CREATING BETTER STEREO 3D FOR ANIMATED MOVIES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Matthew Mangan Low

August 2011

ABSTRACT

Current methods of determining stereoscopic parameters for stereo 3D films abstract away important information about the relationship between the geometry or scene elements, the camera, and the viewer, reducing the problem to a 2-dimensional space of camera separation and convergence distance. This abstraction handicaps the film's artists by narrowing the scope of the stereo aspects of movie making to being a post process, equivalent to applying a 3D lens to a movie already designed to be flat. In an attempt to counter the effect of this abstraction and to better incorporate a stereo mindset into other components of filmmaking, we have developed an interactive visualization and intuitive pipeline toolset for animation artists that allows them to simultaneously view and manipulate the impact of the higher-dimensional parameters normally abstracted away in common stereo techniques. By visualizing the complete, comprehensive stereo pathway and encouraging meaningful interaction with it, artists are better enabled to create fully informed creative and artistic stereo decisions earlier in the preliminary design stages, foresee issues related to inherent or unintentional distortions that may arise further down the production pipeline forcing costly re-renders and re-animations. By providing a quicker comprehension of the stereo space and a better understanding of the terminology used, we hope to show how the study of depth perception can guide the creation of better stereo 3D content.

The toolset also provides for the definition of a metric of stereo distortion as it relates to the perception of volume and shape, which allows artists to make informed judgments on what constitutes perceptually good stereo, either through

automated optimization or constrained manipulation based on the limits of human perception. This tool and its applications were tested in the production animation pipeline at DreamWorks Animation, and provided sufficiently accurate and helpful feedback and aesthetically pleasing suggestions to artists involved in the production of stereoscopic 3D animation.

BIOGRAPHICAL SKETCH

Matthew Mangan Low was born on November 19th 1986 in Indianapolis, Indiana. That is Matthew, in the photo, on the right at his sister's graduation from Carleton College in June 2011. At age three, he and his family moved to Shorewood, Minnesota in the western suburbs of the Twin Cities of Minneapolis and St. Paul where he grew up and calls home. Attending Minnetonka High School he was undecided from an early age, studying everything from music to drafting and animation to French, physics and computer programming, participating in Destination Imagination, playing soccer and traveling the world with family and school, and even attending math courses at the University of Minnesota. Still undecided in college, he attended Cornell University where he graduated Cum Laude with a B.S. in Applied Physics and Computer Science in May of 2009, in an attempt to cover as many engineering disciplines as possible. Having enjoyed the computer graphics courses taken at the intersection of physics and computer science and serendipitously taken Professor Don Greenberg's Johnson Business School course on Disruptive Technologies, Matthew joined the Program of Computer Graphics over the summer of 2009. He has accepted a position in research and development at DreamWorks Animation.

to my family, my teachers, my friends

ACKNOWLEDGEMENTS

I must start by thanking Don Greenberg and his peculiar little course in the Business School on disruptive technologies. The course gets recognition for introducing me to a man I may not have met otherwise, a man who has since inspired and shaped my future academic, professional, and personal lives. Once Don brought me into his utopian bastion of computer graphics atop Rhodes Hall, I was surrounded by an incredible ecosystem of interdisciplinary study and collaborative research which he has developed here over the years. Don, it has been by channeling your motivation, unyielding passion and drive, thirst for learning and experience, and brilliantly sharp mind, that this work was even possible and my time as a graduate student at Cornell made important, memorable, and exciting. I could sit and listen to your incredible life stories and accomplishments, which PK and I did on many occasions in amazement, sopping up as much wisdom as we could. You showed that a balance between technology and art and the remaining little things in life is possible and desirable. I imagine I will continue to be inspired by you personally and professionally for a long time to come.

I would also like to thank and recognize Professors David Field and James Cutting, my minor advisor and proxy, and their courses on art and vision science, for providing support, insight, and inspiration from the perspective of perceptual psychology, something this thesis would otherwise lack.

The inspiration for this work can directly be traced to the moment I met Phil Captain 3D McNally and Marilyn Friedman. Smart and true friends and colleagues, much gratitude is due to them and the rest of the community at DreamWorks Animation, including Lincoln Wallen, Jim Conrads, Neil Okamoto, Tim Kwan and Will Drees. Particularly, I must thank DreamWorks Animation for

their most generous and flexible nature in collaborating on this research and openness in hosting me on their gorgeous Los Angeles campus. Without their permission and access to their pipeline and internal data, this work would not have been possible.

To my fellow PCG student compatriots, I thank you for all your support and friendship. My lock-step academic partner, officemate and housemate, Peter, I wish you all the luck at GE. Thanks for putting up with me. Edgar and Stephen, I can't thank you enough for all of your brainy support, and late night project, coffee, snack, chat, and movie adventures. And thank you to Konstantin, Cari, and Meghan, my fellow TAs.

Larsi, your friendship, knowledge, and technical support have made my time in PCG a most enjoyable one. Best of luck in Rochester. And to the rest of the PCG staff, Hurf, Nathan, Zhao, Kevin, Bruce, Linda, and Peggy, you have kept me guided and sane. To staff and students both, our Friday PCG lunches were quitea bit of fun.

To my academic leaders, teachers and friends while here at Cornell, thank you for broadening my interests and deepening my knowledge. From the world of computers and graphics, thanks are due to KB, Steve, Doug, Noah, and Graeme Bailey. And from my original home in engineering physics, thanks are due to Lovelace, Cool, Velazquez, Kusse, Muller, Xu, Kirkland, and Wickham. From the world of cinema, Chris and Mary and Sabine. The movies, and Cornell Cinema in particular, will remain one of my greatest and longest loves.

To Tipsy Winston and her social club @ 503 and friends, there is no Cornell without you. Eugene, Will, Bill, Ryan, Mike, Bryan, Matt, Ben, Peter, Brad, Becca, Aneesha, Lexi, Ramya, Alison, Tara and Anne, you and our adventures are the memories that I will keep of this place. To all my CS and AEP classmates,

project partners, late night homework mates and friends, thanks. We struggled hard, fought long, and managed to come out ok, together and in one piece.

Recognition is also due for my Minnetonka inspirations, for that is where it begins. Mort, Rosen, Mme. Choffrut, T, Mamma J, the KEY Team and the rest of MHS, MMW, and Clear Springs, thank you for shaping and leading me to who I am today.

To the Tonka friends folk, Fab Six and The World, and my Carlson AP Physics classmates, you are my bedrock foundation. We began and grew up together, and you remain a place I can revisit and feel like I have never left, for Minnesota is our home. To the Johnsons, who are more family than friends, we do good times well.

And to my extended Low and Schwarz families, while distance keeps us apart, our yearly adventures together are the things of memories. To the Scarsellas in particular, your generosity and our great times together in this remote little corner of Upstate New York have been overwhelming and under acknowledged. This thesis is dedicated in part to the memory of my wonderful grandparents, GongGong and Popo, Nana and Papa.

Finally, to Mom, to Dad, and to Elizabeth. Your constant love and guidance and support have always been paramount in my life. You have fostered in me my spirit of adventure, passion for learning, appreciation of art and science, and compassion for others. Thank you.

This research was funded by the Program of Computer Graphics, the Department of Architecture, and the Department of Computer Science and Cornell University, and by DreamWorks Animation. All references, likeness, data, images, and any content related to DreamWorks Animation are used with permission and remain the property and under copyright of DreamWorks Animation.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . viii
List of Tables . . . x
List of Figures . . . xi

1 Introduction **1**
1.1 Organization . . . 4

2 Background **5**
2.1 Human Depth Perception . . . 5
2.1.1 Monoscopic Depth Cues . . . 8
2.1.2 Stereoscopic Depth Cues . . . 12
2.1.3 Combined Cues . . . 14
2.2 Stereoscopy: Mimicking Stereopsis . . . 18

3 Previous Work **23**
3.1 Stereoscopic Transformation . . . 23
3.2 Stereoscopic Distortions . . . 29
3.2.1 Physical and Perceptual Limitations . . . 36
3.2.2 Skewed Rays . . . 42
3.2.3 Epipolar Geometry . . . 43
3.3 Evaluation of Stereoscopic Images . . . 47
3.3.1 Quality and Naturalness of Depth . . . 47
3.4 Visual Comfort and Fatigue . . . 50
3.4.1 The Horopter and Panum's Fusion Area . . . 51
3.4.2 Vergence-Accommodation Conflict . . . 55
3.5 Stereo Parameter Prediction, Control, and Automation . . . 57
3.5.1 Puppet Theater and Cardboard Cutout Effects . . . 57
3.5.2 Bounding Parallax . . . 60

4 Completeness of the Stereoscopic Transformation **63**
4.1 Expansion of the Parameter Space . . . 64
4.1.1 Aspect Ratio - A . . . 64
4.1.2 Horizontal Sensor Resolution - R_c . . . 65
4.1.3 Convergence Method . . . 66
4.1.4 Primary Camera - P_l, P_r . . . 68
4.1.5 Viewer Position - (X_v, Y_v, Z_v) . . . 70
4.1.6 Gaze Direction - (X_g, Y_g, Z_g) . . . 71
4.1.7 Failure of Skewed Rays . . . 71
4.1.8 Allowed Divergence - γ . . . 73

4.1.9 Post Render Shift - s . . . 75
4.2 Alternate Control Mechanisms . . . 77
4.2.1 Direct . . . 78
4.2.2 Convergence Distance . . . 79
4.2.3 Bounded Parallax . . . 80
4.3 Additional Considerations . . . 83
4.3.1 Floating Windows - $(FW_{ul}, FW_{ur}, FW_{lr}, FW_{ll})$ . . . 83
4.3.2 Multirigged Cameras . . . 85
4.4 Stereoscopic Parameter Space Listings and Groupings . . . 86
4.5 Complete Unconstrained Transformation . . . 89

5 Graphical Stereoscopic Distortion Visualization Tool 92
5.1 Visualization . . . 92
5.1.1 Stereoscopic Parameter Groupings . . . 94
5.1.2 Stereoscopically Transformed Geometry . . . 99
5.1.3 Ghosted Geometry . . . 101
5.1.4 Alternative Control Mechanisms and Interfaces . . . 104
5.2 Interactivity, Feedback, and Control . . . 109
5.2.1 Distortion Visualization . . . 109
5.2.2 Floating Window Violations . . . 111
5.2.3 Multirigged Camera Incongruities . . . 114

6 Distortion Metric and Perceptual Constraints 117
6.1 Stereoscopic Distortion Metric . . . 117
6.1.1 Quantization of Distortion . . . 118
6.1.2 Definition of Shape Distortion . . . 120
6.1.3 Culling . . . 123
6.1.4 Geometry Grouping . . . 130
6.1.5 Object Weighting . . . 132
6.2 Perceptual Constraints on the Parameter Space . . . 134
6.2.1 Unconstrained Vs. Constrained Parameter Space . . . 135
6.2.2 Excessive Divergence . . . 137
6.2.3 Accommodation-Vergence Conflict . . . 138

7 Conclusions and Future Work 143

A Optimization and Automation 149
A.1 Optimization . . . 149
A.2 Analysis of Automation . . . 152

Bibliography 156

LIST OF TABLES

3.1 Stereoscopic Transformation Parameters 24
3.2 Stereoscopic Transformation Coordinate Spaces 28

4.1 Primary Camera Specification 70
4.2 Standard Camera Transformation Parameters 87
4.3 Stereoscopic Camera Transformation Parameters 87
4.4 Viewer Transformation Parameters 88
4.5 Additional Stereoscopic Transformation Parameters 88

5.1 Dictionary of Viewer Conditions 99

LIST OF FIGURES

1.1 Interactive Stereoscopic Visualization Tool 1

2.1 Human Depth Perception Cues 6
2.2 Monoscopic Depth Cues 7
2.3 Motion Parallax 10
2.4 Accommodation of the Eye 11
2.5 Vergence of the Eye 13
2.6 Relative Potency of Depth Cues 15
2.7 Accommodation Vergence Feedback Loop 16
2.8 Stereoscopy 19
2.9 Camera Separation 20
2.10 Convergence Distance 21

3.1 Converged Camera Parameters 25
3.2 Parallel Camera Parameters 26
3.3 Viewer Parameters 27
3.4 Converged Cameras and Depth Plane Curvature 32
3.5 Keystone Distortion 32
3.6 Parameter Variation and Resultant Stereoscopic Distortion 35
3.7 Ponzo Illusion 37
3.8 Physical Geometric Limitations 38
3.9 Transformation Boundary Volume 39
3.10 Ambiguous Silhouettes and Boundary Contours 41
3.11 Disparity Fields Produced by Improper Viewing 44
3.12 Skewed Rays and Epipolar Geometry 45
3.13 Quality and Naturalness of Depth 49
3.14 The Horopter and Panum's Fusional Area 52
3.15 Panum's Fusional Area and Percival's Zone of Comfort 53
3.16 Vergence-Accommodation Conflict 56
3.17 Puppet Theater and Cardboarding 58

4.1 Aspect Ratio 64
4.2 Horizontal Sensor Resolution 65
4.3 Convergence Method 67
4.4 Primary Camera 68
4.5 Viewer Position 70
4.6 Allowed Divergence 72
4.7 Infinity in Allowed Divergence 73
4.8 Post Render Shift 76
4.9 Depth Blending 76
4.10 Direct Control Mechanics 78
4.11 Convergence Distance Control Mechanics 79
4.12 Bounded Parallax Control Mechanics 81

4.13 Floating Window Control Mechanics 84
4.14 Floating Window Projection 84

5.1 Visualization Overview 93
5.2 Geometry and Parameter Visualization Grouping 94
5.3 Dictionary of Viewer Configurations 98
5.4 Ghosted Goemetry 101
5.5 Parameter Configuration Display and Interface 105
5.6 DreamWorks Animation Rendered Stereo Stills I 106
5.7 DreamWorks Animation Rendered Stereo Stills II 107
5.8 Distortion Visualization 110
5.9 Stereo Window Violations and Floating Windows I 112
5.10 Stereo Window Violations and Floating Windows II 113
5.11 Multirig Depth Incongruities 114
5.12 Visualization of Multirig Incongruities 115

6.1 Definition of Stereoscopic Shape Distortion 121
6.2 Effect of Parameter Variations on the Distortion Metric 123
6.3 View Frustum Culling 124
6.4 Effect of View Frustum Culling 126
6.5 Back Face Culling 127
6.6 Effect of Back Face Culling 129
6.7 Occlusion Culling 130
6.8 Effect of Occlusion Culling 131
6.9 Distortion Metric Optimization Surface 134
6.10 Orthostereoscopy 136
6.11 Theatrical Comfort Zone 140
6.12 Perceptual Constraints on the Distortion Metric Surface 142

A.1 Automation Results 153

CHAPTER 1

INTRODUCTION

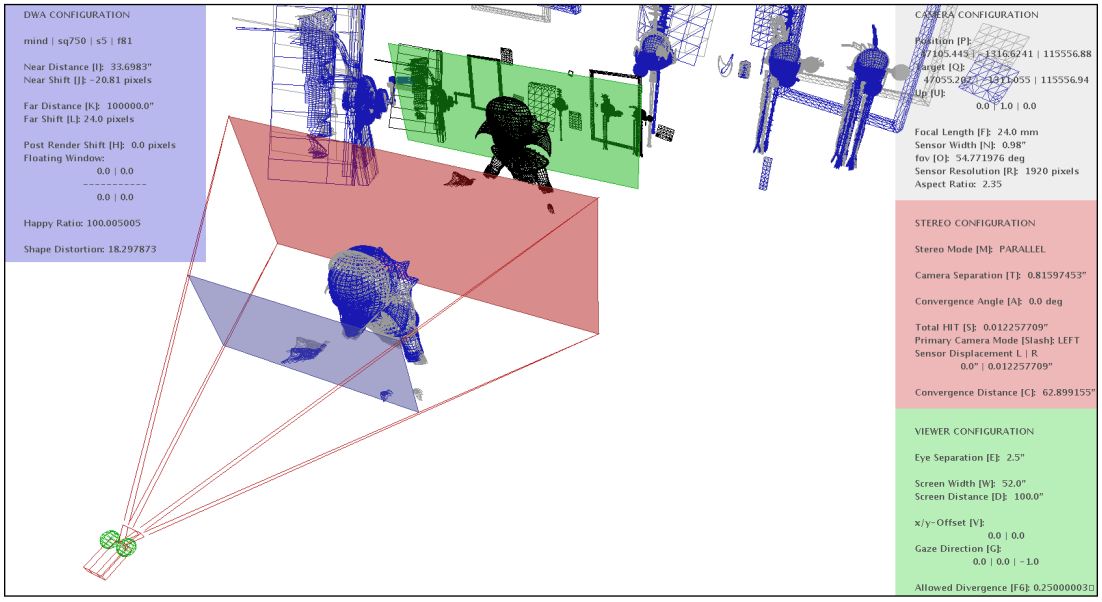


Figure 1.1: Our integrated pipeline tool and interactive visualization allows stereoscopic animation artists to manipulate the entire high-dimensional stereo parameter space and view the resultant impacts the changes have on the stereo image as perceived by the viewer.

As the entertainment industry continues to adopt stereoscopic 3D movie production, the number of 3D movies is increasing exponentially. Ticket prices and 3D surcharges continue to rise as well. As a result, audiences are becoming more discerning when it comes to stereo content in movies, requiring a higher standard of stereo filmmaking in order for a movie to be successful. Yet stereo 3D filmmakers continue to struggle to achieve acceptable stereo that minimizes adverse effects such as eye strain or headaches, or even correct depth perception void of inversions or window violations.

Currently, most 3D productions employ a "stereo consultant" who acts as an appendage to the creative process with minimal influence on the creative core of the film. This relegates stereo as secondary to the composition of the

flat 2D image and other tools such as lens choice, rack focus, blocking and staging, camera movement, editing and shot selection. As such, the stereographer's judgment and influence is traditionally limited to simply supplying a pair of numbers, the distance between the two cameras and the distance to the point of convergence in the scene, that work best given the shot in its pre-configured state. Additionally, these judgments are frequently based on a myriad of different methodologies and rules-of-thumb, and are often inconsistent with one another. This affords little latitude for incorporating stereo-minded reasoning into the core creative decision making process, and the stereo aspects are considered independent of (or worse reliant upon) the other aspects of the film production.

Our goal is to provide a tool which can be embedded into an animated feature film pipeline, one that lets the stereo component be integrated into the creative process of film production and used to help craft the artistry of the film. By allowing artists to interactively visualize the influence that the broader filmic techniques, tools, and geometric, camera, and stereo viewer space properties have on the perception of the stereo image by the audience, they can better grasp how variations in certain aspects of a shot alter the impact on the audience. Artists can then take advantage of this feedback and in turn can utilize an expanded palette of creative tools augmented by the advent of stereoscopic capabilities for the purposes of training, experimentation, and artistic growth.

The interactive graphics tool we have developed takes model and scene geometry for a given animated shot and computes the full stereoscopic transformation from model space to perceived theater space and displays both the original geometry and transformed geometry simultaneously. This allows the artist to visualize both where in theater space the final geometry will be displayed to

ensure that it does not violate any perceptual or physical limits, as well as view the types of distortions in shape and volume inherent in the stereoscopic transformation. The visualization also displays the cameras, view frustums, convergence plane, viewer's eyes, and projection plane for reference. Additionally, all of the parameters, planes, and variables needed for the stereoscopic transformation and display, as well as parameters for a variety of control paradigms currently used to define stereo, are fully manipulatable and controllable, and the resultant transformation is updated and displayed. This provides an intuitive interface and allows for instant iteration and exploration of the full stereo transform space and feedback for artist while laying out a shot.

Finally, a metric for the shape distortion of the geometry is defined using the stereoscopic transformation and provided geometry. The metric allows for the visualization and understanding of how parameter variations affect the quality of the stereo precept in a comprehensible way. The metric quantizes stereo quality, and allows a unified stereo vision across multiple artists, which becomes increasingly important with large scale animated productions where multiple artists might be working on any particular movie. Perceptual constraints are applied to the full parameter space, providing a valid operating subspace for parameter manipulation. Additionally, the potential use of the distortion metric in an optimization technique to automate the determination of stereo parameters is discussed.

The visualization tool can be used as a standalone application as a training tool for artists unfamiliar with stereoscopic filmmaking, or it can be embedded within an animation production pipeline for use by skilled artists to intuitively set the stereo or any other parameters with confidence in how the resultant

stereoscopic images will appear and iterate on variations at earlier stages in the production. This tool allows artists to go beyond simply avoiding imperfections, distortions, or depth violations when determining stereo parameters so they can develop artistic and creative uses for stereoscopic 3D filmmaking based on a better association of the physical parameters with the changes perceived by the audience.

1.1 Organization

This thesis is organized as follows. Chapter 2 introduces concepts paramount to the basic understanding of human depth perception and the craft of stereoscopy, or the creation of stereo 3D films. Chapter 3 covers relevant works from the stereoscopic transformation, to observed distortions and methods of evaluation of stereoscopic images, human visual comfort and fatigue, and finally to a selection of previous attempts to visualize, automate, or control the stereoscopic transformation. Chapter 4 expands the stereoscopic transformation to be more comprehensive, and introduces new concepts and methodologies for stereo control. In Chapter 5, we introduce a visualization tool that can be implemented in a studio animation pipeline which visualizes the capture and display spaces of stereoscopic animation and the virtual precept of the stereoscopically transformed scene by the audience in a theater. Chapter 6 introduces a method of quantizing stereoscopic quality, or a stereo distortion metric. We also discuss perceptual limitations on the stereo transformation parameter space. We conclude and ponder future directions for this research in Chapter 7, and briefly outline one potential application of our visualization tool for stereo automation in Appendix A.

CHAPTER 2

BACKGROUND

2.1 Human Depth Perception

Human depth perception refers to our ability to judge absolute and relative distances in our surrounding environment as processed by our visual system. Absolute, or egocentric distance, is the distance from an observer to an object. Relative distance refers to the distance between objects, or the prepositional ordering of objects in space, like 'in front of' and 'behind', or 'nearer' and 'farther'.

Figure 1 provides a relational hierarchy of the types of depth cues that allow us to see in depth. The primary cues include perspective, motion, relative size, position, familiarity, occlusion, texture gradient, shading, shadows, specular highlights, atmospheric blur, motion parallax, convergence, accommodation, and stereopsis. These cues can be can be categorized as either oculomotor or visual type cues, as well as either binocular or monocular cues.

Oculomotor cues are kinesthetic in nature, whereby the sensation and perception of depth is derived from the contraction of muscular fibers in the eye controlling either the focus of the lens or the orientation of the eye in the socket. All other cues are strictly visual in nature, whereby the sensation and perception of depth is derived from the retinal response of light hitting the back of the eye and the processing of this stimuli by the brain's visual cortex. This distinction is important to make, as the kinesthetic cues are intimately coupled and conflicts between the two can be a source of stereoscopic induced headaches, which we will explain later.

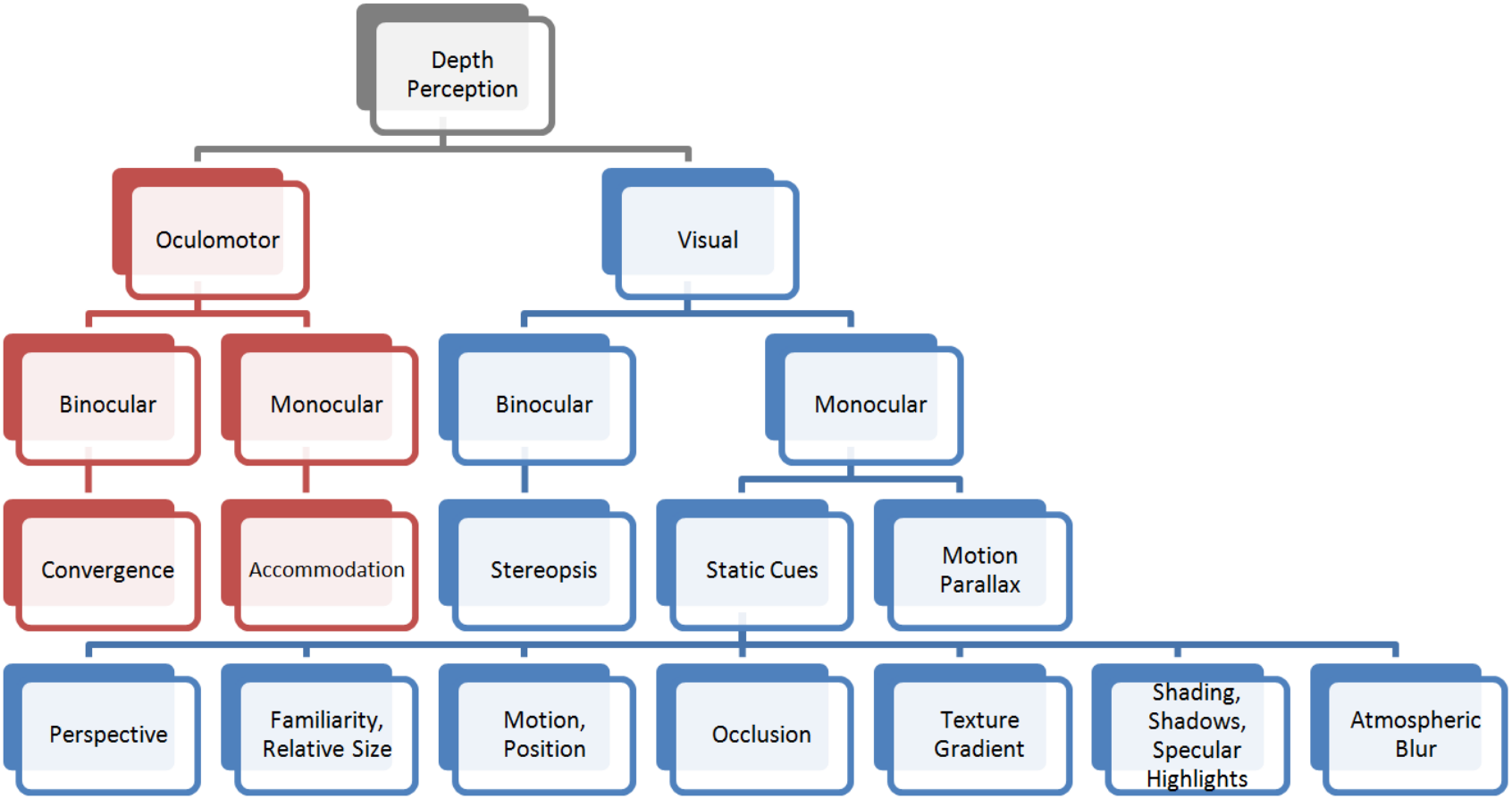


Figure 2.1: Hierarchy of Human depth perception cues. Adapted from [SB06].

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

Figure 2.2: Examples from art, photography, and computer graphics that exemplify monoscopic depth cues. (a) Gustave Caillebotte's 1877 painting 'Paris Street in Rainy Weather' exhibits familiarity, relative size, positioning, implied motion, occlusion, and texture gradient depth cues. (b) A Flickr photo by *local-japantimes* is a dramatic example of linear perspective. (c) Shading, shadows, and specular highlights make the difference between a circle and a sphere. (d) Joaquim Alves Gaspar's photo of the Serra da Estrela in Portugal demonstrates atmospheric blur from an aerial perspective.

The other major distinction is between monocular and binocular vision. Monocular (or monoscopic) cues provide depth information from only a single point of view, or one eye. Binocular (or stereoscopic) cues rely on comparing the views from two different perspectives, one from each eye. We will briefly describe the primary cues.

2.1.1 Monoscopic Depth Cues

Perspective cues describe the way in which the projection of objects onto the retina varies based on the spatial attributes of the object. In particular, as an object become more distant, the visual angle subtended by the object at the eye decreases, and the size of the retinal projection of the object decreases. When looking down a long corridor, for example, parallel objects appear to converge to a point in the eye as they become more and more distant, as in Figure 2.2(b).

Familiarity describes the way in which previous knowledge of an object can impact our assumptions on its position in space, and relative size describes the way in which the differences in apparent size between two familiar or similar objects can be accounted for by positioning in space. If previous knowledge tells us that buildings are traditionally larger than people, an image where a person appears to be as large as a building would lead us to assume that the building is further away from us than the person. Similarly, in an image where one person appears significantly larger than another, it would imply that the larger person is closer to the viewer than the other.

Related to this are motion and positioning cues. Motion cues describe how a change in position over time towards or away from a viewer will alter the projected size of the moving object. Positioning cues describe the way that objects further away converge on the horizon and their positions relative to the horizon in the image plane determines their ordering in depth. In Figure 2.2(a), Gustave Caillebotte's 1877 painting 'Paris Street in Rainy Weather' contains elements of familiarity, relative size, implied motion, and positioning.

The occlusion depth cue is due to the partial obstruction of background ob-

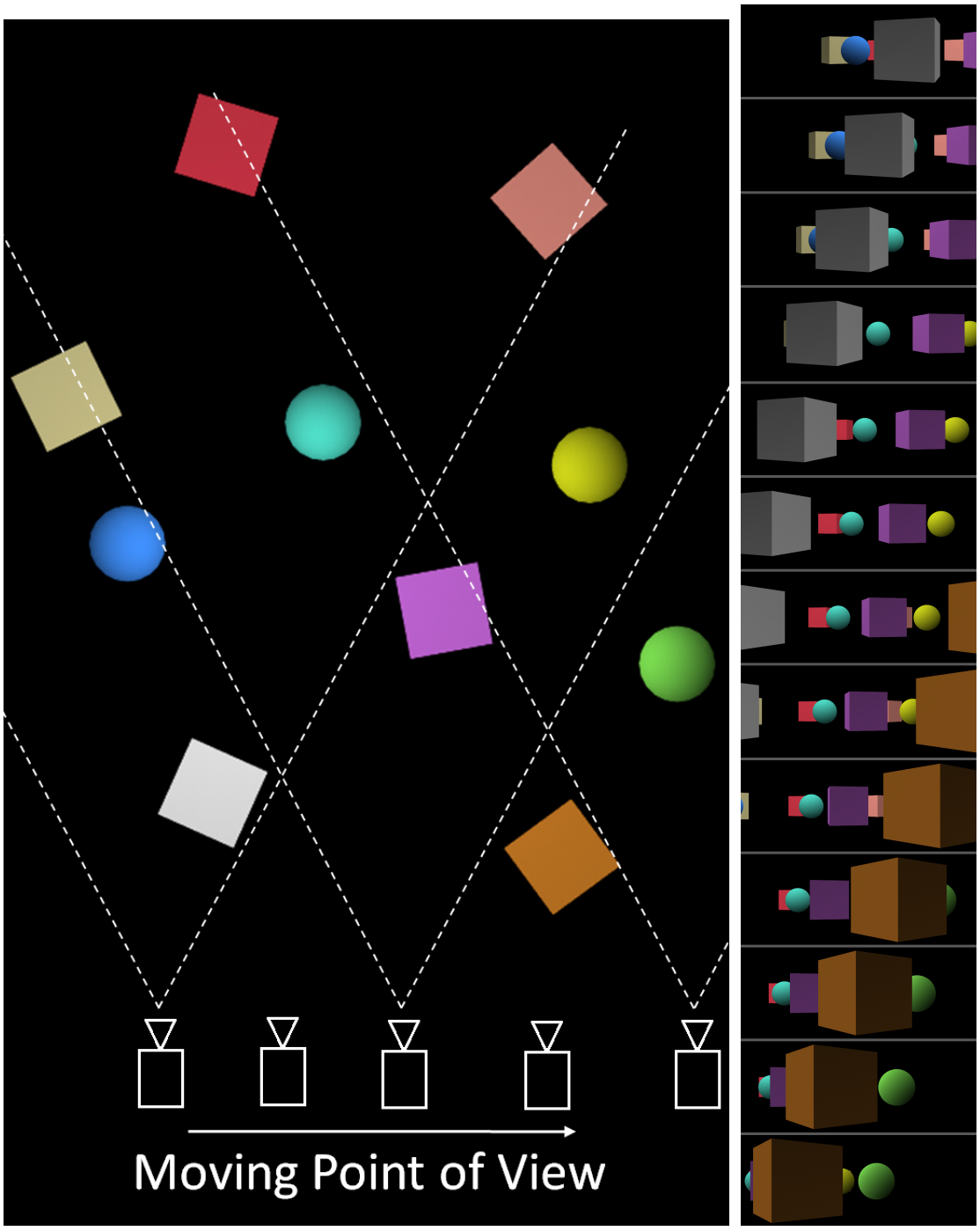
jects by foreground objects. For example, based on the order of obstructions of umbrellas over umbrellas over lamp posts, we can tell that the man with his back to us is closer than the man and woman facing us who are closer than the green lamp post in Caillebotte's painting.

Shading, shadows, and specular highlights describe the way that light interacts with objects and how a human observer can subsequently interpret the shape of the object. Figure 2.2(c) demonstrates how with accurate lighting and a shadow on the ground plane, the observer can interpret the convex geometry of a sphere.

In Figure 2.2(d) we perceive a series of mountains at increasing distances. While this is a prime example of positioning relative to the horizon, we can also note how the color of the ridges shifts from a more saturated blue in the foreground to a less saturated blue in the background. This is an example of atmospheric blur or blue shift, and is a result of reduced contrast at greater distances because of skylight scattering from molecules in the air.

Parallax

The last visual monoscopic depth cue is parallax from motion, and is one of the more powerful depth cues. In the context of depth perception, parallax is the difference in relative position of an object viewed along two different lines of sight. In terms of a monoscopic depth cue, we can obtain different points of view or lines of sight by moving the observer over time, moving it side to side or up and down. As the point of view changes, so too do the relative positions of the objects being viewed: nearer objects will appear to shift more



(a) (b)

Figure 2.3: Motion parallax is the difference in the apparent position of objects when viewed from varying point of view. (a) A plan view of the moving point of view. (b) The scene viewed from varying positions as the point of view moves from left to right.

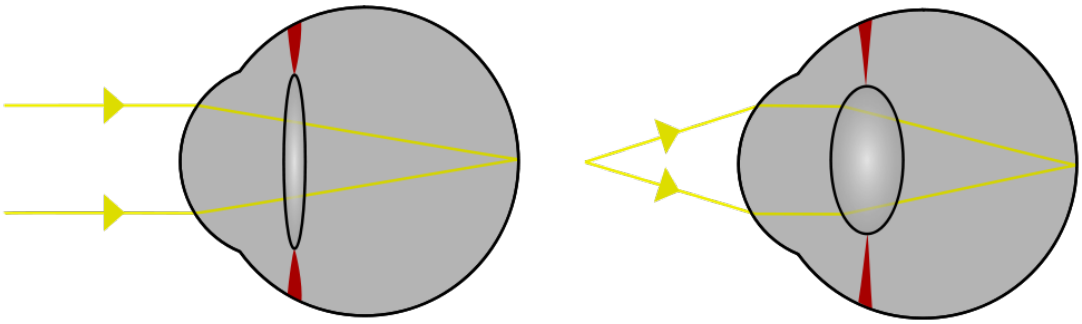


Figure 2.4: Accommodation of the eye. Ciliary muscles alter the shape of the lens in the eye allowing us to focus on objects at different distances.

than distant objects. Figure 2.3 simulates this principle with thirteen still images from a rendered video sequence as the camera tracks from left to right. Notice how the near grey and orange boxes are displaced faster across the image than the further objects, like the purple and red boxes and yellow and teal spheres. This principle is frequently exploited in 2D animation, where layers of drawings are pulled across a capturing plane at differing speeds to simulate depth in the scene.

Accommodation

The final monoscopic depth cue is accommodation, shown in Figure 2.4, which is the physical change in the shape of the lens of the eye by the ciliary muscles to maintain a focused image on the retina as a result of a change in the distance of an object. Accommodation is for the most part a reflex, but can be controlled to some extent. It allows us to change the optical power of our eyes to focus on objects at different depths. To a limited extent depth information can be properly interpreted from the degree of accommodation via muscular tension in the ciliary body for objects within a narrow range close to the viewer [LM66, WF71].

This depth cue is enhanced when conjoined with the other kinesthetic cue of vergence, described below. It should be noted that disorders of accommodation are very common, and include myopia (nearsightedness), hyperopia (farsightedness), and presbyopia (aging and stiffening of the lens), all of which can have adverse affects on vision and the perception of depth.

Depth of focus (DOF) describes the small amount of retinal blur tolerated without adjusting accommodation to perceive a sharp image, and can be thought of as a hysteretic region or image distance that straddles the fixation point. This observation is important when discussing vergence-accommodation coupling and visual comfort around the fixation point.

2.1.2 Stereoscopic Depth Cues

Convergence

The other kinesthetic depth cue is vergence, or the simultaneous motion of the eyes inwards or outwards by the rectus muscles so that the projection of a fixated object is in the center of both retinas. Convergence is movement of both eyes inwards to view a closer object, while divergence is the movement of both eyes outwards to view a further object, as indicated in Figure 2.5. By maintaining this single binocular vision, viewers are triangulating the position of an object with their eyes and the fixated object. Some measure of distance can be interpreted by the angle of muscular contraction up to a distance of 20 feet, after which the eyes have diverged to parallel, effectively fixating at infinity [VH76]. It should be noted that convergence, accommodation, and familiar size are the only depth cues that provide absolute distance information. All others only

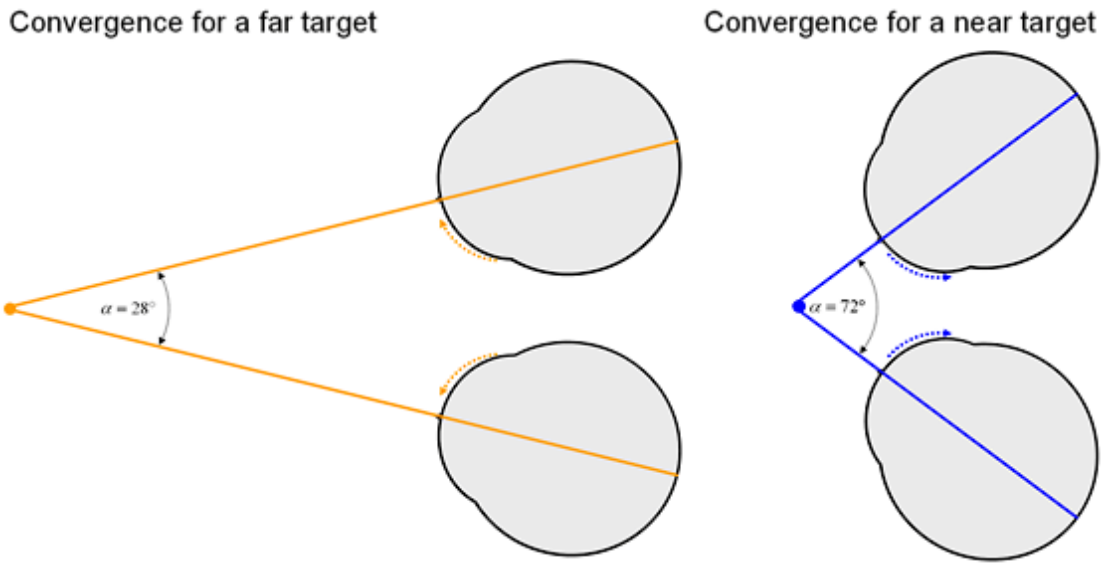


Figure 2.5: Vergence of the eye. Rectus muscles control the inward and outward motion of the eyes to maintain singular binocular vision of a fixed object. [Wal05]

provide only relative information.

Efference Copy

It should be noted that when referring to the perception of depth using the kinesthetic depth cues of accommodation and convergence that the human visual system is not technically capable of detecting or receiving as input the position or muscular tension of the eyes. Instead, the brain is only aware of the control signal, or outflow. This is known as efference copy, which is defined as an internal copy of the predicted movements or sensations resulting from a given motor command [STK+97]. Efference copy, in part, explains why other people are able to tickle us while we are not able to tickle ourselves, because of the efference copy that informs us of self stimulation. While the distinction between control signal outflow and stimulus signal inflow is important to make, it has little impact on the perception of depth as it relates to oculomotor cues

beyond human vision disorders which in general interfere with the perception of depth.

Stereopsis

The final stereoscopic depth cue takes advantage of the fact that our eyes provide two slightly different projections of the world because they are offset in the horizontal direction. The difference between corresponding points in the two projections is called retinal disparity, binocular disparity, or parallax. The brain fuses the left and right images, and using the disparities extracts relative depth information. This process of perceiving depth based on the difference between the views from our two eyes is known as stereopsis.

When we fixate (converge and accommodate) on a particular point, the projection to each eye lands on corresponding areas of the retina, in particular the middle of the fovea, and has zero disparity. Points that have positive retinal disparity are perceived further than the fixation point, and points with negative retinal disparity are perceived nearer than the fixation point. We will revisit the concept of stereopsis later as it relates to visual comfort and fatigue.

2.1.3 Combined Cues

While each depth cue discussed above has been explored independently of all other cues, under normal viewing conditions multiple, if not all, cues are utilized to garner absolute and relative depth perception with higher accuracy, consistency, and a more useful precept of the surrounding environment. While

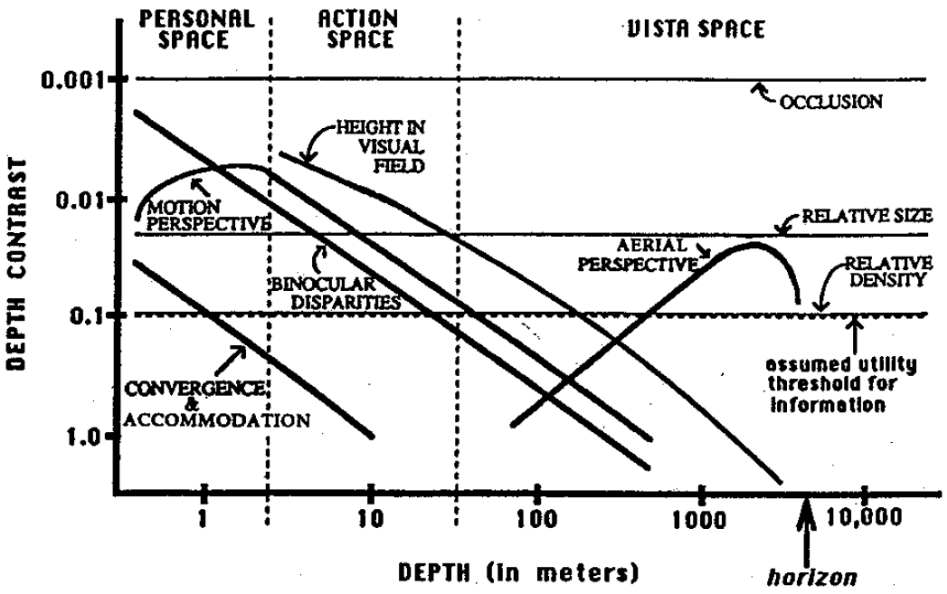


Figure 2.6: The relative potency and spatial importance of different depth cues, as explored by Cutting and Viston. They assert that more potent depth cue information is associated with smaller depth discrimination thresholds, or relative depth contrast. They found that depending on the type of the space being viewed about the viewer, either personal, action, or vista, different sources of depth information have a higher level of importance and validity for depth perception. Notice how some cues, such as occlusion and relative size, have a constant just-discriminable depth threshold, indicating consistent importance across varying depths. Other cues however, such as convergence and accommodation and aerial perspective, are only relevant in certain spaces and vary in potency across their valid range of depths. [CV95]

there is no single unified theory about cue integration, many studies have explored certain aspects of combined cues.

For example, Cutting and Vishton explored the relative potency and spatial importance of different depth cues [CV95]. They assert that more potent depth cue information is associated with smaller depth discrimination thresholds, or relative depth contrast. They found that depending on the type of the space being viewed about the viewer, either personal, action, or vista, different sources of depth information have a higher level of importance and validity for depth perception. Their findings are highlighted in Figure 2.6.

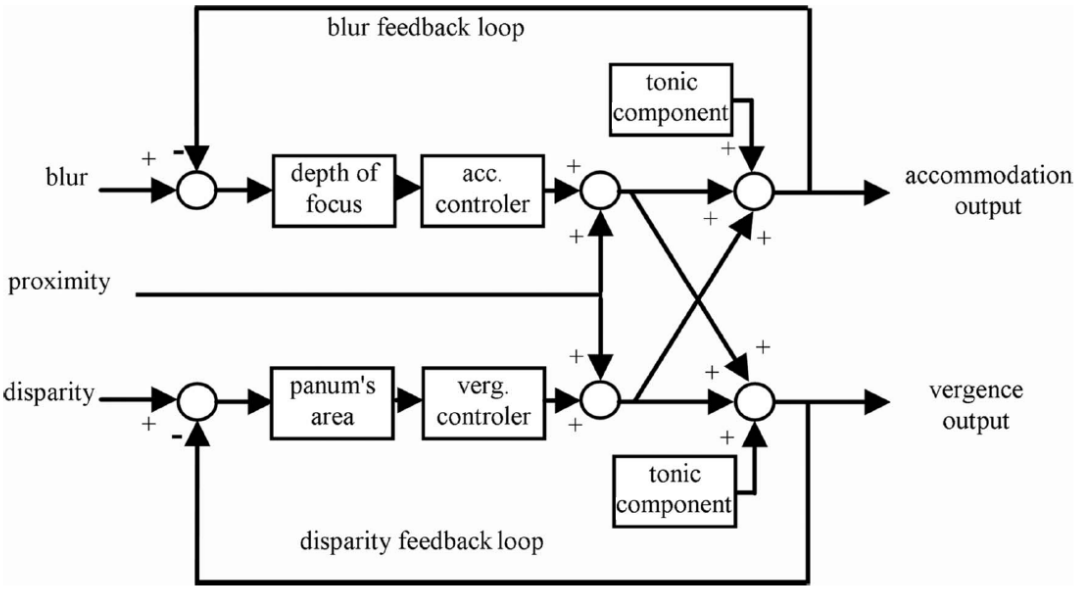


Figure 2.7: Vergence-Accommodation dual parallel feedback control systems that interact via cross-links. [LIFH09]

Combined Oculomotor Cues: Vergence-Accommodation Coupling

Vergence movements are closely accompanied by accommodation in the eye, and the two often operate in a yoked manner since both reflexes occur as a result of a change in distance to a fixation point. Thus, converging to a closer point is often paired with increased accommodation, and diverging to a further point is often paired with decreased accommodation, or relaxation, of the lens. This interaction of accommodation and vergence to the same fixation point provides clear and single binocular vision around the fixation area, where objects are in focus and do not exhibit any parallax or retinal rivalries. Small retinal defocus and disparities do not drive changes in either system, as the depth of field (DOF) and sensoric fusion provide ranges of valid clear and single vision.

Because of this close coupling, vergence and accommodation is generally modeled as two dual parallel feedback control systems that interact through

cross-links, as indicated in Figure 2.7. Accommodation is driven by retinal blur, while vergence is primarily driven by retinal disparities. Both respond to apparent pictorial and motion based depth cues that indicate proximity, and include an adaptive tonic (slow response) component. The two systems interact with each other via a reflexive cross-link, which can also be seen in Figure 2.7. [LIFH09]

Combined Visual Cues

Welchman et al. [WDC+05] studied the neural architecture mediating the combination of different types of visual information about depth structure. They measured the neural responses using functional MRIs while subjects judged the 3D structure of geometric planes defined by perspective construction and by binocular disparity. The planes were depicted as either oriented away from or towards the viewer in the horizontal direction. Consistent-cue stimuli and inconsistent-cue stimuli were provided to the subjects to study which cue dominated when in conflict.

They found that higher visual areas of the brain are involved in the processing of perceived global shape and orientation, and that orientation and depth perception depend on the combination of individual depth cues. They determined a higher reliance on perspective cues (73%) over binocular disparity (24%).

Zhang et al. [ZBS03] explored the thresholds of stereo-slant discrimination, and in particular by the two oculomotor cues. Stereo-slant discrimination is the ability to discern differences in the orientation of two large and separated

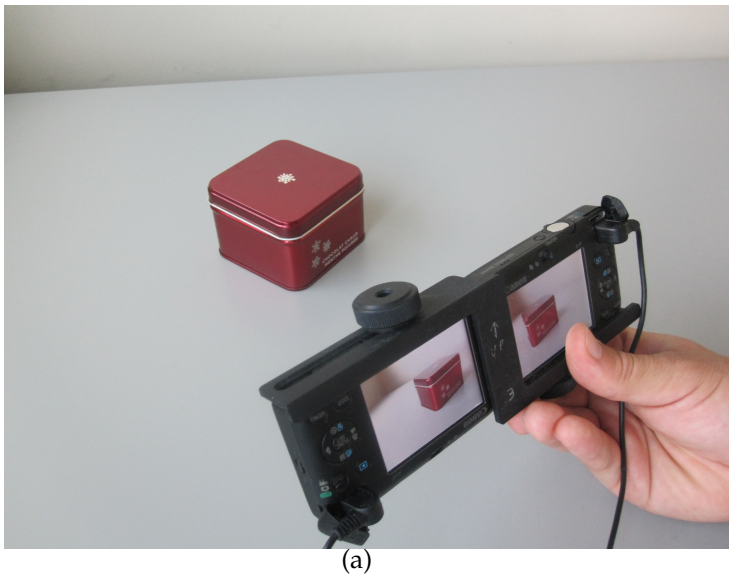
planes in the viewer's field of view. The distance between the planes forces subjects to reconverge their eyes when alternating between observing one plane and the other. To their surprise, they found that while oculomotor errors are an important factor in stereo-depth discrimination, they are not for stereo-slant discrimination. Binocular disparity cues are used to recover slant and variations in surface orientation for large surfaces, and little is derived from accommodation or vergence.

Other techniques have attempted to conceptualize depth cue integration as a statistical inference problem. By casting it as a maximum-likelihood estimation of cue combinations based on reliability, combined cue depth perception becomes an investigation into the resolution of conflicting cues. [HWLB04]

2.2 Stereoscopy: Mimicking Stereopsis

Of the cues for human depth perception explored, all but accommodation, convergence, and stereopsis can be utilized in the generation of two-dimensional imagery. Stereoscopy is any technique that creates the enhanced illusion of depth in an image by utilizing convergence and stereopsis. By capturing and displaying unique image information to the left and right eye, artists can mimic the effects of convergence and stereopsis.

In terms of capturing real world images, this is traditionally accomplished by utilizing two cameras capturing images from slightly different perspectives, as demonstrated in Figure 2.8. The left and right images are captured simultaneously from the pair of cameras. In computer animation, a similar technique is used by creating a second camera that is a copy of and offset from the original



(a)

(b) Left Camera Image

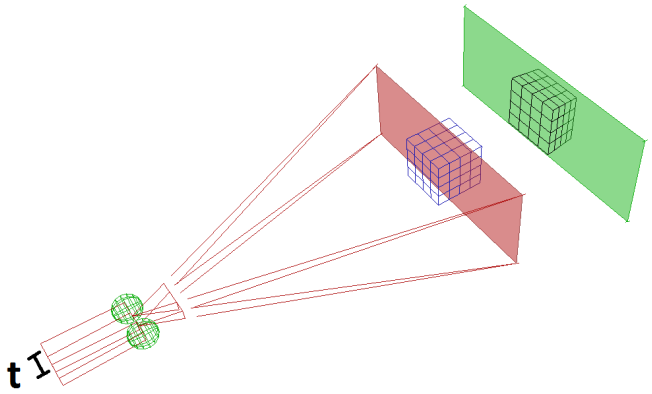
(c) Right Camera Image

Figure 2.8: Stereoscopy is traditionally accomplished by utilizing two cameras capturing images from slightly different perspectives.

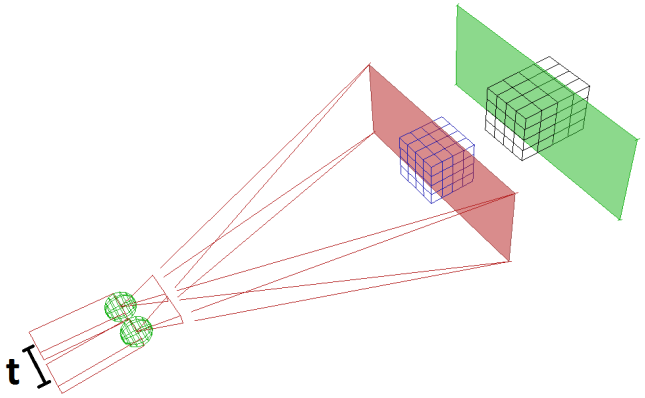
camera, and adheres to similar animation and transformation curves.

The amount of depth and the positioning the depth range are specified by two camera parameters: the camera separation (also known as interocular distance, internodal[1] distance, interaxial distance, etc) which is the distance between the first optical nodes of the two camera lenses, and the convergence distance which is the distance from the midpoint between the first optical nodes and the image plane at which there is zero parallax or disparity between corresponding points in the left and right images. This can be more simply described as the distance at which the optical axes of the cameras intersect.

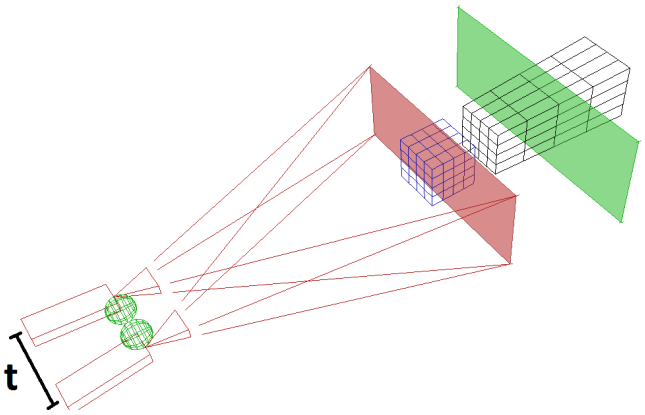
[1]Nodal points, along with focal points and principal points, are the three pairs of cardinal points along the optical axis which in an idealized symmetric Gaussian optical system completely determines the basic imaging properties. Technically, a lens has front and rear nodal points, and are determined such that a ray intersecting one of them will be refracted through the lens as to appear to originate from the other. However, for our purposes and for simplification we use the frequently misunderstood definition, common to the photographer's vernacular, that the nodal point of an optical system is the single point along the optical axis at which all light rays converge, collocated at the iris diaphram of the aperture.



(a) Camera separation = 1

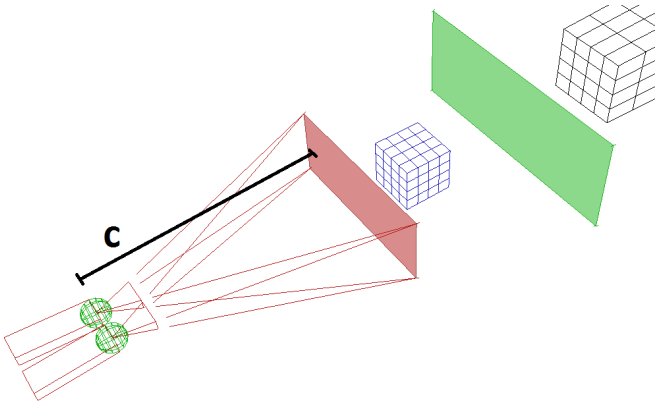


(b) Camera separation = 2.5

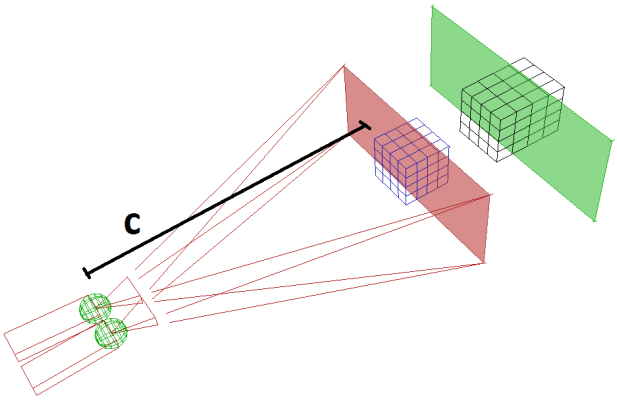


(c) Camera separation = 5

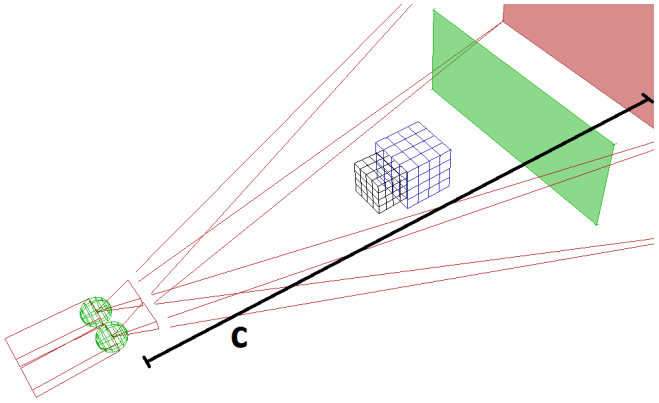
Figure 2.9: Generalizations of the camera separation at a constant convergence distance. The blue geometry indicates the original geometry with respect to the convergence distance plane and camera, in red. The viewer perceives the black geometry with respect to the screen, in green. Notice how the wider the separation between the cameras, the more exaggerated the depth, and the narrower the separation, the flatter the depth.



(a) Convergence distance = 20



(b) Convergence distance = 25



(c) Convergence distance = 50

Figure 2.10: Generalizations of the convergence distance at a constant camera separation. Again, the original geometry is in blue with respect to the convergence distance plane and camera in red, and the viewer perceives the black geometry with respect to the screen in green. Notice how the larger convergence distance shifts the perceived geometry nearer to the viewer while the smaller convergence distance shifts the geometry further from the viewer.

A simple and convenient generalization of these parameters is that the camera separation specifies the amount of depth, either more exaggerated or flatter, and the convergence distance specifies the positioning with respect to the viewer, either nearer or farther. As Figure 2.9 indicates, the greater the camera separation, the more exaggerated the perceived depth effect is by the viewer. Figure 2.10 indicates that a large convergence distance shifts the stereo precept closer to the viewer, while a smaller convergence distance pushes the precept farther from the viewer.

The convergence distance is determined and can be controlled via two distinct shooting methods; converged or parallel (Figures 3.1 and 3.2). Converged shooting is defined by toeing the cameras inwards via physical rotation so their optical axes converge at the convergence distance. Parallel shooting does not rotate the cameras, as the name suggests. Instead, the cameras remain parallel and shift the imaging sensors horizontally in the image plane relative to the lenses. This still has the effect of converging the optical axes of the cameras, but has the added benefit of not rotating the image planes; they still remain perpendicular to the stereo-pair viewing direction. Alternatively, the resultant parallel-captured images can be horizontally translated across each other to achieve the same effect, a common method used in live action filming where the translation of the physical image sensors is difficult. Visual manifestations of the distinction between parallel and converged will be explored later in the discussion on stereoscopic distortions.

CHAPTER 3

PREVIOUS WORK

3.1 Stereoscopic Transformation

One of the primary ways researchers have been able to explore how stereoscopically generated images are perceived and distortions manifested is to work through the geometric optical transformations from real world space to perceived virtual stereo space. This geometric transformation is generally understood and has been derived by many with slight variations and varying limitations [SS53, Lip82, WDK95, HB08]. Below we describe the a general purpose geometric transformation that accommodates for some variability and is adapted from the work of Woods et al. and Held and Banks.

First it is important to specify the stereoscopic camera and viewer position parameters needed to completely and uniquely define the capture and viewing geometry. In particular, six parameters are needed: three for the camera and three for the viewer. As discussed previously, the stereoscopic parameters include the camera separation and the convergence distance. Additionally, the field of view is required. In terms of the viewer configuration, some measure of display size, the position of the viewer with respect to the screen (including distance and any horizontal or vertical offset from the middle of the screen), and the interocular distance, or the distance between the viewer's eyes. Table 3.1 outlines the variable names and descriptions used in this baseline stereoscopic transformation definition.

Woods et al. considers the geometry of a stereo video system as a set of

Stereoscopic Transformation Parameters	
t -	Camera separation. The distance between the first nodal points of the two camera lenses.
c -	Convergence distance. The distance from the midpoint between the first nodal points of the two camera lenses and the convergence point. Also known as ZPS (Zero Parallax Setting, or Zero Pixel Shift). $$c = \frac{t}{2\tan\left(\beta + \arctan\left(\frac{h}{f}\right)\right)} \quad (3.1)$$
α -	Horizontal field of view.* $$\alpha = \arctan\left(\frac{\frac{W_c}{2} + h}{f}\right) + \arctan\left(\frac{\frac{W_c}{2} - h}{f}\right) \quad (3.2)$$
V_z -	Viewing distance. The distance from the viewer to the screen.
e -	Eye separation. The distance between the viewer's eyes.
W_s -	Screen width. The horizontal width of the display surface.
W_c -	Sensor width. The horizontal width of the camera imaging sensor.
f -	Focal length.
β -	Convergence Angle. The angle of inward rotation of the stereo cameras to achieve convergence. For converging camera configurations only.
h -	Sensor offset. The distance each camera sensor is shifted outwards from the optical axis of the lens to achieve convergence. For parallel camera configurations only.

* There is some ambiguity in the definition of the horizontal field of view for stereo cameras, in that there are actually three; one for each of the stereo cameras potentially angled inwards due to convergence, and an abstract one for a theoretical camera positioned between the two stereo cameras whose field of view is the overlap of the first two at the convergence distance. Equation 3.2 is for the third definition.

Table 3.1: Stereoscopic transformation parameters as defined by [WDK95, HB08].

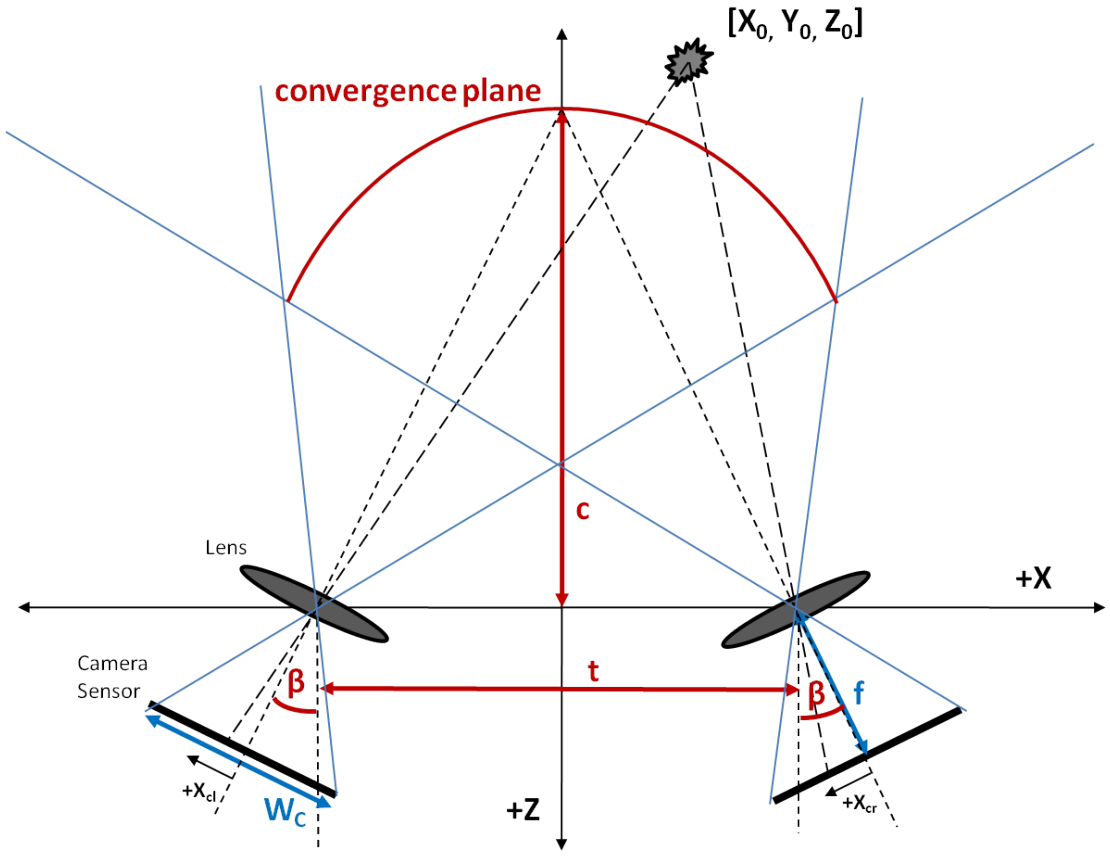


Figure 3.1: The stereoscopic configuration of converged camera parameters. Converged cameras converge at depth by angling the cameras inward, such that the optical axes, which remain perpendicular to their corresponding camera sensor, intersect at the convergence distance.

three separate coordinate transformations: from the X, Y, Z coordinates in camera space to the two (left and right) X, Y positions in the two camera sensors, then to the two (left and right) X, Y positions on the stereo display, and finally to the X, Y, Z coordinates in viewer space. Equation 3.3 and Table 3.2 identify and

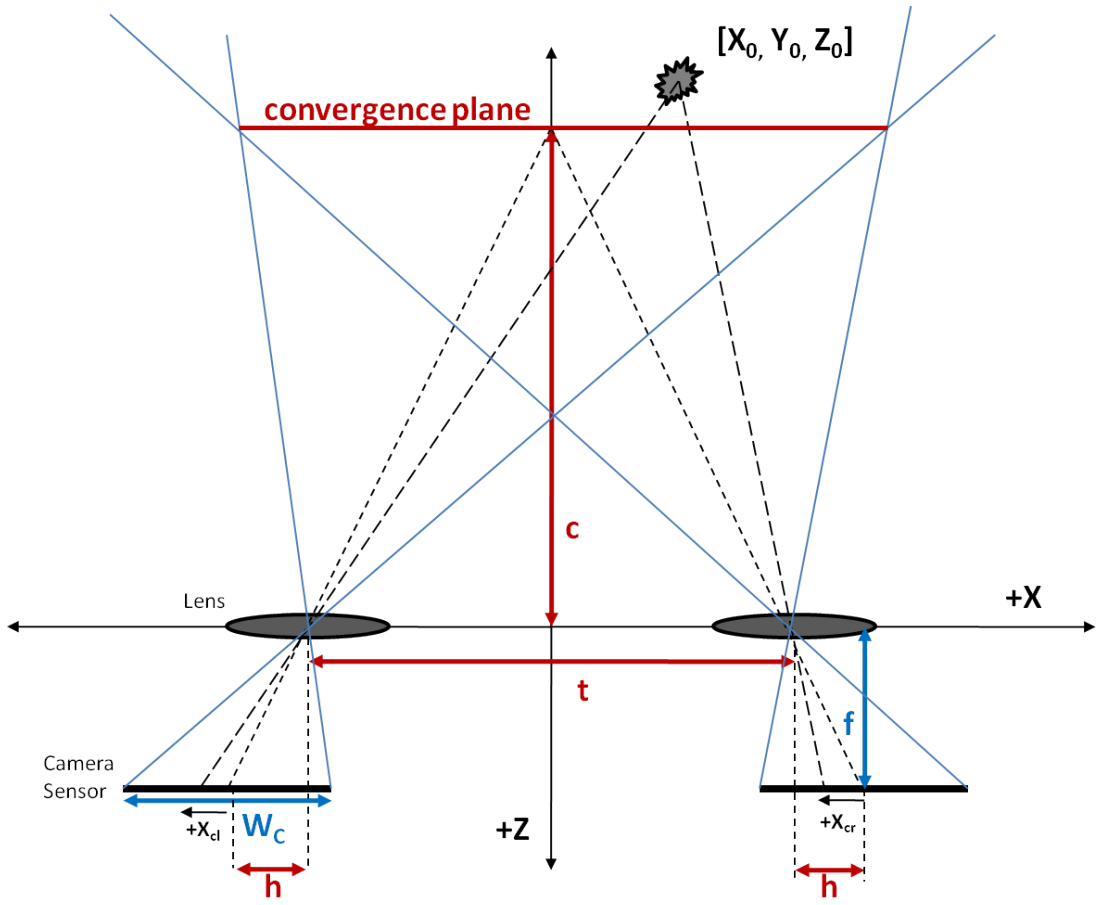


Figure 3.2: The stereoscopic configuration of parallel camera parameters. Parallel cameras maintain parallel camera sensors and converge the optical axes by offsetting the sensors outward with respect to the camera lens. This can also be achieved by translating the resultant images horizontally across each other after being captured in camera.

outline the various coordinate values, their descriptions, and valid spaces.

$$\begin{aligned} \text{World Space: } (X_w, Y_w, Z_w) &\rightarrow \\ \text{Camera Space: } (X_0, Y_0, Z_0) &\rightarrow \\ \text{L/R Camera Sensor Space: } (X_{cl}, Y_{cl}), (X_{cr}, Y_{cr}) &\rightarrow \\ \text{L/R Screen Space: } (X_{sl}, Y_{sl}), (X_{sr}, Y_{sr}) &\rightarrow \\ \text{Virtual Stereo Image Space: } (X_i, Y_i, Z_i) & \end{aligned} \tag{3.3}$$

Having defined the space and variables, each of the individual transfor-

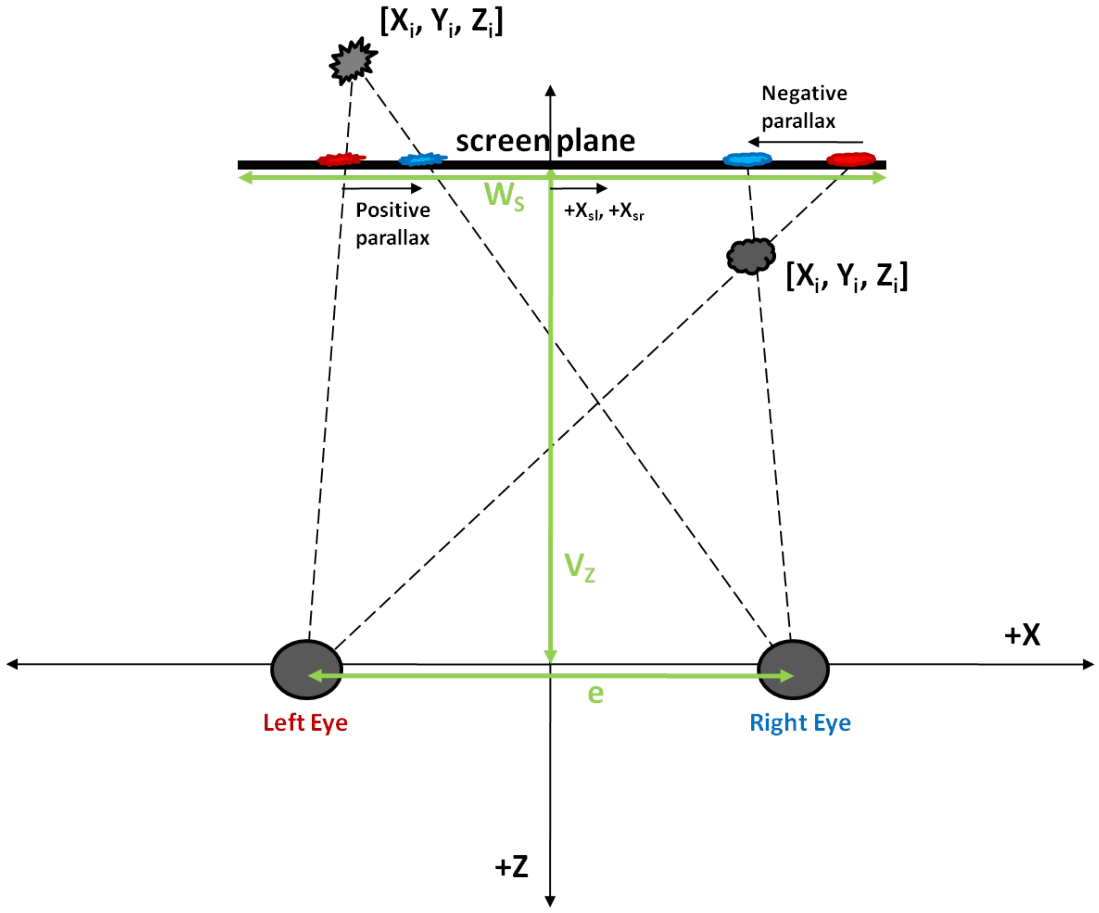


Figure 3.3: The stereoscopic configuration of the viewer parameters in the theater. Notice how objects presented with a positive parallax are perceived by the viewer as being behind the projection plane, while objects with a negative parallax are perceived as being in front of the screen.

mations can now be derived. From object space to camera sensor space, $(X_0, Y_0, Z_0) \rightarrow (X_{cl}, Y_{cl}), (X_{cr}, Y_{cr})$:

$$
\begin{aligned}
X_{cl} &= f \tan\left[\arctan\left(\frac{t + 2X_0}{2Z_0}\right) - \beta\right] - h \\
X_{cr} &= -f \tan\left[\arctan\left(\frac{t - 2X_0}{2Z_0}\right) - \beta\right] + h \\
Y_{cl} &= \frac{Y_0 \cdot f}{Z_0 \cos\beta + \left(X_0 + \frac{t}{2}\right)\sin\beta} \\
Y_{cr} &= \frac{Y_0 \cdot f}{Z_0 \cos\beta - \left(X_0 - \frac{t}{2}\right)\sin\beta}
\end{aligned}
\tag{3.4}
$$

	Stereoscopic Transformation Coordinate Spaces
(X_w, Y_w, Z_w) -	The X, Y, and Z world space coordinates of objects in the scene, including models, cameras, and vectors.
(X_0, Y_0, Z_0) -	The X, Y, and Z camera space coordinates of objects in the scene. This system is from the point of view of the cameras and is centered at the midpoint between the first nodal points of the two cameras.
(X_{cl}, Y_{cl}) -	The X and Y left camera sensor space coordinates of objects projected onto the sensor image plane.
(X_{cr}, Y_{cr}) -	The X and Y right camera sensor space coordinates of objects projected onto the sensor image plane.
(X_{sl}, Y_{sl}) -	The X and Y left screen space coordinates of objects. In general, this is a simple scaling of the left camera sensor space coordinates.
(X_{sr}, Y_{sr}) -	The X and Y right screen space coordinates of objects. In general, this is a simple scaling of the right camera sensor space coordinates.
(X_i, Y_i, Z_i) -	The X, Y, and Z virtual image space coordinates, as stereoscopically viewed by the observer.

Table 3.2: Stereoscopic transformation coordinate spaces as defined by [WDK95].

From camera sensor space to screen space, $(X_{cl}, Y_{cl}), (X_{cr}, Y_{cr}) \rightarrow (X_{sl}, Y_{sl}), (X_{sr}, Y_{sr})$:

$$\begin{aligned} X_{sl} &= X_{cl}\left(\frac{W_s}{W_c}\right) & X_{sr} &= X_{cr}\left(\frac{W_s}{W_c}\right) \\ Y_{sl} &= Y_{cl}\left(\frac{W_s}{W_c}\right) & Y_{sr} &= Y_{cr}\left(\frac{W_s}{W_c}\right) \end{aligned} \tag{3.5}$$

And from screen space to perceived virtual stereo image in viewer space,

$(X_{sl}, Y_{sl}), (X_{sr}, Y_{sr}) \rightarrow (X_i, Y_i, Z_i)$:

$$\begin{aligned} X_i &= \frac{e\,(X_{sl} + X_{sr})}{2\,(e - (X_{sr} - X_{sl}))} \\ Y_i &= \frac{e\,(Y_{sl} + Y_{sr})}{2\,(e - (X_{sr} - X_{sl}))} \\ Z_i &= \frac{eV_z}{e - (X_{sr} - X_{sl})} \end{aligned} \tag{3.6}$$

The full general geometric stereoscopic transformation can then be combined and defined, $(X_0, Y_0, Z_0) \rightarrow (X_i, Y_i, Z_i)$:

$$\begin{aligned} X_i &= \frac{\left(fe\frac{W_s}{W_c}\right)\left(\tan\left[\arctan\left(\frac{t+2X_0}{2Z_0}\right) - \beta\right] - \tan\left[\arctan\left(\frac{t-2X_0}{2Z_0}\right) - \beta\right]\right)}{2e - 4h\left(\frac{W_s}{W_c}\right) + 2f\left(\frac{W_s}{W_c}\right)\left(\tan\left[\arctan\left(\frac{t-2X_0}{2Z_0}\right) - \beta\right] - \tan\left[\arctan\left(\frac{t+2X_0}{2Z_0}\right) - \beta\right]\right)} \\ Y_i &= \frac{e\,(Y_{sl} + Y_{sr})}{e - 2h\left(\frac{W_s}{W_c}\right) + f\left(\frac{W_s}{W_c}\right)\left(\tan\left[\arctan\left(\frac{t-2X_0}{2Z_0}\right) - \beta\right] - \tan\left[\arctan\left(\frac{t+2X_0}{2Z_0}\right) - \beta\right]\right)} \\ Z_i &= \frac{eV_z}{e - 2h\left(\frac{W_s}{W_c}\right) + f\left(\frac{W_s}{W_c}\right)\left(\tan\left[\arctan\left(\frac{t-2X_0}{2Z_0}\right) - \beta\right] - \tan\left[\arctan\left(\frac{t+2X_0}{2Z_0}\right) - \beta\right]\right)} \end{aligned} \tag{3.7}$$

There are several limitations to this general form of transformation. Primarily, it does not accommodate for all variability in stereoscopic configurations such as post rendered image translation, translation and convergence weighting between the left and right cameras, viewer offset from the center of the screen, gaze direction, allowed divergence, and stereoscopic windowed masking, among others.

3.2 Stereoscopic Distortions

Initially explored and documented by Raymond and Nigel Spottiswoode in the 1950's, stereoscopic distortions are deviations of a viewer's virtual perception

of a stereoscopically projected space from their perception of the same space if they were physically present [SS53]. They observed that the reduction of the illusion of three dimensions to the single minor factor of convergence, especially during the Victorian era of small, single viewer stereoscopes, would not scale to larger demonstrations. They suggested the need for a more variable and delicate system of capture and transmission to regulate the apparent distances of images to within physical and comfort limits of the human observer. Additionally they realized that a completely faithful reconstruction of the original physical space was not necessary, and that "the recreated [stereoscopic] image is no more a faithful reproduction of reality than is a 2-dimensional rendering of a film star in a close-up whose head may be twenty feet high, ten feet across, but having no depth whatever". This would allow for artistic and creative flexibility within the realm of comfortable stereo viewing.

As a result of their derivation of the geometric stereoscopic transformations, Andrew Woods et al. were able to objectively detail what the Spootiswoode's had begun to describe 40 years earlier [WDK95]. They observed and discussed the geometric origins, characteristics, and effects of image distortions in stereoscopic video displays, including depth plane curvature, depth non-linearity, depth and size magnification, shearing distortion, and keystone distortion.

One of the more interesting observations is related to the differences between converged and parallel shooting. While both methods can obtain equal convergence distances, toed-in cameras result in a curvature of the depth planes due to the skewed, non-parallel, nature of the left and right image planes. Accordingly, the corners of the images will appear further away from the viewer than objects in the center. Parallel configured cameras have depth planes parallel to

each other and to the surface of the display, thus they do not induce the curved warping seen in Figure 3.4.

Keystone distortion occurs when the projection plane is not parallel to the camera sensor plane or when the two camera sensor planes are not parallel to each other, situations that arise in the converged shooting configuration. This configuration causes trapezoidal image projections, as seen in Figure 3.5. These can traditionally be accommodated for or fixed in monoscopic displays; however keystone distortion causes vertical parallax in stereoscopic images due to the image sensors being on different planes. Accommodation for this sort of distortion can be painful. Also note that a certain amount of horizontal parallax is inherent in keystone distortion as well, and this is the cause of depth plane curvature.

Vertical parallax is also induced by radial lens distortions, such as barrel or pin-cushion distortion. It is caused by the use of spherical lens elements which causes different focal lengths at various radial distances from the center of the lens. In pin-cushion distortion, the focal length increases from center, and in barrel distortion it is the opposite. It was found that radial distortion is most prevalent in shorter lenses due to the extreme optical requirements and glass lens manufacturing processes. Aspherical lenses are capable of reducing the amount of this distortion, as well as digital post-processing techniques. However there are still issues involving other types of lens aberrations and other variances from the idealized geometrical optics simplification and abstraction model (known as paraxial optics), whose impact can be somewhat reduced by utilizing sequentially manufactured lenses. It should be noted that in the fields of computer graphics and animation, simplified idealized lenses are modeled

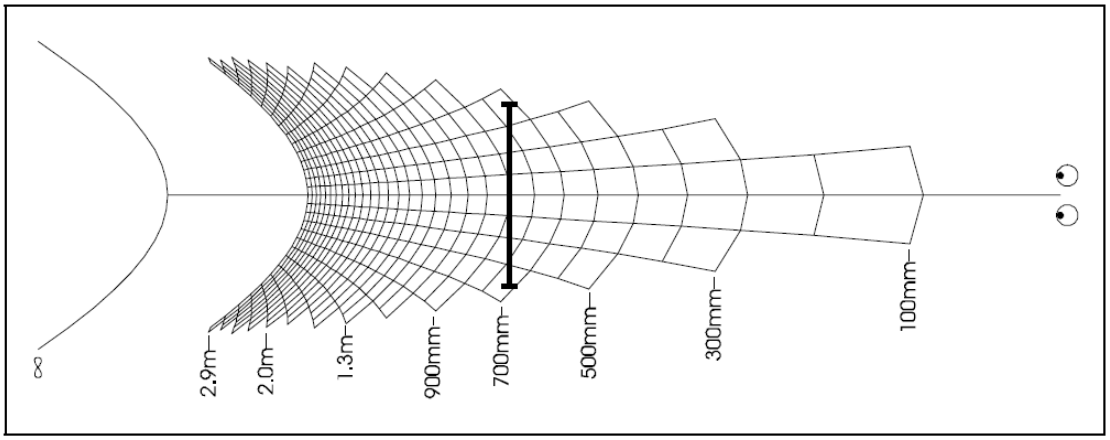


Figure 3.4: Curvature of the depth plane as a result of the toed-in camera configuration. Parallel depth planes in world space are warped when transformed into the virtual viewer stereo space when converged cameras are used. [WDK95]

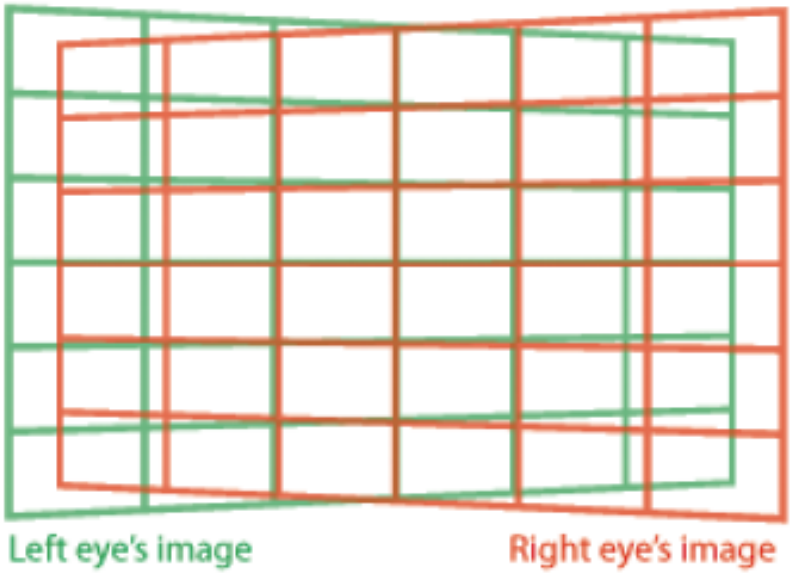


Figure 3.5: Keystone distortion as a result of converged stereo camera configurations. The trapezoidal projections create both vertical and horizontal discrepancies which can lead to visual discomfort and inaccurate depth warping. [HB08]

and used for computation which are free of such artifacts.

Another key observation regarding stereo distortion is that perceived depth as a function of object depth is non-linear (Figure 3.8). Simple examination of the denominators in the derived stereoscopic transformation equations highlights the non-linear and complex nature of the transformations. This non-linearity of depth can lead to shape warp and false estimations of velocity, where an object moving at one velocity may be perceived in stereo to be moving at a different velocity or at a changing velocity due to the non-linear depth transformation. As a result, any disparities are magnified at greater Z_0 values, compounding any perceived distortions unrelated to depth non-linearity [SGC+02]. "A linear relationship between image depth and object depth can only be obtained by configuring the stereoscopic video system such that object infinity is displayed at image infinity on a stereoscopic display" [WDK95]. This type of configuration, however, is difficult to achieve and induces alternate distortions and artifacts.

Distortions in shape perception occur as a result of magnification mismatches between depth in the Z_0 direction and the X/Y planar cross section. Because the perceived virtual stereo image space coordinates X_i and Y_i can be reduced to scaled values of the object space coordinates X_0 and Y_0, the shapes on the X/Y planar cross sections of the perceived images are the same as the original object space shapes, differing only in size. The independent scaling in the Z direction along with the non-linear relationship between image and object depth compresses or expands the X/Y cross sections, creating the magnification mismatches. This can lead to perceived flattening (or cardboarding) or stretching of the image, as can be seen if Figure 2.9(a). [SGC+02, WDK95]

Son et al. [SGC+02] also point out that, compared to 2D images, stereoscopic

images are subject to greater distortion, especially when viewer position is considered. One of the primary distortions related to viewer positioning is shear distortion, where objects appear to pivot about the display surface as a result of the sideways movement in the observer position with respect to the screen [WDK95]. See Figure 3.6(d) and (f) for an example of shear distortion. Son et al. point out that no quantitative analysis relating to viewer position had yet been determined, and subsequently they and others mathematically modeled shear distortions and incorporated them into the stereo transformation equations. [SGC+02, HB08]

Held and Banks [HB08] attempt to visualize these types of distortions by comparing the virtual stereo image projections to the original objects being captured in a variety of different camera/viewer configurations.

They first identified what they call the "proper viewing conditions" for the real world and perceived images to match, which they define as having the eyes positioned at the corresponding center of projections, and the vergence matching between real world and virtual space. These conditions are known as orthostereoscopic, where the captured space is exactly duplicated in the virtual space as perceived by the viewer, and is dependent upon the display size. Held and Banks then independently varied the viewing parameters from the orthostereoscopic conditions to observe the effects of the estimated 3D percept.

Figure 3.6 indicates their results in plan view. The original stimulus is shown in white and the estimated 3D stimulus in blue, while the screen is shown in green and optical axes of the cameras in red. (E) is the proper viewing condition, while (B) and (H) show variations in distance from the screen. (D) and (F) show the off-centered viewing conditions that result in shear distortions, (A) and (I)

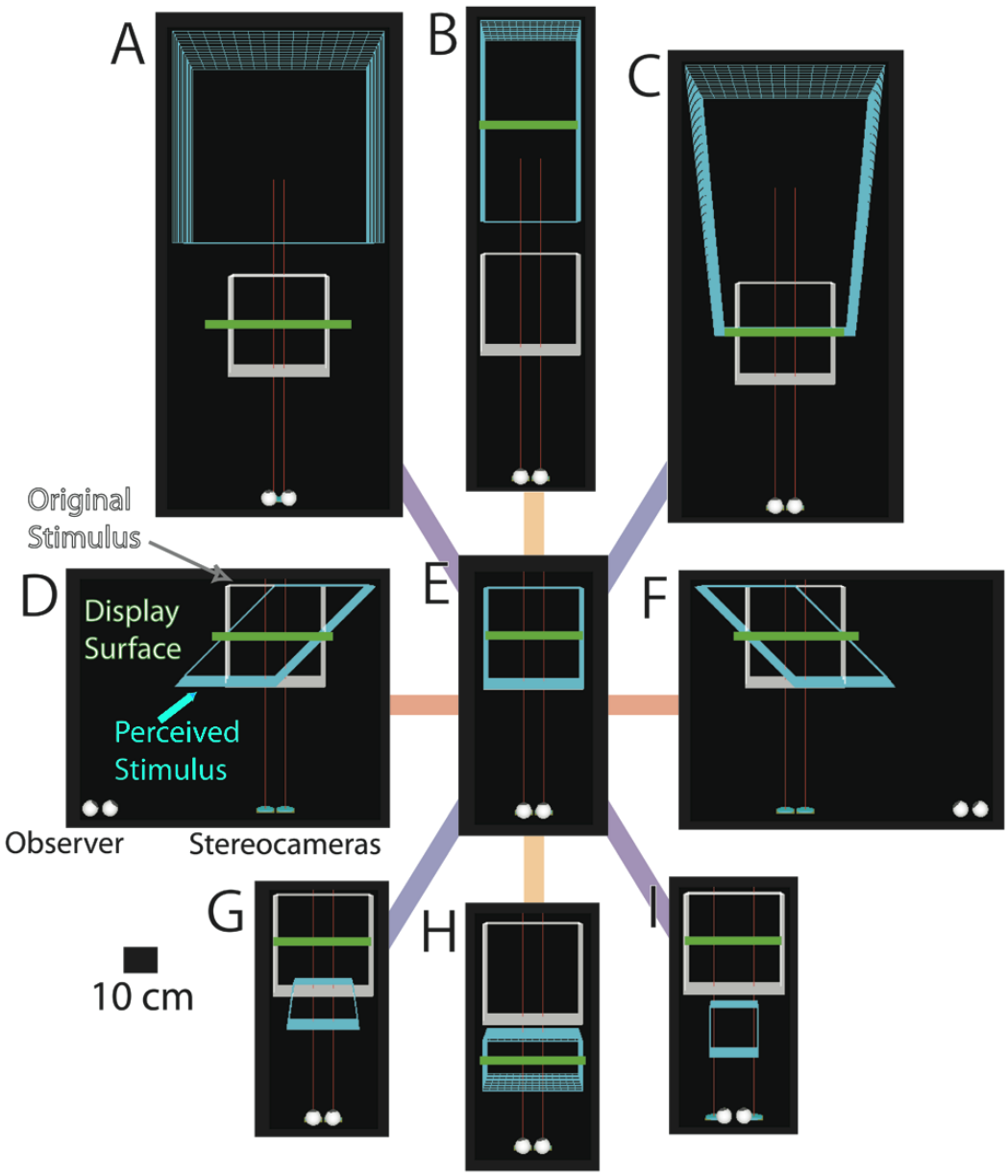


Figure 3.6: Independent variations of viewing parameters and the resulting effects of the estimated 3D precept, as visualized by Held and Banks. The original stimulus is shown in white with the estimated 3D stimulus in blue, while the screen is shown in green, the optical axes in red, the observer in white, and stereo cameras in blue. (E) shows the proper viewing condition. (B) and (H) show variations in viewing distance from the screen. (D) and (F) show off-centered viewing conditions and the resulting shear distortions. (A) and (I) show mismatches in the interocular distance between camera and viewer. (C) and (G) show mismatches in the convergence distance. [HB08]

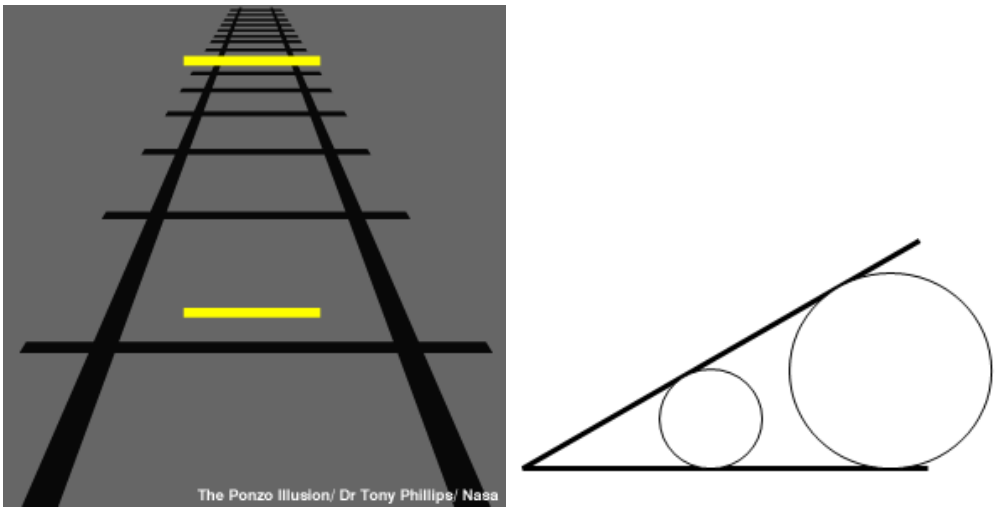
show mismatches in the interocular distances between camera and viewer, and (C) and (G) show mismatches in the convergence distance.

Finally, it has been observed that the non-linear nature of the stereoscopic depth transformation leads to misperceptions in depth, size, and velocity. Kusaka ran psycho-physical experiments with a half mirrored surface in front of a stereo screen and movable LEDs to gauge perceived distance and came to the same determination of the non-linear transformation between world space and image space [Kus92]. They then explored the implications of size perception as it relates to the Ponzo illusion and determined that due to inaccurate perceptions of absolute depth, an object that maintains the same retinal image size will be inaccurately judged in terms of its size. The Ponzo illusion, as demonstrated by Mario Ponzo, suggests that the size of an object as judged by the human mind partially depends on the background, and that objects that are judged at a greater distance will be perceived to be larger if they maintain the same retinal image size. Figure 3.7 provides an example and geometric explanation of this illusion. Diner explored other implications of the non-linear depth transformation as it applies to motion and determined that it leads to inaccurate judgments on velocity and acceleration [Din91].

3.2.1 Physical and Perceptual Limitations

Explorations in the limitations of stereoscopic viewing, both physical and perceptual, have also been conducted.

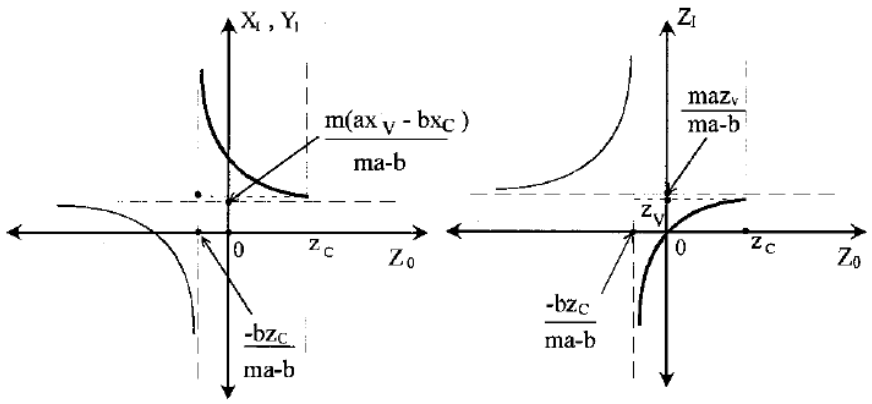
Son et al. re-derived the stereoscopic transformation equations as functions of the original object depth value (Z_0) and the magnification of object size be-



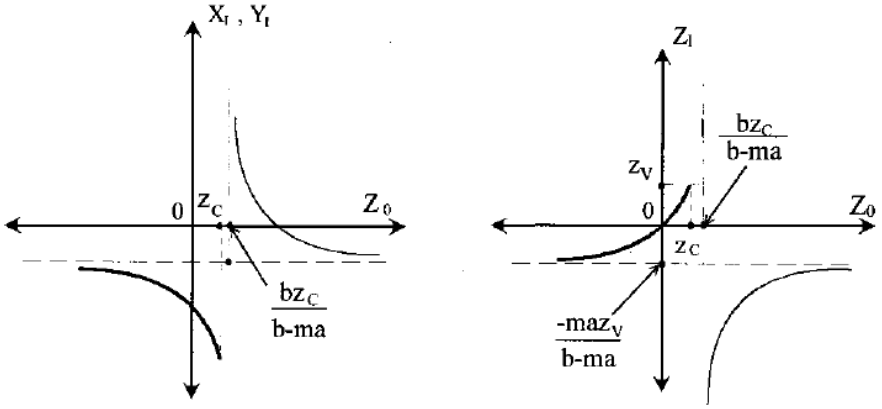
(a) (b)

Figure 3.7: The Ponzo illusion. Figure (a) shows how two lines of equal length are perceived to have different sizes based on the perceived distance along the perspectively viewed tracks. Figure (b) is the geometric explanation of the illusion that demonstrates how images with the same retinal image size, or constant subtended angle, will be perceived as different sizes if perceived at different distances.

tween the original image and perceived image ($m = W_s/W_c$) in order to identify asymptotic limit values [SGC+02]. They identify near and far limitations on the depth and position of real world objects and their corresponding virtual space positions based on the camera and viewer configurations. Their results are summarized in Figure 3.8, and generally identify the geometric limitations in the form of asymptotic limits. These limits often have obvious and practical reasonings behind them, such as the necessity for captured geometry to appear in front of the camera and appear in front of the viewer and not cause an outward divergence of the viewer's eyes. Notice the change in naming conventions from our own. In particular, note that the camera separation $t = 2a$, eye separation $e = 2b$, and the subscripts O, C, V, I are for the object, camera, viewer, and image positions, respectively.



(a) Camera Separation >Interocular



(b) Camera Separation <Interocular

Figure 3.8: Physical geometric limitations of stereo viewing based on camera and viewer configurations, in particular the object depth (Z_0) and image magnification ($m = W_s/W_c$). The left diagrams indicate the relationship between the original object depth (Z_0) and the perceived planar position (X_i, Y_i), while the right diagrams indicate the relationship between the original object depth (Z_0) and the perceived object depth (Z_i). Notice how all parameters are non-linear with respect to the original object depth and also how the asymptotes of each diagram correspond to a computable physical geometric limitation. [SGC+02]

It is of critical note that this study did not accommodate for perceptual allowances, such as the brain's ability to fuse geometrically impossible conditions [HB08, WDK95, Ogl40, ABM+97] or the eyes' ability to diverge outward to a slight degree [Pas95, YS90, CES88, Stu34, WDK95]. Also, this study assumed a parallel, unconverged camera configuration, which is a source of many of their geometric limits.

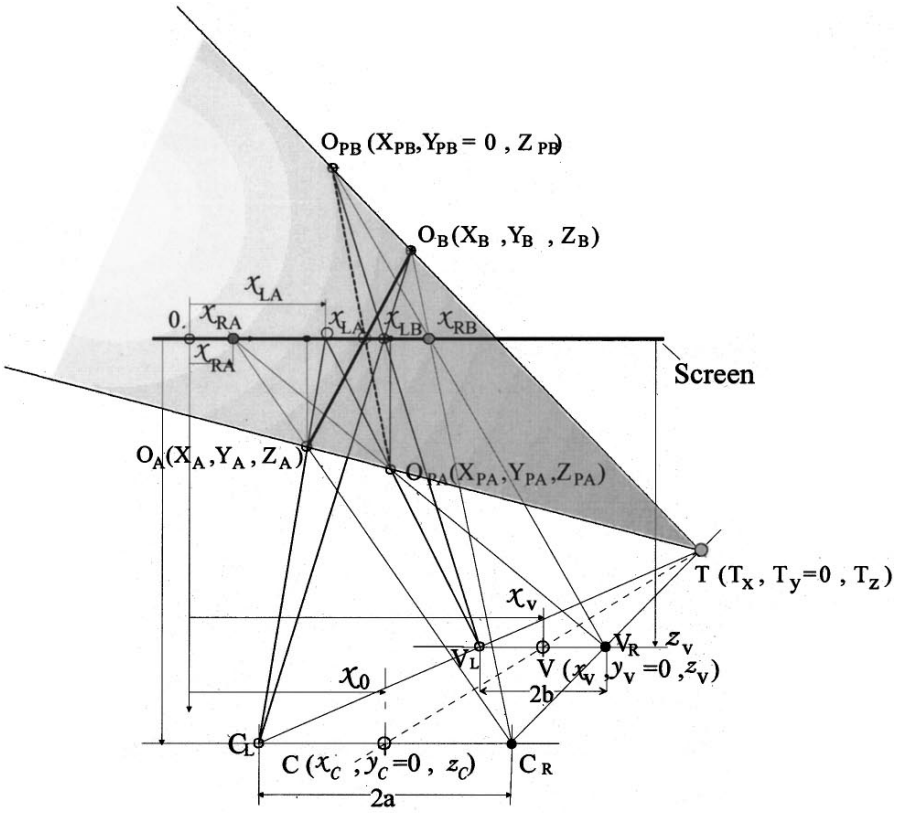


Figure 3.9: Definition and visualization of the stereoscopic transformation boundary volume as defined by [SGC+02]. The shaded region is the boundary volume wherein all stereo-perceived objects will appear to the viewer.

Son et al. also identified a transformation boundary volume, whereby all object space points transformed into stereo image space points lie within a predefined volume (Figure 3.9). By defining the point $T = (T_x, T_y, T_z)$ as the intersection of lines connecting the cameras (C_l/C_r) to their corresponding viewer eyes (V_l/V_r), they were able to identify the transformed boundary volume by projecting from the point T to the original real world boundary defined by the external border of the object. The illustration provides an example of this transformation boundary volume (shaded region), identifying the left and right camera and viewer positions, the point T, and two points (A and B) in both their original (O_A) and transformed (O_B) positions. The shaded transformation boundary volume indicates the possible valid locations where transformed objects can appear

to the viewer and is a helpful diagrammatic way of identifying limits where a full transformation of all object points is not possible.

Another study on the limits of stereoscopic perception was conducted by Ishikawa and Geiger [IG06]. They investigated the perception of shape, and how the ambiguity arising from noise, periodicity, and large regions of constant intensity is resolved and how missing data is interpolated. This ambiguity makes it impossible in general to identify all locations of the two images with certainty. They suggest that any convincing model of stereopsis should detail how ambiguity is resolved and missing data interpolated.

They tested this by generating ambiguous stereoscopic silhouettes or boundary contours, as in Figure 3.10(a). The edges are the only locations that can be matched perfectly and provide exact depth data. They then asked participants to identify which of the three possible resolutions to the ambiguity (Figure 3.10(c)-(e)) they observed. Due to the variation and ambiguity in their results, their findings suggest a refutation of the traditional theories of depth interpolation, including one-dimensional information for depth prediction (horizontal depth interpolation), as well as well as gradient minimization models, convex models, and energy minimizing models.

Ishikawa and Geiger then suggest a minimum disparity gradient or similar model except when overridden by strong prior preference for special spatial features, such as the preference of human perception for elliptical shapes over hyperbolic. They note that the preferred shapes tested were parabolic.

They also propose a Zero Gaussian curvature model, which is characteristic of parabolic points. Such surfaces are developable (can be made by rolling

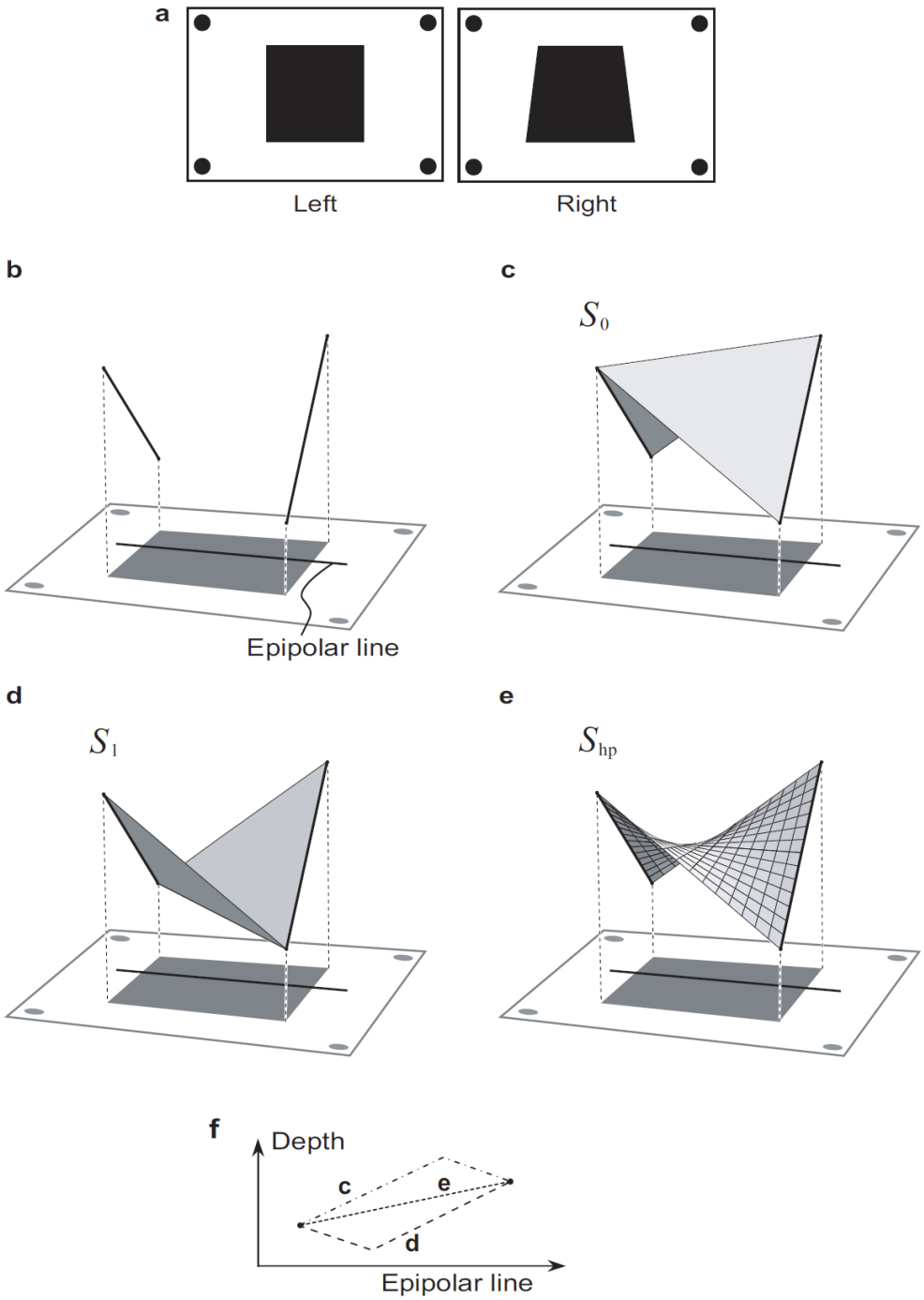


Figure 3.10: An example of the ambiguous stereoscopic silhouettes or boundary contours of the study conducted by Ishikawa and Geiger. (a) shows the stereoscopic stimulus, and (b) indicates the edges, which are the only perfectly discernable depth positions. (c), (d), and (e) indicate possible depth interpolation precept models, and (f) shows the possible illusory volumes along the epipolar line. [IG06]

and bending paper), indicating human vision system attempts to fit a boundary to a wire frame. Functionality would be neither concave nor convex, and is nonlinear which indicates it is dependent on starting location.

A subset of the Zero Gaussian model is the convex hull model, where the perceived surface can be described as the result of wrapping the discernable points and lines with elastic or spandex, creating a minimal surface. This suggests to the authors that the precept of these surfaces arise from later stages in the human visual process and is the simplest 3D shape that is compatible with the data. They conclude by admitting that no developed model or theory completely and accurately matches human vision or the results of their study, and conceded ambiguity in all of their proposed models.

3.2.2 Skewed Rays

As discussed previously, the issue of geometric versus perceptual stereo transformation and interpretation has been identified as an issue in previous stereo perception research. Banks et. al. explore this distinction between geometric and non-geometric distortions in 3D shape perception in stereoscopic displays [HB08]. The standard geometric model involves determining the retinal images by projecting from the stereo images to the viewer's eyes, and the ray intersection determines the 3D shape percept.

Held and Banks point out, however, that there are three instances when this model fails, or, specifically, when the back-traced rays do not intersect and no precept point can be defined: yaw rotation of the viewer, roll rotation of the viewer, and mismatch between camera convergence and the way the stereo im-

ages are displayed.

The non-intersecting rays lead to vertical disparities in the retinal images, and Held and Banks suggest that these disparities are "crucial signals" in the visual system's interpretation of stereo images, pointing out that while many viewing situations do create non-converging rays and vertical disparities, viewers are often able to perceive a coherent 3D scene. This led them to infer that more than just the standard geometric model for distortions is at work within the visual system.

Their findings regarding the vertical and horizontal disparities associated with yaw rotation, roll rotation, and converging cameras are summarized in Figure 3.11, showing the disparity vectors for points on the front ((a), (b), (c)) and back ((d), (e), (f)) planes of a cube when viewed with the varying changes in viewing conditions.

3.2.3 Epipolar Geometry

In an attempt to integrate vertical disparities into a model for stereoscopic viewing, Held and Banks [HB08] define skewed rays using epipolar geometry. Epipolar geometry is defined by the geometric relationships between 3D points and their projections onto 2D image planes, such as corresponding points in stereo images on the projection plane or the back of the retinas. One useful element of epipolar geometry is the epipolar plane which is defined by a point in space and the two centers of the points projection at the viewer's eyes [Figure 3.12(a)].

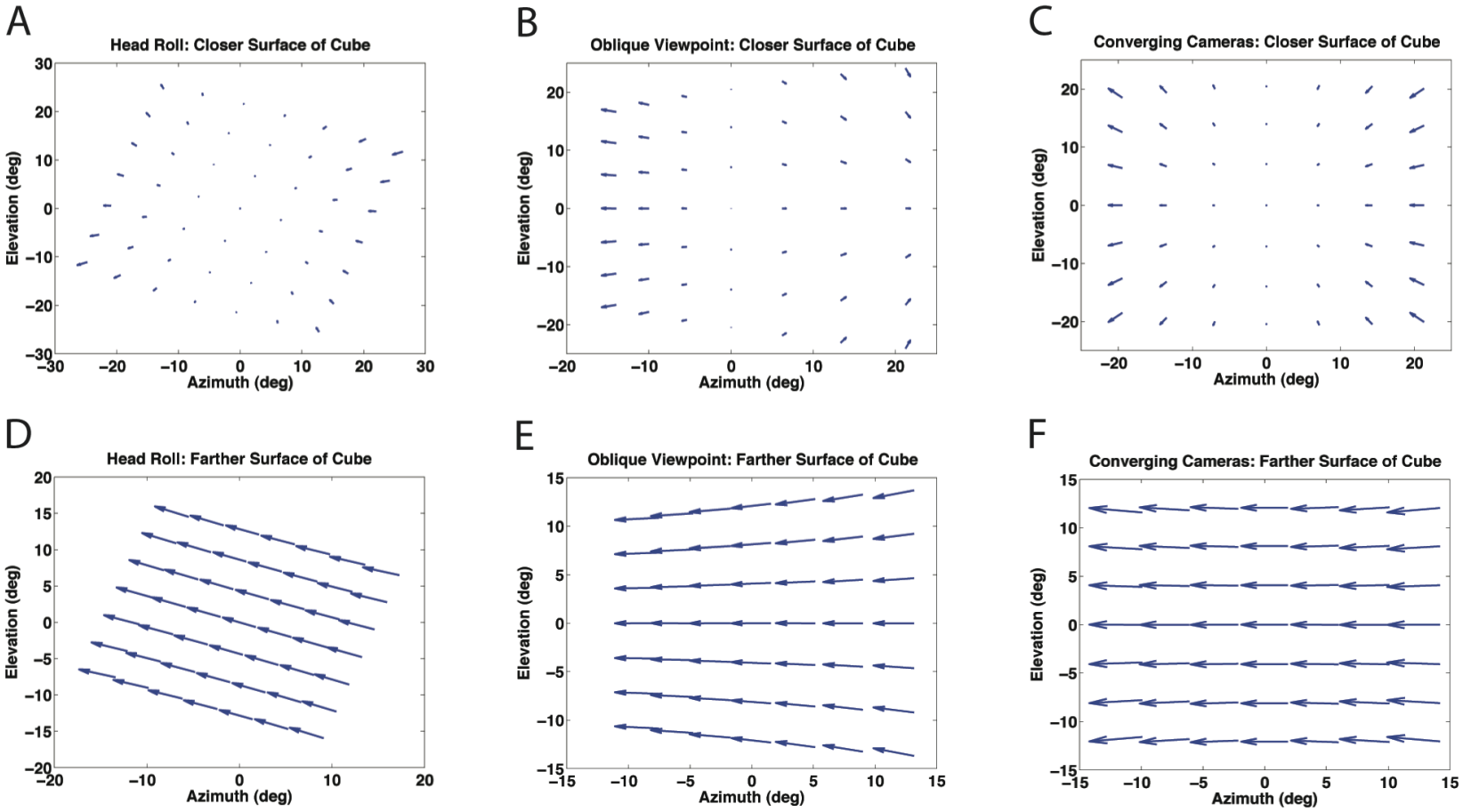


Figure 3.11: Disparity fields produced by improper viewing. The vectors represent the direction and magnitude of retinal disparities produced by the stereoscopic visualization of a cube. Any vector that is not horizontal results in geometrically ambiguous vertical disparities. (A) and (D) are the results of head roll, (B) and (E) are the results of yaw rotation, and (C) and (F) are the results of convergence mismatch. [HB08]

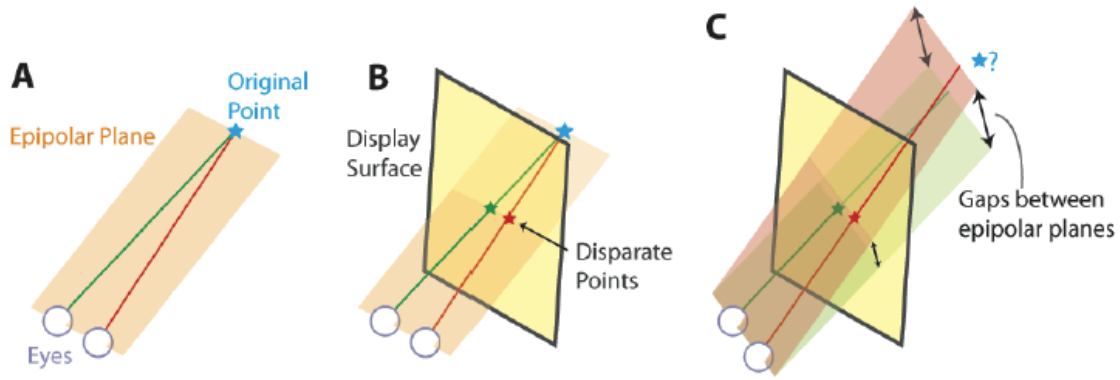


Figure 3.12: Demonstration of epipolar geometry and its applications to skewed rays, by Held and Banks. (a) demonstrates natural viewing, where the epipolar plane is defined by a point in space and the centers of the two eyes. (b) shows how for properly viewed stereoscopic images, the correspondence points lie in the same epipolar plane and thus the corresponding rays intersect in space. (c) shows how for obliquely viewed stereoscopic images, correspondence rays often lie in different epipolar planes and may never intersect in space. [HB08]

Two corresponding points in a stereo picture produce intersecting rays as long as they lie on the same epipolar plane and are non-parallel, as in Figure 3.12(b). Standard transformations described within the classic geometric stereoscopic transform model may rotate or translate the epipolar plane, but these rays still remain epipolar, continuing to provide geometrically valid results. However, rotation of the viewer's head produces rays in different epipoloar planes, as in Figure 3.12(c), resulting in non-intersecting rays and geometrically indeterminable stereo precept points. Note that, as previously discussed, this only applies to yaw and roll rotations, as pitch rotations of the head do not affect the epipolar geometry or the stereo precept by the viewer since the rotation axis is contained within the epipolar plane. Mismatches between camera and display surfaces can also produce differing epipolar planes, such as the case with converging cameras, as the image sensors lie on different planes yet the projected display surface is on one plane and traditionally none of them are parallel with another.

Banks identifies Woods' seminal 1995 paper [WDK95] as the only previous work that attempts to incorporate skewed rays into a geometric model and discusses the possible perceptual consequences. Woods et al. justified vertical disparities by averaging the vertical Y coordinate in mismatching geometric points. However, Banks points out that vision science findings suggest that this 3D estimate does not match human percepts, and explicitly does not apply to yaw and roll rotations.

An example of how stereo mismatches are managed and accommodated to by the human visual precept is the "induced effect". The induced effect occurs when a vertically magnifying lens is placed in front of one eye, creating non-zero vertical disparities. The effect of this is a perceived slanting of the frontoparallel surfaces, despite non-induced horizontal disparities. This led vision scientists to accept that 3D percepts are a product of both horizontal and vertical disparities. [HB08]

Adams et al. conducted a study into the adaptation to the "induced effect" over a period of several days. They found that the process of adaptation and diminishing the perceived distortions is a purely visual change rather than a remapping between the visual stimuli and motor responses. They found that in order to restore veridical perception, the determination of slant from disparity must be recalibrated instead of reweighting or monocular adaptation. [ABVE01]

Backus et al. points out two stereopsis models that may help to incorporate this type of slant precept into a model for stereoscopy. One is based on measuring horizontal disparities and eye position, whereby the slant S of a surface is determined by the horizontal size ratio (HSR - the ratio of horizontally subtended angles between the left and right eyes), the eyes version (γ - off-center

gaze angle), and vergence (μ). [BBvEC99]

$$S \approx -\arctan\left(1/\mu \cdot \ln HSR - \tan\mu\right) \tag{3.8}$$

The other method is based on the horizontal and vertical disparities, without relying on the eye position.

$$S \approx -\arctan\left(1/\mu' \cdot \ln HSR/VSR\right) \tag{3.9}$$

where μ' is the vergence derived from changes in the vertical size ratio (VSR). Banks points to papers suggesting that the visual system uses a weighted average of the two stereoscopic methods when estimating surface orientation from binocular disparity. [BBvEC99]

3.3 Evaluation of Stereoscopic Images

In order to generate the highest caliber stereoscopic images, we need ways to subjectively and objectively evaluate them. The following is a review of previous research in the methodologies and results of a variety of techniques of evaluating stereoscopic images.

3.3.1 Quality and Naturalness of Depth

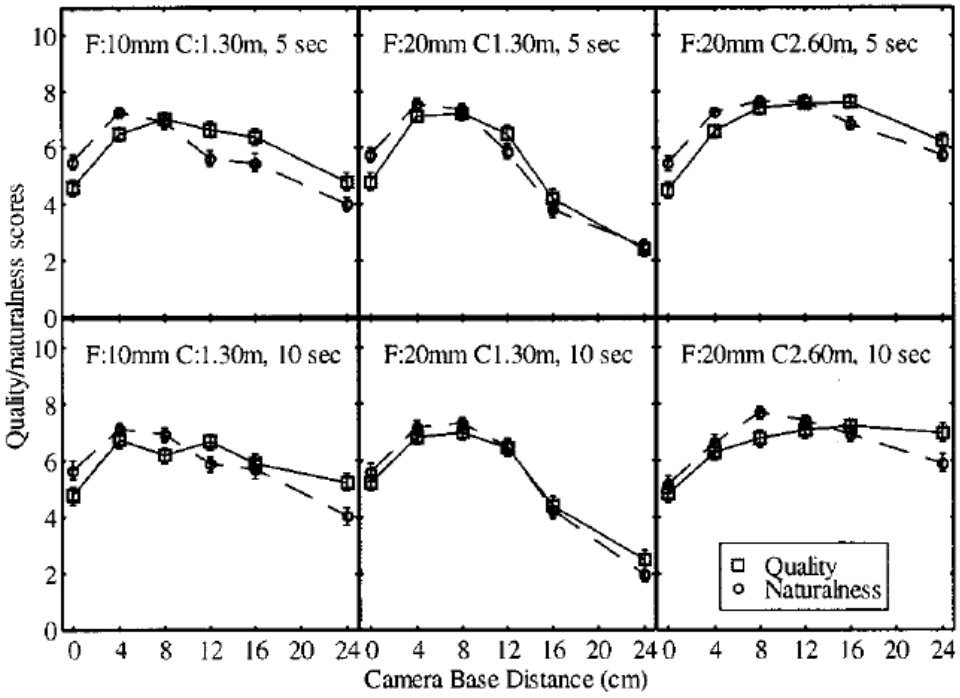
Multiple studies have been conducted investigating the influence of image disparity, convergence distance, and focal length on the subjective assessment of

depth, naturalness, quality, and eyestrain of the resultant stereoscopic images [IdRH98b, IDRV02, TSC98]. In most cases, studies utilized the RANDOT random dot stereo test from Stereo Optical Co to verify user stereoacuity. To distinguish between naturalness and quality as subjective evaluation, subjects were told that naturalness is what they perceive as a truthful representation of reality, while quality is a subjective preference scale. It has been indicated that subjects are able to differentiate between the two [IdRH98b].

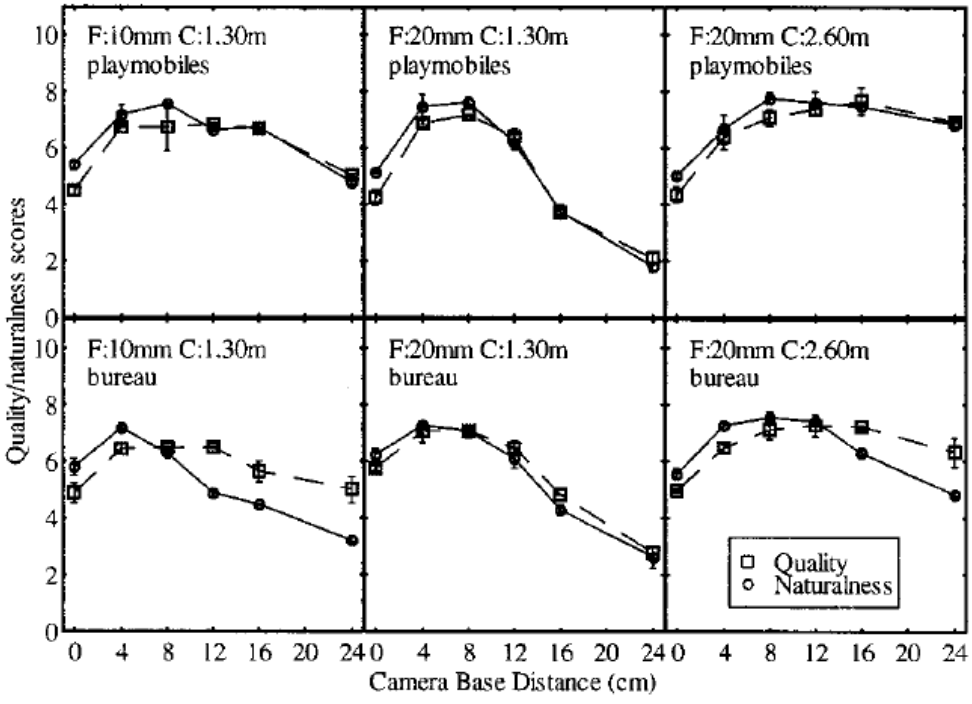
In all studies, a strong preference for stereo over non-stereo images was found. This was expanded upon in one study, in that the preference was diminished if stereoscopic artifacts were noticeable [TSC98]. Another group found that both quality and naturalness of depth increase with the transition from a monoscopic to a stereoscopic mode of presentation within natural bounds of disparities [IDRV02].

In the studies conducted by IJsselsteijn et al., the naturalness and quality evaluations peaked at an intermediate camera separation distance [IdRH98b, IDRV02]. The results are highlighted in Figure 3.13. The authors attribute the decrease in quality and naturalness ratings at larger camera separations to extreme disparity values. They also found that there was a strong linear relationship between naturalness and quality (r=0.96), however a small systematic shift was observed.

There was also a difference between configurations with similar ratios of focal length to object distance, attributed to the greater presence of keystoning due to toe-in at the shorter focal distances, especially at larger stereo bases. The authors calculated that the vertical disparity caused by keystoning perceptually results in depth plane curvature which may have a negative impact on observers



(a) Exposure Time Based



(b) Scene Based

Figure 3.13: Results of quality and naturalness evaluations as a function of stereo parameter variation by IJsselsteijn et al. Note that in the diagrams F is the focal length and C is the convergence distance of the stereo cameras. [IDRV02]

appreciative judgments.

IJsselsteijn et. al. also ran experiments to determine if there is any variation in quality and naturalness over image display duration. Displayed images were of two scenes and varied over durations of 1, 3, 5, 10, and 15 seconds. The results showed no significant main effect of display duration.

Few studies on subjective evaluation of stereoscopic images emphasize the impact of variations in scene content over parameter configurations beyond the brief mention of the ratio between the focal length and object distance.

3.4 Visual Comfort and Fatigue

Another way to evaluate the quality of stereoscopic images is to observe visual comfort and fatigue while viewing stereo images. Visual fatigue and discomfort can be defined by a wide range of visual symptoms, including tiredness, headaches, soreness of the eyes, eyestrain, dried mucus or tears, feeling of pressure in the eyes, aches around the eyes, discomfort when the eyes are open, hot eyes, difficulty in focusing or blurred vision, and stiff shoulders [UH08]. The visual fatigue symptom of eye strain, known as asthenopia, indicates the existence of elastic distortion caused by stress. Visual fatigue can be caused by a demand on the early visual functions or central cortical structures.

In terms of viewing stereoscopic images, visual discomfort and fatigue can be caused by anomalies of binocular vision, binocular asymmetries, accommodation-convergence mismatch, and excessive disparity. Asymmetries are either optical, geometric, or filter-based in nature. Geometric asymmetries

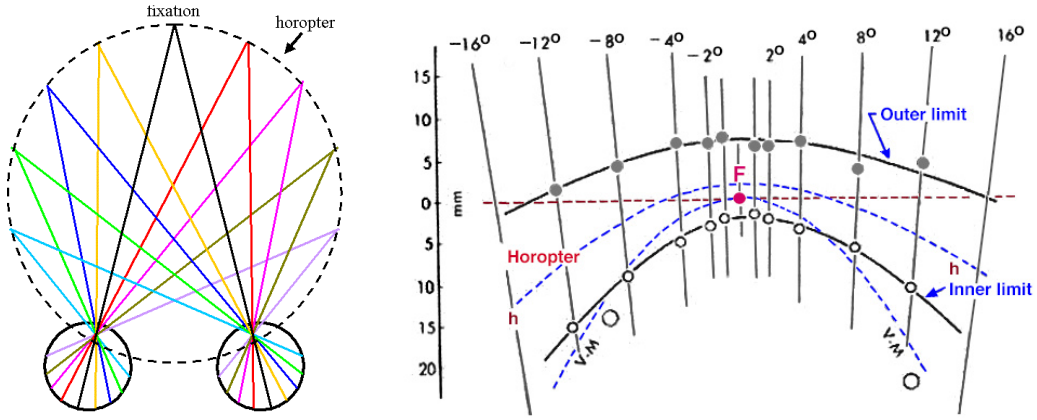
include vertical image shift, as well as rotation, magnification, size, and distortion mismatches. Asymmetric filter characteristics include luminance or contrast mismatches, color differences, and cross-talk, where cross-talk is when the image intended for one eye bleeds into the precept of the other eye, often caused by high contrast in certain 3D display technologies.

Excessive disparity occurs when the projected image disparity is wider than the viewer's interocular base, forcing the viewer to diverge their eyes outward to an uncomfortable degree. Some research has been done to identify if a comfortable outward divergence limit exists, and values ranging from none to 1^o to 3^o have been proposed [Pas95, YS90, CES88, Stu34, WDK95].

These asymmetries and excessive disparities can result in phoria on the eye muscles which can lead to eyestrain. Phoria is any of the tendencies of the axes of vision to deviate from the normal when binocular fusion of the retinal images is prevented, for example in the absence of adequate fusion stimulus [LIH07, UH08].

3.4.1 The Horopter and Panum's Fusion Area

In vision psychology, there are certain regions within our field of view that are defined to bring clarity to the study of binocular vision. The first is Donders' line which represents the correlation and coupling of vergence and accommodation that we have explored previously. Explicitly, it is defined as the relationship between the orientation and rotational movements of the eye for vergence and the focal point using the accommodation function [YEM04]. It can be visualized as a line in the space of focal distance as a function of convergence distance.



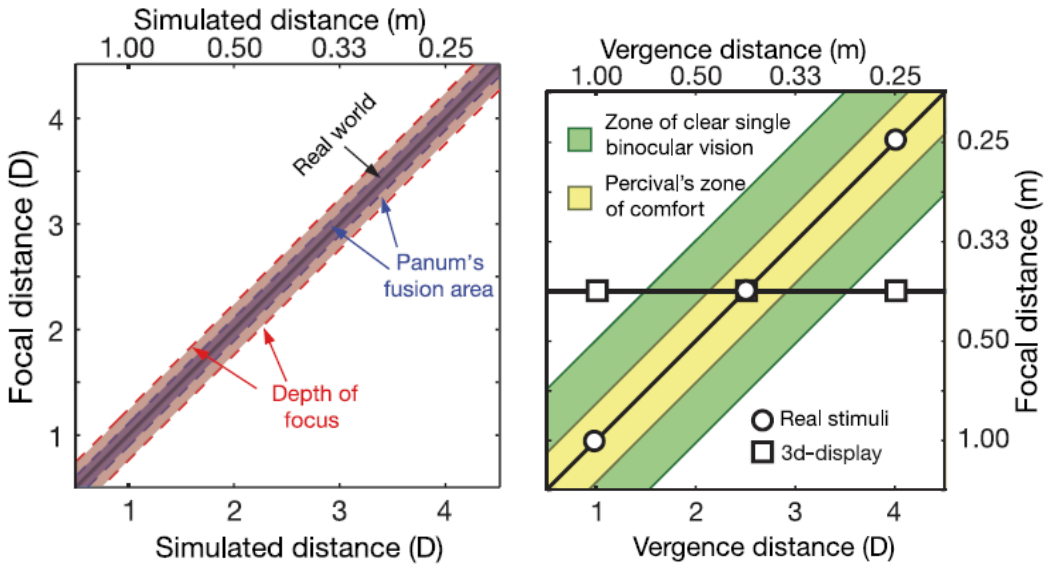
(a) Geometric (b) Empirical

Figure 3.14: The geometric and empirical definitions of the horopter. [Ogl50]

Under normal viewing conditions, this relationship is one-to-one and linear in diopter units, as indicated by the black line in Figure 3.15.

Additionally, we can explore correspondences in the spatial environment surrounding the viewer. Given a particular fixation point, we can identify all of the other points in space whose correspondences in the two retinal images are the same, or in other words where their binocular disparity is zero and fusion is exact. This set of points form a surface, or loci, known as the horopter. This loci has many special properties that are relevant to the study of binocular vision and comfort. Beyond being the surface in space upon which there is zero binocular disparity, the horopter is also the area around which precision of depth estimates is the highest. This is due to the fact that matching solutions during the processing of binocular disparities are biased towards it, and the region of single vision straddles it.

There are two ways to define the horopter; geometrically, as in Figure 3.14(a), and empirically, as in Figure 3.14(b). The shape of the loci is dependent upon which definition is used. The geometric horopter is defined by the points with



(a) Panum's Fusional Area (b) Percival's Zone of Comfort

Figure 3.15: Panum's fusional area and Percival's zone of comfort in vergence-accommodation space. Figure (a) shows Donder's line during natural viewing and Panum's fusional area which is the area around Donder's line where fusion and single vision are still possible. Figure (b) shows Percival's zone of comfort, which is defined as the middle third of the zone of clear and single binocular vision. Notice how real stimuli stay within Percival's zone while 3D-displays traditionally provide vergence-accommodation stimuli that do not. [HGAB08]

the same coordinates in the two retinas, defined mathematically. The empirical horopter is located by experimental measurement. The two surfaces are not coincident with each other.

The geometric horopter contains the Vieth-Muller circle, or the circle containing the fixation point and nodal points of eyes. Vertically, the horopter curves away from the viewer above the fixation point, and curves towards the viewer below the fixation point [SHF+08]. As the distance to the fixation point is increased, the radius of the circle increases, and the horopter becomes more plane-like as the viewing conditions become more parallel.

The empirical definition has been found to contain uncrossed horizontal dis-

parities to the left and right of fixation, thus the horizontal empirical horopter is less concave than the Vieth-Muller circle in the geometric definition. There are also two observed distortions, the Hering-Hillebrand deviation near the horizontal meridian and the Helmholtz shear near the vertical meridian, which imply that the empirical corresponding points are optimized for viewing a ground plane by a standing observer. [SHF+08] Given a particular vergence and accommodative fixation distance, there is a region surrounding the horopter in the focal direction where a certain amount of retinal blur is tolerated but clear vision is still maintained. This is known as depth of field (DOF). Similarly, there is a region surrounding Donders' line in the vergence direction where a certain amount of retinal disparity is tolerated but the fusion and single vision of the two images is still possible. This is known as Panum's fusion area. These ranges are approximately 0.3 diopters and 30 arcminutes, respectively, and are depicted in Figure 3.15(a). Within the intersection of these ranges, clear and single binocular vision is possible, and certain levels of mismatch between accommodation and vergence is tolerated [LIH07, SW83].

A subset of this region is Percival's visual comfort area, and is defined as the middle third of the zone of clear and single binocular vision (green region in Figure 3.15(b)). If Donders' line lies within this zone, there should be no visual discomfort. Under normal real-world viewing conditions, this condition is always met.

Discomfort within this area is still possible but due to alternate issues. Excessive demand on the accommodation-vergence linkage can contribute to visual discomfort within Percival's visual comfort area. This demand includes, but is not limited to, fast tracking depth motion, 3D artifacts from insufficient

depth information yielding spatial and temporal inconsistencies, conflicts between depth cues and geometrical distortions, and unnatural amounts of blur from cross-talk, conversion, or artificially induced DOF. [LIH07]

Alternately, Donders' line may lie outside of Percival's visual comfort area (the horizontal line in Figure 3.15(b)), in which case such viewing configurations may induce visual discomfort or fatigue as a result of the mismatch between the traditionally coupled vergence and accommodation. This decoupling of vergence and accommodation is known as the vergence-accommodation conflict.

3.4.2 Vergence-Accommodation Conflict

Mismatches between focal and vergence distances are significantly different from normal yoked viewing conditions, and were first explored by Charles Wheatstone in the mid 19th century [Whe52b]. Since then, others have explored the effects of the vergence-accommodation conflict and found that it may make the viewing and fusing of stereo images difficult as well as cause visual fatigue [IO97, HGAB08].

As previously discussed, natural real world viewing conditions provide for correlated focal and vergence distances. Most stereoscopic displays, however, force the user to focus on the surface of the screen while converging across a wide spatial range both in front of and behind the screen, creating a conflict of traditionally coupled depth cues, as in Figure 3.16. As a result, accommodation and retinal blur provide incorrect depth information and reduce the ability to fuse retinal disparities, as the eyes are fixated on the traditionally 2D projection plane and not on the implied virtual stereo fixation point.

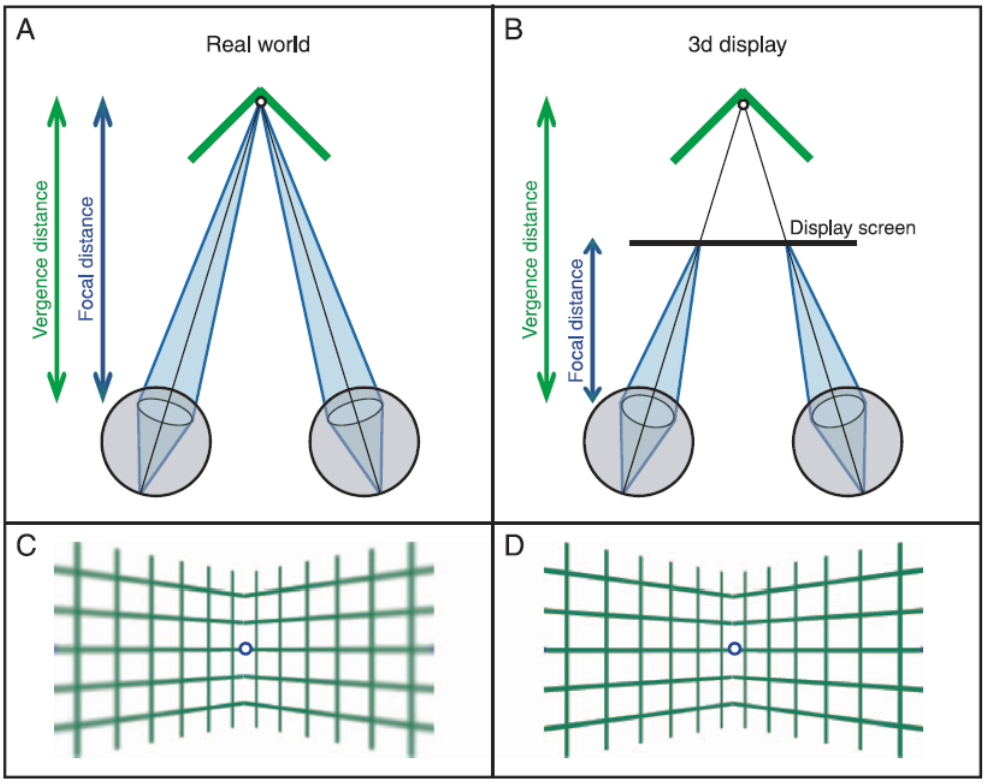


Figure 3.16: Vergence-accommodation conflict. Under normal, real world viewing conditions, vergence and accommodation are traditionally linked and fixate on points at the same depth (A), (C). However, most 3D-displays force the viewer to accommodate to the screen while verging across a wide spatial range both in front of or behind the screen (B), (D). This mismatch creates a conflict between the traditionally coupled depth cues. [HGAB08]

A study by Inoue and Ohzu [IO97] found that correct focus cues can reduce the time to identify a stereo stimulus, increase stereoacuity, reduce perceived distortions in depth, and reduce viewer fatigue and discomfort. They also identified a major difference in perception when viewing identical stimuli, once as a real world stimulus and once on a stereo screen with orthostereoscopic conditions. They found that the depth in the computer display appeared flattened compared to the real world stimulus, implicating a slight amount of distortion when viewing stereo content, even under orthostereoscopic conditions.

The results of this study make logical sense when taken in context with what

is known about the coupling and feedback mechanisms between vergence and accommodation as previously discussed. The double feedback and crosslink mechanisms help to speed up the vergence and accommodation process. When a mismatch is induced while fusing and focusing on 3D display stimuli, the viewer must counteract normal accommodation-vergence coupling, possibly taking longer and causing visual discomfort and fatigue [Uka06, WMW02].

3.5 Stereo Parameter Prediction, Control, and Automation

Finally, we review previous attempts at constructing systems to calculate ideal stereo parameters or to predict potential distortions, as well as currently employed methodologies regarding determining stereo parameters.

3.5.1 Puppet Theater and Cardboard Cutout Effects

The first, by Masaoka et al. [MHE+06], proposed a system that calculates spatial distortion based on shooting, display, and viewing conditions with an emphasis on the puppet-theater and cardboard effects. They also show how the proposed system could be useful in predicting ratings of naturalness and quality of depth in stereoscopic images.

The puppet-theater effect is defined as an unnatural miniaturization effect that results from the reproduced space depth of background objects being emphasized more than the object of focus. In the reproduced image space, the angular retinal size of background objects does not become proportionately smaller, thus miniaturizing objects of interest. One of the authors, Yamanoue

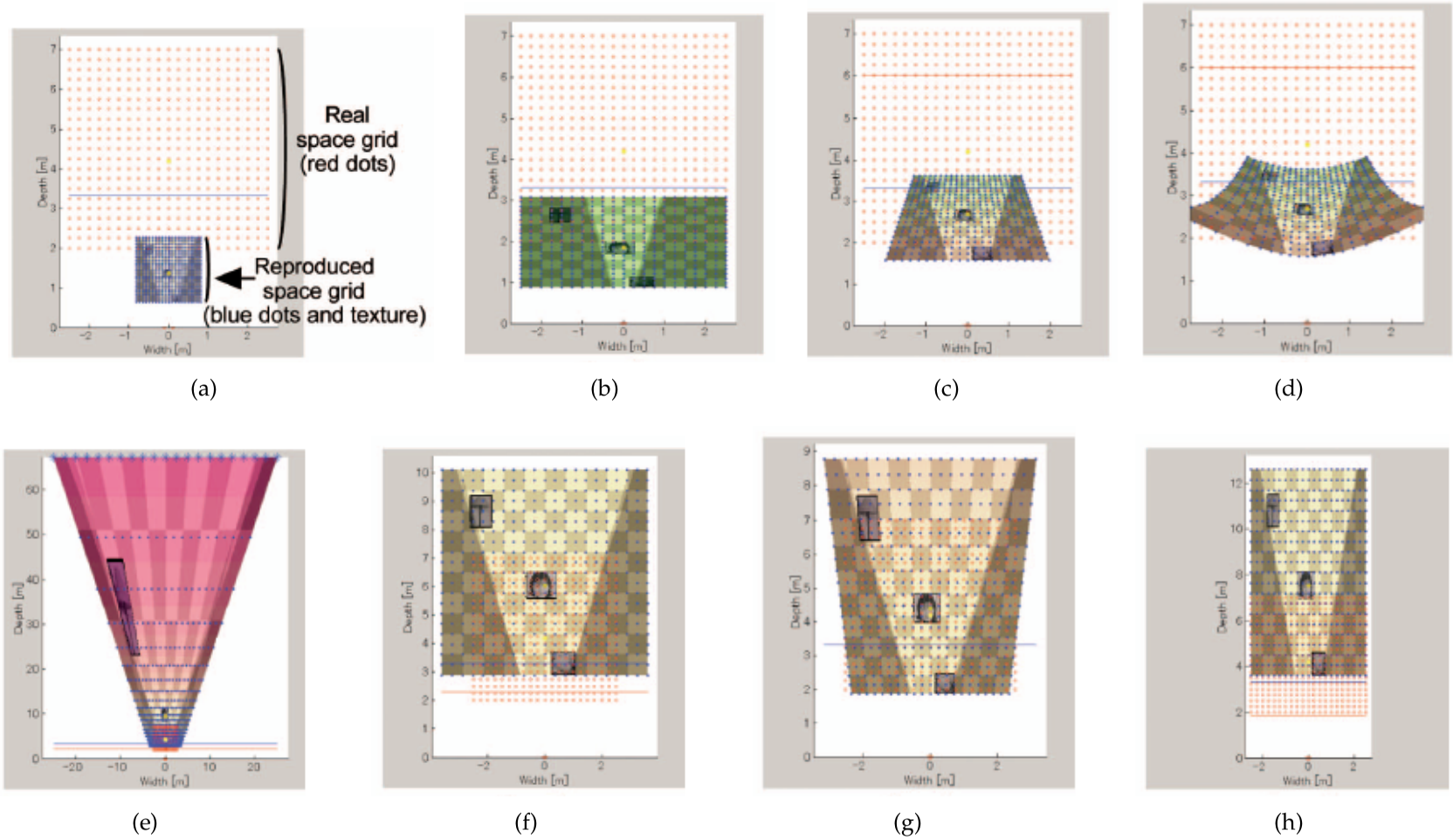


Figure 3.17: The results of the study by Masaoka into the computed anticipation of spatial distortions, particularly as they relate to puppet theater (magenta regions) and cardboarding (green regions) distortions. Seven theoretical cases are shown, computing the precept for a large camera separation (a), narrow field of view (b), long convergence distance in parallel (c) and converged (d) configurations, short convergence distance (e), short convergence distance and small camera separation (f), small eye separation (g), and narrow display viewing angle (h). [MHE+06]

[Yam97], defined a puppet-theater effect parameter, E_p, as the ratio of lateral reproduction magnifications between the object of interest and the background elements. $E_p < 1$ means that an object of interest appears smaller.

The cardboard effect is the result of reproduced space depths being shortened, flattening perceived objects. An example of this can be seen in Figure 2.9(a) where we generalized a shortening of the camera separation to a flattening of the perceived depth. Again, Yamanoue also defined a cardboard effect parameter, E_c, as the ratio of depthwise to lateral reproduction magnifications. Yamanoue determined that an object appears flat when $E_c < 0.75$.

They point out that under orthostereoscopic conditions the puppet-theater and cardboard effects do not occur geometrically, but physically and physiologically these conditions may not be ideal, creating excessive parallax and inducing headaches. They claim their system predicts the extent of excessive parallax distribution and excessive binocular parallax as well as the puppet-theater and cardboard effects.

Masaoka et al. start by suggesting that the amount of parallax in stereoscopic images should be limited to within a defined comfortable range. They define a comfort range for a 90 inch TV viewed from 3.3 meters to be from 30mm of parallax in front of the screen to 65mm of parallax behind for the average eye width, and that parallax distribution should not exceed 0.3 diopters. [NYHO03]

Their system continuously records the shooting parameters as images are captured. Analysis of the captured images gives parallax values, and the real depth is calculated. The user must then specify the maximum and minimum desired depth values, as well as a depth for the objects of interest, which they

specify as the focus plane. The system then provides visual feedback for when the current shooting conditions imply excessive cardboarding or puppet-theater effects.

Their simulation results are shown in Figure 3.17, where a magenta hued texture indicates an increased perception of the puppet theater effect, and a green hued texture indicates an increased perception of the cardboarding effect. They show the results of seven theoretical cases; large camera separation, narrow field of view, long convergence distance, short convergence distance to minimize cardboarding, short convergence distance to minimize excessive parallax, small eye separation, and narrow display viewing angle.

Masaoka et al. then discuss relating their findings to previous work done on describing the naturalness and quality of depth in stereoscopic images. They believe that naturalness will go down as the effects become more pronounced, as would quality of depth except that viewers might prefer stereo images whose depth is enhanced, even if it appears unnatural. To test this, they replicated the conditions used in IJsselsteijn's experiments [IDRV02]. They found a good overall degree of correlation between spatial distortions and perceived naturalness and quality of depth, leading the writers to propose that the spatial distortions measured by their system are those that affect observers' judgments of naturalness and quality of depth.

3.5.2 Bounding Parallax

A separate system was proposed by Robert Akka [Akk92]. He introduces a method of calculating initial stereoscopic settings and ongoing adjustments for

software designers dealing with stereoscopic computer graphics. Part of the issue he identifies is that typically stereo display parameters are interactively user controlled, but often an unviewable starting image and too-frequently needed adjustments prevent manual control from being a viable solution. Akka says that software using his approach should yield stereoscopic scenes that are "comfortable and pleasing to look at" with minimum user adjustment.

Akka defines high-quality stereoscopic computer graphic images. Intuitively, he says, this would imply being geometrically accurate, by estimating the positions of the user's eyes for example; however in many situations where the scale of the scene does not match the scale of the viewer, this fails. He also suggests that algorithms based on object rotations should not be used, as they introduce vertical distortions.

Akka claims that, regardless of the user's viewing position, stereo images tend to look best if a wider field-of-view is used, and defines "wide" as approximately 50 degrees, from Lenny Lipton's The CrystalEyes [LAM91]. He then arrives at a definition for high-quality stereoscopic imagery–"elements in the scene should generally have parallax values that fully span the range from 3% of display width negative parallax to 3% of display width positive parallax."

Next he defines the mathematical stereoscopic transformations using maximum negative and positive parallax values, the range of Z-coordinates where most elements in scene appear, and a "center of interest". He is then able to define a camera separation value that projects objects within the defined Z-coordinate range to within the specified parallax range. The crux of his argument, however, lies in his assumption of a 50^o field of view and a screen size of 12 inches–all of his computed values are derived from these two initial assump-

tions.

Akka then discusses methods for ongoing stereo adjustment. The most popular method he defines as setting a single set of parameters to meet the average condition. This method is good for static scenes. An obvious alternative would be to periodically recalculate the stereo parameters, which requires continuous knowledge of the positioning of the objects within the scene. The final method he defines as simply proportionally scaling the values according to changes in position. This final method assumes that the width-to-depth ratio of the scene remains constant.

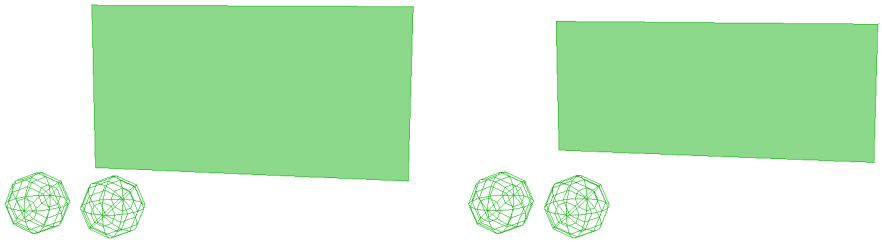
Finally, Akka suggests that the user should have a manual fine-tune control over the degree of the stereoscopic parameters and that such control should not conflict with the software adjustments and act more as minor variations from the computed baseline.

CHAPTER 4

COMPLETENESS OF THE STEREOSCOPIC TRANSFORMATION

The previous chapters have covered the perceptual and computational aspects of stereoscopic animation and cinema, and highlighted many studies and other research endeavors that have explored the understanding, creation, and control of stereoscopic content. We now present an interactive graphical stereoscopic distortion visualization tool for artists working in a 3D animation pipeline. The work takes advantage of many of the recent advancements in stereoscopic filmmaking and research to provide a comprehensive graphical interface for visualizing the resulting stereoscopic precept by the viewer in a theatrical environment.

Through augmentation of the traditionally accepted stereoscopic transformation, we accommodate for perceptual limitations related to human vision and increase the completeness of the parameter space to allow for full variability of all stereoscopic parameters and controls. We provide an interactive visualization of the stereoscopic parameters, as well as the stereoscopically transformed geometry as perceived by the audience simultaneously with the original geometry. The interactive graphic tool thus enhances the artists comprehension of the distortions inherent in the stereoscopic system's configuration. Additionally, the tool has an intuitive interface that allows for easy manipulation and control of all geometry, camera, viewer, and stereo parameters, as well as the implementation of several popular control mechanisms to allow for parameter manipulation by the artist within several commonly used stereo paradigms. Ultimately, these components provide an intuitive way for stereoscopic artists to explore the complete stereoscopic parameter space and understand the effects



(a) 1.85 Aspect Ratio (b) 2.40 Aspect Ratio

Figure 4.1: Aspect ratio variations. Projection surface and viewer shown.

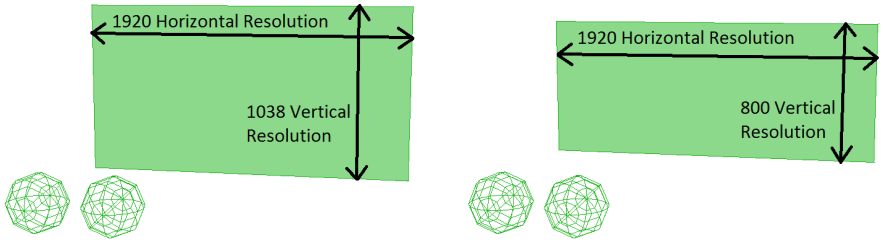
that their decisions have on the ultimate precept by the viewing audience at any stage of the filmmaking pipeline.

In an attempt to remove assumptions, allow for complete configuration, and to conform to a more perceptual model of the stereoscopic transformation, we have augmented the classic geometric stereo transform model through an expansion of the independent parameter space, implementation of indirect stereo camera control paradigms, and the addition of other industry standard stereo techniques. Ultimately, this leads to a new and complete unconstrained stereoscopic transformation model with usability enhancements.

4.1 Expansion of the Parameter Space

4.1.1 Aspect Ratio - *A*

The aspect ratio (A) is the ratio of the width of the camera sensor to the height. It is the same as the ratio of the width of the projection surface to its height. In conjunction with the width of the camera sensor or projection plane it uniquely defines the shape and dimensions of the projected image. This is important as the shape of the camera sensor and projection plane determine what objects are



(a) 1.85 Aspect Ratio (b) 2.40 Aspect Ratio

Figure 4.2: Horizontal sensor resolution at varying aspect ratios.

in the field of view of the camera, which is then used for the framing of a shot by camera operators and artists as well as for computational culling algorithms which discard geometry outside of the field of view of the camera.

Most modern movies are presented in one of two aspect ratios, as determined by the creative lead of the film. The most common format has an aspect ratio of 1.85. It is about the shape of modern widescreen televisions, and is colloquially referred to as "flat". The second most common format, colloquially referred to as "scope", has an aspect ratio of 2.40 and is significantly wider. Our visualization tool allows for variation of the aspect ratio of the camera system to arbitrary user specified values, and can also be snapped to either of the two current industry standards of 1.85 and 2.40, as seen in Figure 4.1.

4.1.2 Horizontal Sensor Resolution - R_c

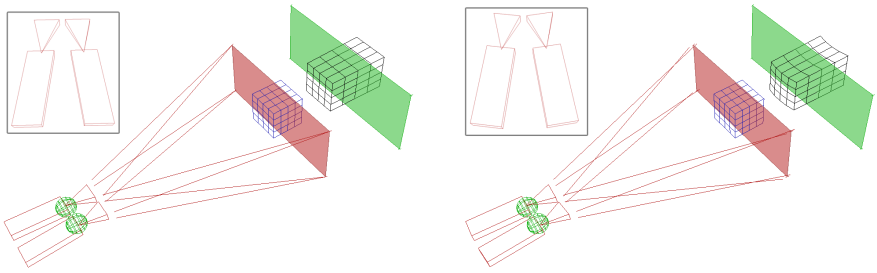
The horizontal sensor resolution (R_c) is an important parameter for the complete definition of a stereoscopic camera. In conjunction with the camera sensor width, screen width, and the aspect ratio, it relates pixels to distance and fully defines the image sensor format, an intrinsic parameter of a camera system, and the projection surface dimensions, as seen in Figure 4.2.

Of significant importance to stereoscopic computations, it allows for a common grammar amongst stereo artists to refer to parallax in pixel units. Since the film industry has currently settled on a standardized high definition horizontal image resolution of 1920 pixels for digital content, a unified reference point has been established so that artists can refer to specific parallax values in pixels and those values will retain their meaning across artists and productions.

Note however that this grammar is reliant upon the horizontal sensor resolution being constant. In today's quickly advancing technological world, digital cinema standards are marching forwards and soon a horizontal sensor resolution of 4096 pixels will take over from 1920 as the new universal horizontal resolution standard. At that point, negative 10 pixels of parallax will no longer have the same meaning in terms of percentage of screen width and stereo precept, and stereoscopic artists will have to relearn their stereo dictionary. Our visualization tool allows for variation of the horizontal sensor resolution to arbitrary user specified values, and defaults to the current industry standard of 1920 pixels.

4.1.3 Convergence Method

As discussed previously, traditional stereoscopic camera configurations are specified as either parallel or converged (Figure 4.3). The inherent distortions between the two configurations has also been presented. Our system allows for the specification of the convergence style as either parallel or converged. The parallel method locks the convergence angle to zero, while the converged method locks the sensor offset to zero. This allows for variation of the con-

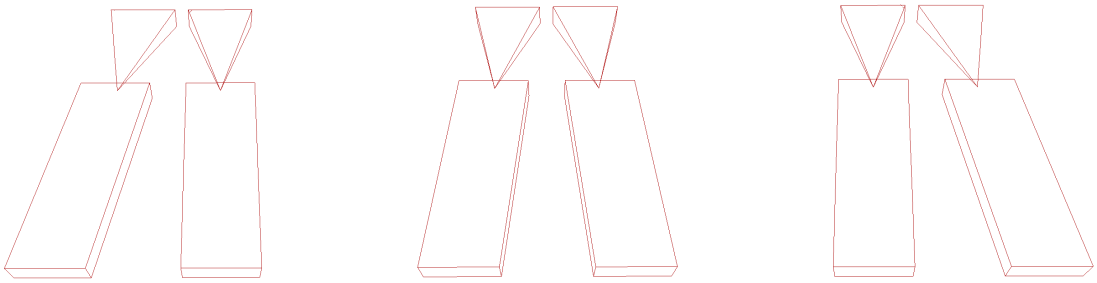


(a) Parallel (b) Converged

Figure 4.3: Convergence methods and their inherent distortions. Notice how depth plane curvature inherent to the converged method is manifest in the curving of the cube.

vergence distance by the artist while maintaining internal consistency of the stereoscopic parameter space. For example, if the system is in parallel mode, changing the convergence distance updates the sensor offset only, while a system in converged mode would update the convergence angle only. The system can be rapidly switched between the two modes, allowing the artist to better visualize the distortions inherent between the two stereo methodologies, all other things equal.

Additionally, our system has a third mode which allows for the utilization of both parameters simultaneously. While this is not currently a standard practice, it does have many theoretical and artistic applications. By removing the restriction of forcing the artist to choose between the two methodologies, it allows them to fully explore the parameter space, understand the differences between the two, and explore potential future artistic uses for combinations of the two modes where the styles could be blended in an effort to minimize or emphasize the impact that one style has over the other.



(a) Right (b) Split (c) Left

Figure 4.4: Primary camera specification and parameter distribution. With a symmetrical split between the left and right camera parameters (b), the math for the stereo transformation is clearer and cleaner. In practice, however, it important to be able to specify the distribution of these parameters between the left and right cameras asymmetrically, as in (a) and (c). This allows artists to use a single camera image for a 2D presentation void of artifacts related to animated stereo parameters.

4.1.4 Primary Camera - P_l, P_r

In most stereoscopic transformation literature, it is often assumed that the camera separation (t), sensor offset (h), and convergence angle (β) are applied uniformly or split evenly between the left and right cameras [SS53, Lip82, WDK95, HB08]. With a symmetrical split, the math is clearer and cleaner, and this decision would seem to be the most obvious choice. In practice, however, it important to be able to specify the distribution of these parameters between the left and right cameras asymmetrically. The issue arises when a stereoscopic film is released for theatrical exhibition or retail distribution in a flat 2D format. In these cases, it is most practical to release one of either the left or right eye versions of the film as the 2D version instead of re-rendering or re-filming. Filmmakers noticed, however, that if the stereoscopic parameters in question were split evenly between the left and right cameras and were also animated across the duration of a shot, the single eye 2D image would appear to drift

unnaturally.

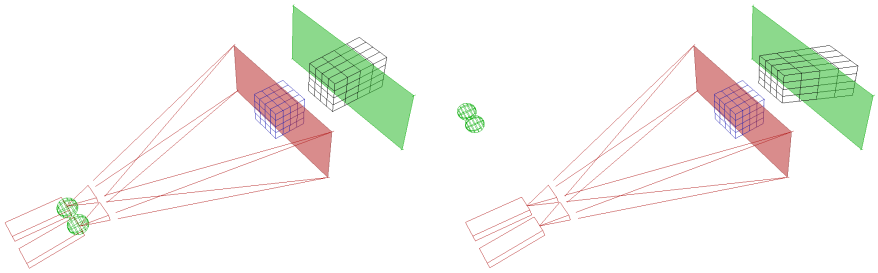
For example, in a static shot where the camera position does not change but the convergence distance is animated, an even split would have both cameras change their convergence angle or sensor offset, and the 3D precept would match the intent of a static camera position with a change in convergence distance. But if only a single eye view is observed, it would appear as if the camera is rotating or translating during the shot where the intent was for a static camera position.

This led industry artists to reweight the distribution of the camera separation, sensor offset, and convergence angle between the left and right cameras such that one camera observes the complete transformation from these parameters, leaving the other camera static and unchanging, perfect for a 2D release version of the film.

Our tool allows for arbitrary weighting of the stereoscopic camera parameters in the form of Primary Left (P_l) and Primary Right (P_r) weighting parameters. The values can be arbitrarily defined by the artist, however the sum of the two must add to 1. The tool also has "snap points", allowing the artists to default to one of the three most common configurations; an even split, left as primary, and right as primary, as outlined in Table 4.1 and demonstrated in Figure 4.4.

Primary Camera Specification		
Configuration	P_l	P_r
Split	0.5	0.5
Left Primary	1	0
Right Primary	0	1

Table 4.1: The primary camera parameter specification. Specifying a primary camera allows for the weighted distribution of stereoscopic camera parameters between the left and right cameras.



(a) Normal Viewer (b) Off-center Viewer

Figure 4.5: Viewer position variations and inherent skew distortions.

4.1.5 Viewer Position - (X_v, Y_v, Z_v)

Beyond defining just the z-distance (Z_v) of the viewer from the center of the projection screen as was previously defined in the classic stereoscopic transformation, our visualization tool allows for the off-centered positioning of the viewer with respect to the screen. By implementing the viewer position extension to the classic model as suggested by [HB08], our tool allows the artist to position the stereoscopic viewer arbitrarily within the theatrical space, as demonstrated in Figure 4.5. Z_v is still the longitudinal distance of the viewer from the screen, while X_v and Y_v specify the horizontal and vertical translations from the center of the projection plane, respectively.

While the artist using the tool ultimately has little control over where the viewer will eventually position him or herself when viewing the stereo content,

this tool does allow the artist to visualize the shear distortions inherent in off-centered viewing as explored previously [WDK95, HB08, SGC+02]. The artist can then tailor the content and stereoscopic configuration of their shot based on the maximum distortions that could be perceived by the viewer from the extremes of the intended exhibition space.

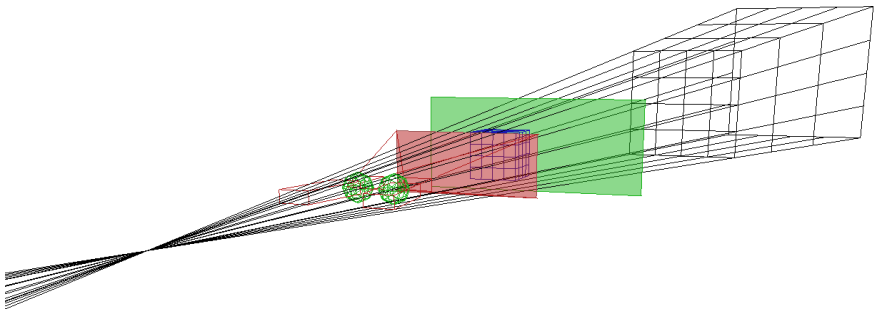
4.1.6 Gaze Direction - (X_g, Y_g, Z_g)

In addition to viewer position within the viewing space, our tool also allows for variation in the viewer orientation. The viewer orientation is specified by a gaze direction, (X_g, Y_g, Z_g), which acts similarly to a target vector for camera specifications.

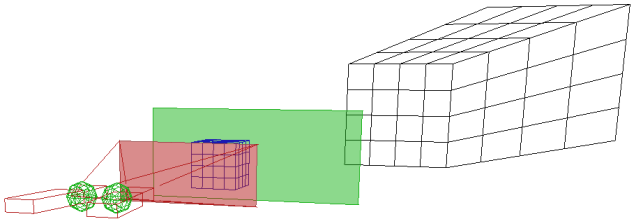
4.1.7 Failure of Skewed Rays

As previously discussed, there are particular situations where geometrically impossible configurations may arise, leading to stereoscopic transformations that are mathematically incomputable (Sections 3.2.2 and 3.2.3). These situations are limited to oblique viewer gaze directions due to the rotation and tilt of the viewer's head, and cause skewed non-intersecting view rays that lie on separate epipolar planes.

Despite the previously explored research into the perception of skewed rays, the resultant definition of the stereo precept as a slant for planar surfaces is not generalizable enough to fit arbitrary geometry for use in our tool. Accordingly, we allow for the specification of a gaze direction but compute the ray intersec-



(a) No Allowed Divergence



(b) 0.25 Degrees of Allowed Divergence

Figure 4.6: Permitting a comfortable amount of outward divergence allows for the proper visualization of objects with parallax wider than the viewer's eyes. Notice how geometrically the cube is perceived behind the viewer in the no allowed divergence instance (a), which clearly does not match the precept of an actual viewer. However, when divergence is allowed (b), the precept appears more credible.

tion and precept point by taking the average of the left and right correspondence points, a commonly used geometrically computable approximation [WDK95].

Finally, there is little an artist can do to predict, correct, or accommodate for such oblique viewing configurations since the extent of this sort of parameter variation is only restricted by the flexibility of the viewer. This makes distortions from variations in viewer gaze direction the least controllable and accountable metric.

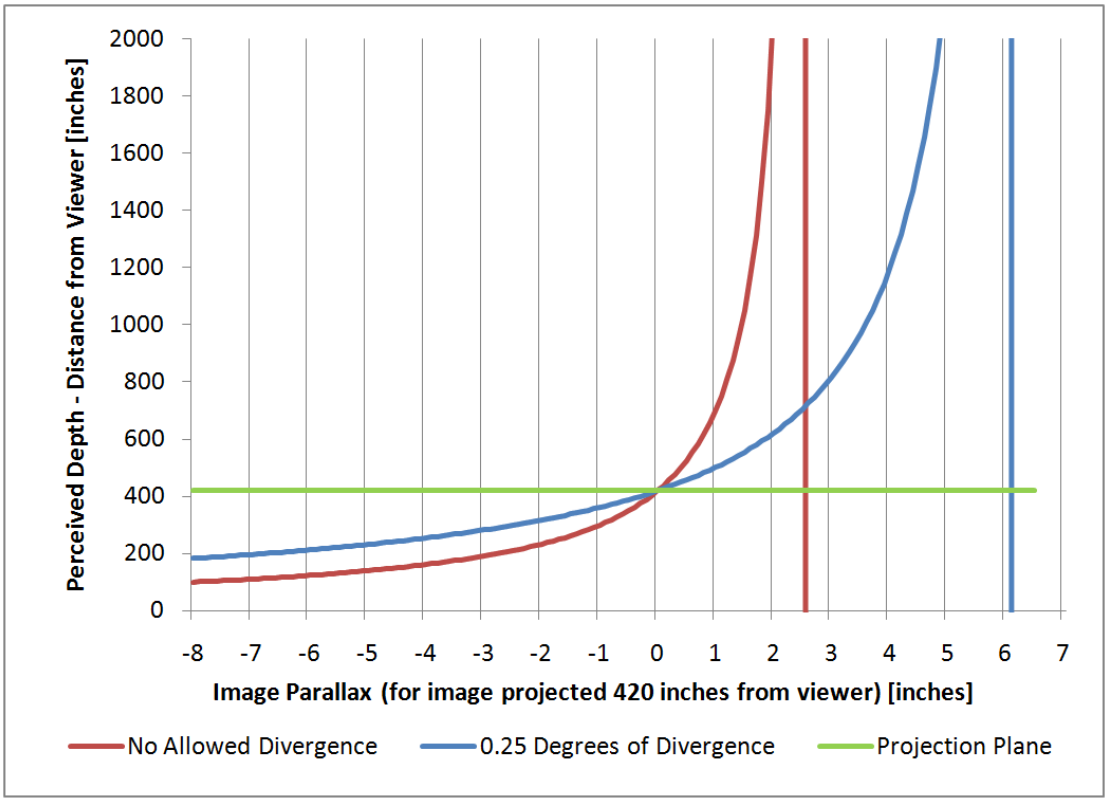


Figure 4.7: A small amount of comfortable divergence of the viewer's eyes allows for deeper stereo perceptions and greater flexibility in parallax ranges on the part of the artist. Notice how allowing a quarter of a degree of divergence per eye on the part of the audience more than doubles the useable virtual stereo real estate behind the screen.

4.1.8 Allowed Divergence - γ

One of the shortcomings of previous stereoscopic transformation models is that they do not allow for the outward divergence of the viewer's eyes. Studies have shown that a small amount of outward divergence of the eyes is not uncomfortable, with comfort ranges cited of up to three degrees of divergence per eye [Pas95, YS90, CES88, Stu34, WDK95]. Geometrically, in the classic stereoscopic transformation model, diverging viewer configurations lead to incorrect stereoscopic precept points behind the viewer, a mathematical fallacy resulting from a inverted triangulation of converging correspondence rays (Figure 4.6).

Beyond academic findings, the expectation that viewers can comfortably diverge their eyes to a small degree has become common practice in many films intended for theatrical release on large screens during moments of accentuated depth. Due to the relationship between screen size and parallax, exhibition on larger screens actually reduces the workable parallax range without allowing comfortable divergence.

For example, an image projected onto a 40 inch wide screen could present an object with a positive parallax of 2.5 inches, or about the width of the human interocular. On that sized screen, this object will be perceived as being at infinity, since the viewer's eye vergence directions are effectively parallel. In the classic model this would be the deepest object permitted. The same image projected onto a 40 foot wide screen, like those found in multiplex theaters, would present the same deep object with a parallax of positive 30 inches, significantly wider than the viewer's eyes and incomputable by the classic stereo transform model. However, viewed from a screen-average and suggested distance of 40 feet away, this configuration would only amount to an outward divergence of 1.6 deg per eye, possibly within the viewer's comfort limit.

Our visualization tool accommodates for this, allowing for the definition and variation of a maximum allowed comfortable outward divergence of the viewer's eyes. This permits artists to correctly visualize the stereoscopically transformed geometry if outward divergence is permitted by the artist.

The threshold for allowed divergence is specified by substituting the interocular distance, e, in the screen space to perceived virtual stereo space transform, Equation 4.10, with

$$e + 2Z_v \tan(\gamma) \tag{4.1}$$

Here, γ is the desired maximum angular divergence in degrees for each eye. This effectively shifts the perceived infinity point beyond geometric infinity and allows for the visualization of positive parallax values greater than the width of the viewer's eyes, as demonstrated in Figure 4.7. Notice how allowing a quarter of a degree of divergence per eye on the part of the audience more than doubles the useable virtual stereo real estate behind the screen. It is because of this flexibility of space that stereo artists are starting to anticipate a small amount of comfortable divergence on the part of the viewers.

As with all of the stereoscopic parameters, γ can be varied by the artist to allow for exploration of possible maximum divergence values as it applies to their working geometry and scene, as well as possible viewer configurations tied to varying outlets for distribution.

4.1.9 Post Render Shift - *s*

In live action stereoscopic filmmaking, converging at a distance that is not infinity using the parallel configuration is made difficult by the physical nature of the camera, which often times prevents physical manipulation of the sensor position to generate a sensor offset. Frequently, then, artists apply a horizontal image translation to the stereo images after capture or render to converge the stereo space. This is known as a post render shift, and is accomplished by horizontally translating the stereoscopic left and right images with respect to each other, as seen in Figure 4.8.

The post render shift is different in mechanics and application from horizontal image translation accomplished via a sensor offset. When utilizing a

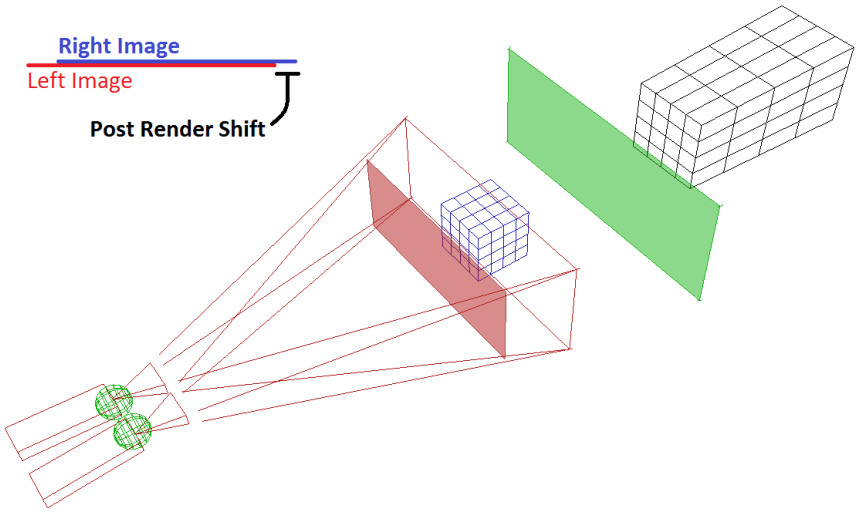
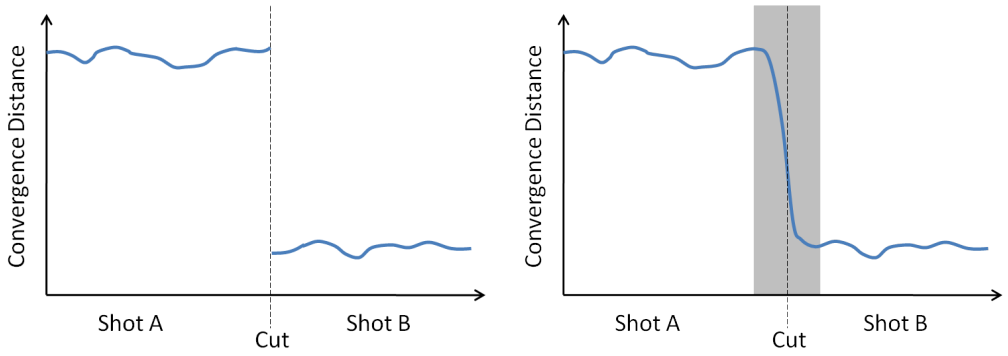


Figure 4.8: Post render shift applied to scene. Notice the loss of horizontal resolution of the image frame, as well as the shift in convergence plane with respect to the original convergence point before the applied shift.



(a) Without Blending (b) With Blending

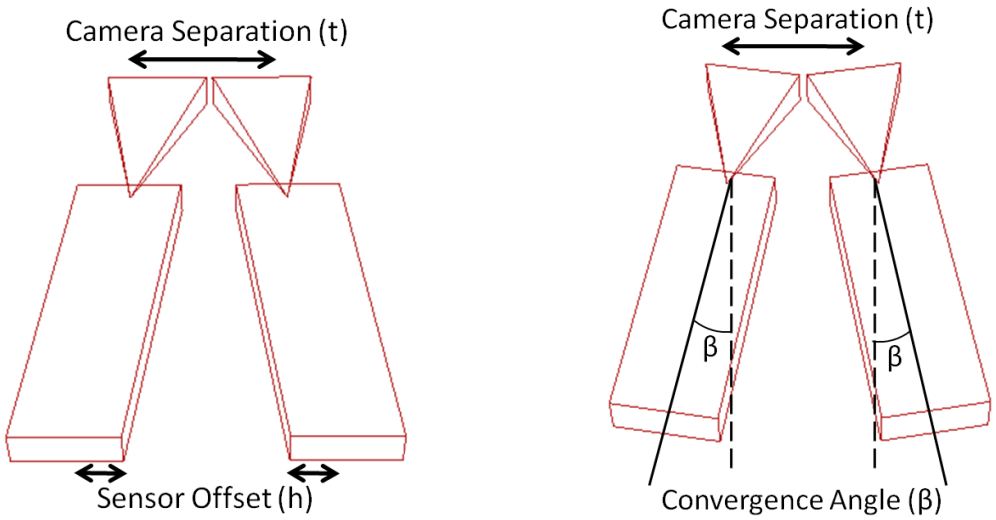
Figure 4.9: Depth blending is a technique which utilizes a post render shift. Consecutive shots across a cut may contain un-matched convergence distances, or depth jumps, which can be difficult for our visual system to instantaneously adjust to. Depth blending is the process of continuously transitioning the convergence distance across the cut via an animated post render shift. The resulting blend is easier to track with the eyes.

sensor offset, this is done in the camera at the time of image capture, and one still has use of the entire sensor area as the horizontal resolution (R_c) remains unchanged. In contrast, post render shift is utilized after image capture on the resultant rendered images, and as a result reduces the final image size by the amount of the post render shift as only the overlapping area of the left and right images is used.

Due to this limitation, the sensor offset acts as the primary convergence method for parallel configurations when it is available, and post render shifts are traditionally used for depth timing or depth blending across cuts during post processing. Depth blending is a technique that artists use when editing to smooth the discrepancies in convergence between two shots across a cut. By applying a brief transitional period between shots where the objects of interest are at different depths across the cut, artists allow viewers to track with the blend, eliminating the stress associated with instantaneously reconverging your eyes to a different depth. Figure 4.9 illustrates this concept. Our visualization tool allows for the specification and variation of the post render shift.

4.2 Alternate Control Mechanisms

There are many methodologies to describe and specify a stereoscopic configuration using one of the many grammars and control mechanisms available. Eventually, however, all of the terminology used must be translated into the two intrinsic properties of a stereoscopic system, the camera separation (t) and either the convergence angle (β) or the sensor offset (h), depending on the convergence method used.



(a) Parallel (b) Converged

Figure 4.10: Direct control of the intrinsic stereo parameters of camera separation (t) and sensor offset (h) or convergence angle (β).

For flexibility and to promote interactivity with artists from varying schools of thought and comfortable with different approaches to stereoscopic specifications, our tool provides interfaces for a number of alternate control mechanisms at various levels of abstraction.

4.2.1 Direct

Primary amongst control mechanisms would be to directly specify camera separation (t) and sensor offset (h) or convergence angle (β), as in Figure 4.10. Our tool allows for control of these parameters directly. Despite its direct correlation to the intrinsic stereo camera system, this method has the disadvantage of being ambiguous when attempting to abstract its relative significance without knowledge of the state of the rest of the parameter space. In other words, while saying that the sensors are offset 1mm might have physical meaning to the cam-

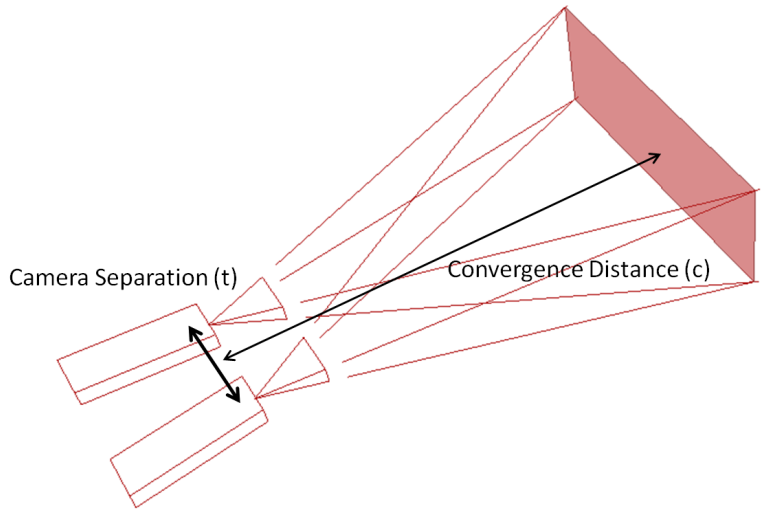


Figure 4.11: Indirect control of the intrinsic stereo parameters by manipulation of the camera separation (t) and convergence distance (c).

era system, without knowledge of the focal distance and sensor size, it has little meaning externally. It is very difficult to understand precisely how such a specification will affect the stereoscopic precept of the viewer.

4.2.2 Convergence Distance

A step up in abstraction is to indirectly specify the sensor offset (h) or convergence angle (β) by directly specifying the convergence distance (c), a dependent variable in the stereoscopic parameter space, as in Figure 4.11. The specification of the convergence distance has more relevance to the captured space external to the camera system rather than to the camera system itself. It specifies in world space units the distance to the convergence plane, an easier metric to conceptualize than the sensor offset or convergence angle, and is frequently representative of where the screen plane will appear to the viewer relative to the objects within the scene. With the ability to directly control the convergence distance,

a derivative parameter, we increase the association between the control point being manipulated and resultant affect on the stereo precept.

As a dependent variable, manipulation of the convergence distance must propagate changes upwards to maintain internal consistency of the entire stereoscopic system. Depending on the convergence method used, the sensor offset or the convergence angle must be updated with any change to the convergence distance. These update equations are outlined below.

$$h = 2f \tan\left(\arctan\left(\frac{t}{2c}\right) - \beta\right) \tag{4.2}$$

$$\beta = \arctan\frac{t}{2c} - \arctan\frac{h}{2f} \tag{4.3}$$

Note that with this method, the camera separation is still directly controlled.

4.2.3 Bounded Parallax

It is often easier to conceptualize the stereoscopic space by describing the amount of parallax, or perceived depth, at a particular distance from the stereo camera, or physical depth. As previously explored, parallax can be referred to in pixel units or as a percentage of the screen width. If parallax is defined in pixels, then parallax as a function of distance is

$$parallax = \left(\frac{fR_c}{W_c}\right)\left(\frac{t}{c} - \frac{t}{distance}\right) \tag{4.4}$$

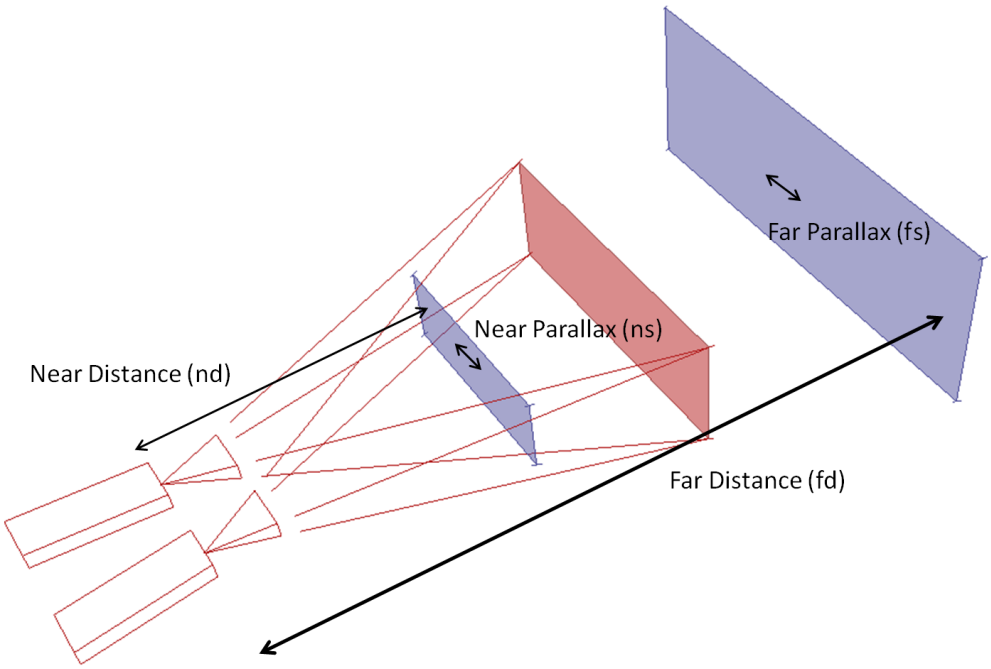


Figure 4.12: Indirect control of the intrinsic stereo parameters by manipulation of bounded parallax values (ns, fs) at specified near (nd) and far distances (fd).

and distance as a function of parallax is

$$distance = c \cdot \frac{tfR_c}{tfR_c - parallax \cdot c \cdot W_c} \tag{4.5}$$

To define parallax as a percentage of screen width, the parallax terms in the equations above are divided by the horizontal image resolution, R_c.

A third and more abstracted methodology for controlling the stereoscopic camera system follows directly from the definition of parallax in pixels or screen width percentages as a function of the object's distance from the camera and vice versa. Both of the intrinsic stereo camera parameters can be indirectly specified by defining a minimum and maximum parallax (ns, fs) for corresponding near and far object distances from the camera (nd, fd), as in Figure 4.12.

Because this abstraction is easy to conceptualize, this paradigm is frequently used as it directly correlates objects in the scene to their desired perceived depth. Additionally, when the near and far distances are defined at the nearest and furthest objects in the field of view respectively, then this methodology applies comprehensible bounds on the total depth perceived in the scene–the artist directly manipulates the nearest and furthest distance any object will be perceived by the viewer via parallax, a very relevant and understandable metric.

By combining and rearranging the equations for parallax as a function of depth (Equations 4.4 and 4.5) with the equation for the convergence distance (Equation 3.1) we arrive at the update equations for the camera separation (t) and convergence distance (c) as a function of the near and far parallax and distances (ns, nd, fs, fd). These equations maintain the internal consistency of the stereoscopic system required for any change to the parameter space.

$$t = \frac{nd \cdot fs \cdot fd - fd \cdot ns \cdot nd}{fd - nd} \cdot \frac{W_c}{f \cdot R_c} \tag{4.6}$$

$$c = \frac{nd \cdot fs \cdot fd - fd \cdot ns \cdot nd}{fd \cdot fs - nd \cdot ns} \tag{4.7}$$

Our visualization tool allows for specification and manipulation of the near and far parallax and distances.

4.3 Additional Considerations

4.3.1 Floating Windows - ($FW_{ul}, FW_{ur}, FW_{lr}, FW_{ll}$)

One of the most noticeable infringements of stereoscopic depth perception is known as a window violation. This occurs when the precept of an object that crosses out of the left or right side of frame is perceived as being in front of the projection plane, as is shown in Figure 5.9(b). This configuration violates the occlusion revelation depth perception cue, wherein one object (the edge of the projection plane, or window) occludes another (the violating object) and the former object is perceived to be nearer than the latter. However, when a window violation occurs, the perceived depth order is reversed and the stereo precept cue conflicts with the occlusion revelation cue.

As a result, the film industry has adopted a technique to alter the perceived position of the projection plane window known as floating windows. For each corner of the projection plane (upper left, upper right, lower right, and lower left), a parallax value is specified which offsets that corner of the window from the physical projection surface, in essence floating the window as seen in Figure 4.13. This is akin to applying a post render shift to each of the window's corners individually, and is achieved by occluding the corresponding portion of the frame linearly interpolated from corner to corner.

Our tool allows for the specification of floating windows via parallax values for the four corners of the projection plane ($FW_{ul}, FW_{ur}, FW_{lr}, FW_{ll}$). Positive values push that corner of the floating window behind the physical projection screen and occlude the left side of the right eye image or the right side of the

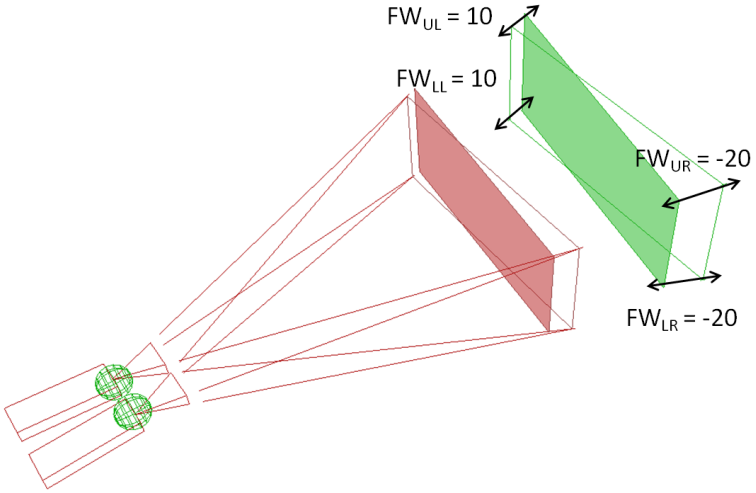
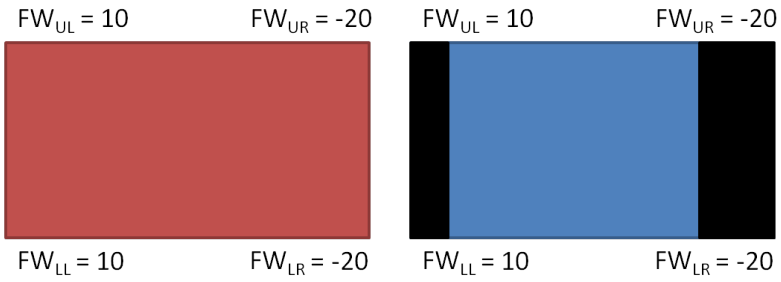


Figure 4.13: The floating window technique where artists alter the perceived position of the projection plane frame. Control of the floating window is achieved by specifying parallax values for each of the four corners of the projection plane ($FW_{ul}, FW_{ur}, FW_{lr}, FW_{ll}$).



(a) Left Eye (b) Right Eye

Figure 4.14: The appearance of the floating window is achieved by occluding the corresponding portion of the projection plane. For the case in Figure 4.13, 10 pixels are occluded on the left side of the right eye image for a positive left side floating window, and 20 pixels are occluded on the right side of the right eye for a negative right side floating window.

left eye image, while negative parallax values float the window in front of the physical projection screen and occlude the left side of the left eye image or the right side of the right eye image, as seen in Figure 4.14.

Notice that vertical window violations are not an issue. This is due to the fact that there are no horizontal parallax cues from the top and bottom edges of the projection frame, thus no cue conflicts arise. The viewer's precept in cases where a theoretical vertical window violation would occur is that of a curved projection surface bowing in to encapsulate such precepts.

4.3.2 Multirigged Cameras

One of the primary challenges of stereoscopy is determining how to fit the expansive depth of the real world into a valid stereoscopic space, restricted by display and perceptual limitations. When additionally considering the non-linear nature of the transform which has the effect of limiting the usable stereo space for an artist, it becomes understandable how many might find it difficult to settle on a single set of stereoscopic parameters that work for the entirety of a shot without sacrificing the quality of the precept. For example, when the framing of a shot dictates the presence of both deep background and close foreground elements, it becomes desirable to be able to specify two sets of parameters, one tailored for the foreground elements, and another for the background.

Live action photographers can sometimes achieve this by shooting the near and far elements separately using a green screen and separate stereo parameters and compositing the two together afterwards. However this can often be difficult as it requires precise alignment of the stereo camera rigs and multiple

takes for a single shot. Computer animated films have an advantage in that they can simply select which scene objects to render and segregate the elements into multiple stereo camera pairs with varying stereo parameters and render each set separately with minimal cost or risk of misalignment. This technique is referred to as multirigging stereo cameras, and is common practice in animated films, particularly in shots with a large range in depth of viewable objects.

Our visualization allows for multirigged cameras, and can associate separate stereoscopic parameters for different subsets of scene objects. We will revisit multirigged cameras later, as it relates to scenegraph representations as well as being aided by the visualization in terms of identifying depth incongruities.

4.4 Stereoscopic Parameter Space Listings and Groupings

Combining the previously explored classical geometric stereoscopic transformation model with the preceding expansion and augmentation parameters, we arrive at the complete listing of all parameters within the stereoscopic transformation space, both dependent and independent. These are all of the parameters which can affect the stereoscopic precept of the original geometry by the viewer. All parameters can be manipulated by the artist using the visualization tool, and all necessary updates are propagated to other parameters to maintain internal consistency as needed. For clarity, we organize the parameter space into traditional 2D camera parameters (Table 4.2), stereoscopic camera parameters (Table 4.3), viewer specific parameters (Table 4.4), and additional considerations and control parameters (Table 4.5).

Intrinsic and Extrinsic Camera Parameters	
(X_w, Y_w, Z_w)	World space coordinates of the geometry.
Camera	Camera extrinsic parameters. The position and orientation of the stereo cameras are identified by an internodal position (X_c, Y_c, Z_c), a look-at/target direction (X_t, Y_t, Z_t), and an up vector (X_u, Y_u, Z_u). Note that the world space coordinates and camera definition are frequently combined to define a point in a coordinate system relative to the camera position and orientation, or camera space (X_0, Y_0, Z_0).
f	Focal length. The focal length of stereo camera lenses.
W_c	Camera sensor width. The width of camera sensor.
R_c	Camera sensor resolution. The horizontal resolution of the camera sensor in pixels.
A	Aspect ratio. The ratio of the width of the camera sensor to its height. This is the same as the projection screen aspect ratio.

Table 4.2: Complete standard camera transformation parameters.

Stereoscopic Camera Parameters	
t	Camera separation. The distance between first nodal points of two camera lenses, aka interocular, interaxial, etc.
h	Sensor offset. The distance each camera sensor is shifted outwards from the optical axis of the lens to achieve convergence, also known as Horizontal Image Translation (HIT).
β	Convergence angle. The angle of inward rotation of the stereo cameras to achieve convergence.
Method	Traditionally either Parallel or Converged, although technically the transformation and the tool support a combination of both.
P_l, P_r	Primary camera. Defines the distribution of t, h, and *PRS* between the two cameras. Either left $(1, 0)$, right $(0, 1)$, or split $(.5, .5)$.
c	Convergence distance. The distance from the midpoint between the first nodal points of the two camera lenses and the convergence point, or point of zero parallax. Used in the indirect alternate control mechanism.

Table 4.3: Complete stereoscopic camera transformation parameters.

Viewer Parameters	
e	The distance between the centers of the viewers eyes.
(X_v, Y_v, Z_v)	The relative position of the viewer to the center of the screen, where Z_v is the distance of the viewer from the screen, and X_v and Y_v are horizontal and vertical translations from the center of the screen.
W_s	Screen width. The width of the display surface.
γ	The maximum comfortable allowed divergence of the viewers eyes.

Table 4.4: Complete viewer transformation parameters.

Additional Configuration Parameters	
s	Post render shift. Horizontal Image Translation applied after image capture, traditionally for depth timing and blending across cuts.
ns, fs	Near and far shift values. The shift values in pixels or percentage of screen width for the near and far points, used to define the Bounded Parallax alternate control mechanism.
nd, fd	Near and far distance values. The distance values for the near and far points, used to define the Bounded Parallax alternate control mechanism.
FW	Floating windows. The parallax in pixels or screen width of the four corners of the projection plane $[FW_{ul}, FW_{ur}, FW_{lr}, FW_{ll}]$used to define the floating window.

Table 4.5: Additional stereoscopic transformation parameters.

4.5 Complete Unconstrained Transformation

Having augmented the stereoscopic parameter space for completeness, perceptual considerations, and alternative control mechanisms, we update and rederive the stereoscopic transformation to reflect these additions. From object space to camera sensor space, $(X_0, Y_0, Z_0) \rightarrow (X_{cl}, Y_{cl}), (X_{cr}, Y_{cr})$:

$$\begin{aligned}
X_{cl} &= f \tan\left[\arctan\left(\frac{t \cdot P_l + X_0}{-Z_0}\right) - \beta\right] - h \cdot P_l \\
X_{cr} &= -f \tan\left[\arctan\left(\frac{t \cdot P_r - X_0}{-Z_0}\right) - \beta\right] + h \cdot P_r \\
Y_{cl} &= \frac{Y_0 \cdot f}{-Z_0 \cos\beta + (X_0 + t \cdot P_l)\sin\beta} \\
Y_{cr} &= \frac{Y_0 \cdot f}{-Z_0 \cos\beta - (X_0 - t \cdot P_r)\sin\beta}
\end{aligned} \tag{4.8}$$

From camera sensor space to screen space, $(X_{cl}, Y_{cl}), (X_{cr}, Y_{cr}) \rightarrow (X_{sl}, Y_{sl}), (X_{sr}, Y_{sr})$:

$$\begin{aligned}
X_{sl} &= X_{cl}\left(\frac{W_s}{W_c}\right) - s \cdot P_l\left(\frac{W_s}{R_s}\right) \\
X_{sr} &= X_{cr}\left(\frac{W_s}{W_c}\right) + s \cdot P_r\left(\frac{W_s}{R_s}\right) \\
Y_{sl} &= Y_{cl}\left(\frac{W_s}{W_c}\right) \\
Y_{sr} &= Y_{cr}\left(\frac{W_s}{W_c}\right)
\end{aligned} \tag{4.9}$$

And from screen space to perceived virtual stereo image in viewer space, $(X_{sl}, Y_{sl}), (X_{sr}, Y_{sr}) \rightarrow (X_i, Y_i, Z_i)$:

$$\begin{aligned}
X_i &= \frac{e\,(X_{sl} + X_{sr})\,/2 - X_v\,(X_{sr} - X_{sl})}{e - (X_{sr} - X_{sl})} \\
Y_i &= \frac{e\,(Y_{sl} + Y_{sr})\,/2 - Y_v\,(X_{sr} - X_{sl})}{e - (X_{sr} - X_{sl})} \\
Z_i &= \frac{-V_z e}{e - (X_{sr} - X_{sl})}
\end{aligned} \tag{4.10}$$

The updated complete geometric stereoscopic transformation can then be combined and defined, $(X_0, Y_0, Z_0) \rightarrow (X_i, Y_i, Z_i)$:

$$
\begin{aligned}
X_i = \Bigg[\frac{e}{2} \cdot \Bigg(\Bigg(\Bigg(f \tan\left[\arctan\left(\frac{t \cdot P_l + X_0}{-Z_0}\right) - \beta\right] - h \cdot P_l\Bigg)\left(\frac{W_s}{W_c}\right) - s \cdot P_l\left(\frac{W_s}{R_s}\right)\Bigg) \\
+ \Bigg(\Bigg(-f \tan\left[\arctan\left(\frac{t \cdot P_r - X_0}{-Z_0}\right) - \beta\right] + h \cdot P_r\Bigg)\left(\frac{W_s}{W_c}\right) + s \cdot P_r\left(\frac{W_s}{R_s}\right)\Bigg)\Bigg) \\
- X_v \cdot \Bigg(\Bigg(\Bigg(-f \tan\left[\arctan\left(\frac{t \cdot P_r - X_0}{-Z_0}\right) - \beta\right] + h \cdot P_r\Bigg)\left(\frac{W_s}{W_c}\right) + s \cdot P_r\left(\frac{W_s}{R_s}\right)\Bigg) \\
- \Bigg(\Bigg(f \tan\left[\arctan\left(\frac{t \cdot P_l + X_0}{-Z_0}\right) - \beta\right] - h \cdot P_l\Bigg)\left(\frac{W_s}{W_c}\right) - s \cdot P_l\left(\frac{W_s}{R_s}\right)\Bigg)\Bigg)\Bigg] \\
* \Bigg[e - \Bigg(\Bigg(\Bigg(-f \tan\left[\arctan\left(\frac{t \cdot P_r - X_0}{-Z_0}\right) - \beta\right] + h \cdot P_r\Bigg)\left(\frac{W_s}{W_c}\right) + s \cdot P_r\left(\frac{W_s}{R_s}\right)\Bigg) \\
- \Bigg(\Bigg(f \tan\left[\arctan\left(\frac{t \cdot P_l + X_0}{-Z_0}\right) - \beta\right] - h \cdot P_l\Bigg)\left(\frac{W_s}{W_c}\right) - s \cdot P_l\left(\frac{W_s}{R_s}\right)\Bigg)\Bigg)\Bigg]^{-1}
\end{aligned}
\tag{4.11}
$$

$$
\begin{aligned}
Y_i = \Bigg[\frac{e}{2} \cdot \Bigg(\Bigg(\frac{Y_0 \cdot f}{-Z_0 \cos\beta + (X_0 + t \cdot P_l)\sin\beta}\left(\frac{W_s}{W_c}\right)\Bigg) \\
+ \Bigg(\frac{Y_0 \cdot f}{-Z_0 \cos\beta - (X_0 - t \cdot P_r)\sin\beta}\left(\frac{W_s}{W_c}\right)\Bigg)\Bigg) \\
- Y_v \cdot \Bigg(\Bigg(\Bigg(-f \tan\left[\arctan\left(\frac{t \cdot P_r - X_0}{-Z_0}\right) - \beta\right] + h \cdot P_r\Bigg)\left(\frac{W_s}{W_c}\right) + s \cdot P_r\left(\frac{W_s}{R_s}\right)\Bigg) \\
- \Bigg(\Bigg(f \tan\left[\arctan\left(\frac{t \cdot P_l + X_0}{-Z_0}\right) - \beta\right] - h \cdot P_l\Bigg)\left(\frac{W_s}{W_c}\right) - s \cdot P_l\left(\frac{W_s}{R_s}\right)\Bigg)\Bigg)\Bigg] \\
* \Bigg[e - \Bigg(\Bigg(\Bigg(-f \tan\left[\arctan\left(\frac{t \cdot P_r - X_0}{-Z_0}\right) - \beta\right] + h \cdot P_r\Bigg)\left(\frac{W_s}{W_c}\right) + s \cdot P_r\left(\frac{W_s}{R_s}\right)\Bigg) \\
- \Bigg(\Bigg(f \tan\left[\arctan\left(\frac{t \cdot P_l + X_0}{-Z_0}\right) - \beta\right] - h \cdot P_l\Bigg)\left(\frac{W_s}{W_c}\right) - s \cdot P_l\left(\frac{W_s}{R_s}\right)\Bigg)\Bigg)\Bigg]^{-1}
\end{aligned}
\tag{4.12}
$$

$$Z_i = -V_z e \cdot \left[e - \left(\left(\left(-f \tan \left[\arctan \left(\frac{t \cdot P_r - X_0}{-Z_0} \right) - \beta \right] + h \cdot P_r \right) \left(\frac{W_s}{W_c} \right) + s \cdot P_r \left(\frac{W_s}{R_s} \right) \right) \right.\right.$$
$$\left.\left. - \left(\left(f \tan \left[\arctan \left(\frac{t \cdot P_l + X_0}{-Z_0} \right) - \beta \right] - h \cdot P_l \right) \left(\frac{W_s}{W_c} \right) - s \cdot P_l \left(\frac{W_s}{R_s} \right) \right) \right) \right]^{-1} \tag{4.13}$$

CHAPTER 5

GRAPHICAL STEREOSCOPIC DISTORTION VISUALIZATION TOOL

5.1 Visualization

In order for our complete expansion of the stereoscopic parameter space to be useful and effective as a tool, we have developed an intuitive interactive visualization for the parameters and their impact on the perception of scene geometry (Figure 1.1). This allows artists to simultaneously visualize the relevant parameter spaces and controls as well as the resulting distortions of object geometry while manipulating the parameters.

The visualization is implemented as a standalone Java application with interfaces for access to animation pipelines to import movie sequence, shot, and frame lists, object, scene, and character geometry and animations, camera transformations and other relevant stereoscopic parameters. The code is compiled with Java 1.6, and utilizes openly available Vecmath and Java OpenGL libraries for vector operations and 3D graphics respectively.

The tool provides for the identification and visualization of all of the stereoscopic parameter subgroups and relevant controls, and implements the stereo transformation model as defined by the parameter space. The transformation is implemented in such a way as to maintain internal consistency within the complete parameter space, and is computed per vertex for each of the vertices of the original geometry. Scene geometry can be selected and loaded into the tool from the animation pipeline, and the transformation is applied to the loaded geometry. Both the original geometry and transformed geometry are visual-

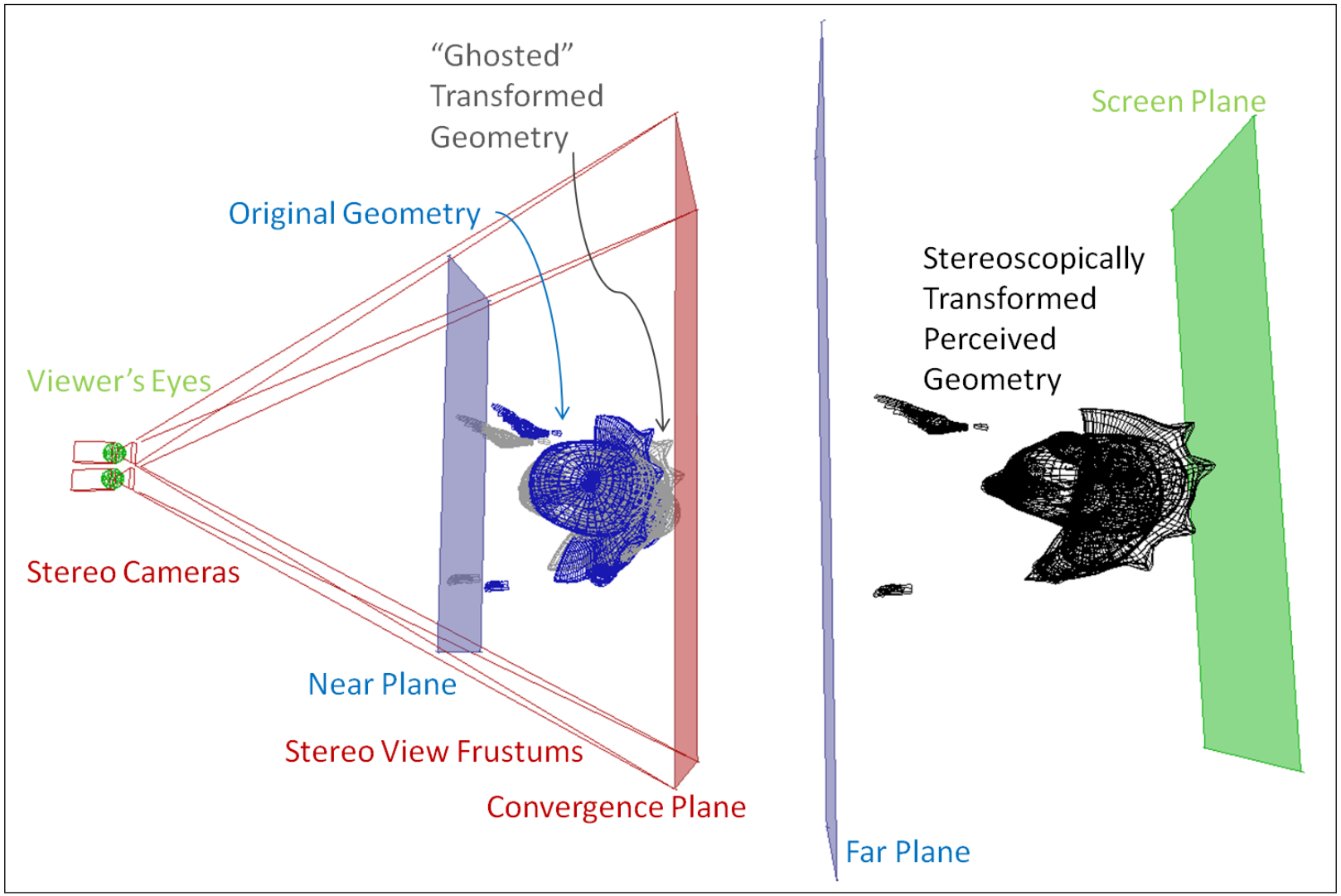
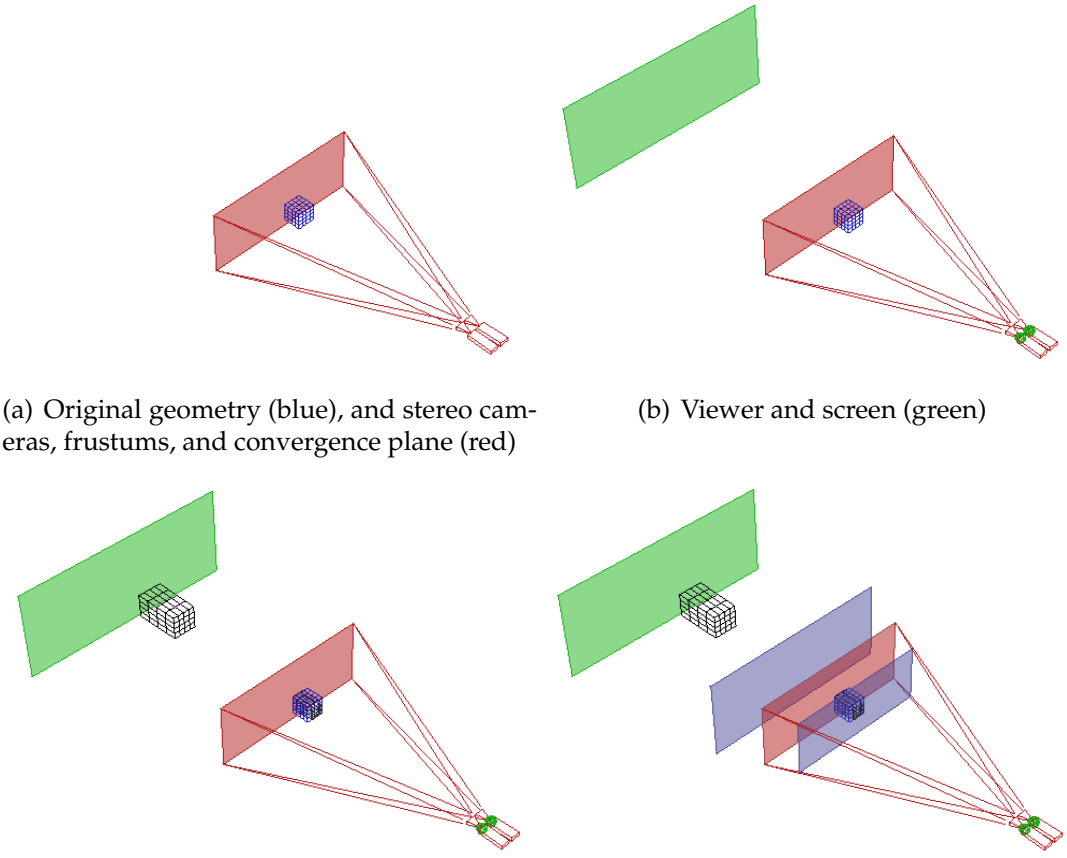


Figure 5.1: The tool uses accurate and informative visualizations for the original geometry, the stereo cameras and their frustums, the convergence plane, near and far planes for parallax specifications, the viewer, the screen, the stereoscopically transformed geometry as it is perceived by the viewer in the theater, and a scale and translationally invariant "ghosted" version of the transformed geometry overlaid on top of the original geometry for more precise interpretations of shape and volume distortions.

ized simultaneously. The viewer can subsequently navigate around the scene to view geometry and parameter visualizations, as well as manipulate the parameters to see the affect on the resultant distorted geometry. Figure 5.1 provides an overview of the tool's geometry, parameter, and distortion visualizations.

5.1.1 Stereoscopic Parameter Groupings



(a) Original geometry (blue), and stereo cameras, frustums, and convergence plane (red)

(b) Viewer and screen (green)

(c) Stereoscopically transformed geometry and "ghosted" geometry (black)

(d) Alternative control mechanics (blue)

Figure 5.2: The visualization tool color-codes geometries, parameters, and control methods into separate groups for ease in identification, distinction, comparison, and control.

For clarity and ease of identification and control, our tool groups the geome-

tries and parameter, as shown in Figure 5.2. These groupings are the same as, or an extension of, the parameter groupings as specified previously in Tables 4.2, 4.3, 4.4, and 4.5. The groupings are color-coded to help distinguish between them. Each of the parameter and geometry visualizations and groupings are described below.

Original Geometry and Stereoscopic Camera

The original geometry, as loaded into the tool by the artist (X_w, Y_w, Z_w), is visualized in blue with respect to the stereoscopic cameras $((X_c, Y_c, Z_c)$, (X_t, Y_t, Z_t), $(X_u, Y_u, Z_u))$ in red (Figure 5.2(a)). The camera icons are visualized as scale versions of the control parameters; the length is relative to the focal length (f), the width and height are relative to the sensor width (W_c) and sensor height (W_c/A) respectively, and the sensor offset (h) is represented by a scale horizontal shearing of the camera body. The camera separation (t) and convergence angle (β) are directly visualized by the translational offset and inwards rotation of the camera representations.

Additionally, the view frustums of both cameras are shown to assist in the framing of the shot. Where the two view frustums intersect is the distance at which there is zero parallax between the left and right images, which indicates the convergence distance. The manipulatable convergence distance (c) is further emphasized in the visualization by the presence of the partially transparent red convergence plane (Figure 5.2(a)). The convergence plane is a good relative reference for artists, as objects that intersect the plane will appear "at screen" in the eventual stereo precept. Objects that fall behind the convergence plane will be perceived as behind the projection plane, and objects that are in front of

the convergence plane will appear in front of the projection plane. Note that the use of a post render shift (s) is visualized directly by an offset of the convergence plane from the intersection of the view frustums, as in Figure 4.8.

Viewer

The counterpart to the image capture component of the stereoscopic system is the image presentation component, which includes the viewer and screen configuration and is visualized by the tool in green (Figure 5.2(b)). While technically the position of the viewer and the position of the camera are completely independent and most likely are not the same, it is easier and more helpful to visualize them as being at the same position. This unifies distances, directions, and orientations between the camera and viewer spaces, so that 10 units in front of the camera is the same position as 10 units in front of the viewer. This validates direct comparisons of size, scale, position, shape, orientation, etc. between the two spaces.

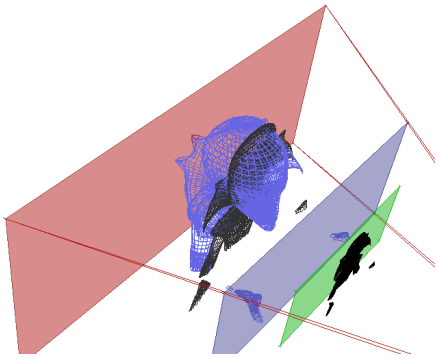
Accordingly, the viewer is positioned and visualized as being centered about the internodal position of the stereo cameras and oriented similarly as well. The interocular distance between the viewer's eyes (e) is directly visualized, as is the positioning of the screen relative to the viewer (X_v, Y_v, Z_v) and the screen's size (W_s, W_s/A). Note that while the screen width may be varied arbitrarily, the height is pre-determined based on the width of the screen and the aspect ratio, which is the same as the aspect ratio of the camera sensor. This ensures a uniform magnification from sensor to screen when the image is projected onto the projection surface.

Also, if a floating window ($FW_{ul}, FW_{ur}, FW_{lr}, FW_{ll}$) is utilized, the corresponding offset of the perceived projection plane from the physical projection screen is directly visualized, as in Figure 4.13.

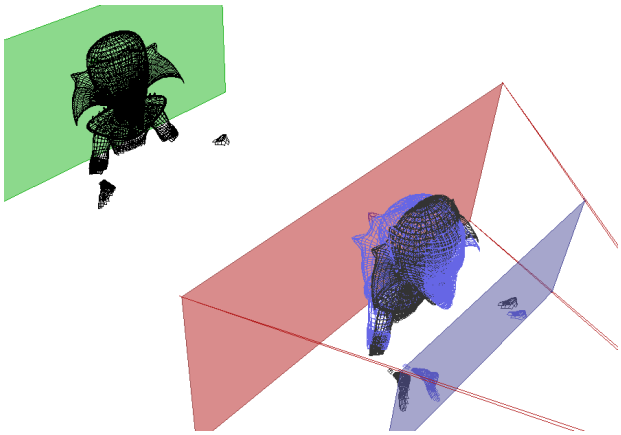
Viewer Dictionary

The viewer configuration is one of the least controllable sets of parameters, in that once a film is finalized and distributed, the artist has minimal direct control over the actual exhibition and projection of the film. Artists may have an ideal, expected, anticipated, or even agreed upon viewing configuration, but the presentation of the film is ultimately up to the theater manager or homeowner or computer user, and variations from one theater or home or computer to the next are common. However, despite the large amount of variability and minimal amount of control, it is absolutely necessary to consider the intended presentation method, as the differences between a large theatrical screen and a personal computer monitor can be dramatic. Figure 5.3 demonstrates how scene elements can be perceived as elongated or compressed, depending on the viewing configuration, despite all other parameters being equal. Accordingly, it has started to become standard practice for artists to reconverge their content for specific distribution outlets, differentiating between a theatrical release print, a DVD or Blu-ray pressing, and potentially a downloaded release for handheld mobile devices.

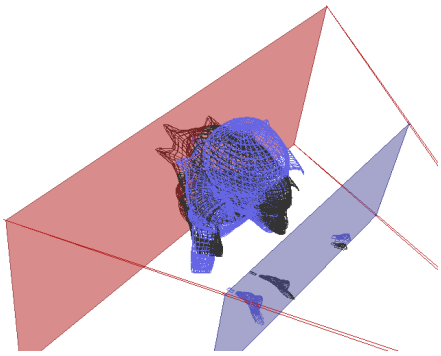
To aid artists in tailoring stereoscopic content for specific distribution outlets, our tool provides a dictionary of various preset viewing configurations. The artist can individually manipulate the viewing parameters, but can also snap the entire viewing configuration to one of a set of predefined viewer con-



(a) PC: 20″ screen viewed from 30″



(b) Home Theater: 52″ screen viewed from 100″



(c) Theatrical: 40′ screen viewed from 35′

Figure 5.3: A set of default viewer configurations allow artists to tailor their content for personal, home, or theatrical viewing. Notice how when all other parameters are held constant, the larger screen viewed at a farther distance exhibits increasing elongation in the perceived geometry (black) when compared to the original geometry (blue). Note that in (c), the screen plane is too far away from the convergence plane and geometry to be visualized simultaneously.

Default Viewer Configurations			
	Screen Width	Viewer Distance	Field of View
Personal Computer	20″	30″	37°
Home Theater	52″	100″	29°
Theatrical	480″	420″	59°

Table 5.1: Default viewer configurations based on manufacturer and retail recommendations, THX standardizations, and average viewing distances and screen sizes.

figuration dictionary entries. The entries can be categorized as generalizations of theatrical big screen exhibition, home theater HDTV viewing, and personal computer viewing on a computer monitor. The specification of the parameter values for these configurations is based on manufacturer and retail recommendations, THX standardizations, as well as average viewing distances and screen sizes, and are outlined in Table 5.1. It should be noted that as advancements in mobile gaming, communications, and computation are made, a fourth case could soon become standard for personal mobile device distribution and exhibition.

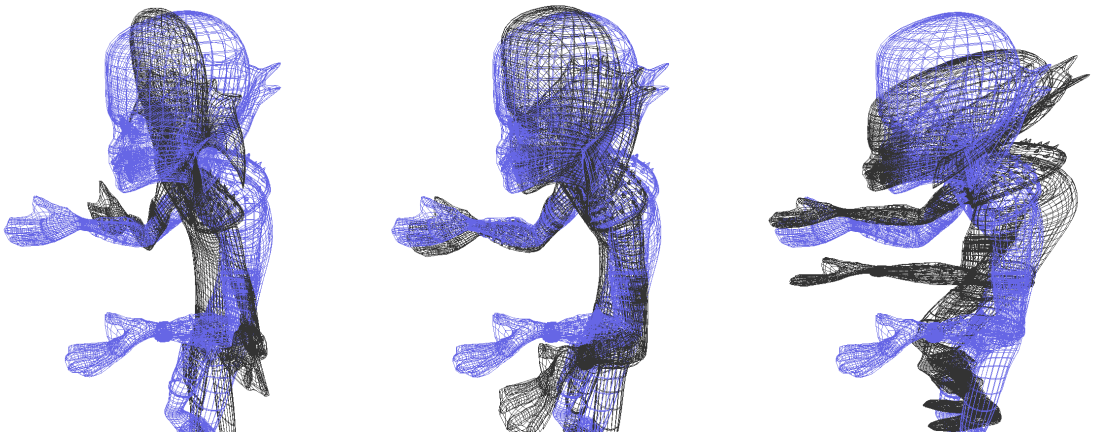
5.1.2 Stereoscopically Transformed Geometry

Finally, the geometry, as transformed by the complete stereoscopic model and defined by the specified parameters, is visualized in black (Figure 5.2(c)). This represents the three dimensional precept by the viewer in the theater of the stereo camera captured original geometry. It exhibits all of the perceptual and geometric distortions inherent in the stereoscopic parameter space as specified by the artist. Any time a parameter is modified that affects the resultant precept, the transformation is recalculated and applied to the original geometry, and the

visualization is updated.

Ultimately, the purpose of this tool is to visualize the distortions that are inherent to the stereoscopic parameters specified, and to do that, a more direct visual comparison between the original geometry and transformed geometry is required. Visualizing the transformed geometry relative to and in context of the viewer configuration is an effective way of ensuring proper positioning within the theatrical space and to verify that no non-diverging objects exist within the scene. However, beyond the relative positionings of transformed objects to the projection plane and original objects to the convergence plane, there are very few directly correlatable elements between the original camera-object space and the viewer-theater space.

As with 2D images, and perhaps because of our extensive training in watching two dimensional films, viewers dissociate the perceived size and distance of a projected object from that of the size and distance of the real object being captured. For example, when viewing a close-up shot of a person's face on the big screen, one does not assume that you are actually seeing a giant head 15 feet tall from 40 feet away. Instead, viewers dissociate their viewing experience from their physical presence in the theater and accept that they are closer to a normal sized human being. This is often referred to as a component of the "4th wall" of cinema, incorporating the suspension of disbelief in that objects are filmed in a way that does not acknowledge the presence of the camera. This assumed dissociation is used to create intimacy and emotional connections between content of the movie and the audience. One explanation for viewers' ease in accepting such discrepancies between size and distance is that familiarity of an object, such as the form of a human face, provides a stronger perceptual cue than the



(a) (b) (c)

Figure 5.4: Close-up examples of "ghosted" geometry. Through the removal of the appearance of size and position distortions, direct comparisons of variations in shape and volume distortion between the original (blue) and ghosted transformed (black) geometries can be made and are more evident. The three displayed cases correspond to the three default viewer conditions, PC (a), home theater (b), and theatrical (c).

perception of size or distance.

5.1.3 Ghosted Geometry

The willingness and ability of audiences to dissociate size and distances between the captured world and the perceived world necessitates the need for a more directly correlatable comparator between the two spaces for the visualization to be effective. This need motivated the development of what we refer to as "ghosted geometry", wherein the stereoscopically transformed geometry is visualized without the impact of size and position discrepancies. The removal of size and position distortions emphasizes the distortions in shape and volume between the original geometry and the stereoscopically transformed geometry.

To remove position and size distortions, the ghosted geometry is defined as

the stereoscopically transformed geometry visualized at the position of the original geometry and uniformly scaled to match the size of the original geometry, as in Figure 5.1 in grey. This enhances the artist's ability to make direct comparisons of orientation, shape, and volume between the two spaces. Discrepancies like shearing, flattening, depth exaggeration, and the non-linear transformation effects are more easily visualized and identified with the ghosted geometry, which accordingly highlights the types of distortions that may more directly impact a viewer's association and familiarity with a particular scene element. Figure 5.4 provides a close-up of the tool which highlights how the ghosted geometry enhances the artists ability to visualize shape and volume distortions between the original and stereoscopically transformed geometries.

Near-Affine Invariance

Position and scale are two of the three components of affine transformations, which, along with rotation, preserve collinearity and ratios of distances between spaces. Since rotational invariance is guaranteed by the orientational alignment of the camera and viewer spaces, our ghosted geometry bares significant resemblance to affine invariance. However, non-uniform scales which introduce shearing of geometry are a component of affine transformations. These are not considered invariant in our ghosted geometry since they violate shape constancy and are an important factor in examining shape distortion. Accordingly, we prescribe our ghosted geometry visualization to be near-affine invariant, accommodating for rotational, translational, and uniform scale transformations.

Since rotational distortions are inherently non-existent, only translational and uniform scale distortions need to be computed to properly display the

ghosted geometry. The translational distortion (D_t) is computed first and is the vector translation between the computed centroids of the two geometries.

$$C_0[X,Y,Z] = \frac{\sum_{vertices} [X_0, Y_0, Z_0]}{\|\text{vertices}\|} \tag{5.1}$$

$$C_i[X,Y,Z] = \frac{\sum_{vertices} [X_i, Y_i, Z_i]}{\|\text{vertices}\|} \tag{5.2}$$

$$D_t[X,Y,Z] = C_0[X,Y,Z] - C_i[X,Y,Z] \tag{5.3}$$

The scale distortion (D_s) is then computed such that the sum of the L^2 norms, or Euclidean lengths, of the offsets between the original and ghosted geometry vertices is minimized. This ensures that the scales of the two geometries are matched. D_s can be computed directly by taking the derivative of the sum of square L^2 norms with respect to D_s, setting it equal to zero, and solving. The sum of square L^2 norms is defined as

$$\begin{aligned} &\sum_{vertices} \Big((X_i - C_i[X]) \cdot D_s + C_i[X] + D_t[X] - X_0\Big)^2 \\ &+ \sum_{vertices} \Big((Y_i - C_i[Y]) \cdot D_s + C_i[Y] + D_t[Y] - Y_0\Big)^2 \\ &+ \sum_{vertices} \Big((Z_i - C_i[Z]) \cdot D_s + C_i[Z] + D_t[Z] - Z_0\Big)^2 \end{aligned} \tag{5.4}$$

Taking the derivative with respect to D_s, setting equal to zero, and solving for D_s defines the scale distortion between the camera and viewer spaces. For brevity, the individually written out sums of the X, Y, and Z coordinates are compacted into sum notation over $W = X, Y, Z$.

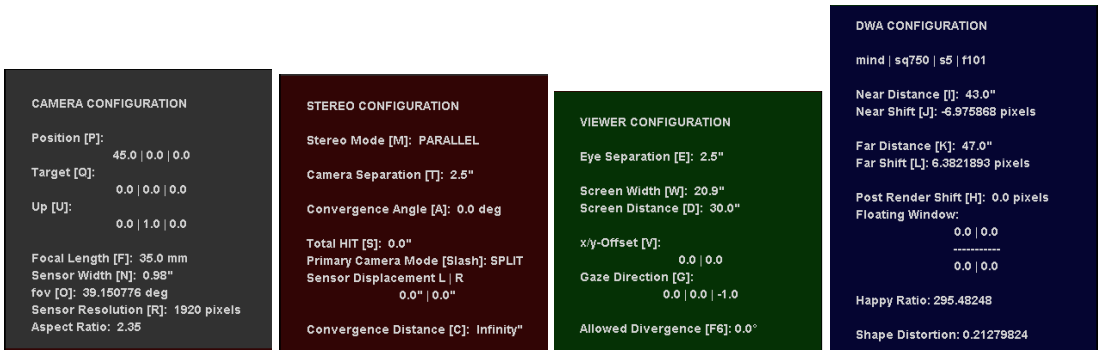
$$D_s = -\frac{\sum_{vertices}\sum_{W=X,Y,Z}\left[(W_i - C_i[W]) \cdot (C_i[W] + D_t[W] - W_0)\right]}{\sum_{vertices}\sum_{W=X,Y,Z}\left[(W_i - C_i[W])^2\right]} \tag{5.5}$$

D_t and D_s are then used to translate and scale the transformed geometry into a scale and translationally invariant, stereoscopically transformed, ghosted version. The visualization of this ghosted geometry is also the basis for a metric to quantize stereoscopic shape distortions. The definition of this metric and its applications are discussed later.

5.1.4 Alternative Control Mechanisms and Interfaces

Visualizing Bounded Parallax

To facilitate the specification of the stereoscopic parameters via alternative, indirect control mechanisms, the tool provides a visual interface for parameter definition. The near and far distances (nd,fd) used in the bounded parallax method of stereo control (Section 4.2.3) are visualized by blue planes, as in Figure 5.2(d). If the near plane is positioned at the object nearest to the camera within its field of view and the far plane is positioned at the object farthest from the camera within its field of view, then the artist-specified near and far shifts (ns,fs) define the maximum and minimum parallax that will be observed by the viewer. This gives the artist a good sense of the extent of the range of perceived depths that the viewer will be exposed to, as visualizing where the nearest and farthest objects are and will be perceived becomes trivial with the addition of the near and far planes. The combination of the conceptually simple and easily abstractable



(a) Camera Controls (b) Stereo Controls (c) Viewer Controls (d) Custom Controls

Figure 5.5: The tool's display interface enumerates the full parameter space, allowing for identification and control. The tool groups parameters into camera, stereo, viewer, and alternative configuration parameters which corresponds to the geometric visualization groupings.

bounded parallax stereo control method with the embedded visualizations of the method's control points provides for a comprehensible and intuitively usable method of stereoscopic manipulation.

Control Mechanics

In order to facilitate the complete customizable and configurable nature of the stereoscopic parameter space beyond the graphical visualizations, the tool also provides textual listings and control points for all manipulatable parameters. As indicated in Figure 5.5, the listings provide a detailed enumeration of the complete parameter space in the proper color-coded groupings. This allows for more precise control of parameter manipulation, numerical feedback of the exact parameter values, and improved nomenclature comprehension.

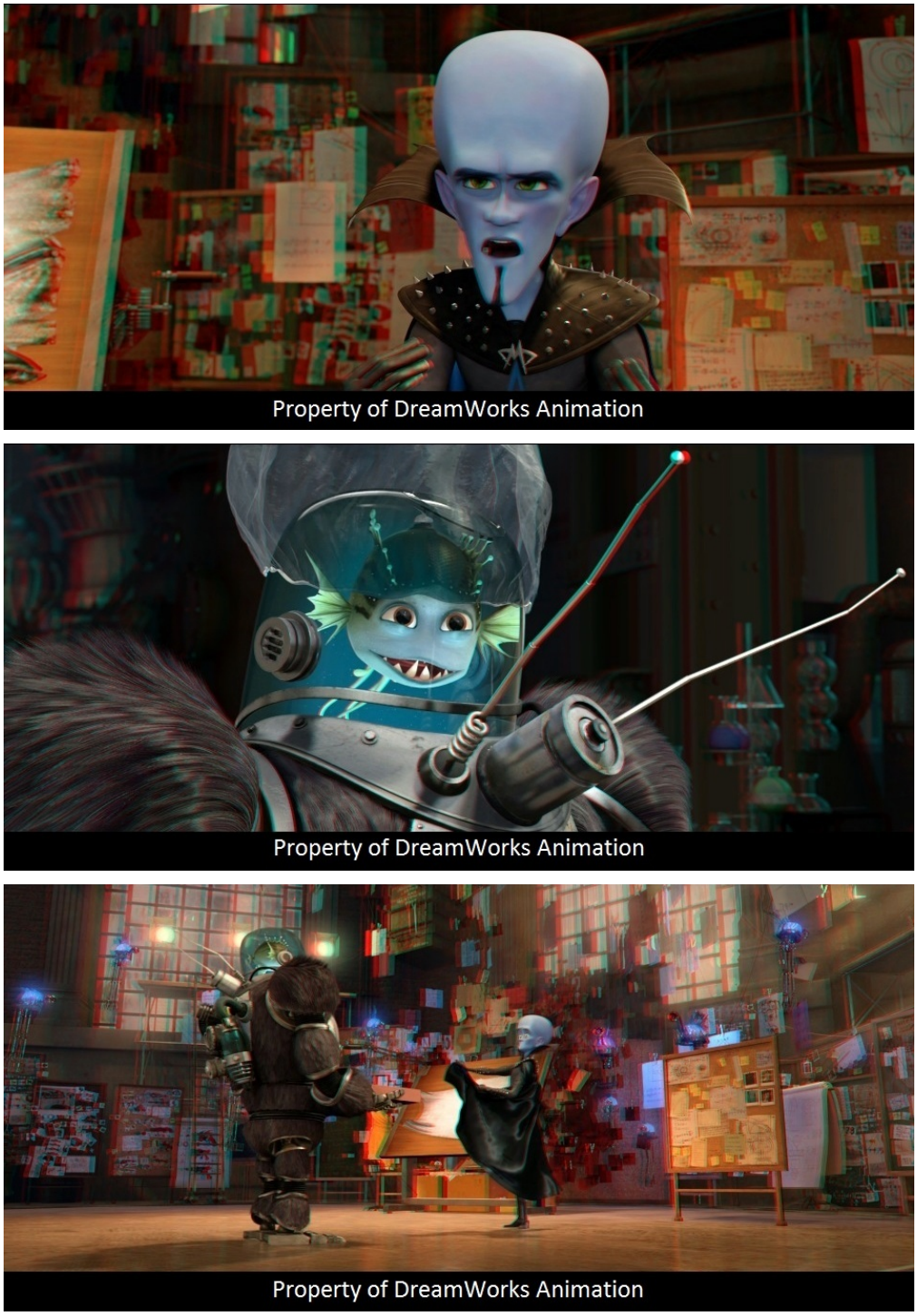


Figure 5.6: Rendered stereoscopic red/cyan anaglyph stills from the DreamWorks Animation 2010 feature film, Megamind. These images and the data that created them are used with the permission of DreamWorks Animation, and remain the property and under copyright of DreamWorks Animation. Duplication is prohibited.

Figure 5.7: Rendered stereoscopic red/cyan anaglyph stills from the DreamWorks Animation 2010 feature film, Megamind. These images and the data that created them are used with the permission of DreamWorks Animation, and remain the property and under copyright of DreamWorks Animation. Duplication is prohibited.

Pipeline Integration

Finally, our stereoscopic distortion visualization tool benefits from the ability to interface with working production animation pipelines. This allows for access to a vast library of animation content, limited only by the library the tool is interfacing with. The content library can provide anything from scene and character geometry, animation and transformation data, camera and stereo specifications, to any other data that relates to the stereoscopic parameter space and visualization as defined by our tool.

Through the support of DreamWorks Animation, we were permitted limited access to their internal data and pipeline to test the abilities of our visualization tool within a production environment. The tool provides an interface to their pipeline, and allows the artist to query and navigate show, sequence, shot, and frame lists. Once a particular frame is selected, the tool can automatically find, load, and convert geometric scene data assets, and parse in animation transform, camera, and stereo parameters.

Figures 5.6 and 5.7 show stills from the DreamWorks Animation 2010 feature film Megamind. It shows example stereoscopically rendered stills from their pipeline data in the red/cyan anaglyph format. These stills exemplify the type, variation, and complexity of data used on a regular basis by production artists, as well as the artistic output of the studio. The geometry and parameter configurations used in the rendering of these stills are used throughout this thesis and visualized in our tool, to test it on real world data and conditions. They can be used to compare the visualization tool geometric representations throughout with the final rendered stills here. For example, Figure 1.1 is our tool's visualization of the geometry present in the first rendered still of Figure 5.6. These images

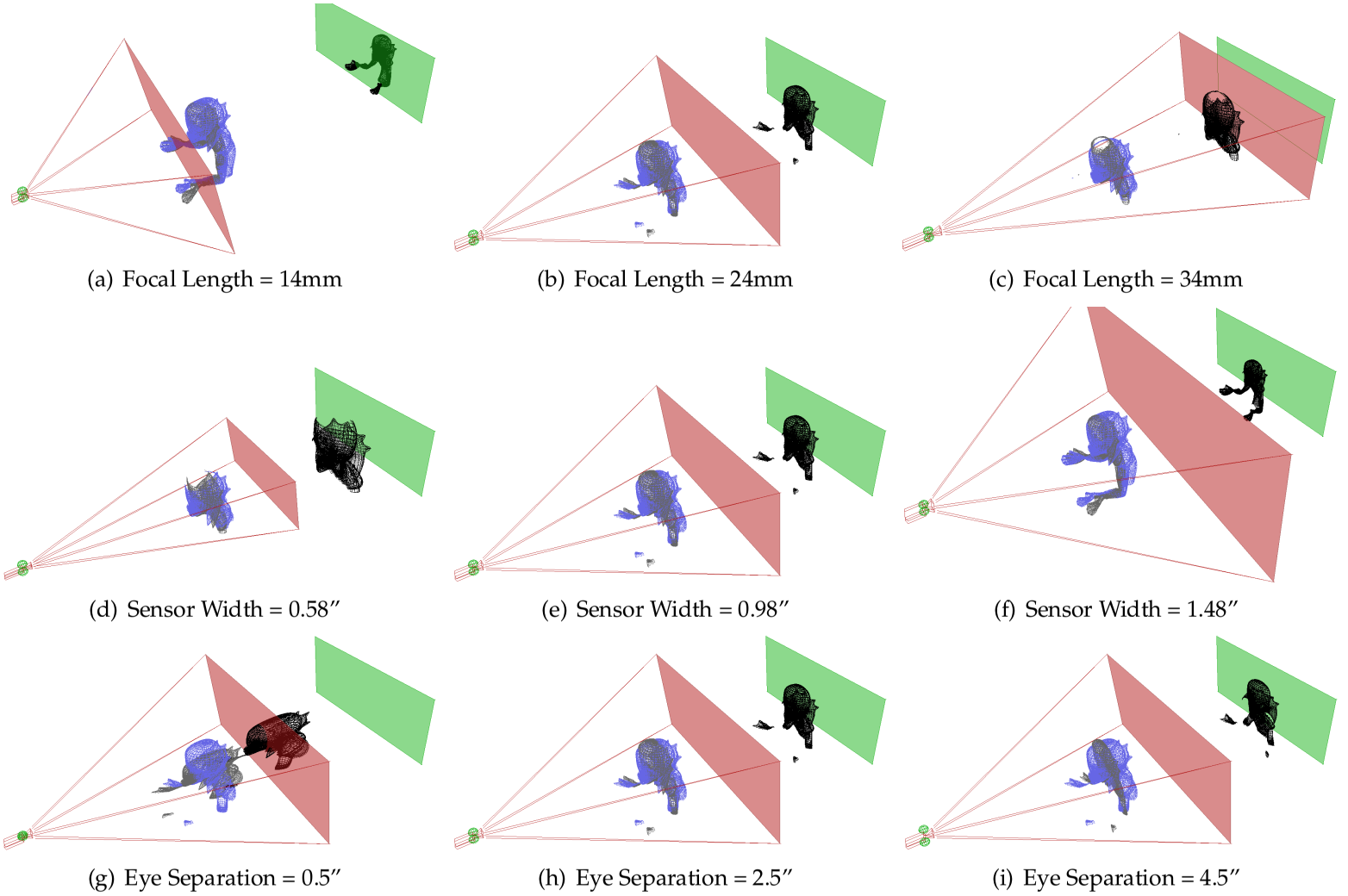
and the data that created them are used with the permission of DreamWorks Animation, and remain the property and under copyright of DreamWorks Animation. Duplication is prohibited.

5.2 Interactivity, Feedback, and Control

5.2.1 Distortion Visualization

Through the process of introducing the expanded stereo parameter space and describing the interactive tool, most of the significant distortion visualizations have been shown and described. We now iterate through examples of stereoscopic distortion visualizations, highlighting some of the remaining unexplored parameters through the use of our tool.

Figure 5.8 demonstrates the visualization of variations in focal length (f), sensor width (W_c), and eye separation (e). Because of the near linear relationship between the convergence distance (c) and the focal length, notice how variations of the focal length directly translate to variations in the convergence distance in Figures 5.8(a)-(c). For the sensor width, the inverse trigonometric relationship between it and the field of view implies an increase in the size of the view frustum as the sensor width is increased, as shown in Figures 5.8(d)-(f). Notice how this is a similar effect to shortening the focal length while simultaneously holding the convergence distance constant. Finally, in Figures 5.8(g)-(i), notice how variations in the eye separation have the near opposite effect as variations in the camera separation (t). Decreases in the eye separation lead to exaggerations in depth, similar to an increase in camera separation, while increased eye sep-



(a) Focal Length = 14mm (b) Focal Length = 24mm (c) Focal Length = 34mm

(d) Sensor Width = 0.58″ (e) Sensor Width = 0.98″ (f) Sensor Width = 1.48″

(g) Eye Separation = 0.5″ (h) Eye Separation = 2.5″ (i) Eye Separation = 4.5″

Figure 5.8: Visualization of distortions inherent in variations of stereoscopic parameters through the use of our tool.

aration leads to a visible flattening of the perceived depth towards the screen plane, similar to decreasing of the camera separation. However, it should be noted that the two parameters are not completely counteractive, as variations in the eye separation have no effect on the convergence distance while variations in the camera separation do.

5.2.2 Floating Window Violations

As discussed previously, our tool allows for the specification, variation, and visualization of the floating window, a technique used to counter the effects of stereo window violations. A side benefit of the combination of these interactions with the simultaneous visualization of scene and transformed geometries is that artists can directly visualize these window violations. Figures 5.9(a)-(b) and 5.10(a) demonstrate how the visualization trivializes the identification of window violations.

Through the feedback loop created by simultaneous interaction and visualization, the artist is capable of intuitively setting the floating window to reduce and eliminate the manifestation of these violations, as in Figures 5.9(c)-(d) and 5.10(b). By directly manipulating the floating window plane, the artist can simply position the plane in a non-violating configuration without reference to or knowledge of the specific values used to define the floating window. Thus, this feedback interaction loop can improve the ability and reduce the time taken by artists to identify and rectify the occurrence of stereo window violations.

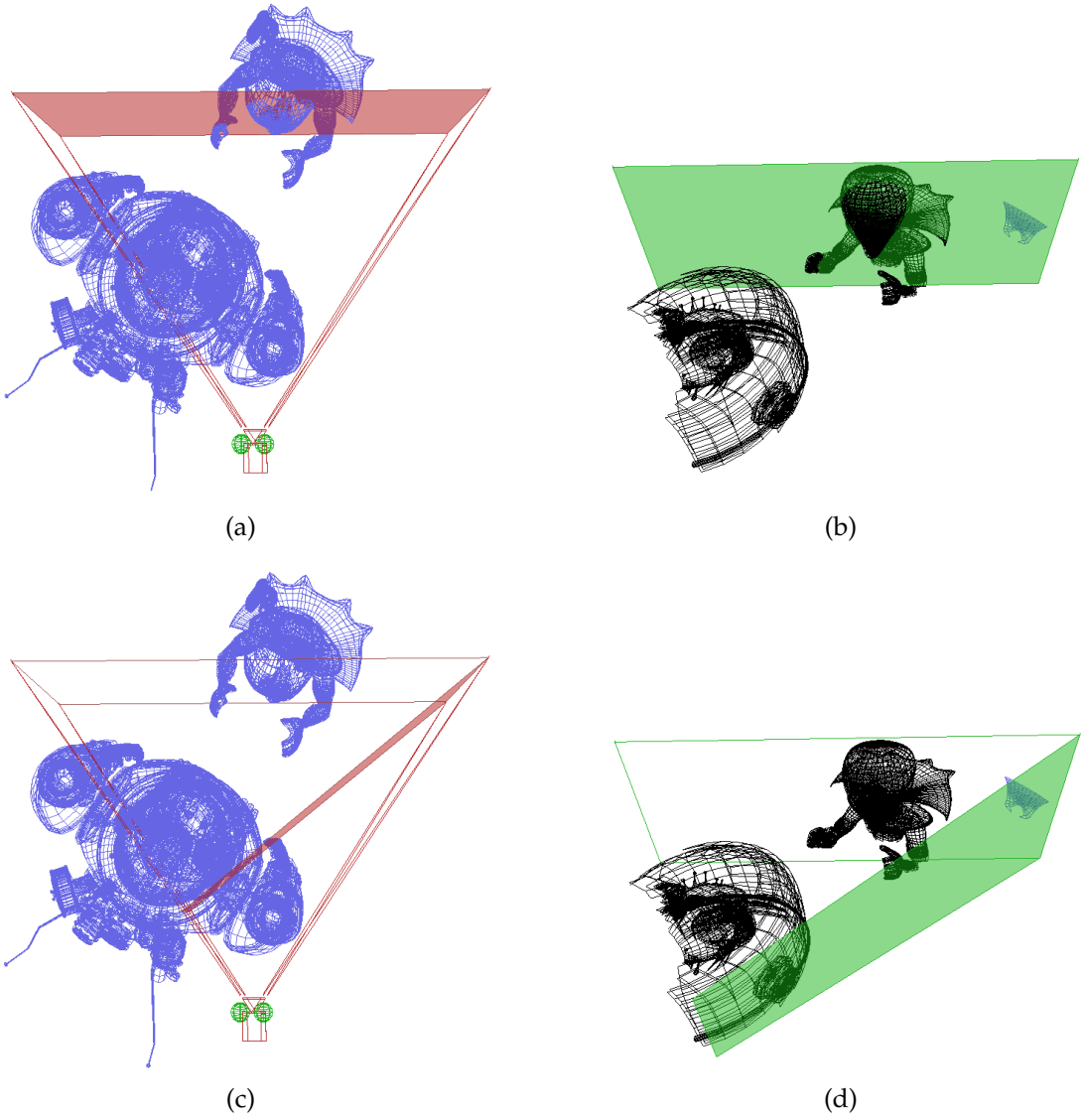
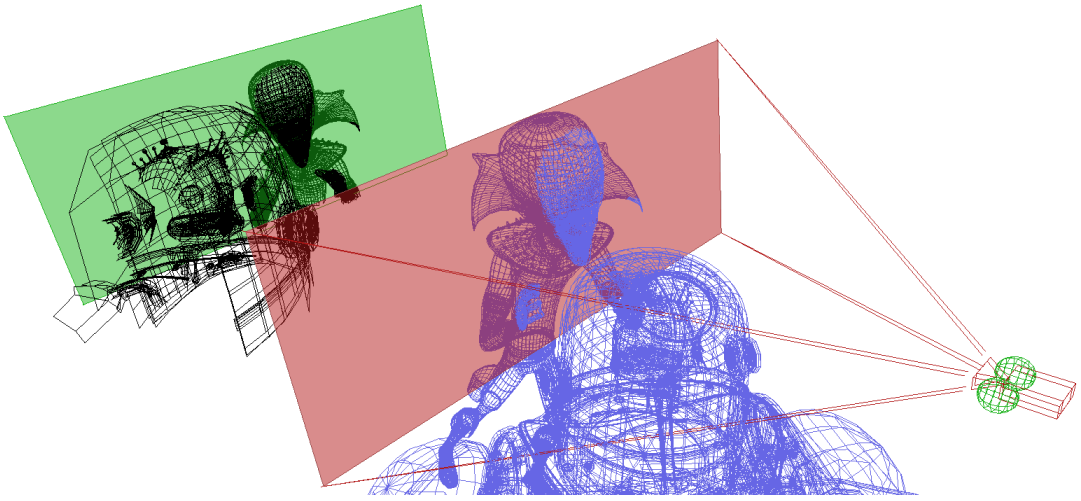
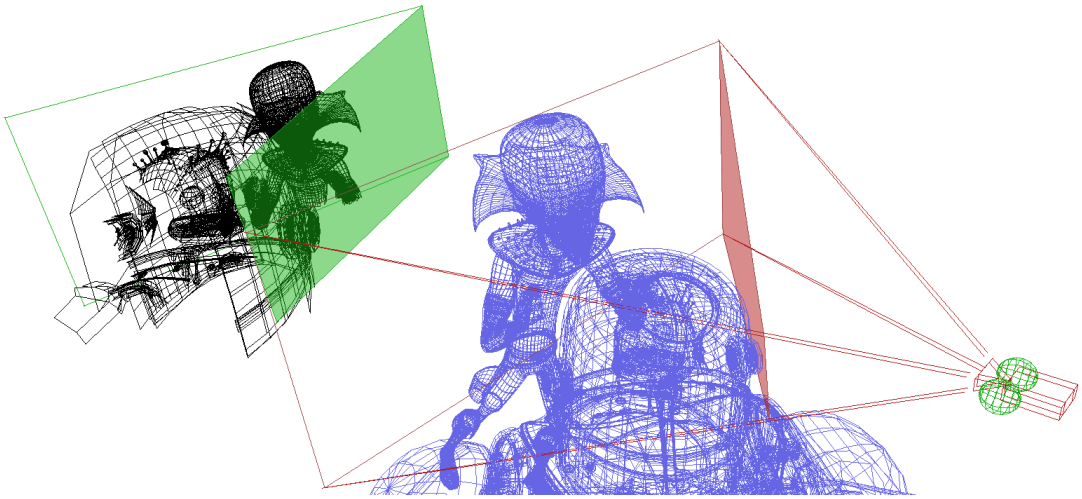


Figure 5.9: Stereo window violations and rectification using floating windows. Figures (a) and (b) demonstrate the existence of a stereo window violation. Notice how the near character breaks the left side of the frame but perceptually is in front of the stereoscopic window, creating a violation. Figures (c) and (d) demonstrate how the violation is corrected by applying a floating window to the left side of the frame so that the character now perceptually appears behind the stereo framed window. Figure 5.10 provides an alternative perspective of the same violation and correction.



(a) Window Violation



(b) Floating Window Correction

Figure 5.10: An alternative view of the same stereo window violation and correction with floating windows as in Figure 5.9.

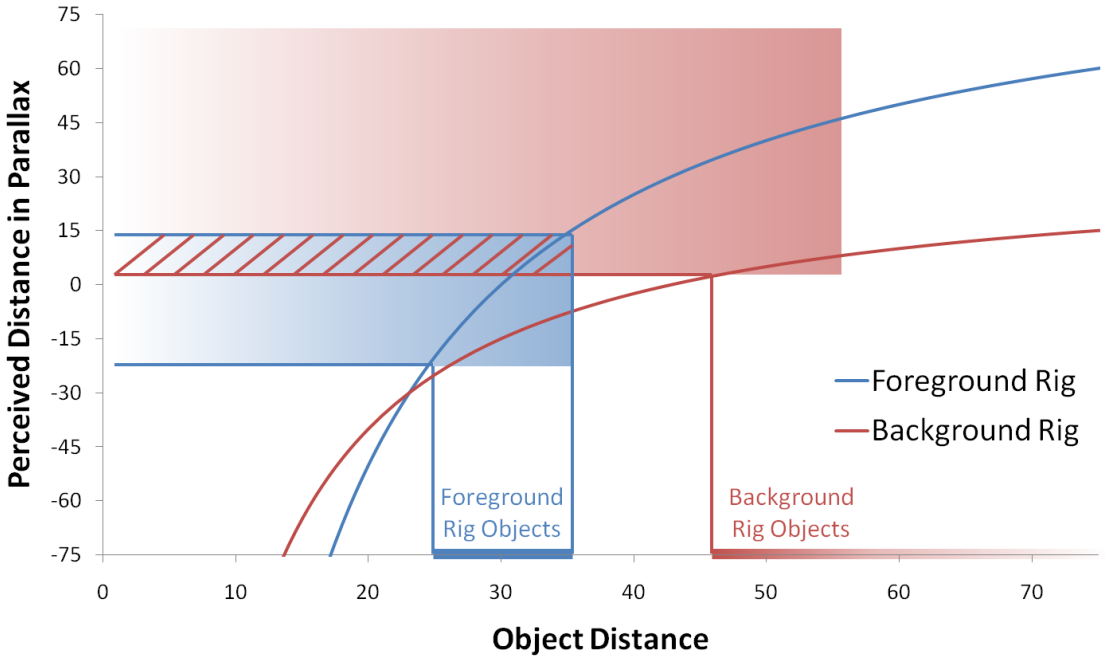


Figure 5.11: Multirig depth incongruities. Chart of real world object distances, and the perceived depth after stereo transformation in parallax. The foreground rig, in blue, contains objects from 25 to 35 units from the camera, and based on its stereoscopic transformation, the objects are mapped to a parallax range of -20 to +15 pixels of parallax. The background rig, in red, contains objects from 47 units or greater, and based on its stereo transform, the objects are mapped to a parallax range of 3 pixels or larger. This leads to an overlap of parallax values in perceived space (red hatched region) of objects in the two rigs that do not overlap in the real world, creating a depth incongruity and possibly indicating an incorrect multirig configuration.

5.2.3 Multirigged Camera Incongruities

A similar interactive feedback loop exists for the identification and rectification of depth incongruities, resulting from misaligned multirigged stereo camera setups. As mentioned previously, the use of multirigged cameras is a technique that segregates scene objects into stereo cameras with varying stereo parameters in an attempt to improve the stereo precept of objects in varying foreground and background regions of the frame. This occurs when a single set of parameters would not work.

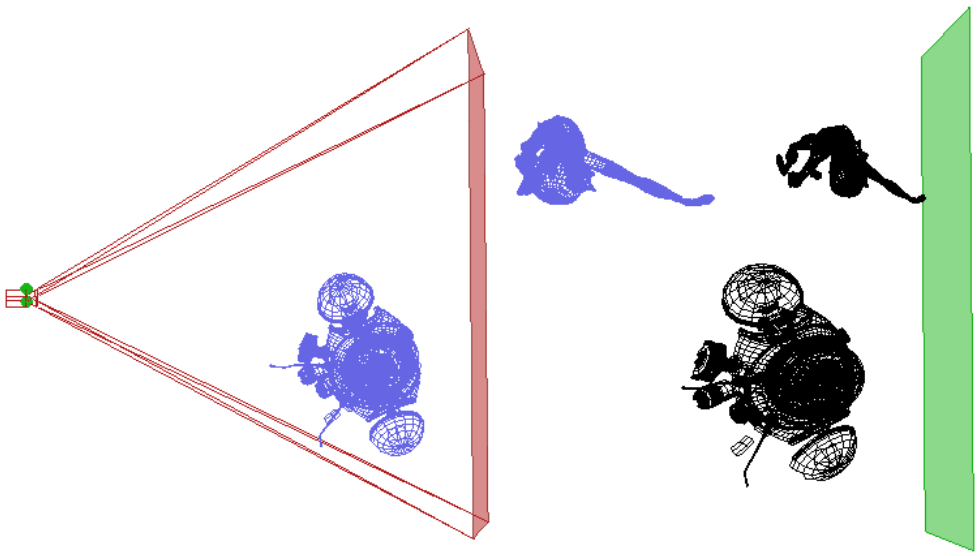


Figure 5.12: Visualization of multirig depth incongruities between the camera space and viewer space, due to simulated incorrect multirig configurations. Notice how while in camera space the characters, which are in separate camera rigs, are at different distances from the camera, while in the viewer perceived space the characters appear to be at the same depth, indicating the existence of an overlapping transformation regions, as in Figure 5.11.

One issue that arises with the use of multiple sets of stereo cameras and parameters is the possibility of incorrectly ordering the perceived depths of objects. This can potentially occur in the overlapping transformed regions between two or more multirigged cameras, as in Figure 5.11. Without being able to simultaneously visualize the complete composited stereo image, it is difficult to identify such depth incongruities by setting the stereo parameters for each rig individually. A frequent issue when using multirig setups is that these depth violations are not immediately identified by artists upon setting the stereo parameters.

Through our tool, by simultaneously visualizing the stereoscopically transformed geometry, identification of depth incongruities due to stereo multirig errors becomes straightforward, as demonstrated in Figure 5.12. While rectify-

ing such depth violations is not as intuitive as it is for stereo window violations, our interactive tool does provide significant feedback, enabling the artist to manipulate the separate rig parameters individually and visualize correct multirig depth ordering.

CHAPTER 6

DISTORTION METRIC AND PERCEPTUAL CONSTRAINTS

One obvious extension to the development of our interactive visualization tool for distortions inherent in the stereoscopic parameter space, is the quantization of the visualized distortion being observed, a metric of stereoscopic quality. Distortion quantization provides numeric confirmation and feedback to the artist about the amount or quality of distortion observed in the visualization, a standardized reference point for distortion comparisons across configurations and scenes. Additionally, it allows for the visualization of the surface of the distortion metric across variations in the stereoscopic parameter space. Once the distortion metric surface is visualized, it is advantageous to consider constraints imposed on the stereo parameter space by human perceptual limitations. These constraints can be visualized as boundaries on the distortion metric surface, indicating the valid operating subspace of the full parameter space.

To demonstrate the capabilities and effectiveness of our visualization tool, we define and implement one prototype of a stereoscopic distortion metric. We also define and implement a set of perceptual boundary constraints to the parameter space. Both the constraints and distortion metric are visualized, and the impact of visualizing a combined parameter and distortion metric space is discussed.

6.1 Stereoscopic Distortion Metric

As previously discussed, there have been previous attempts to quantize the quality of the stereoscopic transformation. However, most have suffered from

assumptions that restrict the configurability of the parameter space, or are defined to measure too particular a metric, such as non-linear depth compression or expansion, cardboarding, or puppet theater effects, all of which limit its applicability and generalization. For example, many metrics are computed based solely on the stereoscopic parameters. In some instances, the positioning of the character or object of focus is incorporated, and in others reference to the near and far distances are made. In either case, the actual content of the scene does not come into play, and thus the impact that the stereoscopic parameters have on the perceived transformation of that particular scene's geometry and its resultant distortions from the original is not considered.

To mitigate these limitations, our stereoscopic distortion metric is based solely on the distortions between the transformed geometry and the original source geometry, only indirectly referencing the stereoscopic parameter space in so much as they affect the scene geometry transformation. Thus, beyond being parameter specific, our quality-of-stereo metric is scene content specific, measuring variations beyond just camera separation or focal and interest distances to encompass all elements that go into the composition of a shot. Accordingly, our metric for stereoscopic quality is a quantization of stereoscopic distortion.

6.1.1 Quantization of Distortion

Stereo distortion can be generally defined as variations in perception between the original geometry as captured by the camera system and the stereoscopically transformed geometry as perceived by the viewer. Because our visualization tool implements a comprehensive model for stereoscopic perception and ap-

plies it to original geometry, we can refine our definition of quantized distortion to be the geometric discrepancies between the two geometries as maintained by our tool.

As our perceptual model works on a vertex-by-vertex basis for computing the stereoscopic transformation between the original and stereo transformed spaces, it is sensible to compute a distortion metric on a similar scale, making comparisons and computing the discrepancies between the two geometries on a vertex-by-vertex basis. Accordingly, our distortion metric is primarily based on the L^2 norm, or Euclidean length, of the distances between points in three-space.

However, previous findings warrant further discussion as to the validity of a naive implementation of a direct measure of vertex distance discrepancies between the original and transformed geometries as defined by our tool. Recall that the camera configuration components and the original geometry are not in the same space as the viewer configuration and stereoscopically transformed geometry. Because of the physical disconnect between the capture and display spaces, the relative positioning and orientation between them can be arbitrarily defined. Our tool visualizes the two spaces in the same space for the benefit of the artist, and for the convenience of viewing them simultaneously. They are positioned relative to each other so that direction, orientation, and scale are consistent between the two spaces, but otherwise they bare no physical proximity. This same realization led to the development of our visualization of the ghosted geometry to allow for more direct comparisons of geometries between spaces, as discussed in Section 5.1.3. Similarly, our definition of a distortion metric should adhere more directly to a more applicable comparator than the proximally unassociated spaces.

We have previously explored in Section 2.1 how the human visual system primarily perceives depth in relative, and not absolute, terms. This suggests that the relative depth, or spatial ordering, as well as the accuracy of the perception of the shape should be prioritized over the exact spatial positioning of the precept. This further emphasizes the utility of the ghosted geometry over the base stereoscopically transformed geometry in the comparison with the original geometry for use as a distortion metric. By reducing the emphasis on the positioning and scale of the transformed geometry and focusing more on its stereoscopically transformed form, our distortion metric can more aptly be considered to measure distortion in shape. Shape distortions are most easily observed and computed by comparing the original geometry with the stereoscopically transformed, and position and scale invariant ghosted geometry.

6.1.2 Definition of Shape Distortion

Ghosted geometry is defined as the near-affine invariant stereoscopically transformed geometry. Visualized at the position and size of the original geometry, it is an attempt to highlight the shape inconstancy inherent in the stereo transformation, precisely what we hope to quantize. The scale and translational distortions used in the ghosted geometry are determined by Equations 5.3 and 5.5, reiterated here.

$$D_t[X, Y, Z] = \frac{\sum_{vertices} [X_0, Y_0, Z_0]}{\|\text{vertices}\|} - \frac{\sum_{vertices} [X_i, Y_i, Z_i]}{\|\text{vertices}\|} \tag{5.3}$$

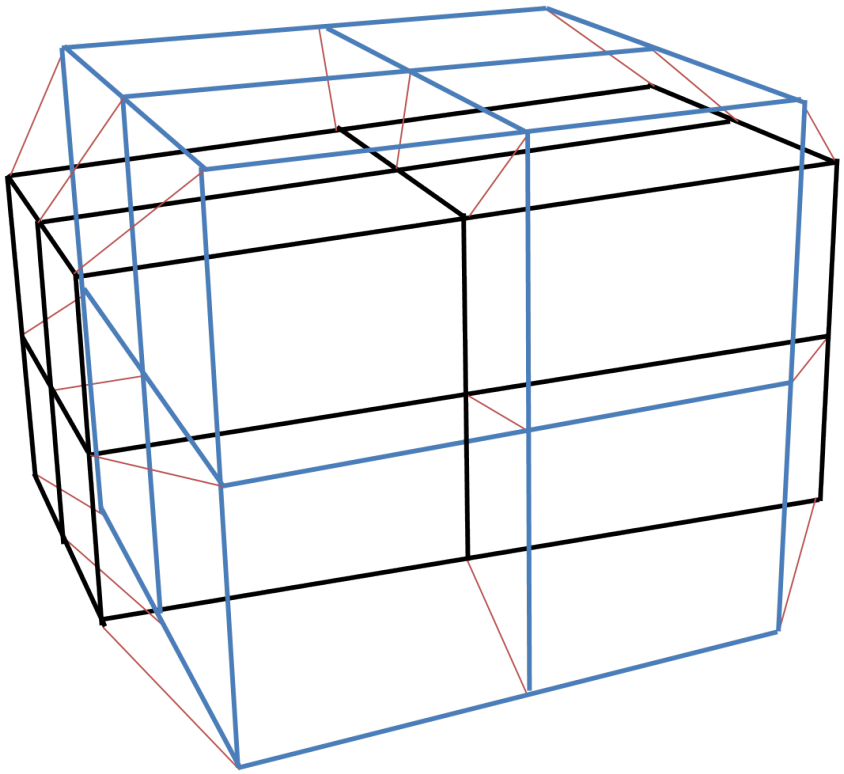


Figure 6.1: Visualization of the definition of stereoscopic shape distortion. The distortion metric is defined as the sum of near-affine invariant square L^2 norms of the vector translational discrepancies (red) between the vertices in the original (blue) and stereoscopically transformed geometries (black).

$$D_s = -\frac{\sum_{vertices}\sum_{W=X,Y,Z}\left[(W_i - C_i[W]) \cdot (C_i[W] + D_t[W] - W_0)\right]}{\sum_{vertices}\sum_{W=X,Y,Z}\left[(W_i - C_i[W])^2\right]} \tag{5.5}$$

Here, C_i is the centroid of the transformed geometry, and for brevity, the individually written out sums of the X, Y, and Z coordinates are compacted into sum notation over $W = X, Y, Z$.

Recall that in the determination of the scale distortion component of the ghosted geometry invariance that a metric of shape discrepancy was minimized with respect to D_s. That metric is the basis of our stereoscopic distortion metric defined here, and is similar except that it is computed after the optimal

translational (D_t) and uniform scale (D_s) invariants are determined. We define our quantization of stereoscopic distortion metric (M) as the sum of near-affine invariant square L^2 norms of the vector translational discrepancies between the vertices in the original and stereoscopically transformed geometries. The lengths of the vertex discrepancies is illustrated in Figure 6.1.

$$\begin{aligned} M = & \sum_{vertices} \Big((X_i - C_i[X]) \cdot D_s + C_i[X] + D_t[X] - X_0 \Big)^2 \\ & + \sum_{vertices} \Big((Y_i - C_i[Y]) \cdot D_s + C_i[Y] + D_t[Y] - Y_0 \Big)^2 \\ & + \sum_{vertices} \Big((Z_i - C_i[Z]) \cdot D_s + C_i[Z] + D_t[Z] - Z_0 \Big)^2 \end{aligned} \tag{6.1}$$

We use the sum of squared distances instead of just distances as it makes iteration for summation simpler as the individual coordinates can be pulled outside of the square root to be grouped and summed by coordinate instead of by vertex. Additionally, it applies a second-order penalty to geometry that is more significantly distorted, which helps to highlight subtle variations in distortion.

Beyond being a numeric representation and reference value for the distortions inherent in the stereoscopic transformation as defined by the complete parameter space, the metric itself can be visualized by plotting the shape distortion at varying points within the parameter space. This provides a visualization of the distortion surface over a parametrically enumerated subset of the parameter space, a graphical representation of the relative goodness or badness of the stereo transformation. For example, considering variations in the camera separation and sensor offset, we can visualize the impact that modification of the intrinsic stereo camera parameter subspace has on the stereo precept, as in Figure 6.2. The X and Y axes represent ranges in the camera separation and

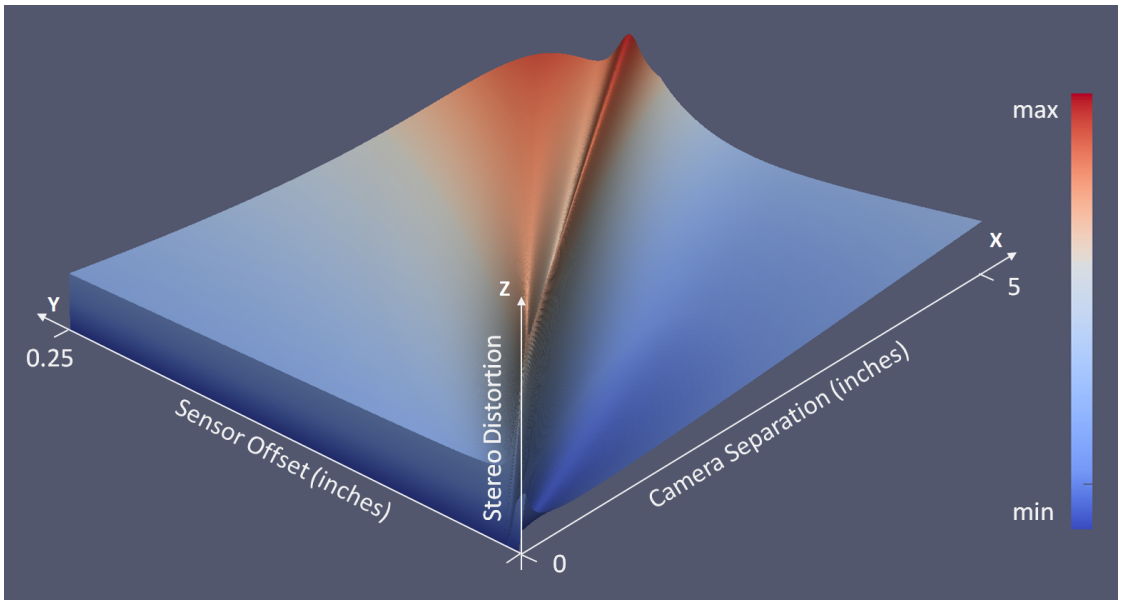


Figure 6.2: The distortion metric surface over variations in the intrinsic stereoscopic camera parameters. The *X* and *Y* axes represent ranges in the camera separation and sensor offset parameters respectively, and the Z axis is the computed stereo distortion metric. Variations of the two parameters can have a drastic impact on the quality of the stereo precept as is seen by the contours of the metric surface, from minimal distortions in blue to significant distortions in red.

sensor offset parameters respectively, and the *Z* axis is the computed stereo distortion metric. The resulting surface shows how variations within the camera separation-sensor offset subspace can significantly vary the quality of the stereo transformation.

6.1.3 Culling

Having defined shape distortion as a function of the stereoscopic transformation of the scene geometry, we need to determine which portions of the scene geometry should be considered for inclusion in the metric. In particular, when working within a production animation pipeline, an individual shot may only

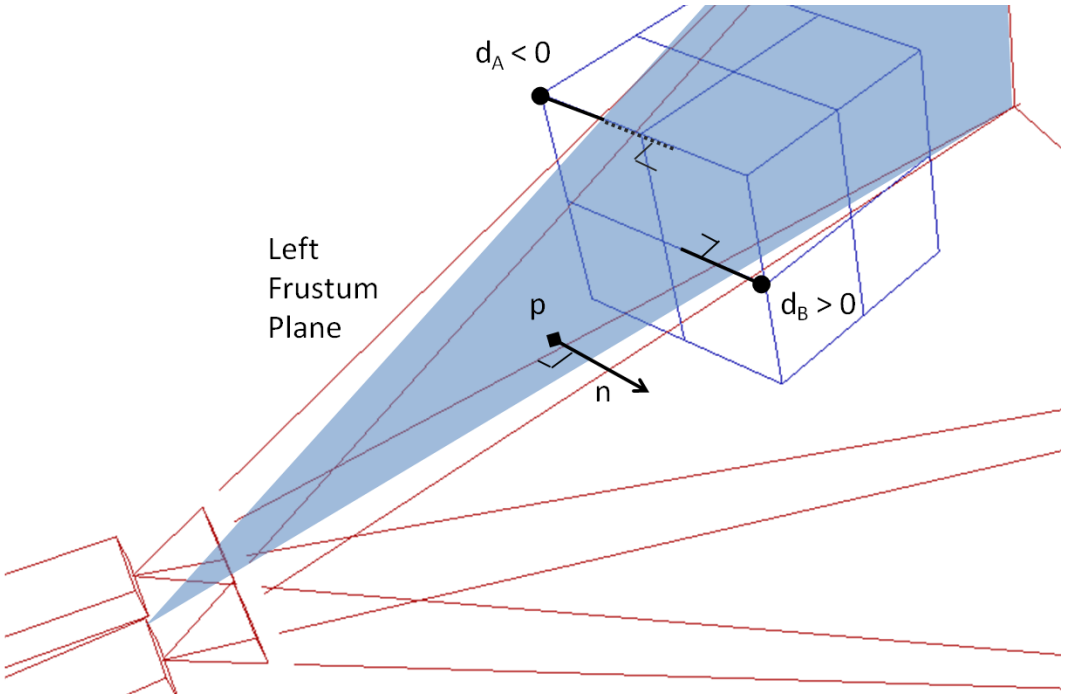


Figure 6.3: View frustum culling. The left plane of the stereo view frustum, in blue, is defined by a point p on the plane and a normal vector n oriented into the frustum. Vertices (v) are defined as inside or outside the frustum based on the sign of their computed distance from the plane. Vertices with a positive distance ($d_B > 0$) may be inside the frustum, while vertices with a negative distance ($d_A < 0$) are outside of the frustum and are culled.

frame a small portion of the available geometry, as entire set locations are traditionally modeled at once to be covered and utilized in multiple shots from varying angles. Because our distortion metric is intended to measure the visual distortions between geometries, it makes sense to only consider visible geometry in its computation. We briefly outline the importance, inclusion, and implementation of common culling and geometry rejection algorithms applicable to the stereo camera and viewer space configurations and the distortion metric.

View Frustum Culling

View frustum culling is a technique of rejecting geometry that lies outside the visible field of view of a camera or viewer. As we assume that the viewer has complete visibility of the entire projection surface, we apply view frustum culling to the stereo camera field of view only. The algorithm works by defining the four planes that determine the left, right, top, and bottom edges of the field of view and rejecting vertices that lie on the non-viewable side of any of the culling planes. This technique is illustrated in Figure 6.3. Note that, as mentioned previously, the definition of the stereo field of view is somewhat arbitrary. As before, we use the definition of the stereo view frustum as the field of view from the point between the stereo cameras of the convergence plane for the determination of the view frustum culling planes.

The technique for determining plane sidedness for the rejection component of the algorithm is defined by the equation for the distance (d) of a point (v) from a plane defined by a point on the plane (p) and the normalized normal vector (n) to the plane.

$$d = n \cdot (v - p) = n_x v_x + n_y v_y + n_z v_z - n_x p_x - n_y p_y - n_z p_z \tag{6.2}$$

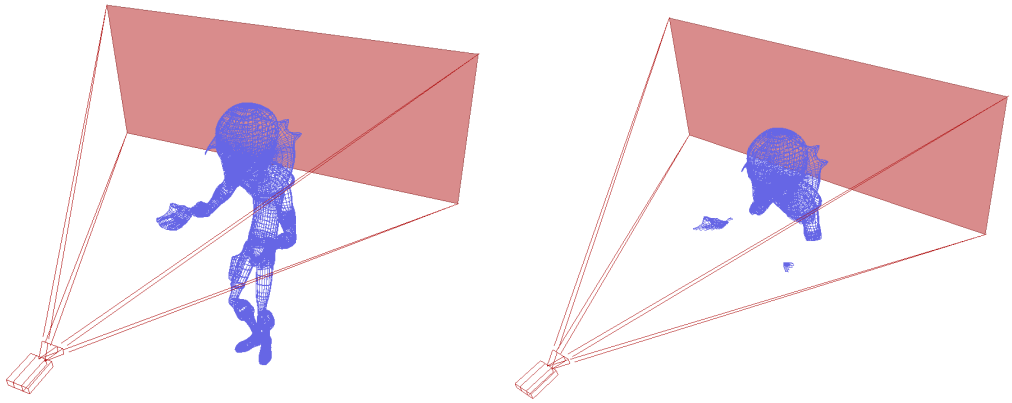
If the normal vector used in the plane definition is the one that points into the view frustum, then vertices that lie outside the field of view are a negative distance away from the plane. Accordingly, the culling criterion can be stated as

$$d(p, n, v) < 0 \tag{6.3}$$

The algorithm for view frustum culling as implemented for geometry inclusion in the distortion metric is summarized in Algorithm 1, and the impact on the geometry included in the distortion metric is highlighted in Figure 6.4.

Algorithm 1 View Frustum Culling

```
for all vertices v do
  for all view frustum planes (p_i, n_i) | i = 1 to 4 do
    if d(p_i, n_i, v) < 0 then
      cull(v)
    end if
  end for
end for
```



(a) (b)

Figure 6.4: The effect of view frustum culling within the visualization tool. Figure (a) shows the geometry of a character partially outside of the view frustum when no view frustum culling is applied. Figure (b) shows the same character geometry with view frustum culling applied. This ensures that the distortion metric is more specific to the geometry visible to the viewer.

Back Face Culling

Back face culling is a technique of rejecting geometry whose normal vector is oriented away from the viewer or camera. Assuming that the normal vectors are properly defined on the outward and visible face of a polygon in a watertight mesh (without holes or gaps), this algorithm rejects geometry on the back, non-visible side of an object from the point of view of the stereo camera. While rejecting geometry on the back side of objects may appear to violate our assertion that the distortion metric quantizes discrepancies in object volume, it is the

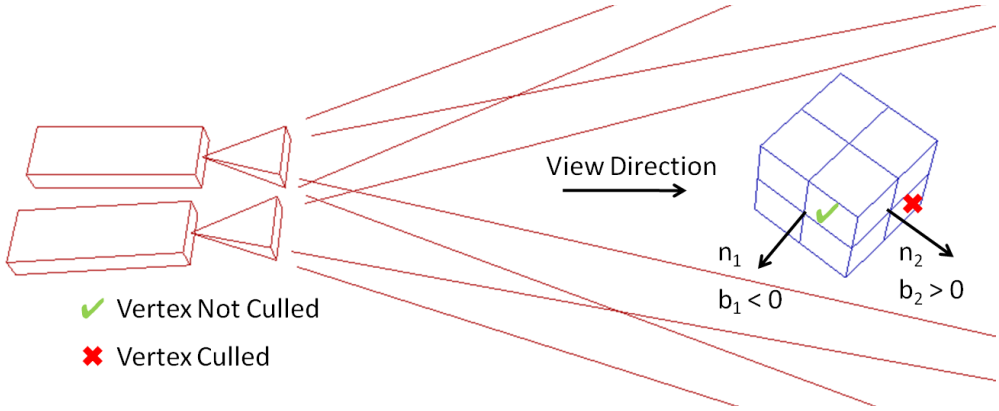


Figure 6.5: Back face culling. Vertices in faces that are oriented away from the camera are not included in the final stereo precept as they are not seen by the camera, and should be culled. The orientation of the face is determined by the dot product of the normal to the face and the camera orientation direction. Faces with normals oriented in the opposite direction as the camera direction have a negative dot product and are visible to the camera, while faces with normals oriented in the same direction as the camera direction have a positive dot product and are not visible to the camera and are culled.

volume as perceived by the viewer that matters, and that perception can only be based upon what is visible to the viewer.

Back face geometry is determined by the orientation of the normal vector. If the normal vector is oriented towards the camera, then the surface is visible, while geometry with normals oriented away from the camera is not visible, as illustrated in Figure 6.5. Orientation is determined by the dot product of the normal (n) with the viewing direction ($[X_t, Y_t, Z_t] - [X_c, Y_c, Z_c]$).

$$b = n \cdot ([X_t, Y_t, Z_t] - [X_c, Y_c, Z_c]) = n_x X_t + n_y Y_t + n_z Z_t - n_x X_c + n_y Y_c + n_z Z_c \tag{6.4}$$

If the dot product is negative, then the geometry faces the camera and is visible. But if the dot product is positive, then the face is oriented along the viewing direction and is not visible and should be culled. Accordingly, the culling criterion can be stated as

$$b(n, Camera) > 0 \tag{6.5}$$

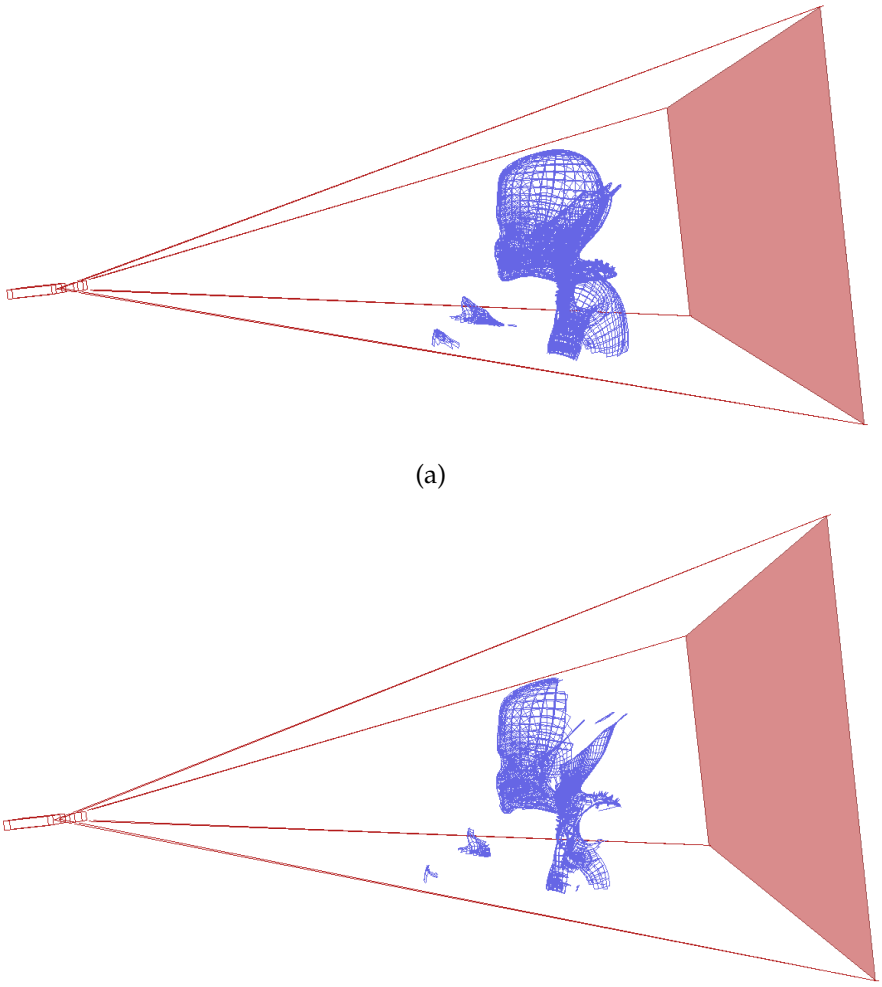
Algorithm 2 Back Face Culling

```
for all vertices v do
  if b(n_v, camera) > 0 then
    cull(v)
  end if
end for
```

The algorithm for back face culling as implemented for geometry inclusion in the distortion metric is summarized in Algorithm 2, and the impact on the geometry included in the distortion metric is highlighted in Figure 6.6.

Occlusion Culling

Occlusion culling is a technique for rejecting geometry that is occluded from the camera or viewer by other geometry. The occluding geometry must be opaque so as to completely block the visibility of the occluded geometry, and can either be geometry of the same object (self-occlusion) or another object. The algorithm works by determining the visibility of each vertex that could potentially be included in the distortion metric with respect to each possibly occluding surface. This involves identifying occluding surfaces or polygons, and for each, iterate over and determine the visibility of every vertex. Because our tool maintains the connectedness graph, or faces, of the mesh vertex geometry, the set of possible occluding surfaces is easily defined and populated by the set of visible faces. A face is considered visible if any of its component vertices is visible. The occlusion culling algorithm is slightly more complex than the previously discussed culling algorithms, as it is $O(n^2)$ in the number of polygons, and should accordingly be computed after the results of the linearly complex culling algorithms are determined.



(a)

(b)

Figure 6.6: The effect of back face culling within the visualization tool. Figure (A) shows the geometry of a character with view frustum culling already applied, but without back face culling. Figure (B) shows the same geometry but with back face culling applied, as portions of his head and back that are not visible from the perspective of the stereo cameras are culled. Again, this ensures that the distortion metric is more specific to the geometry visible to the viewer.

The rejection criterion is similar to view frustum culling, but where the frustum is determined by the edges of the occluding polygon, and a vertex is rejected if it is inside the occluding frustum and behind the surface. Figure 6.7 illustrates an example with one occluding polygon and several rejected and included vertices. Accordingly, the culling criterion for an occluding polygon (g)

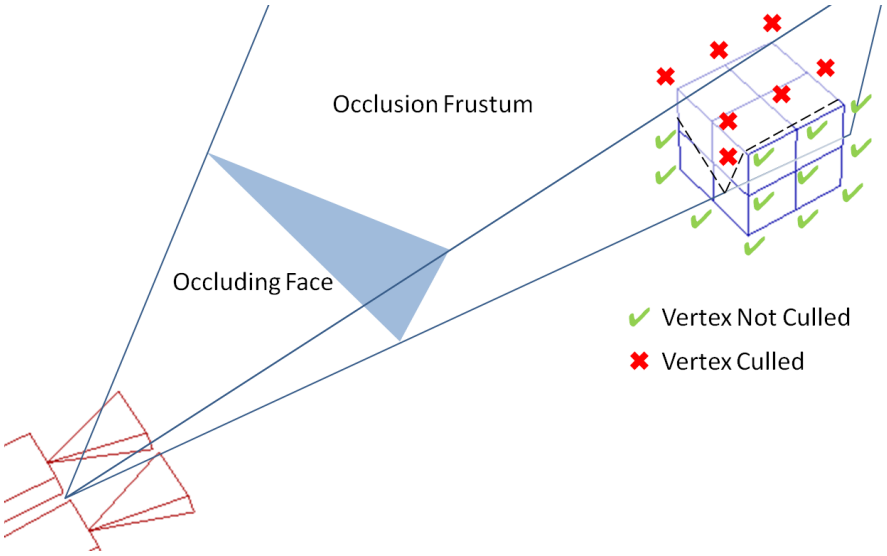


Figure 6.7: Occlusion culling. Vertices that are occluded by other geometry in the scene from the perspective of the stereo cameras will not be visible to the viewer and should be culled. Vertex occlusion is determined for every vertex-face pair in the scene. Vertices within the frustum defined by the edges of the occluding face are culled, where occlusion frustum containment is determined similarly to view frustum culling by determining the sign of the distance of the vertex from the planes of the frustum.

can be stated as

$$d(p_{g,i}, n_{p,i}, v) > 0 \; \forall \text{ frustum planes } i \text{ of } g \tag{6.6}$$

$$v_z > g_z \tag{6.7}$$

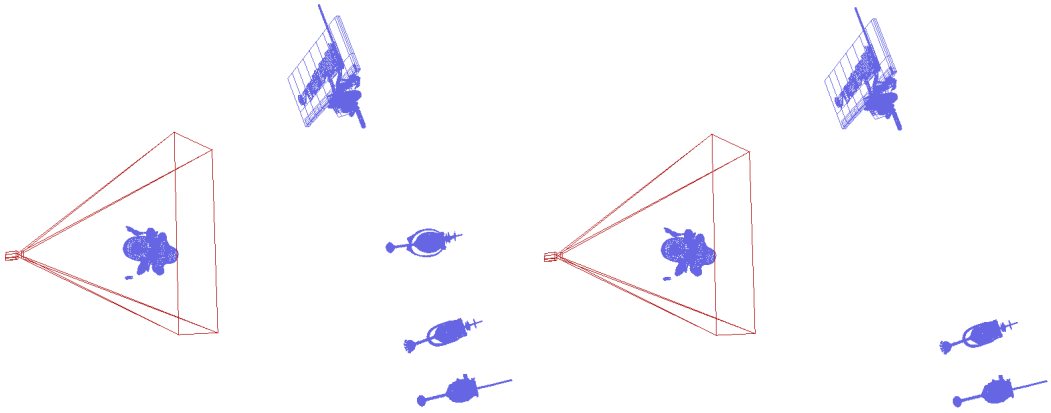
The algorithm for occlusion culling as implemented for geometry inclusion in the distortion metric is summarized in Algorithm 3, and the impact on the geometry included in the distortion metric is highlighted in Figure 6.8.

6.1.4 Geometry Grouping

It is necessary to make the distinction between the full scene geometry and the geometry of individual objects. Individual object geometry is defined as any

Algorithm 3 Occlusion Culling

```
for all vertices v do
  for all occluding polys g and frustum planes i of g, defined by (p_{g,i}, n_{g,i}) do
    if d(p_{g,i}, n_{g,i}, v) > 0 ∀ i then
      if v_z > g_z then
        cull(v)
      end if
    end if
  end for
end for
```



(a) No Culling (b) Occlusion Culling

Figure 6.8: The effect of occlusion culling within the visualization. Figure (a) shows the complete geometry of a shot, without occlusion culling applied. Figure (b) shows the same geometry but with occlusion culling applied. The geometry of one of the robots is culled due to occlusion from the stereo cameras by the geometry of the character inside the viewing frustum as indicated. Occlusion culling ensures that the distortion metric is more specific to the geometry visible to the viewer.

single instance of an individual geometric model, like a rock or a character or a car. Thus the full scene geometry can be segmented into geometrically independent entities. This is important because thus far we have considered the near-affine invariance of the full scene geometry as a whole. This implied that there is a single translational distortion vector (D_t) and a single scale distortion value (D_s) that are applied to all objects collectively. While this might work for

objects located near the centroid of the full scene geometry, it does not for objects at the periphery.

This implies that affine invariance and distortion metrics should be computed on an object by object basis to avoid the bias of the full geometry. With this restriction, it is possible to correctly position and scale the ghosted geometry for visualization of the distortion of object shape. It further emphasizes the use of shape distortion as a metric while still maintaining relative depth and positioning accuracy.

Accordingly, our tool maintains a set of centroid values ($C_{i,n}$), translation ($D_{t,n}$) and scale ($D_{s,n}$) distortion invariance values, and distortion metric values for each object n as defined by the scene and modeling artists.

6.1.5 Object Weighting

As a result of object geometry grouping, our tool maintains a set of distortion metric values, one for each object in the scene. However, a potentially massive set of values is less applicable as a single metric to an artist, and the set of values must be aggregated into a single value for the metric to be effective. As a result, our tool implements a variety of optional weighting functions to determine a single distortion metric.

A trivial method would be to sum the distortion metrics of the individual objects, however the resulting metric would scale with the number of objects in the scene. The result would not be comparable between scenes with varying numbers of objects, negating the intent of the metric.

Alternatively, one could average the individual object distortion values. Since the original distortion metric is defined as a sum of vertex disparities, this weighting would scale with the number of vertices in an object. This effect may be advantageous, as it has the benefit of emphasizing objects with a larger number of vertices, one measure of object complexity which artists may consider important. However, a simple vertex count may not be the most ideal measure of complexity, and complexity may not be the most important factor of object importance.

Another possible weighting scheme is to average the distortion offset across all vertices by summing the total object level distortion values and dividing by total number of non-culled vertices.

$$M = \frac{\sum_{\substack{\text{objects}\\ n}} \sum_{\substack{\text{vertices}\\ v \text{ in } n}} \sum_{W=X,Y,Z} \left[\left(W_{i,v} - C_{i,n}[W] \right) \cdot D_{s,n} + D_{t,n}[W] - W_{0,v} \right]^2}{\|\text{vertices}\|} \tag{6.8}$$

where M is the total overall stereoscopic distortion metric for the scene. This is the most unbiased weighting distribution, and has the advantage of being normalized across arbitrary scene geometries allowing direct relative comparisons between them. As an unbiased value, we use this computed metric as the baseline for any weighted metrics that an artist might decide to use. Possible additional weighting functions, such as angular size subtended, artist specified object importance, distance from point of interest, or other aspects of saliency, can be easily built upon this baseline value and is left for future work.

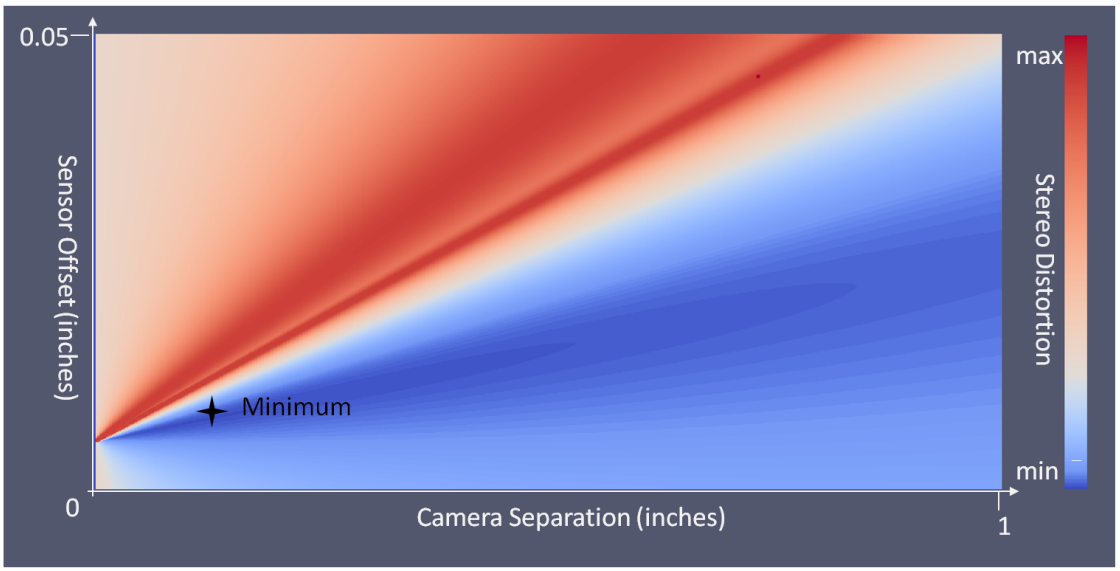


Figure 6.9: Plan view of the distortion metric surface. A visualization of the impact of variations of the intrinsic stereo camera parameters on the perceived stereo distortion. Again, blue indicates minimal distortion, while red indicates significant distortion. The minimal amount of distortion within this parameter subpsace is indicated, and is known as the orthostereoscopic condition.

6.2 Perceptual Constraints on the Parameter Space

Having formally defined a stereo quality metric, we can revisit and further explore the implications of visualizing the contours of the distortion metric surface over variations in a subset of the parameter space, as mentioned previously in Figure 6.2. As a graphical representation of the relative quality of the stereoscopic transformation over parameter variations in a linear space, we can quickly identify parameter configuration states that exhibit large amounts of distortion, and others that exhibit minimal amounts of distortion. Being able to visualize this surface is important, as it can be ideal for an artist to attempt to minimize the amount of distortion inherent in the stereoscopic transformation. However, considerations including the entirety of the distortion metric surface are impractical, as many include physically invalid, perceptually non-

ideal, or eye-straining configurations. Accordingly, it is important to observe a valid subspace of the complete parameter space, as constrained by perceptual limitations, when considering the distortion metric surface. In the following sections, we outline the importance of perceptual constraints and the considerations for their practical implementation.

It becomes apparent that an attempt to computationally minimize the distortion metric with respect to the varying parameters would optimize the perceived distortion in shape resulting from variations in the selected subset of parameters, as shown in Figure 6.9. In Appendix A, we briefly discuss one technique for such an optimization, as well as the necessary considerations for its practical implementation, and discuss its implications.

6.2.1 Unconstrained Vs. Constrained Parameter Space

As previously discussed, there exists a particular stereo parameter configuration where the stereo precept in the display space exactly matches the precept in the capture space, as if the viewer were observing the scene while being captured from the position of the stereo cameras. This is known as the orthostereoscopic condition. Near orthostereoscopic conditions are visualized in Figure 6.10. The defining parameters correspond to the absolute minimum of the distortion metric, or zero distortion, as seen in Figure 6.9. It requires that every parameter be matched between the capture and display spaces. The field of view and convergence distance of the stereo cameras must match the viewer's field of view of and distance to the display surface which involves aligning focal lengths and display widths, and the camera separation must match the interoc-

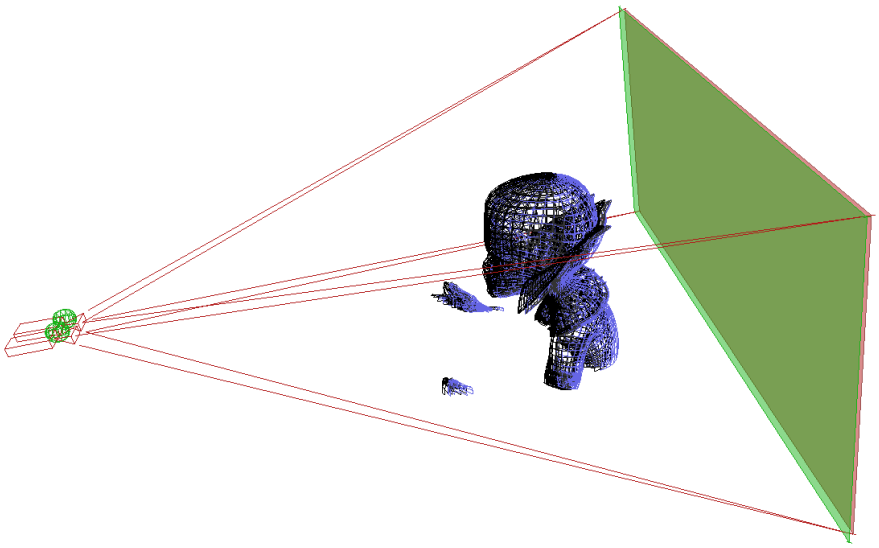


Figure 6.10: The orthostereoscopic condition occurs when all stereo parameters are matched between the capture and display spaces. Namely, the positioning and size of the convergence plane must match the projection surface implying matching fields of view, and the camera separation must match the interocular of the viewer's eyes. Note that while the orthostereosopic condition provides the absolute minimal amount of distortion, it is often not the ideal configuration when human perception and visual comfort are considered.

ular of the viewer's eyes.

By definition, this parameter configuration would automatically minimize the distortion metric by providing zero disparity between the two spaces. If an optimization technique were to freely minimize the distortion metric in a parameter subspace that included the orthostereoscopic point, it would converge to it. However, orthostereoscopy heavily restricts the configuration of the scene by locking the convergence, camera separation, and field of view even before considering the scene content. Even if artists are willing to sacrifice on stereoscopic flexibility in an attempt to achieve orthostereoscopy, without careful attention to the arrangement of the scene, object, and character geometry, the final result can frequently lead to viewer discomfort due to excessive divergence or conflicts in accommodation and vergence beyond the comfort threshold.

Accordingly, we want to limit the parameter space so that the stereo configuration stays within the limitations of visual comfort. On the other hand, we do not want to naively apply visual thresholds to the stereo configuration by simply filling the virtual stereo space within the limits of visual comfort. This sort of methodology would ensure a constant presence of stereo precepts at the thresholds of visual comfort and can cause noticeable fluctuations in stereo parameters to accommodate for any motion of objects or cameras. In turn, this leads to aesthetically unappealing distortions and waverings in perceived volume as the geometry is artificially inflated to maximize the usable stereo space, not to mention constant comfort threshold visualization without relief.

Ultimately, we need to be smarter about how we treat and consider the valid parameter space. This suggests the use of boundary constraints, and considering the stereoscopic distortion metric within the limits of visual comfort. This strikes an appropriate balance between the theoretically perfect orthostereoscopic precept and the perceptual realities of human visual comfort. The two main perceptual comfort limits used boundary constraints relate to excessive divergence and conflicts between accommodation and vergence.

6.2.2 Excessive Divergence

As previously discussed, our tool implements an allowed divergence parameter (γ) inspired by research suggesting that viewers can comfortably diverge their eyes outward by a small degree, anywhere from 0.25° to 3°. The parameter can be specified by the artist, and effectively shifts the perceived infinity point defined in parallax within the virtual viewer space, as illustrated previously

in Figure 4.7. This results in a maximum allowed parallax value in the stereo image pair.

$$p_{max} = e + 2Z_v \tan(\gamma) \tag{6.9}$$

By combining the maximum allowed parallax value (p_{max}) with the distance of the furthest object from the camera (d_{max}) and applying them to the equations relating parallax and distance to the intrinsic stereo camera parameters (Equations 4.5 and 4.4), we can arrive at the definition of our first perceptual constraint.

$$h < \frac{f \cdot t}{d_{max}} + \frac{p_{max} \cdot W_c}{W_s} \tag{6.10}$$

6.2.3 Accommodation-Vergence Conflict

The second constraint is derived from the limits of the accommodation-vergence conflict. As discussed previously, the distance of focus and the distance of convergence of human eyes are the same in natural viewing. However, viewing stereoscopic displays traditionally involves focusing on the plane of the display surface while converging at various distances in space depending on the parallax of the display stimulus. This mismatch between accommodation and vergence distances can be uncomfortable. Percival's zone of comfort describes the range of comfortable accommodation for a fixed convergence distance, or, as is more applicable to stereoscopic displays, the area of comfortable vergence for a fixed accommodation distance. Percival's zone of comfort is defined as the

middle third of the zone of clear and single binocular vision (ZCSBV), which is the region where stereoscopically viewed objects are still fusible, and thus relative depth precept is still valid. The ZCSBV, while usually empirically defined on a person-by-person basis, is traditionally defined as one diopter on either side of the fixation point, as previously illustrated in Figure 3.15(b).

Accordingly, we use the limits of comfortable convergence as defined by the distance of the display surface and Percival's zone of comfort. Figure 6.11 illustrates how Percival's zone can be applied to the stereoscopic viewing space and a constraint derived. The ZCSBV is defined in green surrounding a fixed focal distance, or in our case the screen plane. Percival's comfort zone, as shown in yellow, is the middle third of the ZCSBV. Recall that the ZCSBV and Percival's zone of comfort are defined in diopter units, which are units of inverse distance. Thus, the arrangement of the ZCSBV and Percival's zone are visualized with an inverse relationship to distance in reference to the viewer.

Notice how depending on the parallax of the stereo precept, the convergence distance and horopter can be inside Percival's zone of comfort, inside the zone of clear and single binocular vision (as in Figure 6.11), or outside the ZCSBV. Stereo precepts should ideally be limited to Percival's zone of comfort, or at least to the ZCSBV. Because of the inverse relationship between diopters and distance, any non-diverging parallax value defines a stereo precept at least within the ZCSBV on the far side of a display surface for displays more than one meter from the viewer. This allows us to use excessive divergence as a far distance constraint, and use Percival's comfort zone for a near distance constraint.

To compute the near constraint, we first find the distance of the near boundary of Percival's comfort zone, or one-third of a diopter nearer to the viewer

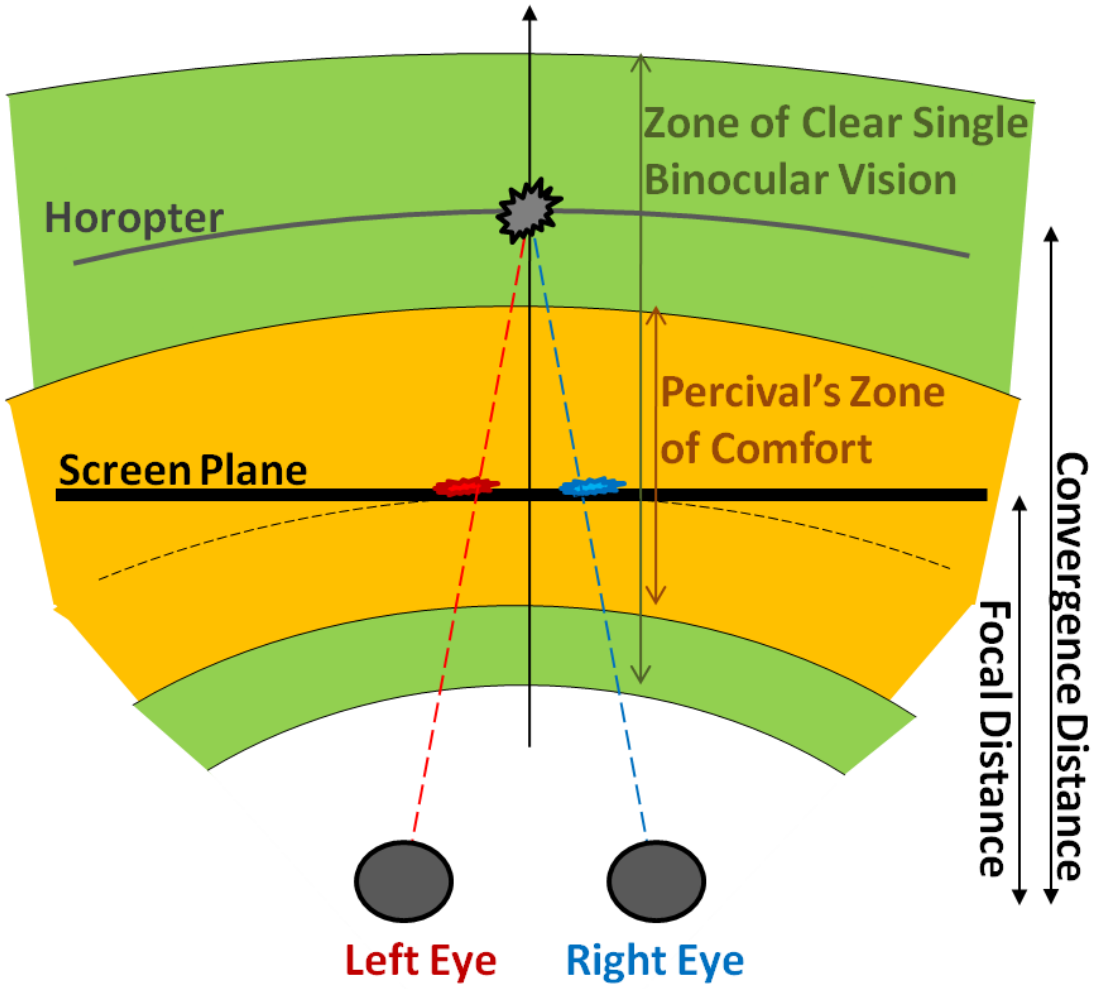


Figure 6.11: When viewing stereoscopic content, the convergence distance frequently mismatches the focal or accommodation distance, a situation that does not arise in natural viewing. Depending on the amount of screen parallax, the convergence point (and subsequently the horopter) could be outside Percival's Zone of Comfort (defined as the middle 1/3 of the ZCSBV in diopter space) or even the Zone of Clear Single Binocular Vision, as determined by the focal distance. The more extreme the situation, the more stressful it is on the viewer's visual system and painful it can be for the viewer.

than the display surface.

$$\left(\frac{1}{Z_v} + \frac{1}{3}\right)^{-1} \tag{6.11}$$

Then, the minimum allowed parallax can be computed and reduced based on the desired distance for the nearest precept and Equation 4.10 which relates parallax to perceived distance.

$$p_{min} = \frac{eZ_v}{3} \tag{6.12}$$

As before, by combining the minimum allowed parallax value (p_{min}) with the distance of the nearest object to the camera (d_{min}) and applying them to the equations relating parallax and distance to the intrinsic stereo camera parameters (Equations 4.5 and 4.4), we can arrive at the definition of our second perceptual constraint.

$$h > \frac{f \cdot t}{d_{min}} + \frac{p_{min} \cdot W_c}{W_s} \tag{6.13}$$

A few final assumed and automatic constraints are that the camera separation is greater than zero, and that the convergence angle or horizontal image translation are non-zero.

$$t > 0 \tag{6.14}$$

$$h \geq 0 \tag{6.15}$$

$$\beta \geq 0 \tag{6.16}$$

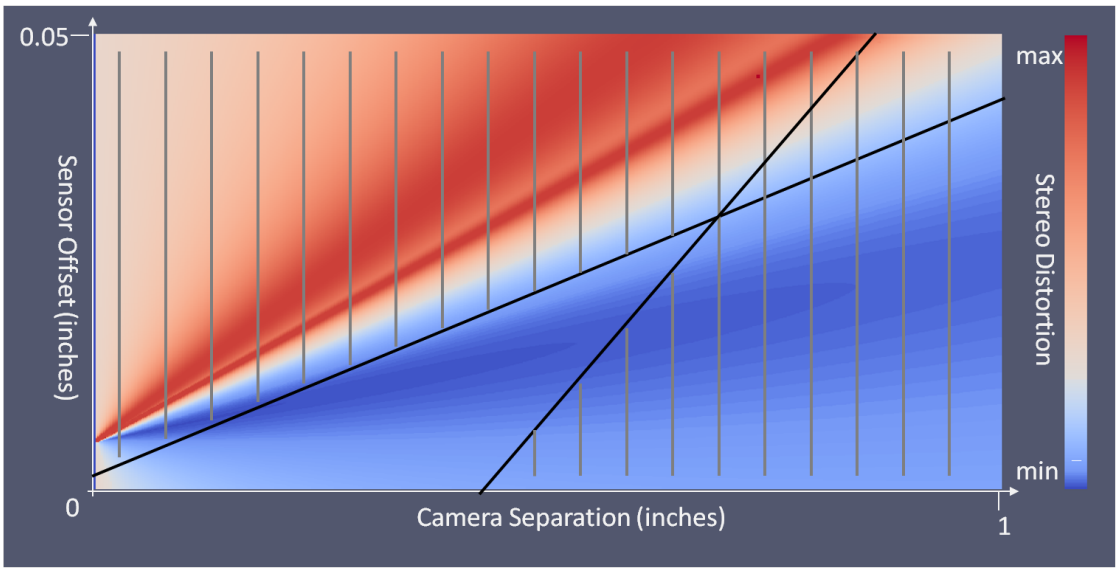


Figure 6.12: Visualization of the perceptual constraints on the distortion metric surface. The black boundaries are the mathematic constraints on the camera separation-sensor offset parameter subspace, derived from the perceptual constraints of excessive divergence and accommodation-vergence mismatch discomfort. The grey striped regions indicate the invalid space. The shallower boundary is the excessive divergence constraint, as defined by Equation 6.10, while the steeper boundary is the accommodation-vergence conflict constraint, as defined by Equation 6.13.

Figure 6.12 illustrates the constraints on the distortion metric surface. These constraints correspond to non-imaginary non-2D precepts, and indicate a valid subspace for parameter variation. Our tool provides the option to constrain the parameter space to within these perceptual boundaries. Appendix A discusses the potential application and implications of computationally minimizing the distortion metric via constrained optimization with respect to a selected subset of stereo parameters for use in automating the determination of stereoscopic parameters in animation pipelines.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

We hope to have shown how the study and understanding of depth perception can guide the creation of better stereo 3D content. We have developed an interactive visualization and interactive pipeline toolset for animation artists that allows them to view and manipulate the impact of the high-dimensional stereo parameter space on the resultant stereo precept. Through expansion of the classically accepted geometric stereoscopic transformation, we have shown how the updated model allows artists to visualize the impact of information traditionally abstracted away. By simultaneously viewing both the capture and display configurations together and by defining the "ghosted" geometry, we have enabled direct comparisons between the original and transformed spaces, allowing easy visualization of any distortions in shape and volume inherent to the expanded stereoscopic transformation. The quantization of stereo quality through the development of a stereo distortion metric and constraints based on perceptual limitations provide a consistent and meaningful way for artists to visualize how variations within a valid parameter subspace affect the quality of the resultant stereo precept.

By promoting meaningful interaction with and a visualization of the complete and comprehensive stereo pathway, we better enable artists to make fully informed creative and artistic stereo decisions. This is particularly true early on in the preliminary production of a film, where the visualization tool can highlight distortions, enabling artists to reduce them and avoid costly errors and preventable mistakes. Hopefully, interaction with this tool and the stereo content it manipulates will allow for greater integration of a stereo mindset throughout

all aspects of a film's production, and promote the development of new and creative stereo techniques and terminology.

While our work in the expansion of the stereoscopic parameter space and resulting transformation have increased the comprehensiveness, effectiveness, and applicability of a distortion visualization tool such as ours, it remains incomplete. There will constantly be new cinematic techniques developed and stereo control paradigms conceived that can be incorporated into the system, allowing for control of the parameters and visualization of the resulting transformation. Many techniques, such as depth of field, racking of focus, camera shake, temporal coherency in camera or object motion, and others, are not currently considered by the visualization tool or the distortion metric.

Having discussed in detail the need, use, implementation, and possible errors of floating windows and multirigged camera setups, one apparent extension would be the automation of the rectification of the errors. Floating window automation could be setup such that it removes all manifestations of window violations. Automated multirig error correction could adjust the near or far pixel shift values of a violating rig so as to remove depth incongruities.

Our quantization of stereo quality takes the form of one measure of distortion, namely shape and volume mismatches. It utilizes many culling, grouping, and weighting techniques to ensure fair and even representation of any distortions visible to the viewer. Other weighting methodologies could be considered based on perceptual, artistic, or scientific motivations to emphasize certain aspects of the scene configuration or other desired elements. Some possibilities include the size or subtended angle of an object, distance from the convergence plane, some measure of object complexity or importance, or any other compo-

nent of saliency.

Alternatively, a completely different approach to defining stereo quality could be taken. Other aspects of distortion could be emphasized, such as the amount of non-linearity in the depth transform, or a non-negative metric could be defined, such as the size of the stereoscopic depth window or the amount of the comfort zone utilized.

As briefly shown in Appendix A, touching on the potential for automating the determination of the intrinsic stereoscopic parameters through optimization of the distortion metric has opened the doors for possible exploration in this area. A plethora of other optimization metrics could be used, other parameters could be optimized, enhanced pipeline integration could be implemented, and so on.

The implications of such automation techniques preliminarily indicated a reduction in artistic control. Accordingly, new ways to integrate stereo-minded creative decision making into a movie production should be considered. For example, artistic input into the automization process could be incorporated, such as a floating target for what optimal stereo quality is, based on the preference of the artist. These inputs could be feed by a depth script that would be defined by an artist, and could indicate periods of accentuation or restraint of the stereo depth based on the emotional flow, amount of action, character personality, or some other determining factor.

Taking multirigging to its natural extreme, it is conceivable that one could specify separate stereo parameters for each object or even vertex individually. This could result in interesting and abstract precepts of the stereo space.

Through some intelligent interface, one could allow for complete control over the non-linearity of the stereo transformation. For example, one might desire to artistically massage the stereo space and define the stereo transformation by physically sketching the z-depth transformation curves, and the parameters could be reverse computed to fit.

Finally, while many of the discussions and techniques uniquely apply to animated filmmaking where access to the exact geometry in a digital format is taken advantage of, there is the possibility for application to live action stereo capture. Through the use of laser range finders, reverse stereo computations via camera arrays, manual input, or some other methodology of digitally acquiring information about the scene content, many of the techniques discussed could apply to live action. Control feedback loops could be established, creating an automation methodology for live broadcasts, such as sporting events, without the need for constant monitoring or hand holding and manipulation.

Interfacing with a working pipeline will certainly increase the value of our tool. With access to a library of completed filmic content, it facilitates the study and understanding of pre-compiled artistically determined parameter configurations, layouts, and animations. For novice artists, the tool can be used as a "sandbox" for exploration, stereo space comprehension, and training. For production artists, the tool can be used at any stage of the production pipeline as a way to verify or get visual feedback on stereo use and visualize distortions specific to the parameter configuration.

Perhaps, most importantly, our visualization tool has great potential for use in the earlier and concept stages of animation production, as it can help emphasize to high-level creative decision makers the impact of stereoscopic choices

on the resulting image. By visualizing distortions in the early stages of filmmaking, artists can prevent the need for costly re-rendering due to situations where the distortions were not otherwise apparent. The tool also allows for quick iteration through multiple stereo configurations in the concept stage of development, which can help set the overall stereo direction and tone of a film and avoid relegating stereo to a 3D-ization post process.

Integration of a stereoscopic mindset early in the creative process allows the stereo elements of a film to be a guiding factor in the creative development of content. Through continued emphasis on interacting with and visualizing the effect of variations within the complete high-dimensional stereoscopic parameter space, artists can understand the impact that even traditionally non-stereo specific parameters, such as focal length, staging, or editing, have on the resultant stereo precept. This increased association between the basic physical parameters and the resulting perceived image enables artists to incorporate stereoscopic understanding into all aspects of filmmaking. Through this process, artists can create and develop a new grammar and visual vocabulary specific to the needs and uses of stereoscopic filmmaking.

Finally, we should not lose sight of the fact that filmmaking is an art. Science and technology can act in support of its creation and study, but appreciation is subjective and certain aspects of quality are ultimately up to the preferences and opinions of the viewer. Thus, we should be attempting to create a space to be viewed by the visual systems of the audience within the limits of perception, instead of replicating and replacing their function. In other words, filmmaking is the craft, and stereoscopy is the tool. There is no one right or wrong answer, and it is up to the artist to decide how best to wield that tool to create something of

interest and value for their particular purpose and intent and to be appreciated and consumed by the audience.

APPENDIX A

OPTIMIZATION AND AUTOMATION

Since our metric for stereoscopic distortion is mathematically defined, the stereo parameter space can be numerically optimized in an attempt to minimize distortions with respect to a defined set of parameters. With such an optimization technique, our visualization tool can be further expanded to automatically configure stereo parameters for large sets of preconfigured geometry and camera data, in effect automating the minimization of stereoscopic distortions with respect to a defined set of parameters. To demonstrate the capabilities and effectiveness of our visualization tool, we minimize the distortion metric through a numeric optimization technique with respect to the intrinsic stereoscopic parameters, and constrained to within the limits of human perception. Finally, we preliminarily test our implementation of a distortion metric and its optimization by implementing it in a working animation production pipeline on existing content and briefly comparing the results to prior artistically determined parameters.

A.1 Optimization

When considering an optimization algorithm to minimize the distortion metric, certain considerations come into play. First, the determination of the stereoscopic distortion metric as defined is complex. The non-linear stereo transformation must first be applied and then the metric must be computed in a two stage process, first for object level near-affine invariance, and then iterated over all non-culled vertices which must also be computed. Due to the complexity of the method, no directly computable gradient or Hessian are available, and must

be estimated via multiple function evaluations if required by the optimization algorithm. It also means that the number of function evaluations to recalculate the distortion metric should be minimized.

Second, while the illustrated optimization surface, as shown in Figure 6.2, seems to be smooth and well behaved, the illustration is actually heavily scaled to allow for a reasonable visualization of the surface. In actuality, the valid region within the constraints is a deep narrow valley. This makes estimations of the gradient difficult, as it varies greatly and changes sign within a very narrow area and even adaptive non-uniform step sizes frequently take traditional optimizers well outside any reasonable converging region. Additionally, the estimation of the gradient involves costly additional distortion metric function evaluations.

Accordingly, great care must be taken in choosing, implementing, and configuring a numeric optimization technique. One optimization technique which addresses all of these considerations is the complex method, a constrained version of the Nelder-Mead direct sequential search technique [NM65]. To provide a first proof-of-concept, we tentatively implemented the complex method of optimization without significant tuning, to simply and quickly gauge the potential application of automated stereo parameter determination. The complex method algorithm and implementation details are omitted here, as they are not critical to the discussion of the implications and relevancy of the minimum, assuming it is correctly determined. Other optimization techniques may be appropriate for our purposes, but currently have not yet been explored.

Having defined a metric for the quality of the stereoscopic precept and implemented one methodology for optimizing this metric with respect to a subset

of stereoscopic parameters, we can explore some of the practical implications of these techniques. To test the effectiveness of our stereoscopic distortion metric optimization, we automated the minimization algorithm to run on a set of shot data within the production pipeline of DreamWorks Animation. The tool iterated over select shots from a particular sequence of the animated film Megamind, one of the studio's 2010 feature releases. For each shot, our visualization tool automatically loaded the scene geometry and non-stereo camera parameters for every 12th frame, or every half second of the movie. The optimization was then automatically run to determine the optimal camera separation and convergence distance for the scene according to the minimum perceived shape distortion as defined by our quality-of-stereo metric and the configuration state of the scene.

This technique attempts to simulate the effect of using the results of the stereoscopic automation as defined by our tool as a first-pass stereo determination methodology, after artists have defined character and object animation and camera movement, but before stereo parameters have been otherwise defined. Note that this technique does not directly support the ideology frequently toted in this thesis of early stage stereo mindset integration, instead allowing stereoscopic thinking to act as a post process. However, exploration of this method of automation is partially justified by the cursory validation of the importance of the distortion metric and perceptual constraints. By comparing the resultant optimized values with those as artistically determined by DreamWorks Animation artists through creative and artistic means, both based on the same source material, we can confirm the significance of the distortion metric.

A.2 Analysis of Automation

As a brief addendum to the major works of the thesis, this analysis is not a comprehensive numerical analysis of the correctness of the optimization technique. Instead, it is a scope-limited first attempt to validate the potential utility of the defined distortion metric within the visualization tool environment. Accordingly, without statistically complete or significant data due to content restriction and time limitations, we only present a graphical representation of the results to highlight and compare magnitudes and trends within scene variations.

Figure A.1 compares the results of our automated simulations with the artistically determined stereoscopic parameters for tested shots in the feature film Megamind from DreamWorks Animation. The top chart shows the camera separation while the lower chart shows the convergence distance for the tested shots, as indicated across the top and separated by dividers. The red trendline shows the results of our distortion metric optimization automation, while the blue trendline shows the creative determination of the DreamWorks Animation artists. Shots and portions of shots with a grey background indicate optimizations that did not converge within a specified maximum number of iterations, and suggests stereo parameters that are undetermined given the optimization coefficients.

There is a reasonable amount of noticeable correlation between the two data sets, both in magnitude and in trending directions of the stereoscopic parameters. In some shots, it appears as if the parameters match well between the automated and artistically determined values suggesting success in optimization and of the distortion metric. In other shots, however, the automation of the

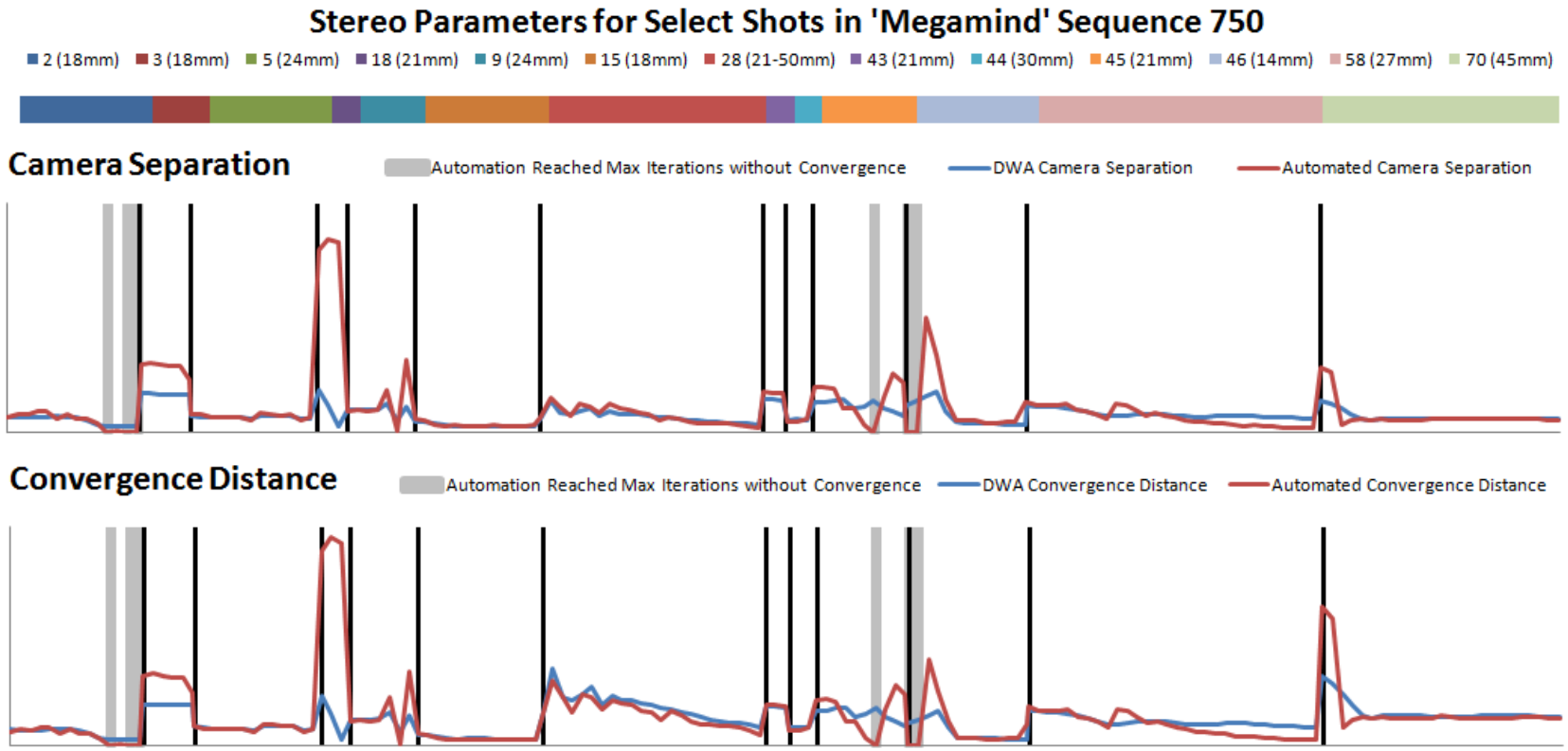


Figure A.1: Comparison of stereo camera parameters between DreamWorks Animation's artistically determined values and the automatically computed values based on minimizing the shape distortion metric. The top chart shows the camera separation while the lower chart shows the convergence distance for the tested shots, as indicated across the top and separated by dividers. The red trendline shows the results of our distortion metric optimization automation, while the blue trendline shows the creative determination of the DreamWorks Animation artists. Shots and portions of shots with a grey background indicate optimizations that did not converge within a specified maximum number of iterations, and suggests stereo parameters that are undetermined given the optimization coefficients.

distortion metric and optimization failed, and there appears to be little correlation between the two sets of values.

Through discussions with the artists involved, the cases where the optimization failed can be categorized into one of two groups. For some shots, artists took creative liberties to intentionally expand or constrict the depth of a shot, varying the stereoscopic parameters from a theoretical artistic norm. This was the case for some extremely long or extremely short lenses, as well as in shots where the use of dramatic stereo effect was desired.

For other shots, discrepancies can be attributed to the distortion metric and optimization process. For example, in some instances mesh geometry is not an ideal representation of the scene content for the purposes of vertex-by-vertex distortion computations or the weighting method does not accurately represent the salient intent of the artist. Additionally, there are several instances where the quick entry or departure of a near object from the field of view over time or the sudden movement of an object resulted in the automation providing coarse and jerky stereo parameter values. Artists traditionally do not use drastic and sudden variations in stereo parameters. The implementation of curve smoothing and temporal coherence methods could potentially ameliorate these unwanted irregularities.

Despite the apparent shortcomings of this preliminary attempt at stereo automation, the potential use for such a tool was highlighted by the artists. It was noted that a large percentage of the simple non-corner case shots emerging from the distortion metric optimizer were production worthy "as is". This suggests the potential utility of the tool as a first-pass baseline method for determining stereoscopic parameters for straightforward uncomplicated shots. This would

allow artists to focus more time on the complicated and more elaborate shots and for greater artistic expression, instead of setting default stereo settings.

While there is not enough data to make statistically supported claims about accuracy or importance, it may suggest, and certainly warrants further exploration into, correlations between what is artistically determined to be aesthetically correct and good, and what is computed to be perceptually appropriate and distortionless in shape and volume.

Finally, it should be reiterated that this technique of optimizing the stereo distortions to the intrinsic stereoscopic parameters limits the scope and impact of stereoscopic thinking. Instead of allowing a stereo oriented mindset to proliferate into other aspects of filmmaking, it relegates it as a post process mechanical appendage to the content creation process. Methods of artistic input and higher level creative control are needed to make any stereo optimization technique artistically effective and impactful. The author intends to continue testing and exploring these venues, professionally and in possible future research endeavors.

BIBLIOGRAPHY

[ABM+97] M. Agrawala, A.C. Beers, I. McDowall, B. Fr
"ohlich, M. Bolas, and P. Hanrahan. The two-user responsive workbench: support for collaboration through individual views of a shared space. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 327–332. ACM Press/Addison-Wesley Publishing Co., 1997.

[ABVE01] W.J. Adams, M.S. Banks, and R. Van Ee. Adaptation to three-dimensional distortions in human vision. *nature neuroscience*, 4(11):1063–1064, 2001.

[Akk92] R.A. Akka. Automatic software control of display parameters for stereoscopic graphics images. In *Proceedings of SPIE*, volume 1669, page 31, 1992.

[AWGB04] K. Akeley, S.J. Watt, A.R. Girshick, and M.S. Banks. A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics (TOG)*, 23(3):804–813, 2004.

[BBvEC99] B.T. Backus, M.S. Banks, R. van Ee, and J.A. Crowell. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39(6):1143–1170, 1999.

[BGL04] M.S. Banks, S. Gepshtein, and M.S. Landy. Why is spatial stereoresolution so low? *Journal of Neuroscience*, 24(9):2077, 2004.

[Box65] MJ Box. A new method of constrained optimization and a comparison with other methods. *The Computer Journal*, 8(1):42, 1965.

[BRVG05] M.S. Banks, H.F. Rose, D. Vishwanath, and A.R. Girshick. Where should you sit to watch a movie? In *Proc. SPIE*, volume 5666, pages 316–325, 2005.

[CES88] H. Collewijn, C.J. Erkelens, and RM Steinman. Binocular coordination of human horizontal saccadic eye movements. *The Journal of physiology*, 404(1):157, 1988.

[CV95] J.E. Cutting and P.M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of dif-

ferent information about depth. *Perception of space and motion*, 5:69–117, 1995.

[DF93] D.B. Diner and D.H. Fender. *Human engineering in stereoscopic viewing devices*. Plenum Press New York, NY, USA, 1993.

[Din91] D. Diner. Danger of collisions for tele-operated navigation due to erroneous perceived depth accelerations in 3-D television. In *Annual Meeting of the American Nuclear Society*, 1991.

[Din02] D.B. Diner. A new definition of orthostereopsis for 3D television. In *Systems, Man, and Cybernetics, 1991.'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on*, pages 1053–1058. IEEE, 2002.

[DM96] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In *PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING*, pages 123–134. Citeseer, 1996.

[DR96] H. De Ridder. Naturalness and image quality: saturation and lightness variation in color images of natural scenes. *Journal of Imaging Science and Technology*, 40:487–493, 1996.

[dRBF95] H. de Ridder, F.J.J. Blommaert, and E.A. Fedorovskaya. Naturalness and image quality: Chroma and hue variation in color images of natural scenes. In *Proceedings of SPIE*, volume 2411, page 51, 1995.

[FAPI99] J. Freeman, SE Avons, D.E. Pearson, and W.A. IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators & Virtual Environments*, 8(1):1–13, 1999.

[GPS94] V.S. Grinberg, G.W. Podnar, and M.W. Siegel. Geometry of binocular imaging. *Stereoscopic Displays and Applications V*, pages 56–65, 1994.

[HB] R.T. Held and M.S. Banks. Visualizing Misperceptions in Stereoscopic Displays.

[HB96] C. Hendrix and W. Barfield. Presence within virtual environments

as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, 5(3):274–289, 1996.

[HB08] 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.

[HGAB08] 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), 2008.

[Hod91] L.F. Hodges. Basic principles of stereographic software development. In *Proceedings of SPIE*, volume 1457, page 9, 1991.

[HPC+02] J. Hsu, Z. Pizlo, D.M. Chelberg, C.F. Babbs, and E.J. Delp. Issues in the design of studies to test the effectiveness of stereo imaging. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 26(6):810–819, 2002.

[HWLB04] J.M. Hillis, S.J. Watt, M.S. Landy, and M.S. Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12), 2004.

[IdRH+98a] W. IJsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman. Perceived depth and the feeling of presence in 3DTV. *Displays*, 18(4):207–214, 1998.

[IdRH98b] W.A. IJsselsteijn, H. de Ridder, and R. Hamberg. Perceptual factors in stereoscopic displays: the effect of stereoscopic filming parameters on perceived quality and reported eyestrain. In *Proceedings of SPIE*, volume 3299, page 282, 1998.

[IDRV02] WA IJsselsteijn, H. De Ridder, and J. Vliegen. Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(2):225–233, 2002.

[IG06] H. Ishikawa and D. Geiger. Illusory volumes in human stereo perception. *Vision research*, 46(1-2):171–178, 2006.

[IO97] T. Inoue and H. Ohzu. Accommodative responses to stereoscopic three-dimensional display. *Applied optics*, 36(19):4509–4515, 1997.

[JLHE01] G. Jones, D. Lee, N. Holliman, and D. Ezra. Controlling perceived depth in stereoscopic images. In *Stereoscopic Displays and Virtual Reality Systems VIII, Proceedings of SPIE*, volume 4297, pages 42–53. Citeseer, 2001.

[KT04] F.L. Kooi and A. Toet. Visual comfort of binocular and 3D displays. *Displays*, 25(2-3):99–108, 2004.

[Kus92] H. Kusaka. Apparent depth and size of stereoscopically viewed images. In *Proceedings of SPIE*, volume 1666, page 476, 1992.

[Kut02] R. Kutka. Reconstruction of correct 3-D perception on screens viewed at different distances. *Communications, IEEE Transactions on*, 42(1):29–33, 2002.

[LAM91] L. Lipton, R. Akka, and L. Meyer. *The CrystalEyes Handbook*. StereoGraphics Corp., 1991.

[LIFH09] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53:030201, 2009.

[LIH07] M. Lambooij, W. IJsselsteijn, and I. Heynderickx. Stereoscopic displays and visual comfort: a review. *Einhoven University of Technology Einhoven Univeristy of Technology Philips Research, SPIE Newsroom2 April*, 2007.

[Lip82] L. Lipton. *Foundations of the stereoscopic cinema: a study in depth*, volume 259. Van Nostrand Reinhold, 1982.

[Liv96] M.S. Livingstone. Differences between stereopsis, interocular correlation and binocularity. *Vision research*, 36(8):1127–1140, 1996.

[LM66] H. Leibowitz and D. Moore. Role of changes in accommodation and convergence in the perception of size. *JOSA*, 56(8):1120–1123, 1966.

[Lun49] R.K. Luneburg. Mathematical analysis of binocular vision. *Bull.*

Amer. Math. Soc. 55 (1949), 866-867. DOI: 10.1090/S0002-9904-1949-09284-4 PII: S, 2(9904):09284–4, 1949.

[LYJ+05] S. Li, M. Yu, G. Jiang, T.Y. Choi, and Y.D. Kim. Approaches to H. 264-based stereoscopic video coding. In *Image and Graphics, 2004. Proceedings. Third International Conference on*, pages 365–368. IEEE, 2005.

[Mac54] DL MacAdam. Stereoscopic perceptions of size, shape, distance and direction. *SMPTE Journal*, 62:271–93, 1954.

[Men09] B. Mendiburu. 3D Movie Making, 2009.

[MHE+06] K. Masaoka, A. Hanazato, M. Emoto, H. Yamanoue, Y. Nojiri, and F. Okano. Spatial distortion prediction system for stereoscopic images. *Journal of Electronic Imaging*, 15:013002, 2006.

[MP79] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 204(1156):301–328, 1979.

[NM65] J.A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308, 1965.

[NYHO03] Y. Nojiri, H. Yamanoue, A. Hanazato, and F. Okano. Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV. In *Proceedings of SPIE*, volume 5006, page 195, 2003.

[Ogl40] K.N. Ogle. The induced size effect. *JOSA*, 30(4):145–151, 1940.

[Ogl50] K.N. Ogle. Researches in binocular vision. 1950.

[Pas91] S. Pastoor. 3D-television: A survey of recent research results on subjective requirements. *Signal processing: Image communication*, 4(1):21–32, 1991.

[Pas95] S. Pastoor. Human factors of 3-d imaging: results of recent research at heinrich-hertz-institut berlin. In *Proc. 2nd Int. Display Workshop*, pages 69–72, 1995.

[PG03] M.J. Pianta and B.J. Gillam. Paired and unpaired features can be equally effective in human depth perception. *Vision research*, 43(1):1–6, 2003.

[RR92] L.B. Rosenberg and STANFORD UNIV CA CENTER FOR DESIGN RESEARCH. The Effect of Interocular Distance upon Depth Perception when Using Stereoscopic Displays to Perform Work within Virtual and Telepresent Environments. 1992.

[SB85] C.M. Schor and D.R. Badcock. A comparison of stereo and vernier acuity within spatial channels as a function of distance from fixation. *Vision Research*, 25(8):1113–1119, 1985.

[SB06] R. Sekuler and R. Blake. Perception (5th edn), 2006.

[SGC+02] J.Y. Son, Y. Gruts, J. Chun, Y.J. Choi, J.E. Bahn, and V.I. Bobrinev. Distortion analysis in stereoscopic images. *Optical Engineering*, 41:680, 2002.

[SHF+08] K.M. Schreiber, J.M. Hillis, H.R. Filippini, C.M. Schor, and M.S. Banks. The surface of the empirical horopter. *Journal of Vision*, 8(3), 2008.

[SS53] R. Spottiswoode and N. Spottiswoode. *The theory of stereoscopic transmission & its application to the motion picture*. University of California Press, 1953.

[SSS52] R. Spottiswoode, NL Spottiswoode, and C. Smith. Basic principles of the three-dimensional film. *SMPTE J*, 59:249–286, 1952.

[STA99] M. Siegel, Y. Tobinaga, and T. Akiya. Kinder gentler stereo. In *SPIE Proc. Stereoscopic Displays and Applications X*, pages 18–27. Citeseer, 1999.

[STK+97] H. Sakata, M. Taira, M. Kusunoki, A. Murata, and Y. Tanaka. The tins lecture the parietal association cortex in depth perception and visual control of hand action. *Trends in Neurosciences*, 20(8):350–357, 1997.

[STM99] L. Stelmach, W.J. Tam, and D. Meegan. Perceptual basis of stereoscopic video. In *Proceedings of the SPIE*, volume 3639, pages 260–265, 1999.

[Stu34] NA Stutterheim. The divergence of the primary position of the eyes. *British Journal of Ophthalmology*, 18(5):256, 1934.

[SW83] C.M. Schor and I. Wood. Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research*, 23(12):1649–1654, 1983.

[Tho87] GA Thomas. Television motion measurement for DATV and other applications. *NASA STI/Recon Technical Report N*, 88:13496, 1987.

[TSC98] W.J. Tam, L.B. Stelmach, and P.J. Corriveau. Psychovisual aspects of viewing stereoscopic video sequences. In *Proceedings of SPIE*, volume 3295, page 226, 1998.

[UBH03] M.T. Ukwade, H.E. Bedell, and R.S. Harwerth. Stereopsis is perturbed by vergence error. *Vision research*, 43(2):181–193, 2003.

[UH08] K. Ukai and P.A. Howarth. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, 29(2):106–116, 2008.

[Uka06] K. Ukai. Human factors for stereoscopic images. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 1697–1700. IEEE, 2006.

[VH76] C. Von Hofsten. The role of convergence in visual space perception. *Vision Research*, 16(2):193–198, 1976.

[VYS08] A. Vetro, S. Yea, and A. Smolic. Towards a 3D video format for auto-stereoscopic displays. *Proceedings of the SPIE: Applications of Digital Image Processing XXXI, San Diego, CA, USA*, 2008.

[Wal05] G. Waloszek. *Vision and Visual Disabilities - An Introduction*. SAP User Experience, June 2005.

[WDC+05] A.E. Welchman, A. Deubelius, V. Conrad, H.H. B
"ulthoff, and Z. Kourtzi. 3D shape perception from combined depth cues in human visual cortex. *Nature neuroscience*, 8(6):820–827, 2005.

[WDK95] A. Woods, T. Docherty, and R. Koch. Image distortions in stereo-

scopic video systems. *Stereoscopic displays and applications IV*, pages 36–48, 1995.

[WDK96] A.J. Woods, T. Docherty, and R. Koch. 3D video standards conversion. In *Proceedings of SPIE*, volume 2653, page 210, 1996.

[WF71] H. Wallach and L. Floor. The use of size matching to demonstrate the effectiveness of accommodation and convergence as cues for distance. *Attention, Perception, & Psychophysics*, 10(6):423–428, 1971.

[Whe52a] C. Wheatstone. LXX. The bakerian lecture.Contributions to the physiology of vision.Part the second. On some remarkable, and hitherto unobserved, phænomena of binocular vision (continued). *Philosophical Magazine Series 4*, 3(21):504–523, 1852.

[Whe52b] C. Wheatstone. XXXVI. Contributions to the physiology of vision.Part the First. On the some remarkable, and hitherto unobserved, phænomena of binocular vision. *Philosophical Magazine Series 4*, 3(18):241–267, 1852.

[WMW02] JP Wann and M. Mon-Williams. Measurement of visual aftereffects following virtual environment exposure. 2002.

[WOM99] A.J. Woods, D. Offszanka, and G. Martin. A PC-based stereoscopic video walkthrough. *Proc. SPIE 1ggg*, 3639:306–312, 1999.

[Woo00] A. Woods. Stereoscopic Presentations–Taking the Difficulty out of 3D. In *6th International Workshop on 3D Imaging Media Technology, Seoul, South Korea*. Citeseer, 2000.

[Woo01] A.J. Woods. Optimal usage of LCD projectors for polarised stereoscopic projection. In *Stereoscopic Displays and Virtual Reality Systems VIII, Proceedings of SPIE*, volume 4297. Citeseer, 2001.

[Woo10] AJ Woods. Understanding Crosstalk in Stereoscopic Displays. 2010.

[WR07] A.J. Woods and T. Rourke. The compatibility of consumer DLP projectors with time-sequential stereoscopic 3D visualisation. *Stereoscopic Displays and Applications XVIII, published in Stereoscopic Displays and Virtual Reality Systems XIV, Proceedings of IS&T/SPIE Electronic Imaging*, 6490:29–31, 2007.

[WT02] A.J. Woods and S.S.L. Tan. Characterising sources of ghosting in time-sequential stereoscopic video displays. *Stereoscopic Displays and Applications XIII, published in Stereoscopic Displays and Virtual Reality Systems IX, Proceedings of SPIE*, 4660:21–23, 2002.

[Yam97] H. Yamanoue. The relation between size distortion and shooting conditions for stereoscopic images. *SMPTE journal*, 106(4):225–232, 1997.

[Yan99] S. Yano. Perception of Sensation of Reality Based on Human Information Processing. *Sanjigen Gazo Konfarensu Koen Ronbunshu*, 1999:199–206, 1999.

[YEM04] S. Yano, M. Emoto, and T. Mitsuhashi. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays*, 25(4):141–150, 2004.

[YLCA08] Y. Yun Lee, T. Chen, and T.L. Alvarez. Quantitative assessment of divergence eye movements. *Journal of Vision*, 8(12), 2008.

[YS90] Y.Y. Yeh and L.D. Silverstein. Limits of fusion and depth judgment in stereoscopic color displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 32(1):45–60, 1990.

[YS02] J.N. Yang and S.K. Shevell. Stereo disparity improves color constancy. *Vision Research*, 42(16):1979–1989, 2002.

[ZBS03] Z.L. Zhang, E.M. Berends, and C.M. Schor. Thresholds for stereoslant discrimination between spatially separated targets are influenced mainly by visual and memory factors but not oculomotor instability. *Journal of Vision*, 3(11), 2003.