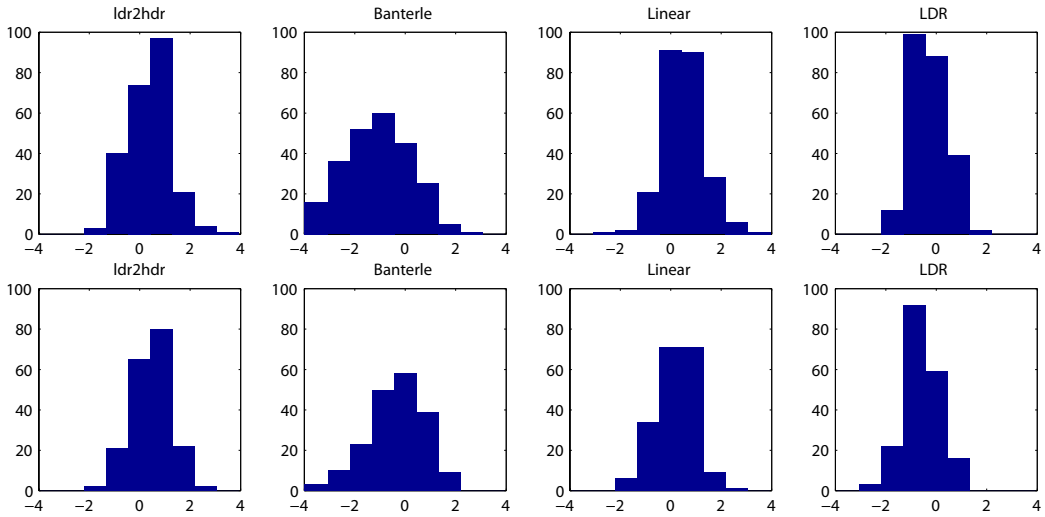


Figure 3.5: Distribution of outlier indices for all four rTMOs. Top: bright series. Bottom: dark series.

tuned, while for Banterle's the spread is much broader. This means that on the one hand, when it performs well, it tends to equal or exceed the others. However, it sometimes introduces substantial artifacts that cause the images to look worse than if they were not reverse tone mapped at all.

Although this seems to contradict a recent study where Banterle's operator actually outperformed other rTMOs [34], it is important to note that the experiments carried out in both cases differ significantly: first of all, in the work by Banterle et al. [34] the LDR source images were again well exposed, which is the regime within which Banterle's rTMO performs well, as we also found. However, when the source material is less flattering, we found that the algorithm sometimes produces clearly visible artifacts, which leads to lower ratings. Second, in [34] the authors used a 2AFC paradigm with direct ground truth comparison, whereas we propose a rating approach, which allows users to report their relative subjective preferences. Both tasks are valid ways of assessing fidelity. However, ours has the advantage that it is closer to the real usage scenario: in general the ground truth is unknown and is not presented for comparison.

3.4 EXPERIMENT TWO: HDR VS. LDR MONITOR

We notice that artifacts produced by LDR2HDR and Banterle's rTMOs are typically visible in low dynamic range renditions of the images. This is because they generally have a spatial component: they are not simply due to inappropriate intensity levels for certain features, but they also include fringes, visibly boosted noise and other artifacts. To analyze this, we performed a second experiment with seven new subjects, which was identical to the first experiment, except that on each trial, the 2×2 stimulus array was tone mapped using histogram adjustment [470][1]. The array was then presented on a standard TFT monitor (note that

1 We have used the *pcond* program in *Radiance* to tone map the stimuli.

this means that the LDR control condition now appears much darker than on a normal TFT).

In Figure 3.6, we plot the average ratings for each image in the LDR control condition against the average ratings in the HDR condition. As can be seen from the scatter plot, the ratings in the LDR control condition correlated extremely strongly with the ratings in the original experiment on the HDR monitor ($r^2 = 0.9018$). We found no significant difference between bright and dark images.

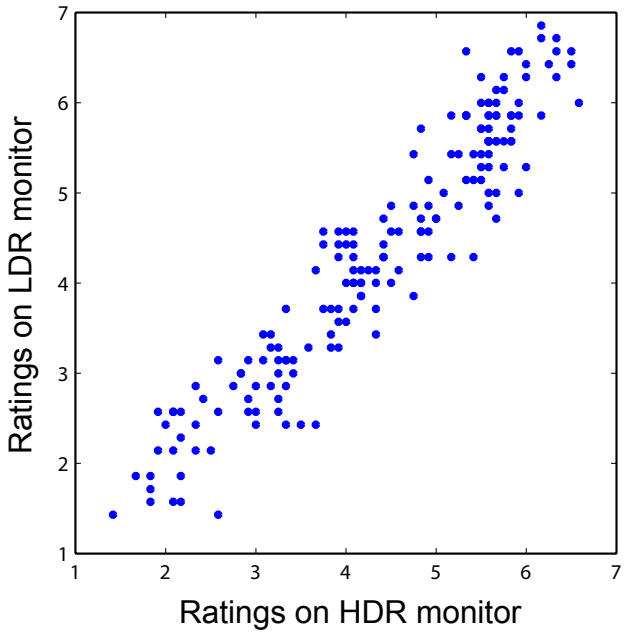


Figure 3.6: Scatter plot showing a strong correlation between ratings on an HDR monitor and ratings when the images were tone mapped back down to LDR and presented on a standard TFT monitor.

This result does not imply that the images look *the same* in LDR as in HDR: the subjects were not asked to compare these conditions directly, and previous studies have confirmed that HDR depictions are preferred over LDR [12]. Indeed, none of the subjects saw both renditions. However, it does demonstrate that the pattern of preferences is extremely well conserved. In other words, the images that were less preferred on the HDR monitor were also less preferred when tone mapped back down to LDR. This has two important implications. First, the strong correlation found suggests that a reasonably predictive evaluation of a rTMO could be made without directly testing on an HDR monitor. Second, as noted, the subjective ratings of HDR images that have been generated from LDR images seem to depend more on the presence or absence of disturbing spatial artifacts than on the exact intensities of different features. A similar observation (confirmed by our test) was made by Aydin et al. [24]: they noted that the key issue in image reproduction is to accurately maintain the important features while preserving overall structure, whereas achieving an optical match becomes relatively less important. This becomes even more salient given that the dark-adaptation state of the observer is typically unknown, making absolute intensities meaningless to the user.

The design philosophy that emerges from these considerations is that it is generally better to apply simpler, less-aggressive rTMO schemes if the original image is imperfect. Failing to fully recreate the HDR experience is less disturbing to users than unintended artifacts that can occur when poorly-exposed images are adjusted too aggressively. In the following section we present a simple and robust approach to boosting the dynamic range of over-exposed images, and show that it is less prone to artifacts than other rTMOs.

3.5 EXPANDING OVER-EXPOSED CONTENT

Our experiments have shown that the danger with computationally sophisticated reverse tone mapping schemes is the potential to make the image appear worse than before processing, through the introduction of objectionable artifacts. However, the goal of a rTMO is to make the image content look better in general and avoid, under any circumstances, making it look worse. Simple global reverse tone mappers, such as linear scaling and gamma boosting, never cause polarity reversals, ringing artifacts or spuriously boost regions well beyond their context. Our first experiment clearly indicates that there is room for improvement in devising an rTMO for bright input images with large saturated areas, whilst darker images turn out much better. We thus focus on the former in this section.

Examining the bright sequence in Figure 3.3 we observe that as exposure increases, more detail is lost as pixel values become saturated, and colors fade to white. It thus seems reasonable to attempt to depict the image in a way that the remaining details become more prominent, as opposed to boosting saturated areas as existing rTMOs do. Note that we do not aim to recover information lost to over-exposure, for which existing hallucination techniques may work [465], but rather to increase perceived quality.

We make the following key observations, which have been confirmed by previous studies on reverse tone mapping: on the one hand, darker HDR depictions are usually preferred for bright input LDR images [316]; on the other hand, in many cases contrast enhancements improve perceived image quality [368]. These suggest expansion of the linearized luminance values following a simple γ curve, which has the desired effect of darkening the overall appearance of the images while increasing contrast. Linearization of the luminance values prior to the dynamic range expansion was done with a gamma curve ($\gamma = 2.2$), following the findings by Rempel et al. [368] which note that simple gamma correction can be used for linearization instead of the inverse of the camera response without producing visible artifacts. To avoid amplifying noise, a bilateral filter [440] can be used prior to expansion [368]. Gamma expansion may potentially boost noise; however, over-exposed images tend to be significantly less noisy than under-exposed ones. Our psychophysical tests confirmed that noise amplification did not affect the final perceived quality.

3.5.1 *Determining the value of γ*

Obviously, the problem with the proposed expansion lies in automatically obtaining an image-dependent suitable γ value, to avoid the cumbersome manual

readjustment of the display settings for each individual image to be shown. We asked users in a pilot study to manually adjust the value of γ in a set of images. Table 3.2 includes the γ values for the image database. Columns 1 to 4 indicate increasing exposure, while rows correspond to the different scenes captured. We

Table 3.2: γ values for the five scenes and the four exposure levels.

Scene	1	2	3	4
Building	1.22	1.5	1.75	2.6
Lake	1.1	1.2	1.5	2.25
Sunset	1.1	1.35	1.4	1.75
Graffiti	1.2	1.35	1.5	1.75
Strawberries	1.22	1.35	1.55	1.9

additionally compute a series of statistics for each image. For all of them, luminance is obtained from sRGB linearized values [365]:

$$L = Y = 0.2126R + 0.7152G + 0.0722B, \tag{12}$$

where L is thus normalized to [0..1].

These statistics include both the arithmetic and the geometric mean luminance (referred to as L_{avg} and L_H, respectively). The arithmetic mean is simply obtained by averaging the luminance value of all pixels ($L_{avg} = 1/N \sum_{i=1}^{N} L(i)$, with N being the total number of pixels in the image); the geometric mean, known to reduce the contribution of outliers, is obtained as follows [37]:

$$L_H = exp\left(\frac{1}{N}\sum_{i=1}^{N} log(L(i) + \varepsilon)\right), \tag{13}$$

where ε is a very small positive number to prevent singularities in black pixels. We additionally compute the logarithm of this quantity, simply $logL_H$. The key of the images is also obtained, using the following equation [11]:

$$k = \frac{logL_H - logL_{min}}{logL_{max} - logL_{min}}. \tag{14}$$

In this equation L_{max} and L_{min} are the maximum and minimum luminance values, respectively, once a percentage of outlier pixels (both on the dark and bright sides) has been eliminated. We calculate two key values, k_5 and k_1, considering 5% or 1% of the pixels as outliers, respectively. Additionally, both the median, L_{med}, and a series of central moments, are computed for the luminance of the images. These include variance V_L (and standard deviation σ_L), skewness ($skew_L$) and kurtosis ($kurt_L$). Finally, we compute the percentage of over-exposed pixels for each of the images, defining over-exposed pixels as those with $L \cdot 255 \geqslant 254$; we will refer to it as p_{ov}. Table A.1 in Appendix A includes the values obtained for each of the aforementioned statistics for the images of our dataset. In the following we show the regressions we explored for the obtention of the γ value of an image from its statistics.

We proceeded to fitting the data with a multilinear regression, that is, a linear regression with multiple variables as predictors. Restricting ourselves to linear regressions was decided to keep the model as simple as possible; if a good model could not be found assuming a linear relationship, we would move on to more complex fittings. We initially used ordinary least squares to do the fittings. This implies a series of assumptions over the errors, mainly that they are normally distributed, with constant variance, and independent of each other. It also implies that the independent variables are free of error, or that their error is insignificant compared to the error of the dependent variable. We tested and analyzed several different fittings, varying the subset of image statistics that constituted the independent variables.

Once the type of model (i.e. linear) has been chosen, the problem which arises when working with multiple predictors is knowing which of the possible predictors (i.e. the independent variables, in our case the calculated image statistics) should be included in the model and which should be left out.

The way in which we deal with this is performing F-tests over the possible models. Computing the R^2 value or another goodness of fit metric and comparing their values for both models is typically not enough. The reason for this is that given two models, A and B, with p_A and p_B terms, respectively, if $p_A > p_B$, model A will always fit the data at least as good as model B. Thus, what has to be found out is if the addition of that extra parameter(s) to model A gives a *significantly* better fitting; as mentioned, we make use of F-statistics to assess that. Appendix B describes the use of F-tests in the construction of multiple variable models.

A stepwise regression is used to build the possible multilinear models [422]. This yields different final models[2], for which a series of metrics are computed in order to evaluate the accuracy of the fitting. In particular we compute the RMSE and the *overall* F-statistic for each model obtained (see Appendix C for a definition of these parameters). Additionally, we check for and remove outliers. After this, the model which yields the best fit to the data—i.e. the one with the lowest RMSE, and with the lowest p-value in the overall F-test—is the one given by the following equation:

$$\gamma = 0.9855 + 2.8972 L_H - 0.8232 L_{med} + 0.2734 skew_L - 0.0898 kurt_L \qquad (15)$$

Figure 3.7 shows the observed γ values in the x-axis, against the γ values predicted by our model. To further illustrate the predictive accuracy of our model, the figure also compiles goodness of fit measures, including RMSE, the F-statistic, R^2 and $\tilde{R}^2$.

Alternatively, we can retain all the observed data but weight their influence when computing the regression. To do this, we perform a new regression using iteratively reweighed least squares. The weight function used is a bisquare function. The new regression is thus given by the following equation:

$$\gamma = 2.4379 + 0.2319 logL_H - 1.1228 k_1 + 0.0085 p_{ov}. \qquad (16)$$

Figure 3.8 shows the predictive accuracy of the model obtained by robust regression compared to ordinary least squares (OLS). Additionally, if we compute

2 More details on possible models and how they are built can be found in [301].

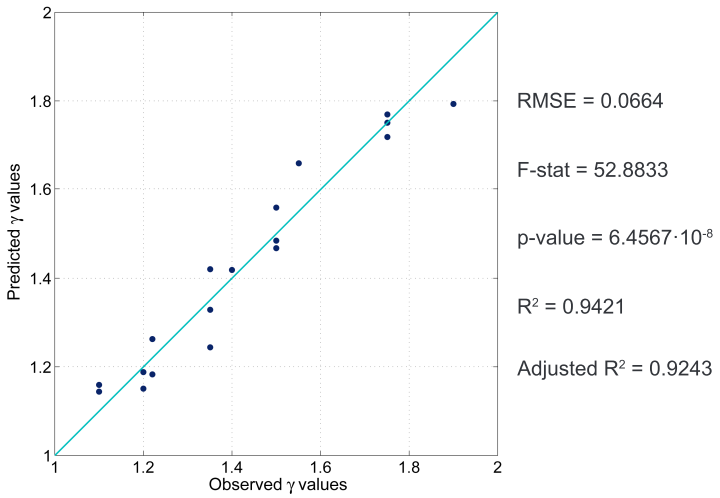


Figure 3.7: Predictive accuracy of the regression shown in Equation 15. The x-axis shows observed γ values, while the y-axis depicts the values predicted by the regression. The cyan line shows the quadrant bisection.

a robust RMSE estimate for this last regression [114], we obtain an estimate of 0.0962 (while estimates for the previous one, OLS without oultiers, was 0.0664).

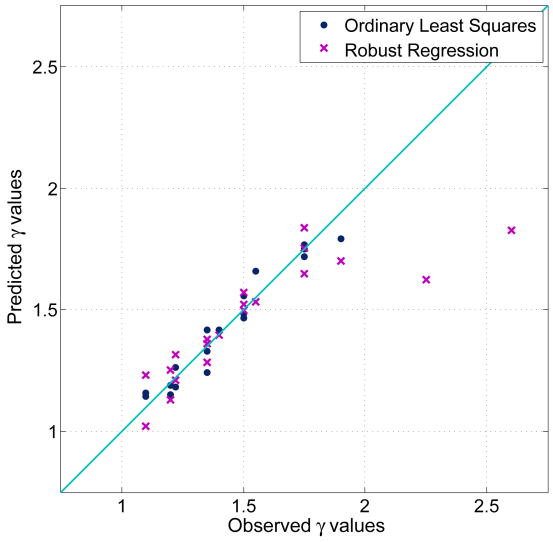
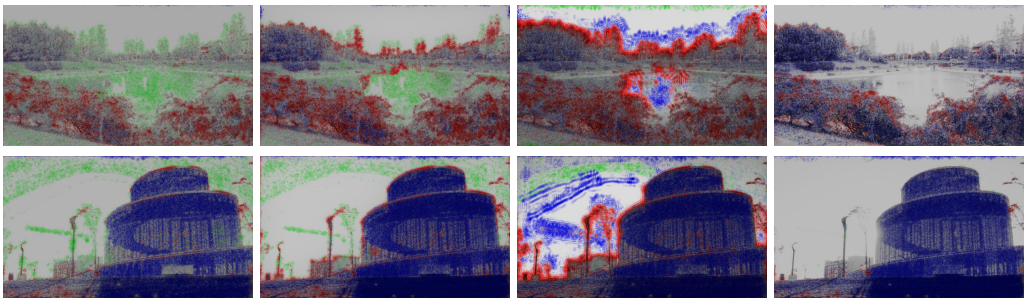


Figure 3.8: Predictive accuracy of the model obtained by robust regression against the one obtained by ordinary least squares. The abscissa shows observed γ values, while the y-axis depicts the values predicted by the regression. The cyan line marks the quadrant bisection.

3.5.2 *Validation*

To provide a subjective evaluation of the performance of our γ-expansion, we repeated Experiment One (Section 3.3), substituting the LDR depiction with our γ-expanded versions (see Table 3.2) in order to maintain the 2×2 stimulus array. The red line in Figure 3.4 shows the results.



Linear expansion | LDR2HDR | Banterle's operator | our γ curve

Figure 3.9: Comparing the results of several rTMOs with the image quality metric from Aydin et al.[24]. The reference LDR images are *Lake* (top) and *Building* (bottom) as depicted in Figure 3.2 (which correspond to the third and second exposure levels in the series. Please refer to Appendix D for all the exposures in all the scenes). Green, blue and red identify loss of visible contrast, amplification of invisible contrast and contrast reversal respectively. Our γ expansion does not lose any contrast, while minimizing gradient reversals. More importantly, it reveals more detail in the most significant areas of the images (trees, grass, bushes and buildings in the images shown).

Experiment One provides useful information about the *subjective* perception of image quality. However, we are also interested in evaluating our approach from an *objective* point of view. The problem is the fact that the intended comparison needs to be performed between an LDR and an HDR image. Aydin and colleagues [24] presented a novel image quality metric which identifies visible distortions between two images, independently of their respective dynamic ranges. The metric uses a model of the human visual system, and classifies visible changes between a reference and a test image. The authors identify three types of structural changes: loss of visible contrast (when contrast visible in the reference image becomes invisible in the second one), amplification of invisible contrast (when invisible contrast in the reference image becomes visible in the second one), and reversal of visible contrast (when contrast polarity is reversed in the second image with respect to the reference). It is important to remember that, as Rempel and colleagues noted [368], contrast enhancement tends to increase perceived quality, and therefore is a desired outcome of the rTMO.

Figure 3.9 shows the results of this metric[3] comparing two of the original LDR images (reference images) with the corresponding outputs using linear expansion, LDR2HDR, Banterle's operator and our proposed γ curve. Our method reveals more detail, shows no loss of contrast and minimizes gradient reversals. Note that while our approach may fail to utilize the dynamic range to its full extent in some cases, it has the important and experimentally validated advantage of avoiding objectionable and unpredictable artifacts.

3 We have used the online implementation provided by the original authors of the paper: `http://drim.mpi-inf.mpg.de/generator.php`

3.6 DISCUSSION

Experiment One shows that performance of rTMOs decreases for input images containing a large number of over-exposed pixels, while they seem to perform significantly better for darker images. This suggests that for bright images the consensual approach of boosting bright areas could be improved. We have shown that a simple rTMO based on γ expansion, without the need for explicitly detecting saturated areas, outperforms existing rTMOs in these cases, and propose an empirical expression to automatically find a suitable γ as a function of the image's key, without user interaction. This rTMO has the desired properties of boosting contrast and detail in non-saturated areas of the image, visually compensating for the lack of information in the saturated ones.

We have performed two validation studies, both subjective and objective. The first one has confirmed that our approach increases the perceived image quality for these kind of images. Pairwise Wilcoxon rank sum tests revealed that the differences in rating were statistically significant with respect to all other rTMOs tested. Given that it produces darker overall images with increased contrast, this result is in accordance with previous suggestions [316, 368]. The second evaluation uses a recently published image quality metric which operates with arbitrary dynamic ranges [24]. The metric concludes that our method reveals more detail in non-saturated areas, does not reduce contrast and shows less gradient reversals than the other rTMOs tested. Thus, the artists' original intentions are better preserved.

In both experiments we used typical numbers of subjects for a within-subject design in psychophysics, and the results were highly coherent across subjects. In Experiment One the reported results are statistically significant to the $p < 0.05$ level, meaning that the chances that the outcome of the pairwise comparisons would change after running more subjects from the same population is less than 5%. Indeed, for many of the results, the probability is many orders of magnitude lower than this, which implies that the qualitative pattern of the results is well conserved across subjects. Likewise, data from Experiment Two exhibit a correlation coefficient of 0.9018, notably conclusive in statistical terms.

We also ran the same expansion on the images from the dark series: as expected, we found no significant improvements over the tested rTMOs, given that our expansion is designed for bright images (see Figure 3.4, bottom).

The results from our second experiment confirm that spatial artifacts are more disturbing than inaccuracy in reproduced intensity levels [24]. We found a very strong correlation in the pattern of preferences when viewing images on HDR and LDR displays. This does not mean that the images looked the same, but it does suggest that the artifacts that emerge with poorly-exposed input images are spatial in nature and severe enough that HDR evaluation is not necessary: they can also be clearly seen in LDR.

3.7 SELECTIVE REVERSE TONE MAPPING

We have seen how, to produce a pleasant HDR image from LDR input, existing rTMOs work under the general assumption that highly saturated pixels need



Figure 3.10: Examples of images containing large saturated areas.

to be expanded much more than the rest. As a result, bright image areas representing features like highlights, or the sun in the sky, are largely boosted, thus counter-parting the clamping of information in the LDR image and better representing the real-world experience. Even though these techniques can produce appealing results for a wide range of LDR content, there are some cases in which the general approach of boosting bright areas may not be the best way to proceed, as shown in the first part of this chapter. These cases include images -such as those shown in Figure 3.10- which contain large saturated areas, either because of artistic purposes or due to a bad exposure.

In this section we show how a tailored approach to dynamic range expansion can be a good alternative in those cases which are unfavorable for existing rTMOs. We present two different techniques, one based in Ansel Adams' Zone System [3], and another based on detection of salient features, which allow the user to control dynamic range expansion based on her own preferences or intended goal. The techniques can also be used in combination with each other. This provides a new method for reverse tone mapping and an artistic tool where tonal balance and mood of the final HDR image can be adjusted by the user (in a similar manner to existing tools for LDR or HDR images [65, 260, 27, 268, 126]).

3.7.1 *Using the Zone System for rTM*

The so-called Zone System was introduced by Ansel Adams as a guide to produce good photographs with correct tonal values [3]. Exposure is the main factor which determines the way in which the luminance values of the scene are finally mapped to the limited range of values which can be reproduced by the photograph; choosing the right exposure is therefore one of the most important concerns of a photographer. Common exposure meters are designed to aid in this task by measuring luminance values of the scene (or object of interest) and suggesting the lens aperture and shutter speed values. However, irrespective of the scene -its lighting or content-, the values provided by an exposure meter are always such that the object of interest will appear as middle gray in the final image, which in many cases will not be the adequate election. A simple example which illustrates this problem is that of photographing a black and white checkerboard and a scene which is all black except for a white square: the same exposure settings should be used in both of them, yet the reading of an exposure meter would give very different exposure settings for each one. Ansel Adams' Zone System provides a simple way of, using this middle gray reading of exposure meters, choosing the best exposure settings. This system is not only a tool for photographers still widely in use today [207, 145], but also a formalization



Figure 3.11: Division of luminance in zones according to Ansel Adams' System.

of sensitometry principles which provides deep insight into how mapping of tonal values works. Reinhard et al. [364] already rely on it as a basis for their well-known tone mapper, and posterior works on interactive tone management have also built on this system [268]. Following Adams' technique the luminance values in a scene can be divided into ten different luminance zones (0 through IX, see Figure 3.11) according to the equation given by Koren [238]:

$$p = \left(\left(\exp(v \sin\left(\pi \frac{zone - 1}{16}\right) - 1\right) / \left(\exp(v) - 1\right)\right)^{\psi}, \tag{17}$$

where p represents the zone limits in normalized pixel luminances and ψ is the encoding function responsible for non-linearities in the LDR values (usually the inverse of a γ function). The value $v = 5.25$ is set so that zone V on a properly calibrated monitor appears as middle gray [3], defined as 21% of the maximum screen brightness level (this is similar to 18% reflectance referenced to 90% white, which is pure white on good photographic paper). Equation 17 is designed so that input values of zero and one map to zones 0 and IX respectively, while the sine function is responsible for the compression required at high pixel levels. Once the luminance range of the LDR input image is divided in zones according

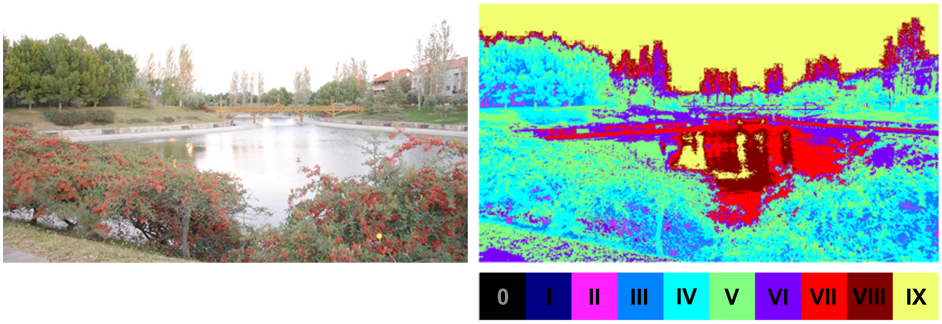


Figure 3.12: *Left*: Input LDR image. *Right*: The result of luminance decomposition for zone-based reverse tone mapping.

to Equation 17 (see Figure 3.12) the reverse tone mapping process is done by assigning different expansion functions to the different zones. Although in theory these functions could be as complex as desired, we choose to use linear functions for each zone, as they offer a good balance between simplicity and control over the expansion. Thus, the resulting rTM function is piece-wise linear. The darkest and the brightest zones (0 and IX, respectively) of the LDR image are mapped to the lowest and the highest luminance values of the HDR display. A second constraint is that the rTM function must be monotonically increasing, as otherwise gradient reversals may appear that spoil the final depiction. Expansion is performed on the luminance channel, and the RGB channels are then recovered. Saturation can be tuned when recovering chromaticities in order to obtain the best depiction. Adjusting the slopes of each of the zones may seem like an in-

volved process; however, in the end it somehow resembles what photographers constantly do, as it translates to assigning ranges of the HDR image luminance to each zone of the LDR input image. Besides, the calculation of the resulting HDR image is almost immediate, thus allowing the user to try different curves before choosing the final one. As an example, Figure 3.13, *right* shows an HDR image obtained by using a piece-wise linear curve on which only three values were specified: Zone IV being assigned 10% of the HDR image luminance range, Zone VI 40% of that range, and Zone VII 60% of it, which translates to adjusting three points of the LDR–HDR curve shown. We can also appreciate how this simple tuning of the rTM function yields a more appealing depiction than the linear scaling (shown to be on par in subject preference with the HDR image itself by Akyüz and colleagues [12]). Additionally, this zone-based expansion can also be used as part of a bigger rTM framework, as examples in Section 8.7 show.

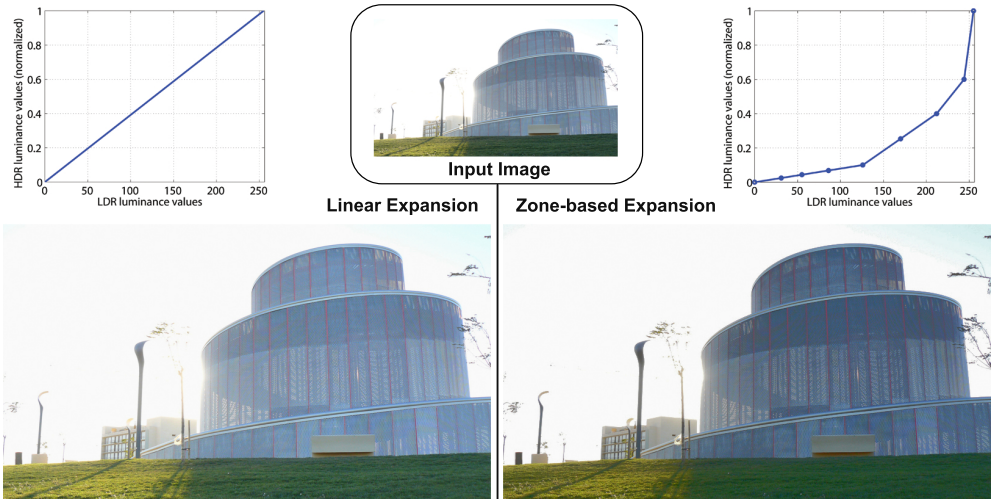


Figure 3.13: Zone-based reverse tone mapping. *Left*: HDR image obtained by linearly expanding luminance values, and corresponding expansion function. *Top center*: Original LDR image. *Right*: HDR image obtained with a piece-wise linear expansion function based on the Zone System, and corresponding graph showing this expansion function.

3.7.2 *Content-aware rTM*

As noted before, the general approach in rTM is to allocate most of the additional dynamic range that an HDR display offers to saturated areas in the scene. However, this may not always be the optimal choice. To our knowledge, none of the previous techniques have taken into consideration the *semantics* of the scene. In an image where a large region of it is saturated, such as the leaf in the snow in Figure 3.10, treating in a different way the object of interest (in this case the leaf) and the saturated background (the snow) can lead to more visually appealing results than boosting the saturated area while leaving the leaf nearly untouched. The same reasoning applies to the rest of the images in Figure 1, and in general to images which, either as a result of the artist's choice, or because of wrong exposure, contain large saturated areas. Moreover, when dealing with these type

of images, linearly expanding the dynamic range (which in general terms is the other rTM alternative offered by the literature) would result in a significant loss of visible contrast, which is a crucial characteristic of these type of images.

We therefore propose to use a higher-level approach in these cases, taking into account the content of the scene and detecting the object of interest in order to use different reverse tone mapping functions for it and for the background. To separate the region of interest from the background a saliency detector can be used.

3.7.2.1 *Detecting salient features*

Saliency detection techniques pursue the objective of detecting those regions where the viewer's attention concentrates when looking at the image. Even though it is an active field where research continues to offer new and improved methods, a series of detectors exist which are able to offer convincing results in a wide variety of images. In general, saliency detection is performed by developing more or less complex models of the human visual system and using them in combination with image metrics. Most models of attention are based on the fact that at the first stage of visual attention low-level visual features (i.e. edges, intensities, orientations) are extracted. Following this, many existing methods obtain low level features on a first stage, and on a second stage they compute saliency based on these features, as does the well-known work of Itti et al. [197]. However, for many purposes it is necessary to perform a third stage in which object segmentation is applied to extract salient *objects* instead of just a map of salient locations. In our case the need for this third stage in the saliency detection is obvious, as we look for an accurate separation between the object of interest and the background. From within the saliency detection techniques developed in the last years, we found two of them to meet our needs and applied them to our content-aware reverse tone mapping framework. They are both briefly summarized below.

LEARNING-BASED SALIENCY DETECTION. This method, introduced by Liu and colleagues [270], delivers, for each input image I, a binary saliency map $A = a_i$, where a_i takes values 1 or 0 depending on whether each pixel belongs or not to the salient object, respectively. In essence, they formulate the problem as a Conditional Random Field (CRF) in which $P(A|I)$ is inferred using a combination of salient features. Learning using a large training database is used to determine the optimal linear combination of the computed salient features.

Given an image I, whose saliency is to be computed, the objective is to obtain a binary saliency mask A. To do this $P(A|I)$ is computed as:

$$P(A|I) = \frac{1}{Z} \exp(-E(A|I)), \tag{18}$$

where Z is the partition function (equivalent to a normalizing factor) and the energy $E(A|I)$ is defined as:

$$E(A|I) = \sum_i \sum_{k=1}^{K} \lambda_k F_k(a_i, I) + \sum_{i,i'} S(a_i, a_{i'}, I). \tag{19}$$

The first term of Equation 19 corresponds to the linear combination of saliency features, so that λ_k are the coefficients which are calculated by learning and $F_k(a_i, I)$ are the K feature functions employed. As for the second term, i and i' denote two adjacent pixels, and $S(a_i, a_{i'}, I)$ is intended so that the pixels in the homogeneous inner part of the salient object are included as salient ones. The function S is thus designed so that the likelihood that two adjacent pixels are assigned different labels decreases the more similar in color the pixels are.

The feature functions F_k follow the expression:

$$F_k(a_i, I) = \begin{cases} f_k(i, I) & a_i = 0 \\ 1 - f_k(i, I) & a_1 = 1 \end{cases}, \tag{20}$$

with $f_k(i, I)$ being a different function depending on the feature computed but always taking values within the $[0, 1]$ interval. In their work, Liu et al. choose to use three different feature functions at different levels: multi-scale contrast at local level, center-surround histogram at regional level, and color spatial distribution at a global level. The experiments performed in their work show how the combination of the three yields an optimal result. Figure 3.14 (bottom row) shows an example of these feature functions used and the final saliency mask obtained from them. Training has to be performed for the CRF in order to obtain

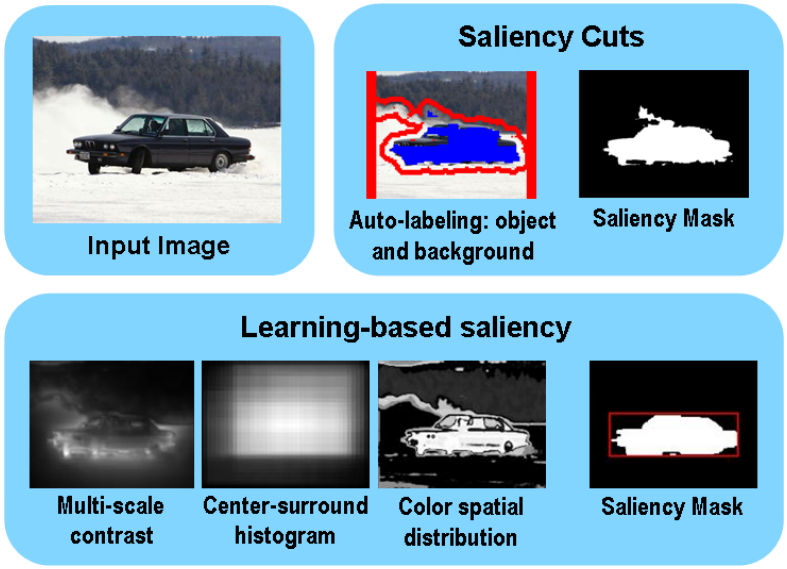


Figure 3.14: Saliency detection with the different methods. *Top left*: Input image. *Top right*: Saliency detection using the Saliency Cuts algorithm. *Bottom row*: Saliency detection using the learning-based saliency detection approach (images from the saliency database publicly available at `http://research.microsoft.com/$\sim$jiansun/`). Further details can be found in the text.

the coefficients λ_k which determine the influence of each feature function. To do this, the training set is a large database of images (ca. 21,000 images) where the salient object has been manually labeled. The obtention of λ_k is posed as a max-

imization problem where the objective function is the sum of the log-likelihood (details on how to solve the optimization problem can be found in their paper):

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \arg\max_{\boldsymbol{\lambda}} \sum_n \log P(A_j | I_j; \boldsymbol{\lambda}), \tag{21}$$

where I_j, $j = 1..N$, refer to the images in the training set and A_j to their corresponding saliency binary masks.

SALIENCY CUTS. This method, presented by Fu et al. [134] is essentially a combination of two techniques: the use of graph cuts for object segmentation [55, 380] and the spectral approach to saliency detection of Hou et al. [182].

Interactive graph cuts methods pose object segmentation as a minimal graph cuts problem. The nodes of the graph are formed by image pixels, and the two terminal nodes $\{s, t\}$ correspond to object and background, respectively. These marking of pixels as object or background by the user constitute the hard constraints on the problem, while soft constraints which take into account boundary and region information are also incorporated. The problem of finding minimal cuts in the graph is then solved via a max–flow algorithm [56].

As for the saliency detection, following the spectral residual approach to the problem the saliency map is computed as:

$$S(x) = g(x) * \mathfrak{F}^{-1}[\exp(R(f) + P(f))]^2, \tag{22}$$

where $R(f)$ is the spectral residual, obtained as $L(f) - A(f)$, $L(f)$ being the log spectrum of the input image (after downsampling it) and $A(f)$ being the general shape of log spectra. $g(x)$ is a Gaussian filter used to smooth the final saliency map, $\mathfrak{F}^{-1}$ denotes the Inverse Fourier Transform, and $P(f)$ represents the phase spectrum of the image (the reader can refer to the original paper for a comprehensive description). In the Saliency Cuts implementation this map $S(x)$ is then binarized to obtain an object saliency map $S^o(x)$. This binary saliency map, together with the auto-labeling used for the background and the salient object when performing the segmentation, can be seen in Figure 3.14 (top right).

The idea behind the Saliency Cuts framework is that even though interactive graph cuts can yield very accurate segmentations when proper priors are used, it usually requires a skillful user to select the appropriate regions to feed the algorithm. However, saliency regions detected by the algorithm by Hou and colleagues can serve as seeds to the graph cuts segmentation process, thus obtaining an automatic and accurate separation between the salient object and the background. The limitations of the method lie in the fact that they can only detect a single object and in their assumption that the salient object is at the center of the image, while the sides are always assumed to belong to the background.

3.7.2.2 *Expanding the dynamic range*

Once the division in object of interest and background has been performed, different expansion functions can be used for each. These expansion functions can

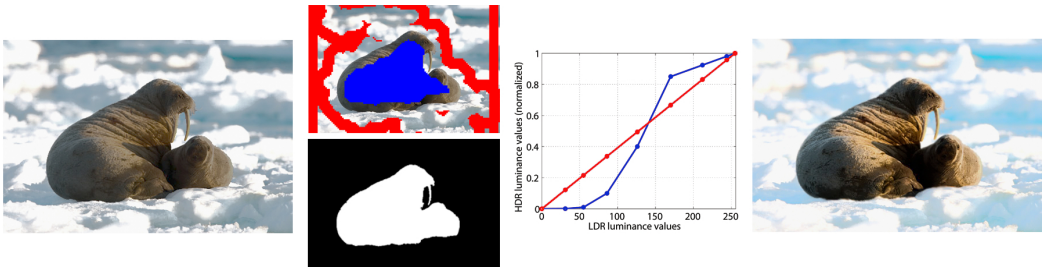


Figure 3.15: Complete pipeline using our rTM approach. *From left to right*: Input LDR image, auto-labeling of salient object (blue) and background (red) and binary saliency mask, expansion functions for the salient object (blue) and the background (red), and final HDR image. Original image copyright of National Geographic.

be of any type. Given that we are focusing on an interactive approach where the user guides the reverse tone mapping process, we choose again to use piece-wise linear functions after a separation in luminance zones as explained in Section 3.7.1. Resulting HDR images obtained with this rTM framework and the corresponding saliencies and expansion functions are shown in the Results section.

3.7.3 *Results and Discussion*

Figure 3.15 shows an example of a complete pipeline using our rTM approach, combining the two techniques described in the previous sections. The original image is segmented yielding a binary mask containing the salient object, and a division in luminance zones of the input image is performed. Next, the user can adjust the range of luminance in the HDR final image that will be assigned to each zone, both for the seals and for the background independently. This allows the user to easily manipulate the tonal balance of the image to get the best depiction. In this case a non-linear curve (shown in blue in the graph) has been applied to the seals, thus increasing their contrast and making them more salient; the snow has been just linearly expanded. Segmentation has been performed using Saliency Cuts (seeds used for the foreground and background are shown in blue and red, respectively). Even though both of the saliency detection methods presented produce segmentations accurate enough for our purposes given input images which are not excessively complex, for increasing complexity (either morphological or related to luminance values) manual segmentation may be necessary. The presence of more than two salient objects in the image also requires a manual segmentation, as the methods discussed cannot segment more than one object. In the results presented in Figure 3.16 the object of interest was segmented manually and, again, different zone-based piece-wise linear expansion functions were used for the salient object and for the background.

The interactive nature of the approach presented implies that the functions for reverse tone mapping, which determine how the high dynamic range image will look, are adjusted and tuned with low dynamic range renditions of the images as feedback. This is reasonable due to the fact that—as shown in the first part of the chapter (Section 3.4) and in [302]—the subjective quality of HDR images that have been generated from LDR images depends more on the presence of absence

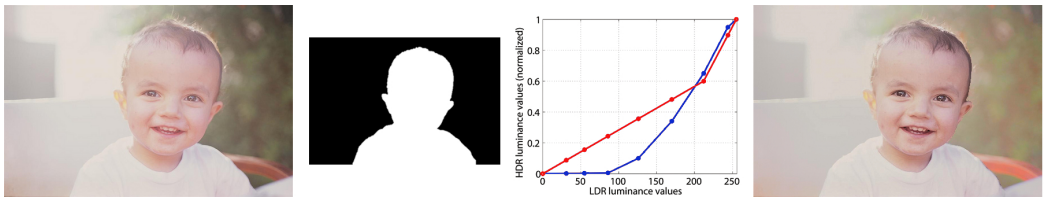


Figure 3.16: Reverse tone mapping using different zone-based expansion functions for the salient object and the background. *From left to right*: Input LDR image, manually obtained saliency mask, expansion functions for the salient object (blue) and the background (red) and final HDR image. Original image courtesy of Leandro Fessia, all rights reserved.

of spatial artifacts than on the exact luminance values, and thus a reasonably predictive evaluation of an HDR image can be done with an LDR depiction of it.

3.8 CONCLUSIONS

Previous works on the perception of HDR images and rTM design have assumed that the input images were, in general, correctly exposed. While these provide valuable knowledge that could guide the development of both HDR display hardware and reverse tone mapping algorithms, existing LDR legacy content actually covers a wide range of exposures, including material that suffers from bad exposure. As currently designed, existing rTMOs tend to boost over-exposed areas more than the rest of the image. The strategy works well for small areas such as light sources or highlights if the rest of the image is correctly exposed, but no performance evaluation on generally over-exposed imagery had been performed.

Our experiments confirmed that performance of rTMOs decreases for over-exposed input images, suggesting that for bright images the consensual approach of boosting bright areas could be improved.We have shown that a rTMO based on γ expansion can outperform existing rTMOs in these cases. Further, our findings indicate seem to indicate that superior rTMOs should take into account global statistics about the image, and not just individual pixel values. We have derived a simple strategy for the expansion based on image statistics, but more sophisticated strategies could also be devised, possibly including high-level semantics.

With the exception of the approach presented in Section 3.7, we do not aim to create new depictions of LDR material, which would potentially interfere with the original intentions and artistic vision. Our goal is much like that of an audio mastering engineer: we wish to increase the illusion of power, presence and fidelity in the final display medium, while preserving the author's original vision of the content. Our results complement those in the work by Akyüz et al. [12], where the authors show that, for *correctly exposed* imagery, a simple linear expansion works well and suggest that sophisticated treatment of LDR data may not be necessary. In fact, our work is consistent with that of Akyüz et al. [12] in the sense that our proposed γ curves approach linear scaling when the image is

approximately correctly exposed. Together, both studies suggest that potentially complex operators might not be needed.

In Section 3.7 have presented an interactive approach to reverse tone mapping which can be useful for a wide variety of images, especially those containing large saturated areas. The basis of our method is inspired by photographer Ansel Adams' well-known Zone System, which allows us to divide the luminance range of the image into zones. With the aid of this division in zones, and in an interactive process, a piece-wise linear function to expand the LDR image can be provided by the user. Furthermore, our technique includes the possibility of using higher-level information as a guide for the expansion, segmenting the image in the object of interest and the background and using different expansion functions for each. This interactive approach offers a tool to expand the dynamic range of a scene with significant yet intuitive control over the final result. Besides, being able to freely adjust the luminance ranges of the zones makes it possible to obtain very different HDR depictions of the same input image, potentially providing an artistic tool for photographers and artists in general.

Regarding future work in this interactive part, adding a fitting step of the piecewise linear rTM functions proposed to smoother ones would be desirable. In the same sense, when dealing with content-aware rTM, taking care of the luminance transitions in the boundary between the objects of interest and the backgrounds would be necessary, either by somehow smoothing the binary mask or by placing constraints to the relationship between both -the object's and the background's- expansion functions. It would also be interesting to work in Yxy color space instead of RGB to automatically keep ratios between color channels constant. Besides, thorough comparison between the proposed rTM technique and existing reverse tone mapping operators by means of psychophysical experiments would certainly be interesting for the field. Finally, salient object detection is an open field of research, and our approach would definitely benefit from future advances in this field. Other lines of future research could involve the design of a contrast-based rTMO, following the findings of the work by Mantiuk et al. [292], which shows promising results in the field of contrast processing of HDR images, working in visual response space.

The conclusions drawn aim to be valuable for further development of HDR display technology, HDR imaging in general and the development of future LDR expansion algorithms in particular. However, further tests on LDR expansion are desirable. As the community investigates this issue further, this and similar studies will surely be extended and updated. Additionally, reverse tone mapping for video content is a key challenge in this field. In order to develop operators that gracefully handle changes in exposure over time, it is crucial to first understand how they fail in the static case.

Part III

COMPUTATIONAL DISPLAYS

This part begins with a survey on computational displays, where the different technologies and algorithms are categorized along the dimensions of the plenoptic function. We then propose a disparity remapping method to deal with the problem of depth of field in automultiscopic displays. Further extensible to stereoscopic displays, this method leverages knowledge from computational models of perception of luminance-contrast and depth. Finally, we focus on comfort when viewing stereo content in motion, and present a set of comprehensive measurements of comfort as a function of a number of parameters, and a metric to assess the degree of comfort for short clips of video.

4

A SURVEY ON COMPUTATIONAL DISPLAYS

ABOUT THIS CHAPTER

This chapter is a review of the state of the art in the field of computational displays, which has undergone great progress over the last few years. One of the main strengths of this survey, we believe, lies in how it is organized: we analyze the advances in the field and categorize them along the dimensions of the plenoptic function. Additionally, the survey pays special attention to the aspects of human perception which are leveraged, or could potentailly be leveraged, to enhance display capabilities and improve the viewing experience. The survey, although led by myself, has been done as a collaboration between Gordon Wetzstein, from MIT Media Lab, Piotr Didyk, from MIT CSAIL, and Diego Gutierrez, with expertise in different areas within the displays field. My focus has been on sections 4.1, 4.2, 4.3, part of Section 4.7 and part of 4.8. The work has been accepted for a special issue on Advanced Displays of the journal Computers & Graphics.

B. Masia, G. Wetzstein, P. Didyk and D. Gutierrez.
A Review on Computational Displays: Pushing the Boundaries of Optics, Computation, and Perception.
Computers & Graphics 2013, *to appear*.

4.1 INTRODUCTION

In 1692, French painter Gaspar Antoine de Bois-Clair introduced a novel technique that would allow him to paint so-called *double portraits*. By dividing each portrait into a series of stripes carefully aligned behind vertical occluding bars, two different paintings could be seen, depending on the viewer's position with respect to the canvas. Figure 4.1 shows the double portrait of King Frederik IV and Queen Louise of Mecklenburg-Gstow [406]. Later, Frederic Ives patented in 1903 what he called the *parallax stereogram*, based also on the idea of placing occluding barriers in front of an image to allow it to change depending on viewer's position [198]. Five years later, Gabriel Lippmann proposed using a lenslet array instead, an approach he called *integral photography* [265].

Both parallax barriers and lenslet arrays shared a common objective: to provide different views of the same scene or, more technically, to increase the range and resolution of the angular dimension(s) of the *plenoptic function*. The plenoptic function [6] represents light observed from every position in every direction, i.e. a complete representation of the light in a scene. It is a multidimensional function that includes information about intensity, color (wavelength), time, position and viewing direction (angle). The previously mentioned techniques, for instance, allow to increase the angular resolution at the cost of reducing the spa-

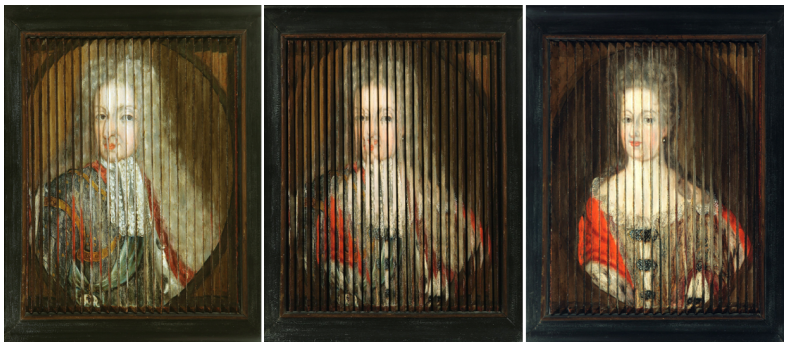


Figure 4.1: Double portrait by Gaspar Antoine de Bois-Clair, as viewed from the left, center and from the right (images courtesy of Robert Simon [406]).

tial resolution (the same image area needs now to be shared between several views); an additional cost is reduced intensity, since parallax barriers block a large amount of light.

Over the past few years we have seen large advances in display technology. These have motivated surveying papers on related topics such as real-time image correction techniques for projector-camera systems [50], parallax capabilities of 3D displays [179, 45], or specialized courses focused on emerging compressive displays in top conferences [477, 483], to name just a few.

In this survey, we provide a holistic view of the field, mainly from a computer graphics perspective, and categorize existing works according to which particular dimension(s) of the plenoptic function is enhanced. For instance, high dynamic range displays improve intensity (luminance) contrast, while automultiscopic displays expand angular resolution. We further note that the recent progress in the field has been spurred by the *joint design* of hardware and display optics with computational algorithms and perceptual considerations. Thus, we identify perceptual aspects of the human visual system (HVS) that are being used by these technologies to yield an *apparent* enhancement, beyond the physical possibilities of the display. Examples of these include wobbling displays, providing higher spatial resolution by retinal integration of lower resolution images, or the apparent increased intensity of some pixels caused by the glare illusion.

Therefore, we provide a novel view of the recent advances in the field taking the plenoptic function as a supporting structure (see Figure 4.2) and putting an emphasis on human visual perception. For each section, each focusing on a dimension of the plenoptic function, we first present perceptual foundations related to that dimension, and then describe display technologies, and software solutions for the generation of content in which the specific dimension being discussed is enhanced. Specifically, we first address expansion on dynamic range in Section 2, followed by color gamut (Section 3), increased spatial resolution (Section 4), increased temporal resolution (Section 5) and finally increased angular resolution, for both stereo (Section 6) and automultiscopic displays (Section 7).

For topics where there is a large body of existing literature, beyond what can be reasonably covered by this survey, we highlight some of the main techniques and suggest alternative publications for further reading (such as tone mapping or superresolution techniques). For other related aspects not covered here, such

as detailed descriptions of hardware, electronics or the underlying physics of the hardware, we refer the interested reader to other excellent sources [272, 155]. Finally, although projection-based display systems are included in this survey whenever they focus on enhancing the aspects of the plenoptic function, there are a number of works which fall out of the scope of this survey. These include works dealing with geometric calibration (briefly discussed in Section 4.3.3), or extended depth-of-field projection [147, 278]. We refer the interested reader to existing books focused on projection systems [48, 50, 284].

4.2 IMPROVING CONTRAST AND LUMINANCE RANGE

The *dynamic range* of a display refers to the ratio between the maximum and minimum luminance that the display is capable of emitting [366]. The advantages and improved quality of High Dynamic Range (HDR) images are by now well established. By not limiting the values of the red, green and blue channels to the range $[0..255]$, physically-accurate photometric values can be stored instead. This yields much richer depictions of the scenes, including more detail in dark areas and avoiding saturated pixels (Figure 4.3).

Many applications can benefit from HDR data, including image-based lighting [97], image editing [219] or medical imaging [49]. The field has been extensively investigated, especially in the last decade, and several excellent books exist offering detailed explanations on related aspects, including image formats and encodings, capture methods, or quality metrics [366, 37, 174, 325].

4.2.1 *Perceptual Considerations*

There are two types of photoreceptors in the eye: cones and rods. Each of the three cone types is sensitive in a wavelength range, the sensitivity of each type peaking at a different wavelength, roughly belonging to red, green and blue; combined, they allow us to see color. They are most sensitive to *photopic* (day light) luminance conditions, usually above 1 cd/m^2, while rods (of which only one type exists) are most sensitive to *scotopic* (night light) conditions, approximately below 10^{-3} cd/m^2. The bridging range where both cones and rods play an active role at the same time is called the *mesopic* range (see Figure 4.4).

On the other hand, luminance values in natural scenes (from moonless night sky to direct sunlight from a clear sky) may span about 12 to 14 orders of magnitude, although simultaneous luminance values usually fall within a more restricted range of about four to six orders of magnitude (for a more exhaustive discussion on luminance ranges in natural scenes the reader may refer to [493]). The HVS can perceive only around four orders of magnitude simultaneously, but it uses a process known as dynamic *adaptation*, effectively shifting its sensitive range to the current illumination conditions [366, 462, 309].

Despite this ability to adapt across a wide dynamic range, our ability to discern local scene contrast boundaries is reduced by veiling caused by light scattering inside the eye (an effect known as veiling glare, or disability glare). Many other luminance-related factors affect our visibility, including the intensity of the background (Weber's law) and the spatial frequency of the stimuli, whose depen-

	Contrast and Luminance Range	Color Gamut	Spatial Resolution	Temporal Resolution	Angular Resolution I - Stereo	Angular Resolution II - Automultiscopic
Perceptual Considerations	Dynamic range of the eye (photopic-mesopic-scotopic vision), dynamic adaptation, CSF, Craik-O'Brien-Cornsweet illusion, visual masking	Dual-process theory, trichromatic & color-opponent stages, chromatic adaptation, standardized observers	Photoreceptor density, foveal and peri-foveal vision, SPEM	Temporal integration, Bloch's law, CFF, hold-type blur, SPEM	Panum's fusional area, zone of comfort, acc.-vergence conflict, DSF, Craik-O'Brien-Cornsweet illusion for depth, disparity models	Disparity models, motion parallax, accommodation, cue integration
Display Architectures	Two-layer optical modulation	Increasing the purity of primaries	Optical superposition	Backlight flashing	Spatial multiplexing (anaglyph, polarization)	Volumetric displays
	Local dimming	Multi-primary displays	Temporal superposition		Temporal multiplexing (shutter glasses)	Light field displays
	Multi-projector systems	Projection systems	Optical Pixel Sharing		Backward-compatible stereo	Compressive light field displays
					Autostereoscopic	Light field displays supporting accommodation
Content Generation	HDR image/video acquisition	Color Appearance Models	Super-resolution	Black data insertion	Camera parameters adjustment	Efficient image synthesis
	Apparent brightness enhancement	Gamut mapping	Sub-pixel rendering	Frame interpolation techniques	Disparity remapping	Light field retargeting
	Tone mapping	Radiometric calibration	Temporal integration	Motion compensated inverse filtering	Microstereopsis and Backward-compatible stereo	Disparity remapping
	Reverse tone mapping			Warping techniques		Multiview from stereo
				Leveraging information from the rendering pipeline		
				Mesh-based temporal upsampling		

Figure 4.2: Overview of architectures and techniques, according to the structure followed in this survey. Columns in the table correspond to dimensions of the plenoptic function, and to the sections of this chapter. For each section, we first discuss relevant perceptual aspects (related keywords are shown in the first row of this table), then describe display architectures (middle row), and finally present software solutions and content generation approaches aimed at improving the corresponding dimension of the plenoptic function (third row).



Figure 4.3: Low dynamic range depictions of a high dynamic range scene, showing large saturated (left) or dark areas (right).

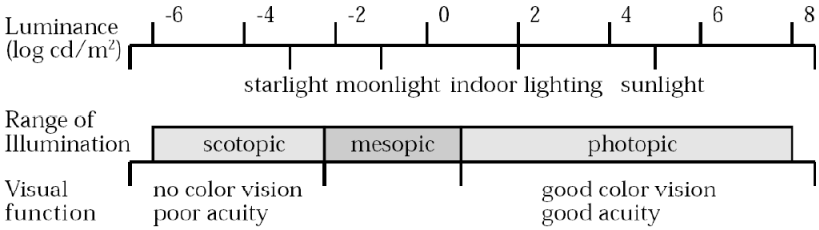


Figure 4.4: Scotopic, mesopic and photopic vision, corresponding to different luminance levels [25].

dency is modeled by the contrast sensitivity function (CSF, see Figure 4.5, right); the bleaching of photoreceptors when exposed to high levels of luminance, which translates into a loss of spectral sensitivity [122]; the Craik-O'Brien-Cornsweet illusion, by which adjacent regions of equal luminance are perceived differently depending on the characteristics of their shared edges [228]; or the effect known as visual masking, where contrast sensitivity loss is induced due to the presence of signal in nearby regions [25]. Researchers have also investigated perceptual aspects of increased dynamic range, including analyzing the subjective preferences of users, to improve HDR display technology [503, 390, 12].

4.2.2 *Display Architectures*

Traditional CRT displays typically show up to two orders of magnitude of dynamic range: Analog display signals are typically 8-bit because, even though a CRT display could reproduce higher bit-depths, it would be including values at levels too low for humans to perceive [366]. LCD displays, although brighter, do not significantly improve that range. HDR displays enhance the contrast and luminance range of the displayed images, thus providing a richer visual experience. A passive HDR stereoscopic viewer overlaying two transparencies was presented by Ledda et al. [256]. Seetzen et al. [392, 389] presented the first two active prototypes, which set the basis for later models that can be now found in the market (Figure 4.5, left). The two prototypes shared the key idea—illustrated in Figure 4.5, center—of optically modulating a high spatial resolution (but low dynamic range) image with an LCD panel showing a grayscale, low spatial resolution (but high intensity) version of the same image. This provides a theoretical

contrast equal to the multiplication of both dynamic ranges. Alternatively, two parallel-aligned LCD panels of equal resolution can be used [377]. A detailed description of the first prototypes and the concept of *dual modulation* of light can be found in Seetzen's PhD. Thesis [391].

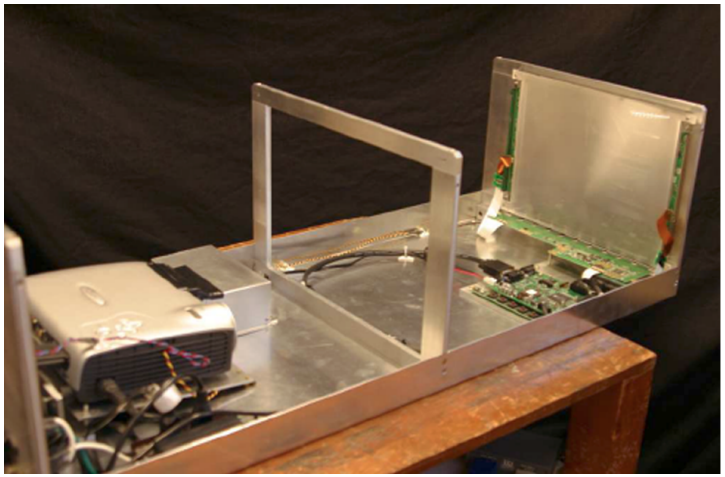
Commercially available displays with increased contrast are mostly based on *local dimming*. This name refers to the particular case of dual modulation in which one of the modulators has a significantly lower resolution than the other [366]. This arises from knowledge of visual perception, and in particular of the effect known as veiling glare. Due to veiling glare, the contrast that can be perceived at a local level is much lower than at a global level, meaning there is no need to have very large local contrast, and thus a lower resolution panel can be used for one of the modulators. The drawback is potentially perceivable halos, whose visibility depends on the particular arrangement of the LED array.

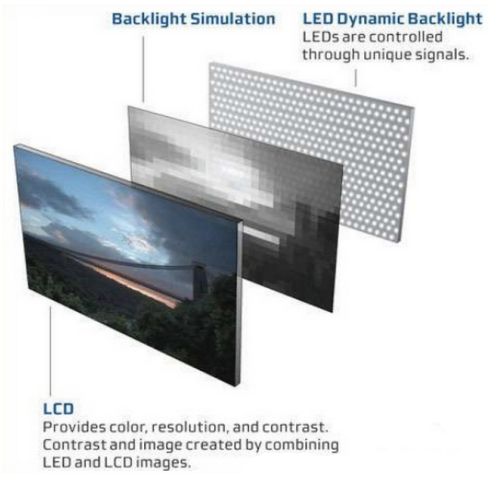
Projector-based systems exist, also based on the principle of double modulation. Majumder and Welch showed how by overlapping multiple projectors, the intensity range (difference between highest and lowest intensity levels; note that is different from contrast, which is defined as the ratio) could be increased [286]. The first contrast expansion technique was proposed by Pavlovych and Stuerzlinger [344], where a small projected image is first formed by a set of lenses, which is subsequently modulated by an LCD panel. A second set of lenses enlarges the final image. Other similar approaches exist, making use of LCD or LCoS panels to modulate the illumination [248, 92]. Multi-projector tiled displays present another problem in addition to limited dynamic range: Brightness and color discontinuities at the overlapping projected areas [475]. Majumder et al. [283] rely on the contrast sensitivity function to achieve a seamless integration with enhanced overall brightness. Recently, secondary modulation of projected light has also been used to boost contrast of paper images and printed photographs [49] (see Figure 4.6).

4.2.3 *HDR Content Generation and Processing*

Contrast and accurate depiction of the dynamic range of real world scenes have been a key issue in photography for over a century (see for instance the work of Ansel Adams [3]). The seminal works by Mann and Picard [289], and by Debevec and Malik [98], brought HDR imaging to the digital realm, allowing to capture HDR data by adapting the multi-bracketing photographic technique. More sophisticated acquisition techniques have continued to appear ever since (see [481] for a compilation), helping for instance to reduce ghosting artifacts in dynamic scenes [397, 510, 184] (see [154] for a recent review on deghosting techniques), using computational photography approaches [381, 330], mobile devices [2, 117] or directly capturing HDR video [439, 325, 450, 244].

Regarding the visualization of such HDR content, we distinguish three main categories: Tone-mapping, by which high dynamic range is scaled down to fit the capabilities of the display; reverse tone mapping, by which low dynamic range is expanded for correct visualization on more modern higher dynamic range displays; and apparent brightness enhancement techniques, which leverage how





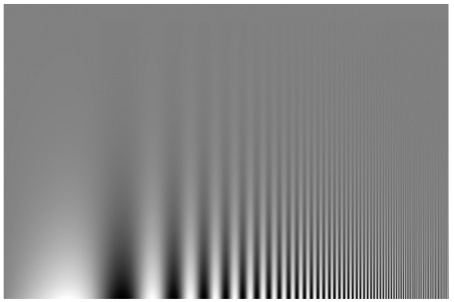


Figure 4.5: Left: The first HDR prototype display, employing dual modulation [389]. Center: Scheme illustrating the functioning of dual modulation. Right: The contrast sensitivity function, represented by a Campbell-Robson chart [365]; the abscissa corresponds to increasing spatial frequencies, the ordinate to decreasing amplitude of the contrast. The chart shows that the sensitivity of the HVS to contrast depends on the spatial frequency of the signal, and follows an inverted U-shape.

our brains interpret some specific luminance cues and translate them into the perception of brightness (but the actual dynamic range remains unchanged).

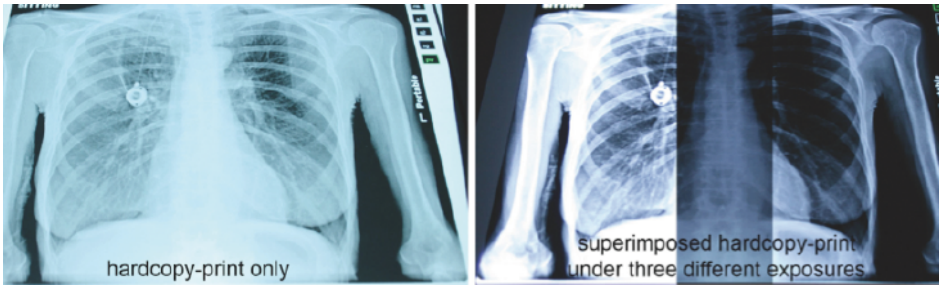


Figure 4.6: Superimposing dynamic range for medical applications. Left: a single hardcopy print. Right: expanded dynamic range by superimposing three different prints with different exposures [49].

Tone Mapping. Over the past few years, many user studies have been performed to understand which tone mapping strategies produce the best possible visual experience [257, 503, 390, 291]. The field has been extremely active over the past two decades, with a proliferation of many algorithms which can be broadly characterized as global or local operators. While a complete survey of all existing tone mapping operators is out of the scope of the work, the interested reader can refer to other sources of information, where many of these algorithms are discussed, categorized and compared [99, 366, 37, 67].



Figure 4.7: Tone mapping allows for a better visualization of HDR images on displays with a limited dynamic range. Left, naive visualization with a simple linear scaling; Right, the result of Reinhard's photographic tone reproduction technique [364].

Global operators apply the same mapping function to all the pixels in the image, and were first introduced to computer graphics by Tumblin and Rushmeier [445]. They can be very simple, although they may fail to reproduce fine details in areas where the local contrast cannot be maintained [471, 388]. To provide results that better simulate how real-world scenes are perceived, usually some perceptual strategies are adopted, based on different aspects of the HVS [132, 470, 111, 362]. Usually these perceptually motivated works rely on techniques like multi-scale representations, transducer functions, color appearance models or retinex-based algorithms [292].

Local operators, on the other hand, tone-map each pixel taking into account information from the surrounding pixels, and thus usually allow for better preservation of local contrast [78]. The main drawback is that the local nature of the algorithms may give rise to unpleasant halos around some edges [366]. Again, perceptual considerations can be introduced in their design to reduce visible artifacts [342, 242]. Other strategies include adapting well known analog tone reproduction techniques from photography [364] (Figure 4.7), while others take into account the temporal domain, being especially engineered for videos [343].

Other operators work from a different perspective, for instance by working in the gradient domain [129] or in the frequency domain [115]. The exposure fusion technique [313] circumvents the need to obtain an HDR image first and *then* apply a tone mapping operator. Instead, the final tone-mapped image is directly assembled from the original multi-bracketed image sequence, based on simple, pixel-wise quality measures. Last, the work by Mantiuk et al. [293] derive a tone mapping operator that takes explicitly into account the different displays and viewing conditions the images can be viewed under.

Reverse Tone Mapping. Somewhat less studied is the problem of *reverse* tone mapping, where the goal is to take LDR content and expand its contrast to recreate and HDR viewing experience (Chapter 3 in this thesis is devoted to this topic). It is gaining importance as more and more HDR displays reach the market, given the large amount of LDR legacy content. Reverse tone mapping involves dealing with clipped data, which makes it a slightly different problem from tone mapping (see Figure 3.1). As before, a number of studies have been recently conducted to understand what the best strategy for dynamic range expansion may be [12, 296, 302, 34, 369].

The first methods were presented by Daly and Feng, and included bit-depth extension techniques [90] followed by techniques to solve subsequent problems such as contour artifacts [91]. Subsequent works have appeared over the years, usually following the approach of identifying the bright areas in the input image and expanding those the most, leaving the rest moderately (if at all) expanded to prevent noise amplification [30, 316, 368, 240, 39]. Other methods require direct user input [101, 465, 303]. Banterle and colleagues proposed one of the first extensions for video [35], while Masia et al. analyzed the problem across varying exposure conditions [302, 301]. In their work, the authors additionally found that the perceived quality of the expanded images depends more on the absence of disturbing spatial artifacts than on the exact contrasts in the image. A more exhaustive presentation on the topic of reverse tone mapping can be found in the recent book by Banterle et al. [37].

Apparent brightness enhancement. A strategy to increase the *apparent* dynamic range of the displayed images is to directly exploit some of the mechanisms of the HVS, and how our brains interpret some luminance cues, and translate them into the perception of brightness. For instance, we have mentioned how some tone mapping operators introduce unwanted halos, that are perceived as artifacts. However, halos have been used for centuries by painters, to create steeper luminance gradients at the edges of objects and increase local image contrast. This technique is known as *countershading*, and it resembles the *unsharp masking* operator, which increases local contrast by adding a high-pass-filtered version

of the image [412, 273, 371, 243]. The potential benefits and drawbacks of this technique have also been recently investigated in this context [444].

Another example is the *bleaching effect*, which was first utilized by Gutierrez and colleagues to both increase apparent brightness of light sources and simulate the associated perceived change of color [149, 152]. The temporal domain was subsequently added, allowing for the simulation of time-varying afterimages [370] (see Figure 4.8). Synthetic glare has also be added around bright light sources in the images, to simulate scattering (both in the atmosphere and in the eye) and thus enhance brightness [502, 372]. Last, binocular fusion has been used by showing two different low dynamic range depictions of the same HDR input image on a binocular display. The fused image presents more visual richness and detail than any of the single LDR versions [496] (Figure 4.9).

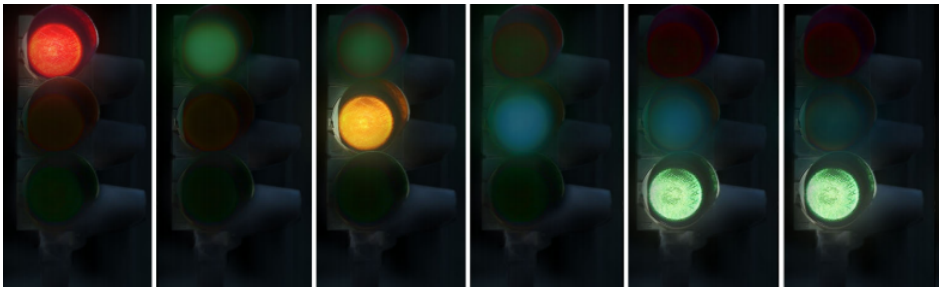


Figure 4.8: Afterimage simulation of a traffic light, showing variations over time of color, degree of blur and shape [370].



Figure 4.9: Binocular tone mapping. The fused image presents more detail than any of the two individual, low dynamic range depictions [496].

4.3 IMPROVING COLOR GAMUT

In 1916 the company Technicolor was granted a patent for “a device for simultaneous projection of two or more images" [82] which would allow the projection of motion pictures in color. Although not the only color film system, it would

be the system primarily used by Hollywood companies for their movies in the first half of the 20th century. Color television came later, starting in 1950 in the United States (although NTSC was not introduced until 1953), and not reaching Europe until 1967 (PAL/SECAM systems). Several standards are in use today, among which YCbCr is the ITU-R recommendation for HDTV (high definition television, with a standard resolution of 720p or 1080p). Until today, the quest to reproduce the whole color range that our visual system can perceive continues.

4.3.1 *Perceptual Considerations*

The *dual-process theory* is the commonly accepted theory that describes the processing of color by the HVS [365]. The theory states that color processing is performed in two sequential stages: a trichromatic stage and an opponent-process stage [189]. The trichromatic stage is based on the theory that any perceivable color can be generated with a combination of three colors, which correspond to the three types of color-perceiving photoreceptors of our visual system (see Section 4.2.1). In the opponent-process stage the three channels of the previous stage are re-encoded into three new channels: a red-green channel, a yellow-blue channel, and a non-opponent channel for achromatic responses (from black to white). These theories, originally developed by psychophysics, are confirmed by neurophysiological results.

The theories which have been mentioned describe the behavior of the HVS for isolated patches of color, and do not take into account the influence of surrounding factors, such as environment lighting. *Chromatic adaptation* (or color constancy), for instance, is the mechanism by which our visual system adapts to the dominant colors of illumination. There are many other mechanisms and effects that play a role in our perception of color, such as simultaneous contrast, the brightness of colors, image size or the luminance level of the surroundings, and many experiments have been carried out to try to quantify them [275, 188, 123, 318, 324]. Recently, edge smoothness was also found to have a measurable impact on our perception of color [225, 61]. Further, color perception has a large psychological component, making it a challenging task to measure, describe or reproduce color. So-called "standardized" observers exist [365, 195], based on measurements of a set of observers, and are used as a reference for display design, manufacturing or calibration.

4.3.2 *Display Architectures*

Increasing the color gamut of displays is typically achieved by using more saturated primaries, or by using a larger number of primaries. The former essentially "pushes further" the corners of the triangle defining the color gamut in a three-primary system (Figure 4.10, left). An alternative technique consists of using negative values for the RGB color signals (Figure 4.10, right). Emitting elements with a broad spectral distribution, as is the case of phosphors in CRTs, severely limit the achievable gamut. Research has been carried out to improve the color gamut of these types of displays [222], but for the last two decades liquid crystal displays have been the most common display technology due to their advantages

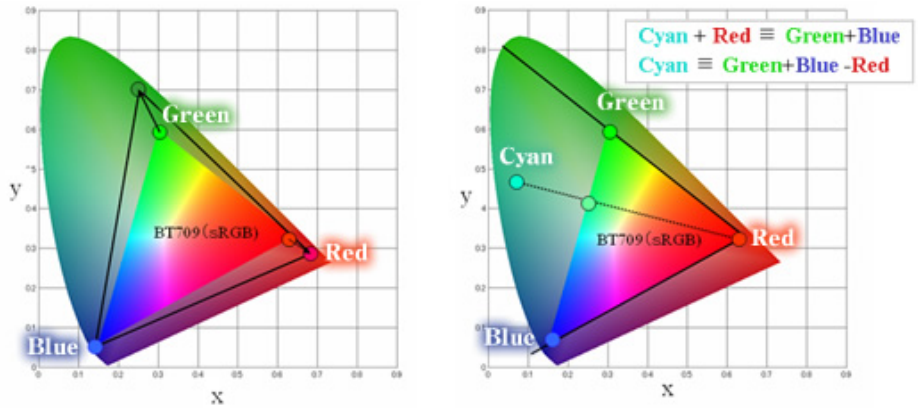


Figure 4.10: Left: Color gamut expanded by raising the saturation of the primary colors. Right: Color gamut expanded by using negative values for the color signals [418] (images copyright of Sony).

over CRTs [402, 312]. Progressively, the traditional CCFL (cold cathode fluorescent lamp) backlights used in these displays are being substituted by LED backlights due to the lower power consumption and the wider color gamuts they can offer because of the use of saturated primaries [413, 213]. LEDs also have some drawbacks, mainly the instability of their emission curves, which can change with temperature, ageing or degradation; color non-uniformity correction circuits are needed for correct color calibration in these displays [425, 430, 426]. Seetzen et al. [393] presented a calibration technique for HDR displays to help overcome degradation problems of the LEDs that cause undesirable color variations in the display over time. Their technique can additionally be modified to extend it to conventional LCD displays. Within this trend of obtaining more pure primaries, lasers have been proposed as an alternative to LEDs due to their extremely narrow spectral distribution, yielding displays that can cover the gamuts of the most common color spaces (ITU-R BT.709, Adobe RGB) [300], or a display offering a color gamut that is up to 190% the color gamut of ITU-R BT.709 [417, 428, 429, 234].

Multiple primary displays result in a color gamut that is no longer triangular, and can cover a larger area of the perceivable horseshoe-shaped gamut. Ultra wide color gamut displays using four [77], five [447, 75], and up to six color primaries [497, 427, 431] have been proposed. Multi-primary displays based on projection also exist [9, 378, 62, 379].

4.3.3 *Achieving Faithful Color Reproduction*

Tone reproduction operators (see Section 4.2) can benefit from the application of color appearance models, to ensure that the chromatic appearance of the scene is preserved for different display environments [11]. Several color appearance models (CAMs) have been proposed, with the goal of predicting how colors will be perceived by an observer [320, 123, 246]. In fact, it has been recently argued that tone reproduction and color appearance, traditionally treated as different problems, could be treated jointly [363] (Figure 4.12. Usually, simple post-processing

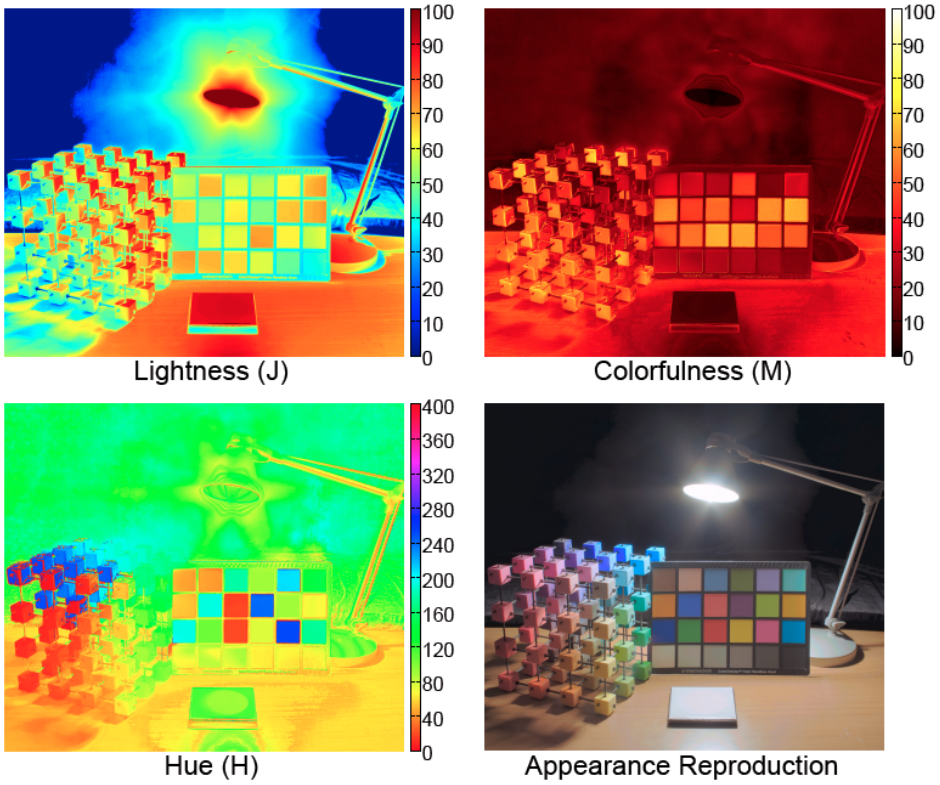


Figure 4.11: Color appearance of a high dynamic range image, based on predicted lightness, colorfulness and hue [224].

steps are performed to correct for color saturation [388, 290]. However, most color appearance models work under a set of simplified viewing conditions; very few, for instance, take into account issues associated with dynamic range. A few notable exceptions exists, such as iCAM [124, 125] or the subsequent iCAM06 [245]. Recently, Kim et al. developed a model of color perception based on psychophysical data across most of the dynamic range of the HVS [224] (Figure 4.11), while Reinhard and colleagues proposed a model that adapts images and video for specific viewing conditions such as environment illumination or display characteristics [367], as shown in Figure 4.13.

From the whole range of colors perceivable by our visual system, only a subset can be reproduced by existing displays. The sRGB color space, which has been the standard for multimedia systems, works well with e.g. CRT displays but falls short for wider gamut displays. In 2003 the scRGB, an extended RGB color space, was approved by the IEC [191], and the extended color space xvYCC [190] followed, which can support a gamut which is nearly double that supported by sRGB.

Faithful color reproduction on devices with different characteristics requires gamut manipulation, known as *gamut mapping*. Gamut mapping can refer both to gamut reduction or expansion, depending on the relationship between the original and target color gamuts [321]; these can further be given by a device or by the content. An example of the latter is the case of image-dependent gamut mapping, where the source gamut is taken from the input image and an optimization is used to compute the appropriate mapping to the target device [139].

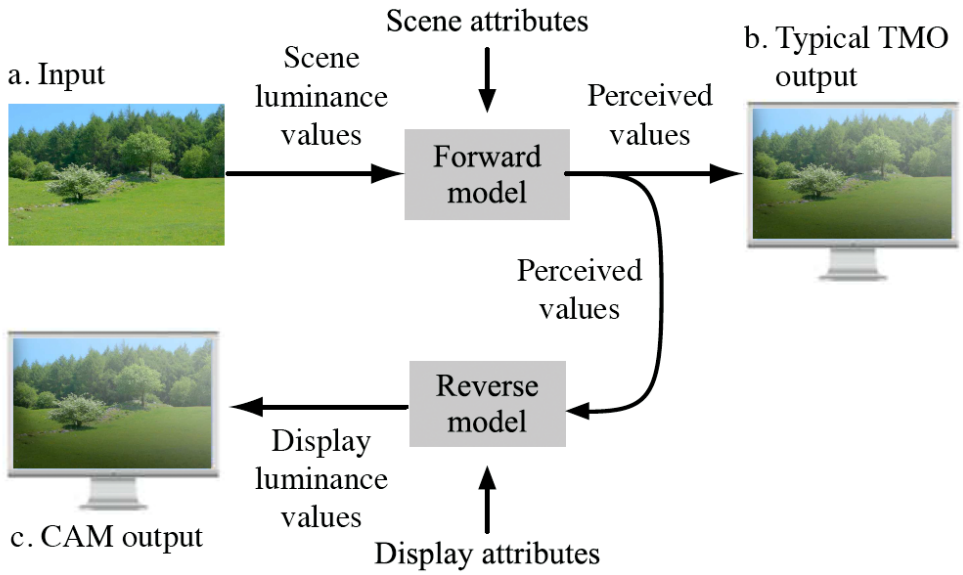


Figure 4.12: Typical processing paths for tone reproduction algorithms and color appearance models (CAMs) [363].

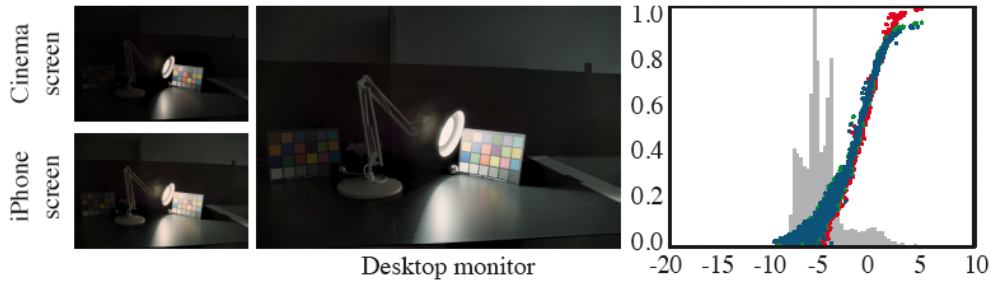


Figure 4.13: Accurate color reproduction, taking into account both display type and viewing conditions (shown here for cinema screen, iPhone and a desktop monitor). The plot shows the image histogram in gray, as well as the input/output mapping of the three color channels [367].

Gamut expansion can be done automatically [69, 162], or manually by experienced artists. The work of Anderson et al. [17] combines both approaches: an expert expands a single image to meet the target display's gamut and a color transformation is learned from that expansion and applied to each frame of the content. The reader may refer to the work by Muijs et al. [323] and by Laird et al. [250] for a description and evaluation of gamut extension algorithms, or to the comprehensive work of Morovič for a more general view on gamut mapping and color management systems [322]. Finally, the concept of display-adaptive rendering was introduced by Glassner et al. [140], applicable to the inverse case of needing to compress color gamut of content to that of the display. Instead of compressing color gamut as a post-process operation on the image [138, 319], they propose to automatically modify scene colors so that the rendered image matches the color gamut of the target display.

Accurate reproduction of color is particularly challenging for projection systems, specially if the projection surface properties are unknown and/or the image is not being displayed on a projection-optimized screen. *Radiometric calibra-*

tion is required to faithfully display an image in those cases. Typically, projector-camera systems are used for this purpose. These compensation is of special importance in screens with spatially varying reflectance [327, 146]. Some authors have incorporated models of the HVS to radiometrically compensate images in a perceptual way, i. e.minimizing *visible* artifacts [463], while others incorporate knowledge of our visual system by computing the differences in perceptually uniform color spaces [22]. Conventional methods usually assume a one-to-one mapping between projector and camera pixels, and ignore global illumination effects, but in the real world there can be surfaces where these effects have a significant influence (e. g.presence of transparent objects, or complex surfaces with interreflections). Wetzstein and Bimber [479] propose a calibration method which approximates the inverse of the light transport matrix of the scene to perform radiometric calibration in real time and being able to account for global illumination effects. These works on radiometric compensation often also deal with geometric correction. Geometric calibration compensates, often by warping the content, for the projection surface being non-planar. An option is to project patterns of structured light onto the scene, as done by e.g. Zollmann and Bimber [511]; an alternative is to utilize features of the captured distorted projection, first introduced by Yang and Welch [495]. Geometric calibration for projectors is out of the scope of this survey, but we refer the interested reader to the book by Majumder and Brown [284].

4.4 IMPROVING SPATIAL RESOLUTION

High spatial definition is a key aspect when reproducing a scene. It is currently the main factor that display manufacturers exploit (with terms such as Full HD, HDTV, UHD, referring to different, and not always strictly defined, spatial resolutions of the display), since it has been very well received among customers. So-called 4K displays, i.e. those with a horizontal resolution of around 4,000 pixels, are already being commercialized, although producing content at such high resolution has now become an issue due to storage and streaming problems; we describe existing approaches in terms of content generation in Section 4.4.3.

4.4.1 *Perceptual Considerations*

Of the two types of photoreceptors in the eye (see Section 4.2.1), cones have a faster response time than rods, and can perceive finer detail. The highest visual acuity in our retina is achieved in the fovea centralis, a very small area without rods and where the density of cones is largest. According to Nyquist's theorem, assuming a top density of cones in the fovea of 28 arc seconds [87], this concentration of cones allows an observer to distinguish one-dimensional sine gratings of a resolution about 60 cycles per degree [102]. Additionally, sophisticated mechanisms of the HVS enhance this resolution, achieving *visual hyperacuity* beyond what the retinal photoreceptors can resolve [476]. In comparison, the pixel size of a typical desktop HD display (a 120 Hz Samsung SyncMaster 2233, 22″), when viewed at a distance of half a meter, covers approximately nine cones [102]. The peri-foveal region is essentially populated by rods; these are responsible for pe-

ripheral vision, which is much lower in resolution. As a consequence, our eyes are only able to resolve with detail the part of a scene which is focused on the fovea; this is one of the reasons for the saccadic movements our visual system performs. Microsaccades are fast involuntary shifts in the gazing direction that our eyes perform during fixation. It is commonly accepted that they are necessary for human vision: if the projection of a stimulus on the retina remains constant the visual percept will eventually fade out and disappear [350].

On the contrary, if the stimulus changes rapidly, the information will be fused in the retina by temporal signal integration [214]. Related to this, the *smooth pursuit eye motion* (SPEM) mechanism in the HVS allows the eyes to track and match velocities with a slowly moving feature in an image [241, 298, 249]. This tracking is almost perfect up to 7 deg/s [249], but becomes inaccurate at 80 deg/s [89]. This process stabilizes the image on the retina and allows to perceive sharp and crisp images.

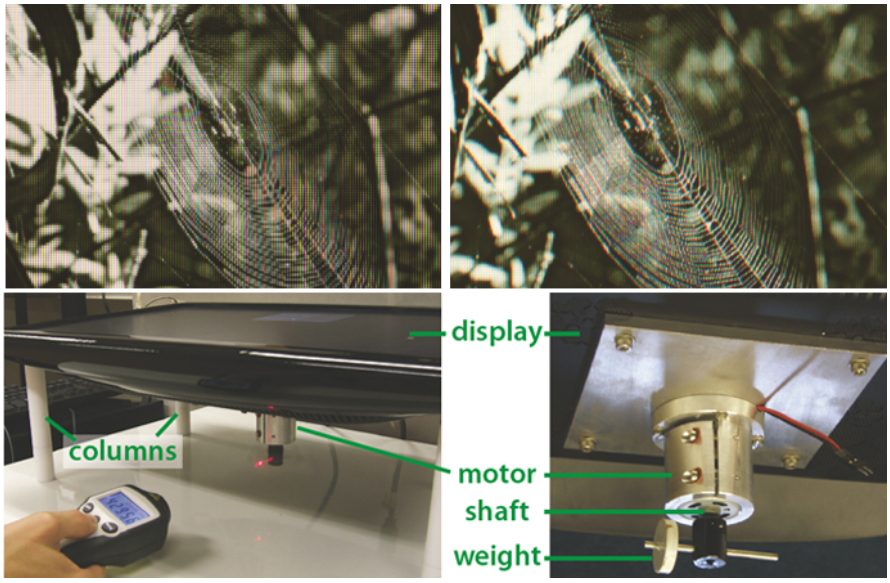


Figure 4.14: Spatial resolution enhancement by temporal superposition in a wobbling display. Top, left: Example image as seen on a conventional (static) display. Top, right: Higher resolution image perceived on a vibrating display. Bottom: to vibrate the display, a motor with an offset weight is attached to its back. Centrifugal forces make the screen vibrate as the motor rotates [47].

4.4.2 *Display Architectures*

There is a mismatch between the spatial resolution of today's captured or rendered images, and the resolution that displays that can currently be found in a typical household can show. This effectively means that captured images need to be downsampled before being shown, which leads to loss of fine details and the introduction of new artifacts. Higher resolution can be achieved by tiling projected images [187, 359, 285, 64, 287]. Another obvious way to increase the spatial resolution of displays is to have more pixels per inch, in order to make the underlying grid invisible to the eye. The Retina display by Apple[6], for instance, packs

about 220 pixels per inch (for a 15″ display). Even though this is a very high pixel density, it is still not enough for a user not to distinguish pixels at the normal viewing distance of 20″[1]. Other alternatives to a very high pixel density have been explored. With the exception of sub-pixel rendering [349] (Section 4.4.3), all superresolution displays require specialized hardware configurations. These can be categorized into *optical superposition* and *temporal superposition*.

Optical Superposition is a projection principle where low-resolution images from multiple devices are optically superimposed on the projection screen. The superimposed images are all shifted by some amount with respect to each other such that one super-resolved pixel receives contributions from multiple devices. Examples of this technique include [93] and [202]. Precise calibration of the projection system is essential in these techniques. The optimal pixel states to be displayed by each projector are usually computed by solving a linear inverse problem. Performance metrics for these types of superresolution displays are discussed in [449].

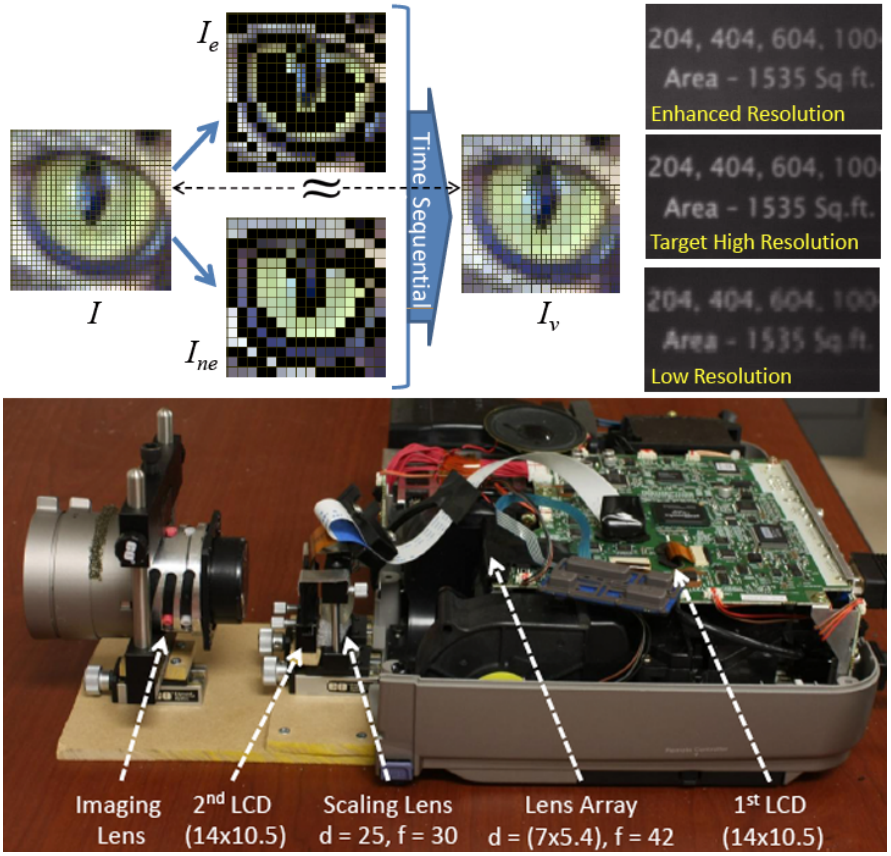


Figure 4.15: Spatial resolution enhancement by optical pixel sharing. Top-left: The Optical Pixel Sharing technique decomposes a target high resolution image I into a high resolution *edge* image I_e and a low resolution *non-edge* image I_{ne}, which are then displayed in a time sequential manner to obtain the *edge-enhanced* image I_v. Top-right: Comparison of the target image with a low resolution and a enhanced resolution version. Bottom: a side view of the prototype projector, achieving enhanced resolution by cascading two lower-resolution panels [385].

1 Pixel density and viewing distance calculator at `http://isthisretina.com/`

Temporal Superposition. Similar to optical superposition techniques, temporal multiplexing requires multiple low-resolution images to be displayed, each shifted with respect to each other. Shown faster than the perceivable flickering frequency of the HVS (which depends on a number of factors, as described in Section 4.5.1), these images will be fused together by the HVS into a higher resolution one, beyond the actual physical limits of the display. This idea can be seen as the dual of the jittered camera for ensembling a high resolution image from multiple low-resolution versions [44]. The shift can be achieved in single display/projector designs using actuated mirrors [14] or mechanical vibrations of the entire display [47] (Figure 4.14). As an interesting avenue of future work, the authors of the latter work outline how the physical vibrations of the display could be avoided, by using a crystal called Potassium Lithium Tantalate Niobate (KLTN), which can change its refractive index [8].

The disadvantage of most existing superresolution displays is that either multiple devices are required, increasing size, weight, and cost of the system, or that mechanically-moving parts are necessary. One approach that does not require either is *Optical Pixel Sharing* (OPS) [385, 384], which uses two LCD panels and a *jumbling* lens array in projectors to overlay a high-resolution edge image on a coarse resolution image to adaptively increase resolution (Figure 4.15). OPS is compressive in the sense that the device does not have the degrees of freedom to represent any arbitrary target image. Much like image compression techniques, OPS relies on the target to be compressible.

4.4.3 *Generation of Content*

We group existing techniques for higher definition content generation into three categories: super-resolution, sub-pixel rendering and temporal integration.

Super-resolution. Increasing spatial resolution is related to super-resolution techniques (see for instance [193, 44, 29]. The underlying idea is to take a signal processing approach to reconstruct a higher-resolution signal from a low-resolution one (or several). It is less expensive than physically packing more pixels, and the results can usually be shown on any low-resolution display. Super-resolution techniques are used in different fields like medical imaging, surveillance or satellite imaging. We refer the reader to recent state of the art reports for a complete overview [341, 454].

Majumder [282] provides a theoretical analysis investigating the duality between super-resolution from multiple captured images, and from multiple overlapping projectors, and shows that super-resolution is only feasible by changing the size of the pixels. In their work on display supersampling [93], the authors present a theoretical analysis to engineer the right aliasing in the low-resolution images, so that resolution is increased after superposition, even in the presence of non-uniform grids. The same authors had previously presented a unifying theory of both approaches, tiled and superimposed projection [94].

Sub-pixel Rendering. Sub-pixel rendering techniques increase the apparent resolution by taking advantage of the display sub-pixel architecture. Instead of assuming that each channel is spatially coincident, they treat each one differently [46]. This approach has given rise to many different pixel architectures and

reconstruction techniques [119, 20, 120]. For instance, Hara and Shiramatsu [157] show that an RGGB pattern can extend the apparent pass band of moving images, improving the perceived quality with respect to a standard RGB pattern.

One of the key insights to handle sub-pixel sampling artifacts like color fringes and moire patterns, is to leverage the fact that human luminance and chrominance contrast sensitivity functions differ, and both signal can be treated differently. Platt [349] and Klompenhouwer and De Haan [230] exploited this in the context of text rasterization and image scaling, respectively. Platt's method, used in the ClearType functionality, is limited to increased resolution in the horizontal dimension; based on this, other different filtering strategies to reduce color artifacts have been tested [128]. Messing and Daly additionally remove chrominance aliasing using a perceptual model [315], while Messing et al. present a constrained optimization framework to mask defective sub-pixels for any regular 2D configuration [314]. These approaches have been recently generalized, presenting optimal, analytical filters for different one- and two-dimensional subpixel layouts [121].

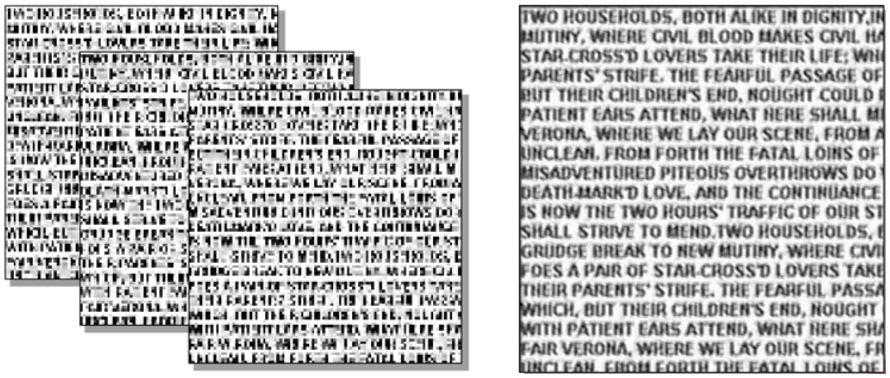


Figure 4.16: Spatial resolution enhancement by temporal superposition in a conventional display. Left: Low resolution images displayed sequentially in time. Right: Corresponding high resolution image perceived as a consequence of the temporal integration performed by the HVS by leveraging SPEM [104].

Temporal integration. An analysis of the properties of the superimposed images resulting from temporal integration appears in [383]. Berthouzoz and Fattal [47] present an analysis of the theoretical limits of this technique. Instead of physically shaking the display, Basu and Baudisch [43] change the strategy and introduce subtle motion to the displayed images, so that higher resolution is perceived by means of temporal integration. Didyk et al [104] project moving low resolution images to predictable locations in the fovea, leveraging the SPEM feature of the HVS (see Section 4.4.1) to achieve perceived high resolution images from multiplexed low resolution content (Figure 4.16). This work is limited to one-dimensional, slow panning movements at constant velocity. In subsequent work, the idea is generalized to arbitrary motions and videos, by analyzing the spatially varying optical flow. The assumption is that between consecutive saccades, SPEM closely follows the optical flow [436].

4.5 IMPROVING TEMPORAL RESOLUTION

Although spatial resolution is one of the most important aspects of a displayed image, temporal resolution cannot be neglected. In this context, it is crucial that the HVS acts as a time-averaging sensor. This has a huge influence in situations where the displayed signal is not constant over time, or there is motion present in the scene. In this section, we will show that the perceived quality can be significantly affected in such situations and present methods that can improve it.

4.5.1 *Perceptual Considerations*

The HVS is limited in perceiving high temporal frequencies, i.e. an elevated number of variations in the image per unit time. This is due to the fact that the response of receptors on the retina is not instantaneous [453]. Also, high-level vision processes further lower the sensitivity of the HVS to temporal changes. As a result, temporal fluctuations of the signal are averaged and perceived as a constant signal. One of the basic findings in this field is Bloch's law [142]. It states that the detectability of a stimuli depends on the product of luminance and exposure time. In practice, this means that the perceived brightness of a given stimuli would be the same if the luminance was doubled and the exposure time halved. Although it is often assumed that the temporal integration of the HVS follows this law, it only holds for short duration times (around 40 ms) [142].

From the practical point of view it is more interesting to know when the HVS can perceive temporal fluctuations and when it interprets them as a constant signal. This is defined by the *critical flicker frequency* (CFF) [214], which defines a threshold frequency for a signal to be perceived as constant or as changing over time. The CFF depends on many factors such as temporal contrast, luminance adaptation, retinal region or spatial extend of the stimuli. For different luminance adaptation levels the CFF was measured yielding a temporal contrast sensitivity function [96]. It is also important that the CFF significantly decreases for smaller stimuli, and that peripheral regions of retina are more sensitive to flickering [310, 288]. Recently, these different factors where incorporated into a video quality metric [26].

In the context of display design, in displays that do not reproduce a constant signal (e.g., CRT displays), low refresh-rate can lead to visible and undesired flickering. Another problem that can be caused by poor temporal resolution is jaggy motion. Instead of smooth motion, which is normally observed in the real world, fast moving objects on the screen appear as they were jumping in a discrete way. Also, when the frame rate of the content does not correspond to the frame rate of the display some frames need to be repeated or dropped. This, similarly to low frame rate, contributes significantly to reduced smoothness of the motion.

Besides the aforementioned issues, low frame rate may introduce significant blur in the perceived image. This type of blur, often called *hold-type blur*, is purely perceptual and cannot be observed in the content: It arises from the interaction between the display and the HVS [339]. In the real world objects move continuously, and they are tracked by the human eyes almost perfectly; this is enabled



Figure 4.17: Simulation of hold-type blur [103]. A user is shown the same animation sequence (sample frame on the left) simultaneously at two different refresh rates. The subject's task is to adjust the blur in the sequence of the right (120 Hz) until the level of blur matches that of the sequence on the left (60 Hz). The average result is shown here: the blurred sequence on the right displayed at 120 Hz is visually equivalent to the sharp sequence on the left displayed at 60 Hz.

by the so-called *smooth pursuit eye motion* (SPEM, please refer to Section 4.4.1 for details). In the context of current display devices, although the tracking still is continuous, the image presented on a screen is kept static for an extended period of time (i. e.the period of one frame). Therefore, due to temporal averaging, the receptors on the retina average the signal while moving across the image during the period of one frame. As a result the perceived image is blurred (see also Figure 4.17). The hold-type blur can be modeled using a box filter [231], its support dependent on object velocity and frame rate. This blur is not the same blur as that due to the slow response of the liquid crystals in LCD panels. Pan et al. [339] demonstrated that only 30% of the perceived blur is a consequence of the slow response (and they assumed a response of 16 ms, whereas in current displays this time does not exceed 4 ms). This, together with overdrive techniques, makes the problem of slow response time of displays negligible compared to the hold-type blur. The hold-type blur is a big bottleneck for display manufacturers, as it can destroy the quality of images reproduced using ultra-high resolutions such as 4K or 8K. Since the strength of the blur depends on angular object velocity, the problem becomes even more relevant with growing screen sizes, which are desired in the context of home cinemas or visualization centers.

4.5.2 *Temporal Upsampling Techniques*

A straightforward solution to all problems mentioned above is higher framerate: It reduces jaggy motion and solves the problem of framerate conversion. For higher frame rates the period for which moving objects are kept in the same location is reduced, therefore, it can also significantly reduce the hold-type blur. However, high frame rate is not provided in broadcasting applications, and in the context of computer graphics high temporal resolution is very expensive. This forced both the graphics community and display manufacturers to devise techniques to increase the frame rate of the content in an efficient manner.

Most of the industrial solutions for temporal upsampling that are used in modern TV-sets are designed to compensate for the hold-type blur. Efficiency is key in these solutions, as they are often implemented in small computational units. These techniques usually increase frame rate to e. g.100 or 200 Hz, by introducing intermediate frames generated from the low frame rate broadcasted signal.

One of the simplest methods in this context is *black data insertion*, i. e.introducing black frames interleaved with the original content. This solution can reduce hold-type blur because it reduces the time during which the objects are shown in the same position. A similar technique, more efficient hardware solution is to turn on and off backlight of LCD panel [339, 131]. This is possible because current LCD panels employing LED backlights can switch at frequencies as high as 500 Hz. These two techniques, although fast and easy to implement, suffer from brightness and contrast reduction as well as possible temporal flickering. To overcome these problems, Chen et al.[74] proposed to insert blurred copies of the original frames. Although this ameliorates the brightness issue, it may produce ghosting, since the additional frames are not motion compensated.

More common solutions in current TV screens are *frame interpolation techniques*. In these techniques, additional frames are obtained by interpolating original frames along motion trajectories [247]. Such methods can easily expand a 24 Hz signal, a common standard for movies, to 240 Hz without brightness reduction or flickering. The biggest limitation of these techniques is related to optical flow estimation, which is required for good interpolation. For efficiency reasons simple optical flow techniques are used, which are prone to errors; they usually perform well for slower motions and tend to fail for faster ones [103]. Additionally, these techniques interpolate in-between frames, which requires knowledge of future frames. This introduces a lag which is not a problem for broadcasting applications, but may be unacceptable for interactive applications. In spite of these problems, motion-based interpolation together with backlight flashing is the most common technique in current display devices. An extended survey on these techniques is provided in [131].

An alternative software solution used in TV-sets to reduce hold-type blur is to apply a filtering step which compensates for the blur later introduced by the HVS. This technique is called *motion compensated inverse filtering* [231, 160]. In practice, it boils down to applying a 1D sharpening filter oriented along motion trajectories, the blur kernel being estimated from optical flow. The effectiveness of such solution is limited by the fact that the hold-type blur removes certain frequencies which cannot be restored using prefiltering. Furthermore, such techniques are prone to clipping problems and oversharpening.

The problem of increasing temporal resolution is also well known in computer graphics. However, in this area, not all solutions need to provide a real-time performance, e. g.some of them were designed to improve low performance of high quality global illumination techniques, where offline processing is not a problem. This, in contrast to previously mentioned industrial solutions, allows for more sophisticated and costly techniques. Another advantage of computer graphics solutions is that they very often rely on additional information that is produced along with the original frames, e. g.depth or motion flow. All this significantly improves the quality of new frames.

One group of methods which can be used for creating additional frames and increasing frame rate are *warping techniques*. The idea of these techniques [416] is to morph texture between two target images, creating a sequence of interpolated images; an extended survey discussing these techniques was presented by Wolberg [487]. Recently Liu et al.[269] presented content-preserving warps for the purpose of video stabilization. Using their technique they can synthesize images as if they were taken from nearby viewpoints. This allows them to create video sequences where the camera path is smooth, i. e.the video is stabilized. Although warping techniques were not originally designed for the purpose of improving temporal resolution, they can be successfully used in this context, taking advantage of the fact that interpolated images are very similar when performing temporal upsampling. An example of this is a method by Mahajan et al.[280].



Figure 4.18: Temporal upsampling: The three frames shown have been synthesized from two input images (not shown), by moving gradients along a path [280].

Their technique performs well for single disocclusions, yielding high quality results for standard content (Figure 4.18). It requires, however, knowledge of the entire sequence, therefore it is not suitable for real-time applications. Although the high quality of interpolated frames is desirable independent of the location, Stich et al.[423] showed that high-quality edges are crucial for the HVS. Based on this observation, they proposed a technique that takes special care of edges, making their movement more coherent and smooth.

For interactive applications, where frame computation costs can limit interactivity, often additional information such as depth or motion flow is leveraged for more efficient and effective frame interpolation. One of the first methods for temporal upsampling for interactive applications was proposed by Mark et al.[295]. They used depth information to reproject shaded pixels from one frame to another. In order to avoid disocclusions they proposed to use two originally rendered frames to compute in-between frames, which significantly decreases the problem of missing information. Similar ideas were used later where re-use of shaded samples was proposed to speed up image generation. In Render Cache, Walter et al.[461] used forward re-projection to scatter the information from previously rendered frames into new ones. Later, forward reprojection was replaced by reversed reprojection [331]. Instead of re-using pixel colors, i. e.the final result of rendering, also intermediate values can be stored and reused for computation of next frames [409], speeding up the rendering process. Another efficient method for temporal upsampling in the context of interactive applications was proposed by Yang et al. [494]. Their method uses fixed-point iteration to find

a correct pixel correspondence between originally rendered views and interpolated ones. Later, this technique was combined with mesh-based techniques by Bowles et al.[54]. The temporal coherence of computer graphics animations was also explicitly exploited by Herzog et al.[167]: They proposed a spatio-temporal upsampling where they not only increased the frame rate, but the also spatial resolution. A more extensive survey on these techniques can be found in [387].

Although techniques developed for computer graphics applications and for TV-sets have slightly different requirements, it is possible to combine these techniques. Didyk et al.[103] proposed a technique which combines blurred frame insertion and mesh-based warping. The method can be performed in a few milliseconds, and the quality is assured by exploring temporal integration of the HVS. The artifacts in generated frames are blurred, and the loss of high frequencies is compensated in the original frames. This solution eliminates artifacts produced by warping techniques as well as blurred frame insertion. Additionally, the technique performs extrapolation instead of interpolation assuming a linear motion. This eliminates the problem of lag, but can create artifacts for a highly nonlinear and very fast motion. The mesh-based temporal upsampling was further improved in [105].

4.6 IMPROVING ANGULAR RESOLUTION I: STEREOSCOPIC DISPLAYS

Recently, due to the success of big 3D movie productions, stereo 3D (S3D) is receiving significant attention from consumers as well as manufacturers. This has spurred rapid development in display technologies, trying to bring high quality 3D viewing experiences into our homes. There is also an increasing amount of 3D content available to customers, e.g., 3D movies, stereoscopic video games, even broadcast 3D channels. Despite the fast progress in S3D, there are still many challenging problems in providing perceptually convincing stereoscopic content to the viewers.

4.6.1 *Perceptual Considerations*

When perceiving the world, the HVS relies on a number of different mechanisms to obtain a good layout perception. These mechanisms, also called depth cues, can be classified as pictorial (e.g., occlusions, relative size, texture density, perspective, shadows), dynamic (motion parallax), ocular (accommodation and vergence) and stereoscopic (binocular disparity) [337]. The sensitivity of the HVS to different cues varies [88], and it depends mostly on the absolute depth. The HVS is able to combine different cues [337, Chapter 5.5.10], which usually strengthen each other; however, in some situations they can also contradict each other. In such cases, the final 3D scene interpretation represents a compromise between the conflicting cues according to their strength. Although much is unknown about cue integration and the relative importance of cues, binocular disparity and motion parallax (see Section 4.7.1) are argued to be the most relevant depth cues at typical viewing distances [88]. Figure 4.19 depicts the influence of depth cues at different distances. A thorough description of all depth cues is outside

the scope of this survey, but the interested reader may refer to [211, 183] for detailed explanations.

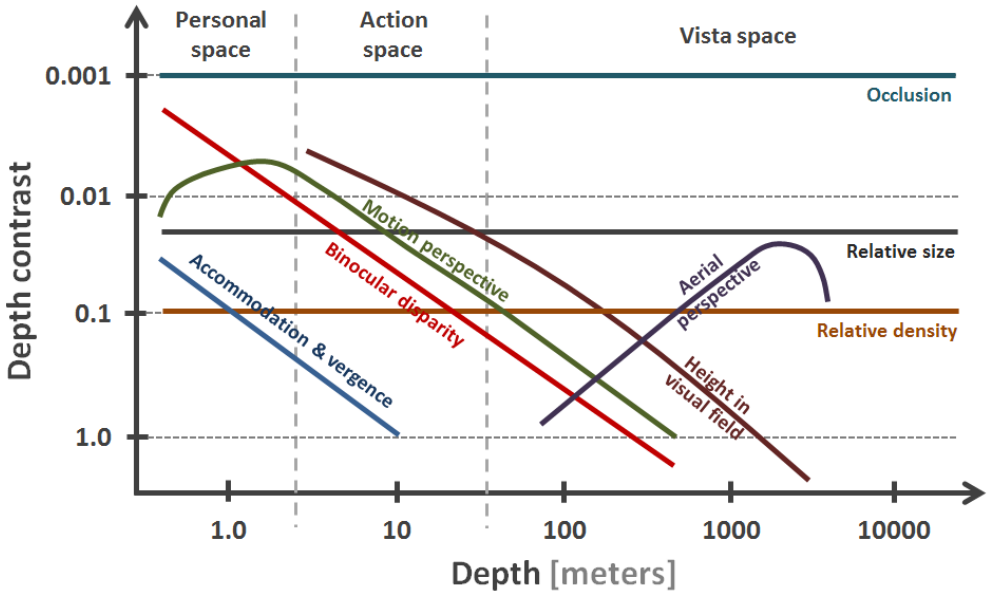


Figure 4.19: Sensitivity (just-discriminable depth thresholds) of the HVS to nine different depth cues as a function of distance to the observer. Note that the lower the threshold (depth contrast), the more sensitive the HVS is to that cue). Depth contrast is computed as the the ratio of the just-determinable difference in distance between two objects over their mean distance. Adapted from [88].

Current 3D display devices take advantage of one of the most appealing depth cues: *binocular disparity*. On such screens the 3D perception is, however, only an illusion created on a flat display by showing two different images to both eyes. In such a case, the conflict between depth cues is impossible to avoid. The most prominent conflict is the *accommodation-vergence mismatch* (Figure 4.20). While vergence—the movement the eyes perform for both to foveate the same point in space—can easily adapt to different depths presented on the screen, accommodation—the change in focus of the eyes—tries to maintain the viewed image in focus. When extensive disparities between left and right eye images drive the vergence away from the screen, the conflict between fixation and focus point arises. It can be tolerated up to the certain degree (within the so-called comfort zone), beyond which it can cause visual discomfort [177]. Based on extensive user studies, Shibata et al. [401] derived a model to predict the zone of comfort. Motion is another potential source of discomfort. Recently, Du and colleagues [112] presented a metric of comfort taking into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast (Figure 4.21).

The fact that the depth presented on the 3D screen fits into the comfort zone does not yet assure a perfect 3D experience. The retinal images created in the left and right eyes are misaligned, since they originate from different viewpoints. In order to create a clear and crisp image they need to be fused. The HVS is able to perform the fusion only in a region called *Panum's fusional area* (Figure 4.20) where relative disparities are not too big; beyond this area double vision

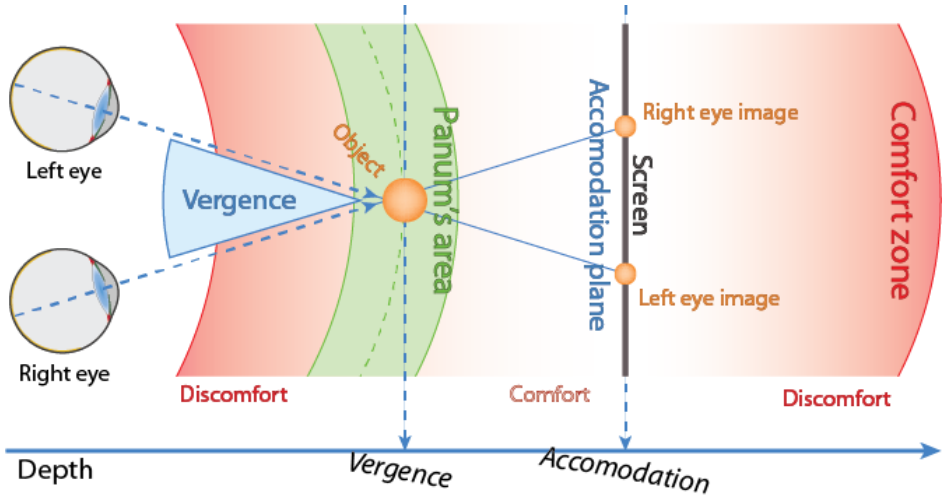


Figure 4.20: Accommodation-vergence conflict in stereoscopic displays. While vergence of the eyes is driven to the 3D position of the object perceived, focus (accommodation) remains on the screen. This mismatch can cause fatigue and discomfort to the viewer.

(diplopia) is experienced (see e. g. [211, Chapter 5.2]). In fact, binocular fusion is a much complex phenomenon, and it depends on many factors such as individual differences, stimulus properties or exposure duration. For example, people are able to fuse much larger relative disparities for low frequency depth corrugations [446]. The fusion is also easier for stimuli which are well illuminated, have strong texture contrast, or are static.

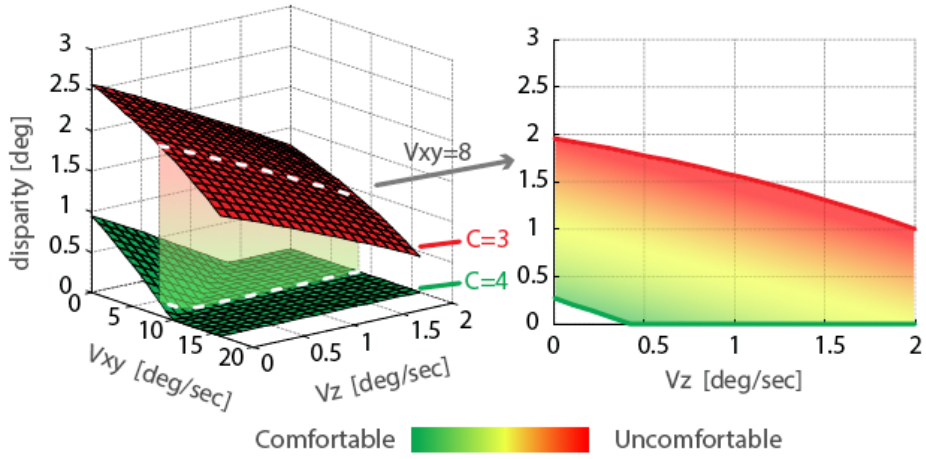


Figure 4.21: Example slice of the comfort zone predicted by Du et al, taking into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast [112].

Assuming that a stereoscopic image is fused by the observer and a single image is perceived, further perception of different disparity patterns depends on many factors. Interestingly, these factors as well as the mechanisms responsible for the interpretation of different disparity stimuli are similar to what is known from luminance perception [63, 274, 58]. One of the most fundamental foundings

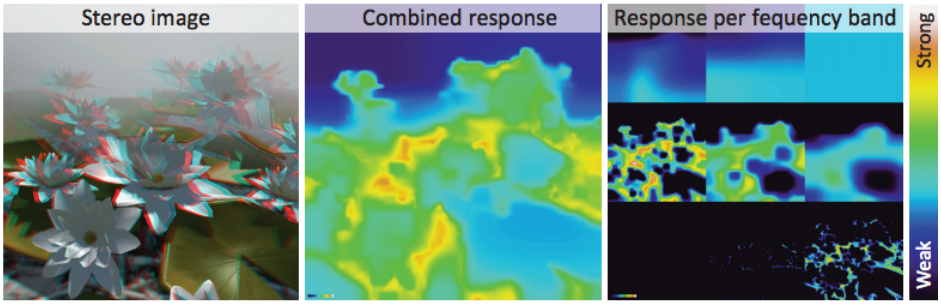


Figure 4.22: Perceived disparity as predicted by a recent metric which incorporates the influence of luminance-contrast in the perception of depth from binocular disparity [108]. From left to right, original stereo image, combined response, and response per frequency band (please refer to the original work for details).

from this field is the contrast sensitivity function (CSF, Section 4.2.1). Similarly, in disparity perception a disparity sensitivity function (DSF) exists. Assuming a sinusoidal disparity corrugation with a given frequency, the DSF function defines a reciprocal of the detection threshold, i. e.the smallest amplitude that is visible to a human observer. Both, CSF and DSF, share the same shape, although the DSF has a peak at a different spatial frequency [58]. Another example of similarities is the existence of different receptive fields tuned to specific frequencies of disparity corrugations [183, Chapter 19.6.3]. Also, similar to luminance perception, apparent depth deduced from the disparity signal is dominated by relative disparities (disparity contrast) rather than absolute depth. Furthermore, illusions which are known from brightness perception exist also for disparity. For example, it turns out that the Craik-O'Brien-Cornsweet Illusion (Section 4.2.1) holds for disparity patterns [18, 375]. These similarities suggesting that brightness and disparity perception undergo similar mechanisms have recently been explored to build perceptual models for disparity [106, 108] (Figure 4.22).

4.6.2 *Display Architectures*

Since in 1838 Charles Wheatstone invented the first stereoscope, the basic idea for displaying 3D images exploiting binocular disparity has not changed significantly. In the real world, people see two images (left and right eye images), and the same has to be reproduced on the screen for the experience to be similar. Wheatstone proposed to use mirrors which reflect two images located off the side. The observer looking at the mirrors sees these two images superimposed. Wheatstone demonstrated that if the setup is correct, the HVS will fuse the two images and perceive them as if looking at a real 3D scene [484, 485].

Since then, people have come up with many different ways of showing two different images to both eyes. The most common method is to use dedicated glasses. A set of solutions employ *spatial multiplexing*: Two images are shown simultaneously on the screen, and glasses are used to separate the signal so that each eye sees only one of them. There are different methods of constructing such setup. One possibility is to use different colors for left and right eye (anaglyph

stereo). The image on the screen is then composed of two differently tinted images (e. g.red and cyan). The role of the glasses is to filter the signal so a correct image is visible by each eye, using different color filters. Although different filters can be used, due to different colors shown to both eyes the image quality perceived by the observed is degraded. To avoid it, one can use more sophisticated filters which let through all color components (RGB), but the spectrum of each is slightly shifted and not overlapping to enable easy separation. It is also possible to use polarization to separate left and right eye images. In such solutions, the two images are displayed on a screen with different polarization and the glasses use another set of polarized filters for the separation. Recently, *temporal multiplexing* gained great attention, especially in the gaming community. In this solution, the left and right eye images are interleaved in the temporal domain and shown in rapid succession. The glasses consist of two shutters which can "open and close" very quickly showing the correct image to each eye. A detailed recent review –which also includes head-mounted displays, not covered here– can be found in [451].

Glasses-based solutions have many problems, e. g.reduced brightness, resolution or color shift. However, a bigger disadvantage is the need to wear additional equipment. Whereas this is not a significant problem in movie theaters, people usually do not feel comfortable wearing 3D glasses at home or in other public places. A big hope in this context is glasses-free solutions. So-called *autostereoscopic displays* can show two different images simultaneously, the visibility of which depends on the viewing position. This is usually achieved by placing a parallax barrier or a lenslet array in front of the display panel. We cover these technologies in detail in Section 4.7, since the main techniques for autostereoscopic displays can be seen as a particular case of those used for automultiscopic displays.

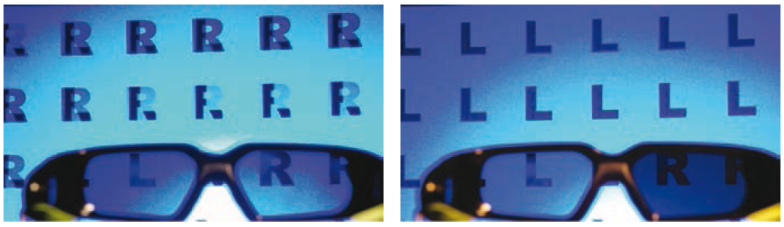


Figure 4.23: 3D+2D TV [386]. Left: A conventional glasses-based stereoscopic display: It shows a different view to each eye while wearing glasses, while without glasses both images are seen superimposed. Right: The 3D+2D TV shows a different view to each eye with glasses, while viewers without glasses see one single image, with no ghosting effect.

A stereoscopic version of the content is not always desired by all observers. This can be due to different reasons, e. g.lack of additional equipment, lack of tolerance for such content, or comfort. An interesting problem is thus to provide a solution which enables both 2D and 3D viewing at the same time, the so-called *backward-compatible stereo* [106]. An early approach in this direction was to use color glasses with color filters which minimize ghosting when the content is observed without them; for example, amber and blue filters can be used (ColorCode 3-D). When the 3D content is viewed with the glasses, enough signal is

provided to both eyes to create a good 3D perception. However, when the content is viewed without the glasses, the blue channel does not contribute much to the perceived image, and the ghosting is hardly visible. Recently, another interesting hardware solution was provided [386] that improves over the shutter-based solution. Instead of interleaving two images, there is an additional third image which is a negative of one of the two original ones. The 3D glasses are synchronized so that the third image is imperceptible for any eye if the glasses are worn. However, when the observer views the content without the glasses, the third image, due to the temporal integration performed by the HVS (Section 4.5.1), cancels one of the images of the stereoscopic pair, and only one of them is visible (see Figure 4.23).

4.6.3 *Software Solutions for Improving Depth Reproduction*

In the real world, the HVS can easily adapt to objects at different depths. However, due to the fundamental limitations of stereoscopic displays, it is not possible to reproduce the same 3D experience on current display devices. Therefore, a special care has to be taken while preparing content for a stereoscopic screen. Such content needs to provide a compelling 3D experience, while maintaining viewing comfort. A number of methods have been proposed to perform this task efficiently. The main goal of all these techniques is to fit the depth range spanned by the real scene to the comfort zone of a display device, which highly depends on the viewing setup [401] (e. g.viewing distance, screen size, etc.). This can be performed at different stages of content creation, i. e.during capture or in a post-processing step.

The first group of methods which enable stereoscopic content adjustment are techniques that are applied during the capturing stage. The adjustments are usually performed by changing camera parameters, i. e.interaxial distance—the distance between cameras—and convergence—the angle between the optical axes of the cameras. Changing the first one affects the disparity range by either expanding it or reducing it (smaller interaxial distances result in smaller disparity ranges). The convergence, on the other hand, is responsible for the relative positioning of the scene with respect to the screen plane. Jones et al.[209] proposed a mathematical framework defining the exact modification to camera parameters that needs to be applied in order to fit the scene into the desired disparity range. More recently, Oskam et al. [335], proposed a similar approach for real-time applications in which they formulated the problem of camera parameters adjustment as an optimization framework. This allowed them not only to fit the scene into a given disparity range but also to take into account additional artists' design constraints. Apart from that, they also demonstrated how to deal with temporal coherence of such manipulations in real-time scenarios. An interesting system was presented by Heinzle et al. [165]. Their complete camera rig provides an intuitive and easy-to-use interface for controlling stereoscopic camera parameters; the interface collects high-level feedback from the artists and adjusts the parameters automatically. In practice, it is also possible to record the content with multiple camera setups, e. g.a different one for background and foreground, and the different video streams combined during the compositing stage. A big advantage of techniques which directly modify the camera parameters is that

they can also compensate for object distortions arising from the wrong viewing position [166].

The above methods are usually a satisfactory solution if the viewing conditions are known in advance. However, in many scenarios, the content captured with a specific camera setup, i. e.designed for a particular display, is also viewed on different screens. To fully exploit the available disparity range, post-processing techniques are required to re-synthesize the content as if it were captured using different camera parameters. Such *disparity retargeting* methods usually work directly on disparity maps to either compress or expand disparity range. An example of such techniques was presented by Lang et al. [252]. By analogy to tone-mapping operators (Section 4.2.3), they proposed to use different mapping curves to change the disparity values. The mapping can be done according to differently designed curves (e. g.linear or logarithmic curves). It can be also performed in the gradient domain. In order to improve depth perception of important objects, they also proposed to incorporate saliency prediction into the curve design. The problem of computing adjusted stereo images is formulated as an optimization process that guides a mesh-based warp according to the edited disparity maps. It is also possible to use more explicit methods which do not involve optimization [105].

Recently, perceptual models for disparity have been proposed [106, 108]. With their aid, disparity values can be transformed into a perceptually uniform space, where they can be mapped to fit a desired disparity range. Essentially, the disparity range is reduced while preserving the disparity signal whenever it is most relevant for the HVS. Perceptual models of disparity can additionally be used to build metrics which can evaluate perceived differences in depth between an original stereo image and its modified version. This allows for defining the disparity remapping problem as an optimization process where the goal is to fit disparities into a desired range while at the same time minimizing perceived distortions [108]. As the metrics can also account for different luminance patterns, such methods were shown to perform well for automultiscopic displays where the content needs to be filtered to avoid inter-view aliasing [512]. More about adopting content for such screens can be found in Section 4.7.3. Disparity models also enable depth perception enhancement. For example, when the influence of luminance patterns on disparity perception is taken into account [108], it is possible to enhance depth perception in regions where it is weakened due to insufficient texture. This can be done by introducing additional luminance information.

One of the most aggressive methods for stereo content manipulation is microstereopsis. Proposed by Siegel et al.[405], this technique reduces the camera distance to a minimum so that a stereo image has just enough disparity to create a 3D impression. This solution can be useful in the context of *backward-compatible stereo* because the ghosting artifacts during monoscopic presentation are significantly reduced. Didyk et al.[106, 107] proposed another stereo content manipulation technique for backward-compatible stereo. Their method uses the Craik-O'Brien-Cornsweet Illusion to reproduce disparity discontinuities. As a result, the technique significantly reduces possible ghosting when the content is viewed without stereoscopic equipment, but a good 3D perception can be achieved when



Figure 4.24: Top: Microstereopsis [405] reduces disparity to the minimum value that would enable 3D perception. Bottom: Backward-compatible stereo [107] aims at preserving the perception of depth in the scene while reducing disparities to enable "standard 2D viewing" (without glasses) of the scene; the Craik-O'Brien-Cornsweet illusion for depth is leveraged in this case to enhance the impression of depth in certain areas while minimizing disparity in others.

the content is viewed with the equipment. It is also possible to enhance depth impression by introducing Cornsweet profiles atop of the original disparity signal. Figure 4.24 shows examples of these techniques.

All aforementioned techniques for stereoscopic content adjustment do not analyze how much such manipulations affect motion perception. Recently, Kellnhofer et al.[216] proposed a technique for preventing visible motion distortions due to disparity manipulations. Besides, previously mentioned techniques are mostly concerned with the disparity signal introduced by scene geometry. However, extensive disparities can also be created by secondary light effects such as reflection. Templin et al. [438] proposed a technique that explicitly accounts for the problem of glossy reflections in stereoscopic content. Their technique prevents viewing discomfort due to extensive disparities coming from such reflections, while maintaining at the same time their realistic look.

4.7 IMPROVING ANGULAR RESOLUTION II: AUTOMULTISCOPIC DISPLAYS

Automultiscopic displays, capable of showing stereo images from different viewpoints without the need to wear glasses or other additional equipment, have been a subject of much research throughout the last century. A recent state-of-the-art review on 3D displays including glasses-free techniques can be found in [451]. We briefly outline these technologies and discuss in more detail the most recent developments on light field displays, both in terms of hardware and of content

generation. In this survey, we do not discuss holographic imaging techniques (e.g., [410]), which present all depth cues, but are expensive and primarily restricted to static scenes viewed under controlled illumination [232].

4.7.1 *Perceptual Considerations*

As discussed in Section 4.6.1, there is a large number of cues the HVS utilizes to infer the (spatial layout and) depth of a scene (Figure 4.19). Here we focus on motion parallax, which is the most distinctive cue of automultiscopic displays, not provided by stereoscopic or conventional 2D displays.

Motion parallax enables us to infer depth from relative movement. Specifically, it refers to the movement of an image projected in the retina as the object moves relative to the viewer; this movement is different depending on the depth at which the object is with respect to the viewer, and the velocity of the relative motion. Depth perception from motion parallax exhibits a close relationship in terms of sensitivity with that of binocular disparity, suggesting similar underlying processes for both depth cues [376, 178]. Existing studies on sensitivity to motion parallax are not as exhaustive as those on disparity, although several experiments have been conducted to establish motion parallax detection thresholds [59]. The integration of both cues, although still largely unknown, has been shown to be non-linear [57].

Consistent vergence-accommodation cues and motion parallax are required for a natural comfortable 3D experience [361]. Automultiscopic displays, potentially capable of providing these cues, are emerging as the new generation of displays, although limitations persist, as discussed in the next subsection. Additional issues that may hinder the viewing experience in automultiscopic displays are crosstalk between views, moire patterns, or the cardboard effect [361, 482].

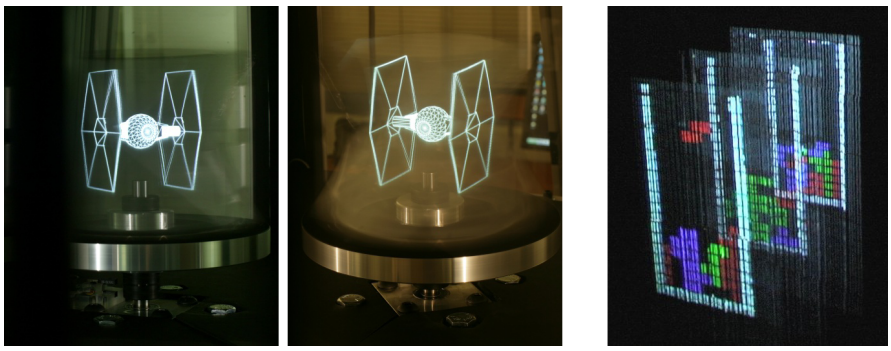


Figure 4.25: Two examples of volumetric displays. Left: Sweeping-based volumetric light field display supporting occlusions and correct perspective [208]. Right: Volumetric display employing water drops as a projection substrate, here showing an interactive Tetris game [41].

4.7.2 *Display Architectures*

VOLUMETRIC DISPLAYS Blundell and Schwartz [52] 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 [130] and have been continuously improved [85]. While requiring similar mechanical motion, Jones et al. [208] instead achieve a light field display, preserving accurate perspective and occlusion cues, by introducing an anisotropic diffusing screen and user tracking. Related designs include the Seelinder [501], exploiting a spinning cylindrical parallax barrier and LED arrays, and the work of Maeda et al. [279], utilizing a spinning LCD panel with a directional privacy filter. Several designs have eliminated moving parts using electronic diffusers [432], projector arrays [7], and beam-splitters [10]. Whereas others consider projection onto transparent substrates, including water drops [41], passive optical scatterers [329], and dust particles [346].

LIGHT FIELD DISPLAYS Light field displays generally aim to create motion parallax and stereoscopic disparity so that an observer perceives a scene as 3D without having to wear encumbering glasses. Invented more than a century ago, the two fundamental principles underlying most light field displays are parallax barriers [198] and integral imaging with lenslet arrays [266]. The former technology has evolved into fully dynamic display systems supporting head tracking and view steering [345, 347], as well as high-speed temporal modulation [255]. Today, lenslet arrays are often used as programmable rear-illumination in combination with a high-speed LCD to steer different views toward tracked observers [424].

COMPRESSIVE LIGHT FIELD DISPLAYS Through the co-design of display optics and computational processing, compressive displays strive to transcend limits set by purely optical designs. It was recently shown that tomographic light field decompositions displayed on stacked films of light-attenuating materials can create higher resolutions than previously possible [482]; and the same underlying approach later applied to stacks of LCDs for displaying dynamic content [253]. A compression is achieved in the number of layer pixels, which is significantly smaller than the number of emitted light rays. Low-rank light field synthesis was also demonstrated for dual-layer [255] and multi-layer displays with directional backlighting [478]. In these display designs, an observer perceptually averages over a number of patterns (shown in Figure 4.26 for a tensor display) that are displayed at refresh rates beyond the critical flicker frequency of the HVS (see Section 4.5.1). The limited temporal resolution of the HVS is directly exploited by decomposing a target light field into a set of patterns, by means of nonnegative matrix or tensor factorization, and presenting them on high-speed spatial light modulators; this creates a perceived low-rank approximation of the target light field.

LIGHT FIELD DISPLAYS SUPPORTING ACCOMMODATION Displays supporting correct accommodation are able to create a light field with enough angular resolution to allow subtle, yet crucial, variation over the pupil. Such displays utilize three main approaches. Ultra-high angular resolution displays, such as super



Figure 4.26: Top row: A prototype tensor display. Middle row: Two different views of a light field as seen on the tensor display. Bottom row: Layered patterns for two different frames [478].

multiview displays [433, 434, 338] (Figure 4.28), take a brute-force approach: all possible views are generated and displayed simultaneously, incurring high hardware costs. In practice, this has limited the size, field of view, and spatial resolution of the devices. Multi-focal displays [10, 175, 400], virtually place conventional monitors at different depths via refractive optics. This approach is effective, but requires encumbering glasses. Volumetric displays [130, 208, 85] also support accommodative depth cues, but usually only within the physical device; current volumetric approaches are not scalable past small volumes. Most recently a compressive accommodation display architecture was proposed [281]. This approach is capable of generating near correct accommodation cues with high spatial resolution over a wide field of view using multilayer display configurations that are combined with high angular resolution backlighting and driven by nonnegative light field tensor factorizations. Finally, Lanman and Luebke recently presented a near-eye light field display capable of presenting accommodation, convergence, and binocular disparity depth cues; it is a head-mounted display (HMD) with a thin form-factor [254].

4.7.3 *Image Synthesis for Automultiscopic Displays*

Stereoscopic displays pose a challenge in what regards to content generation because of the need to capture or render two views, the positioning of the cameras,

or the content post-processing (Section 4.6.3). Multiview content shares these challenges, augmented by additional issues derived from the size of the input data, the computation needed for image synthesis, and the intrinsic limitations that these displays exhibit.

Although targeted only to parallax barriers and lenslet array displays, Zwicker et al. [512] were one of the first to address the problem of reconstructing a captured light field to be shown on light field displays, building on previous work on plenoptic sampling [194, 70]. They proposed a resampling filter to avoid the aliasing derived from limited angular resolution, and derived optimal camera parameters for acquisition.

Ranieri et al. [356] propose an efficient rendering algorithm for multi-layer automultiscopic displays which avoids the need for an optimization process, common in compressive displays. The algorithm is simple, essentially assigning each ray to the display layer closest to the origin and then filtering for antialiasing; they have to assume, however, depth information of the target light field to be known. Similar to this algorithm, but generalized to an arbitrary number of emissive and modulating layers, and with a more sophisticated handling of occlusions, is the decomposition algorithm for rendering light fields in [357].

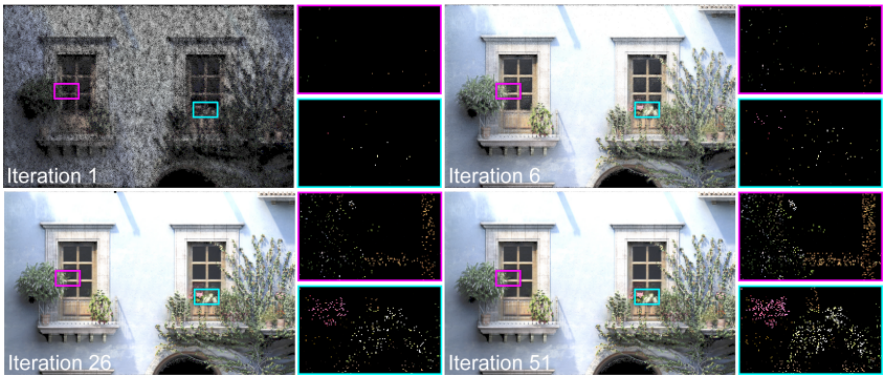


Figure 4.27: Progressive reconstruction of a light field by adaptive image synthesis. In can be seen in the close-ups how the cumulative light field samples used represent a very sparse set of all plenoptic samples [163].

Compressive displays, described in Section 4.7.2, typically require taking a target 4D light field as input and solving an optimization problem for image synthesis. This involves a large amount of computation, currently unfeasible in real time for high angular and spatial resolutions. To overcome the problem, Heide et al. [163] recently proposed an adaptive optimization framework which combines the rendering and optimization of the light field into a single framework. The light field is intelligently sampled leveraging display-specific limitations and the characteristics of the scene to be displayed, allowing to significantly lower computation time and bandwidth requirements (see Figure 4.27). The method is not limited to compressive multiview displays, but can also be applied to high dynamic range displays or high resolution displays.

In the production of stereo content, a number of techniques exist that generate a stereo pair from a single image. This idea has been extended to automultiscopic displays, Singh and colleagues [408] propose a method to generate, from existing stereo content, the patterns to display in a glasses-free two-layer automultiscopic

display to create the 3D effect. Their main contribution lies in the stereo matching process (performed to obtain a disparity map), specially tailored to the characteristics of a multi-layer display to achieve temporal consistency and accuracy in the disparity map. Depth estimation can, however, be a source of artifacts with current methods, resulting in artifacts. To overcome this problem, Didyk et al.[109] proposed a technique that expands a standard stereoscopic content to a multi-view stream avoiding depth estimation. The technique combines both, view synthesis and filtering for antialiasing into one filtering step. The method can be performed very efficiently, reaching a real-time performance.

Content retargeting refers to the algorithms and methods that aim at adapting content generated for a specific display to another display that may be different in one or more dimensions: spatial, angular or temporal resolution, contrast, color, depth budget, etc. [36, 38]. An example in automultiscopic displays is the first spatial resolution retargeting algorithm for light fields, proposed by Birklbauer and Bimber [51]; it is based on seam carving and does not require knowing or computing a depth map of the scene. *Disparity retargeting* for stereo content is discussed in Section 4.6.3. Building on this literature on retargeting of stereo content, a number of approaches have emerged that perform disparity remapping on multiview content (light fields). The need for these algorithms can arise from viewing comfort issues, artistic decisions in the production pipeline, or display-specific limitations. Automultiscopic displays exhibit a limited depth-of-field which is consequence of the need to filter the content to avoid inter-view aliasing. As a result, the depth range within which images can be shown appearing sharp is constrained, and depends on the type and characteristics of the display itself: depth-of-field expressions have been derived for different types of displays [512, 482, 478].

One of the first to address depth scaling in multiview images were Kim et al. [223]. Given the multiview images and the target scaled depth, their algorithm warps the multiview content and performs hole filling whenever disocclusions are present. More sophisticated is the method by Kim and colleagues for manipulating the disparity of stereo pairs given a 3D light field (horizontal parallax only) of the scene [220]. They build an EPI (epipolar-plane image) volume, and compute optimal cuts through it based on different disparity remapping operators. Cuts correspond to images with multiple centers of projection [394], and the method can be applied both to stereo pairs and to multiview images, by performing two or more cuts through the volume according to the corresponding disparity remapping operator. As an alternative, perceptual models for disparity which have recently been developed [106, 108] can also be applied to disparity remapping for automultiscopic displays. This is explained in more detail in Section 4.6.3, but essentially these models allow to leverage knowledge on the sensitivity to disparity of the HVS to fit disparity into the constraints imposed by the display. Leveraging Didyk et al.'s model [106], together with a perceptual model for contrast sensitivity [294], and incorporating display-specific depth-of-field functions, Masia et al. [306, 307] propose a retargeting scheme for addressing the trade-off between image sharpness and depth perception of these displays (see e.g. Figure 5.1, this method is covered in Chapter 5).

4.7.4 *Applications*

In this subsection, we discuss additional applications of light field displays: human computer interaction and vision-correcting image display.

INTERACTIVE LIGHT FIELD DISPLAYS Over the last few years, interaction capabilities with displays have become increasingly important. While light field displays facilitate glasses-free 3D display where virtual objects are perceived as floating in front of and behind the physical device, most interaction techniques focus on either on-screen (multi-touch) interaction or mid-range and far-range gesture-based interaction facilitated by computational photography techniques, such as depth-sensing cameras, or depth-ranging sensors like Kinect™. Computational display approaches to facilitating mid-range interaction have been proposed. These integrate depth sensing pixels directly into the screen of a light field display by splitting the optical path of a conventional lenslet-based light field display such that a light field is emitted and simultaneously recorded through the same lenses [441, 171]. Alternatively, light field display and capture mode can be multiplexed in time using a high-speed liquid crystal panel as a bidirectional 2D display and a 4D parallax barrier-based light field camera [170].



Figure 4.28: Tailored displays can enhance visual acuity. For each scene, from left to right: input image, images perceived by a farsighted subject on a regular display, and on a tailored display [338].

VISION-CORRECTING DISPLAYS Light field displays have recently been introduced for the application of correcting the visual aberrations of an observer (Figure 4.28). Early approaches attempt to filter a 2D image presented on a conventional screen with the inverse point spread function (PSF) of the observer's eye [15, 500, 19]. Although these methods slightly improve image sharpness, contrast is reduced; fundamentally, the PSF of an eye with refractive errors is a low-pass filter—high image frequencies are irreversibly canceled out in the optical path from display to the retina. To overcome this limitation, Pamplona et

al. [338] proposed the use of conventional light field displays with lenslet arrays or parallax barriers to correct visual aberrations. For this application, these devices must provide a sufficiently high angular resolution so that multiple light rays emitted by a single lenslet enter the pupil. This resolution requirement is similar for light field displays supporting accommodation cues. Unfortunately, conventional light field displays as used by Pamplona et al. [338] are subject to a spatio-angular resolution tradeoff: an increased angular resolution decreases the spatial resolution. Hence, the viewer sees a sharp image but at a significantly lower resolution than that of the screen. To mitigate this effect, Huang et al. [186] recently proposed to use multilayer display designs together with prefiltering. While this is a promising, high-resolution approach, combining prefiltering and these particular optical setups significantly reduces the resulting image contrast.

4.8 CONCLUSION AND OUTREACH

We have presented a thorough literature review of recent advances in display technology, categorizing them along the multiple dimensions of the plenoptic function. Additionally, we have introduced the key aspects of the HVS that are relevant and/or leveraged by some of the new technologies. For readers also seeking an in-depth look into hardware descriptions, domain-specific books exist covering aspects such as physics or electronics, particular technologies like organic light-emitting diode (OLED), liquid crystal, LCD backlights or mobile displays [272, 155], or even how to build prototype compressive light field displays [480].

Advances in display technologies run somewhat parallel to advances in *capture* devices: Exploiting the strong correlations between the dimensions of the plenoptic function have allowed researchers and engineers to overcome basic limitations of standard capture devices. Examples of these include color demosaicing, or video compression [481]. The fact that both capture and display technologies are following similar paths makes sense, since both share the problem of the high dimensionality of the plenoptic function. In this regard, both fields can be seen as two sides of the same coin. On the other hand, advances in one will foster further research in the other: For instance, HDR displays have already motivated the invention of new HDR capture and compression algorithms, which in turn will create a demand for better HDR displays. Similarly, a requirement for light field displays to really take off is that light field content becomes more readily available (with companies like Lytro™ and Raytrix™ pushing in that direction).

Our categorization in this survey with respect to the plenoptic function is a convenient choice to support our current view of the field, but it should not be seen as a rigid scheme. We expect this division to become increasingly blurrier over the next few years, as some of the most novel technologies mature, coupled with superior computational power and a better knowledge of the HVS. The most important criteria nowadays for the consumer market seem to be spatial resolution, contrast, angular resolution (3D) and refresh rates.

High definition (ultra-high spatial resolution) is definitely one of the main current trends in the industry. A promising technology is based on IGZO (Indium

Gallium Zinc Oxide), a transparent amorphous oxide semiconductor (TAOS) whose TFT (Thin Film Transistor) performance increases electron mobility up to a factor of 50. This can lead to an improvement in resolution of up to ten times, plus the ability to fabricate larger displays [181]. Additionally, TAOS can be flexed, and have a lower consumption of power during manufacturing, because they can be fabricated at room temperature. The technology has already been licensed by JST (the Japan Science and Technology Agency) to several display manufacturing companies.

Other technologies have their specific challenges to meet before they become the driving force of the industry towards the consumer market. In the case of increased contrast, power consumption is one stumbling block for HDR displays, also shared by some types of automultiscopic displays. LCD panels transmit about 3% of light for pixels that are full on, which means that a lot of light is transduced into heat. For HDR displays, this translates into lots of energy consumed and wasted. OLED technology is a good candidate as a viable, more efficient technology. In the case of automultiscopic displays, parallax barriers entail very low light throughput as well, whereas LCD-based multilayer approaches multiply the efficiency problem times the number of LCD panels needed. While the field is very active, major challenges of automultiscopic displays that remain and have been discussed in this review include the need for a thin form factor, a solution to the currently still low spatio-angular resolution, limited depth of field, or the need for easier generation and transmission of the content.

While we have shown the recent advances and progress lines in each plenoptic dimension, we believe that real advances in the field need to come from a holistic approach to the problem: instead of focusing on one single dimension of the plenoptic function, future displays need to and will tackle several dimensions at the same time. For instance, current state-of-the-art broadcast systems achieve Ultra High Definition (UHD) with 8K at 120Hz progressive, with a deeper color gamut (Rec. 2020) than High Definition standards. This represents a significant advance in terms of spatial resolution, temporal resolution, and color. Similarly, we have seen how dynamic range and color appearance models, formerly two separate fields, are now being analyzed in conjunction in recent works, or how fast changes in the temporal domain can help increase apparent spatial resolution. Stereo techniques can be seen as just a particular case of auto-multiscopic displays, and these need to analyze spatial and angular resolution jointly. Joint stereoscopic high dynamic range displays (SHDR, also known as 3D-HDR) are also being developed and studied. This is and should be the trend for the future.

As technology advances, some of the inherent limitations of current displays (such as bandwidth in the case of light field displays) will naturally vanish, or progressively become less restricting. However, while some advances will rely purely on novel technology, optics and computation, we believe that perceptual aspects will continue to play a key role. Understanding the mechanisms of the HVS will be a crucial factor on which design decisions will be taken. For instance, SHDR directly involves the luminance contrast and angular dimensions of the plenoptic function. However, the perception of depth in high dynamic range displays is still not well known; some works have even hypothesized that HDR content may hinder stereo acuity [23]. In any case it is believed that the study of

binocular disparity alone, on which most of the existing research has focused, is not enough to understand the perception of a 3D structure [16]. Although we are gaining a more solid knowledge on how to combat the vergence-accommodation conflict, or what components in a scene may introduce discomfort to the viewer, key aspects of the HVS such as cue integration, or the interrelation of the different visual signals, remain largely unexplored. As displays become more sophisticated and advanced, a deeper understanding of our visual system will be needed, including hard-to-measure aspects such as viewing comfort.

Last, a different research direction which has seen some first practical implementations aims at integrating the displayed imagery with the physical world, blurring out the boundaries imposed by the form factors of more traditional displays. Examples of this include systems that augment the appearance of objects by means of superimposed projections [466, 13]; compositing real and synthetic objects in the same scene, taking into account interreflections between them [83]; adjusting the appearance of the displayed content according to the incident real illumination [328]; or allowing for gestured-based interaction [170]. Some of these approaches rely on the integration and combined operation of displays, projectors and cameras, all of them enhanced with computational capabilities. This is another promising avenue of future advances, although integrating hardware from different manufacturers may impose some additional practical difficulties. Another exciting, recent technology is *printed optics* [486, 442], which enables display, sensing and illumination elements to be directly printed inside an interactive device. While still in its infancy, this may open up a whole new field, where every object will in the future act as a display.

To summarize, we believe that future displays will rely on joint advances on several different dimensions. Additional influencing factors include further exploration of aspects such as polarization, or multispectral imaging; new materials; the adaptation of mathematical models for high-performance real-time computation; or the co-design of custom optics and electronics. We are convinced that a deeper understanding of the HVS will play a key role as well, with perceptual effects and limitations being taken into account in future display designs. Display technology encompasses a very broad field which will benefit from close collaboration from the different areas of research involved. From hardware specialists to psychophysicists, including optics experts, material scientists, or signal processing specialists, multidisciplinary co-operation will be the key.

5

DISPLAY-ADAPTIVE 3D CONTENT REMAPPING

ABOUT THIS CHAPTER

The work here presented has been done in collaboration with the Camera Culture Group at MIT Media Lab. The collaboration, and the work described in this chapter, started with my first internship in the group as a visiting student during my PhD studies. The group has a renowned expertise in the field of computational displays, and has, in the last years, developed a new generation of light field displays termed *tensor displays*. One of the problems of these tensor displays, shared also by the most common types of light field displays, is their limited depth-of-field. This is the problem we explain and address in this chapter. We proposed a solution, based on an optimization that incorporates computational models of perception, which we describe here and has been accepted for publication in the journal Computers & Graphics in a special issue on Advanced Displays.

B. Masia, G. Wetzstein, C. Aliaga, R. Raskar and D. Gutierrez.
Display Adaptive 3D Content Remapping.
In Computers & Graphics 2013, *to appear*.

5.1 INTRODUCTION

Within the last years, stereoscopic and automultiscopic displays have started to enter the consumer market from all angles. These displays can show three-dimensional objects that appear to be floating in front of or behind the physical screen, even without the use of additional eyewear. Capable of electronically switching between a full-resolution 2D and a lower-resolution 3D mode, parallax barrier technology [198] is dominant for hand-held and tablet-sized devices, while medium-sized displays most often employ arrays of microlenses [266]. Although most cinema screens today are stereoscopic and rely on additional eyewear, large-scale automultiscopic projection systems are an emerging technology [180]. Each technology has its own particular characteristics, including field of view, depth of field, contrast, resolution, and screen size. Counterintuitively, produced content is usually targeted toward a single display configuration, making labor-intense, manual post-processing of the recorded or rendered data necessary.

Display-adaptive content retargeting is common practice for attributes such as image size, dynamic range (tone mapping), color gamut, and spatial resolution [36]. In order to counteract the accommodation-convergence mismatch of stereoscopic displays, stereoscopic disparity retargeting methods have recently been explored [223, 252, 220, 106, 108]. These techniques are successful in modifying the disparities of a stereo image pair so that visual discomfort of the observer is mitigated while preserving the three-dimensional appearance of the scene as

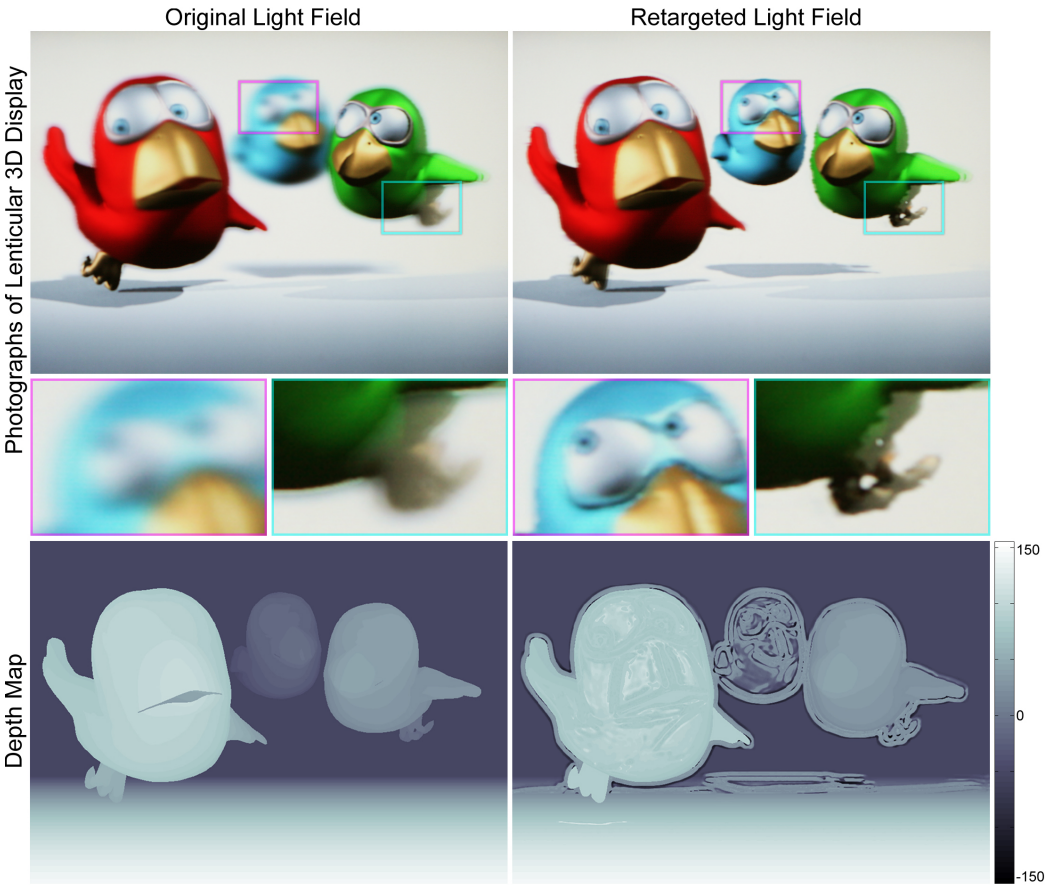


Figure 5.1: Our 3D content retargeting for a glasses-free lenticular display. Due to the limited depth of field of all light field displays, some objects in a 3D scene will appear blurred. Our remapping approach selectively fits the 3D content into the depth budget of the display, while preserving the perceived depth of the original scene. Top: actual photographs of the original and retargeted scenes, as seen on a Toshiba GL1 lenticular display. Notice the improvement in the blue bird or the legs of the green bird in the retargeted version. Middle: close-ups. Bottom: original and retargeted depths yielded by our method.

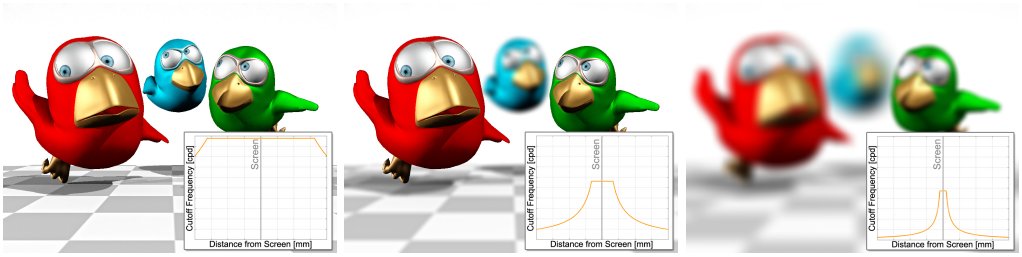


Figure 5.2: Simulated views of the *three-birds* scene for three different displays. From left to right: Holografika HoloVizio C80 movie screen, desktop and cell phone displays. The last two displays fail to reproduce it properly, due to their intrinsic depth-of-field limitations. The insets plot the depth vs. cut-off frequency charts for each display.

much as possible. Inspired by these techniques, we tackle the problem of 3D content retargeting for glasses-free light field (i.e. automultiscopic) displays. These displays exhibit a device-specific depth of field (DOF) that is governed by their limited angular resolution [512, 482]. Due to the fact that most light field displays only provide a low angular resolution, that is the number of viewing zones, the supported DOF is so shallow that virtual 3D objects extruding from the physical display enclosure appear blurred out (see Figs. 5.1, left, and 5.2 for a real photograph and a simulation showing the effect, respectively). We propose here a framework that remaps the disparities in a 3D scene to fit the DOF constraints of a target display by means of an optimization scheme that leverages perceptual models of the human visual system. Our optimization approach runs on the central view of an input light field and uses warping to synthesize the rest of the views.

CONTRIBUTIONS. Our nonlinear optimization framework for 3D content retargeting specifically provides the following contributions:

- We propose a solution to handle the intrinsic trade-off between the spatial frequency that can be shown and the perceived depth of a given scene. This is a fundamental limitation of automultiscopic displays (see Section 5.3).
- We combine exact formulations of display-specific depth of field limitations with models of human perception, to find an optimized solution. In particular, we consider the frequency-dependent sensitivity to contrast of the human visual system, and the sensitivity to binocular disparity. Based on this combination, a first objective term minimizes the perceived luminance and contrast difference between the original and the displayed scene, effectively minimizing DOF blur, while a second term strives to preserve the perceived depth.
- We validate our results with existing state-of-the-art, objective metrics for both image quality and perceived depth.
- We show how our framework can be easily extended to the particular case of *stereoscopic* disparity, thus demonstrating its versatility.
- For this extension, we account for a non-dichotomous zone of viewing comfort which constitutes a more accurate model of discomfort associated with the viewing experience.

As a result of our algorithm, the depth of a given 3D scene is modified to fit the DOF constraints imposed by the target display, while preserving the perceived 3D appearance and the desired 2D image fidelity (Figure 5.1, right).

LIMITATIONS. We do not aim at providing an accurate model of the behavior of the human visual system; investigating all the complex interactions between its individual components remains an open problem as well, largely studied by both psychologists and physiologists. Instead, we rely on existing computational models of human perception and apply them to the specific application of 3D

content retargeting. For this purpose, we currently consider sensitivities to luminance contrast and depth, but only approximate the complex interaction between these cues using a heuristic linear blending, which works well in our particular setting. Using the contrast sensitivity function in our context (Section 5.4) is a convenient but conservative choice. Finally, depth perception from motion parallax exhibits strong similarities in terms of sensitivity with that of binocular disparity, suggesting a close relationship between both [376]; but existing studies on sensitivity to motion parallax are not as exhaustive as those on binocular disparity, and therefore a reliable model cannot be derived yet. Moreover, some studies have shown that, while both cues are effective, stereopsis is more relevant by an order of magnitude [60]. In any case, our approach is general enough so that as studies on these and other cues advance and new, more sophisticated models of human perception become available, they could be incorporated to our framework.

5.2 RELATED WORK

Glasses-free 3D displays were invented more than a century ago, but even today, the two dominating technologies are parallax barriers [198] and integral imaging [266]. Nowadays, the palette of existing 3D display technologies, however, is much larger and includes holograms, volumetric displays, multilayer displays and directional backlighting among many others. State of the art reviews of conventional stereoscopic and automultiscopic displays [451] and computational displays [477] can be found in the literature. With the widespread use of stereoscopic image capture and displays, optimal acquisition parameters and capture systems [267, 311, 209, 335, 165], editing tools [237, 469], and spatial resolution retargeting algorithms for light fields [51] have recently emerged. In this work, we deal with the problem of depth remapping of light field information to the specific constraints of each display.

Generally speaking, content remapping is a standard approach to adapt spatial and temporal resolution, contrast, colors, and sizes of images to a display having limited capabilities in any of these dimensions [36]. For the particular case of disparity remapping, Lang et al. [252] define a set of non-linear disparity remapping operators, and propose a new stereoscopic warping technique for the generation of the remapped stereo pairs. A metric to assess the magnitude of perceived changes in binocular disparity is introduced by Didyk et al. [106], who also investigate the use of the Cornsweet illusion to enhance perceived depth [107]. Recently, the original disparity metric has been further refined including the effect of luminance-contrast [108]. Kim and colleagues [220] develop a a novel framework for flexible manipulation of binocular parallax, where a new stereo pair is created from two non-linear cuts of the EPI volume corresponding to multiperspective images [394]. Inspired by Lang and colleagues [252], they explore linear and non-linear global remapping functions, and also non-linear disparity gradient compression. Here we focus on a remapping function that incorporates the specific depth of field limitations of the target display [306]. Section 5.8 provides direct comparisons with some of these approaches.

5.3 DISPLAY-SPECIFIC DEPTH OF FIELD LIMITATIONS

Automultiscopic displays are successful in creating convincing illusions of three-dimensional objects floating in front and behind physical display enclosures without the observer having to wear specialized glasses. Unfortunately, all such displays have a limited depth of field which, just as in wide-aperture photography, significantly blurs out-of-focus objects. The focal plane for 3D displays is directly on the physical device. Display-specific depth of field expressions have been derived for parallax barrier and lenslet-based systems [512], multilayer displays [482], and directional backlit displays [478]. In order to display an aliasing-free light field with any such device, four-dimensional spatio-angular pre-filters need to be applied before computing the display-specific patterns necessary to synthesize a light field, either by means of sampling or optimization. In practice, these filters model the depth-dependent blur of the individual displays and are described by a depth of field blur applied to the target light field. Intuitively, this approach fits the content into the DOF of the displays by blurring it as necessary. Figure 5.3 illustrates the supported depth of field of various automultiscopic displays for different display sizes.

Specifically, the depth of field of a display is modeled as the maximum spatial frequency f_ξ of a diffuse plane at a distance d_0 to the physical display enclosure. As shown by previous works [512, 482], the DOF of parallax barrier and lenslet-based displays is given by

$$|f_\xi| \leqslant \begin{cases} \frac{f_0}{N_a}, & \text{for } |d_0| + (h/2) \leqslant N_a h \\ (\frac{h}{(h/2)+|d_0|}) f_0, & \text{otherwise} \end{cases}, \tag{23}$$

where N_a is the number of angular views, d_0 is the distance to the front plane of the display (i.e. the parallax barrier or lenslet array plane), h represents the thickness of the display, $f_0 = 1/(2p)$, and p is the size of the view-dependent subpixels of the back layer of the display, making the maximum resolution of the display at the front surface $f_\xi = f_0/N_a = 1/(2pN_a)$. For multilayered displays, the upper bound on the depth of field for a display of N layers was derived by Wetzstein et al. [482] to be

$$|f_\xi| \leqslant N f_0 \sqrt{\frac{(N+1)h^2}{(N+1)h^2 + 12(N-1)d_0^2}}. \tag{24}$$

Note that in this case d_0 represents the distance to the middle of the display, and p the pixel size of the layers.

It can be seen how depth of field depends on display parameters such as pixel size p, number of viewing zones N_a, device thickness h, and number of layers N (for multilayer displays), and thus varies significantly for different displays. It also depends on the viewing distance v_D when expressed in cycles per degree. The above expressions can then be employed to predict an image displayed on a particular architecture, including loss of contrast and blur. Figure 5.2 shows three simulated views of the *three-birds* scene for three different displays: a Holografika HoloVizio C80 movie screen ($h = 100mm$, $p = 0.765mm$, $v_D = 6m$), a

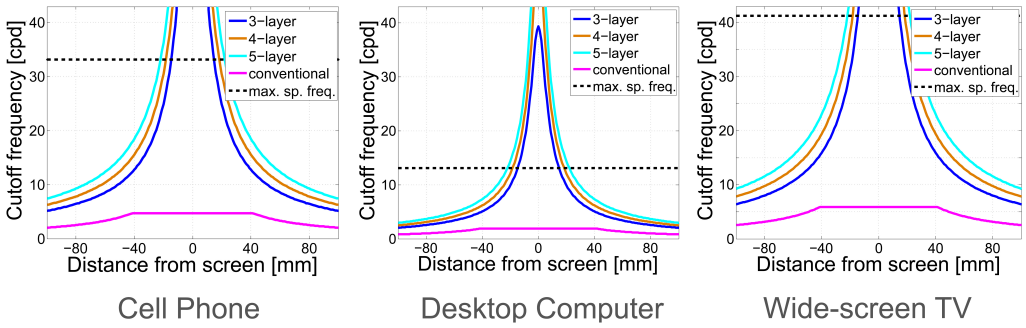


Figure 5.3: Depth of field for different display architectures and target displays. From left to right: cell phone ($p = 0.09mm$, $\nu_D = 0.35m$); desktop computer ($p = 0.33mm$, $\nu_D = 0.5m$); and widescreen TV ($p = 0.53mm$, $\nu_D = 2.5m$). For comparison purposes all depths of field are modeled for seven angular views.

Toshiba automultiscopic monitor ($h = 20$, $p = 0.33$, $\nu_D = 1.5$) and a cell-phone-sized display ($h = 6$, $p = 0.09$, $\nu_D = 0.35$). The scene can be represented in the large movie screen without blurring artifacts (left); however, when displayed on a desktop display (middle), some areas appear blurred due to the depth-of-field limitations described above (see the blue bird). When seen on a cell-phone display (right), where the limitations are more severe, the whole scene appears badly blurred. In the following, we show how these predictions are used to optimize the perceived appearance of a presented scene in terms of image sharpness and contrast, where the particular parameters of the targeted display are an input to our method.

5.4 OPTIMIZATION FRAMEWORK

In order to mitigate display-specific DOF blur artifacts, we propose to scale the original scene into the provided depth budget while preserving the perceived 3D appearance as best as possible. As detailed in Section 5.3, this is not trivial, since there is an intrinsic trade-off between the two goals. We formulate this as a multi objective optimization problem, with our objective function made up of two terms. The first one minimizes the perceived luminance and contrast difference between the original and the displayed scene, for which display-specific expressions of the displayable frequencies are combined with a perceptual model of contrast sensitivity. The second term penalizes loss in perceived depth, for which we leverage disparity sensitivity metrics. Intuitively, the disparity term prevents the algorithm from yielding the obvious solution where the whole scene is flattened onto the display screen; this would guarantee perfect focus at the cost of losing any sensation of depth. The input to our algorithm is the depth map and the luminance image of the central view of the original light field, which we term d_{orig} and L_{orig}, respectively. The output is a retargeted depth map d, which is subsequently used to synthesize the retargeted light field.

OPTIMIZING LUMINANCE AND CONTRAST: We model the display-specific frequency limitations by introducing spatially-varying, depth-dependent convo-

lution kernels $k(d)$. They are defined as Gaussian kernels whose standard deviation σ is such that frequencies above the cut-off frequency at a certain depth $f_\xi(d)$ are reduced to less than 5% of its original magnitude. Although more accurate image formation models for defocus blur in scenes with occlusions can be found in the literature [158], their use is impractical in our optimization scenario, and we found the Gaussian spatially-varying kernels to give good results in practice. Kernels are normalized so as not to modify the total energy during convolution. As such, the kernel for a pixel i is:

$$k(d) = \frac{exp(-\frac{x_i^2+y_i^2}{2(\sigma(d))^2})}{\sum_j^K \left(exp(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2}) \right)} \tag{25}$$

where K is its number of pixels. The standard deviation σ is computed as:

$$\sigma(d) = \frac{\sqrt{-2log(0.05)}}{2\pi p f_\xi(d)} \tag{26}$$

with p being the pixel size in $mm/pixel$.

To take into account how frequency changes are perceived by a human observer, we rely on the fact that the visual system is more sensitive to near-threshold changes in contrast and less sensitive at high contrast levels [293]. We adopt a conservative approach and employ sensitivities at near-threshold levels as defined by the contrast sensitivity function (CSF). We follow the expression for contrast sensitivities ω_{CSF} proposed by Mantiuk et al. [294], which in turn builds on the model proposed by Barten [42]:

$$\omega_{CSF}(l, f_l) = p_4 s_A(l) \frac{MTF(f_l)}{\sqrt{(1 + (p_1 f_l)^{p_2})(1 - e^{-(f_l/7)^2})^{-p_3}}}, \tag{27}$$

where l is the adapting luminance in $[cd/m^2]$, f_l represents the spatial frequency of the luminance signal in [cpd] and p_i are the fitted parameters provided in Mantiuk's paper[1]. MTF (modulation transfer function) and s_A represent the optical and the luminance-based components respectively, and are given by:

$$MTF(f_l) = \sum_{k=1..4} a_k e^{-b_k f_l} \tag{28}$$

$$s_A(l) = p_5 \left(\left(\frac{p_6}{l} \right)^{p_7} + 1 \right)^{-p_8} \tag{29}$$

where a_k and b_k can again be found in the original paper. Figure 5.4 (left) shows contrast sensitivity functions for varying adaptation luminances, as described by Equations 27-29. In our context we deal with complex images, as opposed to a uniform field; we thus use the steerable pyramid [407] $\rho_S(\cdot)$ to decompose a luminance image into a multi-scale frequency representation. The steerable pyramid is chosen over other commonly used types of decomposition (e.g. Cortex Transform) since it is mostly free of ringing artifacts that can cause false masking signals [294].

1 sourceforge.net/apps/mediawiki/hdrvdp/

Taking into account both the display-specific frequency limitations and the HVS response to contrast, we have the following final expression for the first term of our optimization:

$$\left\| \omega_{CSF} \left(\rho_S \left(L_{orig} \right) - \rho_S \left(\phi_b \left(L_{orig}, d \right) \right) \right) \right\|_2^2, \tag{30}$$

where ω_{CSF}, defined by Equation 27, are frequency-dependent weighting factors, and the operator $\phi_b(L, d) = k(d) * L$ models the display-specific, depth-dependent blur (see Section 5.3 and Figure 5.3). Note that we omit the dependency of ω_{CSF} on (l, f_l) for clarity. Figure 5.5 (*left*) shows representative weights ω_{CSF} for different spatial frequency luminance levels of the pyramid for a sample scene.

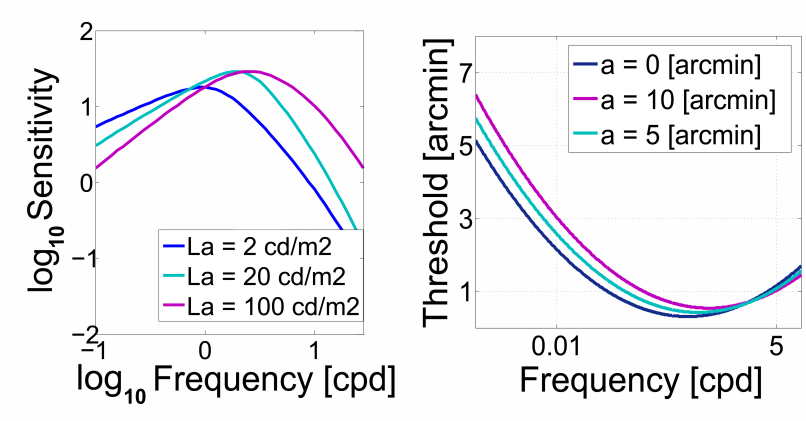


Figure 5.4: Thresholds and sensitivity values from which the weights for our optimization are drawn. Left: Contrast sensitivity functions. Right: Binocular disparity discrimination thresholds (thresholds are the inverse of sensitivities).

PRESERVING PERCEIVED DEPTH: This term penalizes the perceived difference in depth between target and retargeted scene using disparity sensitivity metrics. As noted by different researchers, the effect of binocular disparity in the perception of depth works in a manner similar to the effect of contrast in the perception of luminance [106, 58, 18]. In particular, our ability to detect and discriminate depth from binocular disparity depends on the frequency and amplitude of the disparity signal. Human sensitivity to binocular disparity is given by the following equation [106] (see also Figure 5.4, right):

$$\begin{aligned} \omega_{BD}(a, f) = (&0.4223 + 0.007576a + 0.5593\log_{10}(f) \\ &+ 0.03742a\log_{10}(f) + 0.0005623a^2 + 0.7114\log_{10}^2(f))^{-1} \end{aligned} \tag{31}$$

where frequency f is expressed in [cpd], a is the amplitude in [arcmin], and ω_{BD} is the sensitivity in [arcmin^{-1}]. In a similar way to ω_{CSF} in Equation 30, the weights ω_{BD} account for our sensitivity to disparity amplitude and frequency. Given this dependency on frequency, the need for a multi-scale decomposition of image disparities arises again, for which we use a Laplacian pyramid $\rho_L(\cdot)$ for efficiency reasons, following the proposal by Didyk et al. [106]. Figure 5.5 (*right*), shows representative weights ω_{BD}.

The error in perceived depth incorporating these sensitivities is then modeled with the following term:

$$\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(d_{orig}\right)\right)-\rho_L\left(\phi_v\left(d\right)\right)\right)\right\|_2^2. \tag{32}$$

Given the viewing distance v_D and interaxial distance e, the operator $\phi_v(\cdot)$ converts depth into vergence as follows:

$$\phi_v(d)=\operatorname{acos}\left(\frac{\mathbf{v_L}\cdot\mathbf{v_R}}{\|\mathbf{v_L}\|\,\|\mathbf{v_R}\|}\right), \tag{33}$$

where vectors $\mathbf{v_L}$ and $\mathbf{v_R}$ are illustrated in Figure 5.6. The Laplacian decomposition transforms this vergence into frequency-dependent disparity levels.

OBJECTIVE FUNCTION: Our final objective function is a combination of Equations 30 and 32:

$$\begin{aligned}\arg\min_d \Big(&\mu_{DOF}\left\|\omega_{CSF}\left(\rho_S\left(L_{orig}\right)-\rho_S\left(\phi_b\left(L_{orig},d\right)\right)\right)\right\|_2^2\\ &+\mu_D\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(d_{orig}\right)\right)-\rho_L\left(\phi_v\left(d\right)\right)\right)\right\|_2^2\Big).\end{aligned} \tag{34}$$

For multilayer displays, we empirically set the values of $\mu_{DOF}=10$ and $\mu_D=0.003$, while for conventional displays $\mu_D=0.0003$ due to the different depth of field expressions.

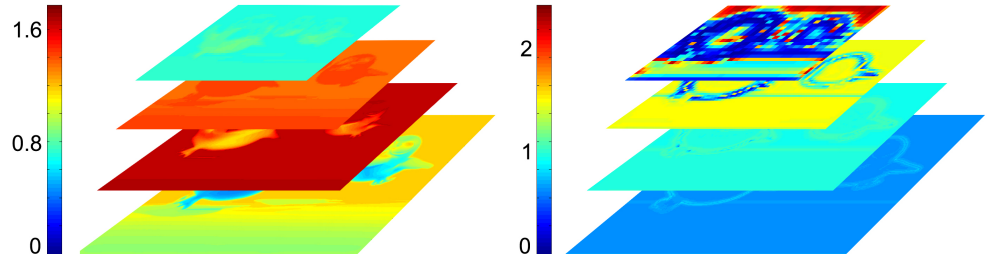


Figure 5.5: Left: Weights ω_{CSF} (contrast sensitivity values) for different luminance spatial frequency levels for a sample scene (*birds*). Right: Weights ω_{BD} (inverse of discrimination threshold values) for different disparity spatial frequency levels for the same scene.

5.5 IMPLEMENTATION DETAILS

We employ a large-scale trust region method [81] to solve Equation 34. This requires finding the expressions for the analytic gradients of the objective function used to compute the Jacobian, which can be found in Appendix E. The objective term in Equation 34 models a single view of the light field, i.e. the central view, in a display-specific field of view (FOV). Within a moderate FOV, as provided by commercially-available displays, this is a reasonable approximation; we obtain the rest of the light field by warping. In the following, we describe this and other additional implementation details.

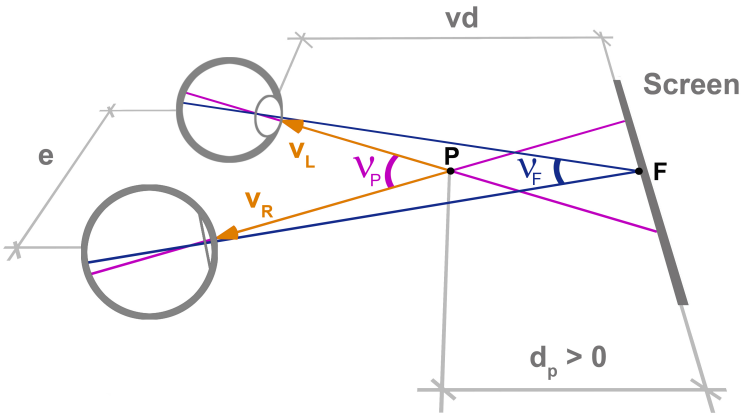


Figure 5.6: Computing vergence values. Vergence ν_P of a point P depends on its position, the viewing distance ν_D and the interaxial e. The corresponding disparity for P is $(\nu_P - \nu_F)$. vd refers to the viewing distance and d_P is the depth of point P.

SENSITIVITY WEIGHTS AND TARGET VALUES: The weights used in the different terms, ω_{CSF} and ω_{BD} are pre-computed based on the values of the original depth and luminance, d_{orig} and L_{orig}. The transformation from d_{orig} to vergence, its pyramid decomposition and the decomposition of L_{orig} are also pre-computed.

CONTRAST SENSITIVITY FUNCTION: As reported by Mantiuk et al. [294], no suitable data exists to separate L- and M-cone sensitivity. Following their approach, we rely on the *achromatic* CSF using only luminance values.

DEPTH-OF-FIELD SIMULATION: The depth-dependent image blur of automultiscopic displays is modeled as a spatially-varying convolution in each iteration of the optimization procedure. Due to limited computational resources, we approximate this expensive operation as a blend between multiple shift-invariant convolutions corresponding to a quantized depth map, making the process much more efficient. For all scenes shown in this chapter, we use $n_c = 20$ quantized depth clusters.

WARPING: View warping is orthogonal to the proposed retargeting approach; we implement here the method described by Didyk et al. [105], although other methods could be employed instead (e.g. [220, 276, 334]). To reduce warping artifacts due to large depth gradients at the limits of the field of view for each light field, we median-filter the depth and constrain depth values around the edges.

5.6 RETARGETING FOR STEREOSCOPIC DISPLAYS

One of the advantages of our framework is its versatility, which allows to adapt it for display-specific disparity remapping of stereo pairs. We simply drop the depth of field term from Equation 34, and incorporate a new term that models

the comfort zone. This is an area around the screen within which the 3D content does not create fatigue or discomfort in the viewer in stereoscopic displays, and is usually considered as a dichotomous subset of the fusional area. Although any comfort-zone model could be directly plugged into our framework, we incorporate the more accurate, non-dichotomous model suggested by Shibata et al. [401]. This model provides a more accurate description of its underlying psychological and physiological effects. Additionally, this zone of comfort depends on the viewing distance v_D, resulting on different expressions for different displays, as shown in Figure 5.7. Please refer to Appendix F for details on how to incorporate the simpler, but less precise, dichotomous model.

Our objective function thus becomes:

$$\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(D_{orig}\right)\right)-\rho_L\left(\phi_v\left(d\right)\right)\right)\right\|_2^2+\mu_{CZ}\left\|\varphi\left(d\right)\right\|_2^2, \tag{35}$$

where $\varphi(\cdot)$ is a function mapping depth values to visual discomfort:

$$\varphi(d)=\begin{cases}1-\frac{s_{far}}{v_D-d}-T_{far} & \text{for} \quad d<0\\ 1-\frac{s_{near}}{v_D-d}-T_{near} & \text{for} \quad d\geqslant 0\end{cases} \tag{36}$$

where v_D is the distance from the viewer to the central plane of the screen and s_{far}, s_{near}, T_{far}, and T_{near} are values obtained in a user study carried out with 24 subjects.

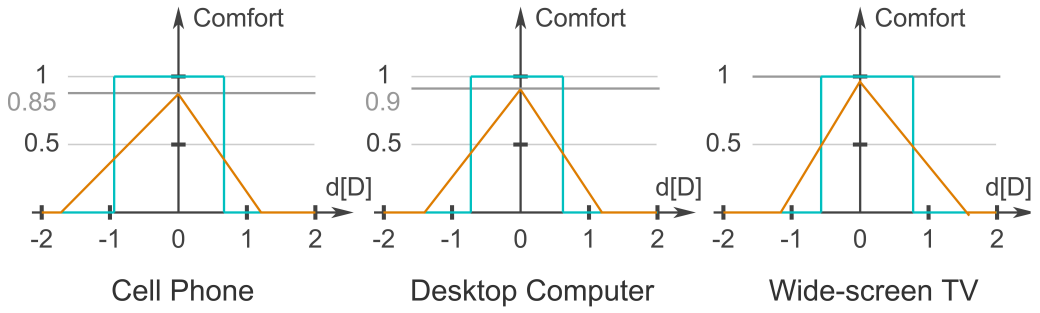


Figure 5.7: Dichotomous (blue) and non-dichotomous (orange) zones of comfort for different devices. From left to right: cell phone ($v_D = 0.35$m), desktop computer ($v_D = 0.5$m) and wide-screen TV ($v_D = 2.5$m).

5.7 RESULTS

We have implemented the proposed algorithm for different types of automultiscopic displays including a commercial Toshiba GL1 lenticular-based display providing horizontal-only parallax with nine discrete viewing zones, and custom multilayer displays. The Toshiba panel has a native resolution of 3840×2400 pixels with a specially engineered subpixel structure that results in a resolution of 1280×800 pixels for each of the nine views. Note that even a highly-engineered device such as this suffers from a narrow depth of field due to the limited angular sampling. We consider a viewing distance of 1.5 m for the Toshiba display and 0.5 m for the multilayer prototypes.

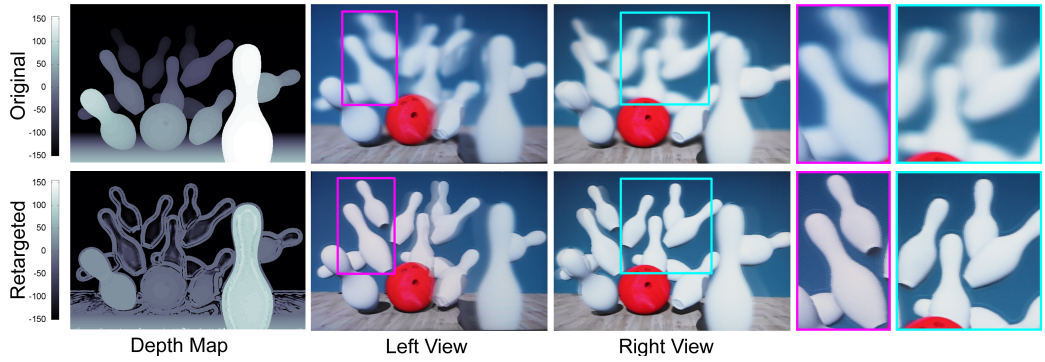


Figure 5.8: Additional results for commercial lenticular display (actual photographs). Top row: depth map, perspective from left, and perspective from right for original scene. Bottom row: depth map and similar perspectives for the retargeted scene. The slight double-view of some of the pins in the left view is due to interview cross-talk in the Toshiba display.

Figures 5.1 and 5.8 show results of our algorithm for the Toshiba display. The target scenes have been originally rendered as light fields with a resolution of 9×9, with a field of view of $10°$. Since the Toshiba display only supports horizontal parallax, we only use the nine horizontal views for these examples. Note how depth is compressed to fit the display's constraints in those areas with visible loss of contrast due to blur (blue bird or far away pins, for instance), while enhancing details to preserve the *perceived* depth; areas with no visible blur are left untouched (eyes of the green bird, for instance). This results into sharper retargeted scenes that can be shown within the limitations of the display. The remapping for the teaser image took two hours for a resolution of 1024×768, using our unoptimized Matlab code.

We have also fabricated a prototype multilayer display (Figure 5.9). This display is composed of five inkjet-printed transparency patterns spaced by clear acrylic sheets. The size of each layer is 60×45 mm, while each spacer has a thickness of 1/8". The transparencies are conventional films for office use and the printer is an Epson Stylus Photo 2200. This multilayer display supports 7×7 views within a field of view of $7°$ for both horizontal and vertical parallax. The patterns are generated with the computed tomography solver provided by Wetzstein et al. [482]. Notice the significant sharpening of the blue bird and, to a lesser extent, of the red bird. It should be noted that these are lab prototypes: scattering, inter-reflections between the acrylic sheets, and imperfect color reproduction with the desktop inkjet printer influence the overall quality of the physical results. In Figure 5.10, we show sharper, simulated results for the *dice* scene for a similar multilayer display.

We show additional results using more complex data sets, with varying degrees of depth and texture, and different object shapes and surface material properties. In particular, we use the Heidelberg light field archive[2], which includes ground-truth depth information. The scenes are optimized for a three-layer multilayer display, similar to the one shown in Figure 5.9. They have been optimized for a viewing distance of 0.5 m and have resolutions ranging from 768×768 to 1024×720. The weights used in the optimization are again $\mu_{DOF} = 10$ and $\mu_D =$

2 http://hci.iwr.uni-heidelberg.de/HCI/Research/LightField/lf_archive.php

0.003. Figure 5.11 shows the results for the *papillon*, *buddha2* and *statue* data sets. Our algorithm recovers most of the high frequency content of the original scenes, lost by the physical limitations of the display. The anaglyph representations allow to compare the perceived depth of the original and the retargeted scenes (the reader may refer to `http://webdiis.unizar.es/~bmasia/downloads/thesis/Displays-DisparityRemapping.zip` for larger versions to ensure proper visualization). Figure 5.12 shows additional views of the *buddha2* and *statue* light fields.

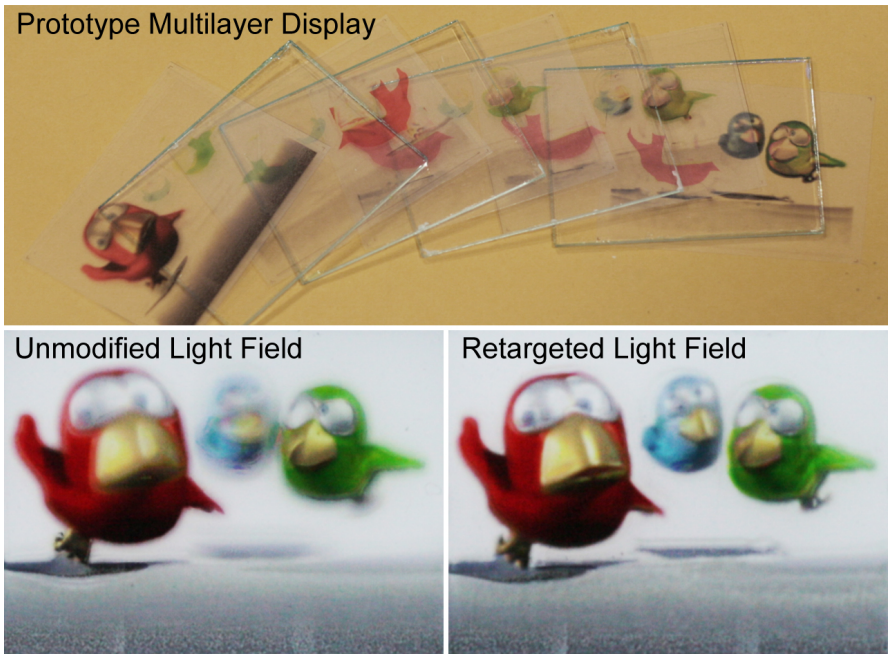


Figure 5.9: 3D content retargeting for multilayer light field displays (actual photographs). Even five attenuating layers (top) can only provide a limited depth of field for a displayed scene (bottom left). Our retargeting algorithm maps the multiview content into the provided depth budget (bottom right).

As shown in this section, our algorithm works well within a wide range of displays and data sets of different complexities. However, in areas of very high frequency content, the warping step may accumulate errors which end up being visible in the extreme views of the light fields. Figure 5.13 shows this: the *horses* data set contains a background made up of a texture containing printed text. Although the details are successfully recovered by our algorithm, the warping step cannot deal with the extremely high frequency of the text, and the words appear broken and illegible.

Finally, Figure 5.14 shows the result of applying our adapted model to the particular case of stereo retargeting, as described in Section 5.6.

5.8 COMPARISON TO OTHER METHODS

Our method is the first to specifically deal with the particular limitations of automultiscopic displays (depth vs. blur trade-off), and thus it is difficult to directly compare with others. However, we can make use of two recently published *objective* computational metrics, to measure distortions both in the observed 2D image fidelity, and in the perception of depth. This also provides an objective background to compare against existing approaches for stereoscopic disparity retargeting, for which alternative methods do exist.

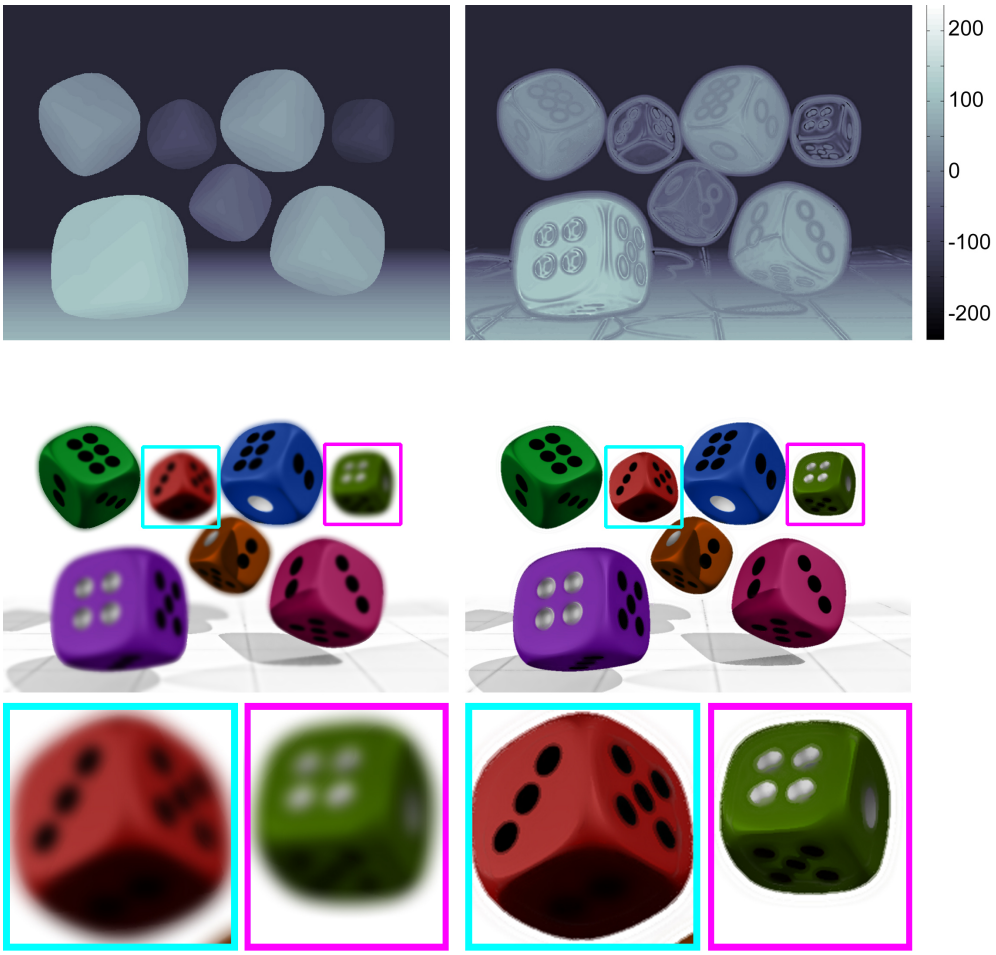


Figure 5.10: Results of simulations for a multilayer display (five layers). Top row: initial and retargeted depth. Middle row: initial and retargeted luminance. Bottom row: close-ups.

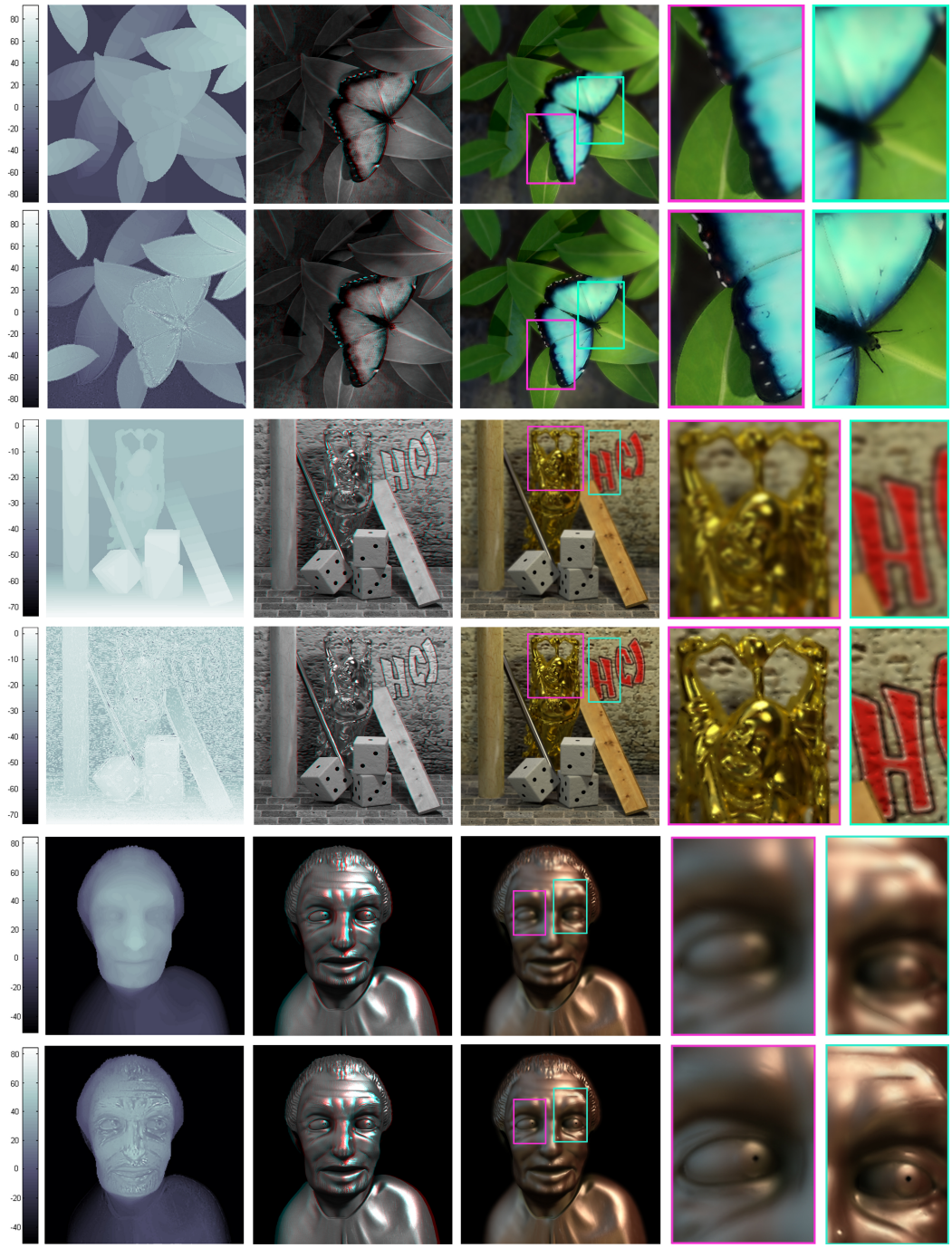


Figure 5.11: Results for the *papillon* (top), *buddha2* (middle) and *statue* (bottom) data sets from the Heidelberg light field archive. For each data set, the top row shows the original scene, while the bottom row shows our retargeted result. From left to right: depth map, anaglyph representation, central view image, and selected zoomed-in regions. Notice how our method recovers most of the high frequency details of the scenes, while preserving the sensation of depth (larger versions of the anaglyphs can be seen in `http://webdiis.unizar.es/~bmasia/downloads/thesis/Displays-DisparityRemapping.zip`). Note: please wear anaglyph glasses with cyan filter on left and red filter on right eye; for an optimal viewing experience please resize the anaglyph to about 10 cm wide in screen space and view it at a distance of 0.5 m.



Figure 5.12: Additional non-central views of the retargeted *buddha2* and *statue* light fields, with corresponding close-ups.

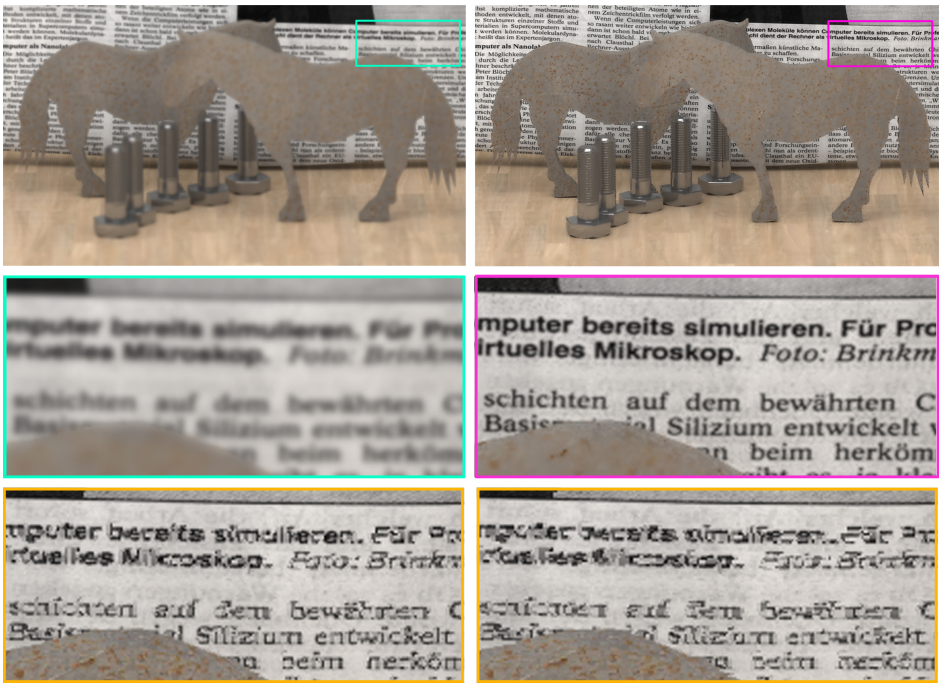


Figure 5.13: Results for the *horses* data set from the Heidelberg light field archive. Very high frequencies that have been initially cut off by the display (green box) are successfully recovered by our algorithm (pink). However, subsequent warping can introduce visible artifacts in those cases, which progressively increase as we depart from the central view of the light field. This progression is shown in the bottom row (yellow boxes).

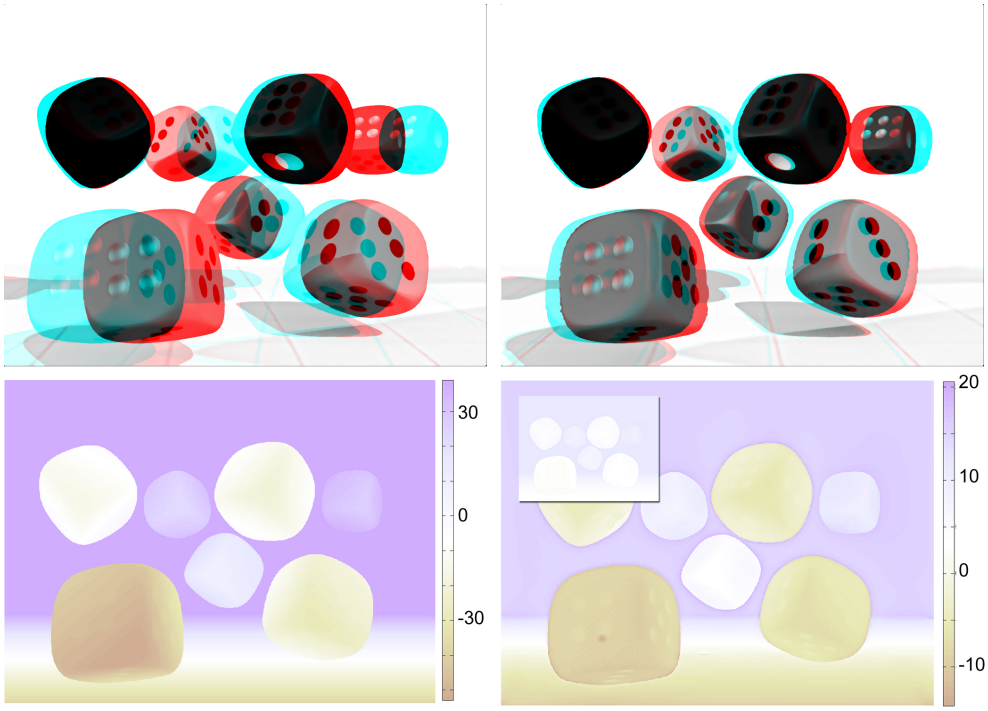


Figure 5.14: Retargeting for stereo content. *Left column:* Anaglyph and corresponding pixel disparity map of the original scene. For a common (around 0.5m) viewing distance on a desktop display, left and right images cannot be fused. *Right column:* Anaglyph and corresponding pixel disparity map of the retargeted scene. Images can now be fused without discomfort, and perception of depth is still present despite the aggressive depth compression. Note that the scales of the disparity maps are different for visualization purposes; the small inset shows the retargeted disparity map for the same scale as the original. Note: please wear anaglyph glasses with cyan filter on left and red filter on right eye; for an optimal viewing experience please resize the anaglyph to about 10 cm wide in screen space and view it at a distance of 0.5 m.

METRICS: We need to measure *both* observed 2D image quality *and* resulting degradations in perceived depth. For image quality, numerous metrics exist. We rely on the HDR-VDP 2 calibration reports provided by Mantiuk and colleagues [294] in their website[3] , where the authors compare quality predictions from six different metrics and two image databases: LIVE [399] and TID2008 [352]. According to the prediction errors, reported as Spearman's correlation coefficient, multi-scale SSIM (MS-SSIM, [509]) performs best across both databases for the blurred image distortions observed in our application. The mapping function we use, $\log(1 - \text{MS-SSIM})$, yields the highest correlation for Gaussian blur distortions.

Fewer metrics exist to evaluate distortions in depth. We use the metric recently proposed by Didyk and colleagues to estimate the magnitude of the perceived disparity change between two stereo images [106]. The metric outputs a heat map of the differences between the original and the retargeted disparity maps in Just Noticeable Difference (JND) units.

ALTERNATIVE METHODS: There is a large space of linear and non-linear global remapping operators, as well as of local approaches. Also, these operators can be made more sophisticated, for instance by incorporating information from saliency maps, or adding the temporal domain [252]. To provide some context to the results of the objective metrics, we compare our method with a representative subset of alternatives, including global operators, local operators, and a recent operator based on a perceptual model for disparity. In particular, we compare against six other results using different approaches for stereo retargeting: a linear scaling of pixel disparity (*linear*), a linear scaling followed by the addition of bounded Cornsweet profiles at depth discontinuities (*Cornsweet* [107])[4], a logarithmic remapping (*log*, see e.g. [252]), and the recently proposed remapping of disparity in a perceptually linear space (*perc. linear* [106]). For the last two, we present two results using different parameters. This selection of methods covers a wide range from very simple to more sophisticated.

The linear scaling is straightforward to implement. For the bounded Cornsweet profiles method, where profiles are carefully controlled so that they do not exceed the given disparity bounds and create disturbing artifacts, we choose $n = 5$ levels as suggested by the authors. For the logarithmic remapping, we use the following expression, inspired by Lang et al. [252]:

$$\delta_o = K \cdot \log(1 + s \cdot \delta_i), \tag{37}$$

where δ_i and δ_o are the input and output pixel disparities, s is a parameter that controls the scaling and K is chosen so that the output pixel disparities fit inside the allowed range. We include results for $s = 0.5$ and $s = 5$. Finally, for the perceptually linear method, disparity values are mapped via transducers into a perceptually linear space, and then linearly scaled by a factor k. The choice of k implies a trade-off between the improvement in contrast enhancement and how faithful to the original disparities we want to remain. We choose $k = 0.75$ and $k = 0.95$ as good representative values for both options respectively.

COMPARISONS: Some of the methods we compare against (*linear*, *Cornsweet* and *log*) require to explicitly define a minimum spatial cut-off frequency, which will in turn fix a certain target depth range. We run comparisons on different data sets and for a varied range of cut-off frequencies: For the *birds* scene, where the viewing distance is $v_D = 1.5$ m, we test two cut-off frequencies: $f_{cpmm} = 0.12$ cycles per mm ($f_{cpd} = 3.14$ cycles per degree), and $f_{cpmm} = 0.19$ ($f_{cpd} = 5.03$), the latter of which corresponds to remapping

3 http://hdrvdp.sourceforge.net/reports/2.1/quality_live/ http://hdrvdp.sourceforge.net/reports/2.1/quality_tid2008/

4 In our tests, this consistently yielded better results than a naive application of unbounded Cornsweet profiles, as originally reported by Didyk and colleagues [107]

to the depth range which offers the maximum spatial resolution of the display (see DOF plots in Figure 5.16b). For the *statue*, *papillon* and *buddha2* scenes, optimized for a multilayer display with $v_D = 0.5$ m, we set the frequencies to f_{cpmm} = 0.4, 0.5 and 1.1, respectively (corresponding f_{cpd} = 3.49, 4.36 and 9.60). The frequencies are chosen so that they yield a fair compromise between image quality and perceived depth, given the trade-off between these magnitudes; they vary across scenes due to the different spatial frequencies of the image content in the different data sets.

Figure 5.15 shows a comparison to the results obtained with the other methods both in terms of image quality and of perceived depth for three different scenes from the Heidelberg data set (*papillon*, *buddha2*, and *statue*). Heat maps depict the error in perceived depth (in JNDs) given by Didyk et al.'s metric. Visual inspection shows that our method consistently leads to less error in perceived depth (white areas mean error below the 1 JND threshold). Close-ups correspond to zoomed-in regions from the resulting images obtained with each of the methods, where the amount of DOF blur can be observed (the complete images can be found in `http://webdiis.unizar.es/~bmasia/downloads/thesis/Displays-DisparityRemapping.zip`). Our method systematically yields sharper images, even if it also preserves depth perception better. Only in one case, in the *statue* scene, perceptually linear remapping yields sharper results, but at the cost of a significantly higher error in depth perception, as the corresponding heat maps show.

To better assess the quality of the deblurring of the retargeted images, Figure 5.16a shows the MS-SSIM metric for the different methods averaged over the scenes tested, together with the associated standard error (we plot the absolute value of $\log(1 - \text{MS-SSIM})$). We have added the result of the original image, without any retargeting method applied (N for *none* in the chart). Our method yields the best perceived image quality (highest MS-SSIM value), and as shown in Figure5.15, the lowest error in depth perception as well. This can be intuitively explained by the fact that our proposed multi-objective optimization (Eq. 34) explicitly optimizes *both* luminance and depth, whereas existing algorithms are either heuristic or take into account only one of the two aspects.

To further explore this image quality vs. depth perception trade-off, we have run the comparisons for the *birds* scene for two different cut-off spatial frequencies. Figure 5.16b shows comparisons of all tested algorithms for the *birds* scene retargeted for a lenslet-based display. For two of the methods, ours and the perceptually linear remapping (with $k = 0.75$ and $k = 0.95$), defining this minimum spatial frequency is not necessary. Error in depth for these is shown in the top row. For the other four methods (*linear*, *Cornsweet*, *log* $s = 0.5$, *log* $s = 5$), the cut-off frequency needs to be explicitly defined: we set it to two different values of $f_{cpmm} = 0.12$ and $f_{cpmm} = 0.19$, which correspond to an intermediate value and to remapping the content to the maximum spatial frequency of the display, respectively. The resulting error in depth is shown in the middle and bottom rows of Figure 5.16b. Error in perceived depth clearly increases as the cut-off frequency is increased. The bar graph at the top left of Figure 5.16b shows image quality results for $f_{cpmm} = 0.12$. Note that for $f_{cpmm} = 0.19$, the methods *linear*, *Cornsweet* and *log* yield perfectly sharp images (since we explicitly chose that frequency to remap to the maximum resolution of the display), but at the cost of large errors in perceived depth.

5.9 CONCLUSIONS AND FUTURE WORK

Automultiscopic displays are an emerging technology with form factors ranging from hand-held devices to movie theater screens. Commercially successful implementations, however, face major technological challenges, including limited depth of field, resolution, and contrast. We argue that compelling multiview content will soon be widely available and tackle a crucial part of the multiview production pipeline: display-adaptive 3D content retargeting. Our computational depth retargeting algorithm extends the capabilities

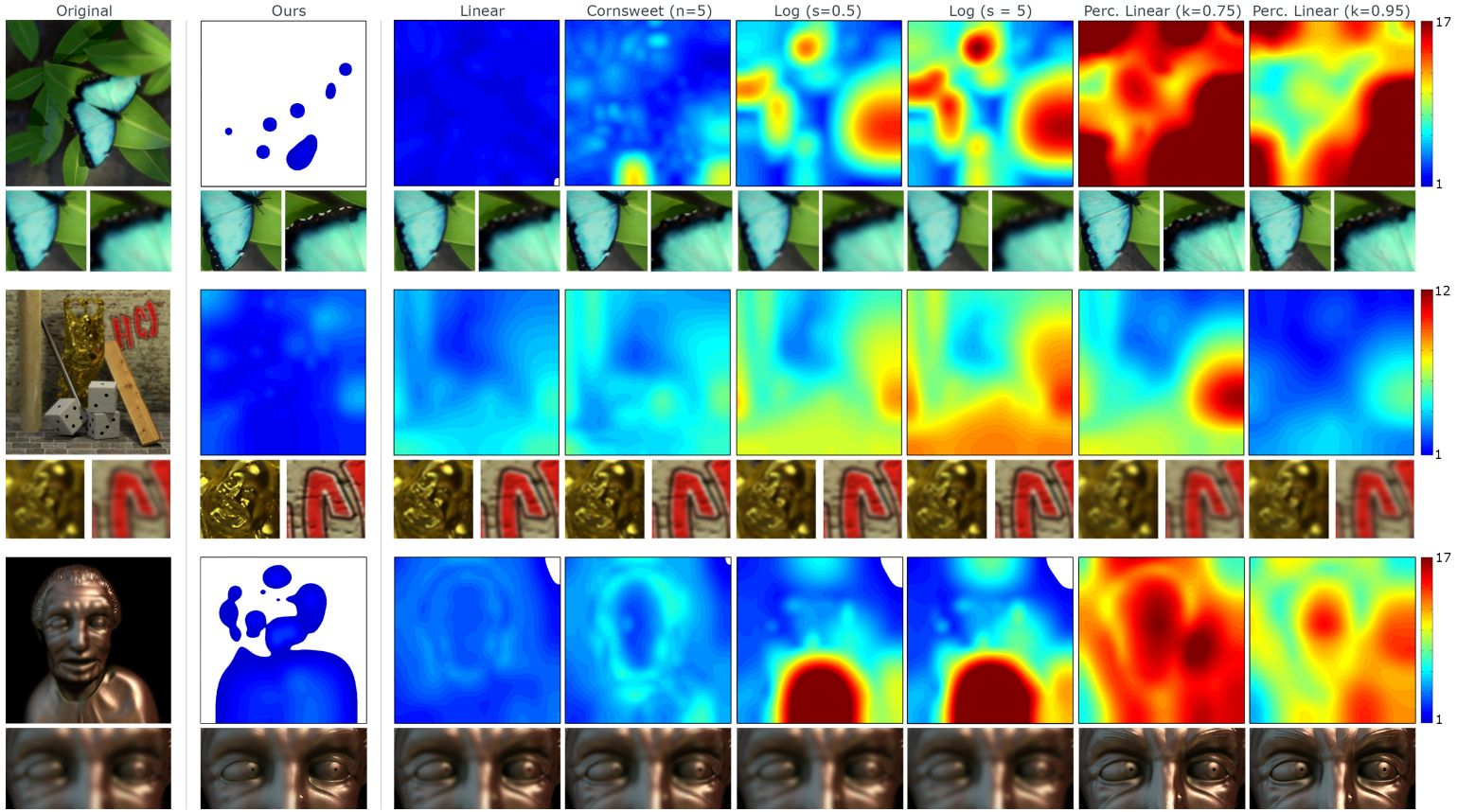


Figure 5.15: Comparison against other methods for three different scenes from the Heidelberg light field archive. From top to bottom: *papillon* ($f_{cpmm} = 0.4$, $f_{cpd} = 3.49$), *buddha2* ($f_{cpmm} = 1.1$, $f_{cpd} = 9.60$), and *statue* ($f_{cpmm} = 0.5$, $f_{cpd} = 4.36$). Errors in depth are shown as heat maps (lower is better) according to the metric by Didyk and colleagues [106]; white areas correspond to differences below one JND. Viewing distance is 0.5 m.

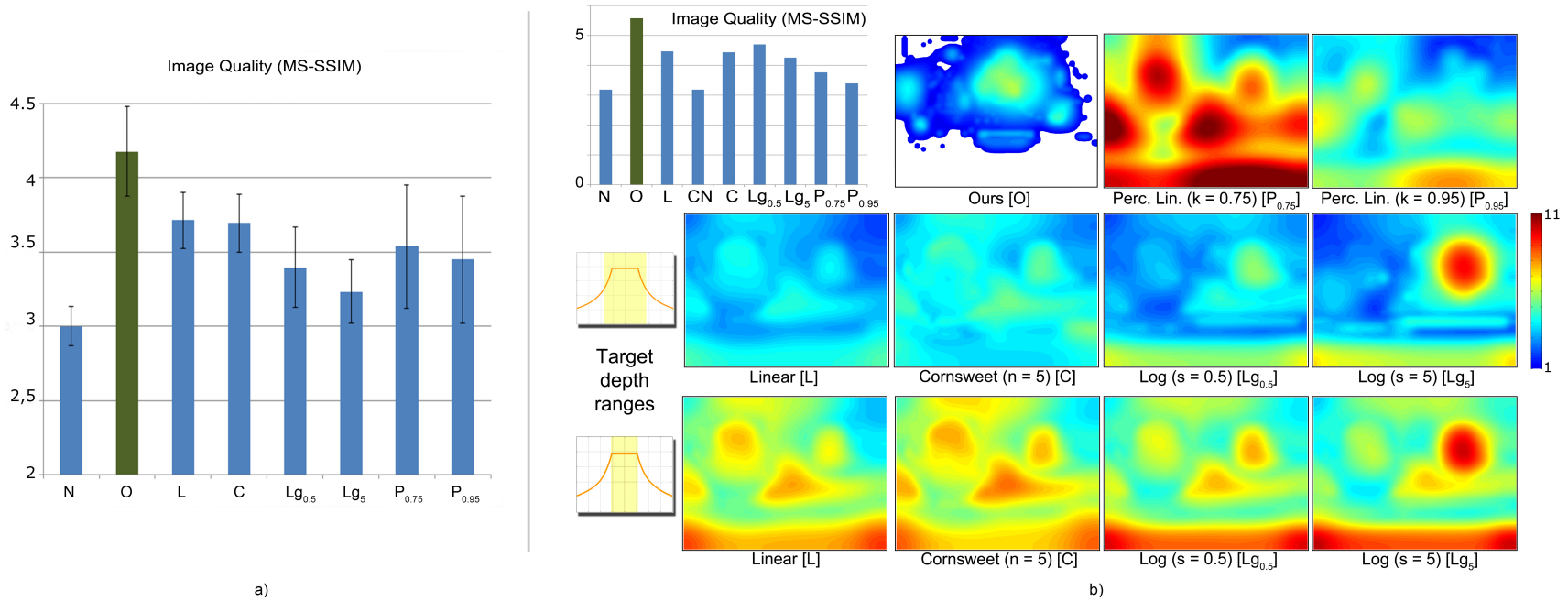


Figure 5.16: (a) Comparison of average luminance quality (lack of blur) according to the MS-SSIM metric for all the data sets used in this comparisons (higher is better). (b) Comparison against other methods for the *birds* scene, for two different cut-off frequencies. Top row, from left to right: resulting image quality as predicted by MS-SSIM for $f_{cpmm} = 0.12$, and error in depth for the two methods that do not require providing a target depth range. Middle row: error in depth for the three methods requiring a target depth range, for a cut-off frequency $f_{cpmm} = 0.12$ ($f_{cpd} = 3.14$). The smaller image represents the depth vs. cut-off frequency function of the display, with the target depth range highlighted in yellow. Bottom row: same as middle row for a cut-off frequency $f_{cpmm} = 0.19$ ($f_{cpd} = 5.03$), corresponding to the maximum spatial frequency allowed by the display (flat region of the DOF function). Errors in depth are shown as heat maps (lower is better) according to Didyk et al's metric [106]; white areas correspond to differences below one JND. Note the intrinsic trade-off between image quality and depth perception for the methods requiring a specific target depth range: when remapping to the maximum spatial frequency of the display, error in perceived depth significantly increases. Viewing distance is 1.5 m.

of existing glasses-free 3D displays, and deals with a part of the content production pipeline that will become commonplace in the future.

As shown in this work, there is an inherent trade-off in automultiscopic displays between depth budget and displayed spatial frequencies (blur): depth *has to* be altered if spatial frequencies in luminance are to be recovered. This is not a limitation of our algorithm, but of the targeted hardware (Figure 3). Our algorithm aims at finding the best possible trade-off, so that the inevitable depth distortions introduced to improve image quality have a minimal perceptual impact. Therefore, the amount of blur (the cut-off frequency) in the retargeted scene depends on the actual visibility of the blur in a particular area, according to the CSF. Should the user need to further control the amount of defocus deblurring, it could be added to the optimization in the form of constraints over the depth values according to the corresponding DOF function.

We have demonstrated significant improvements in sharpness and contrast of displayed images without compromising the perceived three-dimensional appearance of the scene, as our results and validation with objective metrics show. For the special case of disparity retargeting in stereoscopic image pairs, our method is the first to handle display-specific non-dichotomous zones of comfort: these model the underlying physical and physiological aspects of perception better than binary zones used in previous work. The video at `http://webdiis.unizar.es/~bmasia/downloads/thesis/DisplayAdaptive3DContentRemapping_CAG.mov` shows an animated sequence for retargeted content. It is shown as an anaglyph, so it can be seen in 3D on a regular display. Although the frames of this video clip have been processed separately, our algorithm provides temporally stable retargeting results.

A complete model of depth perception remains an open problem. One of the main challenges is the large number of cues that our brain uses when processing visual information, along with their complex interactions [88, 169]. A possible avenue of future work would be to extend the proposed optimization framework by including perceptual terms modeling human sensitivity to accommodation, temporal changes in displayed images, sensitivity of depth perception due to motion parallax or the interplay between different perceptual cues. However, this is not trivial and will require significant advances in related fields. Another interesting avenue of future work would be to extend our optimization framework to deal with all the views in the light field, as opposed to working on the central view, thus exploiting angular resolution and avoiding artifacts for large parallax.

We hope that our work will provide a foundation for the emerging multiview content production pipeline and inspire others to explore the close relationship between light field acquisition, processing, and display limitations in novel yet unforeseen ways.

6

VISUAL COMFORT IN STEREOSCOPIC MOTION

ABOUT THIS CHAPTER

This chapter compiles the work done in measuring the degree of comfort when viewing stereoscopic content in motion. The analysis and measurements are done in a multidimensional space including velocity in depth or disparity, among others. To our knowledge, this work is the most comprehensive study of the influence of stereo motion in viewing comfort up to date. The work has been done in collaboration with Song-Pei Du and Shi-Min Hu, from the *Graphics and Geometric Computing Group* in *Tsinghua University* (Beijing, China), and began as a result of our visit to their lab in November 2012. In this project, my contribution has been in the design of the experiment, methodology and analysis, and also in the design of the validation stage. The work has recently been accepted to SIGGRAPH Asia 2013.

S. Du, B. Masia, S. Hu and D. Gutierrez.
A Metric of Visual Comfort for Stereoscopic Motion.
ACM Transactions on Graphics, 32(6) (Proceedings of SIGGRAPH Asia 2013).

6.1 INTRODUCTION

Over the last few years, there has been a renewed interest in stereoscopic displays. Stereoscopic content is generated for movies, games and visualizations for industrial, medical, cultural or educational applications. This has in turn spurred research on aspects of the human visual system that relate to stereo vision [351]. Recent studies analyze the comfort zone for the vergence-accommodation conflict, the influence of luminance on stereo perception, or the depiction of glossy materials, to name just a few [401, 108, 438]. The goal is to understand different aspects of our visual system in order to produce stereo content that guarantees a comfortable viewing experience.

As opposed to natural viewing of the real 3D world, stereoscopic viewing implies conflicting vergence and accommodation cues, which is widely accepted to be a main cause of visual discomfort (see also Section 4.6.1 in Chapter 4). However, despite recent advances and the extensive existing literature [183, 211, 108, 106], some aspects of binocular vision remain largely unexplored. One of the main reasons is the large number of different factors involved, as well as their complex interaction [88]. As a consequence, generating stereo content that guarantees a comfortable viewing experience remains a challenging task, often reserved to technicians with a large experience in the field [252, 311].

Thus, one of the goals of stereography is to minimize the discomfort that stereoscopic viewing can cause, and numerous works have been devoted to explaining and characterizing the causes [236, 251, 401]. However, fewer have explored how object *motion* affects this discomfort in stereoscopic viewing. Object motion in stereoscopic movies can in fact be a source of discomfort: Researches and experiments have revealed that visual comfort has a close relationship with some oculomotor functions, including eye movements induced by motion in the scene [28, 336]. In this work we analyze visual discomfort due to motion in short stereoscopic movies by means of a comprehensive statistical study. Unlike previous work [499, 212], we take into account the interplay of motion velocity both on the screen plane and on the depth axis, as well as *signed* disparity and luminance spatial frequency. Our goal is not only to help understand the phenomena that

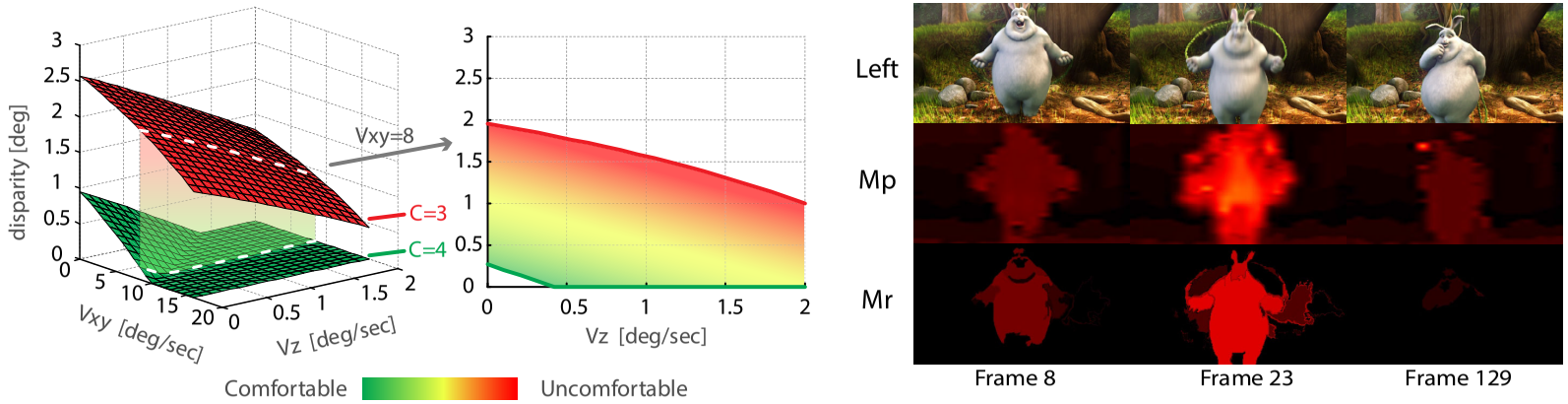


Figure 6.1: Our novel metric of visual comfort for stereoscopic content takes into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast. Based on our measured comfort function, we derive a metric to predict the degree of comfort for short stereoscopic videos. Left: Example slice of the comfort zone computed by our comfort function for a spatial frequency of 1cpd (bounded by comfort values of 3 and 4). Right: comfort maps computed using our metric on three representative frames of the *bunny* movie (© Blender Foundation). From top to bottom, input frames, per-pixel results, and per-region results (brighter red indicates less comfort). Our metric predicts less comfort with faster movement (frame 23), in agreement with the perceptual experiments.

may lead to visual discomfort; we provide a practical metric to assess existing 3D content as well. This can be used as a guideline for the generation of new stereo content, or to keep navigation parameters in virtual reality environments within comfortable limits, for instance.

CONTRIBUTIONS: Specifically, we make the following contributions:

- We show that all the factors included in our study, as well as their interaction, do affect viewing comfort, and should be considered in the design of stereo content.
- We derive a statistical measurement that models the influence of motion, luminance spatial frequency, and signed disparity in visual discomfort.
- We propose a metric to predict potential comfort in short stereoscopic videos, and validate it by means of a user study.
- We propose several direct applications that could benefit from our measurements and metric, including a novel visual comfort zone for stereoscopic production, visualization techniques and stereoscopic retargeting.

LIMITATIONS: Although our measurements and metric are the most complete and exhaustive up to date, we do not aim at providing here the ultimate solution to this problem. Our methodology and results represent a solid step towards fully characterizing discomfort due to motion, and we hope that they will help others build more sophisticated models. However, there are a number of limitations that should be addressed by follow-up work. First, a comfortable stereo viewing experience may be related to many other factors not considered here, such as luminance contrast, disparity spatial frequency, viewing time, flicker, or imperfect content, to name a few [236, 294, 106, 176, 79]. Taking all these factors and their interactions into account would make the problem intractable. Additionally, we limit our study to supra-threshold stimuli. Last, our metric is devised for short video sequences (up to 30 seconds in our results). This is convenient, since the average shot in modern TV and movies is only a few seconds. It would be interesting, however, to analyze how to extend our approach to longer sequences (even entire films) to take into account cumulative discomfort effects.

6.2 RELATED WORK

Many previous works have investigated various aspects of stereoscopic perception (see for instance [183, 337], and Section 4.6.1 in Chapter 4 for a brief overview). Recently, researchers have begun to explore the problem from the perspective of computer graphics and its applications. For instance, Templin et al. [438] introduce a novel technique for stereoscopic depiction of glossy materials. Closer to our approach, Didyk et al. [106] propose a model of disparity based on perceptual experiments, which is later extended to take into account the influence of luminance contrast [108]. We also propose a measurement based on perception-driven studies, although we tackle a different problem, focusing on visual discomfort in the presence of stereoscopic motion.

Existing works have shown that visual discomfort in stereoscopy has a close relationship with oculomotor functions. It is widely accepted that the vergence-accommodation conflict is a key factor of visual discomfort, and that there exists a comfortable zone within which little discomfort occurs [336, 177, 435]. In general, eye movement can be a source of discomfort when viewing stereoscopic content [28], which means that the motion component needs to be explicitly considered when measuring visual discomfort.

Kooi and Toet [236] investigate various factors that may affect the visual comfort of viewing stereo images, including optical errors, imperfect filters and disparity. Hoffman et al. [176] investigate the influence on the stereo viewing experience caused by flicker,

motion and depth artifacts for various temporal presentation methods. Other works offer a quantitative measurement of visual comfort: Jin et al. [204] evaluate the stereoscopic fusion disparity range based on the viewing distance and field of view of the display. Vertical misalignment has also been shown to affect visual comfort, and the maximum tolerable vertical misalignment has been measured as a unified metric based on different kinds of geometric misalignment [205]. Shibata et al. [401] design a series of experiments to evaluate the zone of comfort for different vergence-accommodation combinations, while Yang et al. [496] introduce a binocular viewing comfort predictor. None of them, however, consider motion. Lambooij et al. [251] present a review of causes of visual discomfort, and conclude that visual discomfort might still occur within the so-called comfortable zone because of fast motion, insufficient depth information and unnatural blur.

Existing experiments have also confirmed the correlation between the velocity of moving objects and visual comfort [498, 499, 448]. Speranza et al. [419] investigate the relationship between visual discomfort and object size, motion-in-depth and disparity. Jung et al. [212] introduce a novel visual comfort metric for stereoscopic video based on salient object motion, by computing three different discomfort functions for motion in horizontal, vertical and depth, respectively. The authors then use the mean or min operations to assess the global visual comfort, which is an ad-hoc solution for the co-occurrence of different motion components. Cho and Kang [79] measure the visual discomfort as a function of disparity and viewing time for three levels of motion-in-depth (slow, medium and fast). Last, Li et al. employ pair-comparison experiments and propose a visual discomfort model based on the disparity and motion on the screen plane; they use both experts-only subjects using the Thurstone-Mosteller model [263], and non-experts subjects using the Bradley-Terry model [262].

All previous works on visual discomfort of stereoscopic motion either consider a single component of the motion vector, or simply combine conclusions obtained through separate experiments. In contrast, we offer a comprehensive study and systematically explore a larger parameter space, including the influence of the luminance spatial frequency, which is known to play an important role in depth perception. From our studies, we build a reliable measurement of visual comfort for stereoscopic motion, which we use to derive a predictive metric.

6.3 METHODOLOGY

In this section we describe the subjective experiments performed to measure the subjects' level of comfort when watching stereoscopic motion.

6.3.1 *Parameter Space*

As explained in Section 6.2, two key factors related to visual comfort in stereoscopic images and videos are the disparity value d and the velocity of motion $\mathbf{v} = (v_x, v_y, v_z)$, where subindices x, y refer to the screen plane and z indicates the direction perpendicular to the screen, i.e. depth. Recent studies have found that v_x and v_y have a similar effect on visual comfort [212]; we thus reduce the dimensionality of our problem by focusing on planar motion (v_{xy}) plus motion in depth (v_z). We measure d in terms of angular disparity α (in deg) while v_z and v_{xy} are the derivatives of the angular disparity and viewing direction β, respectively (in deg/sec). These parameters are shown in Figure 6.2. Intuitively, v_{xy} corresponds to a change in the gaze direction, while v_z corresponds to a variation of eye vergence.

The influence of luminance spatial frequency f_l on disparity perception is well known [258, 168], and it was recently used to develop a perceptual disparity model [108]. However, its influence on visual comfort for stereo motion remains largely unexplored; to

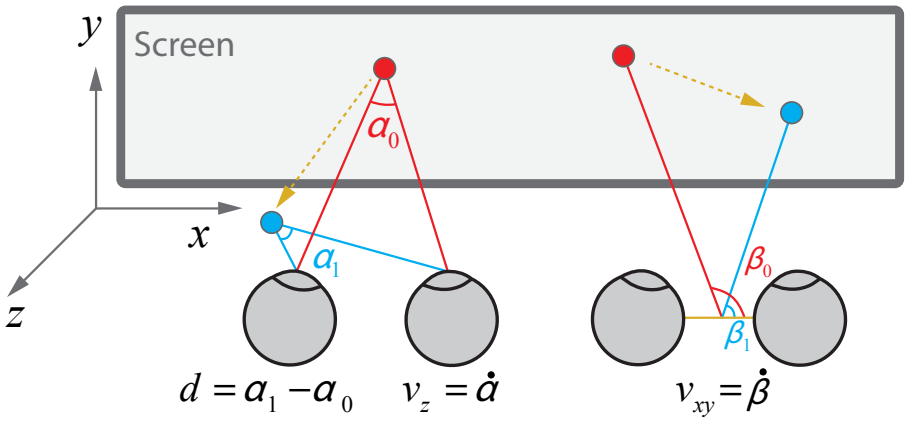


Figure 6.2: Disparity and motion defined as functions of angular disparity α and viewing direction β. The viewing direction is defined from the middle interocular point.

overcome this, we add a parameter for luminance spatial frequency f_l in our experiments. Our parameter space is then four-dimensional: d, v_{xy}, v_z, f_l. Similar to other works, in order to reduce dimensionality, we fix the values of other parameters in our experiments.

6.3.2 *Stimuli*

Each stimulus consists of two-second animations, defined by a sinusoidal depth corrugation textured with a luminance image of noise of spatial frequency f_l. Each corrugation moves at a speed (v_{xy}, v_z). The mean disparity of the corrugation is d, and the amplitude of the sinusoid is fixed for all stimuli at $0.1°$ (6 arcmin), defined as the difference between mean and peak. In the case in which $v_z \neq 0$, and thus the mean disparity of the corrugation changes over time, d is defined as the mean disparity of the whole two-second stimulus. Previous work by Shibata et al. [401] used 4 arcmin, but did not consider motion nor the influence of luminance spatial frequencies. We thus choose a slightly larger value which allows to clearly distinguish the corrugation. The corrugation's disparity spatial frequency is set to 0.3 cpd, which has been reported to be near the peak sensitivity of the human visual system [106]. We sample each dimension of our parameter space as follows:

- $d = \{-2, 0, 2\}$ [°]
- $v_{xy} = \{0, 8, 16\}$ [°/sec]
- $v_z = \{0, 1, 2\}$ [°/sec]
- $f_l = \{1, 4, 16\}$ [cpd]

This makes a total of 81 different stimuli. Additionally, for each stimulus we explore four different corrugation orientations $\psi = \{0, 45, 90, 135\}$ [°], defined as degrees over the horizontal (see Section 6.3.3). The stimuli are shown on a fixed window at the center of the screen (the viewing angle of the window's diagonal is 22°), surrounded by a 50% gray background. To obtain the stereo pair, we use image warping for the left and right views [106]. This is done to avoid the "keystone" distortion in the toed-in camera configuration, and the perspective effect which would make the depth corrugations look non-uniform. We pre-compute the stimuli by warping offline, which works well in practice; no artifacts were reported by the subjects.

We explore v_{xy} by moving the depth corrugation along the 45° diagonal. To avoid potential discomfort from nonlinear motion gradients [212], we only consider positive values of v_z, that is, motion towards the subject. Additionally, as noted by Speranza et al. [419], zero-crossings in the disparity signal will affect the visual comfort: We thus limit depth corrugation motion in the z axis from $d - |v_z|$ to $d + |v_z|$ during the two-second span. Example stimuli are shown in Figure 6.3 while different combinations of d and v_z are illustrated in Figure 6.4.

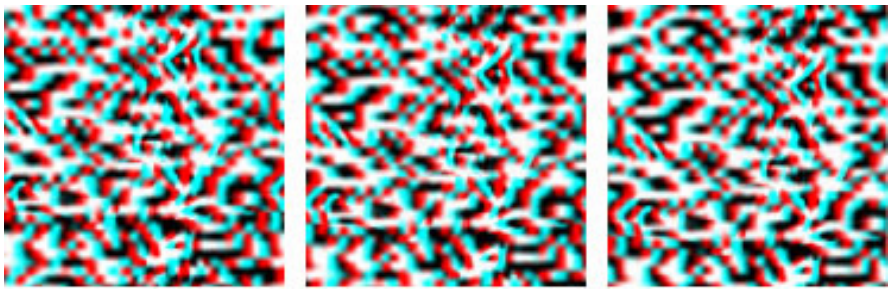


Figure 6.3: Sample stimuli for the case of corrugation orientations $\psi = 0°$. Three successive frames are shown as anaglyphs. Under our viewing configuration, a 0.1° amplitude corresponds to about 7mm in depth.

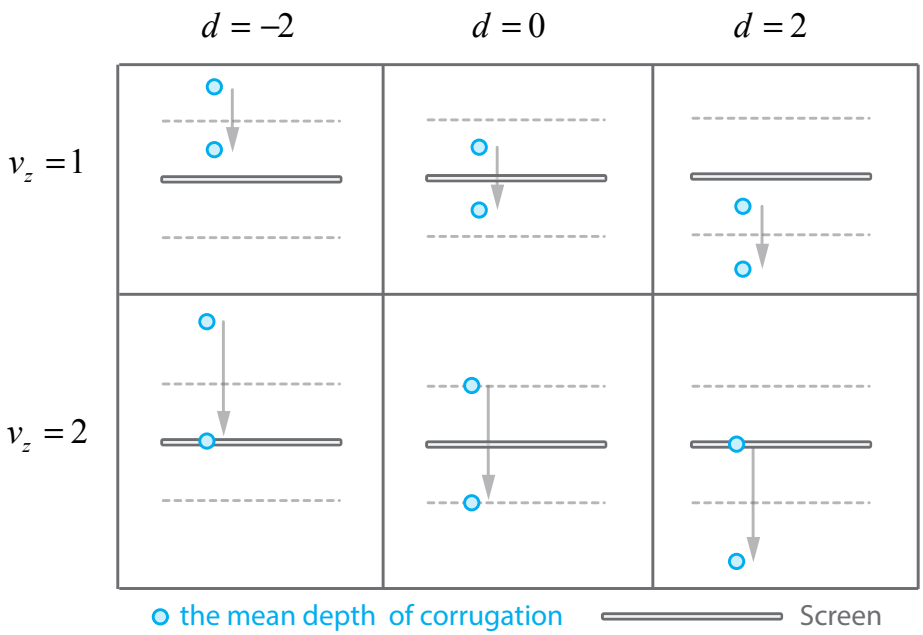


Figure 6.4: Different combinations of d and v_z. There are no non-linear motion components and no sign conversions of d at non-zero disparity. In each case, the upper and lower dashed lines represent $d = -2°$ and $d = 2°$ respectively.

6.3.3 *Procedure*

We use a 23-inch interleaved 3D display (1920 × 1080 pixels, 400 cd/m^2 brightness) with passive polarized glasses. The viewing distance is 50 cm and we assume the interpupillary distance to be 65 mm. Twenty subjects participated in our experiments, all with normal or corrected-to-normal vision, and with no difficulty in stereoscopic fusion. Their

ages range from 20 to 30 years. Horizontal stripes are visible at this viewing distance and this should be universal for all such polarization displays. Subjects were aware of this and did not report complaints in the experiments.

The experiment is divided into 81 sub-sessions, which correspond to all possible combinations of our stimuli with fixed corrugation orientation ψ. In each sub-session, the subject is asked to rate their comfort level after doing a series of *visual oddity* tasks (three-interval, forced choice) [401]. Specifically, for a given sample (d, v_{xy}, v_z, f_l), three stimuli are presented sequentially, with a 0.5 second break between stimuli during which a 50% gray image is shown. Two of the three stimuli have the same corrugation orientation ψ, while the other has a 45° difference. For example, the three sequentially presented orientations may be $(0°, 0°, 45°)$, $(90°, 135°, 90°)$, $(45°, 0°, 45°)$, etc. The order and choice of orientations are random. After presenting the three stimuli ($2 \times 3 + 0.5 \times 2 = 7$ seconds), the subject is forced to select which stimulus had the odd orientation. Each sub-session contains ten such oddity tasks; after completing each sub-session, the subject is asked to rate their comfort level on a 5-point Likert scale, based on the following two questions:

- How do your eyes feel? (From 1 to 5: severe strain, moderate strain, mild strain, normal, very fresh).
- How comfortable was the viewing experience? (From 1 to 5: very uncomfortable, uncomfortable, mildly comfortable, comfortable, very comfortable).

Each sub-session takes about 80 seconds. We split the experiment into three parts and each part contains 27 sub-sessions with the same f_l value (1, 4, or 16 cpd); the three parts are done on three consecutive days. In each part the order of the 27 sub-sessions is random across subjects. To avoid accumulation effects [79], subjects have to take a two-minute rest between sub-sessions, plus a longer, ten-minute break after 13 sub-sessions. They are nevertheless encouraged to take a longer rest if they want. For one subject, the whole experiment takes between 1.5 and 2 hours each day. Our choices are based on pilot tests performed before the regular experiments, which show that after an 80-second sub-session, subjects do perceive discomfort and that they can recover well after a two-minute break between sub-sessions.

6.4 ANALYSIS

We first compute each subject's comfort score for every sub-session by averaging the two scores from the Likert scale, thus obtaining a total of $20 \times 81 = 1620$ scores. Then the comfort score for each session is computed as the average across the 20 subjects. To verify that these averages are statistically reliable, we perform a one-way repeated measures analysis of variance (ANOVA) of our data. This yields an F-value $F(80, 1520) = 24.01$, which is much larger that the F-test critical value for $p = 0.01$. This means the inter stimuli (intra subjects) variances are much larger than intra stimuli (inter subjects) variances, and thus our average scores are statistically reliable. Average ratings per stimuli, together with the standard error of the mean, can be found in Appendix G.

Our data are then used to determine our statistical measurement C of visual comfort for stereoscopic motion. We measure C as $C = C_{\mathbf{v},d} + C_{f_l}$, that is, a function of both the *combination* of velocity and depth, and luminance frequency. Previous works analyzing only velocity and disparity separately suggest that comfort seems to be approximately linear with those parameters [263, 212]. Although Jung et al. [212] fitted their model using logarithmic functions, the non-linear components are relatively small. Thus, we begin by using the following polynomial to fit $C_{\mathbf{v},d}$:

$$C_{\mathbf{v},d} = p_1 v_{xy} + p_2 v_z + p_3 v_{xy} v_z + p_4 d v_{xy} + p_5 d v_z + p_6 d + p_7 \tag{38}$$

Didyk et al. [108] model a discrimination-threshold function s as: $s \approx 0.257 \log^2(f_l) - 0.3325 \log(f_l) + s(f_d, m_d)$ where f_d and m_d represent the frequency and magnitude of

disparity respectively. To measure the influence of luminance spatial frequency f_l in visual discomfort, we therefore define C_{f_l} as a quadratic component of C:

$$C_{f_l} = p_8 \log^2(f_l) + p_9 \log(f_l) \tag{39}$$

Additionally, we want to explore how the *sign* of d affects the comfort score. We expand Equation 38 and include (p_4^+, p_5^+, p_6^+) and (p_4^-, p_5^-, p_6^-) for $d \geqslant 0$ and $d < 0$, respectively. The resulting eleven-dimensional coefficient vector $\mathbf{P}$=[p_1, p_2, p_3, p_4^+, p_5^+, p_6^+, p_4^-, p_5^-, p_6^-, p_7, p_8, p_9] is computed by solving the following quadratic optimization using linear least squares:

$$\arg\min_{\mathbf{P}\in\mathbb{R}^{10}} \sum_{i=0}^{80} (C(\mathbf{x}_i) - C_i)^2 \tag{40}$$

where $\mathbf{x}_i|_{i=0}^{80}$ are the 81 samples described in Section 6.3 and $C_i|_{i=0}^{80}$ corresponds to the 81 average scores across subjects. This yields a vector $\mathbf{P}$ = [−0.0556, −0.6042, 0.0191, 0.0022, 0.1833, −0.6932, −0.0043, −0.1001, 0.2303, 4.6567, 0.9925, −1.1599]. The R^2 measure of goodness of fit is $R^2 = 0.9306$.

6.4.1 *Discussion*

Several slices of our visual comfort function C are visualized in Figures 6.5 and 6.6. Our measurement agrees with previous observations from existing works: Increasing disparity values (Figure 6.6(a)), motion on the screen plane (Figure 6.6(b)) or motion in depth (Figure 6.6(c)) introduce larger discomfort. Additionally, our measurement allows us to infer other important conclusions:

- The *sign* of the disparity also affects visual comfort (see Figure 6.6(a)). This effect was previously reported for the case of static stimuli [401]; our experiments show that this behavior applies to stereoscopic motions as well. Additionally, we provide a quantitative measurement of this difference ($|p_6^+| = 0.6932$ and $|p_6^-| = 0.2303$).
- The *combination* of different values of v_{xy} and v_z has a strong influence in comfort ($p_3 = 0.0191$), as shown in Figure 6.6(b). Comfort decreases differently as v_{xy} *and* v_z increase ($p_1 = -0.0556$, $p_2 = -0.6042$). In particular, the influence of v_{xy} in viewing discomfort diminishes as v_z increases.
- Last, luminance spatial frequency f_l is a non-linear factor in viewing comfort. For $f_l \in$ [1 cpd, 16 cpd] in our experiments, the comfort score has a minimum near 4 cpd ($p_8 = 0.9925$, $p_9 = -1.1599$), as shown in Figure 6.6(d). Didyk et al. [108] computed the influence of f_l on perceived depth, and found a similar minimum. This could mean that, when taking motion into account, smaller perceived depths may produce a more comfortable 3D viewing experience.

6.5 METRIC OF VISUAL COMFORT

Our measurement can be used to predict the level of comfort when viewing short stereoscopic videos. In particular, we derive a metric to compute both a pixel-wise comfort map of each frame in the video $M_p(i,j)$, which allows to identify the particular areas or objects in each frame that are potential sources of discomfort, and a global comfort score M_g for the whole video.

Given an input stereoscopic video consisting of two corresponding left-right image sequences $(I_L(t), I_R(t))|_{t=0,1,\ldots}$, we first compute the motion $\mathbf{v} = (v_x, v_y, v_z)$ and disparity value d at each pixel (i,j) in each frame $I_L(t)$. Since the binocular views have similar content, we assume they share the same comfort map and we will use the left view for

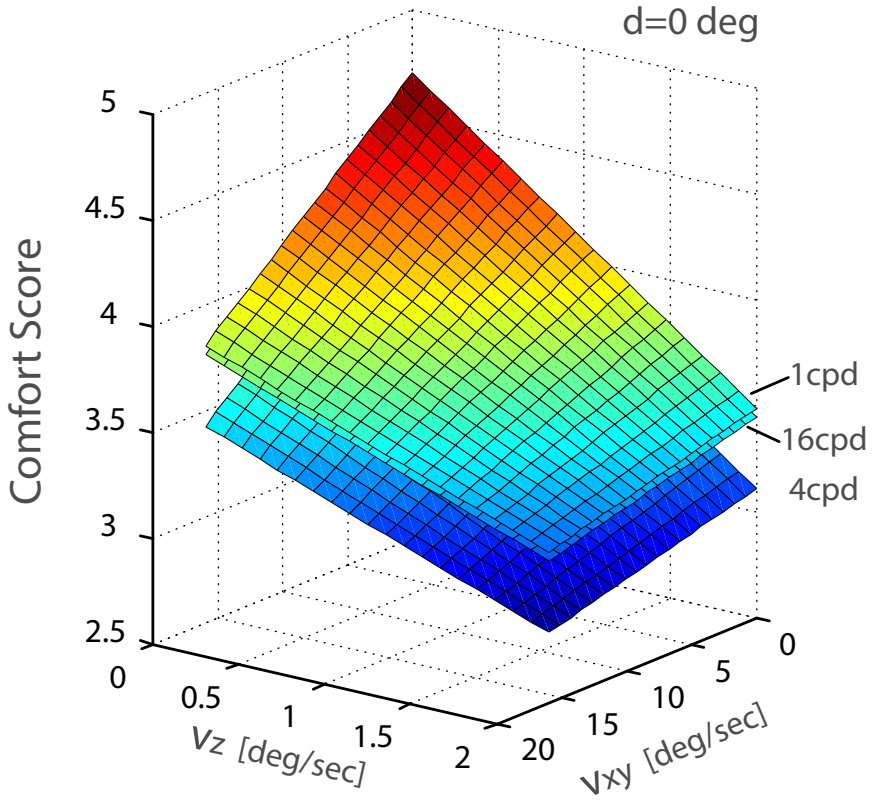


Figure 6.5: Slices of our measurement function C. For $d = 0°$, three slices are shown corresponding to three different luminance frequencies: 1, 4 and 16 cpd. Higher comfort score refers to better visual comfort level predicted by our measurement. We provide the slices for $d = -2°$ and $2°$ in Appendix G.

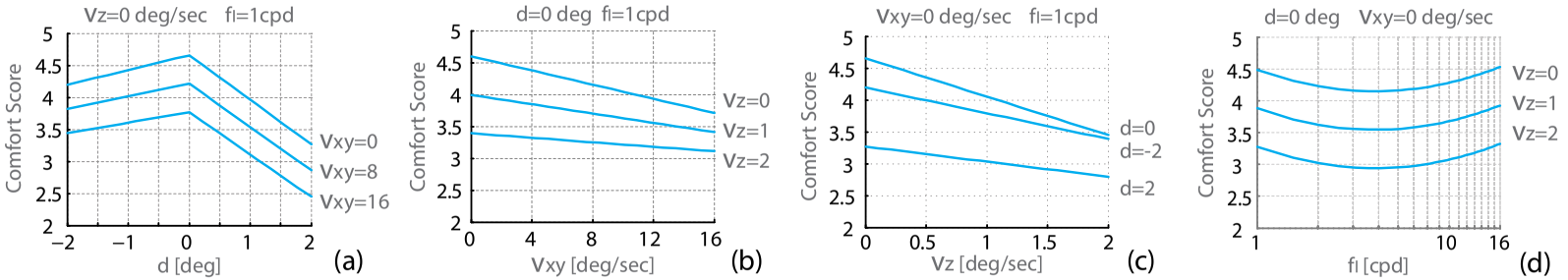


Figure 6.6: Different projections of our measurement function C. We refer the reader to Appendix G for more projections of the function.

computing velocities and f_l, and both views when computing the disparities. For real-world videos where this information is not usually available, we rely on optical flow (for motion), or motion and depth estimation [212]. We define the motion on the screen plane as $v_{xy} = \sqrt{v_x^2 + v_y^2}$.

PIXEL-WISE METRIC Our pixel-wise metric leverages our measurement of comfort C presented in the previous section. To take into account luminance spatial frequency, we construct a Laplacian pyramid of the luminance of $I_L(t)$, from a starting base frequency f_{l_0}. Multi-scale decompositions are frequently used to model the varying sensitivity of the human visual system (HVS) to different spatial frequencies. In the case of luminance-contrast frequencies, Laplacian pyramids are an efficient approximation that works well in practice [292, 108]. Similarly, our local contrast term is an efficient and convenient approximation to assess the influence of each luminance spatial frequency channel on comfort. For each frame, we then compute the comfort score at each (i,j) using our measurement function C as:

$$M_p(i,j) = \sum_{k=0}^{n} C(v_{xy}, v_z, d, \frac{f_{l_0}}{2^k}) \times \frac{L_k(i,j)}{\sum_k L_k(i,j)} \tag{41}$$

where $L_k(i,j)$ is defined as the contrast value of the $(2^{k+1}+1)$- neighborhood at (i,j) at the k-th Laplacian level, and n is the number of Laplacian levels. In practice we select n such that $f_{l_0}/2^n < 1$ cpd. From the resulting $M_p(i,j)$ we obtain a two-dimensional comfort map per frame, which can be used to visualize the spatial location and distribution of the uncomfortable viewing regions. By stacking maps over time, we obtain a three-dimensional $M_p(i,j,t)$ map for the whole video, which allows to visualize the temporal evolution of the discomfort regions. Figure 6.7 shows M_p for two time instants of two different video sequences.

GLOBAL METRIC To compute a global comfort score M_G for the whole video we pool partial metrics both in the spatial and temporal domains. For the spatial, per-frame pooling, existing research suggests that the overall perception of any single frame is dominated by its "worst" area [215]. We thus take a conservative approach and assume that the most uncomfortable region in a frame dictates the discomfort of the whole frame. We further modulate such per-frame discomfort by taking saliency into account. Saliency maps have been employed before in related scenarios, like comfort assessment [212], editing of stereo content [271], or image and video retargeting [382]. We use saliency maps under the reasonable hypothesis that human subjects pay more attention to visually salient regions, and therefore those will have a greater influence in comfort. In our implementation, we obtain a saliency-based segmentation using the method by Cheng et al. [76], which also yields a per-region saliency value between zero and one. We reduce the discomfort in non-salient regions based on the saliency value, and the comfort M_g for frame t_k is then given by:

$$M_g(t_k) = \min_r(5 - S_r(t_k) \cdot (5 - M_r(t_k))) \tag{42}$$

where S_r represents the saliency values of a given region r and $M_r = \frac{1}{|r|}\sum_{(i,j)\in r} M_p(i,j)$ is the average per-pixel comfort in r.

For temporal pooling, various approaches have been proposed for video quality assessment [333, 40]. To obtain our final global metric M_G, we simply pool the results over the whole video by computing the median of the comfort scores over all frames, as validated in previous work [212].

Figure 6.1 (right) shows how the different stages of the metric perform for several sample frames of a stereo movie. For each frame (first row), first a per-pixel map $M_p(i,j)$ is computed (second row), which is then averaged to obtain a per-region comfort measure M_r (third row).

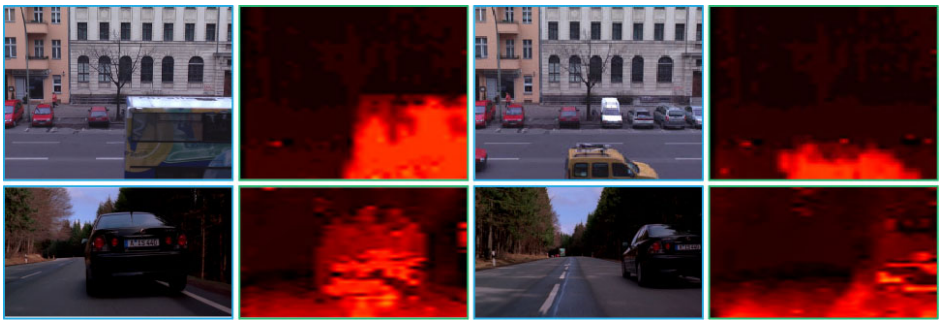


Figure 6.7: Representative frames with their computed pixel-wise comfort map M_p for *bus* (top row, © Fraunhofer HHI) and *car* (bottom row, © KUK Filmproduktion GmbH) scenes.

6.6 VALIDATION

Our metric is based on the comfort measure as derived in Section 6.5. In this section we conduct a set of experiments to validate our assumptions regarding the dependence on luminance contrast and the influence of saliency. Please refer to `http://webdiis.unizar.es/~bmasia/downloads/thesis/Displays-Comfort-ValidationVideos.zip` for videos of the stimuli employed.

CONTROLLED SCENES We first test our metric on four simple, controlled scenes, shown in Figure 6.8. They all consist of a static background ($v_{xy,BG} = 0$) and some moving objects in the foreground. The spatial frequency of luminance contrast f_l varies across stimuli, both for foreground and background. Textures and velocities also vary across stimuli, as detailed in Figure 6.8. We fix $v_z = 0$ in all cases, as well as the disparity d of both foreground and background to $d_{FG} = 0$ and $d_{BG} = -2$, respectively. We explore four combinations of different frequencies for both foreground and background; additionally, the screen plane velocity of the foreground can take two distinct values, $v_{xy,FG} = \{2, 16\}$.

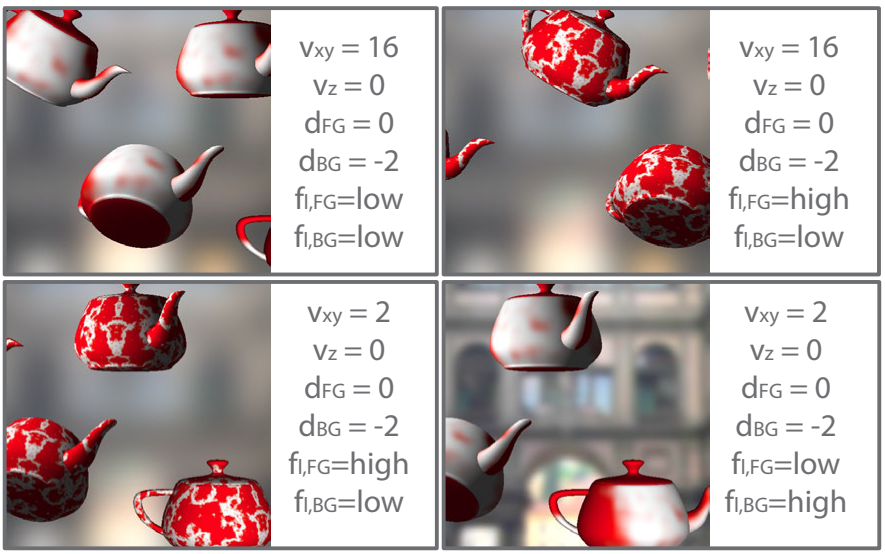


Figure 6.8: Representative frames of the stimuli used to test the validity of the approach used to incorporate the influence of luminance contrast spatial frequency. In reading order, stimulus A, B, C and D.

A user study is run with ten subjects, in which they are asked to rate their comfort when viewing the different clips. Each clip lasts 30 seconds, with a minimum resting period of 60 seconds between videos, and the order of presentation is randomized across subjects. We then compare the comfort values yielded by our metric with assessments given by subjects. Figure 6.11 (left) summarizes the score predicted by our metric and the average score given by the users. It can be seen how our metric follows very closely such user scores. Since our ranking data does not necessarily follow a normal distribution, we compute Spearman's correlation coefficient, which yields $\rho_S = 1$, with $p = 0.0833$. Pearson's linear correlation coefficient is $\rho_P = 0.9916$ with $p = 0.0084$, indicating that the variables are correlated. This test shows that, even in conflicting scenarios such as stimulus B (high frequency foreground with no disparity but high screen plane velocity) and D (high frequency background with high disparity but no screen plane velocity), our metric is able to capture the relative discomfort elicited by the stimuli.

Additionally, we want to test the influence of saliency on viewing comfort. A second experiment was conducted as a separate session with the same users (resting time between sessions was a minimum of 20 minutes), with a procedure analogous to the one described in the previous paragraph. In this case we use two stimuli, shown in Figure 6.9, consisting of a series of objects in the foreground, with disparities d_{FG}, on a constant static mid-gray background with disparity d_{BG}. The objects move erratically with slow screen plane velocity ($v_{xy} = 4$ and $v_z = 0$). In the second clip, we increase the saliency of one of the objects by making it clearly stand out in red. As Figure 6.11 (center) shows, making the object more salient made the user scores drop, since the discomfort caused by the moving (foreground) teapot now becomes more relevant; this behavior was also predicted by our metric.

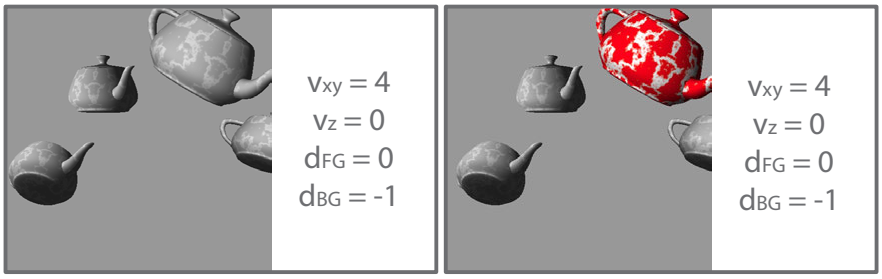


Figure 6.9: Representative frames of the stimuli used to test the validity of the saliency scheme integrated in the global metric.

We also compare our metric with Jung et al.'s [212]. Since they do not consider the influence of luminance, their metric will yield the same values for cases A and B (and C and D) in Figure 6.11 (left), and again for the two cases in Figure 6.11 (center). For the first experiment, Spearman's correlation coefficient is $\rho_S = 0.8944$ ($p = 0.3333$), while Pearson's coefficient is $\rho_P = 0.9045$ ($p = 0.0955$). Recall that our metric, in contrast, yields higher correlations with the measured data: $\rho_S = 1$ and $\rho_P = 0.9916$.

REAL SCENES We use four different scenes –*bus*, *bunny*, *horse* and *car*– to further validate our metric. They exhibit a variety of motion combinations as well as different luminance frequencies and disparity ranges. Figure 6.10 shows representative frames of each one. Ten subjects were asked to rate their comfort level after viewing the clip, and a two-minute rest is forced between two clips. Again, we compute the comfort score according to the presented metric and compare it against the score given by the subjects; results are shown in Figure 6.11 (right). Again, there is a strong correlation between predicted and user scores, with Spearman's rank correlation coefficient yielding a value of $\rho_S = 1$ ($p = 0.0833$), while Pearson's correlation coefficient is $\rho_P = 0.9514$ ($p = 0.0486$). Although our metric tends to slightly overestimate comfort, the predicted

value is generally inside the 95% confidence interval for the mean. The difference for the *car* scene is significantly larger, which is due to its complex v_z velocity field: Objects move in opposite directions, with nonlinear motion in depth and changes in the sign of v_z with respect to the camera. Given the high linear correlation in the data, we can compensate the overestimation by fitting a linear function to the metric, to obtain the final global expected score: $M_{exp} = 3.45M_G - 10.20$ ($R^2 = 0.91$).



Figure 6.10: Representative anaglyph frames of the stimuli used to test the validity of our metric in real scenarios. From left to right, top to bottom: bus, horse (© KUK Filmproduktion GmbH), bunny and car.

6.7 APPLICATIONS

Our work is a contribution towards a comfortable viewing experience. In this section, we describe various applications of our experiments and metric, including: Stereoscopic production, scientific visualization and retargeting.

STEREOSCOPIC PRODUCTION Various rules and guidelines have been proposed for practical use in stereoscopic content production in order to provide a comfortable viewing experience [311, 415, 414]. Often, these guidelines are based on years of experience in the production industry. Our quantitative measurements can complement that know-how. The challenge is the multidimensional nature of the measurements: the function presented in Figure 6.5 may be too elaborate for practical use in a production scenario. However, comfort zones can be derived from the measurements, which can in turn be used as a guide for content design. We can define the comfort zone $Z_{l,h}$ as:

$$Z_{l,h} = \{(d, v_{xy}, v_z, f_l) | l \leqslant C(d, v_{xy}, v_z, f_l) \leqslant h\}$$

where l and h are the given lower and upper bounds. Slices for the sample case of $l = 3$ and $h = 4$, $Z_{3,4}$, are shown in Figure 6.1 (left). Considering, for instance, a moving object with a given v_{xy}, the illustrated comfort zone defines a safe range of disparities on which such object can lie as a function of its velocity in z. Additionally, automatic computation of camera placement [335, 165, 209], could potentially benefit from incorporating information from our measurements.

STEREOSCOPIC RETARGETING The adaptation of stereoscopic content to a disparity range that provides a comfortable viewing experience has motivated a number of recent works that focus on disparity retargeting [252, 220, 307, 216].

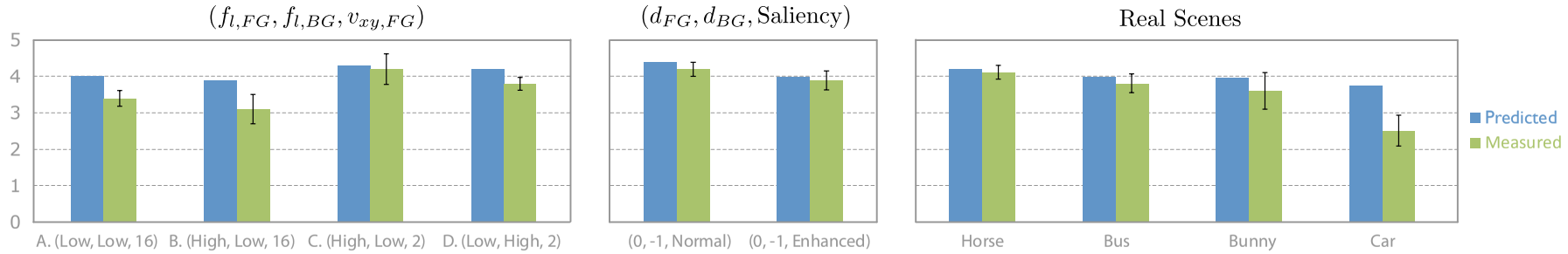


Figure 6.11: Results from the metric validation study. *Left:* Comparison between measured user scores for visual comfort and the predicted score using our metric to assess the validity of the approach used to incorporate the influence of luminance contrast spatial frequency. *Center:* Comparison between predicted and measured scores when testing the influence of saliency in the prediction of viewing comfort. *Right:* Scores measured and predicted by our metric for the videos of real scenes. Error bars show standard error of the mean.

Our work can be applied to this area in two ways. First, as shown in the previous paragraph and in Figure 6.1 (left), our measurements can be used to define zones of comfort which could be incorporated as constraints when defining a retargeting operator $\phi(d)$. Second, our metric can provide information about the change of predicted comfort caused by the retargeting operation, taking into account the motion in the scene; this can in turn be used to evaluate or select a given operator. Figure 6.12 shows our predicted distribution of discomfort for the *horse* video, for two different disparity retargeting operators. This provides users with a more insightful view of potential sources of discomfort, both in the temporal and disparity domains.

VISUALIZATION Visualization of complex, three dimensional data is extensively used in various fields, including engineering, geoscience, medicine, biology, architecture or education. In some cases, this visualization can be improved by employing stereoscopic techniques [116, 206]. In such systems, uncontrolled user navigation may lead to visual discomfort; our measurement can be used to provide constraints on the navigation (motion), which would guarantee a comfortable viewing experience.

Given a desired lower bound of the visual comfort level, our measurement could be used to dynamically map the user's input from the navigation device into a comfortable range of camera motions and velocities. In a simple implementation, a sampling of the scene would be performed: At sampled positions, the disparity d would be queried for the current frame(s), as well as an estimation of the dominant luminance spatial frequency f_l. This would be used to set boundaries on the maximum v_{xy} and v_z allowed for navigating the scene. A simple, conservative approach would take the minimum values from all the sampled data; more sophisticated approaches can incorporate importance sampling strategies, either task-oriented or based on visual saliency. This can be made more practical by processing batches of frames.

ASSESSMENT METRIC A number of tools have recently appeared that focus on the editing of stereoscopic content, such as copy-and-pasting [271, 276], drawing [226], converting 2D images to stereo pairs [235, 110], or warping [334]. While the initial content may be assumed to have been carefully generated, it is still hard to predict how any of these editing operations would affect the resulting viewing experience. The metric we propose in this work can be used to evaluate the discomfort that may arise from post-processing operations such as the ones mentioned above.

6.8 CONCLUSION AND FUTURE WORK

We have introduced a novel measurement for the visual discomfort caused by motion in stereoscopic content. A four-dimensional space is explored which includes disparity, planar and depth velocities, as well as the spatial frequency of luminance contrast. Based on these measurements, a metric is proposed to evaluate the level of comfort associated to viewing short stereoscopic videos.

There is ample opportunity for exciting future work. Given the complexity of the HVS, other factors not taken into account here may affect visual comfort, such as the spatial frequency of the disparity, or the temporal frequency of luminance contrast. Similarly, investigating the effect of higher order components of motion (acceleration) can help analyze more complex scenes. Additionally, our saliency estimation does not consider motion or disparity; this is a possible cause of the current overshooting of our metric, which we fix with a fitting function, but deserves further investigation. Our measurements have been tested for a near viewing distance (50 cm in our experiments): Different viewing conditions could be studied using our methodology. Moreover, more sophisticated metrics and models should probably use qualitative information gathered from industry experts. We hope that our work fosters future research in this area, including

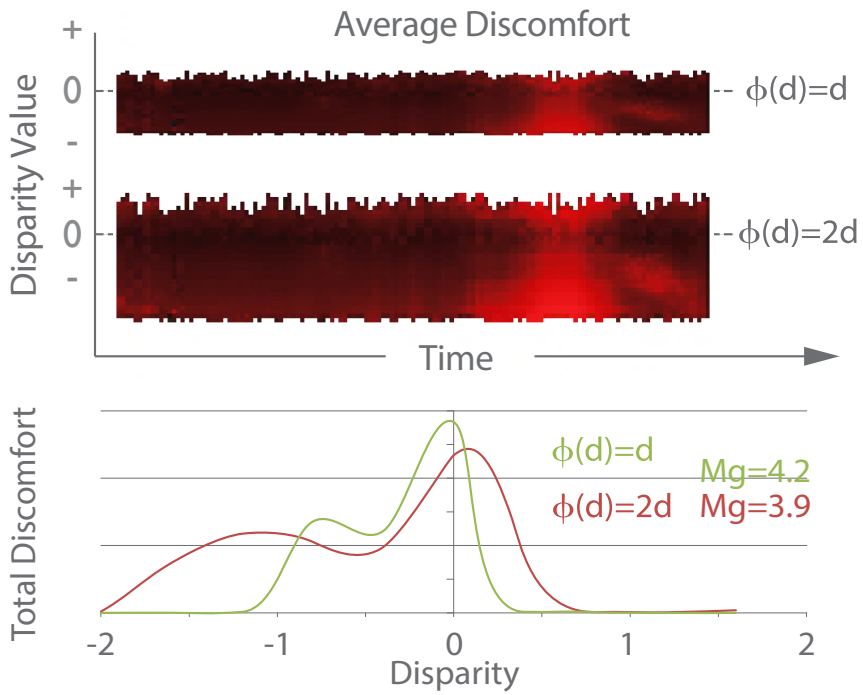


Figure 6.12: For the example video *horse*, we show the distribution of pixel-wise discomfort $(5 - M_p)$ for the original disparity $\phi(d) = d$ and a linear mapping operator $\phi(d) = 2d$. Top: average pixel-wise discomfort. Each vertical line corresponds to one frame. Color indicates discomfort. Bottom: the distribution of total discomfort over d for the whole video. The change of predicted global comfort M_g is also provided.

both stereo applications and a deeper understanding of the mechanisms of our visual system.

Part IV

INTERACTION

This part is motivated by the realization, the thought, that in a near future light fields will play a large role in the imaging pipeline; the fact that they are needed for automultiscopic displays, that they offer much more information from the scene than a photograph, and that plenoptic cameras are already being commercialized are some of the reasons. Here we focus on the problem of how this type of information, this data structure that is a light field, will be edited. We thus study interaction paradigms for light field editing.

7

EVALUATION OF INTERACTION PARADIGMS FOR LIGHT FIELD EDITING

ABOUT THIS CHAPTER

This chapter deals with the study of how would users edit a light field, which has a complex yet very specific structure. The work here presented is not only my work, but also that of Adrian Jarabo (Universidad de Zaragoza), Adrien Bousseau (REVES/INRIA Sophia-Antipolis), Fabio Pellacini (Sapienza University, Rome) and Diego Gutierrez. Not all but the main bulk of the *hands-on* work is being carried out by Adrian Jarabo and myself: Adrian is mostly in charge of the implementation of the interfaces, and I am mostly in charge of the analysis of the experiments. This work is unpublished, and still in progress.

7.1 INTRODUCTION

Light fields [261, 143] are rapidly gaining popularity as an alternative to digital photographs. Light field cameras are already widely available (such as Raytrix™ or Lytro™), and give the user the ability to render a photo of the scene from different perspectives or focus settings. As the light fields usage grows, the need for tools to edit them arises. However, while 2D image editing is well-established, user interfaces to edit light fields remain largely unexplored.

Editing light fields is a challenging task for several reasons. First, a light field is a complex four-dimensional data structure while most existing displays and input devices are designed for manipulating 2D images. Second, light fields are highly redundant which implies that any local edit on a light field needs to be propagated coherently to preserve this redundancy. Finally, while light fields provide a vivid sense of depth, this depth information is not encoded explicitly. Light field user interfaces must take these properties into account to present the visual information in a legible way and to minimize redundant work for the user. In this work, we propose a formal evaluation of interaction paradigms for light field editing. We target the task of drawing strokes in space, analyzing simple operations like painting, dodge and burn and pasting. Since many advanced editing operations rely on the same scribble inputs, our conclusions can be generalized to other tools such as lasso and quick selection, blur, or sharpen.

Existing light field editing tools focus on automatic or user-assisted algorithmic solutions rather than on the interface. Automatic methods are based on edit propagation, or rely on computer vision to estimate depth or *disparity* between views. This information is sufficient to project edits from any view to any another view, but depth estimation algorithms are prone to error on non-diffuse scenes. In most user-assisted approaches, users navigate between the views of the light field to specify correspondences that locate their edits in space. We refer to this class of methods as *multiview interfaces*. A second class of interfaces, which we term *focus interfaces*, leverage the fact that a synthetic shallow depth of field provides both a way to visualize depth in a 2D image and to identify correspondences between views of the light field. Finally, a few prototype *automultiscopic displays* allow light field editing using gesture tracking [299] or 3D light pen [441]. However, we choose to focus on interfaces that are available to a wide audience without the need for specialized hardware.

The paradigms we evaluate in this work rely on different *depth cues* to allow users to specify and visualize the disparity of edits between the views of the light field. While

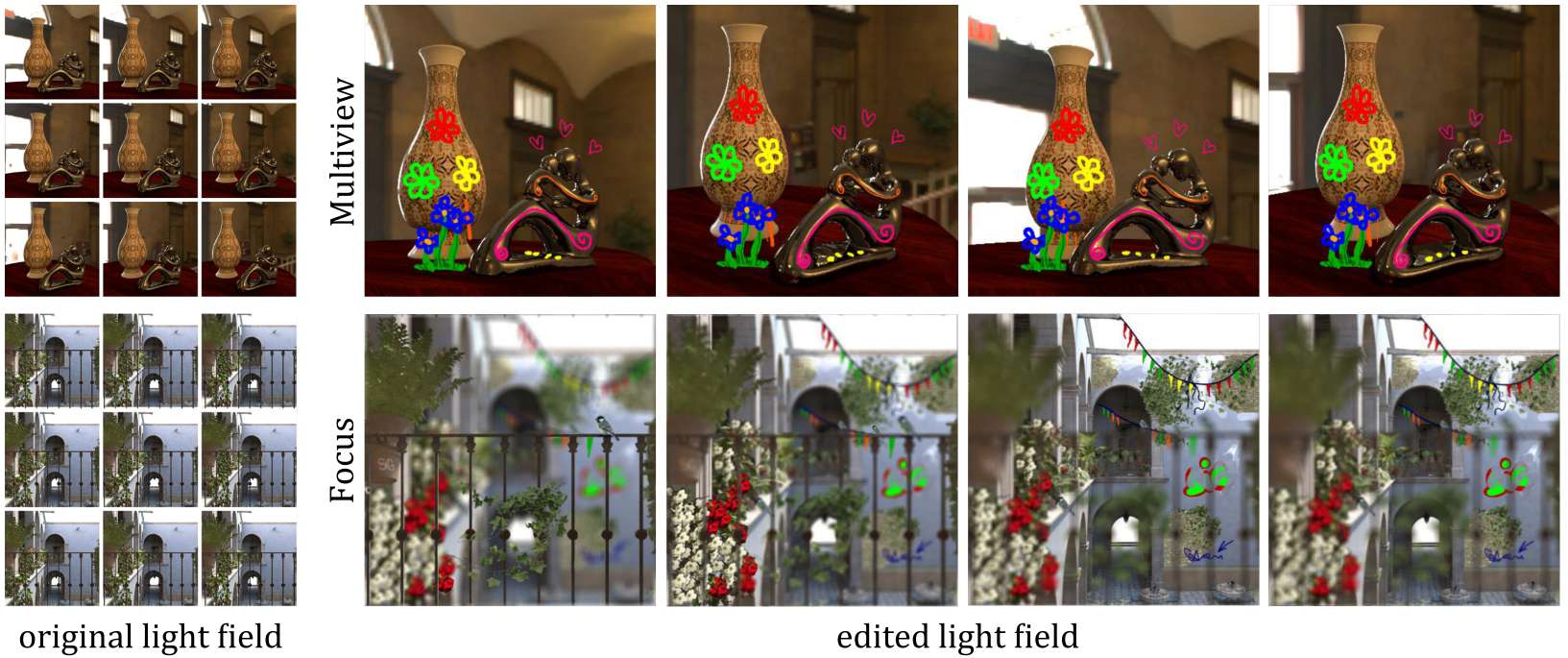


Figure 7.1: Example results of two novice subjects editing light fields using our two interaction paradigms: *multiview*, which relies on parallax cues and *focus*, based on a shallow depth of field representation. In this work we evaluate the relative benefits of these two paradigms, and draw conclusions to help guide future interface designs for light field editing. The examples shown have been obtained with the hybrid interface we describe in Section 7.6.

multiview relies on *parallax* to convey disparity, the *focus* paradigm relies on *depth-of-field blur*. Both cues are reminiscent of the way people experience 3D in the real world, and it is unclear if one cue is preferable to another, even though most existing light field editing tools have chosen so far to rely on the *multiview* approach. Section 7.3 provides detailed descriptions of these interfaces. Additionally, we propose and evaluate two usage scenarios. The first one corresponds to cases where depth information is unreliable or not available, and thus disparity of the edits cannot be automatically inferred from it. This scenario is representative of most existing light field data. The second scenario assumes knowledge of the disparity, that may come from synthetic scenes, depth sensors or advanced computer vision algorithms.

In our study (Section 7.4), we ask the users to complete seven different tasks. Five of them are *directed* tasks, where they are asked to edit an input light field in a similar way to a provided example; the other two are *open* tasks, where we give them reference images for inspiration, but they are free to edit at will. We collected both subjective and objective data during the experiments, the analysis of which can be found in Section 7.5. Finally, Section 7.6 compiles the main findings, and Section 7.7 outlines directions for future work, some of which are already in progress.

7.2 RELATED WORK

Several recent papers evaluate different interface paradigms for specific computer graphics tasks, for example in the case of lighting [217] or material editing [218]. In this work, we perform a similar study focusing on light field interfaces for editing. Light field interfaces can be characterized according to how the correspondences between views are obtained. We discuss here methods based on automatic depth estimation as well as user-assisted methods that rely either on a *multiview* or on a *focus* interaction metaphor.

AUTOMATIC CORRESPONDENCES Seitz and Kutulakos [395] estimate a voxel-based representation of a lightfield to propagate local edits, such as painting and scissoring, between multiple views of a scene. Related methods estimate depth in a stereo pair to perform consistent painting and copy-and-pasting [420, 421, 271, 354, 227]. While depth estimation assumes static scenes, other approaches rely on feature matching to propagate edits over image collections containing deformable objects [159, 153, 504].

Automatic correspondence algorithms are prone to errors inherent in depth estimation and feature matching. Our goal here is to evaluate the performance of user interfaces independently of such errors. To do so, we emulate a perfect depth estimation by rendering synthetic light fields with ground truth depth. While methods that offer automatic correspondence greatly facilitate on-surface editing, our study reveals that they are not suitable to free-space editing. We also complement our study with an evaluation of user-assisted interfaces that could be used to correct for erroneous depth estimation.

MULTIVIEW INTERACTION Most light field editing systems perform consistent operations over multiple views. Jarabo et al. [200] propagate sparse edits in a light field based on pixel affinity. While this approach allows the editing of groups of pixels that share the same appearance, it is not designed to handle local edits. Other systems ask users to indicate correspondences between two or more views. Zhang et al. [505, 464] morph between two light fields by first requiring users to position polygons in several views, constrained by epipolar geometry. Users then indicate corresponding polygons in a second light field to guide the morph. In Pop-Up Light Field [404], users segment the light field into multiple depth layers by adjusting a polygon around the silhouette of each object in multiple views.

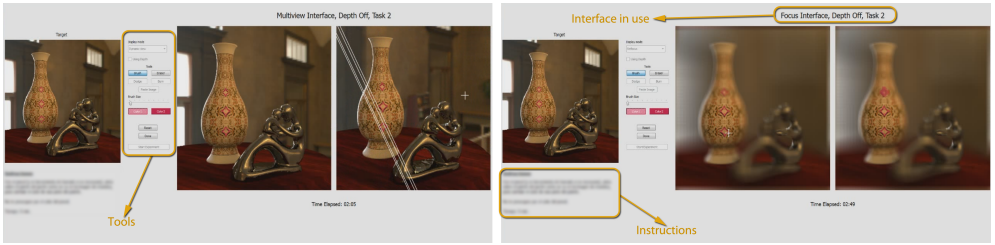


Figure 7.2: User interfaces used in our tests. Left: *multiview* paradigm. Right: *focus* paradigm (instructions blurred out for anonymity reasons; will appear sharp in final version).

FOCUS-BASED INTERACTION Isaksen et al. [194] demonstrate refocusing with a synthetic aperture in their seminal paper on light field reparameterization. A synthetic aperture produces shallow depth of field by blending the multiple views of a light field. As a result, a point appears in focus when its images are aligned, which provides a direct correspondence between the pixels covered by this point in all the views of the light field. While Isaksen et al. noticed that the focal plane provides a simple way to indicate depth, very few papers built on this metaphor. A notable exception is the work of Davis et al. [95] which relies on focus when reconstructing unstructured light fields.

All these papers use either the *multiview* or *focus* interaction paradigms to achieve some form of light field manipulation. However, none of them provide an analysis of the interface itself, nor do they offer insight about which paradigm may be better suited for which specific task.

7.3 INTERFACES

In this section we describe the interfaces being evaluated. The interested reader can refer to the *Directed tasks* video for a practical demonstration of the workflow with each of them; it can be found in: `http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_1.mov`.

OVERVIEW We evaluate two interface paradigms, *multiview* and *focus*, with and without depth information. This yields a total of four interfaces: *multiview without depth*, *multiview with depth*, *focus without depth* and *focus with depth*. All interfaces share the same screen layout, shown in Figure 7.2. On the left, there is a description and example image of the current task. Next to it, a control panel and two working windows, named Window 1 ($W1$) and Window 2 ($W2$). Interface manipulation is performed with a mouse or tablet. For the interfaces without depth information, each user stroke is placed on a plane *parallel* to the camera, so all points share the same depth. We opt for this approach to avoid more complex interactions, such as drawing parametric curves and manipulating their control points. This is not a limitation of the interfaces explored, but a design choice to isolate as much as possible the main object of study. The strokes are modeled as polylines connecting densely sampled points.

MULTIVIEW In the *multiview* paradigm (Figure 7.2, left), the user is presented with two views of the light field, whose viewpoints are independently manipulated by panning and tilting. This allows the artist to view the light field from a different viewpoint than the one used for manipulation, which is a common workflow in 3D software packages. In this interface, the windows are equivalent. All window controls are identical for both windows. To draw a stroke at a given location, the user first draws in one window, where the stroke is created in the view plane at a constant depth that by default is the focal depth of the light field. The stroke depth is then adjusted by translating the

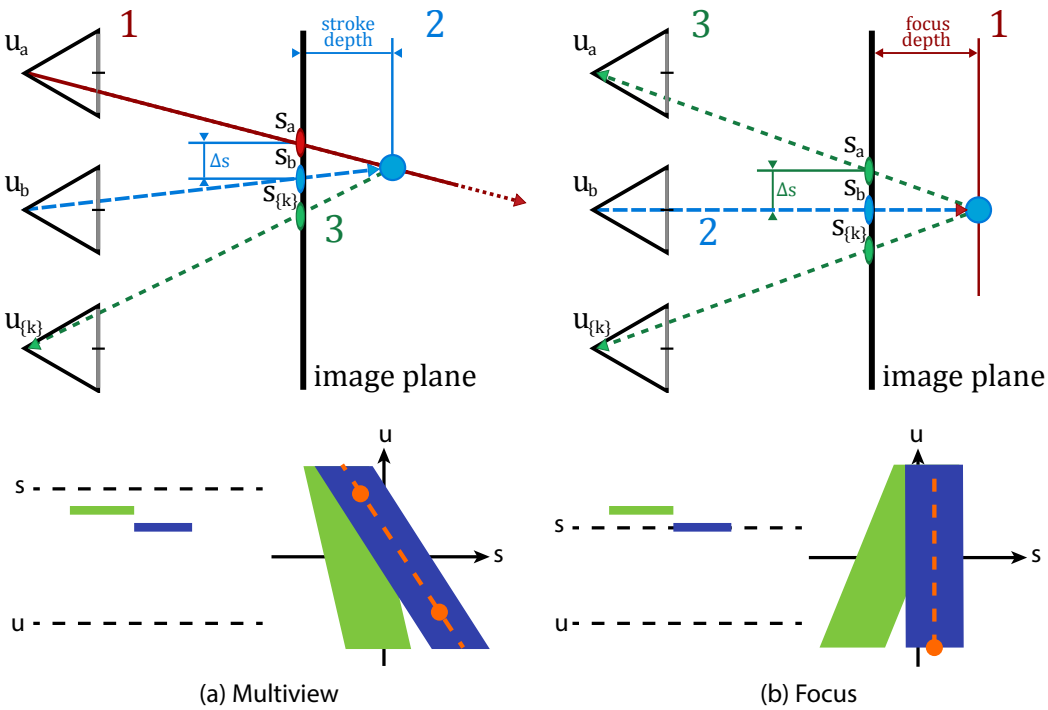


Figure 7.3: Workflow when drawing a stroke in each paradigm. *Top row, left: multiview*. (1) The user first draws a stroke s_a in one view u_a. (2) Depth is then adjusted on a different view u_b by displacing the stroke along the epipolar line. (3) The stroke is then projected onto the other views $u_{\{k\}}$ of the light field, yielding $s_{\{k\}}$. *Top row, right: focus*. (1) The user first specifies depth by placing the focal plane. (2) A stroke s_b is drawn on the central view u_b at the specified depth. (3) The stroke is projected onto the other views of the light field, yielding s_a, and s_b. *Bottom row:* A light field interpretation of the two paradigms. In the *multiview* paradigm, the user specifies two correspondences (orange dots), which provide the disparity of the 3D point (slant in the light field). In the *focus* paradigm, the user first places the point of interest in focus by shearing the light field. As a result, all images of the point are aligned and one scribble is enough to edit all the views.

stroke forward or backward, along the epipolar lines shown as guides, until it lays at the desired location. Typically, this adjustment is performed from another viewpoint (or viewpoints) more suitable for precise placement (see Figure 7.3, top left).

FOCUS In the *focus* paradigm (Figure 7.2, right), the scene is rendered with a wide synthetic aperture that blends all views of the light field [194]. Points that are in focus appear sharp because their images are aligned, while points that are out of focus appear blurry because of the disparity between their images. By construction, this alignment gives us the position of any in-focus point in all views of the light field. In this interface, the user cannot alter the viewpoint, but can adjust the depth of the focal plane of the scene, i.e. the relative disparity of the views [452]. This shallow-depth-of-field visualization and the position of the in-focus plane are common to both windows, but in this interface the windows are not interchangeable. To draw a stroke, the user first adjusts the depth of the focal plane and then draws the desired stroke at such depth in *W1*. Once drawn, the depth of the stroke cannot be modified (see Figure 7.3, top right). As the in-focus plane is moved in depth, both windows show the stroke with the corresponding degree of blur. Since the edit can be completely blurred out when integrating it with the

light field, it is simply pasted on top of it in *W2*, to show its area of influence in other views and help determine occlusions. The light field is hidden in this window when erasing, to facilitate the task (refer to Figure H.1 in Appendix H, and to the example sessions in the videos[1]).

SCENE DEPTH When depth information is used, the users no longer have to adjust the depth of a given stroke; instead, the stroke is projected onto the first visible surface. This allows to draw directly on surfaces (even curved ones) without further adjustment. In the *focus with depth* interface, the views on both windows focus on the visible surface directly under the mouse to further simplify the selection.

EDITING OPERATIONS We make tasks more interesting for users by asking to perform light field editing operations, rather then just positing strokes. We chose to implement *brush painting*, *erasing*, *dodging & burning* and *pasting* of pre-loaded billboards parallel to the camera plane. All these edits are directly controlled by strokes locations. We choose these operations since they are common in most image editing software, they are well-known to users, requiring little training, and they represent simple operations from which more complex edits can be performed (Figure 7.1).

7.4 EXPERIMENTS

GOAL We want to compare all four interfaces with respect to their effectiveness, efficiency and subjective preference. With *effectiveness* we refer to how well the intended task is accomplished, *efficiency* is related to the effort of obtaining a particular output, and *subjective preference* is based on qualitative data, i.e. user opinions on ease of use, learning curve, among others.

LIGHT FIELDS We use three different synthetic light fields, depicting different types of scenes (see Figures 7.1 and I.1 and Appendix L): a complex architectural scene (*San Miguel*), a still life-like scene (*Vase*), and a human head (*Head*). These scenes have different depth, geometry and reflectance complexities. We use synthetic scenes to have precise depth information. We render the scenes with a light field camera implementation in the physically-based renderer PBRT [348]. We use 17×17 views with a resolution of 400×400, in order to achieve real-time interactions at roughly 30 frames-per-second. We up-sample the rendered images to 600×600 during display to facilitate more accurate placement of the strokes.

TASKS We asked users to perform a series of tasks using the editing operations described above. We group tasks in *directed tasks*, where the user has specific instructions on what to edit, and *open tasks*, where the user is only given general guidelines. We refer the reader to Appendix I for the specific instructions and example target images given to the users in each task.

Directed tasks are performed for all four interfaces. Verbal and textual instructions and an example target image are given to the user. We use the central view of an edited light field as the target image. Users are not required to match the target image precisely, but rather to indicate the depth at which the strokes have to be positioned using the target image as a visual reference. Only one tool is available for each task, plus the eraser which is always available; the color brush is limited to one color, to avoid unnecessary

1 The videos showing editing sessions by users can be found in `http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_1.mov`, `http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_2.mov`, and `http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_3.mov`.

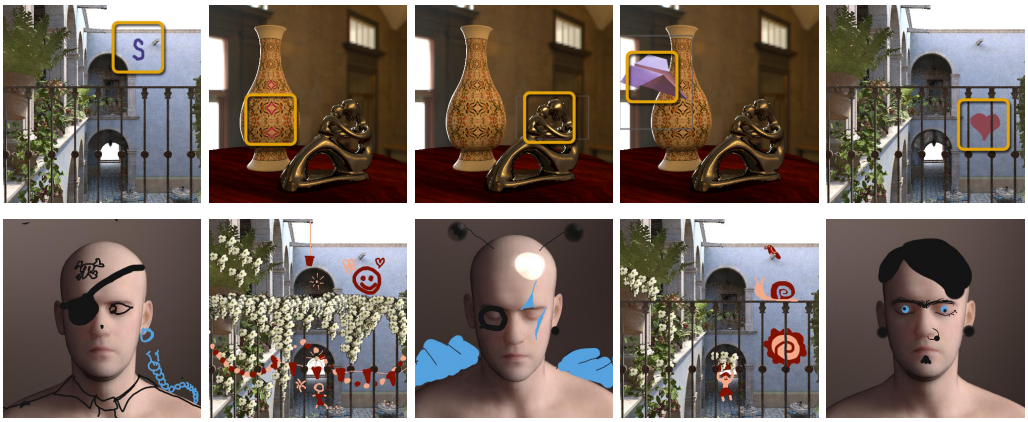


Figure 7.4: Top row: Target images given to users in the *directed* tasks. The small highlighted areas have been added to this figure for visualization purposes and future reference. *Bottom row:* Example results of user edits for the *open* tasks.

distractions. Time to completion is limited to 5 minutes. These tasks are the following (see Figure I.1):

- T1 Draw a simple object on a surface of constant depth (*San Miguel*)
- T2 Paint a simple pattern on a non-planar surface (*Vase*)
- T3 Increase the brightness of the specular reflections on a curved surface (*Vase*)
- T4 Place an object billboard within a certain depth range in free space (*Vase*)
- T5 Draw on a partially occluded surface (*San Miguel*)

The five tasks have been chosen to cover a wide range of use cases: Tasks 1 and 2 are devised to test general editing of surfaces, while Task 3 deals with the particular case of specular highlights, which do not lie on the surface of the object. Task 4 investigates how to work in free space, while Task 5 tests how to best deal with occlusions.

After performing the directed tasks, subjects complete two *open tasks*, where real-world photos are given as a source of inspiration, and participants are free to use all the tools at will, plus two different colors for the brush. Time to completion is limited to 12 minutes. The tasks vary based on interface selection by the user:

- T6 The user is allowed to select whether to use depth information or not during editing. The task is done twice per subject, once with the *multiview* paradigm (toggling freely between using or not depth), and once with *focus* (also with or without depth). The task is carried out on the *Head* light field.
- T7 The user is allowed to freely change between the four interfaces. The task is done on the *San Miguel* light field.

ERROR METRIC To evaluate how well a user can specify locations in the light field, we measure the *error in depth* of the stroke. We choose this over measuring image-based differences since our tasks are not pure matching tasks. Specifically, for each view of the light field, we first compute the L_1 distance between the depth of the stroke and the target depth, for each pixel of the stroke, and divide it by the number of pixels covered by such stroke. We then average across all views of the light field. Our experiments showed that L_1 averaged across views approximates a normal distribution better than other metrics, which facilitates the subsequent analysis. Note that in Task 4 (positioning in free space) there is not a single fixed target depth, but a valid range between the vase and the sculpture. We thus compute error in depth with respect to the limits of such range, assigning a value of zero within it.

EXPERIMENTAL PROCEDURE The study consisted of two main blocks: *multiview* and *focus*, in randomized order for each user. Within each block, the two versions of the interface were used (with and without depth), also in randomized order to compensate for possible learning effects. This yielded a total of four sessions, with each one including all five tasks sequentially (T1 to T5). After each block, subjects were asked to complete Task 6 with the current interaction paradigm. After completing both blocks, subjects additionally performed the final Task 7. We recorded the screen during all the experiments.

After finishing a session with an interface or an open task, users had to fill in a questionnaire and could write free-form comments as well. At the end, subjects were required to fill in a final questionnaire where they had to rate and rank interfaces per task, and also regarding other more general aspects. All questionnaires can be found in `http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Questionnaires.pdf`. Each participant completed the experiment with an informal interview, to collect general impressions and ask about the subject's workflow.

Twenty paid subjects participated in the experiment (6 female, 14 male). All of them had previous knowledge on image editing, 3D modeling or 3D sculpting software, with either an artistic or technical background. Most of them (90%) had no previous knowledge of light fields, which were briefly introduced to them in the beginning. Although the participants were recommended to use a pen on a tablet, they were allowed to use a mouse if they felt more comfortable using it, to ensure that their performance was not affected by the input device. The full experiment took around four hours per subject, including training and short breaks. The training took around one hour, including filling in a preliminary questionnaire, and was performed with an additional light field, shown in Appendix L (Figure L.4).

7.5 ANALYSIS

Throughout the experiment we collected both quantitative data on task errors and timings, qualitative data on performance and difficulty of both tasks and interfaces, and free-form comments on interface effectiveness. In this section, we report the analysis of both quantitative and qualitative data. We report the main findings here and include additional data in Appendix K. For brevity, we will refer to our interfaces in the rest of the chapter as *M* (multiview without depth), *MD* (multiview with depth), *F* (focus without depth) and *FD* (focus with depth).

We use repeated measures ANOVA for the analysis of error, timings and ratings, and Kruskal-Wallis for the analysis of rankings. In all tests, we use a p-value of 0.05 to indicate significance. When sphericity is violated, according to Mauchly's test, we report significance values adjusted with the Greenhouse-Geisser correction [86]. In all figures, error bars represent the standard error of the mean.

OUTLIERS Subjects were comfortably able to complete a task with all interfaces, with only a few exceptions. We detect these exception by performing outlier rejection on the measured error data, based on the interquartile difference, with a factor of 2.2. This is a conservative choice recommended for small sample sizes [173]. This led to dropping one user in Tasks 2, 3 and 5, two users in Task 1 and three users in Task 4.

ERROR IN DEPTH Figure 7.5 shows the per interface mean error for each of the directed tasks (T1–T5). The error is highly dependent on the task, which accounts for 73% of the variance. When taking into account interfaces, the ANOVA yielded significant differences between interfaces for all tasks, as summarized in Table 7.1. Figure 7.5 (bottom row) additionally illustrates significant differences between interfaces according to the pairwise comparisons tests.

Tasks 1 to 3 required drawing strokes onto non-occluded surfaces. *M* yielded a higher error ($p \leqslant 0.018$) than *F* ($p \leqslant 0.018$), showing that users found it more difficult to locate an edit in depth. In these tasks, when interfaces with depth (*MD* and *FD*) are used the error in depth is zero, since strokes directly snap to the surface.

In Task 4, which requires positioning in free space, the trend is reversed: interfaces without depth yield lower errors. An interesting finding is that neither the difference between *M* and *F* nor between interfaces with depth (*MD* and *FD*) is significant. Task 5 is possibly the most complex, since it requires handling occlusions and large depth discontinuities. *F* yields the lowest error, while *M* yields the highest.

This analysis also suggests that while interfaces with depth information work well when manipulating surfaces without occlusions, not using depth information is actually more effective when occlusions are present, or when the editing task requires positioning in free space. This is due to the fact that the edits will snap to the underlying depth of the light field, which is not desirable in those particular cases. The *Directed tasks* video (`http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_1.mov`) shows Task 5 (handling occlusions) being performed with all four interfaces, as well as Tasks 1 to 4 performed with different interfaces.

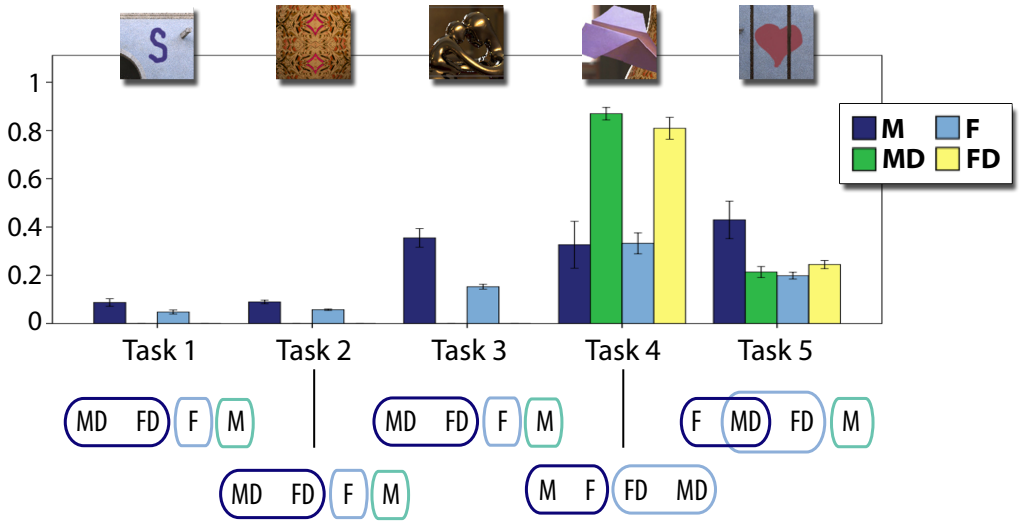


Figure 7.5: Top: Mean error per interface for each task. Bottom: Pairwise comparisons for the error in each task. Items in the same set are statistically indistinguishable.

TIME TO COMPLETION We plot mean times to completion per interface for each directed task (T1–T5) in Figure 7.6 (top row), and also illustrate statistically significant differences between them (bottom row). In general, users were able to complete the tasks in the allocated time with all interfaces. For tasks T1 to T3, which require placing strokes on surfaces, interfaces with depth information (*MD* and *FD*) took less time, although the difference is only significant with respect to *M* ($p \leqslant 0.008$). There is no significant differences in Task 1, due to its simplicity.

Task 4 yields very low times in general, while it was the one with the highest errors. This is interesting, since it is the only task that specifically demands positioning in free space rather than on a surface. For *MD* and *FD*, this is likely due to users realizing that those interfaces are not appropriate for this task and just giving up quickly. This hypothesis seems supported by the low ratings these two interfaces received in this task (Figure 7.8). However, in general, results in this task suggest the difficulty of users to correctly judge depth in free space, using any of the four interfaces: they simply place the billboard at some reasonable point in the scene, occluding the vase. These behaviors

	H	(df_1, df_2)	p	$\eta^2(\%)$
T1	25.230	(1.517, 25.792)	0.000	59.7
T2	138.745	(1.305, 23.491)	0.000	88.5
T3	70.390	(1.612, 20.108)	0.000	79.6
T4	24.951	(1.861, 29.779)	0.000	60.9
T5	6.275	(1.264, 22.760)	0.015	25.8

Table 7.1: Results of the repeated measures ANOVA for the interface factor for the error in depth in each of the tasks. H is the test statistic, df_1 and df_2 the between-group and within-group degrees of freedom, respectively, p the associated significance and η^2 is indicative of the proportion of variance of the data that the interface factor explains.

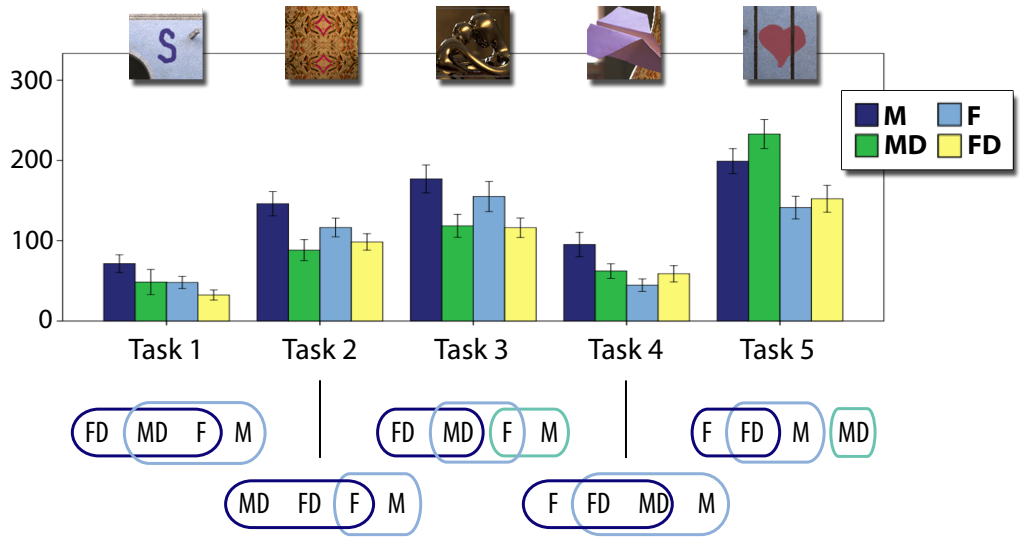


Figure 7.6: Top: Mean time to completion per interface for each task. Bottom: Pairwise comparisons for the time to completion in each task. Items in the same set are statistically indistinguishable.

would also explain the high errors reported in the previous subsection. A closer analysis reveals that *F* takes the least time, although the difference is only significant with respect to *M* ($p \leqslant 0.008$).

Task 5 requires dealing with occlusions. Based on time data, *MD* seems not to be a useful interface, to the point that some subjects did not complete the task in the given time (in particular eight of the subjects, seven of them with the *MD* interface). This is because handling occlusions in *MD* requires erasing occluded parts in various different views, which is time consuming.

Overall, having depth information leads to faster editing when painting on surfaces, as long as no occlusions are present; *M* tends to take longer than the rest, apparently being less intuitive for users, while *F* performs well in most situations, specially when dealing with occlusions or positioning in free space, which are the two most challenging scenarios in our tests.

RANKINGS AND RATINGS The final questionnaire contained eleven questions in which the users had to rank and rate the four interfaces. Five questions referred to the preference of interface for each of the five directed tasks, and one to the overall preference. The remaining five questions investigate preference in more general aspects, namely positioning in depth and on a plane (x-y), erasing, perceived accuracy of the interface and difficulty of use. For each ranking, we also compute the rank product per interface $\Psi(\vartheta) = (\prod_i r_{\vartheta,i})^{1/m}$, where $r_{\vartheta,i}$ is the ranking received by interface ϑ in a specific question and m the number of subjects [382]. We use rank products to sort the interfaces when grouping them in statistically different groups (Figure 7.7, bottom, and Figure 7.9, left). Actual values of the rank products can be found in Appendix K.

Rankings for the different tasks (Figure 7.7) exhibit again a large between-task variability, in accordance with the error and time to completion. In Tasks 2 and 3 *MD* and *FD* are ranked significantly higher ($p \leqslant 0.035$) than no-depth interfaces (*M* and *F*). The difference between the interfaces with depth (*MD* and *FD*) is not significant ($p = 0.160$). In Tasks 4 and 5 the trend is again reversed: there is a clear preference for interfaces without depth, and in particular for *F* ($p \leqslant 0.035$ and $p \leqslant 0.008$ in Tasks 4 and 5, respectively). From Task 4, we can conclude that in the absence of references in free space, the *focus* paradigm offers better depth cues.

When asked about the overall ranking, results show that users do not have a clear preference: *MD* ranks first, significantly higher than *FD* and *M* ($p \leqslant 0.011$), but there is no significant difference between the rest. This is probably due to the large dependency on the task, shown by previous analyses. Despite the similarity of both interfaces in the rest of the tests, the users reported that the *multiview* paradigm allowed them a better visualization of the light field and the edits. Mean ratings for preference per task and overall preference are shown in Figure 7.8, including results of pairwise comparisons between interfaces. These ratings strongly correlate with rankings (Spearman's $\rho = 0.80$, $p \simeq 0.000$). This is meaningful, indicating that users have clear opinions regarding the interfaces for the different tasks.

Users' preferences for the rest of the questions, on general aspects, are illustrated in Figure 7.9, both for rankings (left column) and ratings (right column). For rankings, interfaces are ordered according to their rank product. We see that *F* ranks first in most cases, with no significant difference among the others. When it comes to accuracy, agreement among users decreases, and differences turn out not significant. Overall, what we extract from this analysis is users' inclination towards the *focus without depth* interface. Again, the high correlation between rankings and ratings is apparent.

PREFERENCES IN OPEN TASKS During open tasks, when users can freely toggle the use of depth and interface, we record the time spent on each interface, and what actions are performed in each of them. Specifically, we track the time they spend drawing, erasing, changing the view point, and adjusting depth. The results are shown in Figure 7.10. The *Open tasks* video (`http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_2.mov`) shows sample editing sessions by subjects for these tasks.

In Task 6, the times spent with and without depth for each interface are relatively balanced. This situation changes in Task 7, possibly as a consequence of the different nature of the light fields involved: the *Head* in Task 6 is a large non-planar surface, where having depth information is highly useful, whereas *San Miguel* in Task 7 has many flat surfaces and larger depth discontinuities with free-space in between.

Nevertheless, the analysis of Task 7 reveals a clear general preference for interfaces without depth, and for *F* in particular. It can be seen how most of the pure editing operations are performed in *F*, while *M* is used mainly to change the view point. This is contrary to the assumption that interfaces with depth information would be preferred; when given complete freedom, the preferred workflow is to perform edits with the *focus* paradigm, use *multiview* to inspect the changes from different viewpoints, and then go

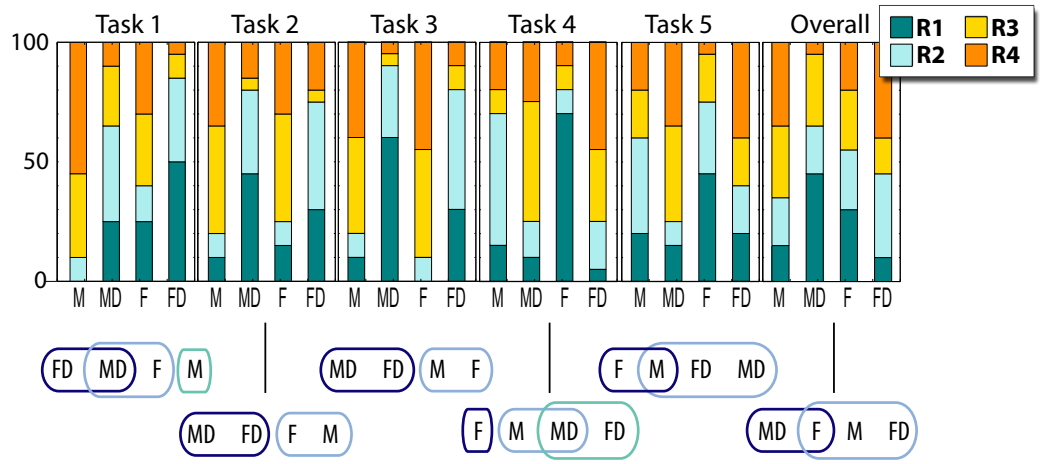


Figure 7.7: Top: Rankings from final questionnaire for questions on preference for each task and overall preference (Ri: rank i). Bottom: Pairwise comparisons between interfaces for the rankings. Items in the same set are statistically indistinguishable.

back to editing with *focus* again. This is supported by the users' feedback in the final interview, and Table K.11 in Appendix K, which shows the total number of times users went from one interface to another during editing.

Based on their answers to the after-task questionnaires, 85% of the users found the possibility to switch between interfaces during Task 7 helpful or very helpful (as indicated by a rating of 4 and 5 on a scale from 1 to 5). In Task 6, both for *focus* and *multiview*, the percentage of people who found toggling depth helpful or very helpful is lower, 75%, but still suggesting that there is not one clear winner.

7.6 DISCUSSION AND CONCLUSIONS

CONCLUSIONS A thorough analysis of the objective and subjective data collected during the user study allows us to draw the following conclusions:

- A future interface for light field editing would need to incorporate both paradigms; our tests on a hybrid interface following our experiments show the viability of such approach
- In general, the *focus without depth* interface is preferred to specify a position of the light field in free space
- For visualizing the light field, *multiview* is preferred by users; parallax cues offer a more intuitive visualization for a volume in space
- When editing on surfaces, *multiview with depth* is the best option, since it combines the benefits of having depth information with a more natural visualization
- Interfaces with depth information are not always preferred, with a high dependency on the task being performed
- Handling occlusions is challenging for users; the *focus without depth* interface is again preferred for this
- The simple tools we include in our tests allow rather complex editing processes in light fields of very different nature
- The vast majority of users were able to perform the tasks with all interfaces

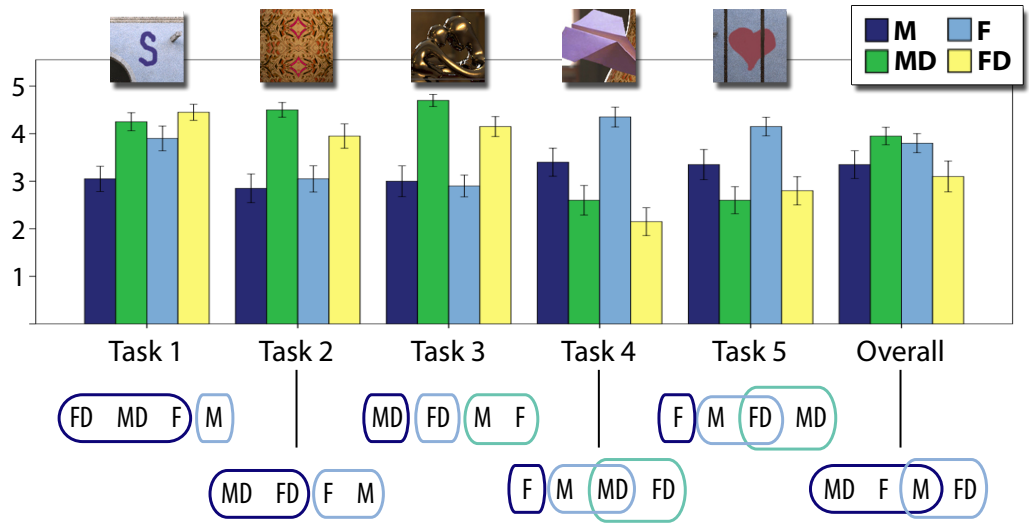


Figure 7.8: Top: Mean ratings from final questionnaire for questions on preference for each task and overall preference. Bottom: Pairwise comparisons between interfaces for the ratings. Items in the same set are statistically indistinguishable.

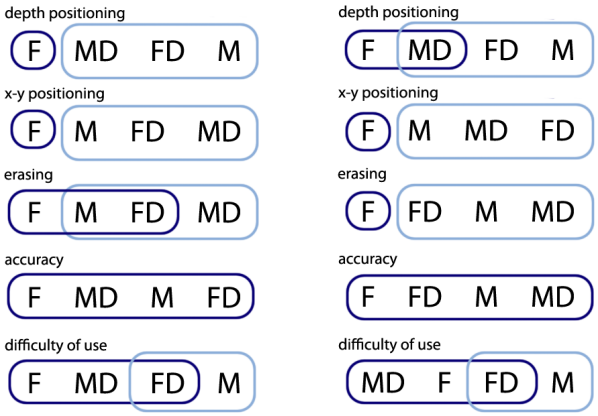


Figure 7.9: Rankings ordered by rank product (left) and ratings by mean (right), for questions on general aspects. Groupings show significant differences between interfaces. Rankings and mean ratings for these questions are shown in Appendix K.

Possibly the most important finding is that a future light field interface will need to offer a combination of both paradigms. There is a clear preference to edit using *focus*, and visualize and navigate using *multiview*. In fact users consistently argued in favor of a hybrid interface. This makes sense, since once the location of the stroke is fixed by adjusting depth-of-field, editing with *focus* is very similar to working with any image editing program, as opposed to having to rely on epipolar lines to locate the edits. On the other hand, interpreting a 3D volume by looking at it from different points of view (which *multiview* allows) is more natural than visualizing it based on shallow depth of field. The workflow in the *focus* paradigm (adjust depth, then edit) seems also more intuitive than in *multiview* (edit, then adjust depth). Also, the data show that relying on depth information is very useful for certain tasks (editing on non-occluded surfaces), but not for others (editing on occluded surfaces or in free space). However, note that, since we are using controlled, synthetic light fields, our depth estimation is more reliable than

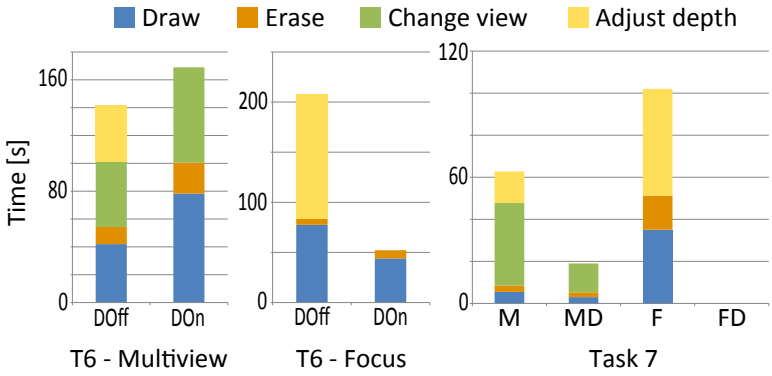


Figure 7.10: From left to right: Distribution of times for Task 6 using the *multiview* and *focus* paradigms, and for Task 7. Note that we do not take idle times into account. We plot median values, which makes *FD* in Task 7 become zero in all four categories.

current state of the art algorithms for real-world light fields. It is remarkable that Task 7 was the most gratifying for the users, possibly because it allowed them to edit with all four interfaces, leveraging the advantages of such a hybrid interface (although this is open to interpretation and may need further studies).

When forced to rank and rate, the two preferred interfaces are *F* and *MD*. This can be explained by analyzing both paradigms individually, with and without depth. For Tasks 1 to 3, which involved editing on surfaces, there is no significant difference between *MD* and *FD*. However, the interviews and rankings and ratings on overall preference show a preference for *MD* over *FD* (see groupings in Figures 7.7 and 7.8). Also, when given the choice in the open tasks 6 and 7, users spend more time working with MD (Figure 7.10).

For tasks not involving editing a visible, non-occluded surface (drawing, dodging, burning, erasing), the users' preference changes to interfaces without depth, and more specifically to *F*. The groupings for tasks 4 and 5 in the previous section clearly show this. Additionally, Figure 7.10 shows that users chose to spend more time working in *F*. In general, *F* is regarded as much more intuitive (requires less of a learning curve) and fast than *M*. Users found *M* very accurate, but only if clear references were present to establish view-to-view correspondences when positioning a stroke in depth. In Task 1, for instance, it ranks very low because of the texture-less wall they have to paint on. In the rest of cases, it required too much view-changing and depth-adjusting. According to users, *F* offered a very simple and appealing way to edit in depth. The only drawback was that they could not properly visualize the whole light field, which again reinforces the idea of a hybrid interface.

Another interesting finding is the fact that all users edited specular highlights as if they were a feature on the surface of the object. This is physically inaccurate since they are actually detached from the reflecting surface [437]. However, nothing was reported in this regard during the interviews, and the users systematically ranked and rated *MD* better (see Figures 7.7 and 7.8). This confirms previous findings on the inability of the human visual system to correctly assess the physical accuracy of reflections and highlights [373, 133].

HYBRID INTERFACE We have implemented a hybrid interface that displays the light field using *focus* in one window, and *multiview* in the other. This combines the advantages reported for both interfaces, based on the analysis of the data and the users' feedback. The use of depth can be switched on and off. Additionally, we introduce two new editing tools: *spline* and *image deformation*, to test how our interfaces generalize to non-scribble-

based interactions. These two tools are based on control points, which can be modified in 3D by the users to perform edits.

We run a pilot study with four experienced participants using this hybrid interface, in a completely open task with any of the three light fields. The feedback from the users was extremely positive: they all agreed that a light field interface should combine both paradigms, and found the new tools also easy to use with this interface. Additional details can be found in Appendix J, and a video example of an editing session can be found in the *Hybrid interface* video (`http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_3.mov`).

7.7 FUTURE WORK

LIMITATIONS Interfaces for light field editing remain largely unexplored, and as such there are many more opportunities for future studies and novel tools. For instance, several users came up independently with the idea of a *depth picker* tool, similar to the color picker in Photoshop™, as a useful tool for light field interaction. As with any user study, our conclusions are only strictly valid for the tested scenarios. For instance, we assume structured light fields [261], such as those acquired with existing commercial plenoptic cameras (Lytro™, Raytrix™); unconstrained light fields (e.g. [95]) are much less common, and out of the scope of this work. However, we argue that the scenes are diverse enough and the editing tools common enough to make extrapolations and aid in the design of future editing tools and interfaces, as our pilot study with an hybrid prototype suggests.

OUTLOOK AND FUTURE WORK We believe that we have taken the first steps towards the design of a light field editing interface, and that our principled evaluation methodology has lead to important findings, and a better understanding of editing in the multidimensional space of light fields. However, this is, as mentioned, just the first step.

This study has allowed us to draw insights on where do the challenges lie, what are the preferred workflows, which are the strengths and weaknesses of each paradigm, and what tools should the final interface have.

We can further summarize the conclusions in two: (a) multiview with depth information is the interface preferred for most editing tasks, its main drawback being occlusion handling, and free space editing; and (b) focus without depth information is the preferred alternative in those cases in which multiview with depth fails, especially because of the high degree of control it allows the users.

Building on this we are currently working on the first version of a new interface. Additionally, we want to analyze a more realistic scenario in which users do not perform the editing with perfect (ground truth) depth information but with imperfect depth maps resulting from current algorithms for depth reconstruction from light fields [468, 221, 277]. The depth information in these light fields is often inaccurate, complicating the editing process and leading to new workflows which we will analyze on a second study. Finally, based on the gathered conclusions, we plan to propose a light field editing interface and evaluate its usability and performance in a third and final study.

Part V

FEMTO-PHOTOGRAPHY AND TRANSIENT IMAGING

Femto-photography is the term used to refer to a new imaging technique capable of capturing a scene with a temporal resolution of less than two picoseconds. As a consequence, in each captured frame, light travels less than a millimeter; this implies that with this technique we can actually see light propagating through a macroscopic scene. We first describe the acquisition system, and the data processing required for adequate visualization, and then present the relativistic effects we need to deal with to visualize the data in the case in which the camera is moved through the scene. This transient imaging technique has opened up a whole new field of possibilities.

8

FEMTO-PHOTOGRAPHY: ACQUISITION AND VISUALIZATION

ABOUT THIS CHAPTER

This chapter describes the system used to capture time-resolved data, and the subsequent processing that this data has to undergo for correct and comprehensible visualization, and the work here presented was accepted to SIGGRAPH 2013 and consequently published on the journal Transactions on Graphics. I started working in this topic during my first stay at the Camera Culture Group at MIT Media Lab, which is the group that led this project. My participation has been on devising visualization methods for the data with Di Wu at MIT Media Lab (Section 8.5), and I also collaborated on the time warping part (Section 8.6), which was carried out in Zaragoza led by Adrián Jarabo. In this chapter, the whole work on acquisition of time-resolved data using femto-photography is described for completeness and context.

I then worked on several projects which further spawned from this work. One of them is described in Chapter 9. The other is a minor collaboration on a project which deals with analysis of light transport using time-resolved data, taking advantage of the high temporal resolution: A model of light transport is presented, together with techniques for the separation of light components, and then a number of applications are demonstrated. My contribution was in the depth recovery in the presence of interreflections, an unsolved problem in Computer Vision which can benefit from this time-resolved data. Since my contribution was minor, and the work is already accepted for publication, this part is not included in this manuscript; instead, we refer the interested reader to the paper [490] (Sections 4 and 6).

A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez and R. Raskar
Femto-Photography: Capturing and Visualizing the Propagation of Light.
ACM Transactions on Graphics 32(4). (Proc. of SIGGRAPH 2013).

8.1 INTRODUCTION

Forward and inverse analysis of light transport plays an important role in diverse fields, such as computer graphics, computer vision, and scientific imaging. Because conventional imaging hardware is slow compared to the speed of light, traditional computer graphics and computer vision algorithms typically analyze transport using low time-resolution photos. Consequently, any information that is encoded in the time delays of light propagation is lost. Whereas the joint design of novel optical hardware and smart computation, i.e, computational photography, has expanded the way we capture, analyze, and understand visual information, speed-of-light propagation has been largely unexplored. In this paper, we present a novel ultrafast imaging technique, which we term *femto-photography*, consisting of femtosecond laser illumination, picosecond-accurate detectors, and mathematical reconstruction techniques, to allow us to visualize movies of light in motion as it travels through a scene, with an effective framerate of about one half trillion frames per second. This allows us to see, for instance, a light pulse scattering inside a plastic bottle, or image formation in a mirror, as a *function of time*.

CHALLENGES Developing such time-resolved system is a challenging problem for several reasons that are under-appreciated in conventional methods: (a) brute-force time

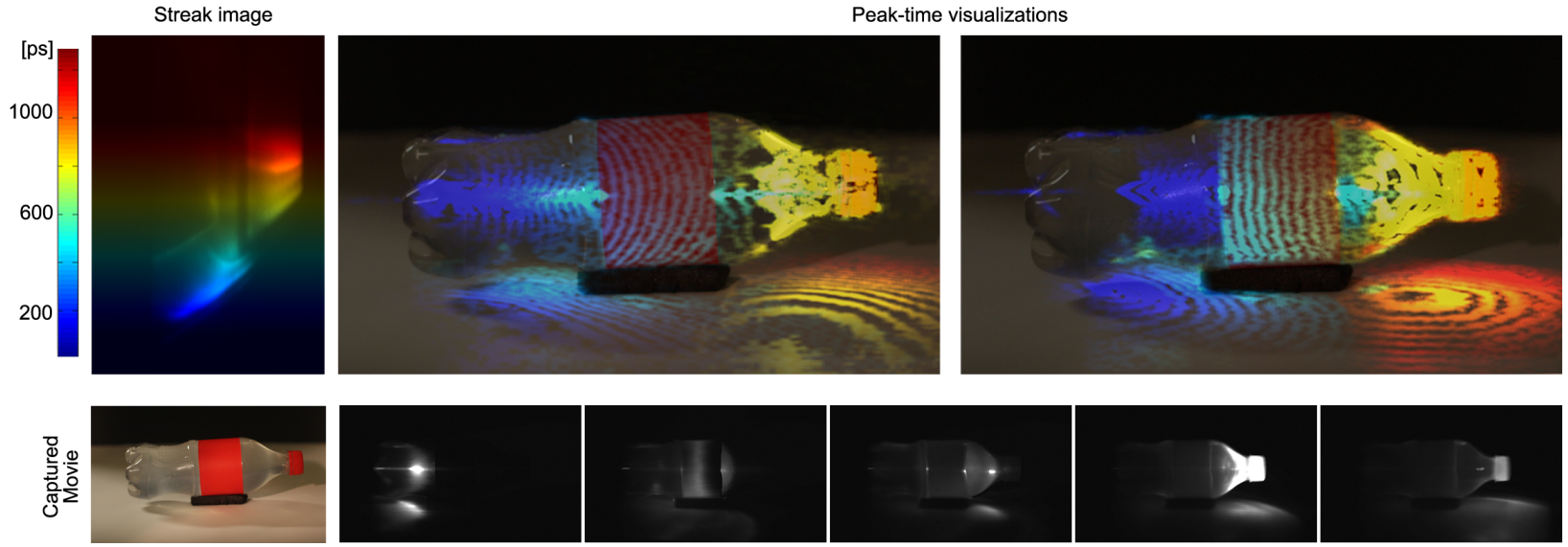


Figure 8.1: What does the world look like at the speed of light? Our new computational photography technique allows us to visualize light in ultra-slow motion, as it travels and interacts with objects in table-top scenes. We capture photons with an effective temporal resolution of less than 2 picoseconds per frame. Top row, left: a false color, single *streak image* from our sensor. Middle: time lapse visualization of the *bottle* scene, as directly reconstructed from sensor data. Right: time-unwarped visualization, taking into account the fact that the speed of light can no longer be considered infinite (see the main text for details). Bottom row: original scene through which a laser pulse propagates, followed by different frames of the complete reconstructed video. For this and other results the reader may refer to the video at `http://webdiis.unizar.es/~bmasia/downloads/thesis/Femto-Main_Video.mp4`.

exposures under 2 ps yield an impractical signal-to-noise (SNR) ratio; (b) suitable cameras to record 2D image sequences at this time resolution do not exist due to sensor bandwidth limitations; (c) comprehensible visualization of the captured time-resolved data is non-trivial; and (d) direct measurements of events appear warped in space-time, because the finite speed of light implies that the recorded light propagation delay depends on camera position relative to the scene.

CONTRIBUTIONS Our main contribution is in addressing these challenges and creating a first prototype as follows:

- We exploit the statistical similarity of periodic light transport events to record multiple, ultrashort exposure times of one-dimensional views (Section 8.3).
- We introduce a novel hardware implementation to sweep the exposures across a vertical field of view, to build 3D space-time data volumes (Section 8.4).
- We create techniques for comprehensible visualization, including movies showing the dynamics of real-world light transport phenomena (including reflections, scattering, diffuse inter-reflections, or beam diffraction) and the notion of *peak-time*, which partially overcomes the low-frequency appearance of integrated global light transport (Section 8.5).
- We introduce a *time-unwarping* technique to correct the distortions in captured time-resolved information due to the finite speed of light (Section 8.6).

LIMITATIONS Although not conceptual, our setup has several practical limitations, primarily due to the limited SNR of scattered light. Since the hardware elements in our system were originally designed for different purposes, it is not optimized for efficiency and suffers from low optical throughput (e.g., the detector is optimized for 500 nm visible light, while the infrared laser wavelength we use is 795 nm), and from dynamic range limitations. This lengthens the total recording time to approximately one hour. Furthermore, the scanning mirror, rotating continuously, introduces some blurring in the data along the scanned (vertical) dimension. Future optimized systems can overcome these limitations.

8.2 RELATED WORK

ULTRAFAST DEVICES The fastest 2D continuous, real-time monochromatic camera operates at hundreds of nanoseconds per frame [141] (about $6 \cdot 10^6$ frames per second), with a spatial resolution of 200×200 pixels, less than one third of what we achieve. Avalanche photodetector (APD) arrays can reach temporal resolutions of several tens of picoseconds if they are used in a photon starved regime where only a single photon hits a detector within a time window of tens of nanoseconds [72]. Repetitive illumination techniques used in incoherent LiDAR [443, 137] use cameras with typical exposure times on the order of hundreds of picoseconds [66, 80], two orders of magnitude slower than our system. Liquid nonlinear shutters actuated with powerful laser pulses have been used to capture single analog frames imaging light pulses at picosecond time resolution [113]. Other sensors that use a coherent phase relation between the illumination and the detected light, such as optical coherence tomography (OCT) [185], coherent LiDAR [492], light-in-flight holography [1], or white light interferometry [491], achieve femtosecond resolutions; however, they require light to maintain coherence (i.e., wave interference effects) during light transport, and are therefore unsuitable for indirect illumination, in which diffuse reflections remove coherence from the light. Simple streak sensors capture incoherent light at picosecond to nanosecond speeds, but are limited to a line or low resolution (20 × 20) square field of view [68, 196, 403, 137, 233, 355]. They

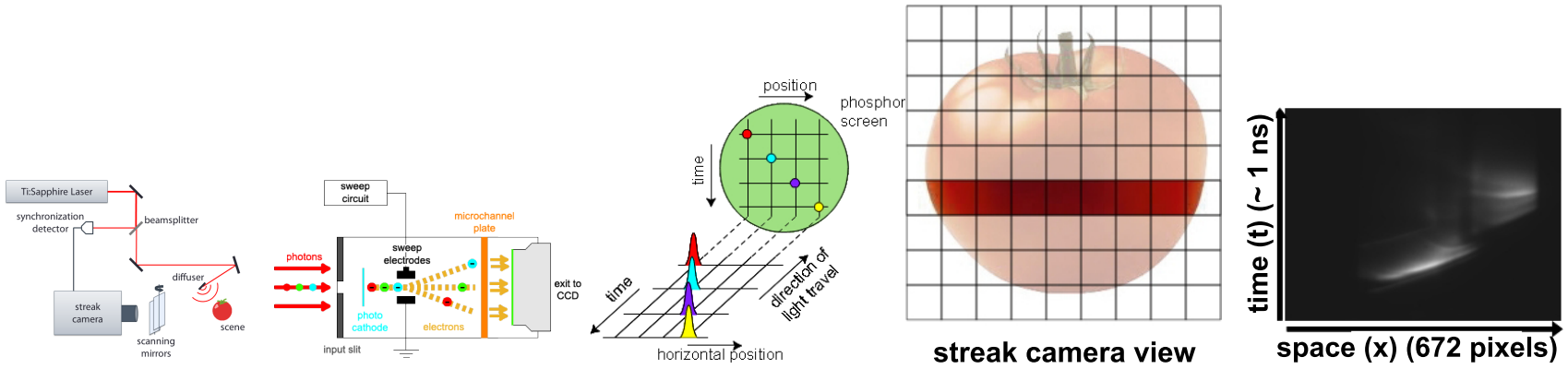


Figure 8.2: Our setup for capturing a single 1D space-time photo. (a) A laser beam strikes a diffuser, which converts the beam into a spherical energy front that illuminates the scene; a beamsplitter and a synchronization detector enable synchronization between the laser and the streak sensor. (b) After interacting with the scene, photons enter a horizontal slit in the camera and strike a photocathode, which generates electrons. These are deflected at different angles as they pass through a microchannel plate, by means of rapidly changing the voltage between the electrodes. (c) The CCD records the horizontal position of each pulse and maps its arrival time to the vertical axis, depending on how much the electrons have been deflected. (d) We focus the streak sensor on a single narrow scan line of the scene. (e) Sample image taken by the streak sensor. The horizontal axis (672 pixels) records the photons' spatial locations in the acquired scanline, while the vertical axis (1 nanosecond window in our implementation) codes their arrival time. Rotating the adjustable mirrors shown in (a) allows for scanning of the scene in the y-axis and generation of ultrafast 2D movies such as the one visualized in Figure 8.1. (Figures (a)-(d), credit: [136])

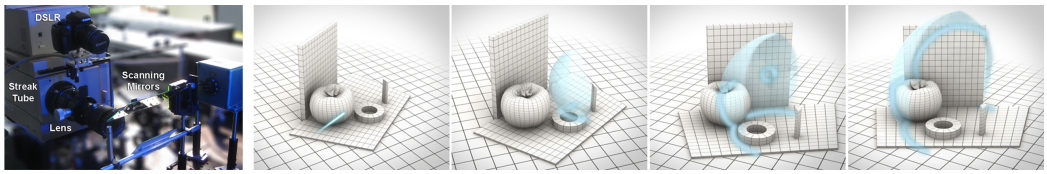


Figure 8.3: Left: Photograph of our ultrafast imaging system setup. The DSLR camera takes a conventional photo for comparison. Right: Time sequence illustrating the arrival of the pulse striking a diffuser, its transformation into a spherical energy front, and its propagation through the scene. The corresponding captured scene is shown in Figure 8.10 (top row).

have also been used as line scanning devices for image transmission through highly scattering turbid media, by recording the ballistic photons, which travel a straight path through the scatterer and thus arrive first on the sensor [161]. The principles that we develop in this paper for the purpose of transient imaging were first demonstrated by Velten et al. [458]. Recently, photonic mixer devices, along with nonlinear optimization, have also been used in this context [164].

Our system can record and reconstruct space-time world information of incoherent light propagation in free-space, table-top scenes, at a resolution of up to 672×1000 pixels and under 2 picoseconds per frame. The varied range and complexity of the scenes we capture allow us to visualize the *dynamics* of global illumination effects, such as scattering, specular reflections, interreflections, subsurface scattering, caustics, and diffraction.

TIME-RESOLVED IMAGING Recent advances in time-resolved imaging have been exploited to recover geometry and motion around corners [358, 229, 457, 456, 148, 340] and albedo of from single view point [326]. But, none of them explored the idea of capturing videos of light in motion in direct view and have some fundamental limitations (such as capturing only third-bounce light) that make them unsuitable for the present purpose. Wu et al. [488] separate direct and global illumination components from time-resolved data captured with the system we describe in this paper, by analyzing the time profile of each pixel. In a recent publication [489], the authors present an analysis on transient light transport in frequency space, and show how it can be applied to bare-sensor imaging.

8.3 CAPTURING SPACE-TIME PLANES

We capture time scales orders of magnitude faster than the exposure times of conventional cameras, in which photons reaching the sensor at different times are integrated into a single value, making it impossible to observe ultrafast optical phenomena. The system described in this paper has an effective exposure time down to 1.85 ps; since light travels at 0.3 mm/ps, light travels approximately 0.5 mm between frames in our reconstructed movies.

SYSTEM: An ultrafast setup must overcome several difficulties in order to accurately measure a high-resolution (both in space and time) image. First, for an unamplified laser pulse, a single exposure time of less than 2 ps would not collect enough light, so the SNR would be unworkably low. As an example, for a table-top scene illuminated by a 100 W bulb, only about 1 photon on average would reach the sensor during a 2 ps open-shutter period. Second, because of the time scales involved, synchronization of the sensor and the illumination must be executed within picosecond precision. Third, standalone streak sensors sacrifice the vertical spatial dimension in order to code the time dimension,

thus producing x-t images. As a consequence, their field of view is reduced to a single horizontal line of view of the scene.

We solve these problems with our ultrafast imaging system, outlined in Figure 8.2. (A photograph of the actual setup is shown in Figure 8.3 (left)). The light source is a femtosecond (fs) Kerr lens mode-locked Ti:Sapphire laser, which emits 50-fs with a center wavelength of 795 nm, at a repetition rate of 75 MHz and average power of 500 mW. In order to see ultrafast events in a scene with macro-scaled objects, we focus the light with a lens onto a Lambertian diffuser, which then acts as a point light source and illuminates the entire scene with a spherically-shaped pulse (see Figure 8.3 (right)). Alternatively, if we want to observe pulse propagation itself, rather than the interactions with large objects, we direct the laser beam across the field of view of the camera through a scattering medium (see the *bottle* scene in Figure 8.1).

Because all the pulses are statistically identical, we can record the scattered light from many of them and integrate the measurements to average out any noise. The result is a signal with a high SNR. To synchronize this illumination with the streak sensor (Hamamatsu C5680 [156]), we split off a portion of the beam with a glass slide and direct it onto a fast photodetector connected to the sensor, so that, now, both detector and illumination operate synchronously (see Figure 8.2 (a)).

CAPTURING SPACE-TIME PLANES: The streak sensor then captures an x-t image of a certain scanline (i.e. a line of pixels in the horizontal dimension) of the scene with a space-time resolution of 672×512. The exact time resolution depends on the amplification of an internal sweep voltage signal applied to the streak sensor. With our hardware, it can be adjusted from 0.30 ps to 5.07 ps. Practically, we choose the fastest resolution that still allows for capture of the entire duration of the event. In the streak sensor, a photocathode converts incoming photons, arriving from each spatial location in the scanline, into electrons. The streak sensor generates the x-t image by deflecting these electrons, according to the time of their arrival, to different positions along the t-dimension of the sensor (see Figure 8.2(b) and 8.2(c)). This is achieved by means of rapidly changing the sweep voltage between the electrodes in the sensor. For each horizontal scanline, the camera records a scene illuminated by the pulse and averages the light scattered by 4.5×10^8 pulses (see Figure 8.2(d) and 8.2(e)).

PERFORMANCE VALIDATION To characterize the streak sensor, we compare sensor measurements with known geometry and verify the linearity, reproducibility, and calibration of the time measurements. To do this, we first capture a streak image of a scanline of a simple scene: a plane being illuminated by the laser after hitting the diffuser (see Figure 8.4 (left)). Then, by using a Faro digitizer arm [127], we obtain the ground truth geometry of the points along that plane and of the point of the diffuser hit by the laser; this allows us to compute the total travel time per path (diffuser-plane-streak sensor) for each pixel in the scanline. We then compare the travel time captured by our streak sensor with the real travel time computed from the known geometry. The graph in Figure 8.4 (right) shows agreement between the measurement and calculation.

8.4 CAPTURING SPACE-TIME VOLUMES

Although the synchronized, pulsed measurements overcome SNR issues, the streak sensor still provides only a one-dimensional movie. Extension to two dimensions requires unfeasible bandwidths: a typical dimension is roughly 10^3 pixels, so a three-dimensional data cube has 10^9 elements. Recording such a large quantity in a 10^{-9} second (1 ns) time widow requires a bandwidth of 10^{18} byte/s, far beyond typical available bandwidths.

We solve this acquisition problem by again utilizing the synchronized repeatability of the hardware: A mirror-scanning system (two 9 cm $\times$ 13 cm mirrors, see Figure 8.3 (left))

rotates the camera's center of projection, so that it records horizontal slices of a scene sequentially. We use a computer-controlled, one-rpm servo motor to rotate one of the mirrors and consequently scan the field of view vertically. The scenes are about 25 cm wide and placed about 1 meter from the camera. With high gear ratios (up to 1:1000), the continuous rotation of the mirror is slow enough to allow the camera to record each line for about six seconds, requiring about one hour for 600 lines (our video resolution). We generally capture extra lines, above and below the scene (up to 1000 lines), and then crop them to match the aspect ratio of the physical scenes before the movie was reconstructed.

These resulting images are combined into one matrix, M_{ijk}, where $i = 1...672$ and $k = 1...512$ are the dimensions of the individual x-t streak images, and $j = 1...1000$ addresses the second spatial dimension y. For a given time instant k, the submatrix N_{ij} contains a two-dimensional image of the scene with a resolution of 672×1000 pixels, exposed for as short to 1.85 ps. Combining the x-t slices of the scene for each scanline yields a 3D x-y-t data volume, as shown in Figure 8.5 (left). An x-y slice represents one frame of the final movie, as shown in Figure 8.5 (right).

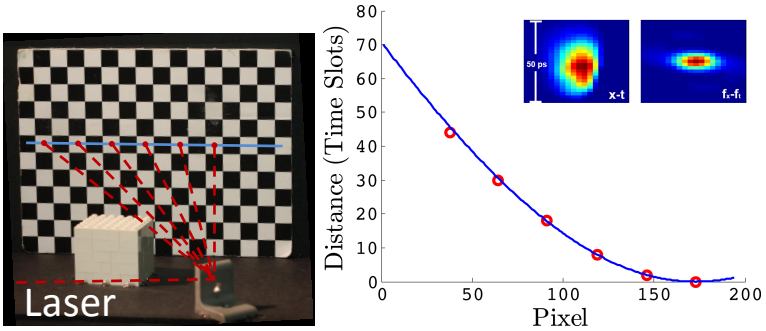


Figure 8.4: Performance validation of our system. Left: Measurement setup used to validate the data. We use a single streak image representing a line of the scene and consider the centers of the white patches because they are easily identified in the data. Right: Graph showing pixel position vs. total path travel time captured by the streak sensor (red) and calculated from measurements of the checkerboard plane position with a Faro digitizer arm (blue). Inset: PSF, and its Fourier transform, of our system.

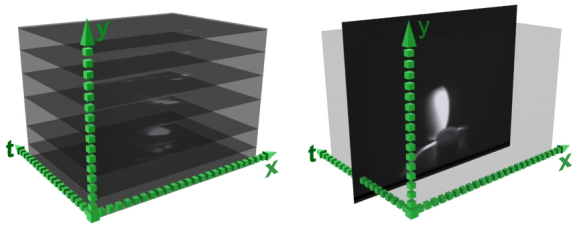


Figure 8.5: Left: Reconstructed x-y-t data volume by stacking individual x-t images (captured with the scanning mirrors). Right: An x-y slice of the data cube represents one frame of the final movie.

8.5 DEPICTING ULTRAFAST VIDEOS IN 2D

We have explored several ways to visualize the information contained in the captured x-y-t data cube in an intuitive way. First, contiguous N_{ij} slices can be played as the frames of a movie. Figure 8.1 (bottom row) shows a captured scene (*bottle*) along with several representative N_{ij} frames. (Effects are described for various scenes in Section 7.)

However, understanding all the phenomena shown in a video is not a trivial task, and movies composed of x-y frames such as the ones shown in Figure 8.10 may be hard to interpret. Merging a static photograph of the scene from approximately the same point of view with the N_{ij} slices aids in the understanding of light transport in the scenes (see movies within the video at `http://webdiis.unizar.es/~bmasia/downloads/thesis/Femto-Main_Video.mp4`). Although straightforward to implement, the high dynamic range of the streak data requires a nonlinear intensity transformation to extract subtle optical effects in the presence of high intensity reflections. We employ a logarithmic transformation to this end.

We have also explored single-image methods for intuitive visualization of full space-time propagation, such as the color-coding in Figure 8.1 (right), which we describe in the following paragraphs.

INTEGRAL PHOTO FUSION By integrating all the frames in novel ways, we can visualize and highlight different aspects of the light flow in one photo. Our photo fusion results are calculated as $N_{ij} = \sum w_k M_{ijk}, \{k = 1..512\}$, where w_k is a weighting factor determined by the particular fusion method. We have tested several different methods, of which two were found to yield the most intuitive results: the first one is *full fusion*, where $w_k = 1$ for all k. Summing all frames of the movie provides something resembling a black and white photograph of the scene illuminated by the laser, while showing time-resolved light transport effects. An example is shown in Figure 8.6 (left) for the *alien* scene. (More information about the scene is given in Section 8.7.) A second technique, *rainbow fusion*, takes the fusion result and assigns a different RGB color to each frame, effectively color-coding the temporal dimension. An example is shown in Figure 8.6 (middle).

PEAK TIME IMAGES The inherent integration in fusion methods, though often useful, can fail to reveal the most complex or subtle behavior of light. As an alternative, we propose peak time images, which illustrate the time evolution of the *maximum* intensity in each frame. For each spatial position (i, j) in the x-y-t volume, we find the peak intensity along the time dimension, and keep information within two time units to each side of the peak. All other values in the streak image are set to zero, yielding a more sparse space-time volume. We then color-code time and sum up the x-y frames in this new sparse volume, in the same manner as in the *rainbow fusion* case but use only every 20th frame in the sum to create black lines between the equi-time paths, or isochrones. This results in a map of the propagation of maximum intensity contours, which we term *peak time image*. These color-coded isochronous lines can be thought of intuitively as propagating energy fronts. Figure 8.6 (right) shows the peak time image for the *alien* scene, and Figure 8.1 (top, middle) shows the captured data for the *bottle* scene depicted using this visualization method. As explained in the next section, this visualization of the bottle scene reveals significant light transport phenomena that could not be seen with the rainbow fusion visualization.

8.6 TIME UNWARPING

Visualization of the captured movies (Sections 8.5 and 8.7) reveals results that are counter-intuitive to theoretical and established knowledge of light transport. Figure 8.1 (top, middle) shows a peak time visualization of the *bottle* scene, where several abnormal light transport effects can be observed: (1) the caustics on the floor, which propagate towards the bottle, instead of away from it; (2) the curved spherical energy fronts in the label area, which should be rectilinear as seen from the camera; and (3) the pulse itself being located behind these energy fronts, when it would need to precede them. These are due to the fact that usually light propagation is assumed to be infinitely fast, so that

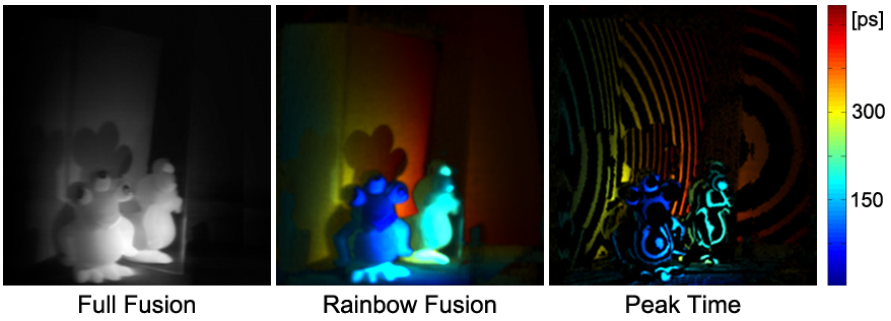


Figure 8.6: Three visualization methods for the *alien* scene. From left to right, more sophisticated methods provide more information and an easier interpretation of light transport in the scene.

events in world space are assumed to be detected simultaneously in camera space. In our ultrafast photography setup, however, this assumption no longer holds, and the finite speed of light becomes a factor: we must now take into account the time delay between the occurrence of an event and its detection by the camera sensor.

We therefore need to consider two different time frames, namely *world time* (when events happen) and *camera time* (when events are detected). This duality of time frames is explained in Figure 8.7: light from a source hits a surface first at point $P_1 = (i_1, j_1)$ (with (i, j) being the x-y pixel coordinates of a scene point in the x-y-t data cube), then at the farther point $P_2 = (i_2, j_2)$, but the reflected light is captured in the reverse order by the sensor, due to different total path lengths ($z_1 + d_1 > z_2 + d_2$). Generally, this is due to the fact that, for light to arrive at a given time instant t_0, all the rays from the source, to the wall, to the camera, must satisfy $z_i + d_i = ct_0$, so that isochrones are elliptical. Therefore, although objects closer to the source receive light earlier, they can still lie on a higher-valued (later-time) isochrone than farther ones.

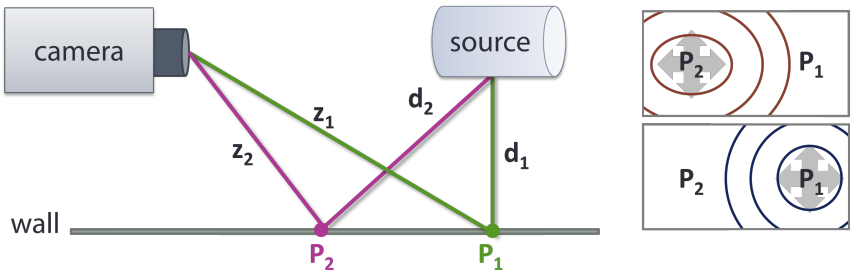


Figure 8.7: Understanding reversal of events in captured videos. *Left:* Pulsed light scatters from a source, strikes a surface (e.g., at P_1 and P_2), and is then recorded by a sensor. Time taken by light to travel distances $z_1 + d_1$ and $z_2 + d_2$ is responsible for the existence of two different time frames and the need of computational correction to visualize the captured data in the world time frame. *Right:* Light appears to be propagating from P_2 to P_1 in camera time (before unwarping), and from P_1 to P_2 in world time, once time-unwarped. Extended, planar surfaces will intersect constant-time paths to produce either elliptical or circular fronts.

In order to visualize all light transport events as they have occurred (not as the camera captured them), we transform the captured data from camera time to world time, a

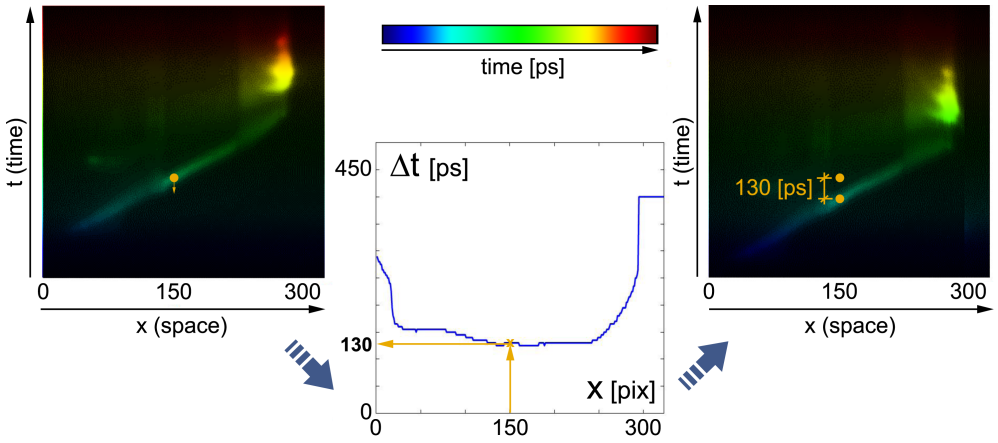


Figure 8.8: Time unwarping in 1D for a streak image (x-t slice). Left: captured streak image; shifting the time profile down in the temporal dimension by Δt allows for the correction of path length delay to transform between time frames. Center: the graph shows, for each spatial location x_i of the streak image, the amount Δt_i that point has to be shifted in the time dimension of the streak image. Right: resulting time-unwarped streak image.

transformation which we term *time unwarping*. Mathematically, for a scene point $P = (i, j)$, we apply the following transformation:

$$t'_{ij} = t_{ij} + \frac{z_{ij}}{c/\eta} \tag{43}$$

where t'_{ij} and t_{ij} represent camera and world times respectively, c is the speed of light in vacuum, η the index of refraction of the medium, and z_{ij} is the distance from point P to the camera. For our table-top scenes, we measure this distance with a Faro digitizer arm, although it could be obtained from the data and the known position of the diffuser, as the problem is analogous to that of bi-static LiDAR. We can thus define light travel time from each point (i, j) in the scene to the camera as $\Delta t_{ij} = t'_{ij} - t_{ij} = z_{ij}/(c/\eta)$. Then, time unwarping effectively corresponds to offsetting data in the x-y-t volume along the time dimension, according to the value of Δt_{ij} for each of the (i, j) points, as shown in Figure 8.8.

In most of the scenes, we only have propagation of light through air, for which we take $\eta \approx 1$. For the *bottle* scene, we assume that the laser pulse travels along its longitudinal axis at the speed of light, and that only a single scattering event occurs in the liquid inside. We take $\eta = 1.33$ as the index of refraction of the liquid and ignore refraction at the bottle's surface. A step-by-step unwarping process is shown in Figure 8.9 for a frame (i.e. x-y image) of the *bottle* scene. Our unoptimized Matlab code runs at about 0.1 seconds per frame. A time-unwarped peak-time visualization of the whole of this scene is shown in Figure 8.1 (right). Notice how now the caustics originate from the bottle and propagate outward, energy fronts along the label are correctly depicted as straight lines, and the pulse precedes related phenomena, as expected.

8.7 CAPTURED SCENES

We have used our ultrafast photography setup to capture interesting light transport effects in different scenes. Figure 8.10 summarizes them, showing representative frames and peak time visualizations. The exposure time for our scenes is between 1.85 ps for the *crystal* scene, and 5.07 ps for the *bottle* and *tank* scenes, which required imaging a longer time span for better visualization. Please refer to the video at `http://`

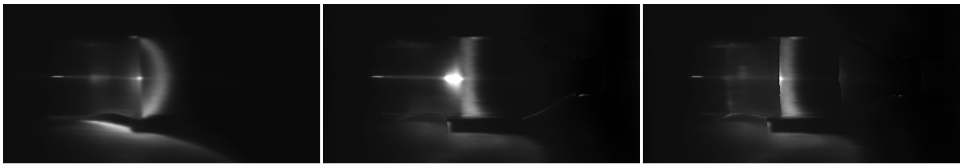


Figure 8.9: Time unwarping for the *bottle* scene, containing a scattering medium. From left to right: a frame of the video without correction, where the energy front appears curved; the same frame after time-unwarping with respect to distance to the camera z_{ij}; the shape of the energy front is now correct, but it still appears before the pulse; the same frame, time-unwarped taking also scattering into account.

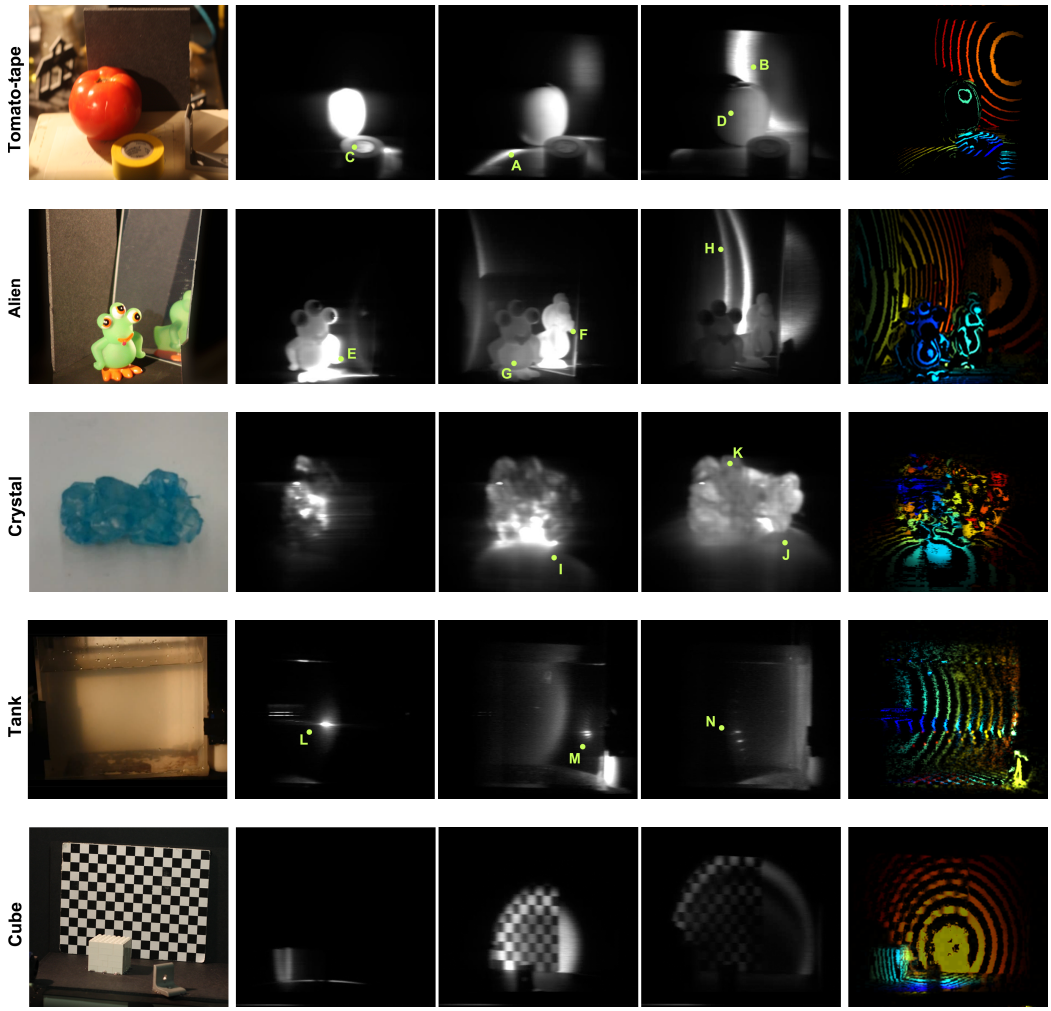


Figure 8.10: More scenes captured with our setup (refer to Figure 8.1 for the *bottle* scene). For each scene, from left to right: photograph of the scene (taken with a DSLR camera), a series of representative frames of the reconstructed movie, and peak time visualization of the data. Please refer to the video at `http://webdiis.unizar.es/~bmasia/downloads/thesis/Femto-Main_Video.mp4` for the full movies. Note that the viewpoint varies slightly between the DSLR and the streak sensor.

webdiis.unizar.es/~bmasia/downloads/thesis/Femto-Main_Video.mp4 to watch the reconstructed movies. Overall, observing light in such slow motion reveals both subtle and key aspects of light transport. We provide here brief descriptions of the light transport effects captured in the different scenes.

BOTTLE This scene is shown in Figure 8.1 (bottom row), and has been used to introduce time-unwarping. A plastic bottle, filled with water diluted with milk, is directly illuminated by the laser pulse, entering through the bottom of the bottle along its longitudinal axis. The pulse scatters inside the liquid; we can see the propagation of the wavefronts. The geometry of the bottle neck creates some interesting lens effects, making light look almost like a fluid. Most of the light is reflected back from the cap, while some is transmitted or trapped in subsurface scattering phenomena. Caustics are generated on the table.

TOMATO-TAPE This scene shows a tomato and a tape roll, with a wall behind them. The propagation of the spherical wavefront, after the laser pulse hits the diffuser, can be seen clearly as it intersects the floor and the back wall (A, B). The inside of the tape roll is out of the line of sight of the light source and is not directly illuminated. It is illuminated later, as indirect light scattered from the first wave reaches it (C). Shadows become visible only after the object has been illuminated. The more opaque tape darkens quickly after the light front has passed, while the tomato continues glowing for a longer time, indicative of stronger subsurface scattering (D).

ALIEN A toy alien is positioned in front of a mirror and wall. Light interactions in this scene are extremely rich, due to the mirror, the multiple interreflections, and the subsurface scattering in the toy. The video shows how the reflection in the mirror is actually formed: direct light first reaches the toy, but the mirror is still completely dark (E); eventually light leaving the toy reaches the mirror, and the reflection is dynamically formed (F). Subsurface scattering is clearly present in the toy (G), while multiple direct and indirect interactions between the wall and the mirror can also be seen (H).

CRYSTAL A group of sugar crystals is directly illuminated by the laser from the left, acting as multiple lenses and creating caustics on the table (I). Part of the light refracted on the table is reflected back to the candy, creating secondary caustics on the table (J). Additionally, scattering events are visible within the crystals (K).

TANK A reflective grating is placed at the right side of a tank filled with milk diluted in water. The grating is taken from a commercial spectrometer, and consists of an array of small, equally spaced rectangular mirrors. The grating is blazed: mirrors are tilted to concentrate maximum optical power in the first order diffraction for one wavelength. The pulse enters the scene from the left, travels through the tank (L), and strikes the grating. The grating reflects and diffracts the beam pulse (M). The different orders of the diffraction are visible traveling back through the tank (N). As the figure (and the video in http://webdiis.unizar.es/~bmasia/downloads/thesis/Femto-Main_Video.mp4) shows, most of the light reflected from the grating propagates at the blaze angle.

cube A very simple scene composed of a cube in front of a wall with a checkerboard pattern. The simple geometry allows for a clear visualization and understanding of the propagation of wavefronts.

8.8 CONCLUSIONS AND FUTURE WORK

Our research fosters new computational imaging and image processing opportunities by providing incoherent time-resolved information at ultrafast temporal resolutions. We hope our work will inspire new research in computer graphics and computational photography, by enabling forward and inverse analysis of light transport, allowing for full scene capture of hidden geometry and materials, or for relighting photographs. To this end, captured movies and data of the scenes shown in this paper are available at `femtocamera.info`. This exploitation, in turn, may influence the rapidly emerging field of ultrafast imaging hardware.

The system could be extended to image in color by adding additional pulsed laser sources at different colors or by using one continuously tunable optical parametric oscillator (OPO). A second color of about 400 nm could easily be added to the existing system by doubling the laser frequency with a nonlinear crystal (about $1000). The streak tube is sensitive across the entire visible spectrum, with a peak sensitivity at about 450 nm (about five times the sensitivity at 800 nm). Scaling to bigger scenes would require less time resolution and could therefore simplify the imaging setup. Scaling should be possible without signal degradation, as long as the camera aperture and lens are scaled with the rest of the setup. If the aperture stays the same, the light intensity needs to be increased quadratically to obtain similar results.

Beyond the ability of the commercially available streak sensor, advances in optics, material science, and compressive sensing may bring further optimization of the system, which could yield increased resolution of the captured x-t streak images. Nonlinear shutters may provide an alternate path to femto-photography capture systems. However, nonlinear optical methods require exotic materials and strong light intensities that can damage the objects of interest (and must be provided by laser light). Further, they often suffer from physical instabilities.

We believe that mass production of streak sensors can lead to affordable systems. Also, future designs may overcome the current limitations of our prototype regarding optical efficiency. Future research can investigate other ultrafast phenomena such as propagation of light in anisotropic media and photonic crystals, or may be used in applications such as scientific visualization (to understand ultra-fast processes), medicine (to reconstruct subsurface elements), material engineering (to analyze material properties), or quality control (to detect faults in structures). This could provide radically new challenges in the realm of computer graphics. Graphics research can enable new insights via comprehensible simulations and new data structures to render light in motion. For instance, relativistic rendering techniques have been developed using our data, where the common assumption of constant irradiance over the surfaces does no longer hold [201]. It may also allow a better understanding of scattering, and may lead to new physically valid models, as well as spawn new art forms.

9

RELATIVISTIC RENDERING FOR TRANSIENT IMAGING

ABOUT THIS CHAPTER

In this chapter we present a method for accurate visualization of time-resolved data. This project originated from the work described in Chapter 8, since we wanted to correctly visualize the data as the camera or observer moved throughout the scene. The core rendering part of this project has been carried out by Adrián Jarabo, whereas my contribution has been on assisting on the research of relativity equations (transformations undergone by observed radiance when moving at relativistic speeds), relativistic rotation, and the adaptation of these to our particular scenario with time-resolved data. This work was presented at the Spanish Conference in Computer Graphics (CEIG) 2013, selected as Best Paper (one of the two papers which were selected as such) and invited to submit an extended version to Computer Graphics Forum, on which we are working at the date of publication of this thesis.

A. Jarabo, B. Masia, A. Velten, C. Barsi, R. Raskar and D. Gutierrez
Rendering Relativistic Effects in Transient Imaging.
In Proc. of CEIG 2013. Selected as Best Paper (1 in 2).

9.1 INTRODUCTION

Analyzing and synthesizing light transport is a core research topic in computer graphics, computing vision and scientific imaging [150]. One of the most common simplifications, rarely challenged, is the assumption that the speed of light is infinite. While this is a valid assumption in most cases, it is certainly not true: light travels extremely fast, but with finite speed. In this part (Part V) of the thesis, we have lifted this assumption and explored the consequences of dealing with time-resolved data (finite speed of light), in this chapter focusing on the relativistic effects that occur when the camera moves at speeds comparable with the speed of light.

Relativistic rendering is not new [71, 474]. However, our time-resolved framework implies by definition that surface irradiance is not constant in the temporal domain, so existing models must be revised and redefined. We describe here our technique to render and inspect scenes where relativistic effects take place: in particular, we address time dilation, light aberration, the Doppler effect and the searchlight effect. Moreover, no existing model of relativistic rotation exists in the literature, which hinders free exploration of scenes; we additionally introduce the first model of relativistic sensor rotation in computer graphics.

To obtain input data, we rely on two sources of information. One the one hand, real-world captured data from femto-photography (Chapter 8 and [459]), which we leverage using image-based rendering techniques. Since the camera cannot be moved in the setup, our technique allows to visualize novel view points, synthesizing light transport in a physically accurate manner. On the other hand, we also employ the transient renderer by Jarabo et al. [201], which allows us to create novel scenes and render simulations of time-resolved light transport. Both approaches can help gain a deeper understanding of light transport at picosecond scale.

In summary, we have developed a rendering and visualization tool for transient light transport, capable of simulating generalized relativistic effects, freed from the restrictions of previous works. Our contributions can be summarized as follows:

- We revise and correct well-established concepts about relativistic rendering, to take into account that irradiance can no longer be assumed to be constant over time
- Previous techniques were also limited by linear velocities of the (virtual) cameras. We propose the first approximate solution for the case of a *rotating* sensor, so the camera can be freely moved in 3D space
- We implement a fully working prototype, which allows interactive visualization and exploration of both real and simulated data

9.2 RELATED WORK

A modified rendering equation can account for the finite speed of light and handle transient effects [21, 411]. However, in previous works no practical rendering framework is derived from the proposed transient rendering framework. A fully functional time-resolved rendering system was recently presented by Jarabo and colleagues [201]; part of the data employed in this chapter (in particular the *bunny* scene) has been generated by that renderer. In addition to the related work described here, we refer the reader to Section 8.2, where the work on time-resolved light transport is compiled.

With regard to relativistic rendering, here we discuss the most relevant work on the field. For a wider survey, we refer to [474], where the different proposed techniques for both general and special relativistic rendering are discussed, including their application as educational tools. Chang et al. [71] introduced the theory of Special Relativity in the field of computer graphics. Their work accounts for geometric and radiance transformations due to fast moving objects or camera. However, their formulation modeled the searchlight and Doppler effects incorrectly; these were later corrected by Weiskopf et al. [472]. Following work [473] simulates relativistic effects in real captured scenes modeled with image-based techniques, by applying the relativistic transformations directly on the light field. However, the authors assume light incoming from infinitely far away light sources with constant radiance, so both the effects of distance and time-varying irradiance are ignored. This allows them to make some simplifying assumptions about the radiance in the scene, which no longer hold in the context of time-resolved data we deal with. Finally, visualization approaches and games have been created with a didactic goal, aiming at helping students in the understanding of relativity. The game *A Slower Speed of Light*, notable among these, uses the open-source toolkit *OpenRelativity* which works with the *Unity* engine and can simulate special relativity effects [239]. However, to our knowledge, they do not deal with time-varying irradiance either.

9.3 RELATIVISTIC RENDERING

Time-resolved data, like that captured with the setup described in Chapter 8, allows us to explore light transport like never before, no longer being constrained by the assumption that light speed is infinite. While this is indeed a valid assumption in most cases, the possibilities that open up analyzing the dynamics of light at pico-second resolution are fascinating.

9.3.1 *Frames of Reference*

Assuming that the geometry in the scene is known (which can be easily acquired with a digitizer arm or from time-of-flight data), we can synthesize new viewpoints and animations of the scene by taking an image-based rendering approach, using x-y textures from the x-y-t data cube and projecting them onto the geometry. This allows us to visualize real-world events from new, interesting angles. However, visualizing light transport

events at this time scale yields counter-intuitive results: Events are not captured in the sensor as they occur, which leads to unexpected apparent distortions in the propagation of light. This effect, which has been explained in Section 8.6 of this thesis, is termed *time warping*. Due to it, two different temporal frames of reference must be employed: one for the world (when the events occur) and one for the camera (when the events are actually captured).

As a consequence, sensor data acquired by the femto-photography technique appears *warped* in the temporal domain, and must be *time-unwarped* to take into account the finite speed of light. So for each frame in the synthesized animations, we access the original warped data and apply the following transformation shown in Equation 43, that we reproduce here again for clarity [459]:

$$t'_{ij} = t_{ij} + \frac{z_{ij}}{c/\eta}$$

where t'_{ij} and t_{ij} are camera and world times respectively, z_{ij} is the depth from each point (i, j) to the new camera position, and η the index of refraction of the medium. Note how a naive approach based on simply sticking the textures from the first frame to the geometry through the animation would produce wrong results; the distance from each geometry point to the center of projection of the camera varies for each frame, and thus a *different* transformation must be applied each time to the original, *warped* x-y-t data (see Figure 9.1). We assume a pinhole model for the camera.

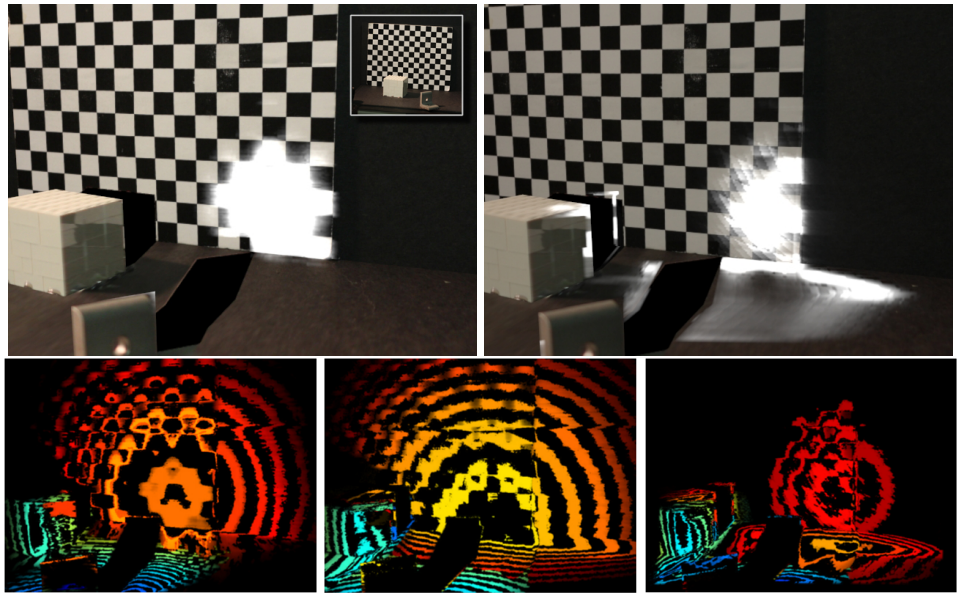


Figure 9.1: Time unwarping between camera time and world time for synthesized new views of a cube scene. Top row, left: Scene rendered from a novel view keeping the unwarped camera time from the first frame (the small inset shows the original viewpoint). Right: The same view, warping data according to the new camera position. Notice the large changes in light propagation, in particular the wavefronts on the floor not visible in the previous image. Bottom row: Isochrones visualization of the cube-scene for a given virtual camera (color encodes time); from left to right: original x-y-t volume in the time-frame of the capturing camera, unwarped x-y-t data in world time frame, and re-warped data for the new virtual camera. Note the striking differences between corresponding isochrones.

9.3.2 *Relativistic Effects*

Apart from the time-warping of data, macroscopic camera movement at pico-second time scales, like the one synthesized in Figure 9.1 would give rise to relativistic effects. This requires a relativistic framework to correctly represent and visualize light traveling through the 3D scene. Although simulations of relativistic effects have existed for a while [71, 474], visualizing our particular time-resolved datasets requires departing from the common simplifying assumption of constant irradiance on surfaces. As we will see in the following paragraphs, this has direct implications on how radiance gets imaged onto the sensor.

According to special relativity, light aberration, the Doppler effect, and the searchlight effect need to be taken into account when simulating motion at fast speeds. *Light aberration* accounts for the apparent geometry deformation caused by two space-time events measured in two reference frames moving at relativistic speeds with respect to each other. The *Doppler effect* produces a wavelength shift given by the Doppler factor. Last, the *searchlight effect* increases or decreases radiance, according to whether the observer is approaching or moving away from a scene. We modify existing models for the three effects to support time-resolved irradiance, and approximate the yet-unsolved solution for camera rotation.

We build our relativistic visualization framework on the derivations by Weiskopf et al. [472]. We consider two inertial frames, O and O', where O' (the sensor) is moving with velocity $v = \beta c$ with respect to O, with $\beta \in [0..\pm 1)$. L represents radiance measured in O, defined by direction (θ, ϕ) (defined with respect to the motion direction) and wavelength λ. The corresponding primed variables (θ', ϕ') and λ' define radiance L' measured in O'. To obtain the modified radiance L' given L and the speed of the sensor, we need to apply the following equation:

$$L'(\theta', \phi', \lambda') = D^{-5} L\left(\arccos\frac{\cos\theta' + \beta}{1 + \beta\cos\theta'}, \phi', \frac{\lambda'}{D}\right) \tag{44}$$

where $D = \gamma(1 + \beta\cos\theta')$ and $\gamma = 1/\sqrt{1 - \beta^2}$. This equation accounts for all three factors: light aberration, the Doppler effect, and the searchlight effect. However, it cannot model explicitly the effect of special relativity on time-resolved irradiance. In the following paragraphs we explain each effect separately, and discuss the modifications needed to handle time-resolved irradiance.

Time dilation: Breaking the assumption of constant irradiance means that we cannot ignore the effect of time dilation [118]. Time dilation relates directly with Lorentz contraction, and is defined as the difference in elapsed time Δt between two events observed in different inertial frames; for our world and camera frames of reference, this translates into $\Delta t' = \gamma \Delta t$. This means that time in these two frames advances at different speeds, making time in the stationary frame (the world) advance faster than in the moving frame (the camera). Thus, we need to keep track of both world t and camera time t', since they differ depending on the motion speed.

Light aberration: An easy example to understand light aberration is to visualize how we see rain drops when traveling on a speeding train. When the train is not moving, raindrops fall vertically; but as the train picks up speed, raindrop trajectories become increasingly diagonal as a function of the train's speed. This is because the speed of the train is comparable with the speed of raindrops. A similar phenomenon occurs with light if moving at relativistic speeds. However, as opposed to rain drops, relativistic light aberration cannot be modeled with classical physics aberration; the Lorentz transformation needs to be applied instead.

Light aberration is computed by transforming θ' and ϕ' with the following equations, which provide the geometric transformation between two space-time events measured in two reference frames which move at relativistic speeds with respect to each other:

$$\cos\theta' = \frac{\cos\theta - \beta}{1 - \beta\cos\theta} \tag{45}$$

$$\phi' = \phi \tag{46}$$

The end result is that light rays appear curved, with more curvature as velocity increases. Given this curvature, light rays reaching the sensor from behind the camera become visible. Finally, as β approaches 1, and thus $v \approx c$, most incoming light rays are compressed towards the motion direction; this makes the scene collapse into a single point as the camera moves towards it (note that this produces the wrong impression that the camera is moving *away* from the scene). The first two rows in Figure 9.2 show the effects of light aberration with increasing velocity as the sensor moves at relativistic speeds, towards and away from the scene respectively.

Doppler effect: The Doppler effect is better known for sound, and it is not a phenomenon restricted to relativistic velocities. In our case, the Doppler effect alters the observed frequency of the captured events in the world when seen by a fast-moving camera, which produces a wavelength shift, as defined by the Doppler factor D:

$$\lambda' = D\lambda \tag{47}$$

The overall result is a color shift as a function of the velocity of the sensor relative to the scene. Somewhat less known, the Doppler effect also creates a perceived speed-up (or down, depending on the direction of camera motion) of the captured events. This means that the *frame rate* of the time-varying irradiance f in world frame is Doppler shifted, making the perceived frame rate f' in camera frame become $f' = f/D$. Figure 9.2 (third row) shows an example of the Doppler effect.

Searchlight effect: Due to the searchlight effect, photons from several instants are captured at the same time differential, in part as a cause of the Doppler shift on the camera's perceived frame rate. This results in increased (if the observer is approaching the scene) or decreased (if the observer is moving away) brightness (see Figure 9.2, bottom row):

$$L'(\theta', \phi', \lambda') = D^{-5} L(\theta, \phi, \lambda) \tag{48}$$

Intuitively, continuing with our previous rain analogy, it is similar to what occurs in a vehicle driving in the rain: the front windshield will accumulate more water than the rear windshield. For our time-varying streak-data, this means that irradiance from several *frames* in world time interval dt is integrated over the same camera differential time dt', such that $dt = dt'/D$. Note that the D^{-5} factor only is valid for the case in which the directions of the velocity vector v and the normal to the detector are parallel. We later show how to approximate a rotation of the sensor.

Finally, Figures 9.3 and 9.4 show the result of combining all these relativistic effects, both for the *cube* scene (data captured with femto-photography techniques) and the *bunny* scene (simulated data by rendering) respectively. The laser wavelength is set at 670 nm for visualization purposes. We refer the reader to the video at `http://webdiis.unizar.es/~bmasia/downloads/thesis/Relativistic_Cube.mov` for a full animation of the *cube* scene.

9.3.3 *Relativistic Rotation*

Providing free navigation of a scene depicting time-resolved light transport implies that the viewers should be allowed to rotate the camera. However, there is no universally accepted theory of relativistic rotation [374]. We propose a suitable approximation based

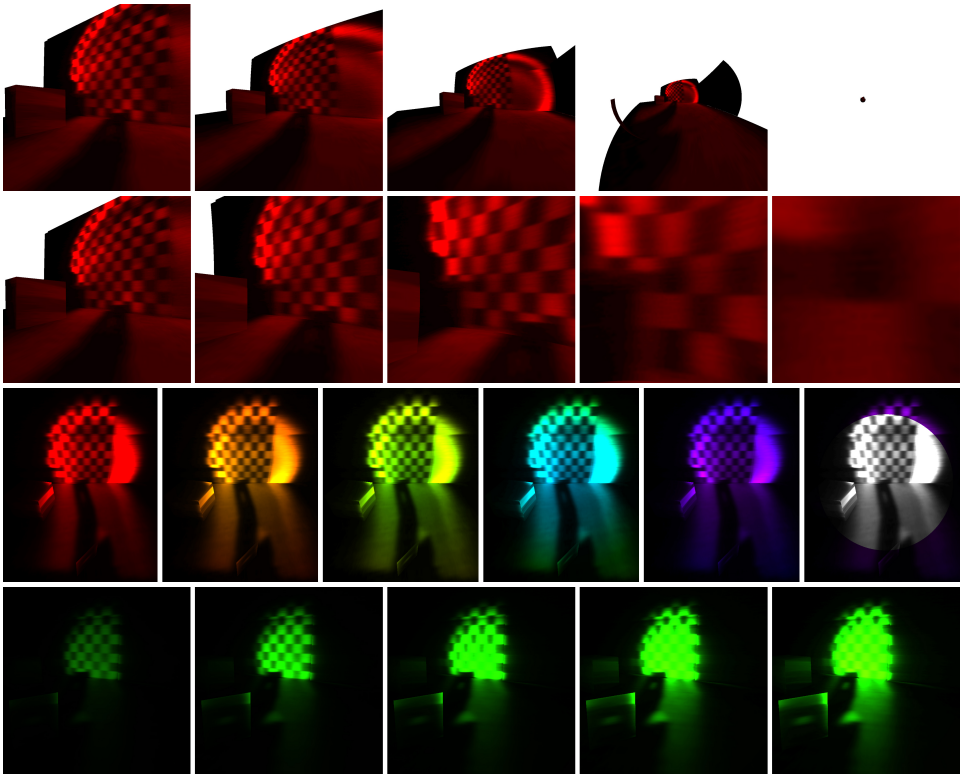


Figure 9.2: Relativistic effects shown separately for the *cube* scene. First row: Distortion due to light aberration as the camera moves towards the scene at different velocities, with $\beta = \{0, 0.3, 0.6, 0.9, 0.99\}$. We assume a laser wavelength of 670 nm for visualization purposes. Second row: The same effect as the sensor moves away from the scene, with the opposite velocity from the previous row. Notice how in both cases light aberration produces counter-intuitive results as the camera appears to be moving in the opposite direction. Third row: Doppler effect, showing the shift in color as a consequence of the frequency shift of light reaching the sensor, with $\beta = \{0, 0.15, 0.25, 0.35, 0.50, 0.55\}$. Fourth row: Searchlight effect, resulting in an apparent increase in brightness as the speed of the approaching camera increases, with $\beta = \{0, 0.2, 0.3, 0.4, 0.5\}$ (simulated laser at 508 nm). All images have been tone-mapped to avoid saturation.

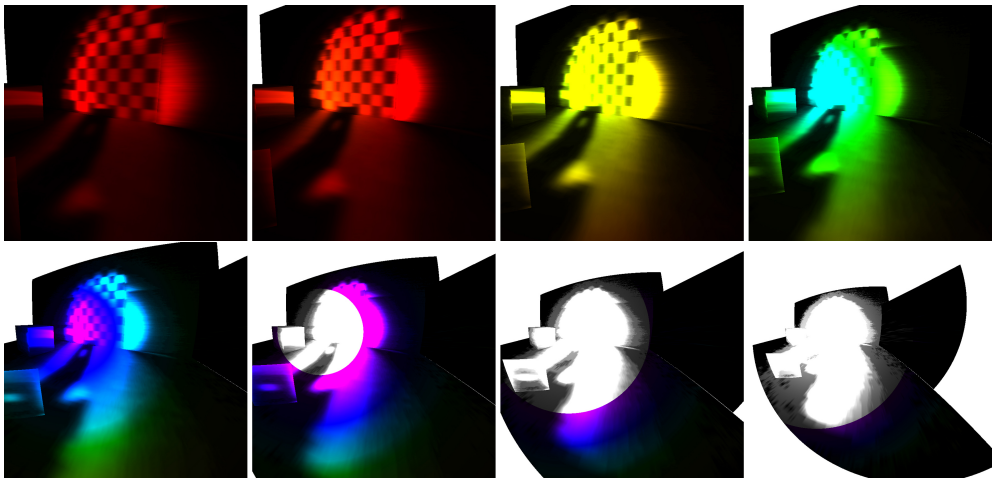


Figure 9.3: Relativistic phenomena for the *cube* scene (real data) including light aberration, Doppler effect and the searchlight effect, as the camera approaches the scene at increasing relativistic velocities $v = \beta c$ (with β increasing from 0 to 0.77).

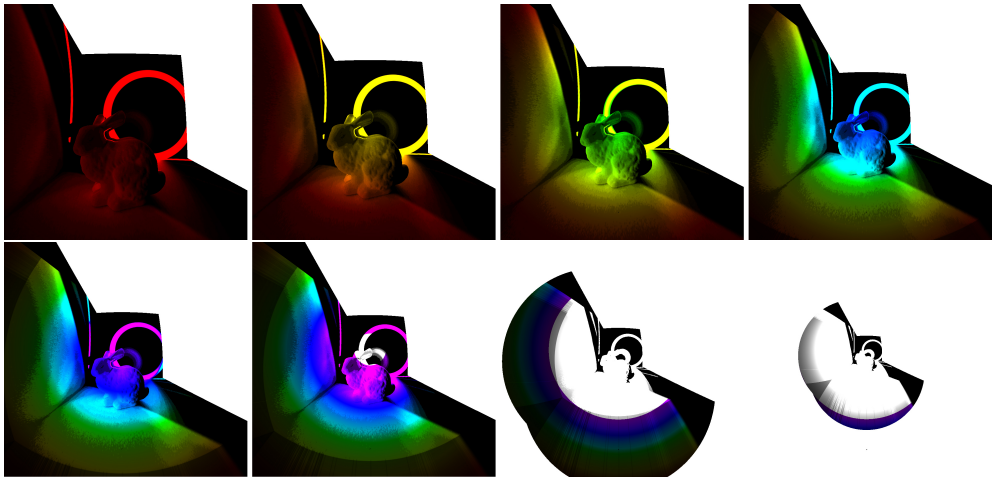


Figure 9.4: Relativistic phenomena for the *bunny* scene (simulated data) including light aberration, Doppler effect and the searchlight effect, as the camera approaches the scene at increasing relativistic velocities $v = \beta c$ (with β increasing from 0.2 to 0.9). Note that we transform the RGB computed radiance into luminance.

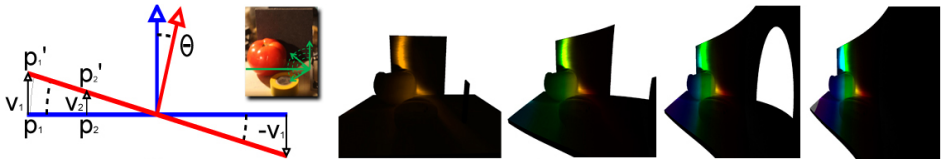


Figure 9.5: Relativistic rotation. Left: assuming that the rotation angle θ can be neglected between frames, we model the rotation as a continuous linear velocity field on the sensor Ψ, so each differential area is assigned a different velocity ψ_s. This causes that depending on the position on the sensor, different relativistic transformations are applied on the scene. The rest of the frames show the effects of a clockwise rotation of the sensor, with $\beta = \{0, 0.4, 0.8, 0.99\}$ (measured at the edge of the sensor). The small inset shows the original scene.

on limiting the rotation to very small angles per frame, so the differential rotation of the camera's viewing direction between frames can be neglected. However, for non-infinitesimal sensors this small rotation causes that the sensor's differential surfaces to move at different speeds: it creates a continuous linear velocity field Ψ on the sensor, with a zero-crossing at the axis of rotation.

To simulate the rotation of the camera we therefore first divide the sensor S in different areas $s \in S$. Our approximation effectively turns each of them into a different translational frame, with linear velocity ψ_s. Then, for each s we render the scene applying the novel relativistic transformations introduced in this section, with a different β_s for each s (trivially obtained from an input β measured at the edge of the sensor). This makes the incoming radiance be deformed differently depending on the position of the sensor where it is imaged. Figure 9.5 shows an example, where the sensor is rotating clockwise.

9.4 IMPLEMENTATION

Our implementation allows for real-time visualization of relativistic effects, both from real and simulated data. It is implemented in OpenGL as an stand-alone application, taking as input the reconstructed geometry of the scene, as well as the time-resolved data. The system is based on classic image-based rendering (IBR) techniques, where the shading of the surface is modeled by the images projected over the surface.

In our case, we use x-y images from the x-y-t data cube to shade the geometry. The cube is stored as a 3D texture on the GPU in *world time* coordinates. This allows us to apply time-warping to adapt it to the new viewpoint in rendering time, by simply applying the transformation defined in Equation 43 (see Section 9.3.2).

Due to light aberration the geometry viewed from the camera is distorted. This distortion causes straight lines to become curved, so the geometry has to be re-tessellated. Image-space warping, which has been used in many scenarios [73, 438, 307] and may appear as an alternative, is not viable in this scenario because of the large extent of the deformations, that make well-known problems of warping such as disocclusions clearly apparent. Our implementation performs the re-tessellation off-line on the CPU, but it is straightforward to tessellate it on the GPU on the fly. Then, in render time, each vertex should be transformed according to Equation 44.

Doppler effect is introduced by modifying the wavelength of the outgoing illumination from the surfaces. To avoid the complexity of a full-fledged spectral renderer, we assume light with energy in only one wavelength of the spectrum. To display radiance we use a simple *wavelenght-to-RGB* conversion encoded as a 1D texture. Wavelengths out of the visible spectrum are displayed as gray-scale values.

Finally, when modeling the searchlight effect, we avoid the straightforward approach to access all frames in the streak data cube, bounded by dt, and integrate them. This would require several accesses to the 3D texture, which would hinder interactivity. Instead, we pre-integrate irradiance values in the temporal domain, and use anisotropic mipmapping to access the pre-integrated irradiance values, using dt to select the mipmap level in the time dimension.

9.5 CONCLUSIONS AND FUTURE WORK

In this chapter we visualize light transport effects from an entirely new perspective, no longer constrained by the assumption of infinite speed of light. We hope this will spur future research and help to better understand the complex behavior of time-resolved interactions between light and matter. We have used real data from the recent femto-photography technique [459] (Chapter 8), as well simulation data produced by a physically-based ray tracing engine especially designed to support transient rendering [201].

To visualize this data, we have developed an interactive image-based rendering application, that allows free navigation through the reconstruction of the captured scenes, including physically-based relativistic effects due to fast camera motion. We have introduced, for the first time in computer graphics, the modified equations necessary to render surfaces when irradiance is not constant over time, as well as an approximate solution for the case of rotation, for which a definite solution does not exist in the physics literature.

Of course there is plenty of exciting future work ahead. Our current implementation assumes Lambertian surfaces, so the viewing angle with respect to the normal has no influence in the result. This assumption can be relaxed by using more sophisticated IBR techniques e.g. [53]. Additionally, right now we only use radiance as captured by the sensor. When camera movement reveals surfaces which were originally occluded, we simply render them black. However, the use of time-resolved photographic techniques has already demonstrated promising results at recovering hidden information, including both geometry [457] and a parametric model of reflectance [326]. A promising avenue of research we are already working on involves generalizing these seminal works to be able to obtain both geometry and reflectance at the same time for hidden objects.

Part VI

CONCLUSION

10 CONCLUSIONS AND FUTURE WORK

In this thesis we have presented a number of contributions providing solutions to existing problems in the different stages of the imaging pipeline. Due to this variety, we will show here conclusions for the different parts separately.

Let aside specific conclusions, our overall conclusion is that the future of imaging, including capture and displays, relies on joint advances on several different dimensions. We are convinced that knowledge of perception should play a key role in future advances, with perceptual effects and limitations of the human visual system (HVS) being taken into account in the design of future algorithms and hardware. We have presented in this thesis a series of examples in which perceptual knowledge was leveraged to improve the viewing experience, by using e.g. existing perceptual metrics (Chapter 2), computational models of perception (Chapter 5), or by making our own measurements (Chapter 6).

However, the challenges in this are not negligible. Despite the vast amount of research devoted to the understanding of how we perceive, the HVS is an extremely complex system based on a combination of physiological and psychological factors. Consequently, presenting computational models of perception which are comprehensive enough has been a challenge, and there is still a lot of work to be done in this realm. At the same time, if the models are too complex, being able to incorporate them into algorithms, some of which need to run in real time (e.g. on the display end), is another challenge. There is thus a trade-off in this process, which requires deep knowledge of both the perceptual side, to know where simplifications can come in, and of the algorithmic and mathematical side; plus knowledge of the hardware and optics. Multidisciplinary cooperation is key.

As for the specific conclusions, we separate them in parts. We compile here a summary of the conclusions which can be found in each part, but also some more general insights and work for the future.

In Part II we presented first our work on coded apertures. We realized that coded apertures for defocus deblurring were based on optimizations where the error was measured using pixel-wise difference metrics; instead, we incorporated perceptual metrics for the measurement of the error, and obtained apertures that performed better than the previous state of the art. However, our exploration of the objective function, which metrics should be incorporated and, if more than one, how should they be weighted, is not definite. We have explored a number of possibilities, but a more thorough exploration would be advisable. We can even see our work as a proof of concept, a demonstration that better results can be obtained by incorporating perceptual metrics, but we by no means claim that our objective function is the definite and optimal objective function. There is also work to do in terms of reconstruction, recovery of the image. We use here a simple prior, but we believe more elaborated priors, either based on perceptual aspects or data-driven, or a combination of both, can yield better results.

The second half of Part II is devoted to reverse tone mapping. We have shown in it how existing reverse tone mapping operators (rTMOs) do not perform well on over-exposed content, or images with large bright areas. For that case, we have proposed a range expansion method based on a gamma curve, where the value of the gamma is computed based on a series of image statistics. Our expansion is very simple, and can thus be done in real time in the display firmware. We show that the method performs better than state-of-the-art algorithms. One aspect that remains unexplored in the work here presented is temporal consistency, how well the method extends to video; this remains

as future work. Additionally, we explored the introduction of an artistic component in the expansion process, presenting a semi-automatic method for range expansion based on the Zone System devised by photographer Ansel Adams. While this approach is not practical for expansion of a large amount of legacy material, this is not the intention either, what we offer is a way to expand the luminance range of the content based on tuning the expansion curve using a division more suited for artistic purposes than e.g. a straightforward uniform division or other manipulations of the curve.

Part III begins with a survey, a state-of-the-art report of the field of computational displays. We believe that categorizing the displays according to the plenoptic dimension they aim to improve offers a new view of the field, and a more practical one. It shows that joint approaches which attend to how the content is generated, and how it will be perceived, allow for improvements which would otherwise not be possible just with advances in hardware. The survey also points out what we believe to be the main challenges for the future, and shows that the future lies in joint advances on different dimensions, as well as additional influencing factors such as new materials, the adaptation of mathematical models for high-performance real-time computation, or the co-design of custom optics and electronics, to name a few.

We have—also in Part III—presented an algorithm for disparity remapping for automultiscopic and stereoscopic displays. Content retargeting, understood widely, is a one of the main problems in imaging, both in industry and in research. It can be, for instance, spatial retargeting, for adaptation of content to different aspect ratios; disparity retargeting, for proper depiction of a 3D scene on different devices; or color retargeting (gamut mapping), for an accurate color appearance. The existing variety of devices, the easy distribution of content, and the large number of players involved makes standardization a huge challenge and makes necessary the development of robust techniques that address these problems. In this case we have focused on automultiscopic displays, which exhibit depth of field, allowing depiction of sharp, in-focus content only in a limited depth range around the screen. This depth of field varies across displays, thus the need for retargeting. Our method incorporates knowledge from luminance-contrast and depth perception, and has been validated with state-of-the-art perceptual metrics. We currently work on the central view of the light field, and in the future approaches which take into account the whole structure of the light field are desirable. On the plus side, we also show retargeting of stereo content; and for the first time take into account a non-dichotomous zone of comfort: The zone of comfort has always been used as a safe dichotomous area, but it is believed to be a continuous area instead. We can incorporate non-dichotomous zones of comfort in our framework.

The end of Part III is devoted to comfort when viewing content in motion in stereo displays. This area had been largely unexplored, and we have proposed here the, to our knowledge, most comprehensive study measuring comfort in stereoscopic motion. We have seen that the interplay of the factors studied has a significant influence in comfort. Seeking for the applicability of such measurements, we have devised a metric of comfort, which correlates well with subjective scores. The challenge in this realm is double, first, performing comprehensive measurements and which involve the relevant factors; second, translating those measurements of comfort to a model which can be used in a practical scenario. To this end we derived zones of comfort from our measurements, and proposed the metric, but what we have done here is only a first step towards the final objective.

In Part IV we have dealt again with light fields, but in a new context: Interaction. We explored interaction paradigms for this new multi-dimensional structure that is a light field. From our first study we have concluded that a paradigm based on parallax alone is insufficient, that focus, and especially the feeling of control it provides, is highly favored by users, that occlusions require more complex handling tools. For the future, we depart from the assumption that depth information –although maybe imperfect– will

be available, given the progress of algorithms for reconstruction from light fields; thus the challenge of handling imperfect depth arises. Also, more sophisticated tools such as a depth picker to select depth ranges, will be incorporated. How to interact with light fields, how to specify what to edit, is what we aim at here; there is also, however, a lot of work to be done in light field editing *per se*, features such as copy and pasting, or morphing, which have received some attention but are still somehow in its infancy compared to the things that could be done.

In the last part, femto-photography (Part V), the main conclusion is that transient imaging has opened a vast, untrodden field with an incredible potential. Not only it can help us better understand how light propagates, being able to sample (as opposed to integrate over) the time dimension can provide solutions to long-standing problems such as depth recovery in the presence of multiple paths, or separation of light transport components. Besides, it enables new applications such as detecting movement or recovering shape around corners. The applications of transient imaging span a wide range of fields, including medicine, surveillance, or material science. A number of works have appeared since the presentation of the femto-photography technique, which are also capable of capturing at very high frame rate at a lower cost; the drawback is that their resolution is in the nanosecond range, and not in the picosecond range, but it does reflect the interest in the field, and we hope these lower-cost techniques will help foster research in time-resolved imaging.

On a personal, or non-technical, level, this thesis has allowed me to gain matureness as a researcher, and as a professional. There is a significant evolution, hard to see while it happens, but clear when you are at the end of these four years and you look back. You improve your analysis skills, you learn the importance of being not only effective, but also efficient in your work. I have additionally had the chance to work with a quite large number of people during my PhD, either remotely or sharing location; sometimes in projects led by myself. This not only teaches you how to work on a team, how to deal with differences in working habits, speeds, level of demand, and even cultural differences, but you also learn how to plan for a project and how to organize a group of people. Of course, this collaborations build up strong teams with expertise in a number of areas, as opposed to working alone with my supervisor, and this enriches and greatly benefits the research being carried out. Supervising students is another opportunity to learn: You have to achieve a combination of desired characteristics: You have to engage the student, motivate him or her, adequate the path to their characteristics if necessary, plan accordingly, and find this compromise between how much you teach him, how much you help him out, and how much you let him find out by himself, even if it costs more. Being at a highly demanding institution like MIT has also been an enriching experience and has had a tremendous impact in my growth as a researcher. Work in a competitive, yet collaborative environment, access to resources beyond what I could expect, constant exposure to world experts in their areas, cooperation with researchers from other fields, and a constant common desire to change the world, and to come up with groundbreaking ideas, inevitably and fortunately shaped my growth as a researcher during those months. Also, having the opportunity to work on a project of the highest impact like that of femto-photography imposes a great challenge, paired up with excitement and satisfaction. All in all, these years have been a tough yet extremely enriching experience.

Part VII

APPENDICES

A

REVERSE TONE MAPPING: IMAGE STATISTICS

Table A.1: Statistics for the images in our dataset. Please refer to Section 3.5.1 and Chapter 3 for details.

Image	γ	L_{avg}	L_H	$\log L_H$	k_5	k_1	L_{med}	V_L	σ_L	$skew_L$	$kurt_L$	p_{ov}
Building01	1.22	0.3493	0.1182	-2.1352	0.5743	0.6019	0.0762	0.1357	0.3684	0.5261	1.6310	5.9764
Building02	1.5	0.4853	0.2485	-1.3922	0.6472	0.6775	0.2076	0.1774	0.4212	0.1879	1.1346	20.0422
Building03	1.75	0.5792	0.4052	-0.9033	0.6865	0.7265	0.4176	0.1527	0.3908	0.0275	1.1750	40.1717
Building04	2.6	0.7105	0.6196	-0.4787	0.7392	0.7806	0.7399	0.0912	0.3020	-0.4495	1.7895	44.3981
Lake01	1.1	0.1188	0.0338	-3.3881	0.4963	0.5612	0.0248	0.0256	0.1601	1.1686	2.6714	0.0003
Lake02	1.2	0.1662	0.0570	-2.8651	0.5151	0.5721	0.0473	0.0450	0.2121	1.1497	2.6493	0.0017
Lake03	1.5	0.3689	0.2020	-1.5996	0.5545	0.6418	0.1827	0.1316	0.3628	0.8654	2.1098	17.8714
Lake04	2.25	0.4977	0.3613	-1.0182	0.5975	0.6874	0.3668	0.1225	0.3500	0.4138	1.5494	22.8514
Sunset01	1.1	0.2088	0.0719	-2.6319	0.4857	0.5622	0.0784	0.0786	0.2803	1.6711	4.7589	4.0931
Sunset02	1.35	0.2633	0.1143	-2.1688	0.5252	0.6170	0.1361	0.0888	0.2980	1.3072	3.5468	5.0731
Sunset03	1.4	0.3930	0.2259	-1.4875	0.5884	0.6907	0.2675	0.1152	0.3394	0.6207	1.9193	9.4525
Sunset04	1.75	0.6633	0.5505	-0.5969	0.6891	0.8009	0.7168	0.1025	0.3202	-0.3523	1.6076	29.5133
Graffiti01	1.2	0.2891	0.1568	-1.8525	0.6213	0.6726	0.2744	0.0557	0.2361	0.3878	1.8997	0.0039
Graffiti02	1.35	0.5020	0.3335	-1.0981	0.6659	0.7405	0.5732	0.1110	0.3332	-0.0818	1.4351	1.0967
Graffiti03	1.5	0.6796	0.5449	-0.6071	0.7201	0.8074	0.8761	0.1158	0.3403	-0.5197	1.5827	21.1744
Graffiti04	1.75	0.8091	0.7415	-0.2991	0.7761	0.8591	0.9949	0.0692	0.2631	-1.0777	2.8062	51.2861
Strawberries01	1.22	0.1718	0.0954	-2.3501	0.5646	0.6074	0.1075	0.0283	0.1681	1.2238	3.6194	0.0000
Strawberries02	1.35	0.3381	0.2240	-1.4962	0.5713	0.6335	0.2544	0.0729	0.2700	0.7483	2.3469	0.0781
Strawberries03	1.55	0.5304	0.4098	-0.8921	0.6235	0.6884	0.5048	0.1037	0.3220	0.1416	1.5362	8.9808
Strawberries04	1.9	0.6961	0.6046	-0.5033	0.6942	0.7524	0.8147	0.0934	0.3057	-0.4674	1.6204	26.5756

B

F-TESTS FOR ASSESSING THE APPROPRIATENESS OF ADDING NEW PREDICTORS TO A MODEL

An F-test is typically performed to decide whether or not a certain null hypothesis can be rejected. To do this, a test statistic (the F-statistic) is needed which under the null hypothesis follows an F-distribution. In our case (Section 3.5.1), the null hypothesis is that, given two models, A and B, with a number of predictors p_A and p_B ($p_A > p_B$), the two models fit equally well the data. The F-statistic is then given by:

$$F_{p_A - p_B, n - p_A} = \frac{(SS_B - SS_A)/(p_A - p_B)}{SS_A/(n - p_A)}, \tag{49}$$

where SS_i, $i = \{A, B\}$, is the sum of squared residuals of model i, and n is the number of data values [422]. It must be noted that in Equation 49, and throughout this thesis, p_i as a measure of the number of terms in the regression includes the constant term (i.e. the intercept).

For the particular case of creating model A by adding one variable to a model B that has p terms, and expressing the formula in terms of R^2, the F-statistic becomes:

$$F_{1, n-p-1} = \frac{R_A^2 - R_B^2}{(1 - R_A^2)/(n - p - 1)} \tag{50}$$

As it is well known, given a value for F in an F-test, the p-value is the probability of obtaining a value as extreme as the F obtained, assuming that the null hypothesis is true. As a consequence, the null hypothesis is typically rejected if the p-value is lower than the significance level alpha (which, in this work, will have the usual value of $\alpha = 0.05$).

C

GOODNESS OF FIT IN MULTILINEAR REGRESSIONS

This appendix includes the description of a series of metrics which are typically used in regression analysis to measure the accuracy of the fitting of a certain model.

RMSE. For a multilinear regression, RMSE is computed as shown in Equation 51, where Y_i are the observed data (i.e. the given γ values) and $\hat{Y}_i$ the data predicted by the model.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2/(n-p)}, \tag{51}$$

where, n is the data size and p the number of terms in the regression. Please recall that in this formulation the intercept is included in p. This metric provides an intuition on the error we would incur in when using a certain regression to estimate the value of a variable.

OVERALL F-STATISTIC. The *overall* F-statistic is simply an F-test in which the null hypothesis is that the data can be explained by a constant (which would be the mean of the observed data), versus the hypothesis that the data can be explained by the selected model. Therefore, a high F-statistic and, specially, a low associated p-value indicate that the hypothesis that our model explains the data (vs. the hypothesis that a constant explains them) is clearly correct.

$\mathbf{R^2}$ AND ADJUSTED $\mathbf{R^2}$. Typically used to assess how well the values predicted by a model will adjust to the real values, in the case of linear regressions R^2 is simply the square of the correlation coefficient between the observed and the predicted data.

However, in the case of multilinear regression, the R^2 value will always increase as new variables are added to the model. For this reason sometimes the *adjusted* R^2 is used, which corrects for the number of explanatory variables in the model. As a result, the adjusted R^2 value will only increase if the new term improves the regression more than would be expected by chance. The adjusted R^2 value is usually denoted by $\tilde{R}^2$ and computed as follows:

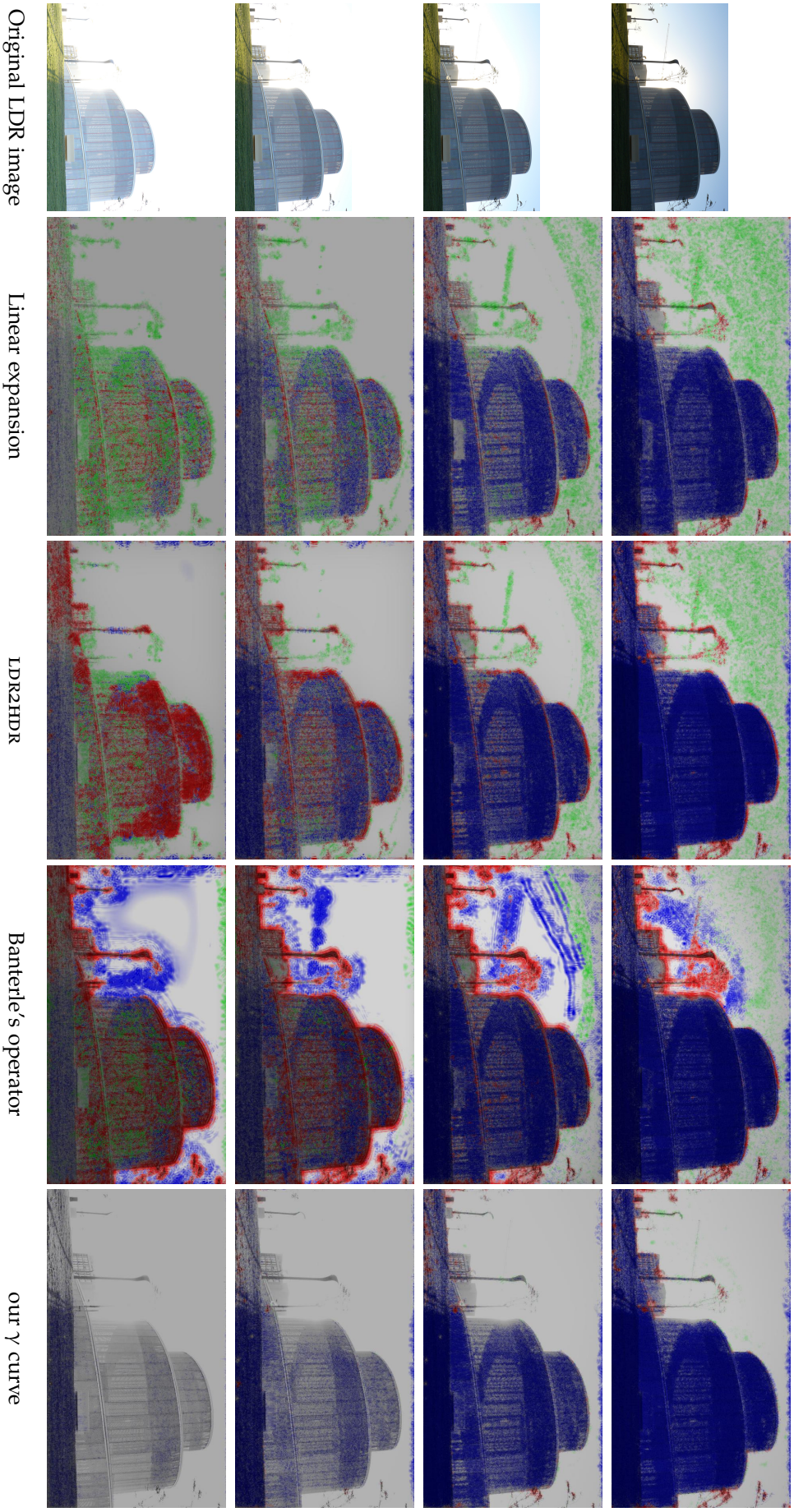
$$\tilde{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p} \tag{52}$$

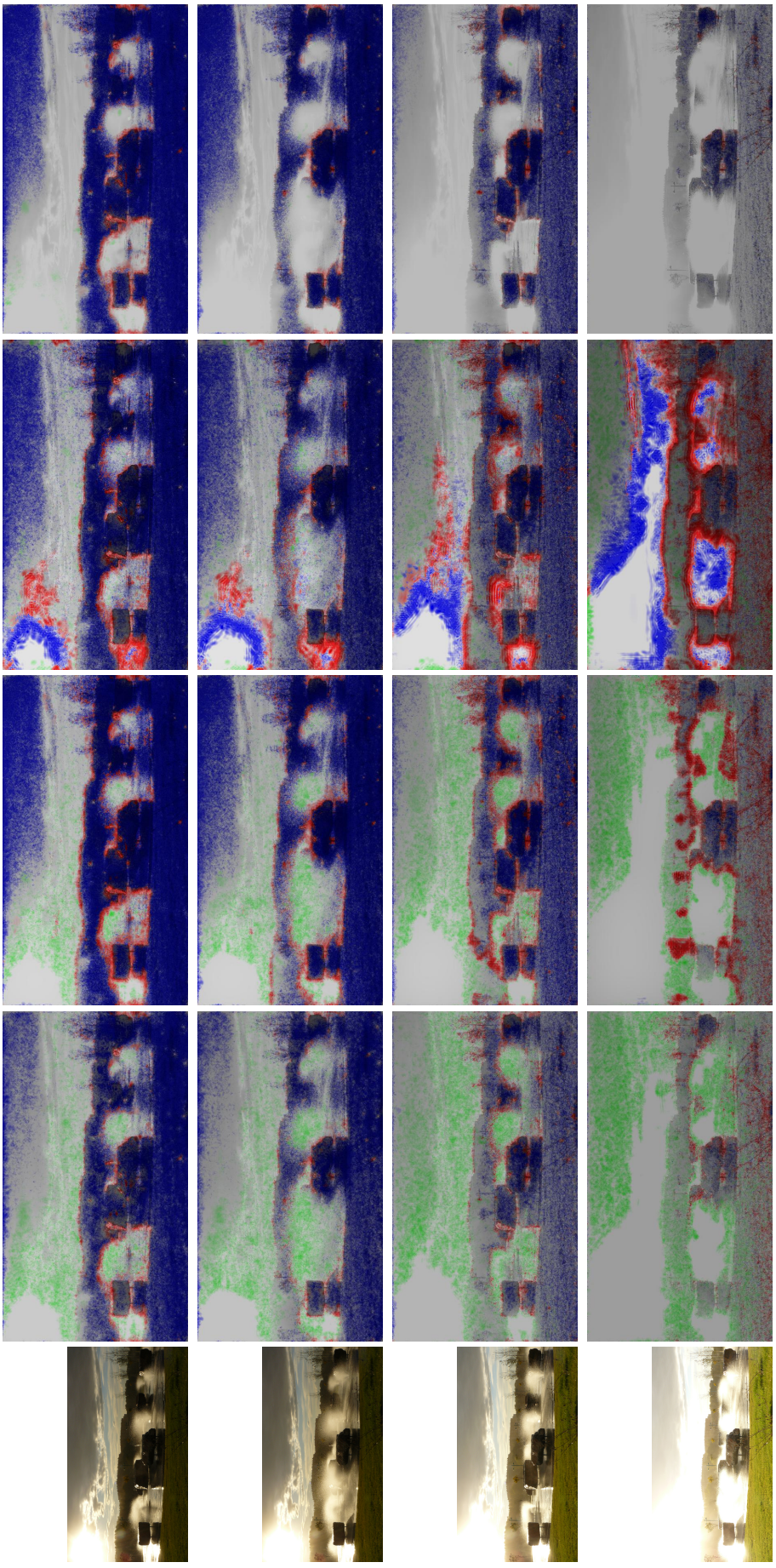
where, again, n is the data size and p the number of terms in the regression. Please recall that in this formulation the intercept is included in p. It is well-known that the higher the R^2 and the adjusted R^2 values, the higher the correlation between the values predicted by the model and the values actually observed.

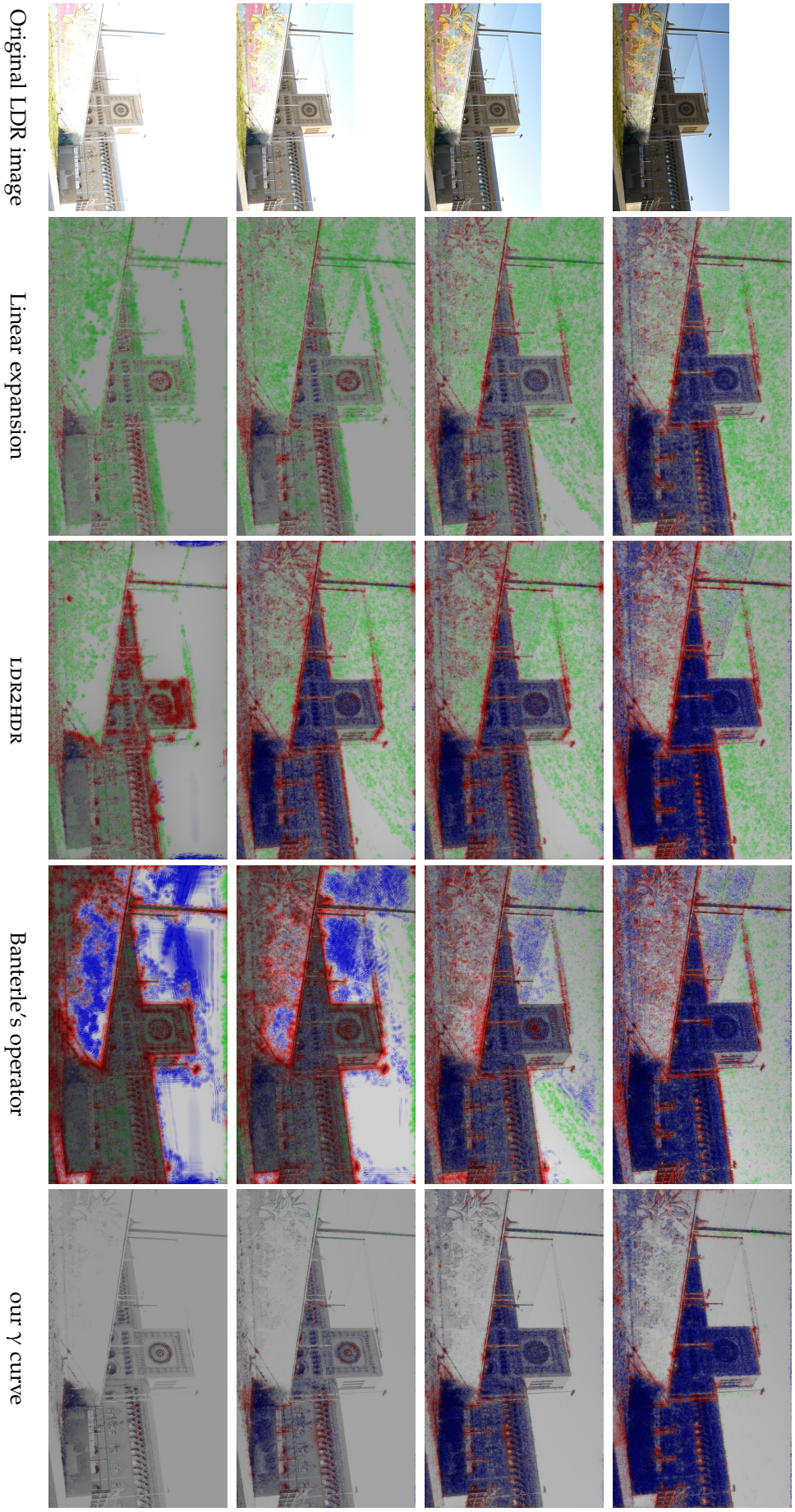
D

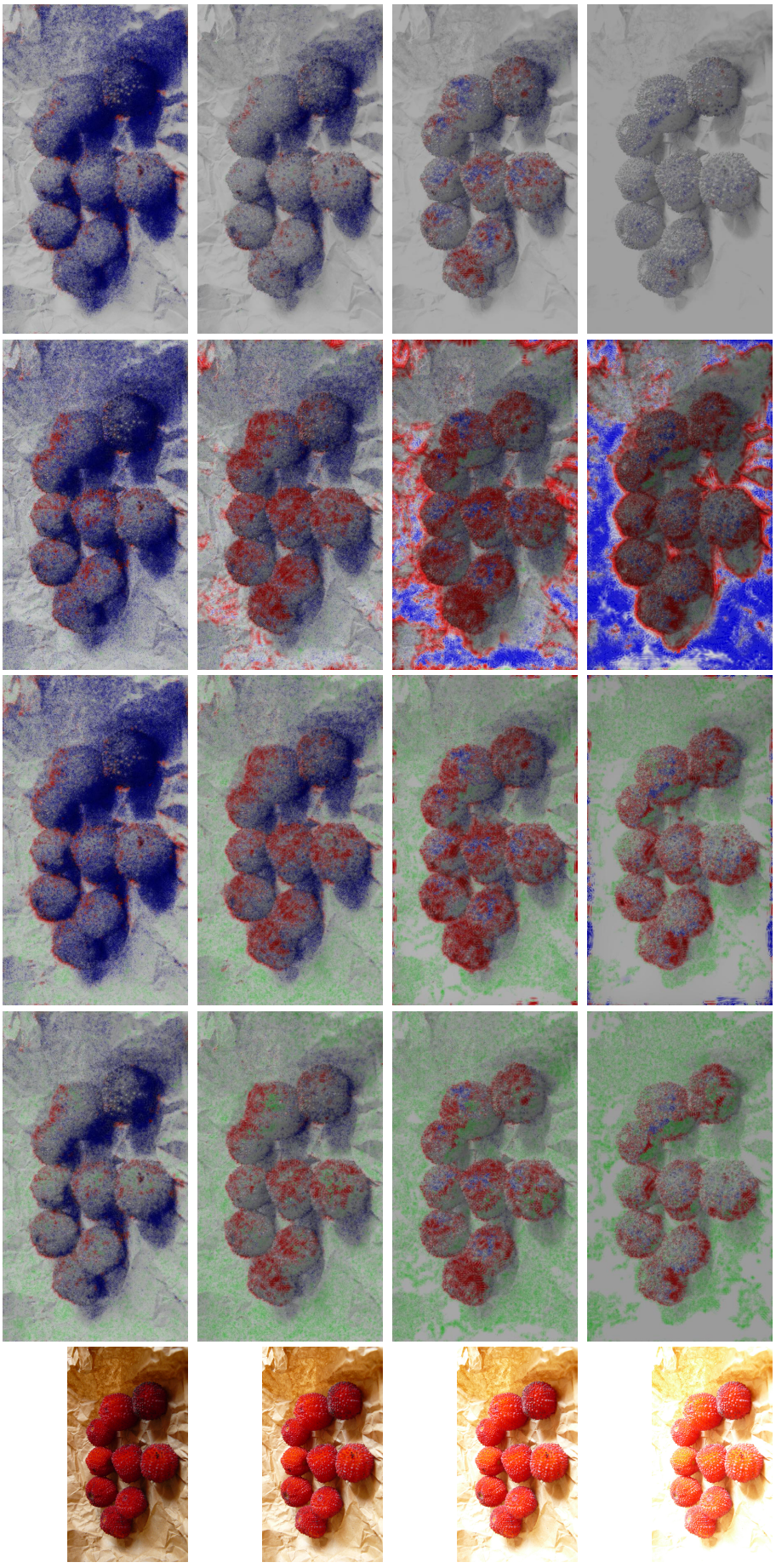
REVERSE TONE MAPPING: RESULTS OF THE OBJECTIVE EVALUATION

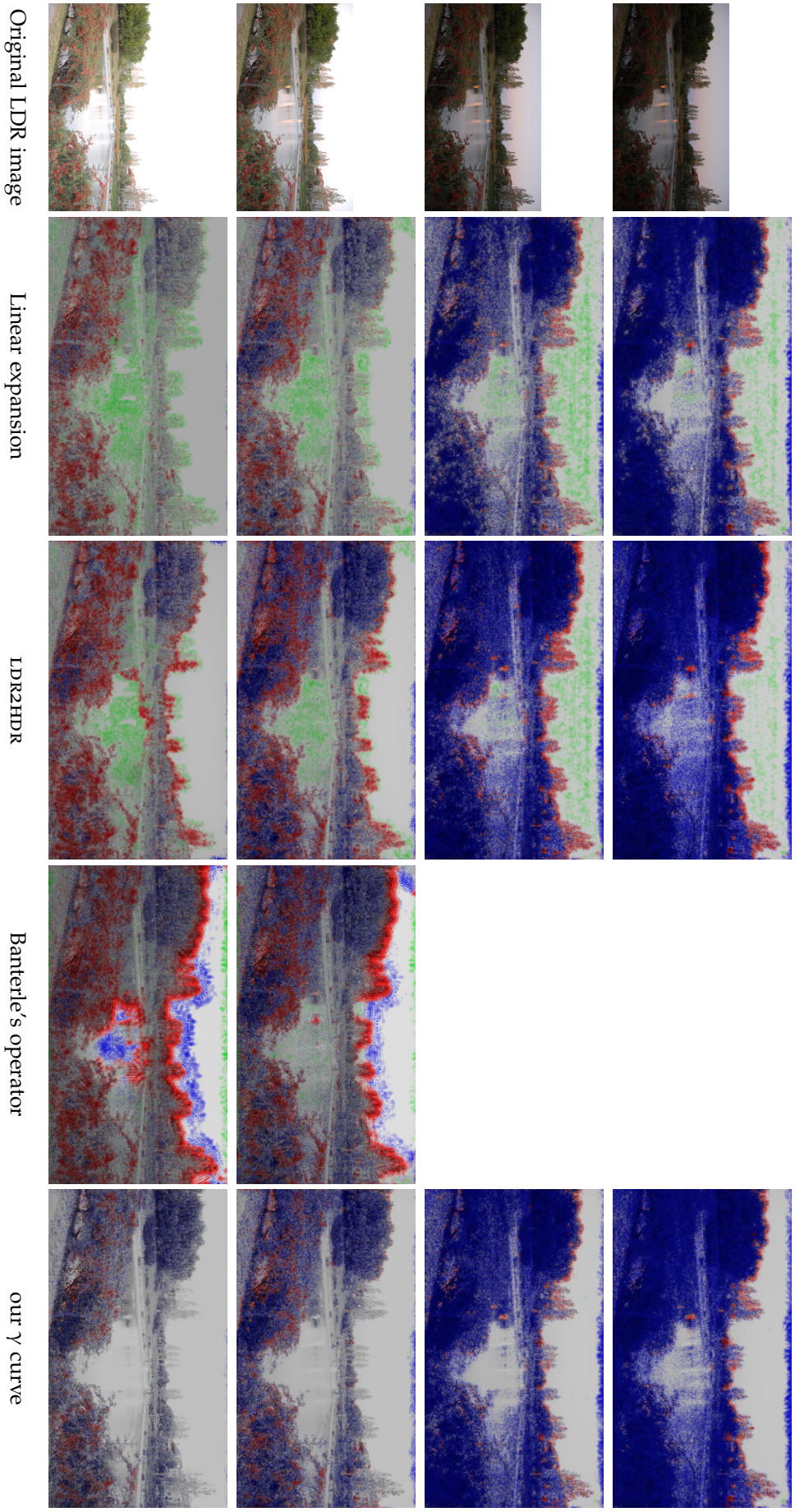
Results of the objective image quality metric developed by Aydin et al. [24] comparing the original LDR images (reference images) with the corresponding outputs after reverse tone mapping them with the different operators. Each row shows the results of the linear expansion, LDR2HDR, Banterle's operator and our gamma expansion respectively. Green, blue and red identify loss of visible contrast, amplification of invisible contrast and contrast reversal, respectively. Please refer to Chapter 3 for details.


Original LDR image
Linear expansion
LDR2HDR
Banterle's operator
our γ curve


our γ curve
Banterle's operator
LDR2HDR
Linear expansion
Original LDR image


Original LDR image
Linear expansion
LDR2HDR
Banterle's operator
our γ curve


our γ curve
Banterle's operator
LDR2HDR
Linear expansion
Original LDR image


Original LDR image
Linear expansion
LDR2HDR
Banterle's operator
our γ curve

E

DISPLAY ADAPTIVE 3D CONTENT REMAPPING: OBJECTIVE FUNCTION AND ANALYTICAL DERIVATIVES IN THE OPTIMIZATION

In this appendix we go through the mathematical expressions of the two terms of the objective function described in Section 5.4 and shown in Equation 34. We also include their derivatives, necessary for computing the analytical Jacobian used in the optimization process.

E.1 TERM 1: OPTIMIZING LUMINANCE AND CONTRAST

This term, as shown in Equation 30 in Section 5.4, has the following form:

$$T_1 = \omega_{CSF}\left(\rho_S\left(L_{orig}\right) - \rho_S\left(\phi_b\left(L_{orig}, d\right)\right)\right) \tag{53}$$

Note that this expression yields a vector of length N_{pyr} (N_{pyr} being the number of pixels in the pyramid $\rho_S\left(L_{orig}\right)$ or
$\rho_S\left(\phi_b\left(L_{orig}, d\right)\right)$), which is a vector of differences with respect to the target luminance L_{orig}, weighted by contrast sensitivity values. This vector of errors thus contains the residuals that `lsqnonlin` optimizes for the depth of field term. The weighting factor μ_{DOF} is left out of this derivation for the sake of simplicity, since it is just a product by a constant both in the objective function term and in its derivatives. This is valid also for the second term of the objective function.

Since the multi-scale decomposition is a linear operation, we can write:

$$T_1 = \omega_{CSF}\left(M_S \cdot L_{orig} - M_S \cdot \phi_b\left(L_{orig}, d\right)\right) \tag{54}$$

where M_S is a matrix of size $N_{pyr} \times N_{im}$, N_{im} being the number of pixels in the luminance image L_{orig}. Substituting the blurring function $\phi_b\left(\cdot,\cdot\right)$ by its actual expression

$$\frac{\partial T_{1,i}}{\partial d} = \omega_{CSF,i}\left(-M_{S,i} \cdot \left(L_{orig} * \frac{\partial k(d)}{\partial d}\right)\right), \tag{55}$$

where $M_{S,i}$ is the $i-th$ row of M_S.

The derivative of the kernels $k(d)$ is:

$$\frac{\partial k(d)}{\partial d} = \frac{\left(\exp\left(-\frac{x_i^2+y_i^2}{2(\sigma(d))^2}\right)\right)\left(\frac{(x_i^2+y_i^2)4\sigma(d)\frac{\partial\sigma(d)}{\partial d}}{(2(\sigma(d))^2)^2}\right)\sum_j^K\left[\exp\left(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2}\right)\right]}{\left(\sum_j^K\left[\exp\left(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2}\right)\right]\right)^2} - \frac{\sum_j^K\left[\left(\exp\left(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2}\right)\right)\left(\frac{(x_j^2+y_j^2)4\sigma(d)\frac{\partial\sigma(d)}{\partial d}}{(2(\sigma(d))^2)^2}\right)\right]\left(\exp\left(-\frac{x_i^2+y_i^2}{2(\sigma(d))^2}\right)\right)}{\left(\sum_j^K\left[\exp\left(-\frac{x_j^2+y_j^2}{2(\sigma(d))^2}\right)\right]\right)^2}. \tag{56}$$

The derivative of the standard deviation σ is straightforward, knowing $\partial(f_\xi(d))/\partial d$. As described in the main text, the expression for $f_\xi(d)$ depends on the type of automultiscopic display. For a conventional display [512]:

$$f_\xi(d) = \begin{cases} \frac{f_0}{N_a}, & \text{for } |d| + (h/2) \leqslant N_a h \\ \left(\frac{h}{(h/2)+|d|}\right) f_0, & \text{otherwise} \end{cases}, \tag{57}$$

where N_a is the number of angular views, h represents the thickness of the display and $f_o = 1/(2p)$ is the spatial cut-off frequency of a mask layer with a pixel of size p. For multilayered displays, the upper bound on the depth of field for a display of N layers is [482]:

$$f_\xi(d) = N f_0 \sqrt{\frac{(N+1)h^2}{(N+1)h^2 + 12(N-1)d^2}}. \tag{58}$$

The derivatives are as follows:

$$\frac{\partial f_\xi(d)}{\partial d} = \begin{cases} 0, & \text{for } |d| + (h/2) \leqslant N_a h \\ \left(\frac{-h d/|d|}{((h/2)+|d|)^2}\right) f_0, & \text{otherwise} \end{cases} \tag{59}$$

for a conventional display and

$$\frac{\partial f_\xi(d)}{\partial d} = N f_0 \frac{12\sqrt{N+1}(N-1)hd}{((N+1)h^2 + 12(N-1)d^2)^{3/2}}. \tag{60}$$

for a multilayered display.

E.2 TERM 2: PRESERVING PERCEIVED DEPTH

This term, introduced in Equation 32 in Section 5.4, is modeled as follows:

$$T_2 = \omega_{BD}\left(\rho_L\left(\phi_v\left(D_{orig}\right)\right) - \rho_L\left(\phi_v\left(d\right)\right)\right) \tag{61}$$

Again, since the multi-scale decomposition is a linear operation, we write:

$$T_2 = \omega_{BD}\left(M_L \cdot \phi_v\left(D_{orig}\right) - M_L \cdot \phi_v\left(d\right)\right), \tag{62}$$

where M_L is a matrix of size $N_{dpyr} \times N_d$, N_d being the number of pixels in the depth map D_{orig}. Taking the derivative with respect to d yields the following expression for each element $T_{2,i}$ of the residuals vector for this term:

$$\frac{\partial T_{2,i}}{\partial d} = \omega_{BD,i}\left(-M_{L,i} \cdot \frac{\partial \phi_v\left(d\right)}{\partial d}\right), \tag{63}$$

where $M_{L,i}$ is the $i-th$ row of M_L. As explained in the main text, $\phi_v\left(d\right)$ converts depth d_P of a point P into vergence v_P. This, given the viewing distance v_D and the interaxial distance e, is done using function $\phi_v\left(\cdot\right)$:

$$\phi_v\left(d\right) = a\cos\left(\frac{\mathbf{v_L} \cdot \mathbf{v_R}}{\|\mathbf{v_L}\| \, \|\mathbf{v_R}\|}\right), \tag{64}$$

where vectors $\mathbf{v_L}$ and $\mathbf{v_R}$ have their origins in P and end in the eyes (please also see Figure 5.6 in Section 5.4). Placing the coordinate origin in the center of the screen (z-axis normal to the screen, x-axis in the horizontal direction) we can rewrite the previous equation for a point $P = (x_i, y_i, d_i)$ as:

$$v_d = \phi_v\left(d\right) = a\cos\left(\frac{\kappa}{\sqrt{\eta}\sqrt{\zeta}}\right), \tag{65}$$

where:

$\kappa = (x_L - x_i)(x_R - x_i) + (v_D - d_i)^2$,

$\eta = (x_L - x_i)^2 + (v_D - d_i)^2$, and

$\zeta = (x_R - x_i)^2 + (v_D - d_i)^2$.

Finally, differentiating Equation 65 with respect to depth:

$$\frac{\partial \phi_v(d)}{\partial d} = -\left(1 - \left(\frac{\kappa}{\sqrt{\eta}\sqrt{\zeta}}\right)^2\right)^{-1/2} \left(\frac{-2(v_D - d_i)\sqrt{\eta}\sqrt{\zeta} - \kappa\Psi(d_i)}{\eta\zeta}\right)$$

where $\Psi(d_i)$ is as follows:

$$\Psi(d_i) = -d_i(v_D - d_i)\eta^{-1/2}\zeta^{1/2} - d_i(v_D - d_i)\zeta^{-1/2}\eta^{1/2}.$$

F

DISPLAY ADAPTIVE 3D CONTENT REMAPPING: A DICHOTOMOUS ZONE OF COMFORT

As explained in Chapter 5 (Section 5.6), Equation 66 describes our objective function for the simplified case of stereo remapping:

$$\left\|\omega_{BD}\left(\rho_L\left(\phi_v\left(D_{orig}\right)\right)-\rho_L\left(\phi_v\left(d\right)\right)\right)\right\|_2^2+\mu_{CZ}\left\|\varphi\left(d\right)\right\|_2^2, \tag{66}$$

where $\varphi(\cdot)$ is a function mapping depth values to visual discomfort. We describe here a means to incorporate a dichotomous model of the zone of comfort, such as those shown in cyan in Figure 5.7 for different devices and viewing distances v_D (we reproduce the figure here again as Figure F.1 for completeness). Instead of the non-dichotomous model described in Section 5.6 (shown in orange in the same figure), we can define a binary indicator function, such as

$$\varphi_{dc}\left(d\right)=\begin{cases}0 & \text{for}\quad d_{comfort}^{min}\leqslant d\leqslant d_{comfort}^{max}\\ \infty & \text{otherwise}\end{cases} \tag{67}$$

For a practical, numerically-robust implementation, a smooth function that approximates Equation 67 is preferable, ensuring C^1 continuity. Our choice for such a function is the Butterworth function which is commonly used as a low-pass filter in signal processing:

$$\varphi_{bf}\left(d\right)=1-\sqrt{\frac{1}{1+\left(\gamma d\right)^{2s}}} \tag{68}$$

where γ controls the position of the cut-off locations and s the slope of such cut-offs.

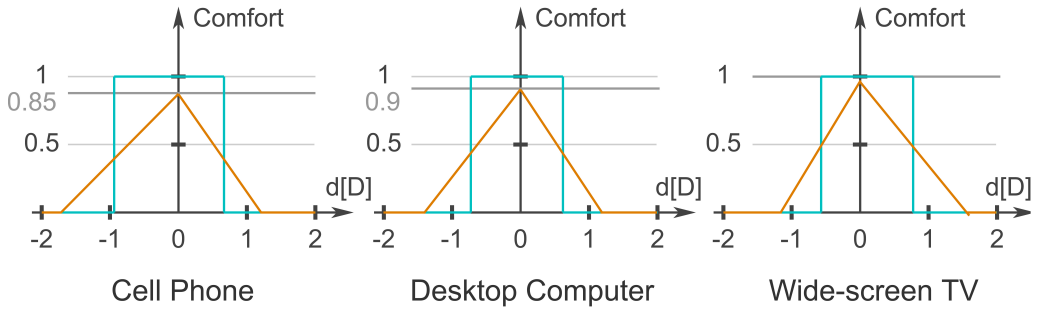


Figure F.1: Dichotomous (blue) and non-dichotomous (orange) zones of comfort for different devices. From left to right: cell phone ($v_D = 0.35m$), desktop computer ($v_D = 0.5m$) and wide-screen TV ($v_D = 2.5m$). This figure is a reproduction of Figure 5.7, reproduced here for completeness.

G

VISUAL COMFORT IN STEREO MOTION: ADDITIONAL DATA

G.1 SLICES OF THE COMFORT FUNCTION

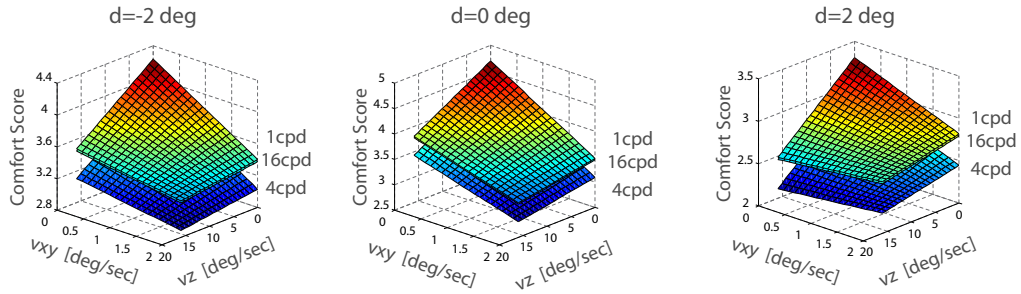


Figure G.1: Slices of our comfort function, from left to right: $d = -2°, 0°, 2°$.

G.2 COMFORT ZONES

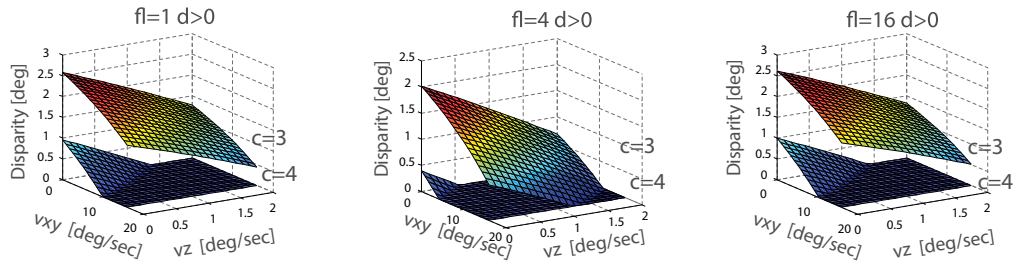


Figure G.2: Comfort zones derived from our comfort function for $d > 0$, from left to right: $f_l = 1cpd, 4cpd, 16cpd$.

G.3 RATINGS FOR THE STIMULI

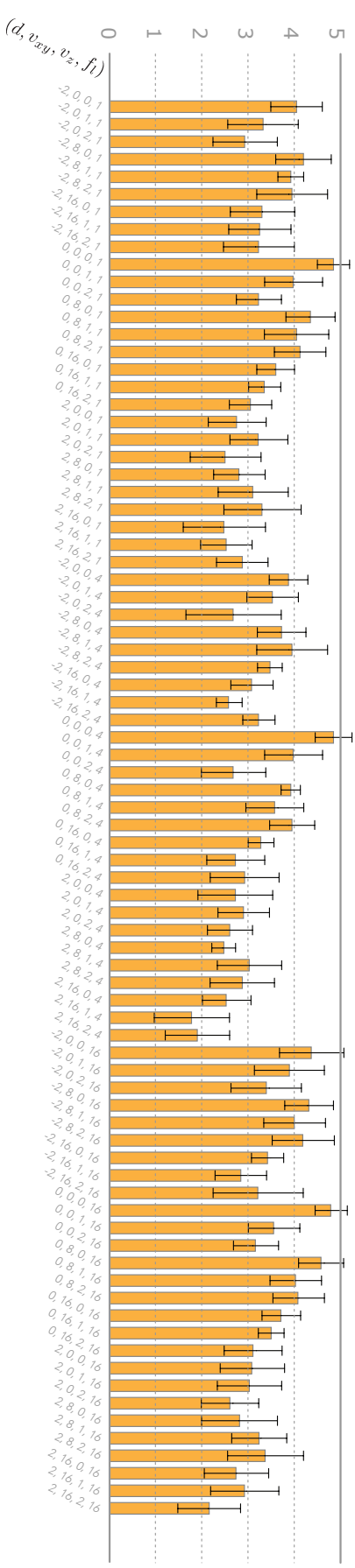


Figure G.3: Ratings for the stimuli in our experiments. Error bars show standard error of the mean.

H

LIGHT FIELD EDITING INTERFACES: INTERFACE IMPLEMENTATION DETAILS

We use screen-space rendering in OpenGL to display both the light field and the edits, using two simple GLSL shaders, one for each interface paradigm (*multiview* and *focus*). The light field and the edits are stored in different 2D textures. These textures store an array of images where each image represents one different view of the light field. The two textures are blended together in rendering time. Strokes are rendered directly on the edits' texture, used as a render target, with a different GLSL shader for each tool.

Depth information is stored as a disparity map, and is computed from the ground truth depth map and the light field camera properties (i.e. number of views, focal distance, FOV and distance between cameras) using the code shown in Listing 1. This disparity map is stored for each view. Storing depth (disparity) information as a map, instead of computing it on-the-fly using e.g. ray casting introduces the problem of quantization, which may lead to small errors due to quantization; nevertheless, these are not statistically significant. We opt for this approach due to its efficiency, to ensure real-time frame rates.

In the second screen *W2* in interfaces based on focus, the edits are not blended with the light field; it is thus used by users mainly when erasing edits. To help them in this process, and based on pilot tests, we do not display the light field in this second screen when erasing, but only the edits (strokes, or pasted images). Figure H.1 shows the view on W1, the "standard" view in W2, and the view in W2 when erasing.

A pilot test showed us that users assume that highlights lie on the surface of the object. In consequence, in Task 3, which requires changing the specular highlights of a *fertility* figurine, we measure error with respect to the surface of the object, in order to provide a more fair comparison between interfaces with and without depth.

Listing 1: Source code of the function used to obtain disparity from the depth map and the light field camera properties.

```
function [ disp ] = depth_to_disparity( depth, size_image,
                nbviews, radius, focaldistance, fov )

        imgsize = tan(fov*pi/180/2)*2*focaldistance;
        pix_size = imgsize / size_image;
        du = radius / floor(nbviews/2);

        disp = -(depth-focaldistance) ./ (depth)
                .* (du/pix_size) + 1;

end;
```

When it comes to occlusion handling, in interfaces with depth information the user can only draw on surfaces that are visible at the time he/she is drawing. The stroke he/she draws, which lies at a certain depth, then gets projected to the rest of the views of the light field. In certain cases editing occluded surfaces may thus require painting strokes on several views. In interfaces without depth information, the user places a stroke at a certain depth and it gets projected to all views; if there are objects closer than that depth in the line of sight of the user, the stroke will be projected on them. Since there is no knowledge of visibility, the user will need to handle that manually, by erasing. For more information on occlusion handling we refer the reader to the *Directed tasks* video (`http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Video_1.mov`), where the work-

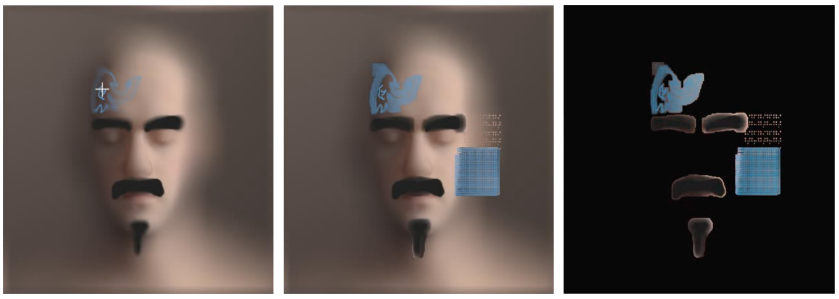


Figure H.1: Screenshot of the interfaces based on the *focus* paradigm. *Left:* View in W1. *Middle:* "Standard" view in W2. *Right:* View in W2 when erasing. To highlight the edits, we do not display the light field in the second window when erasing in this interface.

flow in Task 5 (painting behind a railing, thus handling occlusions) is shown for the four interfaces tested.

I

LIGHT FIELD EDITING INTERFACES: INSTRUCTIONS FOR INTERFACE EVALUATION TASKS

We include here the description given to the users for each *directed* task, together with the sample images in Figure I.1. For open tasks, images given to the users as a source of inspiration can be seen in Figure I.2.

TASK 1 *Draw your initial on the back blue wall approximately in the place indicated in the sample image. Use the brush (and the erase tool if necessary). Do not worry about the color of the brush. Time: 5 minutes.*

TASK 2 *Using the brush (and the erase tool if necessary), paint on the pattern of the vase as shown in the sample image to change the color of that part of the vase. Do not worry about the color of the brush. Time: 5 minutes.*

TASK 3 *Using the dodge tool (and the erase tool if necessary), increase the brightness of the specular highlights in the glossy statue of the image. Change only the specular highlights indicated in the sample image. Time: 5 minutes.*

TASK 4 *Once you press Start, an image will appear joined to the cursor. You have to place that image in the scene, so that to appears to be floating in the air. The image needs to be placed such that in depth it is situated in front of the vase, but behind the glossy statue (see sample image). Time: 5 minutes.*

TASK 5 *Using the brush (and the erase tool if necessary) draw, on the back wall, a heart so that it is partially occluded by the railing (see sample image). The heart needs to be on the wall, and thus occluded by the foreground railing. Time: 5 minutes.*

TASK 6 *In this task you can toggle depth information on/off at any point during the editing process. You are given a set of photos for inspiration. Suggestions: painting on the face, adding glasses, monocle, etc. Time: 12 minutes.*

TASK 7 *You can now choose between any of the four interfaces you have tested so far, that is, focus with or without depth, and multiview with or without depth. You can switch between focus and multiview and activate or deactivate depth information at any point during the editing process. The goal is making the scene more beautiful. Suggestions: adding flowers to the plants (the billboard object to insert are now some flowers), decorating the flower pots, or any other edit you can think of. Time: 12 minutes.*



(a) Task 1



(b) Task 2



(c) Task 3



(d) Task 4



(e) Task 5

Figure I.1: Target images given to users in *directed* tasks.



(a) Task 6



(b) Task 7

Figure I.2: Sample images given to users in *open* tasks.

J

LIGHT FIELD EDITING INTERFACES: HYBRID INTERFACE AND ADVANCED TOOLS

Figure J.1 shows our hybrid interface, using both the *focus* and *multiview* paradigms, which we additionally equip with two new tools: *spline* and *image deformation*. The *spline* tool is implemented as a Hermite spline, where the control points can be modified in 3D. This tool draws on screen using the same GLSL shader as the *brush* tool. The (*image deformation*) tool extends the previous billboard *paste* tool by enabling deformations of the image by moving in 3D control points placed at the corners.

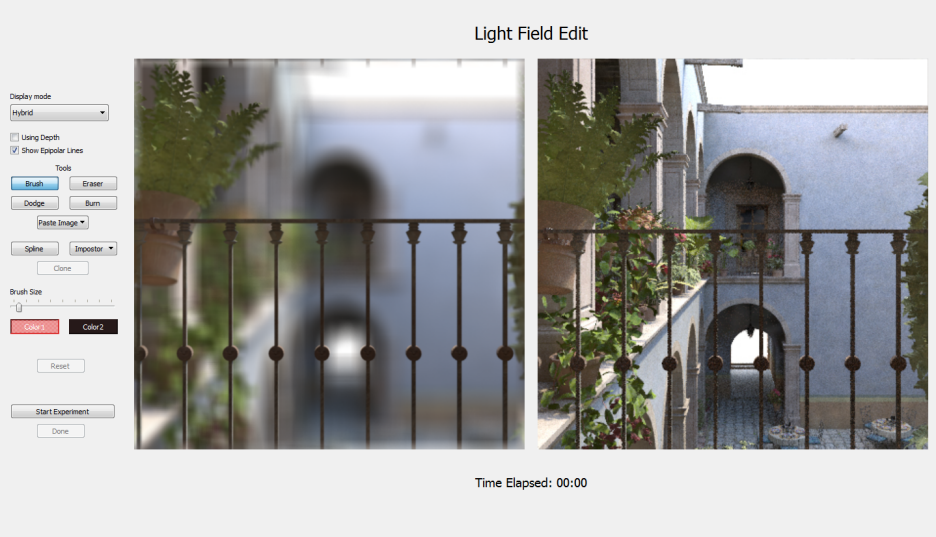


Figure J.1: Screen-shot of our hybrid interface. On the left window (*W1*), the light field is shown using the *focus* paradigm, while on the right window (*W2*) it is depicted with *multiview*.

Last, Figures J.2, J.3 and J.4 show some sample views of light fields edited with this hybrid interface.

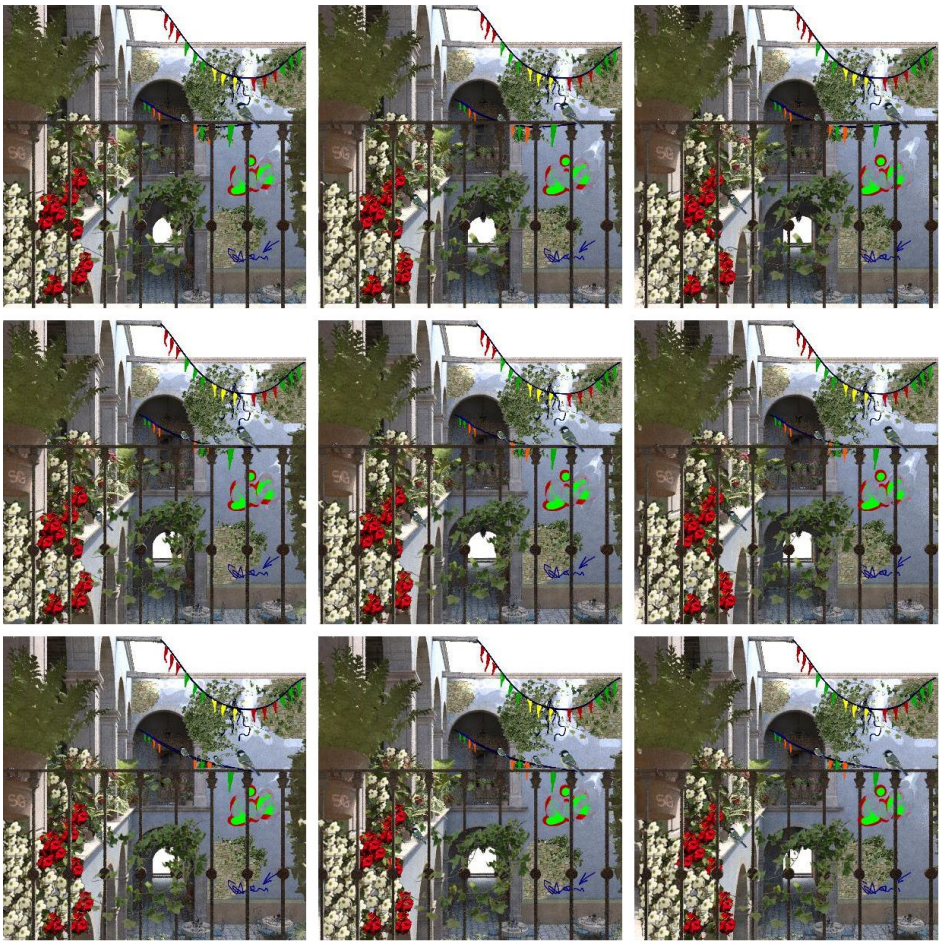


Figure J.2: Sample views of an edited *San Miguel* light field.

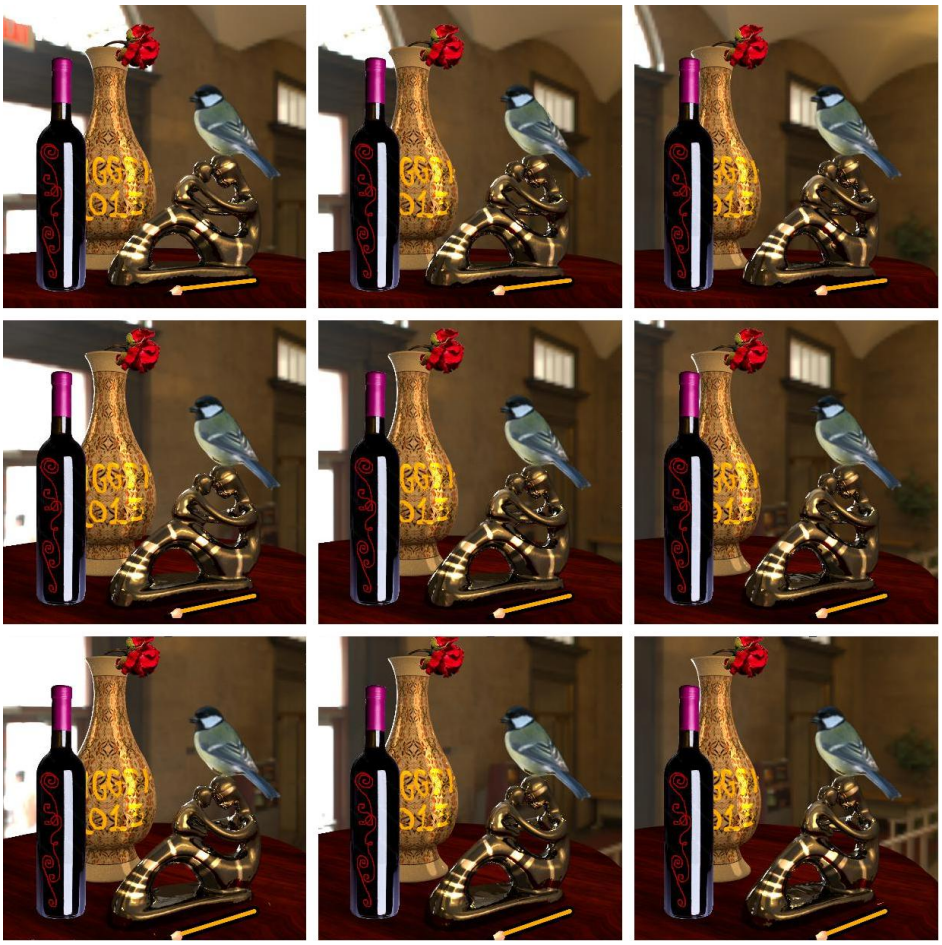


Figure J.3: Sample views of an edited *Vase* light field.

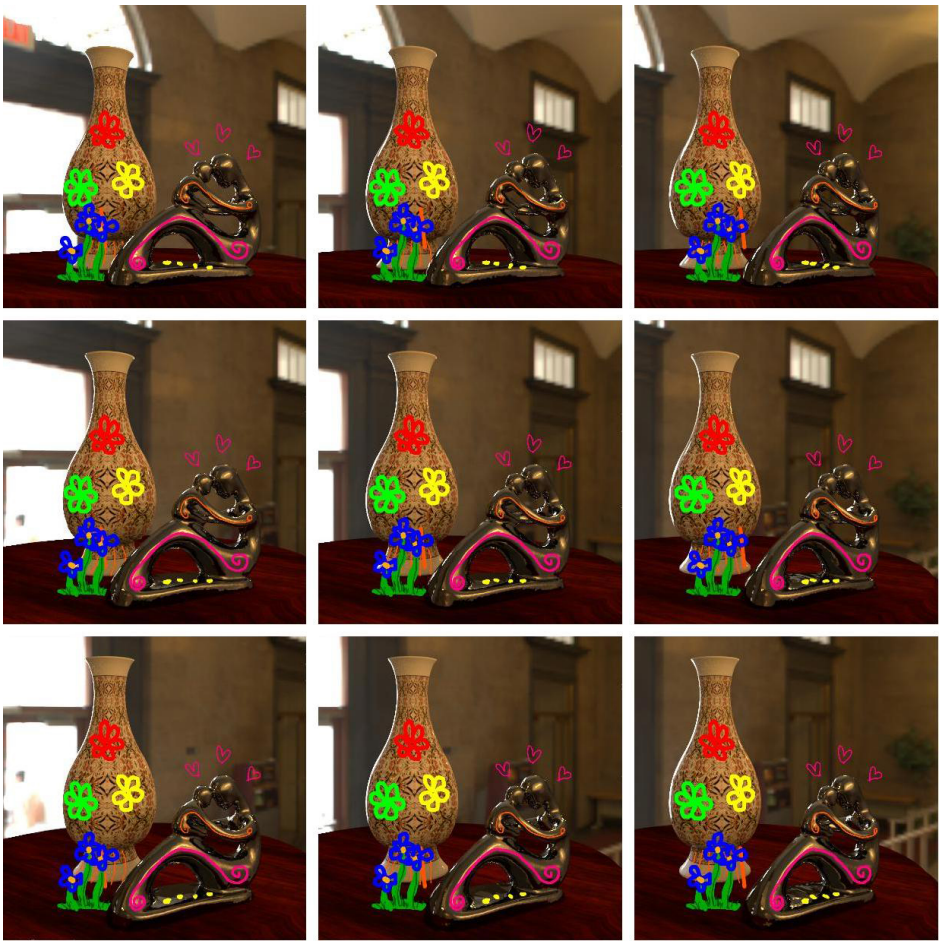


Figure J.4: Sample views of the *Vase* light field edited by a different user.

K

LIGHT FIELD EDITING INTERFACES: ADDITIONAL DATA FROM THE ANALYSIS OF INTERACTION PARADIGMS

K.1 ERROR IN DEPTH

Table K.1 shows pairwise comparisons (p-value) for the error in depth of each of the five *directed* tasks [T1..T5]. For the results of the ANOVA see also Table 1 in the main text. A p-value $\leqslant 0.05$ (marked with a star (*)) indicates the difference between interfaces is significant. Additionally, in Figure K.1 we plot 95% confidence intervals for the difference of the mean between each pair of interfaces. Confidence intervals also show significance (if the interval contains zero, then the difference between the compared interfaces is not significant), but additionally they give an idea of the magnitude of the difference. Since confidence intervals are symmetric for each pair of interfaces (e.g. between $M - F$ and $F - M$ only the sign of the interval changes) we only show half of the pairwise comparisons.

Table K.1: Significance of pairwise comparisons for error in depth in directed tasks.

	M	***MD***	***F***	***FD***
M	-	0.000*	0.018*	0.000*
MD	0.000*	-	0.000*	-
F	0.018*	0.000*	-	0.000*
FD	0.000*	-	0.000*	-

(a) Task 1

	M	***MD***	***F***	***FD***
M	-	0.000*	0.000*	0.000*
MD	0.000*	-	0.000*	-
F	0.000*	0.000*	-	0.000*
FD	0.000*	-	0.000*	-

(b) Task 2

	M	***MD***	***F***	***FD***
M	-	0.000*	0.000*	0.000*
MD	0.000*	-	0.000*	-
F	0.000*	0.000*	-	0.000*
FD	0.000*	-	0.000*	-

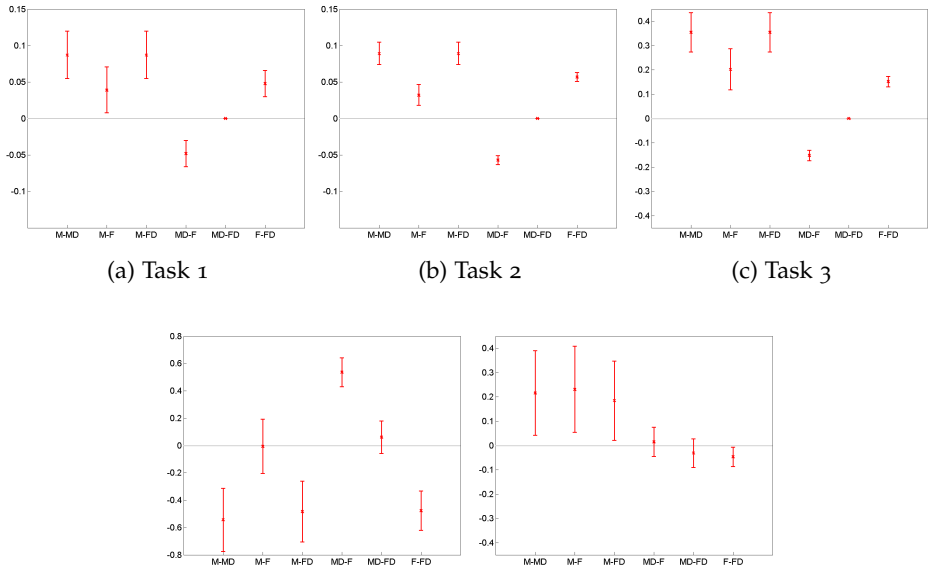
(c) Task 3

	M	***MD***	***F***	***FD***
M	-	0.000*	0.951	0.000*
MD	0.000*	-	0.000*	0.296
F	0.951	0.000*	-	0.000*
FD	0.000*	0.296	0.000*	-

(d) Task 4

	M	***MD***	***F***	***FD***
M	-	0.018*	0.014*	0.028*
MD	0.018*	-	0.606	0.285
F	0.014*	0.606	-	0.024*
FD	0.028*	0.285	0.024*	-

(e) Task 5



(a) Task 1 (b) Task 2 (c) Task 3

(d) Task 4 (e) Task 5

Figure K.1: Confidence intervals at 95% for mean difference of error in depth between interfaces for Tasks 1 to 5.

K.2 TIME TO COMPLETION

We provide in Table K.2 the results of the repeated measures ANOVA performed on the time to completion, from which the main text only reports mean values and significant differences. The table contains the H-test, the between-groups degrees of freedom df_1 (three unless the Greenhouse-Geisser correction is applied because sphericity is violated), the within-groups degrees of freedom df_2, the associated p-value, and the value of the partial eta-squared η^2 for each task, indicative of the proportion of variance that can be attributed to the *interface* factor. Table K.3 contains the pairwise comparisons (p-value) for the time to completion in each of the five *directed* tasks [T1..T5]. A p-value $\leqslant 0.05$ (marked with *) indicates significant difference. Additionally, in Figure K.2 we plot 95% confidence intervals for the difference of the mean between each pair of interfaces (see Section K.1 for details on confidence intervals).

Table K.2: ANOVA results for time to completion in directed tasks.

	T1	***T2***	***T3***	***T4***	***T5***
H	2.048	6.730	5.431	3.986	9.175
(df_1, df_2)	(2.080,35.364)	(3,54)	(1.815,32.669)	(3,48)	(3,54)
p	0.142	0.001*	0.011*	0.013*	0.000*
η^2 (%)	10.8	27.2	23.2	19.9	33.8

Table K.3: Significance of pairwise comparisons for time to completion in directed tasks.

	M	*MD*	*F*	*FD*
M	-	0.293	0.104	0.007*
MD	0.293	-	0.977	0.367
F	0.104	0.977	-	0.115
FD	0.007*	0.367	0.115	-

(a) Task 1

	M	*MD*	*F*	*FD*
M	-	0.001*	0.093	0.006*
MD	0.001*	-	0.061	0.402
F	0.093	0.061	-	0.062
FD	0.006*	0.402	0.062	-

(b) Task 2

	M	*MD*	*F*	*FD*
M	-	0.002*	0.386	0.008*
MD	0.002*	-	0.063	0.850
F	0.386	0.063	-	0.004*
FD	0.008*	0.850	0.004*	-

(c) Task 3

	M	*MD*	*F*	*FD*
M	-	0.056	0.008*	0.068
MD	0.056	-	0.131	0.814
F	0.008*	0.131	-	0.294
FD	0.068	0.814	0.294	-

(d) Task 4

	M	*MD*	*F*	*FD*
M	-	0.050*	0.003*	0.052
MD	0.050*	-	0.000*	0.004*
F	0.003*	0.000*	-	0.590
FD	0.052	0.004*	0.590	-

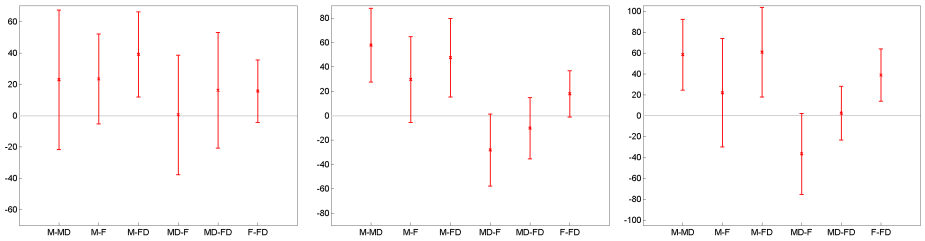
(e) Task 5

K.3 RATINGS

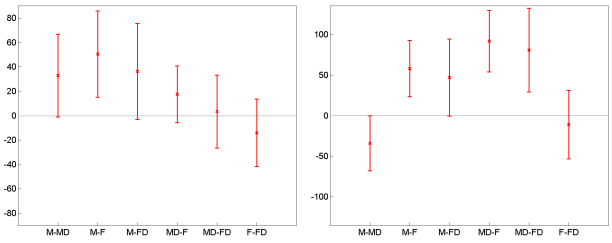
Users were asked to rate their preferences in *directed* tasks (T1..T5), overall preference, and general aspects on a scale [1..5][1]. Mean ratings for directed tasks and for overall preference can be found in the main text (Figure 8), while here in Figure K.3 we show the mean values for the questions on general aspects.

Next we provide the results of the repeated measures ANOVA performed on the ratings, from which the main text only reports which differences between interfaces are significant. Tables K.4 and K.5 provide the H-test, degrees of freedom, and its associated significance p. The between-groups degrees of freedom are three in all cases, since we have four interfaces and sphericity can be assumed, while the within-group degrees of freedom are 57 in all cases. Additionally, we include the partial eta-squared η^2 for each case, and the significance results (p-value) of the pairwise comparisons in Table K.6 (we found no significant difference for *accuracy*, see Table K.5). A p-value $\leqslant 0.05$ (marked with *) indicates the difference between interfaces is significant.

1 The exact questions can be found in `http://webdiis.unizar.es/~bmasia/downloads/thesis/LFEI_Questionnaires.pdf`



(a) Task 1 (b) Task 2 (c) Task 3



(d) Task 4 (e) Task 5

Figure K.2: Confidence intervals at 95% for mean difference in time to completion between interfaces for Tasks 1 to 5.

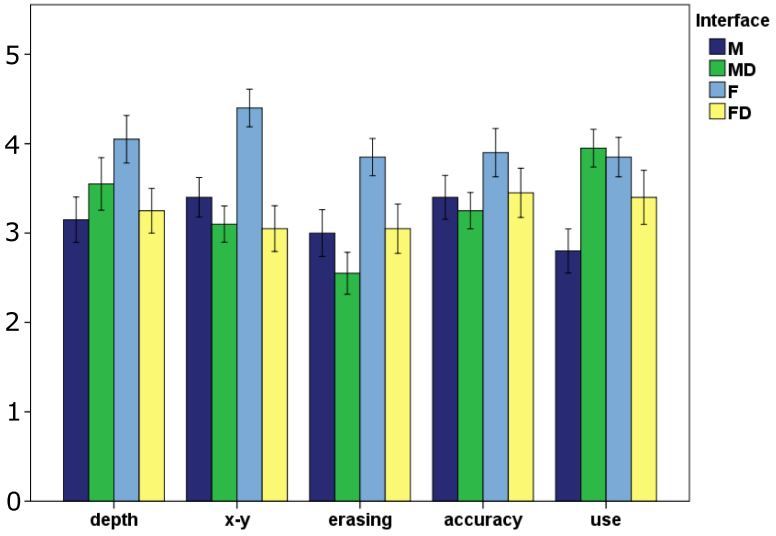


Figure K.3: Mean ratings for each interface for questions on general aspects asked in final questionnaire.

Table K.4: ANOVA results for ratings in final questionnaire (I).

	T1	*T2*	*T3*	*T4*	*T5*	*overall*
$H(3,57)$	7.410	9.251	13.203	12.390	6.218	2.217
p	0.000*	0.000*	0.000*	0.000*	0.001*	0.096
η^2 (%)	28.1	32.7	41.0	39.5	24.7	10.4

Table K.5: ANOVA results for ratings in final questionnaire (and II).

	depth	*x-y*	*erasing*	*accuracy*	*difficulty*
$H(3,57)$	2.053	8.456	4.180	1.119	3.943
p	0.117	0.000*	0.010*	0.349	0.013*
η^2 (%)	9.8	30.8	18.0	5.6	17.2

K.4 RANKINGS

Similarly, users ranked preferences in directed tasks (T1..T5), overall preference, and general aspects[2]. Rankings for preferences in directed tasks and for overall preference can be found in the main text (Figure 7), while here in Figure K.4 we show the ranks for the questions on general aspects.

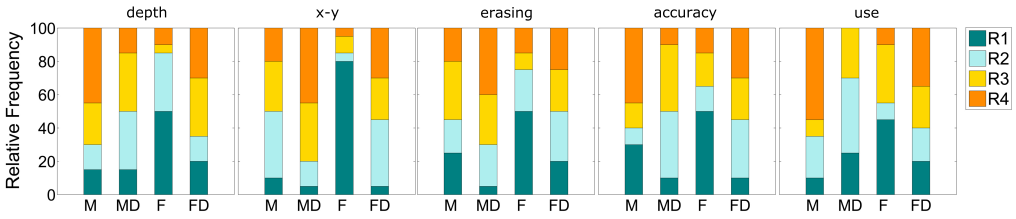


Figure K.4: Rankings for each interface for questions on general aspects asked in final questionnaire.

We provide here the results of the Kruskal-Wallis test performed on the rankings, from which the main text only reports which differences between interfaces are significant. Tables K.7 and K.8 provide the test statistic χ^2, its degrees of freedom (three in all cases, since we have four interfaces) and its associated significance p. We also include the significance results of the pairwise comparisons in Table K.9 (we found no significant difference for *accuracy*, see Table K.8). A p-value $\leqslant 0.05$ (marked *) indicates the difference between interfaces is significant.

For each ranking obtained in each question, we obtain the rank product per interface $\Psi(\vartheta)$ (see main text for details on computation). This rank product is used when sorting the interfaces according to the rankings received. In Table K.10 we include all the rank products per interface per question, highlighting in bold the highest ranked.

K.5 WORKFLOW IN OPEN TASKS

In Table K.11 we show the number of times users switched from one interface to another in Task 7, in which they can freely switch between any of the four interfaces at any time. We show the sum for all subjects. Note that, due to how menus were implemented, users did not select one of four interfaces, but switched between *multiview* and *focus* paradigms, and between depth on or off (there are eight possible interface switches). The high number of switches between M and

2 The exact questions can be found in `http://webdiis.unizar.es/~bmasia/downloads/LFEI_Questionnaires.pdf`

F supports the findings reported in the main text: the preferred workflow was to edit mostly in *F*, then switch to *M* for visualization. To illustrate the workflow of users in this task, we include in Figures K.5, K.6 and K.7 timelines for each subject showing which interface the subject is using and what for (*drawing*, *erasing*, *changing view* or *adjusting depth*). In these figures, in some cases users appear to be *adjusting depth* while interfaces with depth (FD or MD) are activated: this is due to users accidentally touching the *adjusting depth* controls (mouse wheel or equivalent in tablet-pen device); when computing median times for Figure 10 in the main text these spurious times were removed from the computation.

Table K.6: Significance of pairwise comparisons for ratings in final questionnaire.

	M	*MD*	*F*	*FD*
M	-	0.000*	0.025*	0.001*
MD	0.000*	-	0.384	0.494
F	0.025*	0.384	-	0.053
FD	0.001*	0.494	0.053	-

(a) Task 1

	M	*MD*	*F*	*FD*
M	-	0.000*	0.541	0.019*
MD	0.000*	-	0.001*	0.077
F	0.541	0.001*	-	0.016*
FD	0.019*	0.077	0.016*	-

(b) Task 2

	M	*MD*	*F*	*FD*
M	-	0.000*	0.815	0.014*
MD	0.000*	-	0.000*	0.045*
F	0.815	0.000*	-	0.000*
FD	0.014*	0.045*	0.000*	-

(c) Task 3

	M	*MD*	*F*	*FD*
M	-	0.053	0.026*	0.014*
MD	0.053	-	0.000*	0.107
F	0.026*	0.000*	-	0.000*
FD	0.014*	0.107	0.000*	-

(d) Task 4

	M	*MD*	*F*	*FD*
M	-	0.036*	0.035*	0.270
MD	0.036*	-	0.001*	0.618
F	0.035*	0.001*	-	0.002*
FD	0.270	0.618	0.002*	-

(e) Task 5

	M	*MD*	*F*	*FD*
M	-	0.131	0.275	0.566
MD	0.131	-	0.614	0.047*
F	0.275	0.614	-	0.044*
FD	0.566	0.047*	0.044*	-

(f) Overall

	M	*MD*	*F*	*FD*
M	-	0.338	0.041*	0.823
MD	0.338	-	0.220	0.410
F	0.041*	0.220	-	0.049*
FD	0.823	0.410	0.049*	-

(g) Depth Positioning

	M	*MD*	*F*	*FD*
M	-	0.316	0.001*	0.309
MD	0.316	-	0.001*	0.871
F	0.001*	0.001*	-	0.000*
FD	0.309	0.871	0.000*	-

(h) x-y Positioning

	M	*MD*	*F*	*FD*
M	-	0.154	0.034*	0.910
MD	0.154	-	0.003*	0.234
F	0.034*	0.003*	-	0.022*
FD	0.910	0.234	0.022*	-

(i) Erasing

	M	*MD*	*F*	*FD*
M	-	0.002*	0.015*	0.219
MD	0.002*	-	0.785	0.102
F	0.015*	0.785	-	0.216
FD	0.219	0.102	0.216	-

(j) Difficulty of Use

Table K.7: Kruskal-Wallis results for rankings in final questionnaire (I).

	T1	*T2*	*T3*	*T4*	*T5*	*overall*
$\chi^2(3)$	26.149	14.931	35.313	22.357	11.455	9.006
p	0.000*	0.001*	0.000*	0.000*	0.008*	0.028*

Table K.8: Kruskal-Wallis results for rankings in final questionnaire (and II).

	depth	*x-y*	*erasing*	*accuracy*	*difficulty*
$\chi^2(3)$	13.825	28.440	10.507	5.925	12.403
p	0.002*	0.000*	0.014*	0.116	0.005*

Table K.9: Significance of pairwise comparisons for rankings in final questionnaire.

	M	***MD***	***F***	***FD***
M	-	0.000*	0.025*	0.000*
MD	0.000*	-	0.206	0.160
F	0.025*	0.206	-	0.008*
FD	0.000*	0.160	0.008*	-

(a) Task 1

	M	***MD***	***F***	***FD***
M	-	0.001*	0.673	0.011*
MD	0.001*	-	0.005*	0.482
F	0.673	0.005*	-	0.035*
FD	0.011*	0.482	0.035*	-

(b) Task 2

	M	***MD***	***F***	***FD***
M	-	0.000*	0.482	0.002*
MD	0.000*	-	0.000*	0.206
F	0.482	0.000*	-	0.000*
FD	0.002*	0.206	0.000*	-

(c) Task 3

	M	***MD***	***F***	***FD***
M	-	0.122	0.035*	0.025*
MD	0.122	-	0.000*	0.482
F	0.035*	0.000*	-	0.000*
FD	0.025*	0.482	0.000*	-

(d) Task 4

	M	***MD***	***F***	***FD***
M	-	0.122	0.122	0.261
MD	0.122	-	0.002*	0.673
F	0.122	0.002*	-	0.008*
FD	0.261	0.673	0.008*	-

(e) Task 5

	M	***MD***	***F***	***FD***
M	-	0.011*	0.160	1.000
MD	0.011*	-	0.261	0.011*
F	0.160	0.261	-	0.160
FD	1.000	0.011*	0.160	-

(f) Overall

	M	***MD***	***F***	***FD***
M	-	0.160	0.000*	0.482
MD	0.160	-	0.035*	0.482
F	0.000*	0.035*	-	0.005*
FD	0.482	0.482	0.005*	-

(g) Depth Positioning

	M	***MD***	***F***	***FD***
M	-	0.092	0.001*	0.574
MD	0.092	-	0.000*	0.261
F	0.001*	0.000*	-	0.000*
FD	0.574	0.261	0.000*	-

(h) x-y Positioning

	M	***MD***	***F***	***FD***
M	-	0.122	0.092	0.888
MD	0.122	-	0.001*	0.160
F	0.092	0.001*	-	0.068
FD	0.888	0.160	0.068	-

(i) Erasing

	M	***MD***	***F***	***FD***
M	-	0.003*	0.005*	0.325
MD	0.003*	-	0.888	0.049*
F	0.005*	0.888	-	0.068
FD	0.325	0.049*	0.068	-

(j) Difficulty of Use

Table K.10: Rank products per interface for rank scores on final questionnaire.

	T1	*T2*	*T3*	*T4*	*T5*	*overall*	*depth*	*x-y*	*erasing*	*accuracy*	*difficulty*
M	3.37	2.85	2.90	2.16	2.17	2.59	2.72	2.42	2.23	2.36	2.85
MD	1.99	**1.66**	**1.39**	2.72	2.70	**1.71**	2.30	3.04	2.88	2.35	1.90
F	2.34	2.66	3.28	**1.37**	**1.64**	2.07	**1.55**	**1.24**	**1.63**	**1.70**	**1.81**
FD	**1.52**	1.90	1.81	2.98	2.49	2.62	2.47	2.63	2.29	2.54	2.46

Table K.11: Switching between interfaces in Task 7.

	F→M	*M→F*	*F→FD*	*FD→F*	*M→MD*	*MD→M*	*FD→MD*	*MD→FD*
$N_{switches}$	58	52	29	27	25	15	18	31

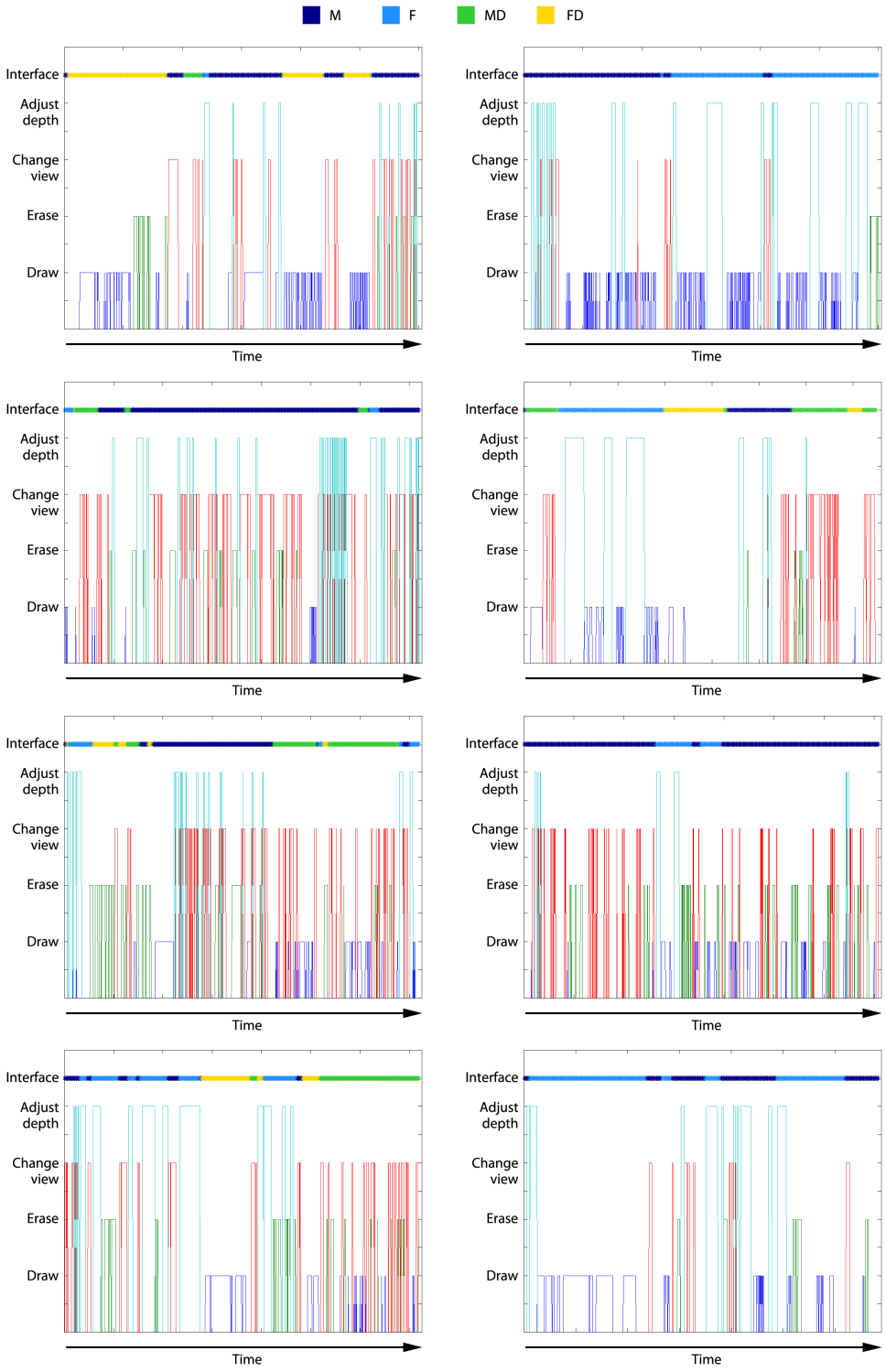


Figure K.5: Workflow for Task 7, subjects 1-8.

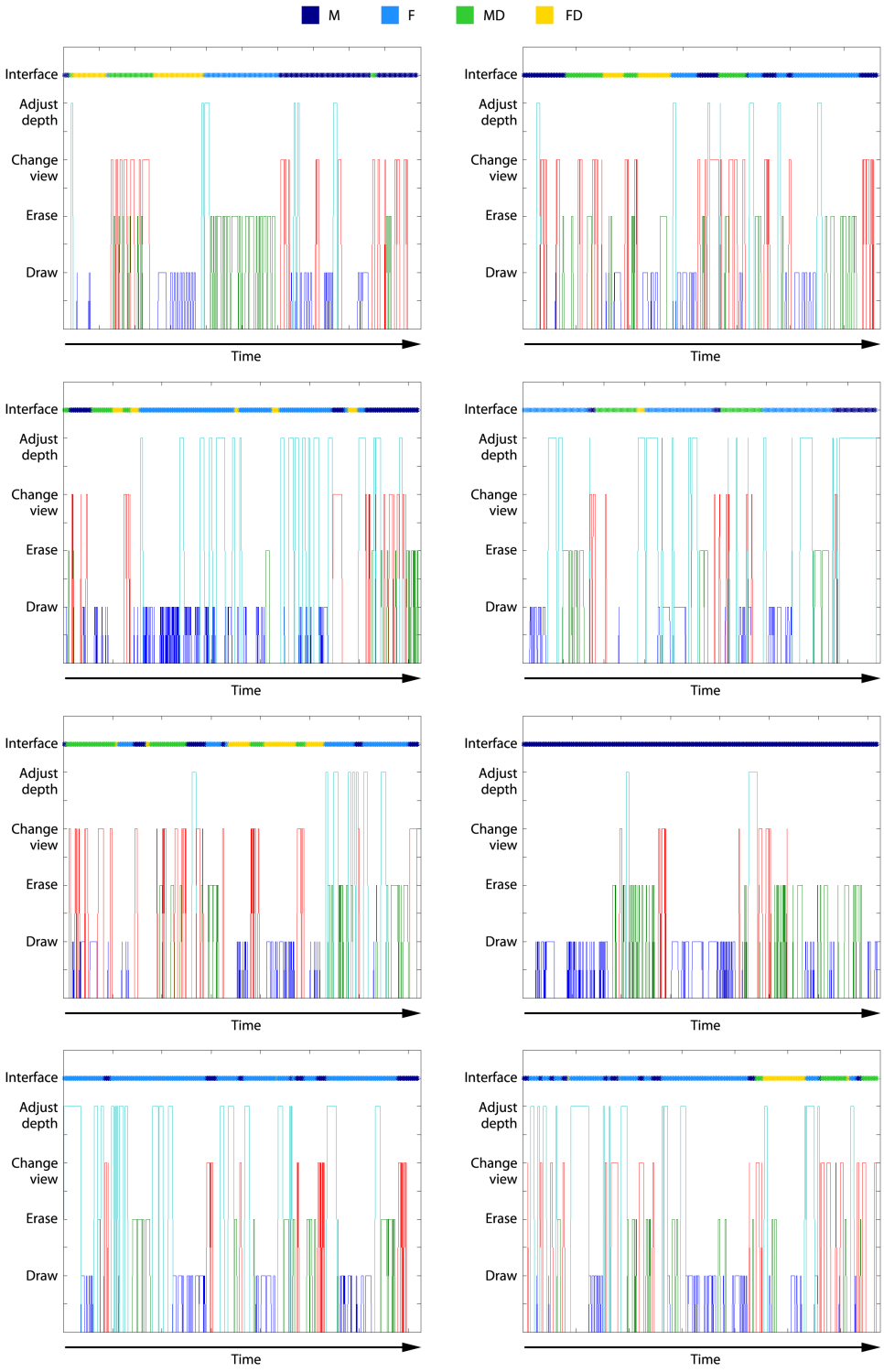


Figure K.6: Workflow for Task 7, subjects 9-16.

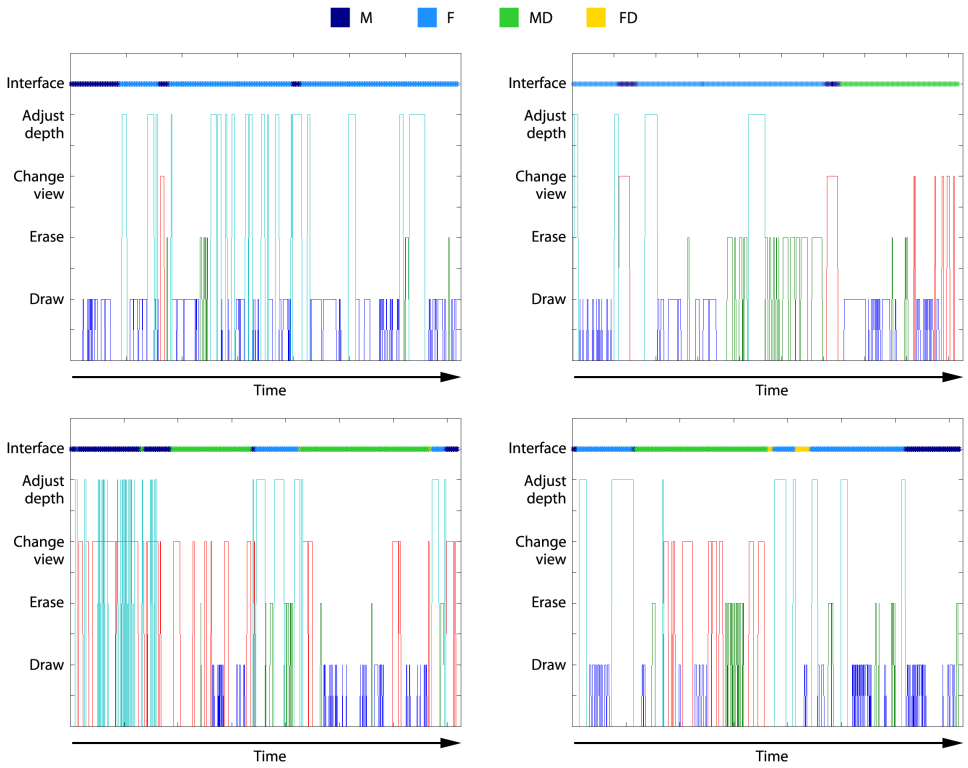


Figure K.7: Workflow for Task 7, subjects 17-20.

L

LIGHT FIELD EDITING INTERFACES: LIGHT FIELDS USED IN THE EVALUATION

We include representative views of the light fields used in the experiments in Figures L.1, L.2 and L.3. All light fields have 17×17 views; we only show here 5×5 views, obtained from uniformly sampling the light field in both dimensions.

For the training session we employ the *Teapots* light field, shown in Figure L.4. The choice is motivated by the following: it has clear texture on both floor and wall, facilitating the explanation of both the focus and multiview paradigm; it contains curved and flat surfaces, the latter both parallel to the camera and slanted; and it has a number of occlusions of varying degree.



Figure L.1: Sample views of the *San Miguel* light field.

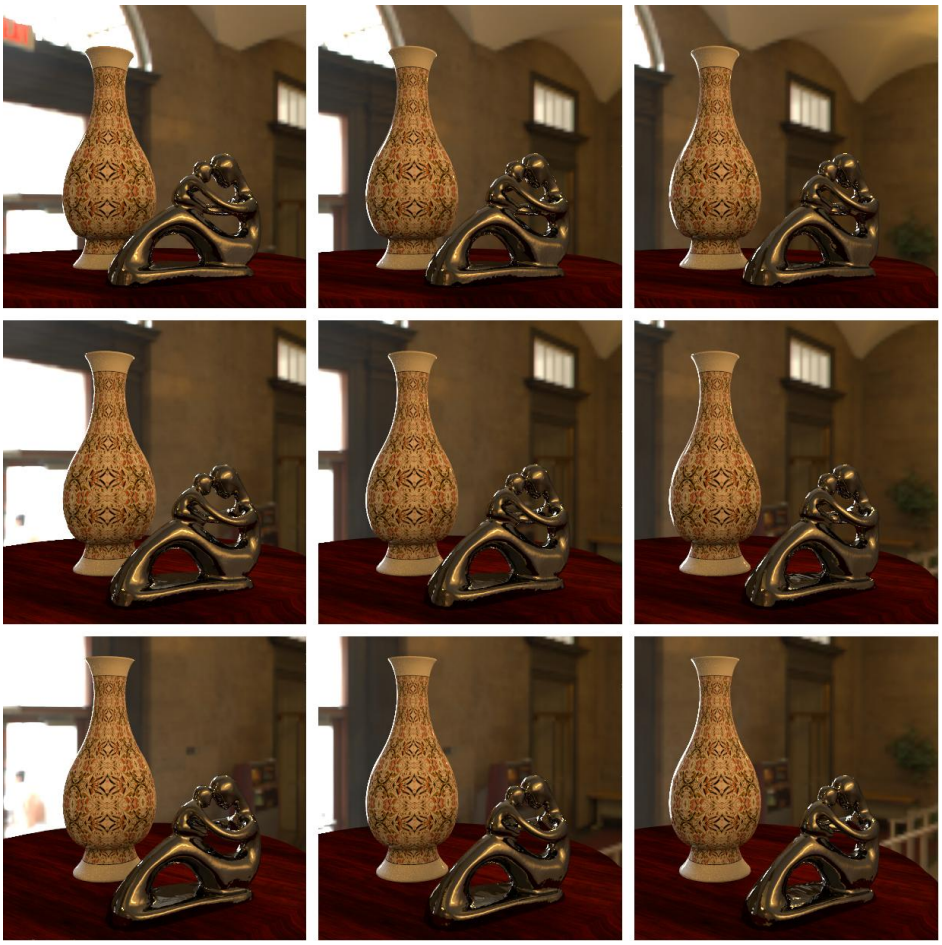


Figure L.2: Sample views of the *Vase* light field.

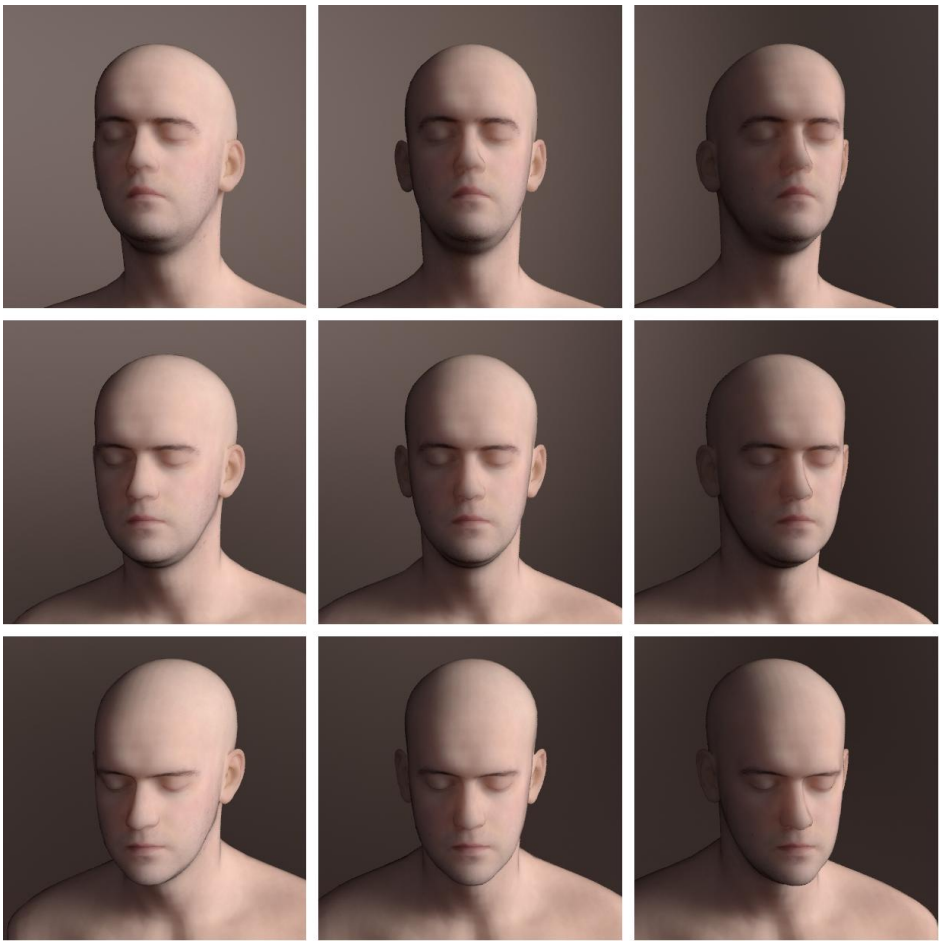


Figure L.3: Sample views of the *Head* light field.

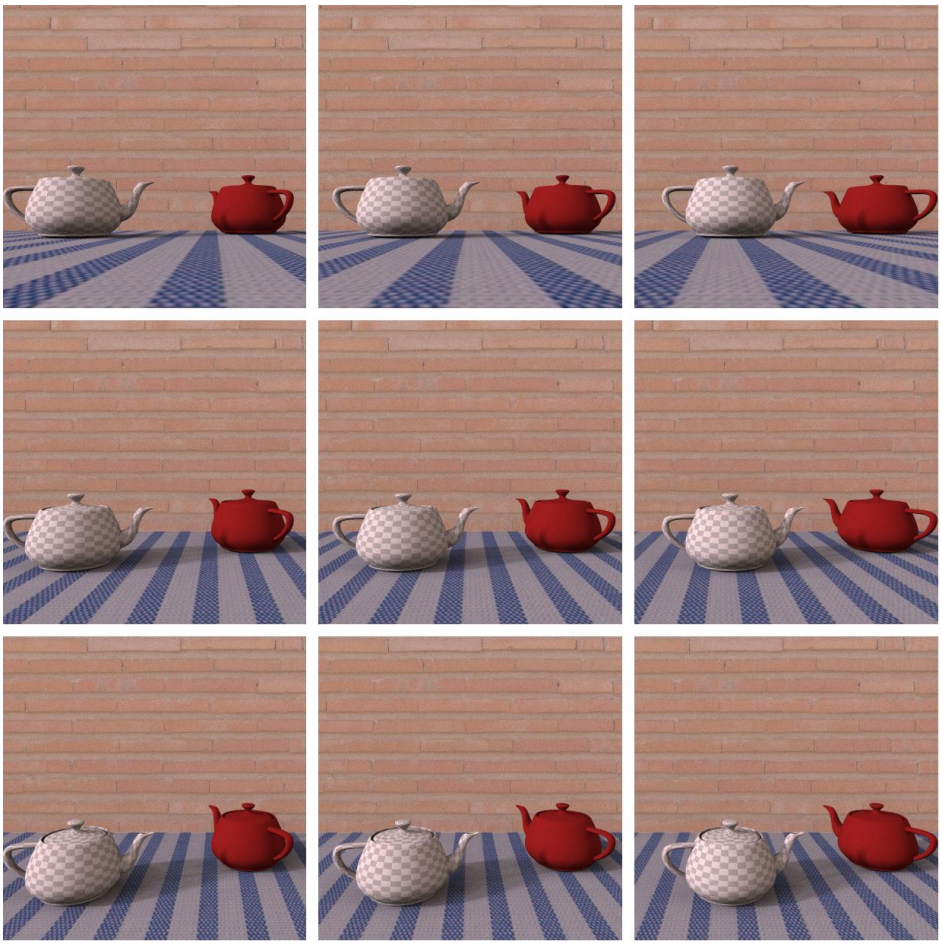


Figure L.4: Sample views of the *Teapots* light field.

I have my mind... And a mind needs books as a sword needs a whetstone, if it is to keep its edge.
— Tyrion Lannister to Jon Snow.
A Game of Thrones, by George R. R. Martin.

BIBLIOGRAPHY

[1] Nils Abramson. Light-in-flight recording by holography. *Optics Letters*, 3(4):121–123, 1978.

[2] Andrew Adams, Eino-Ville Talvala, Sung Hee Park, David E. Jacobs, Boris Ajdin, Natasha Gelfand, Jennifer Dolson, Daniel Vaquero, Jongmin Baek, Marius Tico, Hendrik P. A. Lensch, Wojciech Matusik, Kari Pulli, Mark Horowitz, and Marc Levoy. The frankencamera: an experimental platform for computational photography. *ACM Trans. Graph.*, 29(4):29:1–12, July 2010.

[3] Ansel Adams. *The Print*. The Ansel Adams Photography series. Little, Brown and Company, 1983.

[4] Edward H. Adelson. Image statistics and surface perception. In *Human Vision and Electronic Imaging XIII, Proceedings of the SPIE*, number 1. SPIE, 2008.

[5] Edward H. Adelson. Checkershadow Illusion. `http://persci.mit.edu/gallery/checkershadow`, 2013.

[6] E.H. Adelson and J.R. Bergen. The plenoptic function and the elements of early vision. *Computational models of visual processing*, 1:3–20, 1991.

[7] Agocs et al. A large scale interactive holographic display. In *IEEE Virtual Reality*, pages 311–312, 2006.

[8] Aharon J. Agranat, Alexander Gumennik, and Harel Ilan. Refractive index engineering by fast ion implantations: a generic method for constructing multi-components electro-optical circuits. 76040Y: 1–17, 2010.

[9] Takeyuki Ajito, Takashi Obi, Masahiro Yamaguchi, and Nagaaki Ohyama. Expanded color gamut reproduced by six-primary projection display. 3954:130–137, 2000.

[10] Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. A stereo display prototype with multiple focal distances. *ACM Trans. Graph. (SIGGRAPH)*, 23:804–813, 2004.

[11] A. O. Akyüz and E. Reinhard. Color appearance in high dynamic range imaging. *SPIE Journal of Electronic Imaging*, 15(3):033001–1–12, 2006.

[12] Ahmet Oğuz Akyüz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H. Bülthoff. Do HDR displays support LDR content?: a psychophysical evaluation. *ACM Trans. Graph.*, 26(3), July 2007.

[13] Daniel G. Aliaga, Yu Hong Yeung, Alvin Law, Behzad Sajadi, and Aditi Majumder. Fast high-resolution appearance editing using superimposed projections. *ACM Trans. Graph.*, 31(2):13:1–13, April 2012.

[14] W. Allen and R. Ulichney. Wobulation: Doubling the addressed Resolution of Projection Displays. In *Proc. SID 47*, 2005.

[15] M. Alonso Jr. and A. B. Barreto. Pre-compensation for high-order aberrations of the human eye using on-screen image deconvolution. In *IEEE Engineering in Medicine and Biology Society*, volume 1, pages 556–559, 2003.

[16] Barton L. Anderson. Stereovision: beyond disparity computations. *Trends in Cognitive Sciences*, 2:214–222, 1998.

[17] Hyrum Anderson, Eric Garcia, and Maya Gupta. Gamut expansion for video and image sets. In *Proceedings of the 14th International Conference of Image Analysis and Processing - Workshops*, ICIAPW '07, pages 188–191, Washington, DC, USA, 2007. IEEE Computer Society.

[18] S.M. Anstis, I.P. Howard, and B. Rogers. A Craik-O'Brien-Cornsweet illusion for visual depth. *Vision Research*, 18(2):213–217, 1978.

[19] Pierre Archand, Eric Pite, Herve Guillemet, and Loic Trocme. Systems and methods for rendering a display to compensate for a viewer's visual impairment. International Patent Application PCT/US2011/039993, 2011.

[20] A. D. Arnold, P. E. Castro, T. K. Hatwar, M. V. Hettel, P. J. Kane, J. E. Ludwicki, M. E. Miller, M. J. Murdoch, J. P. Spindler, S. A. Van Slyke, K. Mameno, R. Nishikawa, T. Omura, and S. Matsumoto. Full-color amoled with rgbw pixel pattern. *Journal of the Society for Information Display*, 13(6):525–535, 2005.

[21] James Arvo. Transfer equations in global illumination. In *Global Illumination, SIGGRAPH ï¿œ93 Course Notes*, 1993.

[22] M. Ashdown, T. Okabe, I. Sato, and Y. Sato. Robust content-dependent photometric projector compensation. In *IEEE International Workshop on Projector-Camera Systems (PROCAMS)*, 2006.

[23] Philipp R. Aumayr. Stereopsis in the Context of High Dynamic Range Stereo Displays. Master Thesis. Johannes Kepler Universität Linz, Germany, 2012.

[24] Tunç Ozan Aydin, Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Dynamic range independent image quality assessment. *ACM Trans. Graph*, 27(3):69, 2008.

[25] Tunç Aydın. *Human Visual System Models in Computer Graphics*. PhD thesis, Max Planck Institute for Computer Science, 2010.

[26] Tunç O. Aydın, Martin Čadík, Karol Myszkowski, and Hans-Peter Seidel. Video quality assessment for computer graphics applications. In *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, volume 29, pages 161:1–161:12, 2010.

[27] Soonmin Bae, Sylvain Paris, and Frédo Durand. Two-scale tone management for photographic look. *ACM Trans. Graph.*, 25(3):637–645, 2006.

[28] A. T. Bahill and L. Stark. Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Exp Neurol*, 48(1):95–106, 1975.

[29] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1167–1183, September 2002.

[30] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Inverse tone mapping. In *Proc. of the 4th Intnl. Conf. on Comp. Graph. and Interactive Tech. in Australasia and Southeast Asia*, GRAPHITE '06, pages 349–356, 2006.

[31] Francesco Banterle, Patrick Ledda, Kurt Debattista, Alan Chalmers, and Marina Bloj. A framework for inverse tone mapping. *Vis. Comput.*, 23(7):467–478, 2007.

[32] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Expanding low dynamic range videos for high dynamic range applications. In *Proc. of the Spring Conference on Computer Graphics*, New York, NY, USA, 2008. ACM.

[33] Francesco Banterle, Kurt Debattista, Alessandro Artusi, Sumanta Pattanaik, Karol Myszkowski, Patrick Ledda, Marina Bloj, and Alan Chalmers. High dynamic range imaging and LDR expansion for generating HDR content. Annex Eurographics 2009, April 2009.

[34] Francesco Banterle, Patrick Ledda, Kurt Debattista, Marina Bloj, Alessandro Artusi, and Alan Chalmers. A psychophysical evaluation of inverse tone mapping techniques. *Comput. Graph. Forum*, 28(1):13–25, 2009.

[35] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Expanding low dynamic range videos for high dynamic range applications. In *Proceedings of the 24th Spring Conference on Computer Graphics*, SCCG '08, pages 33–41, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-957-2.

[36] Francesco Banterle, Alessandro Artusi, Tunc Aydin, Piotr Didyk, Elmar Eisemann, Diego Gutierrez, Rafal Mantiuk, and Karol Myszkowski. Multidimensional image retargeting. ACM SIGGRAPH Asia Course Notes, 2011.

[37] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice*. A.K. Peters, Ltd., 2011.

[38] Francesco Banterle, Alessandro Artusi, Tunc Aydin, Piotr Didyk, Elmar Eisemann, Diego Gutierrez, Rafal Mantiuk, and Tobias Ritschel. Mapping images to target devices: spatial, temporal, stereo, tone, and color. In *Eurographics 2012 Tutorials*, May 2012.

[39] Francesco Banterle, Alan Chalmers, and Roberto Scopigno. Real-time high fidelity inverse tone mapping for low dynamic range content. In Olga Sorkine Miguel A. Otaduy, editor, *Eurographisc 2013 Short Papers*. Eurographics, Eurographics, May 2013.

[40] M. Barkowsky, J. Bialkowski, Bjorn Eskofier, R. Bitto, and A. Kaup. Temporal trajectory aware video quality measure. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):266–279, 2009.

[41] Peter C. Barnum, Srinivasa G. Narasimhan, and Takeo Kanade. A multi-layered display with water drops. *ACM Trans. Graph.*, 29:1–7, 2010.

[42] Peter G. J. Barten. *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*. SPIE Press, 1999.

[43] S. Basu and P. Baudisch. System and process for increasing the apparent resolution of a display. US Patent 7548662, 2009.

[44] M. Ben-Ezra, A. Zomet, and S.K. Nayar. Jitter camera: high resolution video from a low resolution detector. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages 135–142, 2004.

[45] Benton SA (editor). *Selected Papers on Three-Dimensional Displays*. SPIE Press, 2001.

[46] J. L. Benzschawel and W. E. Howard. Method of and apparatus for displaying a multicolor image. U.S. Patent 5341153, 1994.

[47] Floraine Berthouzoz and Raanan Fattal. Resolution Enhancement by Vibrating Displays. *ACM Trans. Graph.*, 31(2):15:1–15:14, 2012.

[48] O. Bimber and R. Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters/CRC Press, 2005. ISBN 978-1568812304.

[49] Oliver Bimber and Daisuke Iwai. Superimposing dynamic range. *ACM Trans. Graph.*, 27(5), December 2008.

[50] Oliver Bimber, Daisuke Iwai, Gordon Wetzstein, and Anselm Grundhöfer. The visual computing of projector-camera systems. In *ACM SIGGRAPH 2008 classes*, SIGGRAPH '08, pages 84:1–25, New York, NY, USA, 2008. ACM.

[51] C. Birklbauer and O. Bimber. Light-field retargeting. *Computer Graphics Forum*, 31(2):295–303, May 2012.

[52] Barry Blundell and Adam Schwartz. *Volumetric Three-Dimensional Display Systems*. Wiley-IEEE Press, 1999.

[53] Samuel Boivin and Andre Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 107–116, 2001. ISBN 1-58113-374-X.

[54] H. Bowles, K. Mitchell, R.W. Sumner, J. Moore, and M. Gross. Iterative image warping. In *Computer Graphics Forum*, volume 31, pages 237–246. The Eurographics Association and Blackwell Publishing Ltd., 2012.

[55] Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *ICCV*, pages I: 105–112, 2001.

[56] Y. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9): 1124–1137, 2004.

[57] M. F. Bradshaw and B. J. Rogers. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research*, 36(21):3457–3468, 1996.

[58] M F Bradshaw and B J Rogers. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research*, 39(18):3049–56, 1999. ISSN 0042-6989.

[59] Mark F. Bradshaw, Paul B. Hibbard, Andrew D. Parton, David Rose, and Keith Langley. Surface orientation, modulation frequency and the detection and perception of depth defined by binocular disparity and motion parallax. *Vision Research*, 46:2636–2644, 2006.

[60] Marius Braun, Ulrich Leiner, and Detlef Ruschin. Evaluating motion parallax and stereopsis as depth cues for autostereoscopic displays. In *Proc. SPIE 7863, Stereoscopic Displays and Applications XXII*, volume 7863, 2011.

[61] Eli Brenner, Jesus S. Ruiz, Esther M. Herraiz, Frans W. Cornelissen, and Jeroen B.J. Smeets. Chromatic induction an the layout of colours within a complex scene. *Vision Research*, 43:1413–1421, 2003.

[62] M. S. Brennesholtz, S. C. McClain, S. Roth, and D. Malka. A Single Panel LCOS Engine with a Rotating Drum and a Wide Color Gamut. In *SID Digest*, volume 36, pages 1814–1817, 2005.

[63] A. Brookes and K.A. Stevens. The analogy between stereo depth and brightness. *Perception*, 18(5): 601–614, 1989.

[64] Michael Brown, Aditi Majumder, and Ruigang Yang. Camera-based calibration techniques for seamless multiprojector displays. *IEEE Transactions on Visualization and Computer Graphics*, 11(2):193–206, March 2005.

[65] Peter J. Burt and Edward H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. URL `citeseer.ist.psu.edu/burt83laplacian.html`.

[66] J Busck and H Heiselberg. Gated viewing and high-accuracy three-dimensional laser radar. *Applied optics*, 43(24):4705–4710, 2004.

[67] Martin Čadík, Ondrej Hajdok, Antonín Lejsek, Ondřej Fialka, Michael Wimmer, Alessandro Artusi, and Laszlo Neumann. Evaluation of Tone Mapping Operators. `http://dcgi.felk.cvut.cz/home/cadikm/tmo/`, 2013.

[68] A. Campillo and S. Shapiro. Picosecond streak camera fluorometry: a review. *IEEE Journal of Quantum Electronics*, 19(4):585–603, 1987.

[69] Stacey E. Casella, Rodney L. Heckaman, and Mark D. Fairchild. Mapping Standard Image Content to Wide-Gamut Displays. In *Sixteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, pages 106–111, 2008.

[70] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum. Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 307–318, 2000.

[71] Meng-Chou Chang, Feipei Lai, and Wei-Chao Chen. Image shading taking into account relativistic effects. *ACM Trans. Graph.*, 15(4):265–300, October 1996. ISSN 0730-0301.

[72] E. Charbon. Will avalanche photodiode arrays ever reach 1 megapixel? In *International Image Sensor Workshop*, pages 246–249, 2007.

[73] Gaurav Chaurasia, Olga Sorkine-Hornung, and George Drettakis. Silhouette-aware warping for image-based rendering. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering)*, 30 (4), 2011.

[74] Hanfeng Chen, Sung-Soo Kim, Sung-Hee Lee, Oh-Jae Kwon, and Jun-Ho Sung. Nonlinearity compensated smooth frame insertion for motion-blur reduction in LCDs. In *Proceedings of Multimedia Signal Processing, 2005 IEEE 7th Workshop on*, pages 1–4, 2005.

[75] Hui-Chuan Cheng, Ilan Ben-David, and Shin-Tson Wu. Five-Primary-Color LCDs. *J. Display Technol.*, 6 (1):3–7, Jan 2010.

[76] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J. Mitra, Xiaolei Huang, and Shi-Min Hu. Global contrast based salient region detection. In *IEEE CVPR*, pages 409–416, 2011.

[77] E. Chino, K. Tajiri, H. Kawakami, H. Ohira, K. Kamijo, H. Kaneko, S. Kato, Y. Ozawa, T. Kurumisawa, K. Inoue, K. Endo, H. Moriya, T. Aragaki, and K. Murai. Development of Wide-Color-Gamut Mobile Displays with Four-Primary-Color LCDs. *SID Symposium Digest of Technical Papers*, 37(1):1221–1224, 2006.

[78] K Chiu, M Herf, P Shirley, S Swamy, C Wang, and K Zimmerman. Spatially nonuniform scaling functions for high contrast images. In *In Proceedings of Graphics Interface Õ93*, pages 245–253, 1993.

[79] Sang-Hyun Cho and Hang-Bong Kang. Subjective evaluation of visual discomfort caused from stereoscopic 3d video using perceptual importance map. In *TENCON 2012 - 2012 IEEE Region 10 Conference*, pages 1–6, 2012.

[80] A. Colaço, A. Kirmani, G. A. Howland, J. C. Howell, and V. K. Goyal. Compressive depth map acquisition using a single photon-counting detector: Parametric signal processing meets sparsity. In *IEEE Computer Vision and Pattern Recognition, CVPR 2012*, pages 96–102, 2012.

[81] T.F. Coleman and Y. Li. A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6:1040–1058, 1996.

[82] Technicolor Motion Picture (Inventor: Daniel Comstock). Auxiliary registering device for simultaneous projection of two or more pictures. Patent US1208490 (A), 1916.

[83] O. Cossairt, S. K. Nayar, and R. Ramamoorthi. Light Field Transfer: Global Illumination Between Real and Synthetic Objects. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Aug 2008.

[84] O. Cossairt, D. Miau, and S. K. Nayar. Gigapixel Computational Imaging. In *IEEE International Conference on Computational Photography (ICCP)*, Mar 2011.

[85] Oliver S. Cossairt, Joshua Napoli, Samuel L. Hill, Rick K. Dorval, and Gregg E. Favalora. Occlusion-capable multiview volumetric three-dimensional display. *Applied Optics*, 46:1244–1250, 2007.

[86] Douglas Cunningham and Christian Wallraven. *Experimental Design: From User Studies to Psychophysics*. A K Peters/CRC Press, 2011.

[87] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson. Human photoreceptor topography. *The Journal of Comparative Neurology*, 292(4):497–523, 1990.

[88] James E. Cutting and Peter M. Vishton. *Perception of Space and Motion*, chapter Perceiving Layout and Knowing Distances: The integration, relative potency, and contextual use of different information about depth. Academic Press, 1995.

[89] S. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models. In *Human Vision and Electronic Imaging III*, volume 3299 of *SPIE Proceedings Series*, pages 180–191, 1998.

[90] Scott Daly and Xiaofan Feng. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Proc. of Color Imaging VIII: Processing, Hardcopy, and Applications*, volume 5008, 455. SPIE, June 2003.

[91] Scott Daly and Xiaofan Feng. Decontouring: prevention and removal of false contour artifacts. In *Proc. of Human Vision and Electronic Imaging IX*, volume 5292, 130. SPIE, June 2004.

[92] Gerwin Damberg, Helge Seetzen, Greg Ward, Wolfgang Heidrich, and Lorne Whitehead. High dynamic range projection systems. In *SID Digest*, volume 38, pages 4–7, 2007.

[93] N. Damera-Venkata and N. Chang. Display supersampling. *ACM Trans. Graph.*, 28(1):9:1–19, February 2009.

[94] Niranjan Damera-Venkata, Nelson Chang, and Jeffrey Dicarlo. A unified paradigm for scalable multi-projector displays. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1360–1367, November 2007.

[95] Abe Davis, Marc Levoy, and Fredo Durand. Unstructured light fields. *Comp. Graph. Forum*, 31(2pt1): 305–314, May 2012. ISSN 0167-7055. doi: 10.1111/j.1467-8659.2012.03009.x. URL `http://dx.doi.org/10.1111/j.1467-8659.2012.03009.x`.

[96] H. de Lange. Research into the dynamic nature of the human fovea - Cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America*, 48(11):777–783, 1958.

[97] Paul E. Debevec. Image-based lighting. *IEEE Computer Graphics and Applications*, 22(2):26–34, 2002.

[98] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of the 24th annual conference on computer graphics and interactive techniques*, SIGGRAPH '97, pages 369–378, 1997.

[99] K. Devlin, A. Chalmers, A. Wilkie, and W. Purgathofer. STAR Report on Tone Reproduction and Physically Based Spectral Rendering. In *Eurographics 2002*, 2002.

[100] Piotr Didyk. *Perceptual Display: Exceeding Display Limitations by Exploiting the Human Visual System*. PhD thesis, Max-Planck-Institute Informatik, 2012.

[101] Piotr Didyk, Rafal Mantiuk, Matthias Hein, and Hans-Peter Seidel. Enhancement of bright video features for HDR displays. *Computer Graphics Forum*, 27(4):1265–1274, 2008.

[102] Piotr Didyk, Elmar Eiseman, Tobias Ritschel, Karol Myszkowski, and Hans-Heper Seidel. Apparent Resolution Display Enhancement for Moving Images. *ACM Trans. Graph. (SIGGRAPH)*, 29(3), 2010.

[103] Piotr Didyk, Elmar Eisemann, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Perceptually-motivated real-time temporal upsampling of 3D content for high-refresh-rate displays. *Computer Graphics Forum (Proc. Eurographics)*, 29(2):713–722, 2010.

[104] Piotr Didyk, Elmar Eisemann, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Apparent display resolution enhancement for moving images. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2010, Los Angeles)*, 29(4):113:1–8, 2010.

[105] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. Adaptive image-space stereo view synthesis. In *Vision, Modeling and Visualization Workshop*, pages 299–306, Siegen, Germany, 2010.

[106] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. A perceptual model for disparity. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2011, Vancouver)*, 30(4):96:1–96:10, 2011.

[107] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. Apparent stereo: The cornsweet illusion can enhance perceived depth. In *Human Vision and Electronic Imaging XVII, IS&TSPIE's Symposium on Electronic Imaging*, pages 1–12, Burlingame, CA, 2012.

[108] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, Hans-Peter Seidel, and Wojciech Matusik. A luminance-contrast-aware disparity model and applications. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, 31(6):184:1–184:10, 2012.

[109] Piotr Didyk, Pitchaya Sitthi-Amorn, William Freeman, Frédo Durand, and Wojciech Matusik. Joint view expansion and filtering for automultiscopic 3d displays. *ACM Transactions on Graphics (SIGGRAPH Asia))*, 32(6), 2013.

[110] Luat Do, Sveta Zinger, and Peter H. N. de With. Warping error analysis and reduction for depth-image-based rendering in 3dtv. volume 7863, pages 78630B–78630B–9, 2011.

[111] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22(3):419–426, 2003. ISSN 1467-8659.

[112] Song-Pei Du, Belen Masia, Shi-Min Hu, and Diego Gutierrez. A metric of visual comfort for stereoscopic motion. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 32(6), 2013.

[113] M. A. Duguay and A. T. Mattick. Pulsed-image generation and detection. *Applied Optics*, 10:2162–2170, 1971. doi: http://dx.doi.org/10.1364/AO.10.002162.

[114] William Dumouchel and Fanny O'Brien. *Integrating a robust option into a multiple regression computing environment*, pages 41–48. Springer-Verlag New York, Inc., New York, NY, USA, 1991. ISBN 0-387-97633-7. URL `http://portal.acm.org/citation.cfm?id=140806.140809`.

[115] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.*, 21(3):257–266, July 2002. ISSN 0730-0301.

[116] David S. Ebert, Christopher D. Shaw, Amen Zwa, and Cindy Starr. Two-handed interactive stereoscopic visualization. In *IEEE Visualization*, 1996.

[117] Jose I. Echevarria and Diego Gutierrez. Mobile computational photography: Exposure fusion on the n900. In *Proc. of SIACG*, 2011.

[118] Albert Einstein. *Relativity: the special and the general theory*. Crown Publishers, 1961.

[119] Candice H. Brown Elliott, Seokjin Han, Moon H. Im, Michael Higgins, Paul Higgins, MunPyo Hong, Nam-Seok Roh, Cheolwoo Park, and Kyuha Chung. Co-optimization of color amlcd subpixel architecture and rendering algorithms. In *SID Digest*, volume 33, pages 172–175, 2002.

[120] Candice H. Brown Elliott, Thomas L. Credelle, and Michael F. Higgins. Adding a white subpixel. *Information Display*, 21(5):26–31, 2005.

[121] Thomas Engelhardt, Thorsten-Walther Schmidt, Jan Kautz, and Carsten Dachsbacher. Low-cost subpixel rendering for diverse displays. *Computer Graphics Forum*, 2013.

[122] G. L. Fain, H. R. Matthews, M. C. Cornwall, and Y. Koutalos. Adaptation in vertebrate photoreceptors. *Physiological reviews*, 81(1):117–151, January 2001.

[123] Mark D. Fairchild. *Color appearance models (The Wiley-IS&T Series in Imaging Science and Technology)*. Wiley; 2nd edition, 2005.

[124] M.D. Fairchild and G.M. Johnson. Meet iCAM: An Image Color Appearance Model. In *IS&T/SID 10th Color Imaging Conference*, pages 33–38, 2002.

[125] M.D. Fairchild and G.M. Johnson. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging*, 2004.

[126] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.*, 27(3):67, 2008. ISSN 0730-0301. doi: http://doi.acm.org/10.1145/1360612.1360666.

[127] Faro. Faro Technologies Inc.: Measuring Arms. `http://www.faro.com`, 2012.

[128] Joyce Farrell, Shalomi Eldar, Kevin Larson, Tanya Matskewich, and Brian Wandell. Optimizing subpixel rendering using a perceptual metric. *Journal of the Society for Information Display*, 19(8):513–519, 2011.

[129] Raanan Fattal, Dani Lischinski, and Michael Werman. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21(3):249–256, July 2002. ISSN 0730-0301.

[130] Gregg E. Favalora. Volumetric 3D displays and application infrastructure. *IEEE Computer*, 38:37–44, 2005.

[131] Xiao-Fan Feng. LCD motion blur analysis, perception, and reduction using synchronized backlight flashing. In *Human Vision and Electronic Imaging XI*, volume 6057, pages M1–14. SPIE, 2006.

[132] James A. Ferwerda, Sumanta N. Pattanaik, Peter Shirley, and Donald P. Greenberg. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, pages 249–258, New York, NY, USA, 1996. ACM. ISBN 0-89791-746-4.

[133] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5):347–368, 2003.

[134] Y. Fu, J. Cheng, Z. Li, and H. Lu. Saliency cuts: An automatic approach to object segmentation. In *International Conference on Pattern Recognition (ICPR)*, 2008.

[135] L. Garcia, L. Presa, D. Gutierrez, and B. Masia. Analysis of coded apertures for defocus deblurring of hdr images. In *In Proc. of CEIG*, 2012.

[136] Greg Gbur. A camera fast enough to watch light move? `http://skullsinthestars.com/2012/01/04/a-camera-fast-enough-to-watch-light-move/`, 2012.

[137] Asher Gelbart, Brian C. Redman, Robert S. Light, Coreen A. Schwartzlow, and Andrew J. Griffis. Flash lidar based on multiple-slit streak tube imaging lidar. volume 4723, pages 9–18. SPIE, 2002. doi: 10.1117/12.476407. URL `http://link.aip.org/link/?PSI/4723/9/1`.

[138] Ronald S. Gentile, Jan P. Allebach, and Eric Walowit. A comparison of techniques for color gamut mismatch compensation. In *Proc. SPIE 1077*, pages 342–354, 1989.

[139] J. Giesen, E. Schuberth, K. Simon, P. Zolliker, and O. Zweifel. Image-dependent gamut mapping as optimization problem. *Image Processing, IEEE Transactions on*, 16(10):2401–2410, 2007.

[140] Andrew S. Glassner, Kenneth P. Fishkin, David H. Marimont, and Maureen C. Stone. Device-directed rendering. *ACM Trans. Graph.*, 14(1):58–76, January 1995. ISSN 0730-0301.

[141] K. Goda, K. K. Tsia, and B. Jalali. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature*, 458:1145–1149, 2009.

[142] Andrei Gorea and Christopher W. Tyler. New look at Bloch's law for contrast. *Journal of the Optical Society of America A*, 3(1):52–61, 1986.

[143] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *Proc. of the 23rd annual conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 43–54, New York, NY, USA, 1996. ACM. ISBN 0-89791-746-4. doi: http://doi.acm.org/10.1145/237170.237200. URL `http://doi.acm.org/10.1145/237170.237200`.

[144] S. Gottesman and E. Fenimore. New family of binary arrays for coded aperture imaging. *Applied Optics*, (20):4344–4352, 1989.

[145] C. Graves. *The zone system for 35mm photographers*. Focal Press, 1997.

[146] M.D. Grossberg, H. Peri, S.K. Nayar, and P.N. Belhumeur. Making one object look like another: controlling appearance using a projector-camera system. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages 452–459, 2004.

[147] M. Grosse, G. Wetzstein, A. Grundhöfer, and O. Bimber. Coded Aperture Projection. *ACM Transactions on Graphics*, pages 22:1–12, July 2010.

[148] Otkrist Gupta, Thomas Willwacher, Andreas Velten, Ashok Veeraraghavan, and Ramesh Raskar. Reconstruction of hidden 3D shapes using diffuse reflections. *Optics Express*, 20:19096–19108, 2012. doi: http://dx.doi.org/10.1364/OE.20.019096.

[149] D. Gutierrez, F. J. Seron, O. Anson, and A. Munoz. Chasing the green flash: a global illumination solution for inhomogeneous media. In *Proc. Spring Conference on Computer Graphics*, 2004.

[150] D. Gutierrez, S.G. Narasimhan, H.W. Jensen, and W. Jarosz. Scattering. In *ACM SIGGRAPH Asia Courses, 18*, 2008.

[151] D. Gutierrez, B. Masia, and A. Jarabo. Computational photography. In *CEIG Tutorials*, 2012.

[152] Diego Gutierrez, Oscar Anson, Adolfo Munoz, and Francisco J. Seron. Perception-Based Rendering: Eyes Wide Bleached. In *Eurographics Short Papers*, 2005.

[153] Yoav HaCohen, Eli Shechtman, Dan B Goldman, and Dani Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. on Graph.*, 30(4):70:1–70:9, 2011.

[154] K. K. Hadziabdic, J. H. Telalovic, and R. Mantiuk. Comparison of deghosting algorithms for multi-exposure high dynamic range imaging. In *Proc. of Spring Conference on Computer Graphics*, 2013.

[155] Rolf R. Hainich and Oliver Bimber. *Displays: Fundamentals and Applications*. CRC Press/A. K. Peters, 2011.

[156] Hamamatsu. Guide to Streak Cameras. `http://sales.hamamatsu.com/assets/pdf/catsandguides/e_streakh.pdf`, 2012.

[157] Zen-ichiro Hara and Naoki Shiramatsu. Improvement in the picture quality of moving pictures for matrix displays. *Journal of the Society for Information Display*, 8(2):129–137, 2000.

[158] Samuel W. Hasinoff and Kiriakos N. Kutulakos. Multiple-aperture photography for high dynamic range and post-capture refocusing. Technical report, University of Toronto, Dept. of Computer Science, 2009.

[159] S.W. Hasinoff, M. Jozwiak, F. Durand, and W.T. Freeman. Search-and-replace editing for personal photo collections. In *ICCP 2010*, pages 1–8, 2010.

[160] Haiyan He, L.J. Velthoven, E. Bellers, and J.G. Janssen. Analysis and implementation of motion compensated inverse filtering for reducing motion blur on lcd panels. In *Consumer Electronics, 2007. ICCE 2007. Digest of Technical Papers. International Conference on*, pages 1–2, 2007.

[161] Jeremy C. Hebden. Line scan acquisition for time-resolved imaging through scattering media. *Opt. Eng.*, 32(3):626–633, 1993.

[162] Rodney L. Heckaman and James Sullivan. Rendering digital cinema and broadcast tv content to wide gamut display media. 42(1):225–228, 2011.

[163] F. Heide, G. Wetzstein, R. Raskar, and W. Heidrich. Adaptive Image Synthesis for Compressive Displays. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 32(4):1–11, 2013.

[164] Felix Heide, Matthias Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Trans. Graph.*, 32(4), 2013.

[165] Simon Heinzle, Pierre Greisen, David Gallup, Christine Chen, Daniel Saner, Aljoscha Smolic, Andreas Burg, Wojciech Matusik, and Markus Gross. Computational stereo camera system with programmable control loop. *ACM Transactions on Graphics*, 30:94:1–10, 2011.

[166] R.T. Held and M.S. Banks. Misperceptions in stereoscopic displays: A vision science perspective. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 23–32. ACM, 2008.

[167] Robert Herzog, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. Spatio-temporal upsampling on the GPU. In *Proceedings of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 91–98, 2010.

[168] Robert F. Hess, Frederick A. A. Kingdom, and Lynn R. Ziegler. On the relationship between the spatial channels for luminance and disparity processing. *Vision Research*, 39:559–68, 1999.

[169] James M Hillis, Simon J Watt, Michael S Landy, and Martin S Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4:967–992, 2004.

[170] Matthew Hirsch, Douglas Lanman, Henry Holtzman, and Ramesh Raskar. Bidi screen: a thin, depth-sensing lcd for 3d interaction using light fields. *ACM Trans. Graph. (SIGGRAPH Asia)*, 28(5):1–9, 2009.

[171] Matthew Hirsch, Shahram Izadi, Henry Holtzman, and Ramesh Raskar. 8d: interacting with a relightable glasses-free 3d display. In *CHI*, pages 2209–2212, 2013.

[172] Shinsaku Hiura and Takashi Matsuyama. Depth measurement by the multi-focus camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, Washington DC, USA, 1998. IEEE Computer Society.

[173] David C. Hoaglin and Boris Iglewicz. Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82(400):pp. 1147–1149, 1987. ISSN 01621459. URL `http://www.jstor.org/stable/2289392`.

[174] B. Hoefflinger. *High-Dynamic-Range (HDR) Vision: Microelectronics, Image Processing, Computer Graphics (Springer Series in Advanced Microelectronics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007. ISBN 3540444327.

[175] David M. Hoffman and Martin S. Banks. Stereo display with time-multiplexed focal adjustment. In *SPIE Stereoscopic Displays and Applications XX*, volume 7237, pages 1–8, 2009.

[176] David M Hoffman, Vasiliy I Karasev, and Martin S Banks. Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth. *Journal of the Society for Information Display*, pages 271–297, 2011.

[177] D.M. Hoffman, A.R. Girshick, K. Akeley, and M.S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):1–30, 2008.

[178] M A Hogervorst, M F Bradshaw, and R A Eagle. Spatial frequency tuning for 3-D corrugations from motion parallax. *Vision Research*, 40:2149–2158, 2000.

[179] N.S. Holliman, N.A. Dodgson, G.E. Favalora, and L. Pockett. Three-dimensional displays: A review and applications analysis. *Broadcasting, IEEE Transactions on*, 57(2):362–371, 2011.

[180] Holografika. HoloVizio C80 3D Cinema System. `http://www.holografika.com`, 2012.

[181] Hideo Hosono. Running electricity through transparent materials: triggering a revolution in displays! JST Breakthrough Report 2013, Vol. 6, 2013.

[182] X. D. Hou and L. Q. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition*, 2007. URL `http://dx.doi.org/10.1109/CVPR.2007.383267`.

[183] Ian P. Howard and Brian J. Rogers. *Seeing in Depth*, volume 2: Depth Perception. I. Porteous, Toronto, 2002.

[184] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. Hdr deghosting: How to deal with saturation? In *CVPR*, 2013.

[185] D Huang, EA Swanson, CP Lin, JS Schuman, WG Stinson, W Chang, MR Hee, T Flotte, K Gregory, and CA Puliafito. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[186] Fu-Chung Huang, Douglas Lanman, Brian A. Barsky, and Ramesh Raskar. Correcting for optical aberrations using multilayer displays. *ACM Trans. Graph. (SIGGRAPH Asia)*, 31(6):185:1–185:12, 2012.

[187] Greg Humphreys and Pat Hanrahan. A distributed graphics system for large tiled displays. In *Proceedings of the conference on Visualization '99: celebrating ten years*, VIS '99, pages 215–223, Los Alamitos, CA, USA, 1999. IEEE Computer Society Press.

[188] R. Hunt. *The Reproduction of Colour (The Wiley-IS&T Series in Imaging Science and Technology)*. Wiley; 6th edition, 2004.

[189] L. M. Hurvich and D. Jameson. An Opponent-process Theory of Color Vision. *Psychological Review*, 64 (1(6)):384–404, Nov. 1957.

[190] IEC 61966-2-4 First edition. Multimedia systems and equipment- Color measurement and management- Part 2-4: Color management- Extended-gamut YCC colour space for video applications- xvYCC, 2006.

[191] IEC 61996-2-2. Multimedia systems and equipment- color measurement and management- part 2-2: Color management- extended rgb color space- scrgb.

[192] J.J.M. In 't Zand. *Coded Aperture Imaging in High-Energy Astronomy*. PhD thesis, University of Utrecht, 1992.

[193] M. Irani and S. Peleg. Super Resolution From Image Sequences. In *ICPR-C*, pages 115–120, 1990.

[194] Aaron Isaksen, Leonard McMillan, and Steven J. Gortler. Dynamically reparameterized light fields. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 297–306, 2000.

[195] ISO 11664-1:2007(E)/CIE S 014-1/E:2006. Joint ISO/CIE Standard: Colorimetry Part 1: CIE Standard Colorimetric Observers, 2007.

[196] J. Itatani, F. Quéré, G. L. Yudin, M. Yu. Ivanov, F. Krausz, and P. B. Corkum. Attosecond streak camera. *Phys. Rev. Lett.*, 88:173903, 2002.

[197] L. Itty, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.

[198] Frederic E. Ives. Parallax stereogram and process of making same. U.S. Patent 725,567, 1903.

[199] A. Jarabo, C. Barsi R. Raskar B. Masia, A. Velten, and D. Gutierrez. Rendering relativistic effects in transient imaging. In *Proc. of CEIG*, 2013.

[200] Adrian Jarabo, Belen Masia, and Diego Gutierrez. Efficient propagation of light field edits. In *Proc. of the V Ibero-American Symposium in Computer Graphics*, SIACG 2011, pages 75–80, 2011. ISBN 978-972-98464-6-5.

[201] Adrian Jarabo, Belen Masia, and Diego Gutierrez. Transient rendering and relativistic visualization. Technical Report TR-01-2013, Universidad de Zaragoza, April 2013.

[202] Christopher Jaynes and Divya Ramakrishnan. Super-resolution composition in multi-projector displays. *IEEE PROCAMS*, 2003.

[203] J. Jimenez, B. Masia, J. I. Echevarria, F. Navarro, and D. Gutierrez. Practical morphological anti-aliasing. In *GPU Pro 2: Advanced Rendering Techniques*. A. K. Peters, 2011.

[204] Elaine W. Jin, Michael E. Miller, Serguei Endrikhovski, and Cathleen D. Cerosaletti. Creating a comfortable stereoscopic viewing experience: effects of viewing distance and field of view on fusional range. *Proc. SPIE*, 5664:10–21, 2005.

[205] Elaine W. Jin, Michael E. Miller, and Mark R. Bolin. Tolerance of misalignment in stereoscopic systems. *Proc. ICIS*, pages 370–373, 2006.

[206] Andrew E. Johnson, Jason Leigh, Paul Morin, and Peter Van Keken. GeoWall: Stereoscopic Visualization for Geoscience Research and Education. *IEEE Computer Graphics and Applications*, 26:10–14, 2006.

[207] C. Johnson. *The practical zone system*. Focal Press, 1999.

[208] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. Rendering for an interactive 360° light field display. *ACM Trans. Graph. (SIGGRAPH)*, 26:40:1–40:10, 2007.

[209] Graham Jones, Delman Lee, Nicolas Holliman, and David Ezra. Controlling Perceived Depth in Stereoscopic Images. In *SPIE Stereoscopic Displays and Virtual Systems VIII*, volume 4297, pages 42–53, 2001.

[210] N. Joshi, R. Szeliski, and D. J. Kriegman. PSF Estimation Using Sharp Edge Prediction. In *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, 2008. IEEE Computer Society.

[211] Bela Julesz. *Foundations of Cyclopean Perception*. MIT Press, 2006.

[212] Yong Ju Jung, Seong-il Lee, Hosik Sohn, Hyun Wook Park, and Yong Man Ro. Visual comfort assessment metric based on salient object motion information in stereoscopic video. *Journal of Electronic Imaging*, 21(1):011008–1–011008–16, 2012.

[213] Koichiro Kakinuma. Technology of Wide Color Gamut Backlight with Light-Emitting Diode for Liquid Crystal Display Television. *Japanese Journal of Applied Physics*, 45:4330–4334, 2006.

[214] Michael Kalloniatis and Charles Luu. Temporal Resolution. `http://webvision.med.utah.edu/temporal.html`, 2009.

[215] Brian W. Keelan. Predicting multivariate image quality from individual perceptual attributes. In *PICS 2002: IS&T's PICS Conference, An International Technical Conference on Digital Image Capture and Associated System*, pages 82–87, 2002.

[216] Petr Kellnhofer, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Optimizing disparity for motion in depth. *Computer Graphics Forum (Proc. EGSR 2012)*, 32(4), 2013.

[217] William B. Kerr and Fabio Pellacini. Toward evaluating lighting design interface paradigms for novice users. *ACM Trans. Graph.*, 28(3):26:1–26:9, July 2009. ISSN 0730-0301. doi: 10.1145/1531326.1531332. URL `http://doi.acm.org/10.1145/1531326.1531332`.

[218] William B. Kerr and Fabio Pellacini. Toward evaluating material design interface paradigms for novice users. *ACM Trans. Graph.*, 29(4):35:1–35:10, July 2010. ISSN 0730-0301. doi: 10.1145/1778765.1778772. URL `http://doi.acm.org/10.1145/1778765.1778772`.

[219] E. A. Khan, E. Reinhard, R. W. Fleming, and H. H. Bülthoff. Image-based material editing. *ACM Trans. Graph.*, 25(3):654–663, July 2006.

[220] Changil Kim, Alexander Hornung, Simon Heinzle, Wojciech Matusik, and Markus Gross. Multi-perspective stereoscopy from light fields. *ACM Trans. Graph.*, 30:190:1–190:10, December 2011. ISSN 0730-0301.

[221] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4):73:1–73:12, July 2013.

[222] Jee-Hong Kim and Jan P. Allebach. Color Filters for CRT-based Rear Projection Television. *IEEE Transactions on Consumer Electronics*, 42(4):1050–1054, Nov. 1996.

[223] Manbae Kim, Seno Lee, Changyeol Choi, Gi-Mun Um, Namho Hur, and Jinwoong Kim. Depth Scaling of Multiview Images for Automultiscopic 3D Monitors. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008.

[224] Min H. Kim, Tim Weyrich, and Jan Kautz. Modeling human color perception under extended luminance levels. *ACM Trans. Graph.*, 28(3):27:1–9, July 2009.

[225] Min H. Kim, Tobias Ritschel, and Jan Kautz. Edge-aware color appearance. *ACM Transactions on Graphics (presented at SIGGRAPH2011)*, 30(2):13:1–9, 2011.

[226] Yongjin Kim, Yunjin Lee, Henry Kang, and Seungyong Lee. Stereoscopic 3d line drawing. *ACM Trans. Graph.*, 32(4):57:1–57:13, July 2013.

[227] Yongjin Kim, Holger Winnemöller, and Seungyong Lee. WYSIWYG stereo painting. In *Proceedings of ACM Symposium on Interactive 3D Graphics and Games 2013*, 2013.

[228] Fred Kingdom and Bernard Moulden. Border effects on brightness: A review of findings, models and issues. *Spatial Vision*, 3(4):225–262, 1988.

[229] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using ultrafast transient imaging. *International Journal of Computer Vision*, 95(1):13–28, 2011.

[230] Michiel A. Klompenhouwer and Gerard De Haan. Subpixel image scaling for color-matrix displays. *Journal of the Society for Information Display*, 11(1):99–108, 2003.

[231] Michiel A. Klompenhouwer and Leo Jan Velthoven. Motion blur reduction for liquid crystal displays: Motion-compensated inverse filtering. In *Proceedings of SPIE*, volume 5308, 2004.

[232] Michael Klug, Mark Holzbach, and Alejandro Ferdman. Method and apparatus for recording one-step, full-color, full-parallax, holographic stereograms. U.S. Patent 6,330,088, 2001.

[233] R. Kodama, K. Okada, and Y. Kato. Development of a two-dimensional space-resolved high speed sampling camera. *Rev. Sci. Instrum.*, 70(625), 1999.

[234] Kuniko Kojima and Akihisa Miyata. Laser TV. Technical report, Mitsubishi Electric ADVANCE, December 2009.

[235] J. Konrad, G. Brown, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee. Automatic 2d-to-3d image conversion using 3d examples from the internet. volume 8288, pages 82880F–82880F–12, 2012.

[236] Frank L. Kooi and Alexander Toet. Visual comfort of binocular and 3D displays. *Displays*, 25:99–108, 2004.

[237] Sanjeev J. Koppal, C. Lawrence Zitnick, Michael F. Cohen, Sing Bing Kang, Bryan Ressler, and Alex Colburn. A viewer-centric editor for 3d movies. *IEEE Computer Graphics and Applications*, 31:20–35, 2011.

[238] Norman Koren. A simplified zone system. URL `www.normankoren.com/zonesystem.html`. www.normankoren.com/zonesystem.html.

[239] Gerd Kortemeyer, Philip Tan, and Steven Schirra. A slower speed of light: Developing intuition about special relativity with games. In *Proceedings of the International Conference on the Foundations of Digital Games (FDG)*, 2013.

[240] Rafael Pacheco Kovaleski and Manuel M. Oliveira. High-quality brightness enhancement functions for real-time reverse tone mapping. *The Visual Computer*, 25(5-7):539–547, April 2009.

[241] R.J. Krauzlis and S. G. Lisberger. Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys. *J. Neurophysiol.*, 72(1):150–162, July 1994.

[242] Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. Lightness perception in tone reproduction for high dynamic range images. *Computer Graphics Forum*, 24(3):635–645, 2005.

[243] Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. Contrast restoration by adaptive countershading. *Computer Graphics Forum*, 26, 2007.

[244] J. Kronander, S. Gustavson, G. Bonnet, and J. Unger. Unified HDR reconstruction from raw CFA data. In *IEEE International Conference on Computational Photography (ICCP)*, 2013.

[245] Jiangtao Kuang, Garrett M. Johnson, and Mark D. Fairchild. iCAM06: A refined image appearance model for {HDR} image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414, 2007.

[246] T. Kunkel and E. Reinhard. A neurophysiology-inspired steady-state color appearance model. *Journal of the Optical Society of America A*, 26:776–782, March 2009.

[247] T. Kurita. Moving picture quality improvement for hold-type AM-LCDs. In *Society for Information Display (SID) '01*, pages 986–989, 2001.

[248] Yuichi Kusakabe, Masaru Kanazawa, Yuji Nojiri, Masato Furuya, and Makato Yoshimura. A yc-separation-type projector: High dynamic range with double modulation. *Journal of the Society for Information Display*, 16(2):383–391, 2008.

[249] J. Laird, M. Rosen, J. Pelz, E. Montag, and S. Daly. Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli. In *Human Vision and Electronic Imaging XI*, volume 6057 of *SPIE Proceedings Series*, pages 32–43, 2006.

[250] Justin Laird, Remco Muijs, and Jiangtao Kuang. Development and evaluation of gamut extension algorithms. *Color Research & Application*, 34(6):443–451, 2009.

[251] M.T.M. Lambooij, W.A. IJsselsteijn, and M.F. Fortuin. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Technology and Science*, 53:1–14, 2009.

[252] Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross. Nonlinear disparity mapping for stereoscopic 3D. *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, 29(4):75:1–75:10, 2010.

[253] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar. Polarization Fields: Dynamic Light Field Display using Multi-Layer LCDs. *ACM Trans. Graph. (SIGGRAPH Asia)*, 3:1–9, 2011.

[254] Douglas Lanman and David Luebke. Near-eye light field displays. In *ACM SIGGRAPH 2013 Emerging Technologies*, SIGGRAPH '13, pages 11–11:1, 2013.

[255] Douglas Lanman, Matthew Hirsch, Yunhee Kim, and Ramesh Raskar. Content-adaptive Parallax Barriers: Optimizing Dual-Layer 3D Displays using Low-Rank Light Field Factorization. *ACM Trans. Graph. (SIGGRAPH Asia)*, 29:163:1–163:10, 2010.

[256] Patrick Ledda, Greg Ward, and Alan Chalmers. A wide field, high dynamic range, stereographic viewer. In *Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, GRAPHITE '03, pages 237–244, New York, NY, USA, 2003. ACM.

[257] Patrick Ledda, Alan Chalmers, Tom Troscianko, and Helge Seetzen. Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph.*, 24(3):640–648, July 2005. ISSN 0730-0301.

[258] Billy Lee and Brian Rogers. Disparity modulation sensitivity for narrow-band-filtered stereograms. *Vision Research*, 37:1769–1777, 1997.

[259] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3), 2007.

[260] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. *ACM Trans. Graph.*, 23 (3):689–694, August 2004. ISSN 0730-0301.

[261] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proc. of SIGGRAPH'96*, pages 31–42. ACM, 1996. ISBN 0-89791-746-4. doi: http://doi.acm.org/10.1145/237170.237199. URL `http://doi.acm.org/10.1145/237170.237199`.

[262] Jing Li, M. Barkowsky, and P. Le Callet. The influence of relative disparity and planar motion velocity on visual discomfort of stereoscopic videos. In *Third International Workshop on Quality of Multimedia Experience (QoMEX), 2011*, pages 155–160, 2011.

[263] Jing Li, Marcus Barkowsky, Junle Wang, and Patrick Le Callet. Study on visual discomfort induced by stimulus movement at fixed depth on stereoscopic displays using shutter glasses. In *International Conference on Digital Signal Processing*, pages 1–8, 2011.

[264] C. Liang, T. Lin, B. Wong, C. Liu, , and H. Chen. Programmable aperture photography: multiplexed light field acquisition. *ACM Transactions on Graphics*, 27(3), 2008.

[265] Gabriel Lippmann. La photographie integrále. *Academie des Sciences*, 146:446–451, 1908.

[266] Gabriel Lippmann. Épreuves réversibles donnant la sensation du relief. *Journal of Physics*, 7(4):821–825, 1908.

[267] Lenny Lipton. *Foundations of the Stereoscopic Cinema: a study in depth*. Van Nostrand Reinhold, 1982.

[268] Dani Lischinski, Zeev Farbman, Matt Uyttendaele, and Richard Szeliski. Interactive local adjustment of tonal values. *ACM Trans. Graph.*, 25(3):646–653, 2006. doi: http://doi.acm.org/10.1145/1179352.1141936.

[269] Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala. Content-preserving warps for 3D video stabilization. *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, 28, 2009.

[270] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. In *IEEE Computer Vision and Pattern Recognition*, 2007.

[271] Wan-Yen Lo, Jeroen van Baar, Claude Knaus, Matthias Zwicker, and Markus H. Gross. Stereoscopic 3D copy & paste. *ACM Trans. Graph.*, 29(6):147, 2010.

[272] Ernst Lueder. *3D Displays (Wiley Series in Display Technology)*. Wiley; 1st edition, 2012.

[273] Thomas Luft, Carsten Colditz, and Oliver Deussen. Image enhancement by unsharp masking the depth buffer. *ACM Trans. Graph.*, 25(3):1206–1213, July 2006.

[274] P.D. Lunn and M.J. Morgan. The analogy between stereo depth and brightness: a reexamination. *Perception*, 24(8):901–4, 1995.

[275] M. Ronnier Luo, Anthony A. Clarke, Peter A. Rhodes, AndrÃ© Schappo, Stephen A. R. Scrivener, and Chris J. Tait. Quantifying colour appearance. part i. lutchi colour appearance data. *Color Research & Application*, 16(3):166–180, 1991.

[276] Sheng-Jie Luo, I-Chao Shen, Bing-Yu Chen, Wen-Huang Cheng, and Yung-Yu Chuang. Perspective-aware warping for seamless stereoscopic image cloning. *Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2012)*, 31(6):182:1–182:8, 2012.

[277] Lytro Inc. The Lytro camera. `http://www.lytro.com`, 2012.

[278] C. Ma, J. Suo, Q. Dai, R. Raskar, and G. Wetzstein. High-rank coded aperture projection for extended depth of field. In *International Conference on Computational Photography (ICCP)*, 2013.

[279] Hiroyuki Maeda, Kazuhiko Hirose, Jun Yamashita, Koichi Hirota, and Michitaka Hirose. All-around display for video avatar in real world. In *IEEE/ACM ISMAR*, pages 288–289, 2003.

[280] Dhruv Mahajan, Fu-Chung Huang, Wojciech Matusik, Ravi Ramamoorthi, and Peter Belhumeur. Moving gradients: A path-based method for plausible image interpolation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, 28(3):42:1–42:11, 2009.

[281] A. Mainmone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs. Focus 3D: Compressive Accommodation Display. *ACM Trans. Graph.*, pages 1–12, 2013.

[282] A. Majumder. Is spatial super-resolution feasible using overlapping projectors? In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 4, pages 209–212, 2005.

[283] Aditi Majumder. Contrast enhancement of multi-displays using human contrast sensitivity. In *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 377–382, 2005.

[284] Aditi Majumder and Michael S. Brown. *Practical Multi-projector Display Design*. A. K. Peters, Ltd., Natick, MA, USA, 2007. ISBN 1568813104.

[285] Aditi Majumder and Rick Stevens. Perceptual photometric seamlessness in projection-based tiled displays. *ACM Trans. Graph.*, 24(1):118–139, January 2005.

[286] Aditi Majumder and Greg Welch. Computer Graphics Optique: optical superposition of projected computer graphics. In *Proceedings of the 7th Eurographics conference on Virtual Environments & 5th Immersive Projection Technology*, EGVE'01, pages 209–218, 2001.

[287] Aditi Majumder, Zhu He, Herman Towles, and Greg Welch. Achieving color uniformity across multi-projector displays. In *Proceedings of the conference on Visualization '00*, VIS '00, pages 117–124, Los Alamitos, CA, USA, 2000. IEEE Computer Society Press.

[288] P. Mäkelä, J. Rovamo, and D. Whitaker. Effects of luminance and external temporal noise on flicker sensitivity as a function of stimulus size at various eccentricities. *Vision Research*, 34(15):1981–91, 1994.

[289] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, pages 442–448, 1995.

[290] Radoslaw Mantiuk, Rafal Mantiuk, Anna Tomaszewska, and Wolfgang Heidrich. Color correction for tone mapping. *Computer Graphics Forum (Proc. of EUROGRAPHICS 2009)*, 28(2):193–202, 2009.

[291] Rafal Mantiuk and Hans-Peter Seidel. Modeling a generic tone-mapping operator. *Computer Graphics Forum (Proc. of Eurographics)*, 27(3), 2008.

[292] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.*, 3(3):286–308, July 2006. ISSN 1544-3558.

[293] Rafal Mantiuk, Scott Daly, and Louis Kerofsky. Display adaptive tone mapping. *ACM Trans. Graph. (SIGGRAPH)*, 27:68:1–68:10, 2008.

[294] Rafal Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30 (4):40:1–40:14, 2011.

[295] William R. Mark, Leonard McMillan, and Gary Bishop. Post-rendering 3D warping. In *Proceedings of ACM I3D*, pages 7–16, 1997.

[296] Miguel Martin, Roland Fleming, Olga Sorkine, and Diego Gutierrez. Understanding exposure for reverse tone mapping. In *Congreso Español de Informática Gráfica*, pages 189–198, 2008.

[297] Susana Martinez-Conde and Stephen L. Macknik. The Neuroscience of Illusion. Scientific American. `http://www.scientificamerican.com/article.cfm?id=the-neuroscience-of-illusion`, 2013.

[298] Susana Martinez-Conde, Stephen L. Macknik, and David H. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240, March 2004.

[299] Fabio Marton, Marco Agus, Enrico Gobbetti, Giovanni Pintore, and Marcos Balsa Rodriguez. Natural exploration of 3d massive models on large-scale light field displays using the fox proximal navigation technique. *Computers & Graphics*, 36(8):893 – 903, 2012.

[300] K. Masaoka, Y. Nishida, M. Sugawara, and E. Nakasu. Design of primaries for a wide-gamut television colorimetry. *Broadcasting, IEEE Transactions on*, 56(4):452–457, 2010.

[301] Belen Masia and Diego Gutierrez. Multilinear regression for gamma expansion of overexposed content. Technical Report RR-03-11, Department of Computer Science and Systems Engineering, Universidad de Zaragoza, July 2011.

[302] Belen Masia, Sandra Agustin, Roland Fleming, Olga Sorkine, and Diego Gutierrez. Evaluation of reverse tone mapping through varying exposure conditions. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, 28(5), 2009.

[303] Belen Masia, Roland Fleming, Olga Sorkine, and Diego Gutierrez. Selective reverse tone mapping. In *Proc. of CEIG (Congreso Español de Informatica Grafica)*, 2010.

[304] Belen Masia, Adrian Corrales, Lara Presa, and Diego Gutierrez. Coded apertures for defocus deblurring. In *Proc. of SIACG (Iberoamerican Symposium on Computer Graphics)*, Faro, Portugal, 2011.

[305] Belen Masia, Lara Presa, Adrian Corrales, and Diego Gutierrez. Perceptually-optimized coded apertures for defocus deblurring. *Computer Graphics Forum*, 31(6), 2012.

[306] Belen Masia, Gordon Wetzstein, Carlos Aliaga, Ramesh Raskar, and Diego Gutierrez. Perceptually-optimized content remapping for automultiscopic displays. In *ACM SIGGRAPH 2012 Posters*, SIGGRAPH '12, pages 63:1–63:1, 2012.

[307] Belen Masia, Gordon Wetzstein, Carlos Aliaga, Ramesh Raskar, and Diego Gutierrez. Display Adaptive 3D Content Remapping. *Computers & Graphics, to appear*, 37(8), 2013.

[308] Belen Masia, Gordon Wetzstein, Piotr Didyk, and Diego Gutierrez. Computational displays: Pushing the boundaries of optics, computation and perception. *Computers & Graphics, to appear*, 2013.

[309] George Mather. *Foundations of Perception*. Psychology Press, 2006. ISBN 9780863778346.

[310] Suzanne P. McKee and Douglas G. Taylor. Discrimination of time: comparison of foveal and peripheral sensitivity. *Journal of the Optical Society of America A*, 1(6):620–628, 1984.

[311] Bernard Mendiburu. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, 2009.

[312] M. Menozzi, F. Lang, U. Näpflin, C. Zeller, and H. Krueger. CRT versus LCD: effects of refresh rate, display technology and background luminance in visual performance. *Displays*, 22:79–85, 2001.

[313] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, PG '07, pages 382–390, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 0-7695-3009-5.

[314] D. S. Messing and L. J. Kerofsky. Using optimal rendering to visually mask defective subpixels. In *SPIE Conference Series*, volume 6057, pages 236–247, February 2006.

[315] D.S. Messing and S. Daly. Improved display resolution of subsampled colour images using subpixel addressing. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages 625–628, 2002.

[316] Laurence Meylan, Scott Daly, and Sabine Süsstrunk. The reproduction of specular highlights on high dynamic range displays. In *IS&T/SID 14th Color Imaging Conference*, 2006.

[317] Laurence Meylan, Scott Daly, and Sabine Süsstrunk. Tone mapping for high dynamic range displays. In *Proc. IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII*, volume 6492, 2007.

[318] Patrick Monnier and Steven K Shevell. Large shifts in color appearance from patterned chromatic backgrounds. *Nature Neuroscience*, 6(8):801–802, 2003.

[319] Ethan Montag and Mark Fairchild. Gamut mapping: evaluation of chroma clipping techniques for three destination gamuts. In *IS&T/SID Sixth Color Imaging Conference: Color Science, Systems and Applications*, pages 57–61, 1998.

[320] Nathan Moroney, Mark D. Fairchild, Robert W. G. Hunt, Changjun Li, M. Ronnier Luo, and Todd Newman. The CIECAM02 color appearance model. In *IS&T/SID 10 th Color Imaging Conference*, pages 23–27, 2002.

[321] J. Morovic and M. R. Luo. The Fundamentals of Gamut Mapping: A Survey. *Journal of Imaging Science and Technology*, 45(3):283–290, 2001.

[322] Ján Morovič. *Color Gamut Mapping*. Wiley, 2008.

[323] R. Muijs, J. Laird, J. Kuang, and S. Swinkels. Subjective evaluation of gamut extension methods for wide-gamut displays. In *IDW*, 2006.

[324] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1):381–400, 1985.

[325] K. Myszkowski, R. Mantiuk, and G. Krawczyk. *High Dynamic Range Video*. Morgan & Claypool Publishers, 2007. ISBN 1598292145.

[326] Nikhil Naik, Shuang Zhao, Andreas Velten, Ramesh Raskar, and Kavita Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. Graph.*, 30(6):171:1–171:10, 2011.

[327] S. K. Nayar, H. Peri, M. D. Grossberg, and P. N. Belhumeur. A projection system with radiometric compensation for screen imperfections. In *First IEEE International Workshop on Projector-Camera Systems (PROCAMS-2003)*, 2003.

[328] Shree K. Nayar, Peter N. Belhumeur, and Terry E. Boult. Lighting sensitive display. *ACM Trans. Graphics*, pages 963–979, 2004.

[329] S.K. Nayar and V.N. Anand. 3D display using passive optical scatterers. *IEEE Computer Magazine*, 40 (7):54–63, 2007.

[330] S.K. Nayar and T. Mitsunaga. High dynamic range imaging: spatially varying pixel exposures. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 472–479, 2000.

[331] Diego F. Nehab, Pedro V. Sander, Jason Lawrence, Natalya Tatarchuk, and John Isidoro. Accelerating real-time shading with reverse reprojection caching. In *Graphics Hardware*, pages 25–35, 2007.

[332] Ren Ng. Fourier slice photography. *ACM Trans. Graph.*, 24(3):735–744, July 2005.

[333] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):253–265, 2009.

[334] Yuzhen Niu, Wu-Chi Feng, and Feng Liu. Enabling warping on stereoscopic images. *Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2012)*, 31(6), 2012.

[335] Thomas Oskam, Alexander Hornung, Huw Bowles, Kenny Mitchell, and Markus H Gross. OSCAM - optimized stereoscopic camera control for interactive 3D. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, 30:189:1–8, 2011.

[336] O. Ostberg. Accommodation and visual fatigue in display work. *Displays*, 2(2):81 – 85, 1980.

[337] Stephen E. Palmer. *Vision Science: Photons to Phenomenology*. The MIT Press, 1999.

[338] V. Pamplona, M. Oliveira, D. Aliaga, and R. Raskar. Tailored displays to compensate for visual aberrations. *ACM Trans. Graph. (SIGGRAPH)*, 2012.

[339] Hao Pan, Xiao-Fan Feng, and Scott Daly. LCD motion blur modeling and analysis. In *Proceedings of ICIP*, pages 21–24, 2005.

[340] Rohit Pandharkar, Andreas Velten, Andrew Bardagjy, Moungi Bawendi, and Ramesh Raskar. Estimating motion and size of moving non-line-of-sight objects in cluttered environments. In *IEEE Computer Vision and Pattern Recognition, CVPR 2011*, pages 265–272, 2011.

[341] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[342] Sumanta N. Pattanaik, James A. Ferwerda, Mark D. Fairchild, and Donald P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, pages 287–298, New York, NY, USA, 1998. ACM. ISBN 0-89791-999-8.

[343] Sumanta N. Pattanaik, Jack Tumblin, Hector Yee, and Donald P. Greenberg. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 47–54, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN 1-58113-208-5.

[344] A. Pavlovych and W. Stuerzlinger. A high-dynamic range projection system. In *Proc. SPIE*, volume 5969, 2005.

[345] Ken Perlin, Salvatore Paxia, and Joel S. Kollin. An autostereoscopic display. In *ACM SIGGRAPH*, pages 319–326, 2000.

[346] Kenneth Perlin and Jefferson Y. Han. Volumetric display with dust as the participating medium. U.S. Patent 6,997,558, 2006.

[347] Tom Peterka, Robert L. Kooima, Daniel J. Sandin, Andrew Johnson, Jason Leigh, and Thomas A. DeFanti. Advances in the Dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. *IEEE TVCG*, 14(3):487–499, 2008.

[348] M. Pharr and G. Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2010.

[349] John C. Platt. Optimal filtering for patterned displays. *Signal Processing Letters, IEEE*, 7(7):179–181, 2000.

[350] M. Poletti and M. Rucci. Eye movements under various conditions of image fading. *Journal of Vision*, 10(3), 2010.

[351] Brice T Pollock, Melissa Burton, Jonathan W. Kelly, Stephen Gilbert, and Eliot Winer. The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 18:581–588, 2012.

[352] N Ponomarenko, V Lukin, A Zelensky, K Egiazarain, M Carli, and F Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10: 30–45, 2009.

[353] T. Pouli, D. Cunningham, and E. Reinhard. Statistical regularities in low an high dynamic range images. *ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*, July 2010.

[354] Brian L. Price and Scott Cohen. Stereocut: Consistent interactive object selection in stereo image pairs. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1148 –1155, November 2011. doi: 10.1109/ICCV.2011.6126363.

[355] Junle Qu, Lixin Liu, Danni Chen, Ziyang Lin, Gaixia Xu, Baoping Guo, and Hanben Niu. Temporally and spectrally resolved sampling imaging with a specially designed streak camera. *Optics Letters*, 31: 368–370, 2006.

[356] Nicola Ranieri, Simon Heinzle, Quinn Smithwick, Daniel Reetz, Lanny S. Smoot, Wojciech Matusik, and Markus Gross. Multi-layered automultiscopic displays. *Computer Graphics Forum*, 31(7pt2):2135–2143, 2012.

[357] Nicola Ranieri, Simon Heinzle, Peter Barnum, Wojciech Matusik, and Markus Gross. Light-field approximation using basic display layer primitives. *SID Symposium Digest of Technical Papers*, 44(1):408–411, 2013. ISSN 2168-0159.

[358] R. Raskar and J. Davis. 5D time-light transport matrix: What can we reason about scene properties? Technical report, MIT, 2008.

[359] Ramesh Raskar, Michael S. Brown, Ruigang Yang, Wei-Chao Chen, Greg Welch, Herman Towles, Brent Seales, and Henry Fuchs. Multi-projector displays using camera-based registration. In *Proceedings of the conference on Visualization '99: celebrating ten years*, VIS '99, pages 161–168, Los Alamitos, CA, USA, 1999. IEEE Computer Society Press.

[360] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graph.*, 25(3):795–804, July 2006.

[361] Stephan Reichelt, Ralf Häussler, Gerald Fütterer, and Norbert Leister. Depth cues in human visual perception and their realization in 3d displays. In *Proc. SPIE*, volume 7690, pages 76900B–76900B–12, 2010.

[362] E. Reinhard and K. Devlin. Dynamic range reduction inspired by photoreceptor physiology. *Visualization and Computer Graphics, IEEE Transactions on*, 11(1):13–24, 2005. ISSN 1077-2626.

[363] Erik Reinhard. Tone Reproduction and Color Appearance Modeling: Two Sides of the Same Coin? In *19th Color and Imaging Conference*, pages 7–11, 2011.

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

[365] Erik Reinhard, Erum Arif Kahn, Ahmet Oguz Akyuz, and Garrett M. Johnson. *Color Imaging: Fundamentals and Applications*. A.K. Peters, 2008.

[366] Erik Reinhard, Greg Ward, Sumanta N. Pattanaik, Paul E. Debevec, and Wolfgang Heidrich. *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting (2. ed.)*. Academic Press, 2010. ISBN 9780123749147.

[367] Erik Reinhard, Tania Pouli, Timo Kunkel, Ben Long, Anders Ballestad, and Gerwin Damberg. Calibrated image appearance reproduction. *ACM Trans. Graph.*, 31(6):201:1–11, November 2012.

[368] Allan G. Rempel, Matthew Trentacoste, Helge Seetzen, H. David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph.*, 26(3), July 2007.

[369] Allan G. Rempel, Wolfgang Heidrich, Hiroe Li, and RafałMantiuk. Video viewing preferences for hdr displays under varying ambient illumination. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, APGV '09, pages 45–52, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-743-1.

[370] Tobias Ritschel and Elmar Eisemann. A computational model of afterimages. *Comp. Graph. Forum*, 31: 529–534, 2012.

[371] Tobias Ritschel, Kaleigh Smith, Matthias Ihrke, Thorsten Grosch, Karol Myszkowski, and Hans-Peter Seidel. 3d unsharp masking for scene coherent enhancement. *ACM Trans. Graph.*, 27(3), August 2008.

[372] Tobias Ritschel, Matthias Ihrke, Jeppe Revall Frisvad, Joris Coppens, Karol Myszkowski, and Hans-Peter Seidel. Temporal glare: Real-time dynamic simulation of the scattering in the human eye. *Comput. Graph. Forum*, 28(2):183–192, 2009.

[373] Tobias Ritschel, Makoto Okabe, Thorsten Thormählen, and Hans-Peter Seidel. Interactive reflection editing. *ACM Transactions on Graphics, Proceedings Siggraph Asia*, 28(5), 2009.

[374] Guido Rizzi and Matteo Luca Ruggiero. *Relativity in Rotating Frames*. Kluwer Academic, 2004.

[375] B. Rogers and M. Graham. Anisotropies in the perception of three-dimensional surfaces. *Science*, 221 (4618):1409–11, 1983. ISSN 0036-8075.

[376] Brian Rogers and Maureen Graham. Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, 22:261–270, 1982.

[377] J. Rosink, D. Chestakov, R. Rajae-Joordens, L. Albani, M. Arends, and G. Heeten. Innovative lcd displays solutions for diagnostic image accuracy. In *Proc. RadiologicalSociety of North America annual meeting*, 2006.

[378] S. Roth, I. Ben-David, M. Ben-Chorin, D. Eliav, , and O. Ben-David. Wide gamut, high brightness multiple primaries single panel projection displays. In *SID Digest*, pages 118–121, 2003.

[379] Shmuel Roth and Walt Caldwell. Four primary color projection display. In *SID Digest*, pages 1818–1821, 2005.

[380] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph*, 23(3):309–314, 2004.

[381] M. Rouf, R. Mantiuk, W. Heidrich, M. Trentacoste, and C. Lau. Glare encoding of high dynamic range images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 289–296. IEEE Computer Society, 2011.

[382] Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir. A comparative study of image retargeting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 29(5):160:1–160:10, 2010.

[383] A. Said. Analysis of subframe generation for superimposed images. In *Image Processing, 2006 IEEE International Conference on*, pages 401–404, 2006.

[384] B. Sajadi, A. Majumder, G. Meenakshisundaram, D. Lai, and A. Thler. Image Enhancement in Projectors via Optical Pixel Shift and Overlay. In *Proc. ICCP*, 2013.

[385] Behzad Sajadi, M. Gopi, and Aditi Majumder. Edge-guided Resolution Enhancement in Projectors via Optical Pixel Sharing. *ACM Trans. Graph. (SIGGRAPH)*, 31(4):79:1–79:122, 2012.

[386] Steven Scher, Jing Liu, Rajan Vaish, Prabath Gunawardane, and James Davis. 3D+2DTV: 3D displays with no ghosting for viewers without glasses. *ACM Trans. Graph.*, 32(3):21:1–21:10, July 2013.

[387] Daniel Scherzer, Lei Yang, Oliver Mattausch, Diego Nehab, Pedro V. Sander, Michael Wimmer, and Elmar Eisemann. A survey on temporal coherence methods in real-time rendering. In *Eurographics 2011 State of the Art Reports*, pages 101–126, 2011.

[388] Christophe Schlick. Quantization techniques for visualization of high dynamic range pictures. In Georgios Sakas, Stefan MÃŒller, and Peter Shirley, editors, *Photorealistic Rendering Techniques*, Focus on Computer Graphics, pages 7–20. Springer Berlin Heidelberg, 1994. ISBN 978-3-642-87827-5.

[389] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Trans. Graph.*, 23(3):760–768, 2004.

[390] H. Seetzen, H. Li, L. Ye, G. Ward, L. Whitehead, and W. Heidrich. Guidelines for contrast, brightness, and amplitude resolution of displays. In *SID Digest*, pages 1229–1233, 2006.

[391] Helge Seetzen. *High dynamic range display and projection systems*. PhD thesis, University of British Columbia, 2009.

[392] Helge Seetzen, Lorne A. Whitehead, and Greg Ward. A High Dynamic Range Display Using Low and High Resolution Modulators. In *SID Digest*, volume 34, pages 1450–1453. Blackwell Publishing Ltd, 2003.

[393] Helge Seetzen, Samy Makki, Henry Ip, Thomas Wan, Vincent Kwong, Greg Ward, Wolfgang Heidrich, and Lorne Whitehead. Self-Calibrating Wide Color Gamut High Dynamic Range Display. In *Human Vision and Electronic Imaging XII: Proc. of SPIE-IS&T Electronic Imaging, SPIE*, pages 64920Z–1–64920Z–9, 2007.

[394] Steven M. Seitz and Jiwon Kim. The space of all stereo images. *Int. J. Comput. Vision*, 48:21–38, June 2002. ISSN 0920-5691.

[395] Steven M. Seitz and Kiriakos N. Kutulakos. Plenoptic image editing. *Int. Journal of Computer Vision*, 48 (2):115–129, 2002. ISSN 0920-5691.

[396] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R. Marschner, Mark Horowitz, Marc Levoy, and Hendrik P. A. Lensch. Dual photography. *ACM Trans. Graph.*, 24(3):745–755, July 2005.

[397] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B. Goldman, and Eli Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6): 203:1–203:11, November 2012. ISSN 0730-0301.

[398] Q. Shan, J. Jia, and A. Agarwala. High-quality Motion Deblurring from a Single Image. *ACM Transactions on Graphics*, 27(3), August 2008.

[399] H Sheikh, M Sabir, and A Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Processing*, 15:3440–3451, 2006.

[400] Takashi Shibata, Takashi Kawai, Keiji Ohta, Masaki Otsuki, Nobuyuki Miyake, Yoshihiro Yoshihara, and Tsuneto Iwasaki. Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence. *SID*, 13(8):665–671, 2005.

[401] Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):1–29, 2011.

[402] Kong-King Shieh and Chin-Chiuan Lin. Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics*, 26:527–536, 2000.

[403] H. Shiraga, M. Heya, O. Maegawa, K. Shimada, Y. Kato, T. Yamanaka, and S. Nakai. Laser-imploded core structure observed by using two-dimensional x-ray imaging with 10-ps temporal resolution. *Rev. Sci. Instrum.*, 66(1):722–724, 1995.

[404] Heung-Yeung Shum, Jian Sun, Shuntaro Yamazaki, Yin Li, and Chi-Keung Tang. Pop-up light field: An interactive image-based modeling and rendering system. *ACM Trans. Graph.*, 23(2):143–162, 2004. ISSN 0730-0301. doi: 10.1145/990002.990005. URL `http://doi.acm.org/10.1145/990002.990005`.

[405] M. Siegel and S. Nagata. Just enough reality: comfortable 3-d viewing via microstereopsis. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(3):387–396, 2000.

[406] Robert Simon. Gaspar antoine de bois-clair. robert simon fine art. `http://www.robertsimon.com/pdfs/boisclair_portraits.pdf`, 2013.

[407] Eero P. Simoncelli and William T. Freeman. The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation. In *Proc. ICIP*, pages 444–447, 1995.

[408] Darryl S. K. Singh and Jung Shin. Real-time handling of existing content sources on a multi-layer display. 8648:86480I–86480I–8, 2013.

[409] Pitchaya Sitthi-amorn, Jason Lawrence, Lei Yang, Pedro V. Sander, Diego Nehab, and Jiahe Xi. Automated reprojection-based pixel shader optimization. *ACM Transactions on Graphics*, 27(5), December 2008. ISSN 0730-0301.

[410] C. Slinger, C. Cameron, and M. Stanley. Computer-generated holography as a generic display technology. *Computer*, 38(8):46–53, 2005.

[411] Adam Smith, James Skorupski, and James Davis. Transient rendering. Technical Report UCSC-SOE-08-26, School of Engineering, University of California, Santa Cruz, February 2008.

[412] K. Smith, G. Krawczyk, K. Myszkowski, and H. P. Seidel. Beyond tone mapping: Enhanced depiction of tone mapped hdr images. *Computer Graphics Forum (Proceedings of Eurographics)*, 25(3):427–438, 2006.

[413] Robert Smith-Gillespie. Design Considerations for LED Backlights in Large Format Color LCDs. In *LEDs in Displays SID Technical Symposium*, pages 1–10, 2006.

[414] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang. Three-dimensional video postproduction and processing. *Proceedings of the IEEE*, 99(4):607–625, 2011.

[415] A. Smolic, S. Poulakos, S. Heinzle, P. Greisen, M. Lang, A. Hornung, M. Farre, N. Stefanoski, O. Wang, L. Schnyder, R. Monroy, and M. Gross. Disparity-aware stereo 3d production tools. In *Visual Media Production (CVMP), 2011 Conference for*, pages 165–173, 2011.

[416] D.B. Smythe. A two-pass mesh warping algorithm for object transformation and image interpolation. *Rapport technique*, 1030, 1990.

[417] Jun Someya, Yoko Inoue, Hideki Yoshii, Muneharu Kuwata, Shuichi Kagawa, Tomohiro Sasagawa, Atsushi Michimori, Hideyuki Kaneko, and Hiroaki Sugiura. Laser TV: Ultra-Wide Gamut for a New Extended Color-Space Standard, xvYCC. In *SID Digest*, pages 1134–1137, 2006.

[418] Sony Inc. Extended-gamut Color Space for Video Applications. `http://www.sony.net/SonyInfo/technology/technology/theme/xvycc_01.html`, Last accessed July 2013.

[419] Filippo Speranza, Wa J. Tam, Ron Renaud, and Namho Hur. Effect of disparity and motion on visual comfort of stereoscopic images. In *Proceedings of the SPIE*, volume 6055, pages 94–103, 2006.

[420] Efstathios Stavrakis and Margrit Gelautz. Image-based stereoscopic painterly rendering. In *Proc. of EGSR'04*, pages 53–60, 2004. ISBN 3-905673-12-6. doi: 10.2312/EGWR/EGSR04/053-060. URL `http://dx.doi.org/10.2312/EGWR/EGSR04/053-060`.

[421] Efstathios Stavrakis and Margrit Gelautz. Interactive tools for image-based stereoscopic artwork. In *Proceedings of SPIE - Stereoscopic Displays and Applications XIX*, volume 6803, 2008.

[422] James H. Steiger. Introduction to multiple regression. `http://www.statpower.net/Content/312/Lecture%20Slides/MultipleRegressionIntro.pdf`. [Online, last accessed 20-October-2013].

[423] Timo Stich, Christian Linz, Christian Wallraven, Douglas Cunningham, and Marcus Magnor. Perception-motivated interpolation of image sequences. *ACM Transactions on Applied Perception*, 8(2): 11:1–11:25, 2011.

[424] Hagen Stolle, Jean-Christophe Olaya, Steffen Buschbeck, Hagen Sahm, and Armin Schwerdtner. Technical solutions for a full-resolution autostereoscopic 2D/3D display technology. In *Proc. SPIE*, pages 1–12, 2008.

[425] H. Sugiura, H. Kaneko, S. Kagawa, M. Ozawa, H. Tanizoe, H. Katou, T. Kimura, and H. Ueno. Wide Color Gamut and High Brightness Assured by the Support of LED Backlighting in WUXGA LCD Monitor. In *SID Digest*, pages 1230–1233, 2004.

[426] H. Sugiura, S. Kagawa, H. Kaneko, M. Ozawa, H. Tanizoe, T. Kimura, and H. Ueno. Wide color gamut displays using led backlight - signal processing circuits, color calibration system and multi-primaries. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages 9–12, 2005.

[427] H. Sugiura, H. Kaneko, S. Kagawa, M. Ozawa, J. Someya, H. Tanizoe, H. Ueno, and T. Kimura. Improved Six-Primary-Color 23-in. WXGA LCD using Six-Color LEDs. In *SID Digest*, pages 1126–1129, 2005.

[428] H. Sugiura, M. Kuwata, Y. Inoue, T. Sasagawa, A. Nagase, S. Kagawa, N. Watanabe, and J. Someya. Laser TV - Ultra Wide Color Gamut in Conformity with xvYCC. In *SID Digest*, pages 12–15, 2007.

[429] H. Sugiura, T. Sasagawa, A. Michimori, E. Toide, T. Yanagisawa, S. Yamamoto, Y. Hirano, M. Usui, S. Teramatsu, and J. Someya. 65-inch, Super Slim, Laser TV with Newly Developed Laser Light Source. In *SID Digest*, pages 854–857, 2008.

[430] Hiroaki Sugiura, Hideyuki Kaneko, Shuichi Kagawa, Masahiko Ozawa, Hideki Tanizoe, Hiroshi Ueno, Taro Kimura, and Hiroshi Katou. Wide-color-gamut and high-brightness WUXGA LCD monitor with color calibrator. In *Proc. SPIE*, volume 5667, pages 554–561, 2005.

[431] Hiroaki Sugiura, Hideyuki Kaneko, Shuichi Kagawa, Jun Someya, and Hideki Tanizoe. Six-primary-color LCD monitor using six-color LEDs with an accurate calibration system. 6058:60580H–60580H–8, 2006.

[432] Alan Sullivan. A solid-state multi-planar volumetric display. In *SID Digest*, volume 32, pages 207–211, 2003.

[433] Y. Takaki. High-Density Directional Display for Generating Natural Three-Dimensional Images. *Proc. IEEE*, 94(3), 2006.

[434] Yasuhiro Takaki, Kosuke Tanaka, and Junya Nakamura. Super multi-view display with a lower resolution flat-panel display. *Opt. Express*, 19(5):4129–4139, 2011.

[435] Wa James Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono. Stereoscopic 3d-tv: Visual comfort. *IEEE Transactions on Broadcasting*, 57(2):335–346, 2011.

[436] Krzysztof Templin, Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. Apparent resolution enhancement for animations. In *27th Spring Conference on Computer Graphics*, pages 85–92, Vinicne, Slovak Republic, 2011.

[437] Krzysztof Templin, Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. Highlight microdisparity for improved gloss depiction. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2012, Los Angeles, CA)*, 31(4):1–5, 2012.

[438] Krzysztof Templin, Piotr Didyk, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Highlight microdisparity for improved gloss depiction. *ACM Trans. Graph.*, 31(4):92:1–92:5, July 2012.

[439] Michael D. Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen. A versatile hdr video production system. *ACM Trans. Graph.*, 30(4):41:1–41:10, July 2011.

[440] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.

[441] James Tompkin, Samuel Muff, Stanislav Jakuschevskij, Jim McCann, Jan Kautz, Marc Alexa, and Wojciech Matusik. Interactive light field painting. In *ACM SIGGRAPH Emerging Technologies*, 2012.

[442] James Tompkin, Simon Heinzle, Jan Kautz, and Wojciech Matusik. Content-adaptive lenticular prints. *ACM Trans. Graph.*, 32(4):133:1–133:10, July 2013.

[443] T. Y. Tou. Multislit streak camera investigation of plasma focus in the steady-state rundown phase. *IEEE Trans. Plasma Science*, 23:870–873, 1995.

[444] Matthew Trentacoste, Ratal Mantiuk, Wolfgang Heidrich, and Florian Dufrot. Unsharp masking, countershading and halos: Enhancements or artifacts? *Comp. Graph. Forum*, 31:555–564, May 2012.

[445] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *IEEE Comput. Graph. Appl.*, 13(6):42–48, November 1993.

[446] C.W. Tyler. Stereoscopic vision: cortical limitations and a disparity scaling effect. *Science*, 181(4096): 276–278, 1973.

[447] S. Ueki, K. Nakamura, Y. Yoshida, T. Mori, K. Tomizawa, Y. Narutaki, Y. Itoh, and K. Okamoto. Five-Primary-Color 60-Inch LCD with Novel Wide Color Gamut and Wide Viewing Angle. In *SID Digest*, pages 927–930, 2009.

[448] Kazuhiko Ukai and Peter A. Howarth. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, 29:106–116, 2008.

[449] R. Ulichney, A. Ghajarnia, and N. Damera-Venkata. Quantifying the Performance of Overlapped Displays. In *IS&T/SPIE Electronic Imaging*, pages 7529–27, 2010.

[450] J. Unger and S. Gustavson. High-dynamic-range video for photometric measurement of illumination. In *SPIE*, volume 6501, 2007.

[451] H. Urey, K. V. Chellappan, E. Erden, and P. Surman. State of the Art in Stereoscopic and Autostereoscopic Displays. *Proc. IEEE*, 99:540–555, 2011.

[452] Vaibhav Vaish, Bennett Wilburn, Neel Joshi, and Marc Levoy. Using plane + parallax for calibrating dense camera arrays. In *In Proc. CVPR*, pages 2–9, 2004.

[453] J. H. van Hateren. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of Vision*, 5(4):331–347, 2005.

[454] J.D. van Ouwerkerk. Image super-resolution survey. *Image and Vision Computing*, 24(10):1039–1052, 2006.

[455] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*, 26, July 2007.

[456] Andreas Velten, Amy Fritz, Moungi G. Bawendi, and Ramesh Raskar. Multibounce time-of-flight imaging for object reconstruction from indirect light. In *Conference for Lasers and Electro-Optics*, page CM2F.5. OSA, 2012.

[457] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G. Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3(745), 2012. doi: 10.1038/ncomms1747.

[458] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Everett Lawson, Chinmaya Joshi, Diego Gutierrez, Moungi G. Bawendi, and Ramesh Raskar. Relativistic ultrafast rendering using time-of-flight imaging. In *ACM SIGGRAPH 2012 Talks*, 2012.

[459] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: Capturing and visualizing the propagation of light. *ACM Trans. Graph.*, 32(4), 2013.

[460] N. J. Wade and S. Finger. The eye as an optical instrument: from camera obscura to helmholtz's perspective. *Perception*, 30(10):1157–1177, 2001.

[461] Bruce Walter, George Drettakis, and Steven Parker. Interactive rendering using render cache. In *Proceedings of EGSR*, pages 19–30, 1999.

[462] Brian A. Wandell. *Foundations of Vision*. Sinauer Associates Inc., 1995. ISBN 9780878938537.

[463] Dong Wang, Imari Sato, Takahiro Okabe, and Yoichi Sato. Radiometric compensation in a projector-camera system based on the properties of human vision system. In *IEEE International Workshop on Projector-Camera Systems (PROCAMS)*, Washington, DC, USA, 2005. IEEE Computer Society.

[464] Lifeng Wang, Stephen Lin, Seungyong Lee, Baining Guo, and Heung-Yeung Shum. Light field morphing using 2d features. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):25–34, January 2005. ISSN 1077-2626. doi: 10.1109/TVCG.2005.11. URL http://dx.doi.org/10.1109/TVCG.2005.11.

[465] Lvdi Wang, Li-Yi Wei, Kun Zhou, Baining Guo, and Heung-Yeung Shum. High dynamic range image hallucination. In *Eurographics Symposium on Rendering*, pages 321–326, 2007.

[466] O. Wang, M. Fuchs, C. Fuchs, J. Davis, H. P Seidel, and H. Lensch. A context-aware light source. In *Computational Photography (ICCP), 2010 IEEE International Conference on*, pages 1–8, 2010.

[467] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), April 2004.

[468] S. Wanner and B. Goldluecke. Globally consistent depth labeling of 4D lightfields. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[469] Ben Ward, Sing Bing Kang, and Eric P. Bennett. Depth director: A system for adding depth to movies. *IEEE Computer Graphics and Applications*, 31:36–48, 2011.

[470] G. Ward, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. on Visualization and Computer Graphics*, 3(4):291–306, 1997.

[471] Greg Ward. Graphics gems iv. chapter A contrast-based scalefactor for luminance display, pages 415–421. Academic Press Professional, Inc., San Diego, CA, USA, 1994. ISBN 0-12-336155-9.

[472] Daniel Weiskopf, Ute Kraus, and Hanns Ruder. Searchlight and doppler effects in the visualization of special relativity: A corrected derivation of the transformation of radiance. *ACM Trans. Graph.*, 18(3), 1999.

[473] Daniel Weiskopf, Daniel Kobras, and Hanns Ruder. Real-world relativity: Image-based special relativistic visualization. In *IEEE Visualization*, pages 303–310, 2000.

[474] Daniel Weiskopf, Marc Borchers, Thomas Ertl, Martin Falk, Oliver Fechtig, Regine Frank, Frank Grave, Andreas King, Ute Kraus, Thomas Muller, Hans-Peter Nollert, Isabel Rica Mendez, Hanns Ruder, Tobias Schafhitzel, Sonja Schar, Corvin Zahn, and Michael Zatloukal. Explanatory and illustrative visualization of special and general relativity. *IEEE Transactions on Visualization and Computer Graphics*, 12: 522–534, 2006.

[475] Greg Welch, Henry Fuchs, Ramesh Raskar, Herman Towles, and Michael S. Brown. Projected imagery in your "office of the future". *IEEE Comput. Graph. & Appl.*, 20(4):62–67, July 2000. ISSN 0272-1716.

[476] G. Westheimer. Hiperacuity. In Le Squire (ed.), editor, *Encyclopedia of Neuroscience*. Academic Press, Oxford, 2008.

[477] G. Wetzstein, D. Lanman, D. Gutierrez, and M. Hirsch. Computational Displays. ACM SIGGRAPH Course Notes, 2012.

[478] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH)*, 31(4):1–11, 2012.

[479] Gordon Wetzstein and Oliver Bimber. Radiometric compensation through inverse light transport. In *Proceedings of Pacific conference on computer graphics and applications*, pages 391–399, 2007.

[480] Gordon Wetzstein and Matt Hirsch. Display Blocks: Build your own display. http://displayblocks.org/, 2013.

[481] Gordon Wetzstein, Ivo Ihrke, Douglas Lanman, and Wolfgang Heidrich. Computational plenoptic imaging. *Computer Graphics Forum*, 30(8):2397–2426, 2011.

[482] Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar. Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. *ACM Trans. Graph. (SIGGRAPH)*, 30(4):1–12, 2011.

[483] Gordon Wetzstein, Douglas Lanman, and Piotr Didyk. Computational displays. In *Eurographics 2013 Tutorials*, 2013.

[484] Charles Wheatstone. Contributions to the Physiology of Vision. Part the First. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. *Philosophical Transactions of the Royal Society of London*, 128:371–394, 1838.

[485] Charles Wheatstone. Contributions to the Physiology of Vision. Part the Second. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision (continued). *Philosophical Transactions of the Royal Society of London*, 142:1–17, 1852.

[486] K. D.D. Willis, E. Brockmeyer, S. E. Hudson, and I. Poupyrev. Printed Optics: 3D Printing of Embedded Optical Elements for Interactive Devices. In *Proc. ACM UIST*, 2012.

[487] George Wolberg. Image morphing: A survey. *The Visual Computer*, 14(8):360–372, 1998.

[488] D. Wu, M. O'Toole, A. Velten, A. Agrawal, and R. Raskar. Decomposing global light transport using time of flight imaging. In *IEEE Computer Vision and Pattern Recognition, CVPR 2012*, pages 366–373. IEEE, 2012.

[489] D. Wu, G. Wetzstein, C. Barsi, T. Willwacher, M. O'Toole, N. Naik, Q. Dai, K. Kutulakos, and R. Raskar. Frequency analysis of transient light transport with applications in bare sensor imaging. In *European Conference on Computer Vision, ECCV 2012*, pages 542–555. Springer, 2012.

[490] D. Wu, A. Velten, M. O'Toole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar. Decomposing global light transport using time of flight imaging. *International Journal of Computer Vision (IJCV), to appear*, 2013.

[491] J. C. Wyant. White light interferometry. In *SPIE*, volume 4737, pages 98–107, 2002.

[492] Haiyun Xia and Chunxi Zhang. Ultrafast ranging lidar based on real-time Fourier transformation. *Optics Letters*, 34:2108–2110, 2009.

[493] F. Xiao, J. DiCarlo, P. Catrysse, and B. Wandell. High dynamic range imaging of natural scenes. In *The Tenth Color Imaging Conference*, 2002.

[494] Lei Yang, Yu-Chiu Tse, Pedro V Sander, Jason Lawrence, Diego Nehab, Hugues Hoppe, and Clara L Wilkins. Image-based bidirectional scene reprojection. *ACM Transactions on Graphics*, 30(6), 2011.

[495] R. Yang and G. Welch. Automatic and continuous projector display surface calibration using every-day imagery. In *WSCG 2001 Conference Proceedings*, 2001.

[496] Xuan Yang, Linling Zhang, Tien-Tsin Wong, and Pheng-Ann Heng. Binocular tone mapping. *ACM Trans. Graph.*, 31(4):93:1–93:10, July 2012.

[497] Y.-C. Yang, K. Song, S. Rho, N.-S. Rho, S. Hong, K. B. Deul, M. Hong, K. Chung, W. Choe, S. Lee, C. Y. Kim, S.-H. Lee, and H.-R Kim. Development of Six Primary-Color LCD. In *SID Digest*, pages 1210–1213, 2005.

[498] Sumio Yano, Shinji Ide, Tetsuo Mitsuhashi, and Hal Thwaites. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays*, 23:191–201, 2002.

[499] Sumio Yano, Masaki Emoto, and Tetsuo Mitsuhashi. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays*, 25:141–150, 2004.

[500] John I. Yellott and John W. Yellott. Correcting spurious resolution in defocused images. *Proc. SPIE*, 6492, 2007.

[501] Tomohiro Yendo, Naoki Kawakami, and Susumu Tachi. Seelinder: the cylindrical lightfield display. In *ACM SIGGRAPH Emerging Technologies*, 2005.

[502] Akiko Yoshida, Matthias Ihrke, Rafał Mantiuk, and Hans-Peter Seidel. Brightness of the glare illusion. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, APGV '08, pages 83–90, New York, NY, USA, 2008. ACM.

[503] Akiko Yoshida, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum*, 25(3):415–426, 2008.

[504] Kaan Yücer, Alec Jacobson, Alexander Hornung, and Olga Sorkine. Transfusive image manipulation. *ACM Trans. Graph. (proceedings of ACM SIGGRAPH ASIA)*, 31(6):176:1–176:9, 2012.

[505] Zhunping Zhang, Lifeng Wang, Baining Guo, and Heung-Yeung Shum. Feature-based light field morphing. *ACM Trans. Graph.*, 21:457–464, July 2002. ISSN 0730-0301. doi: http://doi.acm.org/10.1145/566654.566602. URL `http://doi.acm.org/10.1145/566654.566602`.

[506] C. Zhou and S. K. Nayar. What are Good Apertures for Defocus Deblurring? In *IEEE International Conference on Computational Photography*, San Francisco, CA, USA, 2009.

[507] Changyin Zhou, Stephen Lin, and Shree Nayar. Coded aperture pairs for depth from defocus. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.

[508] Changyin Zhou, Stephen Lin, and Shree Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. *International Journal of Computer Vision (IJCVs*, 93(1):53–72, 2011.

[509] Eero P. Simoncelli Zhou Wang and Alan C. Bovik. Multi-scale Structural Similarity for Image Quality Assessment. In *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, 2003.

[510] H. Zimmer, A. Bruhn, and J. Weickert. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum (Proceedings of Eurographics)*, 30(2):405–414, 2011.

[511] S. Zollmann and O. Bimber. Imperceptible calibration for radiometric compensation. In *Proc. of Eurographics (Short Paper)*, 2007.

[512] M. Zwicker, W. Matusik, F. Durand, H. Pfister, and C. Forlines. Antialiasing for automultiscopic 3D displays. In *Proc. of EGSR*, pages 73–82, 2006.

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX: `http://code.google.com/p/classicthesis/`