Diss. ETH No. 23029

Perceptual Enhancements for Novel Displays

A dissertation submitted to
ETH Zurich

for the Degree of
Doctor of Sciences of ETH Zurich
(Dr. sc. ETH Zurich)

presented by
Alexandre Chapiro
MSc in Mathematics, IMPA, Rio de Janeiro, Brazil.
born on 17.02.1990
citizen of Brazil and Russia.

accepted on the recommendation of
Prof. Dr. Markus Gross, examiner
Dr. Aljoša Smolić, co-examiner
Prof. Dr. Carol O'Sullivan, co-examiner

2015

Abstract

Novel display technologies give us the chance to enjoy content with additional degrees of realism. These advantages do not always come easily, however. Changing the technologies used in our display systems creates new challenges for content creation, post-production and distribution. In this thesis, we examine two special cases of novel display: stereoscopic 3D and high dynamic range. In each case, the richness of the natural world needs to be compressed for effective reproduction. Our goal is to understand the requirements for optimal presentation on these displays and leverage this knowledge to generate practical systems that can be used to operate on high-quality content in professional film production pipelines. Special emphasis is placed on perceptual techniques both as a means of modeling the response of the human visual system to stimuli presented by novel display technologies as well as being employed to validate that the goals of the designed methods are reached successfully.

We begin by focusing our attention on 3D displays. Existing 3D technology is limited in the amount of depth that can be shown without discomfort or visual artifacts, but when addressed this limitation can produce an unnatural sensation of flatness. Three main contributions to stereoscopic display are presented in this thesis. First, perceptual measurements of the pervasive cardboarding artifact are performed. The perceptual thresholds found in this step can be useful for content creators and computational methods that seek to avoid cardboarding. Second, a stereo-to-multiview conversion system is presented with the aim of producing content for autostereo displays which utilizes the severely limited depth capabilities of these screens in an optimal fashion. Lastly, we present a novel method that enhances the 3D sensation of viewers by inducing stereo from binocular variations in shading.

Next, our work touches on high dynamic range display. We identify two important challenges facing the expected launch of various high dynamic range screens into the consumer market: that of content creation and distribution. In order to address these difficulties we propose a novel approach to color-grading where a continuous dynamic range video is generated simultaneously in the course of the standard post-production pipeline. Finally, we demonstrate an efficient and perceptually transparent representation of this information that requires only a fraction of the bandwidth of a traditional video.

Zusammenfassung

Neue Bildschirmtechnologien ermöglichen es heutige Inhalte mit einem hohen Grad an Realitätsnähe zu genießen. Diese Vorteile haben jedoch oft einen hohen Preis. Eine Veränderung der Technologien, die zum Anzeigen verwendet werden, erzeugt ebenfalls neue Herausforderungen für die Erzeugung der Inhalte, Nachbearbeitung und den Vertrieb. Diese Doktorarbeit betrachtet zwei spezielle neue Bildschirmtechnologien: stereoskopische 3D sowie High Dynamic Range (engl.: hoher Dynamik Umfang) Darstellung. In beiden Fällen muss die Vielfalt der Umwelt komprimiert werden, um sie für die Darstellung reproduzieren zu können. Unser Ziel ist es, zu verstehen, was für eine optimale Präsentation mittels solch moderner Bildschirme notwendig ist. Dieses Wissen kann dann dazu verwendet werden, Systeme für die Erzeugung von qualitativ hochwertigen Inhalten zu entwickeln, die in der Filmindustrie zum Einsatz kommen. Ein besonderes Augenmerk liegt dabei auf perzeptuellen Ansätzen. Zum einen, um eine Modellierung der Reaktion des menschlichen Sehsystems auf visuelle Reize durch solch neue Bildschirmtechnologien zu ermöglichen. Zum anderen, kann somit überprüft werden, ob die entwickelte Methodik die gesteckten Ziele auch erreicht.

Zu Beginn werden 3D Bildschirme genauer erläutert. Existierende Technologien zur Darstellung von 3D Inhalten sind beschränkt in dem Umfang an Tiefe der dargestellt werden kann, ohne Unbehagen oder visuelle Artefakte zu erzeugen. Eine naive Herangehensweise diese Einschränkungen zu beheben führt jedoch zu einem unnatürlichen Abflachen von Objekten (Cardboarding). Diese Doktorarbeit präsentiert drei Hauptbeiträge zur stereoskopischen Darstellung. Zunächst wird eine perzeptuelle Messung der allgemeinen Cardboarding-Artefakte vorgenommen. Die Schwellwerte aus diesen Messungen können für das Erzeugen von Inhalten verwendet werden, um Cardboarding zu vermeiden oder zu reduzieren. Als nächstes wird ein System zur Stereo-to-Multiview-Konvertierung vorgestellt. Dieses soll das Erzeugen von Inhalten erleichtern, welche den stark eingeschränkten Tiefenumfang von Autostereo-Bildschirmen optimal ausnutzen. Zuletzt wird eine neue Methode zur Verbesserung der 3D Wahrnehmung durch binokulare Variationen im Shading präsentiert.

Nachfolgend, geht diese Doktorarbeit auf High Dynamic Range Bildschirme ein. Es werden zwei wichtige Herausforderungen identifiziert, welche die erwartete Einführung von High Dynamic Range Bildschirmen auf den Verbrauchermarkt beeinflussen: die Erzeugung sowie der Vertrieb von Inhalten. Vor diesem Hintergrund stellt diese Doktorarbeit einen neuen Ansatz für das Color-Grading vor, der Videos aus einem kontinuierlich Dynamikbereich während der Nachbearbeitung generieren kann. Zuletzt wird eine effiziente und perzeptuell transparente Repräsentation dieser Informationen demonstriert, die lediglich einen Bruchteil der Bandbreite von traditionellen Videos benötigt.

Acknowledgments

First of all, I would like to thank my advisor Prof. Markus Gross for maintaining the incredible work environments at the Computer Graphics Laboratory and Disney Research Zurich and giving me the opportunity to come study at ETHZ. Spending my PhD here was an incredible privilege and allowed me to greatly broaden my horizons. I thank my co-advisor Dr. Aljoša Smolić for the great support throughout my PhD as part of his team at Disney, the thoughtful discussions and for giving me the freedom to pursue my own ideas and learn.

I would like to thank my many co-authors, without whom the work in this thesis would have been impossible. Aljoša Smolić, Markus Gross and Carol O'Sullivan, whom I also thank for taking part in my thesis committee, Miquel Farre, Nikolce Stefanoski, Simon Heinzle, Steven Poulakos, Tunç Aydın, Robert Sumner, Alexander Sorkine-Hornung, Amit Bermano, Benjamin Scheibehenne, Fabio Zund, Jordi Pont-Tuset, Manuel Lang, Marc Junyent, Mattia Ryffel, Matthias Zwicker, Michael Schaffner, Olga Diamanti, Oliver Wang, Pablo Beltran, Pierre Greisen, Pascal Berard, Rafael Huber, Seth Frey, Simone Croci, Stefan Schmid and Wojciech Jarosz - thank you! I thank Maurizio Nitti for generating beautiful illustrations for the works presented here. I thank all the user study participants for their efforts.

During my PhD, I spent many great coffee breaks at the Institute of Visual Computing. I would like to thank all my past and present colleagues at CGL, IGL and CVG for creating this genial atmosphere. My time as a member of Disney Research has been equally pleasant, for which I thank my colleagues. I thank all the members of the Stereo and Displays group for their enthusiasm and for always being ready to help. In particular, I am grateful to Simon Heinzle for providing me with down to earth guidance during the first part of my PhD and Tunç Aydın for his continuous support during the second part, and for selflessly reading my drafts.

I would like to thank Kaan Yucer, Yagiz Aksoy and Andy Cohen for always keeping their doors open in the haze of harsh deadlines; you guys are being really obstinately superb! I thank Kaan for spotting my burdens and for being a great office-mate. I thank Yagiz for his transparent reasoning. I thank Andy for helping me with a bind. I am grateful to Marios Papas, Georgia Koula and Aparna Taneja for their kind friendship and I thank Amit Bermano for improving my pool.

Finally, I am very grateful for all the support from my friends and family. My parents spent a considerable amount of time transferring their passion for knowledge and learning to me during childhood, and are without a doubt my greatest influence for pursuing a PhD. I thank my parents for all their effort to help me succeed and unending encouragement. I thank my brother Grigori and my friend Renata for always being supportive. I thank Olga for being there for me through thick and thin.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Digital display devices have become ubiquitous in our daily lives. Modern surveys show that TV screens, cinema projectors, smartphone displays and computer monitors are used during a significant portion of our daily lives [Ofcom, 2015]. With such prevalence of digital imagery, it is natural that significant research efforts are spent in obtaining novel and improved methods of delivery for this visual content. In this thesis, we place special focus on novel forms of visual display, and the associated challenges for post-production and distribution.

Due to its importance, visual content has been a rapidly evolving medium since its conception. From mechanically moving pictures as presented on a Zoopraxiscope to early projection systems to modern screens, improved realism and expanding capabilities to reproduce the world as it can be naturally perceived by observers is sought. As expected, this tendency did not change, and brings both industrial and academic interest in finding additional improvements to existing forms of display. Many directions of advance can be seen in today's display scene: stereo 3D is used in cinemas and televisions to add the realism of stereoscopic presentation, high dynamic range cameras and displays help mitigate limitations on brightness contrast, high resolution and frame rate display improves the sampling of the presentation, while alternative modes of display such as virtual reality strive to fully immerse users into the content.

These drastic changes in the manner of showing visual content will also necessarily impact the way it is created. In this thesis, the main focus is on a specific aspect of content creation: namely high quality cinematic film production. In this context, great value is given to maximizing visual quality. In addition, content producers value artistic intent as an irreplaceable facet of their work and each portion of a given asset must correspond exactly to the desired outcome. Unlike many traditional applications in computer graphics or vision, a practical system targeted at high quality content creation must be completely failsafe, or at the very least a human operator should be able to manually override our methods at any point. In fact, the movie industry employs a multitude of artistic professionals that perform subjectively motivated tasks which are unlikely to be fully replaced by automated methods in the near future. This means that differently from the standard goal of automation present in most of Computer Science we instead aim to provide artists with the understanding of underlying problems and tools

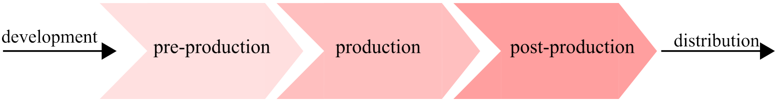


Figure 1.1: *Traditional steps of a film-making pipeline.*

to help in their work, while retaining high levels of human control over the final product.

Traditional film production follows a series of general steps, as illustrated in Figure 1.1. While computer graphics techniques are often present throughout the pipeline, novel display methods are most likely to affect the post-production and distribution steps. As an example, let us consider 3D displays. In order to properly display 3D content, one must either capture the scene using a dedicated stereo camera setup, or convert a monoscopic view into stereo in post-production. Furthermore, once the content has been generated, at least two views need to be transmitted to consumers, which requires additional bandwidth. Such scenarios often require the use of image processing and computer graphics techniques employed as tools for content creators. On the other hand, physically accurate content is not always pleasing or even possible to show on novel display technologies. In order to make the best use of the capabilities of contemporary displays, an understanding of what is visually important is also necessary. At this point, applied perception techniques can be used to measure and model the human visual system, with the goal of obtaining the best tradeoff between the limitations of modern technology and the requirement of showing the best perceived image.

This work is targeted towards stereo 3D displays and high dynamic range displays. In Section 1.1.1 we will discuss limitations in showing depth on a 3D screen. Section 1.1.2

will touch on high dynamic range post-production and distribution challenges. Both of these problems can be seen as variations of the same theme: natural scenes possess attributes (such as stereoscopic depth or brightness contrast) that cannot be fully reproduced by modern display technologies. We therefore seek to obtain the best perceptual representation of such scenes, bounded by the technological limitations of our visual output devices.

1.1.1 Stereo 3D

When discussing the human visual system (HVS), the term *stereopsis* refers to the ability of interpreting an observed scene tridimensionally. In order to achieve this effect the HVS leverages several aspects of the available visual information, interpreting the scene and obtaining depth information. In this context, *depth* is considered to be the shortest distance from the observer to the observed point. An interesting introduction to depth perception from a cognitive perspective can be found in Chapter 2 of the classical book by Sternberg [2008]. In this work, a number of depth cues that affect our perception are presented (illustrated here in Figure 1.2).

Monocular cues include:

- texture gradients,
- relative size,
- interposition,
- linear perspective,
- aerial perspective,

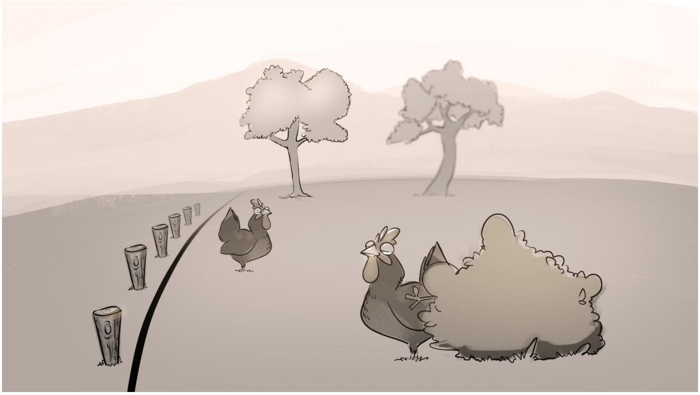


Figure 1.2: *This figure showcases a number of monocular depth cues. The reader is encouraged to try to find as many depth cues as possible.*

- location in the picture plane relative to the horizon,
- and motion parallax

while binocular cues are less numerous:

- binocular convergence,
- and binocular disparity.

Of these, *binocular disparity* is often considered to be the strongest cue [Epstein and Rogers, 1995]. While the importance of stereopsis is well known, historically most media of communication have been restricted to planar representations of the world. When striving for realism, such images can be enhanced by emulating some monocular depth cues (see Figure 1.3 for some interesting examples). While the lack of disparity does not preclude users from experiencing

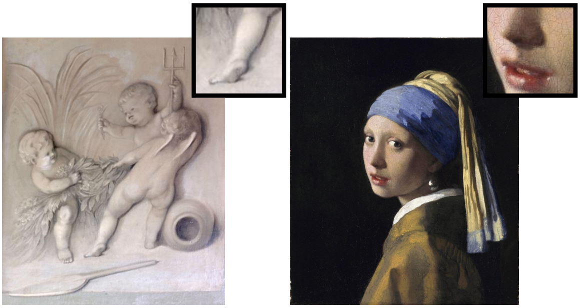


Figure 1.3: *Left: a fresco makes heavy use of non-photorealistic shadows to give a sensation of depth (unknown author); Right: shadows enhance the shape of the nose, that otherwise has very little contrast ('Girl with a Pearl Earring', Vermeer)*

traditional content, stereoscopy has been shown to be an influential factor for immersive experiences in virtual reality environments [Bowman et al., 2007]. In fact, stereoscopic content has even been shown to generate a stronger experience than traditional display with the use of fMRI [Gaebler et al., 2014]. We conclude that it is desirable to have the ability to represent the world more faithfully by adding 3D-capabilities to our content.

Since the introduction of the *stereoscope* by Wheatstone in the mid-19th century [Brewster, 1856], methods for emulating binocular cues, and therefore presenting viewers with a more realistic scene, have been devised and improved upon. The use of *stereoscopy*, that is, presenting different views to the observers' eyes, has become familar in some industries. In the context of digital content production and consumption, we refer to such content as *stereoscopic 3D* (S3D). Recent statistics from the MPAA [Motion Picture Association of America, 2014] show that S3D content is present in a significant share of cinemas. Globally, 51% of all cinemas now have 3D-capable digital screens. In addition, 15 out of 25 of the top grossing movies last year (2014) were released using stereo 3D. Another area in which S3D is of vast importance is *virtual reality* (VR). VR is expected to evolve rapidly in the immediate future [Abrash, 2014], and has attracted significant interest from the IT industry. Modern VR displays, traditionally feature dedicated screen space for each of the observers' eyes, generating a stereo experience. Without stereo, VR headsets would not be able to generate the immersive realism expected by users of the technology. However, with screens placed very close to the users' faces, VR displays must be particularly careful to avoid creating an unpleasant viewing experience. These global trends indicate that S3D is now a common technology, and learning to create good quality S3D content is an important challenge.

In spite of its ubiquity, the technology used to display S3D content does not provide users with a natural representation of the world. Most 3D displays consist of a flat screen that uses some sort of device to filter the light that reaches users' eyes in order to present them with different images. This filtering can be done using color (anaglyph glasses), time multiplexing (shutter glasses), polarization (polarized glasses) or even optical elements attached to the screen (autostereo screens). All these filtering approaches have limitations - color differences may generate binocular rivalry [Tong et al., 2006] and time interlaced presentation may generate the sensation of movement when there is none [Kim et al., 2014]. The limitations of autostereoscopic displays are discussed in depth in Chapter 4. In addition, all the methods mentioned filter a significant portion of the displayed light, resulting in a dimmer image, which affects the perceived quality of the content. Incorrect positioning relative to the screen has been shown to make the HVS misrepresent the shape and size of objects by Held and Banks [2008].

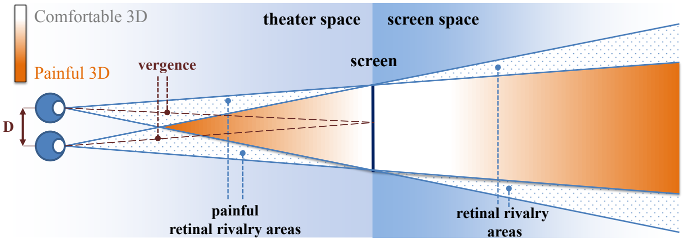


Figure 1.4: *This figure illustrates the stereoscopic comfort zone. Image taken from the work of Lang et al. [2010] with the permission of the authors.*

Most importantly, however, 3D displays represent a tridimensional scene on a flat panel. This difference between the display and what is shown generates a mismatch, namely the vergence-accomodation conflict [Shibata et al., 2011a].

Vergence-accomodation conflicts limit the gamut of the depth we can represent on standard S3D displays [Hoffman et al., 2008]. In practice, this means that a *comfort zone* exists for a pleasant 3D experience viewing S3D content. Content with depth exactly at the display location will be perceived with no discomfort, but as depth shifts further in and out of the screen, uncomfortable viewing may occur (this process is illustrated in Figure 1.4).

A natural solution to this problem would be to only display content that has a shallow depth image, that is, all objects should be located somewhere near the screen plane. This compressed presentation generates an artifact known as *cardboarding*. When cardboarded, objects appear flat and without substance, presenting an unnatural perceptual artifact. Cardboarding has been documented for a significant time [Valyus and Asher, 1966] and is expanded upon in detail in Chapter 3 of this work. We are faced with a conundrum: S3D is desirable, but may become uncomfortable to view. In order to allow for a comfortable experience, we must compress the content in depth, which reduces the strength of the stereoscopic effect.

We finish this section with the following questions: Can depth be compressed in a smarter way, while avoiding cardboarding? This question and a technique that attempts to maximize the perceived depth without exceeding the comfort zone is discussed in Chapter 4. Although binocular disparity is very important for stereopsis, could we leverage other stereo cues in order to improve the depth perception of cardboarded scenes? The use of an alternative stereo cue to enhance perceived depth is discussed in Chapter 5.

1.1.2 High Dynamic Range

The term *dynamic range* denotes the range of luminance present in an image. Luminance, or the amount of light per unit area coming from a certain direction, is a measureable physical quantity, with the SI unit being the candela per square meter (cd/m^2), also known as a *nit*. The scale shown in Figure 1.5 shows some example values for luminance present in daily life.

A well known fact about the HVS is that its sensitivity to luminance is approximately logarithmic [Fechner, 1858][Dehaene, 2003]. Previous research has shown that the human eye can adapt to a wide range of luminance levels [Ferwerda and others, 2001], and can adapt to a scene containing about four orders of magnitude simultaneously [Reinhard et al., 2010]. This adaptation interval is termed the dynamic range of the HVS. While a traditional camera sensor can be adjusted to measure light at different luminance levels by adjusting exposure time, ISO or aperture, its dynamic range

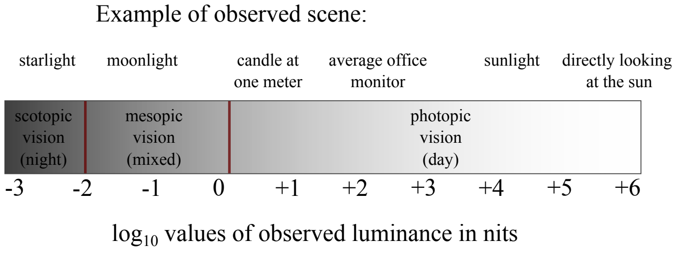


Figure 1.5: *This figure shows approximate luminance values of some everyday scenes on a logarithmic scale.*

is significantly lower than that of the HVS. Similarly, if we were given a digital image with a wide dynamic range (for example, a computer generated image), in order to correctly visualize it a display would need to be able to reproduce very different levels of brightness simultaneously. In practice, however, standard display technologies are limited by their design and allow only a modest dynamic range. This asymmetry results in a well known problem in computational photography: how can we capture and display images that are closer to something we would naturally perceive in nature?

High dynamic range (HDR) imaging was introduced to deal with this deficiency. The capture of HDR content has received a significant amount of attention by researchers in the last decade. Many methods have been proposed, for example, using bracketed exposures [Debevec and Malik, 2008], dual-iso [Hajsharif et al., 2014] or even novel camera designs [Tocci et al., 2011]. Similarly, HDR displays were introduced by Seetzen et al. [2004] and HDR display technology has been an active topic of contemporary research. Recent announcements show that displays with higher than usual dynamic ranges will be released by prominent manufacturers in the near future. Displays with maximum brightness exceeding 500 nits at reference [Canon Inc., 2015; Sony Corporation, 2015a; Sim2, 2015] and consumer [Samsung Electronics, 2015; Sony Corporation, 2015b; Vizio, 2015] levels are currently in production. While the capabilities of these displays differ and present a number of technical challenges, in this work we will focus mostly on content creation. Much like the 3D displays discussed in Section 1.1.1, HDR displays have a lot of variety in their characteristics. By way of example, the maximum brightness of the mentioned displays goes from 1000 nit [Sony Corporation, 2015b] to 4000 nit [Sim2, 2015]. Notice that while even the dimmer of these displays is significantly brighter than an average office display (which normally has a maximum brightness of 100 nit), the difference in log space between the two is almost as big! Knowing that color perception will vary significantly in this space [Kim et al., 2009], the task of broadcasting content that provides a good viewing experience for a wide range of displays becomes critically important.

Traditional post-production pipelines involve a color grading step, performed by professionals. In this step, the colors of the raw content are adjusted on a reference monitor in order to generate a more pleasant image. It is assumed, however, that consumers' monitors are tight sets of similar technologies: for example, one color grade may be performed for cinemas (which have relatively dim projectors) and another for televisions. With the advent of HDR displays, a third HDR grade is sometimes generated as well. Importantly, the process of color grading video assets is costly and time consuming, and it would be impractical to perform separate gradings for many different ranges of consumer-level displays. While it is inevitable that at least one color grading step must be performed, it would be beneficial if content could be graded for several ranges of display luminance levels simultaneously.

Another important issue is content distribution. A standard commercial business-to-consumer distribution bandwidth used for 1080i50 television signals is in the range of 12 *Mbit/s*. If additional gradings for different dynamic range displays are generated, more bandwidth will be required to transmit them to consumers. As resources are limited, any added information is a costly requirement. While previous work [Mantiuk et al., 2006a] indicates that an HDR video can be efficiently compressed in tandem with the corresponding LDR version costing approximately 30% of the original bandwidth, it is still unclear how an efficient system would operate. Sending multiple HDR versions color-graded for sufficiently separated displays would incur unacceptable costs. On the other hand, sending only a couple additional gradings might significantly impair the quality of the content seen by viewers. An ideal system would be able to efficiently encode a large number of differing versions of the same content while using a reasonable fraction of the available bandwidth.

In conclusion, it is clear that in the rapidly changing landscape of HDR display, some big questions remain: How can content be generated for the emerging variety of HDR displays? Furthermore, how can this content be distributed without requiring a prohibitive bandwidth? Further discussion on these questions can be seen in Chapter 6, where a novel end-to-end system for HDR content grading, representation and encoding is introduced.

1.2 Project Goals

The general goal of this work is to get the most of the capabilities of novel display technologies while maintaining high content quality. For 3D displays, this translates into showing depth in a comfortable range in the most efficient way. Good use of disparity and other cues can thus prevent the cardboarding artifact. For HDR displays, the objective is to optimize the luminance contrast to each target display's specification. This thesis is organized as follows:

- Chapter 2 outlines the previous art in the areas of 3D perception, stereo and autostereoscopic screens, multiview content creation, use of diverse 3D cues in computer graphics and finally HDR content creation and display.

- Chapter 3 discusses perceptual experiments targeted towards understanding and measuring the cardboarding artifact and their results.

- Chapter 4 presents a practical system for autostereo content creation that aims to prevent cardboarding. Perceptual limits of depth presentation on an autostereo display are established. An image-based algorithm uses depth maps obtained through an optimization procedure to generate multiview content that appears rounder than linearly mapped versions.

- Chapter 5 explores a relatively unknown binocular depth cue as a means to enhancing the depth sensation generated by S3D content. Perceptual experiments that explore the feasibility of our method are presented and results are shown for both computer generated and live-action content.

- Chapter 6 describes an end-to-end pipeline for HDR content creation and distribution. A user interface for the generation of a novel data structure, namely Continuous Dynamic Range video, is presented, followed by an efficient representation of this information that can be used to transmit it in practice.
- Chapter 7 Concludes this thesis with a discussion of the accomplished results and outlines some possible directions for future work.

1.3 Principal Contributions

In this section, we will briefly outline the main novel contributions of this thesis.

1.3.1 Perceptual Evaluation of Cardboarding in 3D Content Visualization

A pervasive artifact that occurs when visualizing 3D content is the so-called “cardboarding” effect, where objects appear flat due to depth compression, with relatively little research conducted to perceptually quantify its effects. In chapter 3 we aim to shed light on the subjective preferences and practical perceptual limits of stereo vision with respect to cardboarding. We present three experiments that explore the consequences of displaying simple scenes with reduced depths using both subjective ratings and adjustments and objective sensitivity metrics. Our results suggest that compressing depth to 80% or above is likely to be acceptable, whereas sensitivity to the cardboarding artifact below 30% is very high. These values could be used in practice as guidelines for commonplace depth mapping operations in 3D production pipelines.

1.3.2 Optimizing Stereo-to-Multiview Conversion for Autostereoscopic Displays

In chapter 4, we present a novel stereo-to-multiview video conversion method for glasses-free multiview displays. Different from previous stereo-to-multiview approaches, our mapping algorithm utilizes the limited depth range of autostereoscopic displays optimally and strives to preserve the scene's artistic composition and perceived depth even under strong depth compression. We first present an investigation of how subjective perceived image quality relates to spatial frequency and disparity. The outcome of this study is utilized in a two-step mapping algorithm, where we (i) compress the scene depth using a non-linear global function to the depth range of an autostereoscopic display, and (ii) enhance the depth gradients of salient objects to restore the perceived depth and salient scene structure. Finally, an adapted image domain warping algorithm is proposed to generate the multiview output, which enables overall disparity range extension.

1.3.3 Stereo from Shading

Chapter 5 presents a new method for creating and enhancing the stereoscopic 3D (S3D) sensation without using the parallax disparity between an image pair. S3D relies on a combination of cues to generate a feeling of depth, but only a few of these cues can easily be modified within a rendering pipeline without significantly changing the content. We explore one such cue—shading stereopsis—which to date has not been exploited for 3D rendering. By changing only the shading of objects between the left and right eye renders, we generate a noticeable increase in perceived depth. This effect can be used to create depth when applied to flat images, and to enhance depth when applied to shallow depth S3D images. Our method modifies the shading normals of objects or materials, such that it can be flexibly and selectively applied in complex scenes with arbitrary numbers and types of lights and indirect illumination. Our results show examples of rendered stills and video, as well as live action footage.

1.3.4 Art-Directable Continuous Dynamic Range Video

We present a novel, end-to-end workflow for content creation and distribution to a multitude of displays that have different dynamic ranges. The emergence of new, consumer level HDR displays with various peak luminance levels expected in 2015 gives rise to two new research questions: (i) how can the raw source content be graded for a diverse set of displays both efficiently and without restricting artistic freedom, and (ii) how can an arbitrary number of graded video streams be represented and encoded in an efficient way. In chapter 6 we propose a new editing paradigm which we call *dynamic range mapping* to obtain a novel *Continuous Dynamic Range (CDR)* video representation, where the luminance of the video content, instead of being a scalar value, is defined as a continuous function of the display dynamic range. We present an interactive interface where CDR videos can be efficiently created while providing full artistic control. In addition, we discuss the efficient approximation of CDR video using a polynomial series approximation, and its encoding and distribution to an arbitrary set of target displays. We validate our workflow in a subjective study, which suggests that a visually lossless CDR video representation can be achieved with little bandwidth overhead. Our solution can be implemented easily in the current distribution infrastructure and consists of transmitting two gradings and an additional meta-data stream, which occupies less than 13% current standard video distribution bandwidth.

1.3.5 Publications

In the context of this thesis, the following work has been published:

[2014a] A.Chapiro, O.Diamanti, S.Poulakos, C.O'Sullivan, A.Smolic and M. Gross. Perceptual Evaluation of Cardboarding in 3D Content Visualization. In *Proceedings of ACM Symposium on Applied Perception*.

[2014b] A.Chapiro, S.Heinzle, T.Aydin, S.Poulakos, M.Zwicker, A.Smolic and M. Gross. Optimizing Stereo-to-Multiview Conversion for Autostereoscopic Displays. In *Computer Graphics forum, Proceedings of Eurographics*.

[2015b] A.Chapiro, C.O'Sullivan, W.Jarosz, M. Gross and A.Smolic. Stereo from Shading. In *Proceedings of Eurographics Symposium on Rendering*.

[2015a] A.Chapiro, T.Aydin, N.Stefanoski, S.Croci, M.Gross and A.Smolic. Art-Directable Continuous Dynamic Range Video. In *Computers & Graphics, Elsevier*.

In addition, the following co-authored works have been published but will not be discussed in this thesis:

[2014] A.Smolic, O.Wang, M.Lang, N.Stefanoski, M.Farre, P.Greisen, S.Heinzle, M.Schaffner, A.Chapiro, A.Sorkine-Hornung and M.Gross. Image Domain Warping for Advanced 3D Video Applications. In *IEEE COMSOC MMTC E - Letter*.

[2015] R.Huber, B.Scheibehenne, A.Chapiro, S.Frey and R.Sumner. The Influence of Visual Salience on Video Consumption Behavior A Survival Analysis Approach. In *Proceedings of ACM Web Science*.

[2015] M.Junyent, P.Beltran, M.Farre, J.Pont-Tuset, A.Chapiro and A.Smolic. Video Content and Structure Description Based on Keyframes, Clusters and Storyboards. In *Proceedings of IEEE International Workshop on Multimedia Signal Processing*.

[2015] F.Zund, P.Berard, A.Chapiro, S.Schmid, M.Ryffel, A.Bermano, M.Gross and R.Sumner. Unfolding the 8-bit Era. In *Proceedings of the European Conference on Visual Media Production*.

CHAPTER 2

Related Work

This chapter deals with previous art on novel displays: in particular 3D and HDR. Relevant work is presented in the order in which it appears in this thesis.

2.1 Stereo 3D

In this section, a review of the state of the art in S3D is presented. In section 2.1.1 special focus is given to the artifacts that arise from *depth compression*, i.e. the flatenning in depth of a scene. Autostereoscopic displays offer the option of watching 3D content without the need for glasses. This freedom comes with severe limitations for these displays and makes content creation difficult. Section 2.1.2 discusses previous work on autostereo content creation. Finally, while disparity is widely considered to be the dominant binocular stereo cue, other cues can be leveraged in S3D content creation. We discuss these questions in section 2.1.3.

2.1.1 Cardboarding

The cardboarding effect, along with other stereoscopic distortions, is believed to influence both perceived image quality and visual comfort [Meesters et al., 2004; Lambooij et al., 2009a]. One factor influencing the perception of cardboarding is the mismatch between perception of object size and object disparity with distance. Howard and Rogers [2002] point out that size sensitivity is inversely proportional to distance, while disparity sensitivity is inversely proportional to the squared distance. This results in a conflict between size and depth scaling.

Another significant factor that influences cardboarding is a geometric mismatch between the stereoscopic capture, display and viewing conditions. These geometric relationships have been well studied [Woods et al., 1993; Jones et al., 2001; Masaoka et al., 2006; Yamanoue et al., 2006; Zilly et al., 2011]. Masaoka et al. [2006] sought to develop a spatial distortion prediction system to determine the extent of the stereoscopic cardboarding effect. However, they developed geometric relations without taking the subjective perception of the artifact into account.

Yamanoue et al [2000] experimentally evaluated perceived cardboarding by exploring several factors including lighting and variation of spatial thickness. They observed a significant effect of spatial thickness in the subjective rating of perceived cardboarding. Only one object with three spatial thickness values was evaluated, thereby making it difficult to draw more general conclusions regarding fine-scale changes in spatial thickness. Yamanoue et al. [2006] later modeled the cardboarding effect as the ratio of size and depth magnification. They observed a good correlation with their previous experimental observations from one object [2000]. In Chapter 3 we aim to further build on this work by introducing more objects and to rigorously observe the effects of cardboarding effect using several experimental paradigms.

With the goal of staying well within the zone of comfort [Shibata et al., 2011a], Siegel and Nagata [2000] proposed the concept of microstereopsis, in which small interocular separation is combined with alignment of interesting content about the zero parallax plane. Their informal experiments demonstrated sensitivity to small disparities and they hypothesize that minimal detectable disparity is sufficient when combined with other visual cues for depth. Didyk et al [2011; 2012b] formulated depth discrimination thresholds and demonstrated an application of minimal stereopsis.

Finally, depth adaptation is often necessary for various applications, with a range reduction being the standard. This means that cardboarding is a significant concern in practice. Previous works, such as that of Didyk and colleagues [2012a; 2012b], re-map depths based on models that take depth perception into account. They do not, however, target cardboarding specifically. Some works [Lang et al., 2010], like that described in Chapter 4 [2014b] of this thesis interpret a scene and re-target depth based on the importance of different areas. Our work in particular is aimed specifically at avoiding cardboarding when generating content for auto-stereo displays (that have a particularly small depth budget), but no quantitative characterization of the effect is provided. These mapping operations could therefore benefit from a better understanding of cardboarding.

2.1.2 Content Creation for Autostereoscopic Displays

The area of **glasses-free multiview displays** has been researched extensively in the last decade, and existing manuscripts provide a good overview on the huge body of previous work [Lueder, 2011; Wetzstein et al., 2012a; Masia et al., 2013b]. Most commercial displays are based on parallax barriers [Ives, 1903] and integral imaging [Gabriel, 1908]. Since then, much work has been devoted to improve on these glasses-free displays, with a recent trend towards computational displays [Wetzstein et al., 2011a; Wetzstein et al., 2012b; Ranieri et al., 2012; Tompkin et al., 2013]. A new method for showing stereo video on multilayer displays was introduced in [Singh and Shin, 2013]. Unfortunately, their approach cannot deal with multi-view dispalys.

Sampling and depth of field. Similar to 2D displays, multiview displays provide a sampled approximation to continuous light fields. Chai et al. [Chai et al., 2000] presented the first analysis on sampling requirements for light field signals. Durand et al. [2005] extended their work to a fundamental analysis of light transport and its sampling requirements. Based on both analyses, Zwicker et al. [2006] determine the limits of light field displays in terms of depth of field. One of their key findings shared by all multiview displays is the very shallow, device-specific depth of field. Scenes exceeding these boundaries will lead to aliasing artifacts, which can only be avoided by pre-filtering these scenes. Other researchers [Jain and Konrad, 2007; Ramachandra et al., 2011; Masia et al., 2013a] extended this work to include aliasing on light field displays in the presence of visual crosstalk.

Content creation for multiview displays still poses an unresolved challenge. These displays require multiple input views, whereas the number of views and depth of field limitations are often not known during production time. A much more promising approach is to generate multiview images from stereo footage or video+depth, using techniques such as depth-image based rendering (DIBR) [Smolic et al., 2008] or image domain warping (IDW) [Stefanoski et al., 2013]. These techniques determine how to warp the input images to new viewing positions, between the input views. However, they do not consider appropriate mapping of disparity ranges, which can lead to flattening of the perceived image, and thus reduce the depth experience. Recent work [Didyk et al., 2013] addresses content creation for MAD using phase-based motion magnification. Compared to our work presented in Chapter 4, their method does not require disparity information but only supports small disparity ranges, does not allow for local disparity manipulations, and may prefilter visually important content.

Depth adaptation has been proposed to adjust existing stereo images based on various remapping operators [Lang et al., 2010; Didyk et al., 2012b; Didyk et al., 2012a]. Our approach is similar to Lang et al. [2010] in the sense that we use IDW, which they introduced initially, as well as the notion of saliency to control the warping. However, they did not target MAD and the particular specifics of view interpolation with overall disparity range expansion. Further, they did not cover local disparity gradient enhancements. Didyk et al. [2012b] targets MAD among other applications, but filtering images is not always acceptable. Our method puts artistic intent over perception, as we try to preserve volume of important scene elements, while accepting to lose some JND of depth perception in less important image regions.

2.1.3 Stereo from Shading

By displaying two photographs of a scene between which the light was shifted horizontally, Puerta [1989] generated a 3D effect when viewing stereo images without disparity. The author suggests that the effect is caused by the difference in cast shadows between views, but mentions that shading could possibly also act as a stereo cue. Langer and Bülthoff [Langer and Bülthoff, 1999] hypothesize that shading could be an effective cue to communicate the shape of objects, particularly under natural lighting conditions such as diffuse light or lighting from above. While differences in lighting have not previously been used to augment the depth perception of S3D, researchers have shown that shadows can increase speed and reduce error rates of depth ordering tasks [Bailey et al., 2003] or to improve visual processing of technical content [Šoltészová et al., 2011]. For a full discussion of the role of cast shadows in 3D perception, please see the work of Kersten and Mamassian [2014].

Recent research has focused on enhancing perceived depth by augmenting the disparity cue. Lang et al. [2010] propose a mapping function that maps stereo into a target space non-linearly. Our own work presented in Chapter 4 [2014b] suggests a depth re-mapping approach to increase perceived depth under extreme compression. Masia et al. [2013a] present a content re-mapping approach to retarget stereo content into the zone of comfort or retarget autostereo while avoiding blurriness induced from limited angular resolution. Didyk et al. [2012a] take advantage of the Cornsweet illusion to create a convincing stereo experience and reduce the overall depth range. They also propose a perceived disparity model that takes into account both contrast and disparity [Didyk et al., 2012b]. All these methods target parallax disparity as the main source of depth perception and do not consider light and shadows. Some work has addressed the influence of color contrasts [Ichihara et al., 2007] and, more recently, luminance differences [Vangorp et al., 2014] to S3D, but shadows and shading are not directly addressed. View-dependent effects like specular and refractive materials are particularly challenging when displaying stereoscopic content and have been addressed by several researchers [Dabala et al., 2014; Templin et al., 2012]. Shading stereo could also be combined with these methods to enhance depth perception and handle highlights effectively.

In order to create stereoscopic content that is backwards compatible, Didyk et al. [2011] present stereo techniques that compress disparity until is it barely noticeable without glasses, while maintaining a small noticeable depth effect. Others aim to create display technologies that show artifact-free content that can be viewed with or without glasses, while sacrificing some contrast [Scher et al., 2013]. Shading stereo could be used for this purpose either by itself or in combination with such approaches, since the lack of disparity and identical direct shadows between views means that the mixed view seen without glasses is only barely distinguishable from a regular monoscopic image.

Finally, with respect to perception and comfort in stereo, the vergence-accommodation problem [Shibata et al., 2011a], cardboarding (as described in Chapter 3 [2014a]) and motion and luminance effects [Du et al., 2013] have all been investigated. In addition, Siegel and Nagata [Siegel and Nagata, 2000] propose using microstereopsis to view 3D content comfortably. A comprehensive survey of comfort in stereo was published by Lambooij and colleagues [2009b].

The brain has been demonstrated to fuse different low dynamic range images presented to each eye to a higher dynamic range impression [Yang et al., 2012]. Changing the normals of objects in order to exaggerate details, or to help visually parse complex information [Rusinkiewicz et al., 2006], has also been proposed, albeit not for S3D applications. In this paper, however, we present subtle lighting changes to each eye to create an S3D effect.

2.2 High Dynamic Range

High dynamic range techniques try to bring the dynamic range of displays and cameras closer to what can be perceived by the human eye. With the introduction of several lines of consumer-level HDR-capable displays to the market expected in the next few years, many challenges arise for HDR content creation. In this section we discuss relevant work on HDR image and video tone mapping, HDR display and distribution.

2.2.1 Tone Mapping

Tone mapping of HDR images has been studied extensively in the literature. A comprehensive overview can be found in Reinhard et al. [2010]. Early image tone mapping operators have been heavily influenced by the photographic film development process. The photographic tone mapping operator [Reinhard et al., 2002] utilizes an S-shaped curve to globally compress the input dynamic range, as well as dodging and burning operations to control local details. Another tone mapping approach aimed to produce natural looking results by modeling various mechanisms of the human visual system [Ferwerda et al., 1996; Pattanaik et al., 2000;

Reinhard and Devlin, 2005]. Durand and Dorsey [2002] proposed decomposing HDR images into base and detail layers by utilizing edge-aware filtering. They showed that local image details can be preserved by restricting tonal compression to the base layer while keeping the detail layer intact. Similar effects were also achieved by processing the input HDR image in the gradient domain [Fattal et al., 2002; Mantiuk et al., 2006b].

While most tone mapping operators target a single hypothetical SDR display, the display adaptive tone mapping [Mantiuk et al., 2008] approach tailors its outcome for a user-selected display dynamic range. Our CDR video representation can be thought of the union of content tone mapped for all possible displays. Additionally, our dynamic range mapping workflow does not restrict the user to a single tone mapping approach, as the source content graded for the smallest and largest dynamic range can be generated manually or using any tone mapping operator.

Tone mapping of HDR video has recently become an active field of research. The majority of the various video tone mapping operators have been discussed and subjectively evaluated by Eilertsen et al. [2013]. More recently, Boitard et al. [2014] proposed segmenting each video frame (typically to $2-4$ segments) and applying a global tone curve to each segment individually. Local adaptation is introduced at a segment level at the cost of more complex processing involving video segmentation. Another recent operator achieved temporally coherent local tone mapping through efficient spatiotemporal filtering [Aydın et al., 2014].

2.2.2 Display and Distribution

While tone mapping is a useful tool for displaying HDR content on SDR devices, the research community has long aspired to develop displays that can natively reproduce HDR images. While reproducing the entire range that the human eye can see may prove difficult, it has been subjectively shown that a luminance range from $0 - 10,000$ nits satisfies 90% of the viewers who were asked to select an ideal range [Dolby Laboratories, 2015]. The first prototype HDR display has been introduced by Seetzen et al. [**?**], which was then followed by multiple custom-built research prototypes [Wanat et al., 2012; Ferwerda and Luka, 2009; Zhang and Ferwerda, 2010; Guarnieri et al., 2008; Kim et al., 2009]. For a detailed discussion on the various HDR display approaches we refer the reader to Reinhard et al. [2010]. In parallel, experimental HDR displays have been introduced by private enterprises such as Brightside, SIM2 and Dolby. More recently, major TV manufacturers including LG, Sony, Samsung, Panasonic and TCL have announced the upcoming release of their consumer-level HDR displays. While many of these displays are being marketed using the term HDR, their dynamic ranges are quite different from each other (peak luminances varying from $800 - 4000$ nits). As a consequence of these emerging displays, traditional content production and distribution methods have to be revisited.

The efficient distribution of HDR content has also been investigated by various researchers. Mantiuk et al. [2006a] proposed encoding HDR video as a residual stream over its SDR counterpart with an overhead of 30%. More recent work proposed an optimized bit-depth quantization and human visual system based wavelet transform denoising for HDR compression [Zhang et al., 2011], and also investigated the distribution of HDR video using existing codecs such as H.264/AVC [Touze et al., 2013].

CHAPTER 3

The Cardboarding Artifact

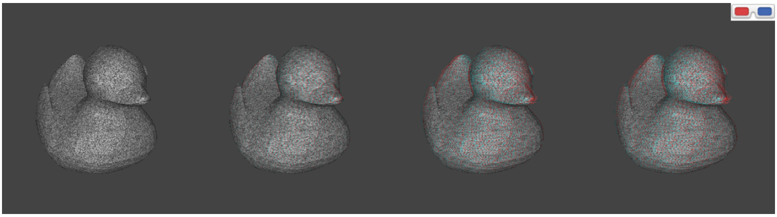


Figure 3.1: *The cardboarding effect is illustrated in these anaglyphs, with depth compression levels of* $\alpha = 0.0$ *(completely flat),* $\alpha = 0.2$, $\alpha = 0.8$, *and* $\alpha = 1.0$ *(fully 3D). In our studies, we found that differences between the left three images were detected significantly often, whereas the right two appeared to be the same and equally acceptable to our participants.*

3.1 Introduction

Creating high-quality 3D content is a challenging task, with many efforts in academia and industry directed towards the development of an effective pipeline for 3D content production and delivery. Unlike with regular displays, 3D viewing can often be physically uncomfortable when unsuitable depth volumes are displayed. To avoid this discomfort, depth limitations on displayed content for comfortable watching have been determined [Shibata et al., 2011a].

However, these inherent limitations of 3D-capable displays in showing depth are not uniform, and may change to a large degree depending on the technology used (for example, auto-stereo screens have much smaller ranges than most displays that use glasses). From a content production perspective, this means that content depths must often be adapted before they can be displayed. When voluminous objects are shown with a reduced depth profile, such as one that could result from depth re-mapping to suit a display's capabilities, the reduced depth profile appears unnaturally flat and results in a disturbing perception of the scene geometry known as "cardboarding" (see Figure 3.2). Although this perceptual artifact is very common, it has been relatively unexplored.

It follows that when 3D content is compressed in depth, cardboarding should be avoided if possible. Since the effect has not yet been fully explored in the research literature, content creators struggle to make well-informed decisions when implementing mapping methods and must follow heuristic solutions or adjust content manually. The contribution of our paper is a perceptual exploration of preferences and thresholds for cardboarding effects in simple scenes, which can be applied as guidelines to improve exist-

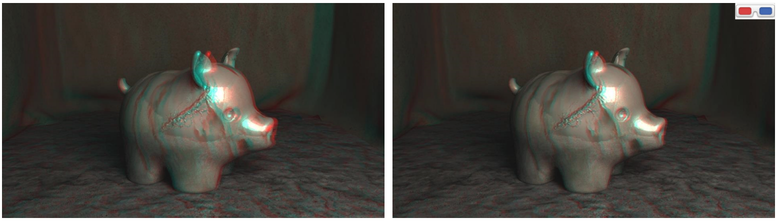


Figure 3.2: *This anaglyph image showcases cardboarding. The right image has a starkly reduced depth profile, resulting in an unnatural perception of the scene's geometry. The left image has a more natural depth profile, and is provided as reference. Notice how the perception of the room's size changes when looking at the back wall.*

ing methods in depth re-mapping. We present three experiments, the results of which each painted a consistent picture of the effects of cardboarding on four models. This methodology can now be used to explore the cardboarding effect further in more complex scenes.

In this chapter we will present some 3D examples using anaglyph. Such figures are marked with this icon 👓. They can be viewed in 3D using anaglyph glasses (red - left, cyan - right). Please note that to get a better depth perspective you can zoom in on the figures.

3.2 Experiments

We conducted three perceptual experiments in order to explore the effects of cardboarding. In the *Dial* experiment, we aimed to determine whether preferences for the appearance of stereo scenes could be self-selected by our participants. We found that this was a difficult task, with much variation

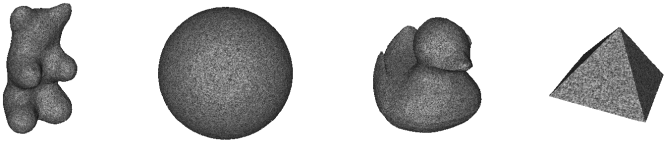
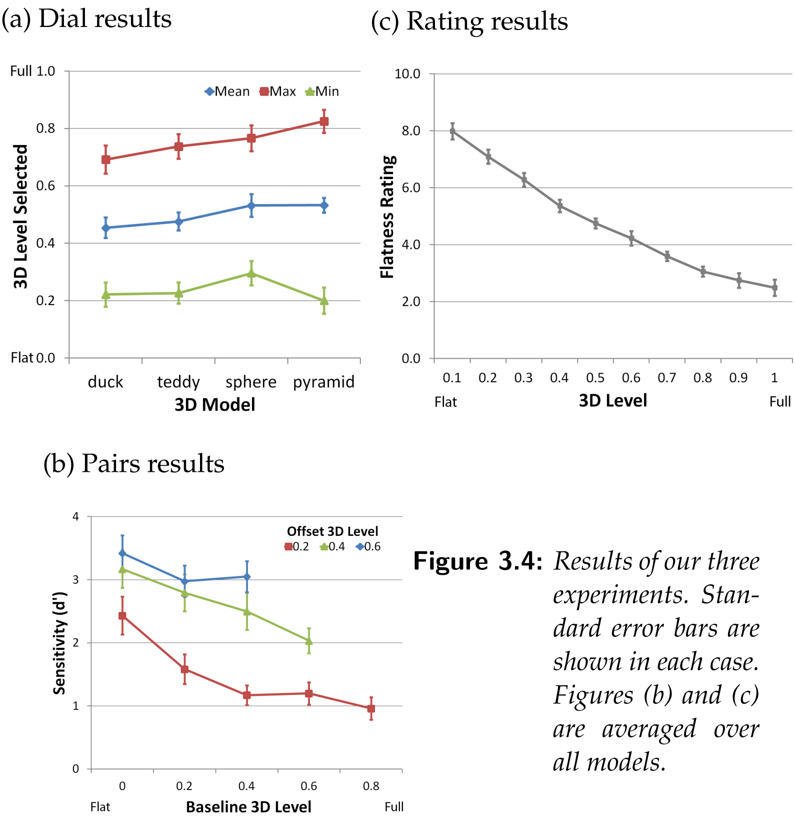


Figure 3.3: *This figure shows a monoscopic view of the stereo scenes used in the experiment described in section 3.2.1. The meshes shown were displayed with varying depth profiles on a solid gray background.*

in the quality levels selected, even for a single participant. However, the flatness was almost never disturbing above 80% and nearly always noticed below 30%. We followed up with a *Pairs* study, to determine whether this wide range of preferences was due to a lack of sensitivity to the cardboarding artifact. We found that participants were relatively efficient at detecting differences between more flattened images, but less sensitive the fuller i.e., more 3D, the images became. Finally, we ran a subjective *Ratings* experiment, and found that the results were consistent with the previous two studies. In particular, we found that both objective sensitivity performance and subjective preference rating indicate a lack of sensitivity and hence similar ratings for compression to 80% and above of the full, 3D model, whereas cardboarding up to 30% almost always noticeable.

3.2.1 Method

We recruited 19 naive participants (2F,17M) aged between 23 and 34 with normal or corrected-to-normal vision in both eyes. Of this group, 15 participants performed all three experiments in random order, while four performed only the Pairs and Rating studies. The observers viewed 3D scenes consisting of the simple objects shown in Figure 3.3



(a) Dial results

(c) Rating results

(b) Pairs results

Figure 3.4: *Results of our three experiments. Standard error bars are shown in each case. Figures (b) and (c) are averaged over all models.*

on an Alienware 2310 23″ 3D capable monitor with the help of time-multiplexed glasses, and sat approximately 60 centimeters from the screen. A mix of geometric and natural objects was selected, with both angular and round appearance. The standard setup for each experiment mimicked the position of the observer's eyes as cameras in the renderer, which were located 6 centimeters apart and 60 centimeters away from the objects being rendered. In this way, the ren-

Base	**Offset**	*PAIR Groups*								**Rating**	*LEVEL Groups*							
0.8	0.2	†								1.0	†							
0.4	0.2	†	†							0.9	†							
0.6	0.2	†	†							0.8	†	†						
0.2	0.2		†	†						0.7		†	†					
0.6	0.4			†	†					0.6			†	†				
0.0	0.2				†	†				0.5				†	†			
0.4	0.4				†	†	†			0.4					†			
0.2	0.4					†	†	†		0.3						†		
0.2	0.6						†	†	†	0.2							†	
0.4	0.6							†	†	0.1								†
0.0	0.4							†	†									
0.0	0.6								†									

Table 3.1: *Homogeneous groups calculated using Fisher's LSD post-hoc analysis for: Pair effect in the Pairs experiment (l); Level effect in the Rating experiment (r). Each column indicates which pairs, or levels, were found to not be significantly different from each other. The values are graphed in Figure 3.4.*

dered unmodified 3D scene showed objects with similar 3D characteristics as those of a real-world object at the center of the screen. At the start of each experiment, cardboarding was explained to each participant and they received training on each task, and written instructions were available throughout for reference.

The rendering cameras were oriented parallel to each other along the z axis and the resulting stereo images were re-converged around the center of coordinates, i.e., the center of coordinates always had zero disparity and appeared to be at the screen's depth. For the camera baseline β and point $p = (x_p, y_p, z_p)$, the disparity between the rendered views is d_p. If our camera baseline was changed to be $\alpha * \beta$ with $\alpha \in [0, 1]$, the disparity of p would become $\alpha * d_p$. This effectively gives us the freedom to linearly control the overall disparity compression of our scene by changing the baseline by the factor α. Figure 3.1 shows an example of a mesh mapped with $\alpha = 0.0$, $\alpha = 0.2$, $\alpha = 0.8$ and $\alpha = 1.0$.

In the *Dial* experiment, a method-of-adjustment process was performed where the same object was displayed twice, once with $\alpha = 0$ and the other with $\alpha = 1$. Pressing one button increased α by 0.02 and another decreased it by the same amount. This gave a total of 16 stimuli (4 models X 2 directions X 2 repetitions). When the object began flat, the task was to select the point when cardboarding stopped being disturbing; when the object began full, the point where cardboarding started to become disturbing was selected. In the *Pairs* study, users were shown two versions of the same model side by side. The shapes were shown with either the same or different levels of fullness (i.e. the same α value), with random left-right placement. One object was known as the baseline level, with possible values of $\alpha \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$ and was compared with another with one of the offsets: $0.0, 0.2, 0.4, 0.6$ added, with three repetitions of each pair. The task was to answer yes or no to the question: "Are the objects the same in terms of cardboarding?" Finally, for the *Ratings* experiment, a single object was displayed in the center of the screen with a random α baseline factor from 0 to 1 with a 0.1 step. Each stimulus was repeated twice, totaling 20 stimuli for each of the four models. The task was rating the scene on a scale of 1 to 10 with respect to cardboarding, with 1 = "Not disturbing at all" and 10 = "Very disturbing, completely flat".

3.2.2 Results

We performed Repeated Measures Analysis of Variance (ANOVA) on participant responses to test for statistically significant effects, and performed post-hoc analysis using Fisher's LSD (Least Significant Difference) test for pair-wise comparisons of means. Effects are considered to be significant at the 95% level ($p < 0.05$). The results are summarized in Figure 3.4 and Table 3.1.

For the *Dial* experiment (Figure 3.4(a)), we performed single factor (4 Model) repeated measures ANOVAs on the Min, Max and Mean of each participant's selected levels, averaged over all participants. There was a main effect of Model for the means ($F(3, 42) = 3.05, p < 0.05$), where the duck was set to a significantly lower level on average than the sphere or pyramid, but not the teddy. This is probably due to the beak of the duck where the change in 3D was much more obvious, in that participants reported that it "came out" of the screen more and contrasted more with the tail in the background. We can see that each participant selected a wide range of acceptable levels, indicating that the decision was a difficult one for them. However, the Max values rarely exceeded 80%, indicating that compression to that level and above was not found to be disturbing. The averages were around 50% and the lowest Min values were around 20%, meaning that in some cases, they accepted very high compression levels for some stimuli.

The task in the *Pairs* experiment (Figure 3.4(b)) is a signal detection one, so we calculated the sensitivity of each participant to a difference between the two images. The d-prime (d') metric is commonly used in psychophysics to reliably measure sensitivity to a signal, as it takes response bias into account (i.e., the tendency to be over-conservative or over-discriminative) by considering both the Hit Rate (e.g., percentage of time a difference is correctly reported) and the False Alarm Rate (e.g., percentage of time the images are incorrectly reported to be different when they are the same). High values indicate that participants are very sensitive to a difference being present between the stimuli, whereas values of 1 and below are considered to be guessing. We performed a two-way (4 Model x 12 Pair) repeated measures ANOVA on the d' values. A main effect of Model ($F(3, 54) = 7.05, p < 0.0005$) was found, where differences for the sphere were most easily detected, and of Pair ($F(11, 198) = 25.5, p \approx 0.0$), where the same differences between fuller stimuli were far less detectable than between those that were very compressed. This result is expected, as low α values incurred a larger relative change. Again, when compression was to 80% or above, sensitivity was at its lowest, whereas when compression was to 20% or below, performance was above chance. These results are consistent with our findings in the Dial experiment. Please see the homogeneous groups in Table 3.1(left).

Finally, we performed a two-way (4 Model x 10 Level) repeated measures ANOVA on the results of the *Rating* experiment (Figure 3.4(c)) and found a main effect of the preference Level ($F(9, 171) = 64.3, p \approx 0.0$). From the homogeneous groups shown in Table 3.1(right), we can see that the flattest levels 0.1-0.3 are all significantly different, whereas differences between the fuller 0.8-1 stimuli are much smaller, and not statistically significant, indicating a plateauing effect at about 80%. Compression to 40% was rated on average just above 5, indicating that this is the point after which the flatness became noticeable more often than not. From 30% it was clearly rated flat far more often. From these and the results of the other two experiments, we can conclude that depth compression to 80% fullness or above is likely to be acceptable, whereas below 30% it is probably never going to be acceptable. Of course, we cannot generalize from the four simple scenes we presented to more complex scenes, though it seems possible that we have presented a worst-case scenario, and more complexity might mask cardboarding artifacts further, allowing higher compression rates below the conservative 80% limit than we found here, as several mid-range levels were acceptable at least some of the time.

3.3 Conclusions

We have shown that, at least for the simple scenes depicted in our experiments, depth can be safely compressed by up to 20% without significantly affecting perceived cardboarding. It may be possible to compress at much higher rates, as there appears to be a wide range of compression ratios that appear acceptable to some viewers at least some of the time. However, it appears that below 30% of the natural depth, cardboarding is significantly disturbing. The results obtained in this work could be directly applied to guide existing depth remapping methods. Further studies are needed to examine the effects of many other factors (e.g., lighting effects, scene complexity, motion) and also to determine more subjective preference measures, in addition to the simple ratings we recorded here. Our findings may provide information that could be used to map depth into a smaller range while avoiding as much as possible the introduction of disturbing cardboarding artifacts. Previous approaches such as those of Lang et al. [2010] and the approach described in Chapter 4 [2014b] could use the cluster boundaries we have found as targets for the depth budget given to a salient region.

CHAPTER 4

Autostereo Content Creation

4.1 Introduction

Multiview autostereoscopic displays (MADs) are expected to make their way into the households in the near future, and major display manufacturers are intensively working towards consumer-grade screens. A significant limitation of the current autostereo technology is the display's depth range. While the emergence of very high resolution displays (4k and beyond) can alleviate this problem to a certain degree, the constraints on the display depth range will remain as an inherent limitation of the MAD techologies.

In contrast to the recent progress on the display side, autostereoscopic content creation still lacks the tools and standards for the mainstream deployment of MAD technolo-

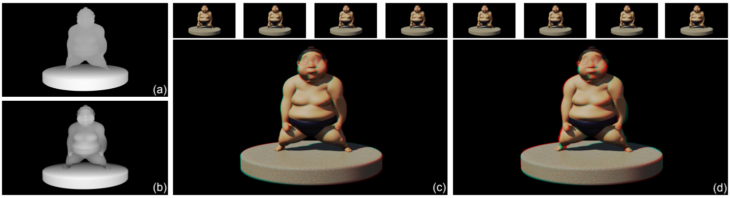


Figure 4.1: *Our method produces depth-enhanced multiview content from stereo images while preserving the original artistic intent. (a) and (b) show the linearly mapped disparities as well as enhanced disparities computed using our method. (c) and (d) show the result of stereo-to-multiview conversion using (a) and (b), respectively. Our method avoids the cardboarding effect that can be seen in the linearly mapped version.*

gies. In fact, content creation for 2-view stereo (S3D) for glasses-based systems is just developing and maturing. Even with the emergence of MAD technologies, stereo will remain in use for the foreseeable future, as content creators cannot change rapidly and completely. Consequently, support for legacy stereo content through stereo-to-multiview conversion will likely be a key feature for ensuring a graceful and backward compatible transition from 2-view stereo to multiview autostereo.

The main technical challenge in faithful stereo-to-multiview conversion is that the disparity range of many S3D scenes often exceeds the limitations of MADs. However, current stereo-to-multiview conversion methods such as depth-image based rendering (DIBR) [Smolic et al., 2008] and image domain warping (IDW) [Stefanoski et al., 2013] directly interpolate between the two input views and do not take the inherent depth limitations of autostereoscopic screens into account. Moreover, unlike in the early days of S3D where the technology was used mainly as a "wow factor", more recently the depth layout is being used as an artistic element to support the content's narrative and action. Thus, any autostereoscopic content creation workflow should not only reduce the content's depth range to the limits of the MAD technology, but also preserve the artistic intent and *perceived* depth layout as much as possible.

In their basic work, Zwicker et al. [2006] evaluate the bounds on content creation for multiview displays and propose filtering the content as solution for the limited depth range. Didyk et al. recently proposed a framework for depth remapping based on just noticeable differences (JND) of depth perception[Didyk et al., 2012b]. They identify content creation for multiview as one of the use cases, and also propose blurring the content in addition to depth compression. However, from a creative point of view, filtering the content in this way is unsuitable. For instance those objects that are far off screen are in many cases the most important by artistic intent, and blurring a character that is in center of attention is undesirable. Furthermore, previous methods have not been evaluated for video and live action footage so far. Inspired by this previous work, we address stereo-to-multiview conversion from the point of view of content creation. Rather than JND, we introduce a notion of saliency to capture and characterize artistic intent. Filtering important image content is avoided and instead we rather sacrifice noticeable disparity differences in non-salient regions.

We start by investigating the influence of disparity and texture frequency on the perceived picture quality through a subjective study. Based on the study, we choose a range of disparities that are perceived as pleasant but exceed the theoretical limit of multiview displays. We then compute a disparity mapping that retains the overall depth layout but strives to keep the volume of salient objects to avoid cardboarding. To achieve this, we perform our mapping in two steps. A global non-linear mapping operator first transforms the overall depth range to the range of the display. In a second step, we locally enhance the depth gradients to reduce the effect of cardboarding. Both global mapping and local gradient enhancement are based on saliency. We then generate multiview content directly from the input views using an extended version of image domain warping (IDW) [Stefanoski et al., 2013], which is applicable for disparity range extension. We investigate the suitability of the different mapping strategies for synthetic content (where perfect disparity is given) and for live action content (where imperfect disparities pose additional challenges). In a final user study we validate our approach on a variety of live action and synthetic video sequences.

In summary, our paper makes the following contributions

- Subjective user study on perceived quality versus disparity on a multiview autostereoscopic display.
- Global and local disparity mapping algorithms based on saliency for stereo-to-multiview conversion.
- Extended IDW algorithm for optimized disparity mapping, which supports overall disparity range extension.
- Validation of the approaches using a variety of live action and synthetic video content.

4.2 Calibration

Multiview displays usually exhibit a very shallow depth of field, but content is often displayed using substantially bigger depth ranges. Despite the violation of the sampling requirements, only a small amount of aliasing artifacts is usually perceived. We therefore investigate the relationship between image disparity and perceived quality with a subjective user study. The goal of the study is to determine the sensitivity of spectators to such depth ranges that exceed the display's depth of field. The outcome of this study is then used as a guideline for disparity mapping in our content creation pipeline.

In our experiment, the *stimuli* consist of the 8 synthesized views of a simple disc with the radius of 100 pixels at a certain distance, displayed against a background positioned at the display plane (Figure 4.2).

Both the disc and the background were covered with a number of different grayscale textures that varied in spatial contrast frequency and stereoscopic disparity. The textures were generated by applying various low-pass filters to random per-pixel noise in the frequency domain utilizing the Discrete Cosine Transform.

Our *setup* was chosen to resemble a regular viewing experience at a home theater system. All stimuli were presented on an 8-view, 47″ Alioscopy display with an approximate depth of field of ±98mm ([Zwicker et al., 2006]), which corresponds to a disparity maximum of ±2.66 pixels between two consecutive views.

During the experiment the subjects were comfortably seated on a chair 4.3 meters away from the display. Each subject was given a task that consisted of rating the perceived

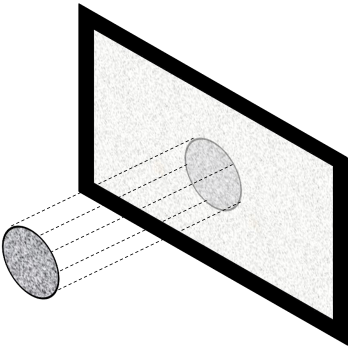


Figure 4.2: *A depiction of how our stereoscopic stimuli is perceived by the subjects.*

crosstalk and angular aliasing on a scale of 0 to 9 using a computer keyboard. Our subjects were 10 males and 6 females from age 25 to 36. In order to prevent the commonly encountered anchoring problems in rating studies, each subject performed the entire experiment twice, and only the results of the second iteration were used. The subjects were free to spend as much time as they needed at each trial, and most subjects finished the experiment in 20-25 minutes.

The mean preference scores over all subjects are shown in Figure 4.3. The main finding of this study is that disparity has a significant influence on preference score, which is a direct result of depth of field of the multiview display (see bandwidth analysis of Zwicker et al. [2006]), and not due to other effects such as vergence-accomodation conflict which has a much larger comfort zone of about 306 pixels [Shibata et al., 2011b]. We also found that, to a lesser degree, spatial frequency also has a statistically significant influence on preference score, especially for the middle frequency range.

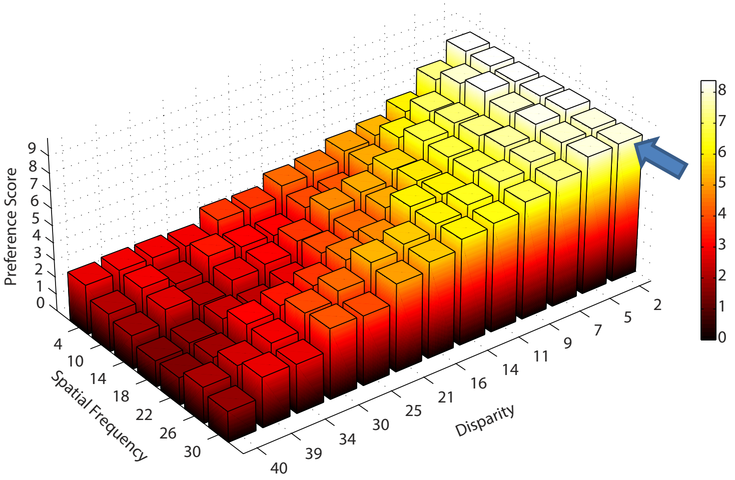


Figure 4.3: *Subjective data showing the relation of spatial frequency and disparity to mean preference score. The blue arrow denotes the depth of field of our display*

Furthermore, our study shows that disparity ranges of ×2 the display depth of field only create noticeable artifacts for higher texture frequencies. The quality then degrades almost linearly for even higher disparities. Other lenticular or parallax-barrier based multiview displays will most likely exhibit similar characteristics.

Using the data shown in Figure 4.3 we can estimate pleasant disparity ranges by taking the spatial frequency of the content into account and choosing a suitable threshold preference score. In practice, we chose a value of maximum ±5 pixels disparity for our display to achieve good image quality while allowing for twice the supported depth range.

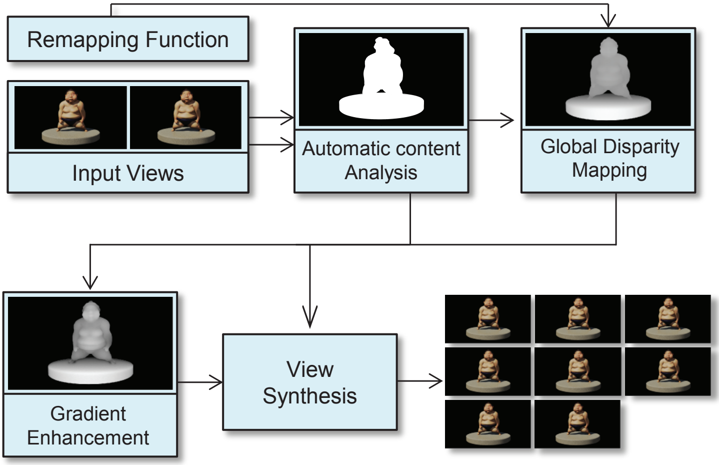


Figure 4.4: *Overview of our stereo-to-multiview conversion pipeline. The input stereo and disparity video is analysed for saliency, and edges in a first step. Next, a global non-linear mapping transforms the input disparity space into the disparity range of the target display. The subsequent local gradient enhancement step then recovers flattened image regions of important objects in a third step. Finally, our optimized image domain warping is used to synthesize the output views for the multiview display.*

4.3 Method

Our algorithm converts stereo 3D input into multiview video output optimized for autostereoscopic displays. The overall pipeline is illustrated in Fig. 4.4. In a first step, the overall disparity space is globally transformed to a new disparity range suitable for the display device limits. Next, the globally transformed disparities are locally enhanced for salient objects. Then, the transformed disparity map is used to perform view interpolation to generate the final multi-view output. A detailed description of our inputs can be seen in Section 4.4. In the following, we will give more details on the individual steps.

4.3.1 Global Disparity Mapping

The disparity range of professional stereo content is usually not very well suited for MADs, which tend to support a significantly smaller disparity range. Due to inherent difficulties of disparity estimation, conversion of live action content creates specific challenges. Estimated disparity maps may contain many kinds of artifacts and imperfections, of which cardboard effects and estimation failures are most severe. In the case of cardboarding, gradients across objects are often missing, which can result in flat disparity regions partitioned into multiple layers. Furthermore, estimation failures can lead to drastic changes in disparity or holes in the estimation [1]. Gradient-based approaches such as described in the next section will not work well alone with such non-continuous content, and we therefore propose to use a two-step mapping.

Our pipeline starts by globally transforming the input disparity space into a new piece-wise linear disparity space that better suits the device-dependent limits of MADs. Our mapping works equally well for live-action input (with piece-wise linear disparity maps) as well as rendered content (with continuous disparity maps). Our piece-wise linear mapping uses saliency characteristics of the input content to keep important regions as uncompressed as possible.

[1] Please compare input disparity maps of synthetic vs. live action content in the supplemental video of the relevant publication [Chapiro et al., 2014b].

The unavoidable distortion is hidden in areas which are less important.

In the following, we will describe the global mapping. Assuming the original disparity map contains values in a space $[d_{\min}, d_{\max}]$, the mapping is then a function $f : [d_{\min}, d_{\max}] \rightarrow [d'_{\min}, d'_{\max}]$. For our piecewise linear approach, we divide the domain of f into n equally sized bins which are linearly mapped to bins in the co-domain. Thus the linear function $f_i : [d^i_{min}, d^{i+1}_{max}] \rightarrow [d'^i_{min}, d'^{i+1}_{max}]$ is of the form $f_i(x) = \Delta_i x + \alpha_i$. If we define $R = d^{i+1}_{max} - d^i_{min}$ and $R' = d'^{i+1}_{max} - d'^i_{min}$, a linear function would be equivalent to a single bin with $\Delta = R'/R$. We would like our Δ_i to satisfy the following conditions:

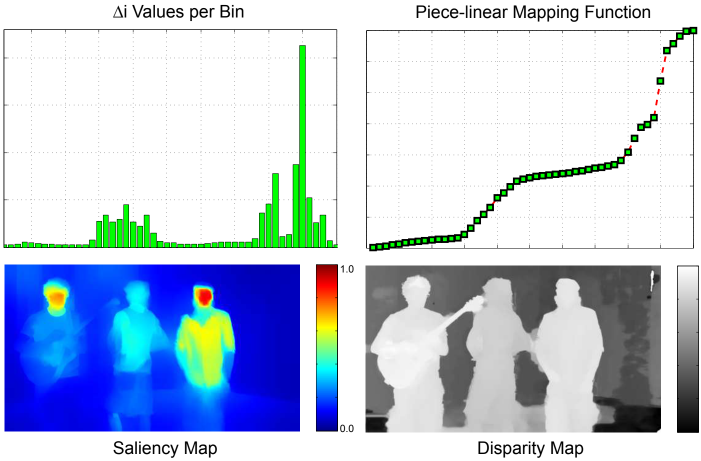
$$\sum_{i=1}^{n} \Delta_i = \Delta \,, \tag{4.1}$$

$$\Delta_i \geq 0, \forall i. \tag{4.2}$$

This ensures that we map our disparities exactly into the target space and that the three dimensional position of pixels is never reversed, i.e. a pixel will never be mapped to a position in front of another pixel if it was behind it originally. Given these conditions and naming s_i the sum of the saliency values of all pixels in the bin i, we propose the solution:

$$\Delta_i = \frac{s_i}{\sum_{j=1}^{n} s_j} \Delta\alpha + (1 - \alpha)\frac{\Delta}{n} \tag{4.3}$$

The coefficient $\alpha \in [0, 1]$ controls by how much a given bin can be compressed. A value of 0 defaults the mapping into a simple linear mapping and a value of 1 means bins with no salient pixels will have their disparity completely removed. An illustration showing the result of this algorithm can be seen in Fig. 4.5. The method described above is directly related to the saliency provided for the scene, and as such is


Δi Values per Bin

Piece-linear Mapping Function

Saliency Map

Disparity Map

Figure 4.5: *The top right image shows our piecewise linear mapping function with the respective Δ_i values per bin shown on the top left. The bottom images depict the saliency map (left) and the corresponding disparity map (right).*

very sensitive to temporal instability in saliency. To prevent the global mapping function from becoming temporally unstable, saliencies are filtered out over several frames which ensures that the disparity re-mapping is similar for consecutive images. A related approach was proposed in [Lang et al., 2010], which integrates over saliency instead of computing a piece-wise linear function, and does not provide a compression safeguard parameter k.

Fig. 4.7 shows a result of such piecewise linear global mapping. The left side shows results of a linear mapping. Our results are shown on the right. Both are mapped into a fraction of the input disparity range. Our mapping function nicely compresses empty space, while retaining disparity in

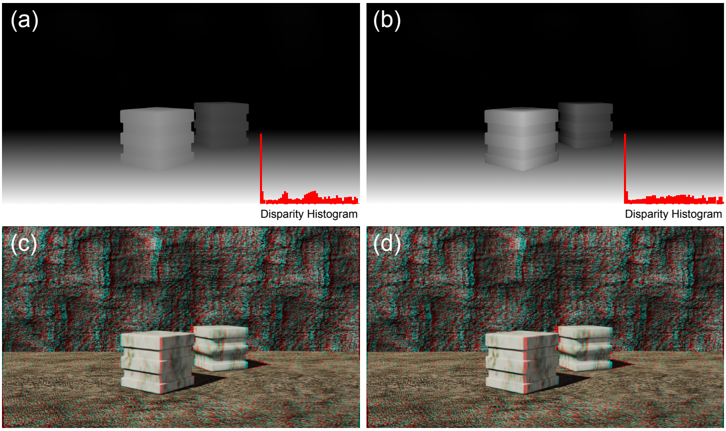


Figure 4.6: *Local disparity gradient enhancement. (a) shows a linearly mapped disparity map of a scene; (b) shows the disparity map adapted by our algorithm; in (c) and (d) two generated views are displayed in anaglyph, in a disparity range similar to two adjacent views of a MAD display. Notice how the cardboarding effect flattens out the cube in (c).*

more salient regions leading to an enhanced depth experience.

Other operators. Similar to Lang et al. [2010] our pipeline supports arbitrary global mappings which can be specified by the user or predefined for a certain system. In principle, all C0-continous and monotonically increasing functions are allowed, such as operators proposed by Didyk and colleagues [2012b], or non-linear operators proposed in the work of Lang et al. [2010].

4.3.2 Local Disparity Gradient Enhancement

After the global disparity mapping we perform an additional, local mapping step. Our main goal is to locally enhance disparity gradients in important image regions for an increased depth perception. We formulate our goal as set of constraints, that can then be solved for the locally enhanced disparity map D_L with a least-squares energy minimization. A result of this mapping can be seen in Figure 4.6. In the following, we will use the ensuing notation. Let $\mathbf{x} \in \mathbf{R}^2 = (x, y)$ be an image position, and $D(\mathbf{x}) \in \mathbf{R}$ be a disparity map.

Gradient constraints. As our central constraint, we enforce the mapped disparity gradients of salient image regions to be similar to the gradients of the input disparity map D_I:

$$\frac{\partial}{\partial x} D_\mathrm{L}(\mathbf{x}) = \lambda \, \frac{\partial}{\partial x} D_\mathrm{I}(\mathbf{x}) \, , \tag{4.4}$$

$$\frac{\partial}{\partial y} D_\mathrm{L}(\mathbf{x}) = \lambda \, \frac{\partial}{\partial y} D_\mathrm{I}(\mathbf{x}) \, . \tag{4.5}$$

The global parameter λ is then a constant factor to control the overall disparity enhancement, and is dependent on the disparity compression from the previous global mapping. In general, we propose to use a factor of $\lambda = 2({}^{d_\mathrm{range}}/_{d'_\mathrm{range}})$, where d_range and d'_range are the disparity ranges before and after the global mapping, respectively.

Global mapping constraints. In addition, we enforce the overall mapping to follow the global mapped disparity D_G as closely as possible:

$$D_\mathrm{L}(\mathbf{x}) = D_\mathrm{G}(\mathbf{x}) \, . \tag{4.6}$$

Least squares energy minimization. The constraints defined in the above equations can then be rewritten as constraints of a linear least squares energy minimization. Let

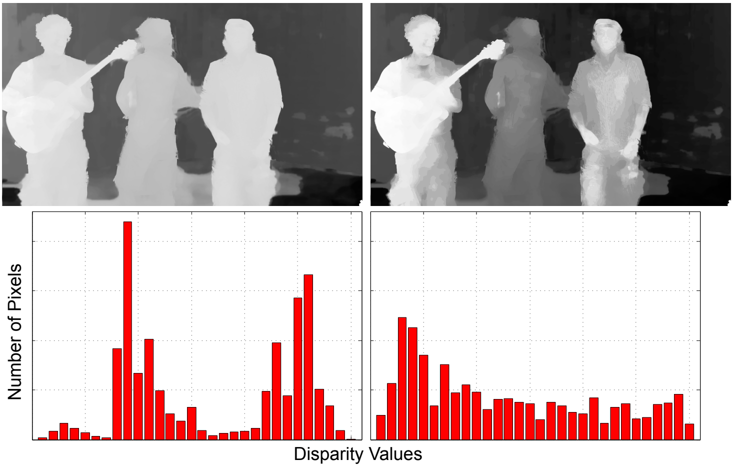


Figure 4.7: *Results of linear (left) and our saliency-based piece-wise linear (right) global mapping. The histograms show how our approach compresses unimportant space, while retaining volume of salient objects as well as possible.*

$S(x,y) : \mathbf{R}^2 \to (0,1]$ be a saliency map that classifies important image regions. A small amount of saliency is added to all pixels to prevent null weights in the constraints. Equations (4.4) and (4.5) can then be rewritten as

$$E_{\mathrm{g}}(D_{\mathrm{L}}) = \sum_{\mathbf{x}} S(\mathbf{x})\, ||\nabla D_{\mathrm{L}}(\mathbf{x}) - \lambda \nabla D_{\mathrm{I}}(\mathbf{x})||^2 \tag{4.7}$$

where ∇ is the vector differential operator, and $||\cdot||$ defines the vector norm. The global mapping constraints (4.6) are reformulated as

$$E_{\mathrm{l}}(D_{\mathrm{L}}) = \sum_{\mathbf{x}} (D_{\mathrm{L}}(\mathbf{x}) - D_{\mathrm{G}}(\mathbf{x}))^2 \tag{4.8}$$

The optimum linear least squares solution for $D_{\mathrm{L}}(\mathbf{x})$ can then be found by minimizing

$$\underset{D_{\mathrm{L}}}{\operatorname{argmin}} \left(w_{\mathrm{g}} E_{\mathrm{g}} \left(D_{\mathrm{L}} \right) + w_{\mathrm{l}} E_{\mathrm{l}} \left(D_{\mathrm{L}} \right) \right) \tag{4.9}$$

Note, that this minimum can be computed by solving a linear system, see the work of Greisen et al. [2013] for a good overview. The system defined in (4.9) will try to enhance the gradients of the salient regions, while trying to enforce all other disparity values towards their globally mapped version D_{G}. Disparity edges, i.e. strong disparity gradients between objects at different depths, can lead to a high contribution to the squared error, and thus such disparity edges would be enforced strongly as well. As we are only interested in gradients within the objects, these disparity edges need special treatment which we will discuss in the following. This step is different from previous methods, such as that of Lang et al. [2010].

Disparity edges. The gradient constraint can lead to artifacts around disparity edges due to very high disparity gradients between objects. We thus remove the influence of such disparity edges to enforce the gradient enhancement within objects only. Luckily, disparity edges can usually be detected quite robustly on the disparity map. We use a combination of a simple threshold function and a more sophisticated Canny edge detector on the input disparity D_{I} to determine the set of edge pixels E. Subsequently, we enforce the salience value to be zero at these edge pixels $S(\mathbf{x}) = 0$ for $\mathbf{x} \in E$.

Fig. 4.6 shows a result of local disparity gradient enhancement. In our result the cubes have more volume, while the linearly mapped version appears more flat.

4.3.3 View Interpolation

We developed an extension of existing work on image domain warping [Lang et al., 2010; Stefanoski et al., 2013] for stereo-to-multiview interpolation. The previously computed optimized disparity maps are used as main input to control this process. Based on optimized disparity, we formulate a constrained energy minimization problem, which is solved by linear least squares optimization (similar to previous section). The results are warping functions, which deform the input views to generate the novel in-between views. In addition to disparity, we apply conformal constraints that penalize local deformations, and line constraints that disadvantage line bending.

In the following, we will describe the particular constraints in more detail. The optimized disparity map $D(\mathbf{x}) : \mathbf{R}^2$ is used to compute a warp $w(\mathbf{x}) : \mathbf{R}^2 \rightarrow \mathbf{R}^2$, where the deformations should be hidden in visually less important regions. The warp $w(\mathbf{x})$ will then describe the optimal transformation of the input view corresponding to $D(\mathbf{x})$.

Disparity constraints. The disparity constraints can be viewed as positional constraints: every point in the image should be translated to the position described by its disparity:

$$w(\mathbf{x}) = \begin{bmatrix} x + D(\mathbf{x}) \\ y \end{bmatrix} \tag{4.10}$$

Conformal constraints. The conformal constraints penalize deformations, and are mainly evaluated on visually salient image regions. A constraint of the form $\frac{\partial}{\partial x} w(\mathbf{x})^{(x)} = 1$ prescribes to avoid any compression or enlargement along the x-direction, whereas a constraint of the form $\frac{\partial}{\partial x} w(\mathbf{x})^{(y)} = 0$

penalizes deformations that result in a pixel-shear operation. All four constraints are then formulated as:

$$\frac{\partial}{\partial x} w(\mathbf{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \frac{\partial}{\partial y} w(\mathbf{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{4.11}$$

Depth ordering constraints. Because disparity edges often have gradients that are very different from their neighbors, we use edge detection to reduce their saliency values, often preventing artifacts. For this we use the same edge map as the one previously calculated for Section 4.3.2. Additionally, pixel overlaps where the correct order of pixels along a line is reversed often occurs for large warps. This may generate large distortions in the warp optimization step. To resolve such conflicts we perform a simple check where, in case of an overlap, the occluded pixel is moved to the coordinate of its occluder.

Temporal constraints. Applying all constraints results in output images that have correct disparity values and hide distortions in visually non-salient areas. However, when applying this method for each frame of a video sequence, small changes in the input might result in larger changes within the optimization, which may lead to disturbing temporal artifacts. We try to remedy this by introducing a temporal constraint that takes into account the warps calculated for previous frames as an additional disparity constraint. This effectively makes these constraints three-dimensional, linking temporally separated pixels, as shown below:

$$w_t(\mathbf{x}) = w_{t-1}(\mathbf{x}) \tag{4.12}$$

Multiview generation for remapped disparity. Our final goal is to generate multiview content related to the new, optimized disparity maps. This is done in a 2-step approach, as outlined in Figure 4.8. The first step maps the input

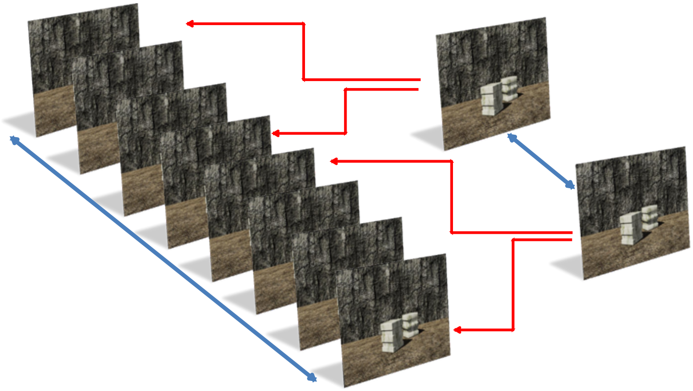


Figure 4.8: *A pair of input figures is warped to a set of multiview results. Notice that the results are now mapped to a completely new disparity range. The original figures are not among the results, which are created by interpolating between two warps, represented here with red arrows.*

images to new virtual images corresponding to the optimized disparity. The second step then does the actual interpolation. This distinction is only conceptional. In practice both warps are done at once. Typically the overall disparity range from leftmost to rightmost view of an MAD is larger than the disparity range of the input stereo pair. Our optimized disparity maps carry the information about this necessary expansion of the overall disparity range, which is illustrated by the blue arrows in Figure 4.8. Within the expanded range we then linearly interpolate from left and right input view as illustrated by the red arrows in Figure 4.8. The disparity range between each image pair of the resulting multiview image set is then a fraction of the input disparity range. Such expansion of the overall disparity range with intermediate view rendering would not be easily possible with DIBR, due to dis-occlusions, which require in-painting. For this reason we use IDW instead, which does not create dis-occlusions and can handle disparity range expansions without noticeable artefacts. This step is an extension of the algorithm presented by Lang and colleagues [2010].

Assume we are generating the first set of multiview images based on the left input image only. In the first step, the adjustments to the input image according to the disparity change have to be determined. To achieve this, we compute a first warp $w_{\text{ext}}(\mathbf{x})$ using the disparity map $D_{\text{ext}} = D_{\text{I}} - D_{\text{L}}$. This warp then describes the transformation of the left image to its adjusted new left image that corresponds to the disparity map D_{L}. In a second step, a warp $w_{\text{cen}}(\mathbf{x})$ is computed that determines the transformation from the left input image to the center view between the left camera and right camera.

Both warps $w_{\text{ext}}(\mathbf{x})$ and $w_{\text{cen}}(\mathbf{x})$ can then be used to compute warps $w(a)$ that transform the left input image to a first set of multiview images

$$w(a) = a w_{\text{ext}}(\mathbf{x}) + (1-a) w_{\text{cen}}(\mathbf{x}) \quad \text{for } a = [0..1] \tag{4.13}$$

whereas $a = 0$ corresponds to the left most image, and $a = 1$ corresponds to the center image. The second set of multiview images can then be generated in the same manner based on the right input view.

4.4 Experiments and Results

We evaluated our pipeline on a variety of synthetic and filmed stereoscopic video sequences. For synthetic scenes, we use ground truth disparity maps and saliency maps rendered from object annotations. This allows the artist to decide which objects should retain as much depth as possible by assigning an importance value to these objects. The importance values are then rendered into a saliency map. For the filmed scenes, we either use automatically generated depth maps [Zilly et al., 2014] (Musicians, Band, Poker) or computed and additionally hand-tuned depth maps [Wildeboer et al., 2010] (Ballons, Kendo). All filmed scenes use an extended version of a contrast-based saliency algorithm [Perazzi et al., 2012] that employs an edge-aware spatiotemporal smoothing [Lang et al., 2012] to achieve temporal consistency. Most steps of our pipeline are implemented in Matlab, only the actual warp rendering to generate the interpolated views has been implemented in OpenGL in C++. Multiview image sequences can be generated in 15 - 600 seconds per frame, depending on the input size and resolution of the image warp grid.

For all scenes, we evaluated the simple linear mapping and our saliency-based mapping. The view to view disparity range for our target display is determined using the results of our user study as ±5 pixels. Figure 4.9 shows anaglyph results accompanied with the associated disparity maps and histograms. Our method clearly enhances the depth for the salient image regions, and effectively compresses less salient image regions as well as empty disparity ranges. The results generated using our method show rounder, more voluminous objects, and are thus able to convey a deeper depth experience even for such small depth ranges. Figure 4.10 shows generated multiview images for additional scenes. As can be seen, our adapted warping method is able to hide distortions in visually unimportant regions, and avoids distracting artifacts even for scenes with inaccurate estimated depth maps.

Figure 4.11 shows a comparison between linear mapping, our mapping algorithm, and another perceptually-based disparity compression algorithm [Didyk et al., 2012b]. Linearly mapping the input range results in flattening the whole scene uniformly, which results in loss of depth perception and cardboarding. Both our method and the method of Didyk and colleagues [2012b] compresses to the same overall disparity range, but provide more depth perception. In contrast to our method, Didyk's method uses a perceptual model for noticeable differences based on disparity, luminance and contrast, whereas our model focuses on salient image regions. While both methods lead to an increased depth perception, our method enhances the depth on the front-most persons better while flattening less salient parts. Didyk's method on the other hand is able to retain "just enough" disparity to perceive depth uniformly across the image.

While our method generates improved results compared to a simple linear mapping, there are also some drawbacks. First, our method relies completely on saliency and will not be able to produce improved results if the saliency computation fails. Fortunately, our method will fall back to a simple linear mapping in the worst case, due to our compression safeguard. Second, our method is computationally expensive and not yet ready for real-time applications. In addition, our rendering method tries to minimize distortions by distributing the error over possibly large, unimportant backgrounds. As a result, our constraints might lead to a small jump between the two middle views, which could be resolved at the expense of other artifacts.

4.5 Subjective Validation

We validated our method using subjective testing, where we showed multiview video content on an 8-view 47″

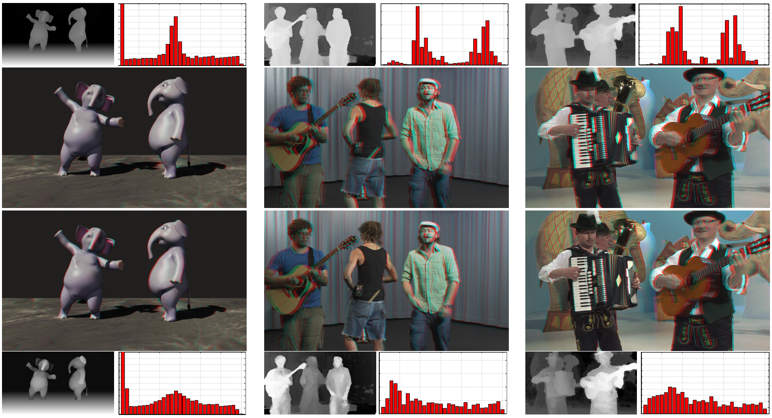


Figure 4.9: *Comparison between linear mapping (top) and our saliency based mapping (bottom), shown in anaglyph with their associated disparity map and histogram. Our method retains more depth volume for the important parts of the scene while flattening out less important parts as well as empty space. Our mapping effectively creates more apparent depth within the same overall depth limits.*

Alioscopy display to our subjects. Our stimuli comprised of result pairs using naïve linear mapping as well as our method, presented in random order. In total, 7 video sequences where displayed. After watching the two stimuli in each trial, the subjects were queried on (i) which of the two stimuli has more depth, and (ii) which one has more artifacts. Our validation experiment had 20 participants naïve to the purpose of the study.

Figure 4.12 shows the responses of each subject averaged over the video scenes. The top figure shows that among the tested subjects there was a strong opinion that our results have more depth. Pearson's chi-square goodness-of-fit analysis demonstrated a statistically significant opinion that our

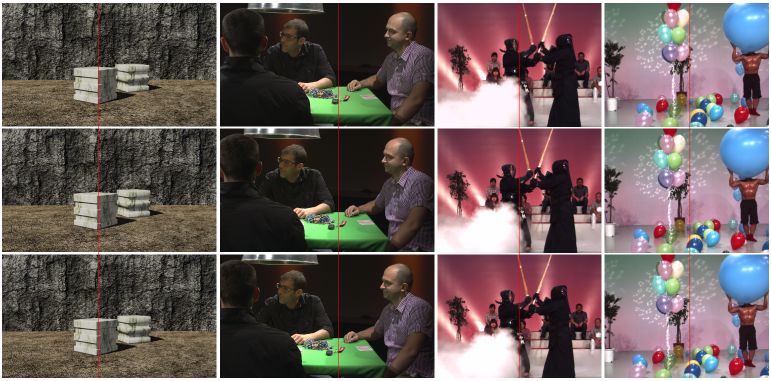


Figure 4.10: *Three views for one frame, generated using our pipeline. Despite the challenging disparity maps, our method is able to hide distortion in visually less important regions and is able to generate novel views without many noticable artifacts.*

method has more depth, $\chi^2(1, 140) = 37.03$, $p<.01$. In total, 76% of test subjects stated our method to have more depth. The bottom of Figure 5.8 shows the result for the second question. There was no statistically significant preference, $\chi^2(1, 140) = 1.83$, $p>.1$. Among all votes 56% indicated our method has more artifacts, and 46% indicated that the naïve mapping had more artifacts.

We performed *Anova* analysis to determine if there is a main effect due to either subjects or video sequences. For the depth assessment task, the p values were found 0.9717 for subjects and 0.3036 for video sequences, indicating that both factors do not have a significant effect on our results (both $\gg$ 0.05). Same was found to be true for the artifact assessment task, where the p values were 0.8361 and 0.7352 respectively.

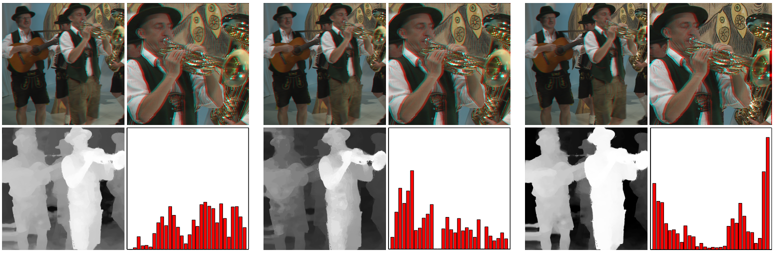


Figure 4.11: *Comparison between linear mapping, our saliency-based mapping, and the mapping of Didyk et al. [2012b] (from left to right). Our mapping is able to increase the perceived depth best, while flattening out lesser important regions. Didyk's method on the other hand retains a noticeable depth difference across the image. Notice the carboarding effect happening in the insets on the left and right.*

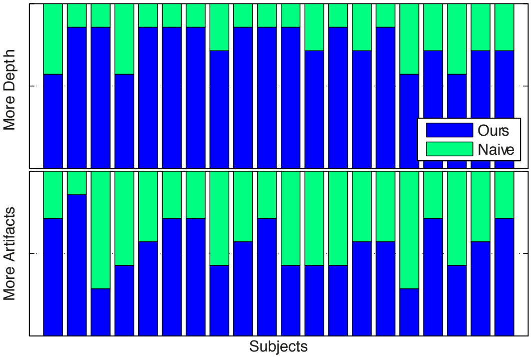


Figure 4.12: *Our validation study revealed a strong opinion among the tested subjects that our method resulted in more perceived depth compared to the naïve mapping (top). The study also showed no clear trend on either methods producing more artifacts than the other.*

In conclusion, our subjective data shows that our method consistently produces results with more perceived depth compared to the naïve mapping, without causing a significant difference in image quality.

4.6 Conclusion

We presented a saliency-based stereo-to-multiview conversion method that generates optimized content for autostereoscopic multiview displays. In a first step, we perform a global disparity mapping that flattens out unimportant regions while trying to retain important image regions. In a second step, we locally enhance the disparity gradients for visually salient regions. Finally, we employ an extended image domain warping algorithm to render the output views according to the modified disparity maps.

As shown in our final validation study, our method clearly improves the amount of perceived depth compared to a simple linear mapping. Compared to other state-of-the-art methods, our approach is more faithful and retains the artistic intent. In addition, our extended image domain warping is robust to temporally unstable and inaccurate disparity maps. In our initial user study we validated theoretical limitations on disparity ranges of autostereoscopic displays, while showing that those can be relaxed in practice to some extent. Nevertheless for conversion of typical stereoscopic input, significant disparity remapping is necessary.

CHAPTER 5

Stereo from Shading

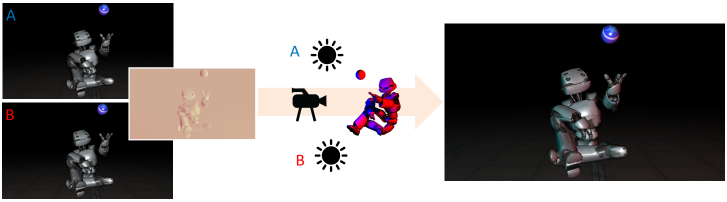


Figure 5.1: *Stereo from Shading: the anaglyph image on the right is assembled from a single view (center) with no parallax disparity, but with different shading (left), producing a 3D impression. The inset represents the difference between eye renders, with the background color showing no difference.*

5.1 Introduction

In natural perception, the human visual system combines several depth cues in order to create a sensation of depth. These include the binocular cues of convergence and disparity, as well as a variety of monocular cues, such as motion parallax, interposition, occlusion, size, perspective, accommodation, and others. Stereoscopic 3D (S3D) as enabled by stereo 3D displays, supports binocular cues and from that creates a sensation of depth. Presenting a separate view to each eye does not, however, perfectly emulate real life depth perception so limitations such as the vergence-accommodation conflict [Shibata et al., 2011a] and other imperfections remain problematic. Furthermore, all 3D displays are limited to a feasible depth volume. In particular, autostereoscopic displays are still very restricted in the depth range they can reproduce without artifacts due to limited angular resolution. Therefore, enhancement of S3D perception has been a very active field of recent research, which has been focused mainly on disparity enhancement [Didyk et al., 2012b; Chapiro et al., 2014b].

In this paper we explore the novel use of lighting variations between eye renders as a means to enhance S3D. We are inspired by Puerta [1989], who showed that a 3D illusion is created when images with no parallax but different cast shadows are displayed stereoscopically.

The goal of our new method, *shading stereo*, is to generate an increase in the S3D sensation by leveraging differences in shading between views, with cast shadows left unchanged between renders. Our algorithm does not change the position of light sources [Medina Puerta, 1989], but rather modifies the vertex normals of selected objects in a scene. This method, not previously used in 3D rendering, allows us to choose where in a scene to add shading stereo and to apply the method to complex scenes with multiple and varied light sources. Our main contributions are:

- The use of variable shading between eye renders as a tool to create or enhance S3D.
- An algorithm that applies shading stereo to scenes with arbitrary lighting by manipulating normals rather than lights.
- The application of shading stereo to live action scenes in scenarios where relighting is feasible.

All examples in this chapter can be viewed with anaglyph glasses 👓 (red-l, cyan-r). The supplemental images can be optimally viewed using color neutral glasses (such as time-multiplexed or polarized variants) if available, or alternatively anaglyph glasses.

5.2 Shading Stereo

Figure 5.2 (top) shows a traditional S3D camera setup, with light positions static between views and two shifted cameras. Simply shifting light sources between views as in Figure 5.2 (middle) would cause the same object to cast shadows in different directions because of varying illumination (see Figure 5.3-A). Furthermore, since the lighting of the scene as a whole is changed between renders, some objects with complex geometries may acquire disturbingly different colors that are hard to fuse stereoscopically. Reflections and refractions may be affected, and multiple lights, or lights that are not conveniently located near the camera, would also be difficult to handle.

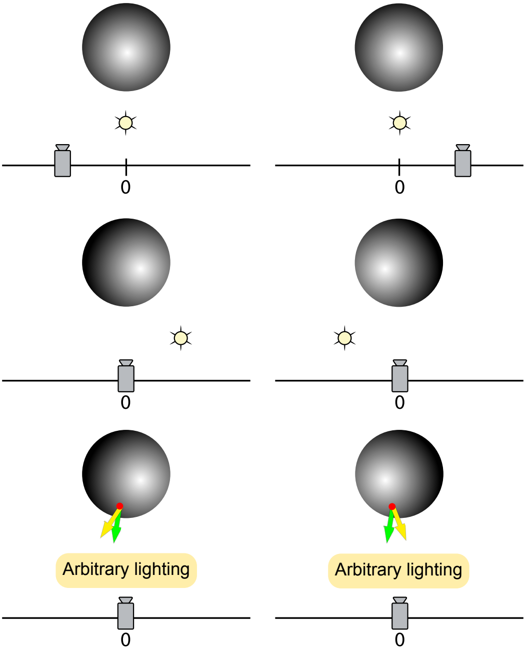


Figure 5.2: *Top: stereo from disparity. Bottom: shading stereo modifies normals to simulate shifting lights.*

Our solution to this problem, illustrated in Figure 5.2 (bottom), consists of shifting the edits from the scene illumination to specific objects in the scene. In particular, since Lambert's cosine law states that the diffuse reflection off a point is directly proportional to the cosine of the angle of incident light to the surface normal, we can directly edit the normals at all shade points of an object so that the illumination from a light positioned at 0 matches exactly that of a shifted light by employing an appropriate rotation. Now, arbitrary lighting of the scene can be employed, as the new lighting will affect the rest of the scene normally, while the object of shad-

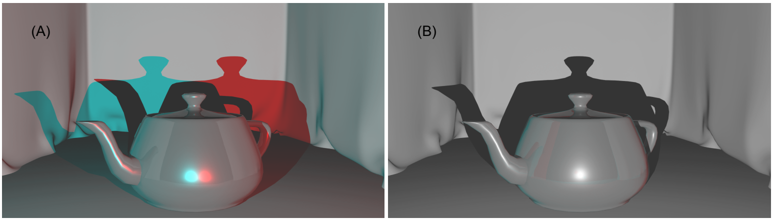


Figure 5.3: *Left: rendering S3D with shifted lights. Right: shading stereo is applied to the teapot only.*

ing stereo will have modified shading. If the scene is lit by a single point light positioned at the origin (in view space), the new method on the edited object will match the shifted lighting described in the previous paragraphs exactly.

We begin by defining the 3D camera position $\mathbf{c}$ and two reference points $\mathbf{s}_l$ and $\mathbf{s}_r$ that are shifted symmetrically from $\mathbf{c}$ along the camera's horizontal axis. For each eye, we use one of these reference points as appropriate, and denote the reference point used for the current render is $\mathbf{s}$.

Given a surface $\mathcal{M}$, for each point $\mathbf{x} \in \mathcal{M}$, we define $\vec{v}_\mathbf{c} = \mathbf{c} - \mathbf{x}$ and $\vec{v}_\mathbf{s} = \mathbf{s} - \mathbf{x}$ and perform the following operations:

- The axis of rotation is defined as $\vec{a}_\mathbf{x} = \vec{v}_\mathbf{c} \times \vec{v}_\mathbf{s}$.
- If the original normal of $\mathbf{x}$ is $\vec{n}_\mathbf{x}$, the angle by which the normal is rotated is defined as $\theta_\mathbf{x} = \theta_\mathbf{xc} - \theta_\mathbf{xs}$, where $\theta_\mathbf{xc} = \arccos(\vec{n}_\mathbf{x} \cdot \vec{v}_\mathbf{c})$ and $\theta_\mathbf{xs} = \arccos(\vec{n}_\mathbf{x} \cdot \vec{v}_\mathbf{s})$, i.e., how different the angle between the vertex normal and the camera is from the angle between the vertex normal and the reference point. All vectors above are considered normalized, so a dot product is the cosine of the factors.

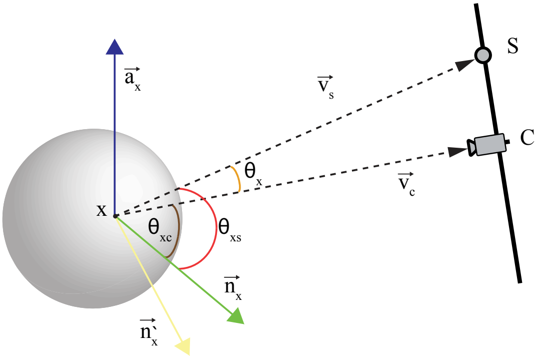


Figure 5.4: *Shading stereo: the normal of* **x** *is adjusted to match the angle of a reference point.*

- The new normal $\vec{n}'_{\mathbf{x}}$ is the result of rotating $\vec{n}_{\mathbf{x}}$ around the axis $\vec{a}_{\mathbf{x}}$ by an angle of $\theta_{\mathbf{x}}$.

This operation is performed for each point, thus obtaining the new normals. Figure 5.4 illustrates these operations. We assume that a shading model with a view-independent diffuse and a view-dependent specular component (such as Phong) is used, and shading stereo is applied to the diffuse component. As for S3D rendering in general, the problem remains of handling specularities and reflections, which often produce incorrect depth sensations. We render specularities separately and adjust them to what is defined as a "flat" setting by Templin et al. [2012].

Finally, while the procedure described above requires knowledge and manipulation of lighting and normals, it is possible to obtain S3D effects in live-action sequences through image relighting (see the work of von der Pahlen [2014]). As a proof of concept, we demonstrate shading stereo on scenes where geometry was known a priori. We overlay the shaded geometry on the original camera views, thus generating a 3D sensation when viewed stereoscopically. The shading from an additional virtual light can thus be used to generate the shading stereo effect (see Figure 5.5). Alternatively, the lighting of a scene could be emulated in a rendering environment for a better match between shading and image.

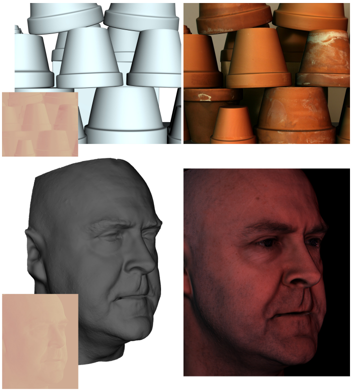


Figure 5.5: *(L) geometry proxy re-lit using our method (left eye view) (R) anaglyph image showing stereo from shading. The insets represent the difference between left and right eye images for each example.*

5.3 Feasibility Study

Our method follows the premise that S3D sensation can be generated or increased by simply rendering the left and right views with different shading. However, only a small subset of all possible variations of illumination between both eyes will cause a plausible depth sensation. For example, we observed that the retinal rivalry caused by an exaggerated difference in shading generates discomfort. We therefore conducted a perceptual experiment to explore the feasibility of using shading stereo for S3D without causing disturbing retinal rivalry.

Stimuli: Scenes were presented on an Alienware 2310 23", 3D capable display with the help of time-multiplexed glasses. Participants were seated at a comfortable distance of about 60 centimeters from the screen. The virtual camera setup attempted to emulate natural viewing by placing cameras on a scene 60 centimeters away from the stimuli and 6 centimeters away from each other, so that the virtual stimuli would have 3D characteristics of size and shape similar to a real world objects (orthostereo).

The rendering setup consisted of a stereo camera pair oriented in parallel along the depth axis, with the resulting images re-converged, thereby displaying the center plane of the stimuli with null parallax disparity. We consider this to be the unit baseline. Given a point with coordinates (x, y, z) and parallax disparity β, if the interaxial distance is reduced to $\alpha \in [0, 1]$, the disparity will change to be $\alpha \cdot \beta$. As such, depth can be controlled proportionally in our rendering system for experimental purposes, so α can be thought of as the depth compression factor.

Since the reference points can be placed arbitrarily far from an object, an absolute distance such as that used for cameras would not be suitable for parametrization. We therefore parameterized the difference in shading by angle θ, the angle between the reference points and the center. Given a maximum angle (in our experiments, $\pi/6$), we introduce a light disparity $dLighting \in [0, 1]$ so that the angle between the reference points is always $dLighting \cdot \theta$.

Method: Fifteen naive volunteers (3F, 12M, ages 24–35,

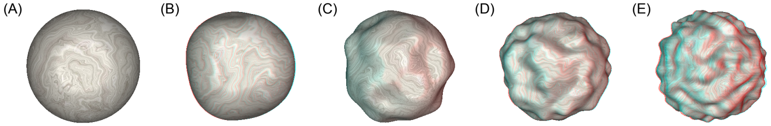


Figure 5.6: *Stimuli shown in the experiment, where* (α, *dLighting*)= *A:* $0.0, 0.0$, *B:* $0.2, 0.0$, *C:* $0.0, 0.4$, *D:* $0.2, 0.6$, *E:* $0.6, 0.6$

with normal or corrected-to-normal vision) participated in the experiment. We displayed simple geometric models similarly to Ramanarayanan and colleagues [2007], ranging from flat to bumpy (Fig. 5.6), at compression ratios of $\alpha \in \{0.0, 0.2, 0.4, 0.6\}$. We hypothesized that irregularities in the shape would generate more pronounced self-shading and hence seem more 3D. Users were given control of the parameter *dLighting* using a method-of-adjustment procedure. They were tasked with setting the following three light disparity measures in turn: (i) the smallest *dLighting* where they perceived a just noticeable difference in depth, compared to the presented α baseline; (ii) the smallest *dLighting* where they first perceived retinal rivalry; and (iii) a preferred *dLighting* setting between the two values found above.

We performed a $5(\mathit{model}) \times 4(\alpha) \times 3(\mathit{measure})$ Repeated Measures Analysis of Variance (ANOVA) on participants' responses (see Fig. 5.7) and found a main effect of measure ($F(2, 28) = 156, p \approx 0$). Newman-Keuls post-hoc tests of differences between means showed that all three measures were significantly different ($p < 0.05$). We found no significant main or interaction effects of α or model. This simple experiment demonstrates that shading stereo does affect the perception of depth, within a limited range, but further studies are required to fully evaluate and quantify the effect.

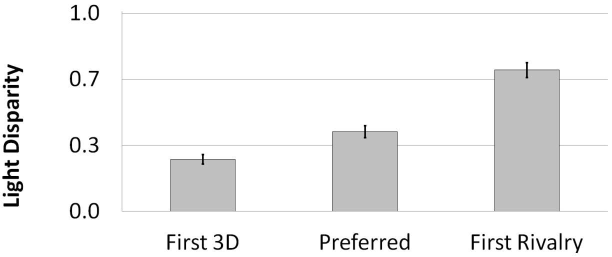


Figure 5.7: *Results of our feasibility study to determine acceptable perceptual limits for shading stereo. Error bars show standard errors.*

5.4 Perceptual Validation

In addition to the work presented in Section 5.3, we performed two more perceptual studies to explore some basic questions about shading stereo. We will present a brief outline of these studies here in an informal manner. These results were not included in our previous work detailing shading stereo [Chapiro et al., 2015b].

5.4.1 Pseudo shading-stereo

A possible concern regarding the use of shading stereo is that our results could be attributed to color contrast rather than a truly stereoscopic effect. Medina Puerta [MP89] tested a similar hypothesis by reversing right and left images, and found that subjects reported that the S3D sensation in the original condition disappeared in this "pseudo-stereo" scenario. To test whether this also holds for shading stereo, we repeated the experiment presented in Section 5.3 with 9 additional subjects, but this time the virtual shading points were swapped (i.e., the shading point intended for the left view was used for the right view and vice versa). A 3-way (5 Model X 4 Disparity X 3 Measure), between-groups ANOVA with categorical predictor StereoType (Real, Pseudo) found a main effect of StereoType ($F(1, 21) = 11.99$, $p < 0.005$), and an interaction of StereoType with Disparity($F(3, 63) = 3.3943$, $p < 0.05$). The reason for these effects is as follows: shading stereo was perceptible earlier for the Real stereo condition than for the Pseudo stereo one, and was equally salient across disparity levels (including α=0). However, in the Pseudo stereo condition, values were set slightly higher than Real for the no-disparity case, but significantly higher for all three cases with disparity.

We believe that shading stereo enhanced the impression of depth in all cases, whereas the pseudo stereo behaved in the opposite manner. We suggest that, as in Medina's work, this difference in performance shows that participants perceived a true stereo effect with shading stereo, rather than a simple contrast enhancement (in which case the Real and Pseudo cases should have similar results). Indeed, let us assume that the effect was due to a monocular contrast cue. Swapping eye views should maintain identical contrast characteristics, but our results suggest a difference between the standard and "swapped" modes. Due to the constraints of the original study presented in Section 5.3, subjects were forced to give a value at which a 3D sensation is achieved. Because the swapped imagery resulted in an unnatural binocular stereo cue, subjects continued to increase light disparity until binocular rivalry became obvious, at which point it was interpreted as a 3D sensation. One might then wonder if binocular rivalry could be responsible for the totality of the results of the Feasibility study: but this cannot be so, as swapping the shading points provides observers with the same color differences, and thus the same rivalry, but results were found to be very different between standard and pseudo conditions.

5.4.2 Validation Experiment

Another important question for shading stereo can be formulated as follows: will it work with complex scenes? What about complicated (concave?) shapes? How about materials with different colors, reflective and transparent materials? Naturally these questions can be presented for any stereoscopic method. In fact, for many of these questions the state of the art in S3D perception is vastly insufficient to provide a satisfactory answer in any case. In this chapter, shading stereo is presented as a novel technique, and we are fully aware that sometimes these and many other questions must go unanswered in a first visit of a new topic in S3D perception. We performed a simple experiment with the goal of determining whether observers could perceive a difference in 3D profiles between objects enhanced by shading stereo and unenhanced references.

13 naive volunteers (4F, 13M, ages 24-36) participated in a *Validation* experiment to assess the shadow stereo effect on three scenes with simple lighting: Dragons (as shown in Fig. 5.10), Sumo and Elephants (variants with single light and simple background, available in supplementary). We used a two alternative forced choice (2AFC) procedure, where at each trial participants selected which of a pair of stereoscopic images had more depth. The two images were presented one at a time, giving subjects the ability to switch back and forth (although a blank image was shown briefly when images were switched). Each pairwise comparison involved one stereoscopic image that relied *only* on parallax disparity ($l = 0.0$) for stereopsis, and another augmented

by *light disparity* ($l \in \{0.0, 0.2, 0.4, 0.6\}$). Both images had the same *parallax disparity* ($d \in \{0.0, 0.2\}$). An additional 17 participants (3F, 14M, ages 24-36) did the same task with three scenes with complex lighting: Antenna, Elephants and Sumo (as shown in Fig. 5.10). These scenes were lit by, respectively: environment lighting; three area and two point lights; a directional light, two spotlights, two area lights and four point lights. Four participants took part in both trials.

The results are shown in Fig. 5.8. A Repeated Measures Analysis of Variance (ANOVA) found a main effect of disparity, with $d = 0.0$ having a significantly stronger effect from shadow stereo than $d = 0.2$ ($F(1, 29) = 16.939$, $p < 0.01$), which was not found in the Dial experiment, probably because of the less complex scenes. We also found a main effect of light difference ($F(3, 87) = 6.011$, $p < 0.01$), with post-hoc tests showing that $l = 0.0$ was significantly different from $l \in \{0.2, 0.4\}$, but not from $l = 0.6$. This could be due to the highest light disparity value producing rivalry that broke the depth sensation, as it is close to the threshold of disturbing rivalry found in the *dial* experiment. Overall, participants preferred images augmented with shadow stereo to their corresponding versions without shadow stereo. When $l = 0.0$, the stimuli presented are identical and the results were at chance performance. Our results therefore suggest that shading stereo could in fact be used effectively to create and enhance depth perception compared to unprocessed references, and that the effects are more noticeable when applied to images completely devoid of disparity, but still incremental for shallow depth ranges.

At this point, a note on the metodology could be helpful. In particular, two points should be explained. Firstly, why have values of $\alpha = 0$ and 0.2 been tested, but not further values? This questions has a simple answer - in our experience adding parallax disparity to the scene reduced the effectiv-

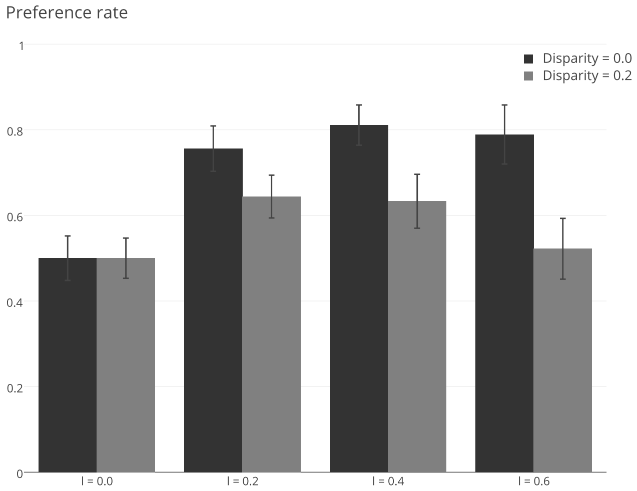


Figure 5.8: *Results of the Validation experiment. Error bars show standard deviations. Preference rate reflects the rate at which the imagery enhanced by shading stereo was chosen over the unenhanced version.*

ity of shading stereo. Our goal was to test a few hypothesis that we considered most important, but practical time concerns did not allow us to test the full spectrum of possibilities of shading stereo. Knowing that the shading stereo effect tended to disappear in the presence of the stronger binocular cue, we limited our test cases to *no disparity* and *some disparity*. Another question is this: why has the relative strength of shading stereo as compared to disparity not been measured? The reason for this is somewhat more complex. While shading stereo provides a semblance of 3D, it is not enough for a fully stereoscopic scene. A good example would be to compare it to other stereoscopic cues, such as linear perspective. If shown side-by-side, it is unlikely that subtle changes in perspective would be matched to specific binocular disparity variations by observers. Instead, both scenes would be perceived as simply different. The same is true for shading stereo. For this reason, images in our *Validation* study are not presented simultaneously, but one at a time, and participants are required to maintain a mental model of their stereoscopic depth sensation.

5.5 Results and Discussion

We have explored the feasibility of using a lesser known stereo cue to create or enhance depth perception. Our algorithm is easy to implement and does not add significant complexity to the rendering process and can be applied to both images and video. The effect can be tuned and adjusted flexibly on a per-object or even per-material basis and can be used on scenes with arbitrary lighting.

Figure 5.10 reveals some interesting features of shading stereo. Firstly, the first and second columns show that shading stereo creates a noticeable depth illusion which is clearly visible when compared to the original 2D image. The third column shows a version of the content with no shading stereo, but with a small amount of parallax disparity ($\alpha = 0.2$). Note that the depth range of these examples corresponds approximately to the depth reproduction limits of current autostereoscopic displays. Finally, the third and fourth columns show that using shading stereo in combination with disparity enhances the depth illusion.

The last comparison above hints at interesting practical uses for shading stereo. While in general the depth range that can be generated by shading stereo without causing discomfort is significantly more limited than disparity, shading stereo can still be used to enhance the depth sensation on depth limited display devices (e.g. autostereoscopic displays) to go beyond the devices' capabilities. Furthermore, shading stereo images without disparity can be viewed without glasses and still provide reasonable 2D image quality. This enables backward compatible stereo applications, where the same image can be viewed in both 2D (without glasses) and 3D (with glasses) with good quality.

Our method has several limitations. Firstly, views have to be re-lit, which requires access to the renderer or image re-lighting methods. If geometric approximations of a scene are used, some degree of precision and alignment is required. Excessive use of light disparity causes retinal rivalry, and while we found some thresholds in this paper, complex scenes might have significant variation (see Fig. 5.9). A failure case where geometry is estimated poorly is also shown.

Finally, we have presented an initial exploration of shading stereo as a factor for 3D perception, but many questions remain. Further studies are required to determine under which conditions shading stereo works best, how shading stereo compares to disparity, what kind of lighting, geometry, and materials should be used, and other interesting directions for future work.

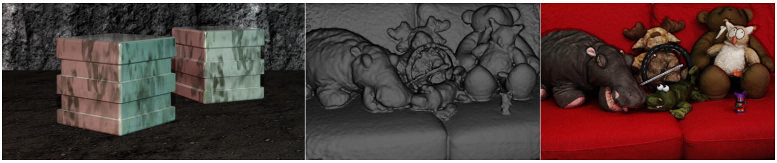


Figure 5.9: *Objects with sharp piecewise-planar features can sometimes generate color contrasts through color variation quickly, leading to a disturbing viewing experience. Additionally, planar objects will sometimes not benefit from shading stereo, as self-shading might not change significantly with light disparity (L) Faulty geometry (M) for live-action scenes can lead to misplaced shading, degrading image quality (R)*

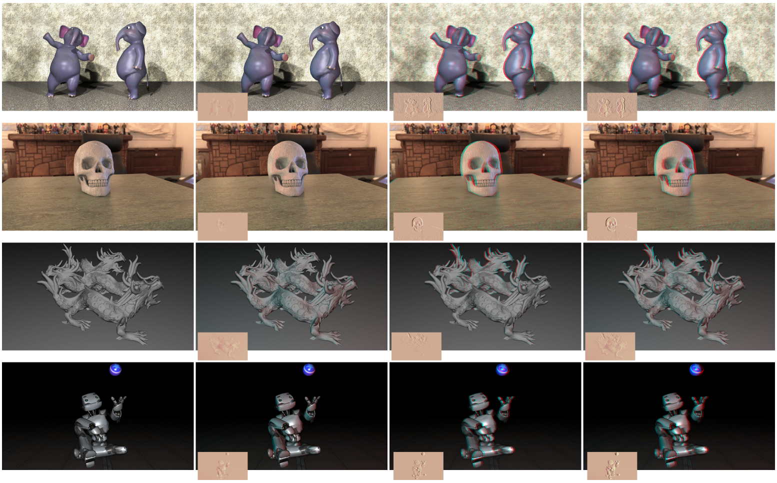


Figure 5.10: *Each set shows, from left to right: (i) no disparity and no shading stereo, (ii) some shading stereo, but no disparity, (iii) some disparity, but no shading stereo, and (iv) the same disparity, augmented by shading stereo. The insets represent the differences between the left and right eye images for each image.*

CHAPTER 6

Continuous Dynamic Range Video

6.1 Introduction

After years of research and development, we are finally about to witness the emergence of High Dynamic Range (HDR) content distribution and display at the consumer level. While high-end cameras (such as the Red Epic Dragon, SonyF55 and F65, and ARRI Alexa XT) have been able to natively capture HDR video, up to now displaying HDR content has only been possible through research prototypes or custom built hardware. This landscape is rapidly changing as TV manufacturers including LG, Sony, Samsung, Panasonic and TCL have announced HDR displays with various peak luminances and black levels, which they

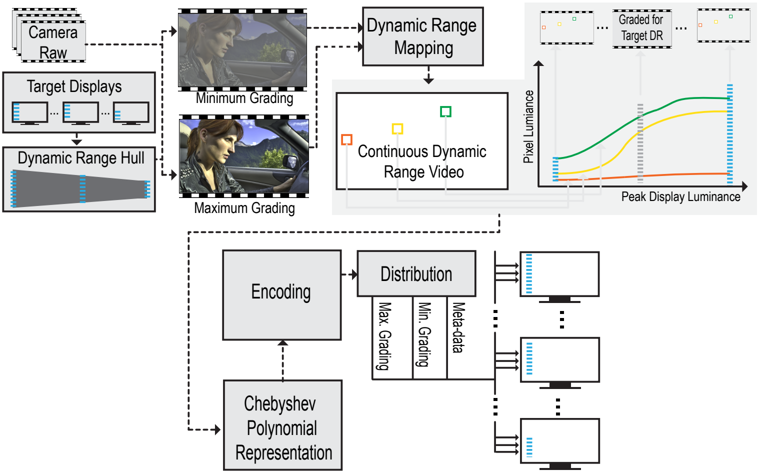


Figure 6.1: *We present an end-to-end pipeline for efficient content creation and distribution of HDR content for a multitude of target display dynamic ranges. Our workflow starts by color grading the source content for the largest and smallest of the target dynamic ranges among the set of targeted dynamic ranges. Next, through an interactive "dynamic range mapping" we obtain "continuous dynamic range" (CDR) video, where each pixel contains an art-directable function rather than a scalar value. We approximate this CDR video using Chebyshev Polynomials, and encode it for efficient distribution to the target displays.*

plan to release in 2015. On the content creation side, Technicolor and the Sinclair Broadcast Group successfully demonstrated over-the-air broadcast of UltraHD HDR content and Technicolor now offers HDR grading services. Netflix and Amazon announced the upcoming start of HDR content streaming services. In the meantime, experimental HDR short films [Lukk, 2014; Schriber, 2014] explored the creative use of HDR imaging in film making.

These efforts towards realizing an HDR content production and distribution pipeline from capture to display are fueled by the massive difference that HDR makes in viewing experience [Hanhart et al., 2014]. The significance of the increase in experience quality provided by HDR over *standard dynamic range* (SDR) is becoming widely accepted as common knowledge. As a result, the focus point of the next generation viewing experience is shifting from *more* pixels to obtaining *better* pixels by extending their dynamic range, among other factors.

The emergence of HDR displays from multiple vendors with different dynamic ranges creates some significant challenges for content production and distribution. Specifically, the *production challenge* is tailoring HDR content to a number of upcoming displays which are announced to have peak luminances ranging from 800-4000 nits, as well as future HDR displays with different dynamic ranges. The straightforward approach of grading content for each specific display dynamic range does not scale well due to the required additional manual labor. Methods proposed in literature, such as display adaptive tone mapping [Mantiuk et al., 2008], can alleviate this issue, but do not allow for precise artistic freedom in the expression of brightness variations.

The *distribution challenge* is the task of efficiently coding and transmitting a large number of HDR streams graded for different display dynamic ranges. Previous work proposed distributing a single HDR stream efficiently as a residual signal over the SDR content [Mantiuk et al., 2006a]. This approach, however, is not well suited for application in the emerging landscape where numerous HDR streams are required to be transmitted simultaneously.

The content creation and distribution challenges force us to rethink the way raw source content is graded for a multitude of target displays, and how graded content can be efficiently represented. In this work we propose a new content creation paradigm which we call *Dynamic Range Mapping*, where raw source content is graded not only for a single display (as in traditional tone mapping), but for a dynamic range continuum that entails the dynamic ranges of an arbitrary number of target displays. Unlike tone mapping where the resulting pixels have scalar luminance values, dynamic range mapped pixels are defined by art-directable functions of display dynamic range. We call this new data structure *Continuous Dynamic Range Video* and propose a method for its efficient representation and distribution. Specifically, our work makes the following contributions:

- A practical dynamic range mapping workflow, allowing the creation of continuous dynamic range video with full artistic control.
- An efficient representation of continuous dynamic range video using a polynomial series approximation.
- A demonstration of efficient encoding of continuous dynamic range video.

The individual components presented in this paper form an end-to-end solution for efficiently creating, representing and distributing content graded for an arbitrary number of target displays with different dynamic ranges. In our solution, continuous dynamic range content is efficiently represented by two video streams (graded for the highest and lowest target dynamic ranges) and an additional meta-data stream that occupies less than 13% of the current standard business-to-consumer video distribution bandwidth.

6.2 Continuous Dynamic Range Video

We propose Continuous Dynamic Range (CDR) as a novel way of representing video within a continuum of dynamic ranges. For practical purposes, it is important that the CDR representation is both efficient and allows full artistic freedom. In this section we will introduce key concepts and components of CDR video and discuss artistic control. The efficient approximation and encoding of CDR video will be discussed in Section 6.3.

6.2.1 Key Concepts

A high-level overview of our pipeline is illustrated in Fig. 6.1. The goal of our method is distributing the input source video to a number of *target displays*, where the grading for each of the target displays can be art directed. The first input to our method is the source video in camera raw format. While the dynamic range of the source video can be arbitrary, in this work we used HDR content with up to 14 f-stops. Formally, we denote a frame of the raw input video as $\mathtt{I}$, its color at pixel p as $\mathtt{I}^p$, and the corresponding luminance as $\mathcal{L}(\mathtt{I}^p)$.

Since the dynamic range continuum encompassed by a CDR video is a superset of the dynamic ranges of all target displays, we also require the user to specify a *dynamic range (DR) hull*. The DR hull defines a dynamic range continuum between the *minimum* and *maximum dynamic ranges*. The minimum dynamic range is bounded by the min peak luminiance and max black level among the set of all target display dynamic ranges. Analogously, the maximum dynamic range is bounded by the max peak luminance and min black level (Fig. 6.2).

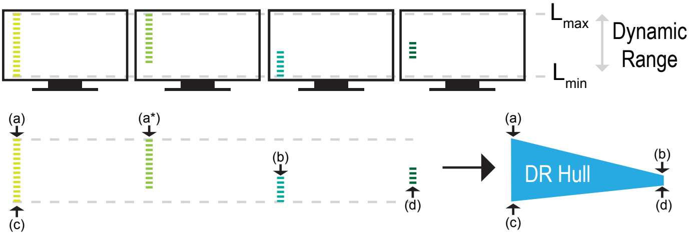


Figure 6.2: *The dynamic range hull is a superset of the dynamic ranges of all target displays. Colored bars represent the dynamic range of a display.*

The first artistic interaction in our pipeline is the grading of the raw content for the minimum and maximum dynamic ranges to obtain *minimum* and *maximum gradings*, which we denote with $\mathtt{I}_\alpha$ and $\mathtt{I}_\beta$. Here, the user has full freedom in terms of tools to be used and edits to be performed, as long as the spatial correspondence between the pixels of the minimum and maximum gradings are preserved. We denote the minimum and peak luminance of $\mathtt{I}_\alpha$ by η_α and π_α, respectively, and the minimum and peak luminance of $\mathtt{I}_\beta$ by η_β and π_β, respectively.

After these traditional grading processes, the next artistic interaction step is *dynamic range mapping* where the user specifies how the pixel luminances change across the dynamic range hull. Functions of the following type have to be generated

$$h^p : [\eta_\alpha, \eta_\beta] \times [\pi_\alpha, \pi_\beta] \to [\mathcal{L}(\mathtt{I}_\alpha^p), \mathcal{L}(\mathtt{I}_\beta^p)], \qquad (6.1)$$

which associate with each pixel p and dynamic range (η, π) a unique luminance value $h^p(\eta, \pi)$. To reduce the complexity of generating these functions and the amount of distributed data, we restrict the domain to $[\pi_\alpha, \pi_\beta]$ and define the associated minimum luminace for any $\pi \in [\pi_\alpha, \pi_\beta]$ by $\eta(\pi) := \eta_\alpha + (\eta_\beta - \eta_\alpha)\frac{\pi - \pi_\alpha}{\pi_\beta - \pi_\alpha}$. Thus, $(\eta(\pi), \pi)\ \forall\ \pi \in [\pi_\alpha, \pi_\beta]$ defines the considered DR hull.

Consequently, at each pixel a user-defined function, which we call a *lumipath*, represents the pixel's luminance value as a function of the peak luminance π of a target display. Formally, we define a lumipath as

$$g^p : [\pi_\alpha, \pi_\beta] \rightarrow [\mathcal{L}(\mathtt{I}_\alpha^p), \mathcal{L}(\mathtt{I}_\beta^p)], \tag{6.2}$$

where π_α and π_β are the peak luminances corresponding to the maximum and minimum dynamic ranges. The end result of the dynamic range mapping process, namely the *continuous dynamic range* video, stores a lumipath at each pixel rather than a scalar luminance value. In our current implementation we transform the graded image pair $\mathtt{I}_\alpha$ and $\mathtt{I}_\beta$ to the CIE YUV color space, where the variation of the Y channel across the dynamic range hull is controlled by the user defined lumipaths, and the chrominance channels are linearly interpolated.

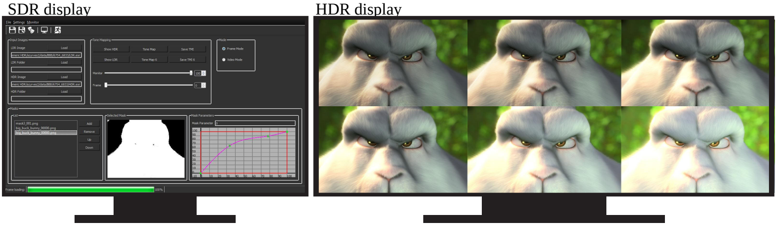
6.2.2 Dynamic Range Mapping

We implemented a user interface (Fig. 6.3) for convenient dynamic range mapping. Users can select desired image regions by using masks and adjust the corresponding lumipaths by modifying the control points of a third degree polynomial spline interface. While we chose this particular representation on the based on the standard tone curve interfaces in commercial color grading software, other tools could be employed to a similar effect.

Formally, given a series of image masks $\mathtt{M}_j$ with values $\mathtt{M}_j^p \in [0, 1]$, the user manually specifies functions $k_j : [\pi_\alpha, \pi_\beta] \to [\pi_\alpha, \pi_\beta]$ with the user interface. When applied to each pixel, the function is modulated at each pixel position by the mask, and k_j^p is obtained as follows:

$$k_j^p(\pi) = \mathtt{M}_j^p\, k_j(\pi) + (1 - \mathtt{M}_j^p)\pi. \tag{6.3}$$

This defines a blending between the artist defined curve and a linear curve based on the weights specified by the mask, allowing for smoothly varying edits. By employing n masks and specifying n such functions, the corresponding lumipaths g^p are obtained by applying all functions successively



LDR Image: eneric HDR/scurves2/data/BBB/6754_6833/LDR.exr
HDR Image: eneric HDR/scurves2/data/BBB/6754_6833/HDR.exr
Monitor: 100
Frame: 0
● Frame Mode
○ Video Mode
mask3_001.png
big_buck_bunny_00000.png
big_buck_bunny_00000.png
Mask Parameter: 0
Frame loading: 100%

Figure 6.3: *This figure shows our luminance grading interface. On the SDR display, masks with values in $[0, 1]$ can be loaded in the bottom left menu, and are displayed in the bottom-middle window. A cubic spline interface shown on the bottom right allows the user to manually input lumipaths. Different visualization options can be selected from the menu on the top. On the HDR display, users can visualize their edits in an interactive manner. To see a standard work session using our interface, we point the reader to the supplementary material of the corresponding publication [Chapiro et al., 2015a]*

(layer based grading) and scaling the result:

$$g^p = \frac{k_1^p \circ \ldots \circ k_n^p - \pi_\alpha}{\pi_\beta - \pi_\alpha} \left(\mathcal{L}(\mathtt{I}_\beta^p) - \mathcal{L}(\mathtt{I}_\alpha^p) \right) + \mathcal{L}(\mathtt{I}_\alpha^p). \tag{6.4}$$

The lumipath $g^p : [\pi_\alpha, \pi_\beta] \to [\mathcal{L}(\mathtt{I}_\alpha^p), \mathcal{L}(\mathtt{I}_\beta^p)]$ is the desired curve defining the luminance of the pixel p for any display with maximum brightness between the two analyzed extremes. This process is illustrated in Fig. 6.4.

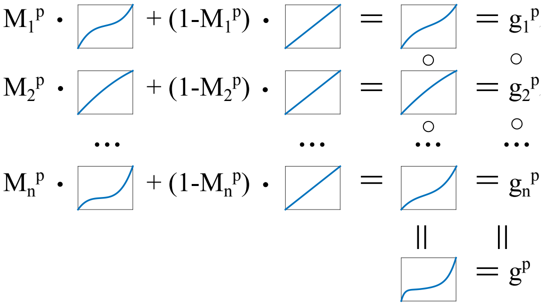


Figure 6.4: *A visual representation of the process of obtaining a numerical lumipath (Eqs. 6.3 and 6.4) is shown here. Lumipaths input by an artist are averaged with linear functions according to the weights specified in the interface and subsequently concatenated to obtain the final per-pixel lumipath g^p.*

In practice, the dynamic range mapping process begins by specifying maximum and minimum gradings to our tool. The user can additionally import multiple masks that can be generated using modern video editing software (e.g. Resolve, Nuke, etc.). Our user interface, which is rendered on a standard LCD display, provides interactive visual feedback on an external HDR display as the lumipaths for the selected region are modified. Visualization is provided by computing and rendering a user defined number of gradings (Fig. 6.3-right). Since (i) there are no restrictions in how the input gradings are obtained (except preserving pixel correspondences), (ii) any number of pixel-level masks for region selection can be used, and (iii) the lumipaths can be defined precisely using any number of control points, our method allows for significant artistic freedom during dynamic range mapping.

6.3 Efficient Approximation and Coding

CDR video in its raw form is represented by a considerable amount of data, where each frame f comprises (i) an LDR image $\mathtt{I}_{\alpha}^{f}$, (ii) an HDR image $\mathtt{I}_{\beta}^{f}$, and (iii) lumipaths $g^{p,f}$ for every pixel of a frame. In this section we describe how we solve the distribution challenge by efficiently approximating and coding CDR video.

LDR and HDR image sequences can be jointly compressed with dedicated coding methods like the work of Mantiuk et al. [2006a] or other methods which are currently the subject of intensive investigations in MPEG [ISO/IEC MPEG, 2015]. In this work, we assume that the image sequences of $\mathtt{I}_{\alpha}^{f}$ and $\mathtt{I}_{\beta}^{f}$ are already encoded. In this section a first exploration of the compressibility of the remaining data - the lumipaths - is performed.

Our compression approach can be subdivided into two parts: we begin by approximating the lumipath functions in a perceptually lossless way using a polynomial series, followed by a representation of the coefficients in an image-like format and encoding using a video compression method.

6.3.1 Approximation

The first step towards the efficient compression of this information is a suitable approximation of the individual lumipath functions. The goal is to find a representation of all lumipaths, which should be both compact and visually indistinguishable from the original. Our approach consists of approximating each lumipath by a series of functions. The series is truncated at a point where the resulting output is visually lossless based on a human visual system model. The result is a representation of each lumipath by a finite set of coefficients with respect to a polynomial basis. These coefficients are later further compressed with the help of a standard video codec.

Our human visual system model consists of a *threshold-versus-intensity (*tvi*)* function that computes an approximate threshold luminance, given the level of luminance adaptation ($\mathcal{L}_a$). The tvi function is computed by finding the peak contrast sensitivity at each luminance level as described in previous work [Mantiuk et al., 2011; Aydın et al., 2008]:

$$\mathrm{tvi}(\mathcal{L}_a^p) = \frac{\mathcal{L}_a^p}{\max_x \left(\mathrm{CSF}\left(x, \mathcal{L}_a^p\right)\right)}, \tag{6.5}$$

where CSF denotes the contrast sensitivity function, and $\mathcal{L}_a^p$ denotes the adaptation luminance for pixel p. To avoid introducing visual artifacts, we make the conservative assumption that the human eye can adapt perfectly to a single pixel p. In practice, we found that even such a conservative threshold estimation can significantly reduce the number of required polynomial basis coefficients. In our experiments, the number of coefficients did not exceed 20.

In mathematical analysis, the *Weierstrass approximation theorem* shows that a continuous real-valued function f :

$[a, b] \to [c, d]$ can always be uniformly approximated by a polynomial series. Approximation by simple functions is desirable because they can be easily computed and evaluated. Several bases of the space of polynomials can be used for such an approximation, but while all of them may converge, not all perform equally well for a given problem.

A common method for approximating functions with a polynomial basis consists of using Chebyshev series [Davis, 1975]. Chebyshev polynomials have some very useful properties that make them desirable for our problem, namely (i) they are guaranteed to minimize Runge's phenomenon when approximating in an interval (this is particularly important since in practice most displays are located near the minimum end of the examined dynamic range hulls), (ii) they can be quickly computed numerically, and (iii) the error of the approximated function as compared to the original can be estimated from the calculated coefficients, which is important as a stopping criterion.

Our goal is to approximate a lumipath $g^{p,f}$ at a given pixel in a perceptually lossless way by a truncated Chebyshev series $\bar{g}^{p,f}$ such that $\left\| g^{p,f} - \bar{g}^{p,f} \right\|_\infty < \text{tvi}(\mathcal{L}_a^p)$, i.e. the deviation is smaller than the threshold computed by our model of the human visual system. The truncated Chebyshev series is represented by

$$\bar{g}^{p,f}(x) = \sum_{k=0}^{N_{p,f}} c_k^{p,f} \psi_k(x) \tag{6.6}$$

where $\psi_k(x)$ is the k-th Chebyshev polynomial, $c_k^{p,f}$ the corresponding Chebyshev coefficient at pixel p of frame f, and $N_{p,f}$ is the smallest degree required to obtain an error $\left\| g^{p,f} - \bar{g}^{p,f} \right\|_\infty$ which is smaller than $\text{tvi}(\mathcal{L}_a^p)$. This defines a perceptually lossless approximation of $g^{p,f}$ which is determined by $N_{p,f}+1$ coefficients $c_0^{p,f}, \dots, c_{N_{p,f}}^{p,f}$.

We implement our computation of the Chebyshev series as outlined by Broucke [1973]. For simplicity we consider normalized lumipaths, i.e. the domain and the range of all lumipaths is scaled such that all of them lie in the standard Chebyshev domain $g^{p,f} : [-1,1] \rightarrow [-1,1]$. This normalization process can be easily inverted based on the provided peak luminances π_α and π_β and the images $\mathtt{I}_\alpha$ and $\mathtt{I}_\beta$.

Note that since each basis polynomial ψ_k has a domain $\mathcal{D} := [-1,1]$ and its range $\psi_k(\mathcal{D})$ is also a subset of $[-1,1]$, the total $\|g^p - \bar{g}^p\|_\infty$ error of the approximation is bounded by the sum of the absolute values of the infinite remaining coefficients of the series. When approximating a function with m continuous derivatives on $[-1,1]$, the approximation error of a Chebyshev series truncated at n elements has a convergence rate of $\mathcal{O}(n^{-m})$ when $n \rightarrow \infty$ [Gil et al., 2007]. As such, when operating on "well-behaved" functions, a common stopping criterion is given by the sum of the absolute values of a small number of elements. In practice, our algorithm truncates the series when the absolute sum of the next three elements is below the allowed error threshold. An example of an approximation of a function by Chebyshev polynomials of different orders is illustrated in Fig. 6.5.

In our unoptimized Matlab implementation, computing lumipaths for every pixel of a FullHD image takes approximately 3-5 seconds. Decoding this information to reconstruct the represented functions for all pixels takes an additional 0.1-1 seconds. Importantly, this computation could be significantly sped up through parallelization as each pixel is independent of the rest of the image.

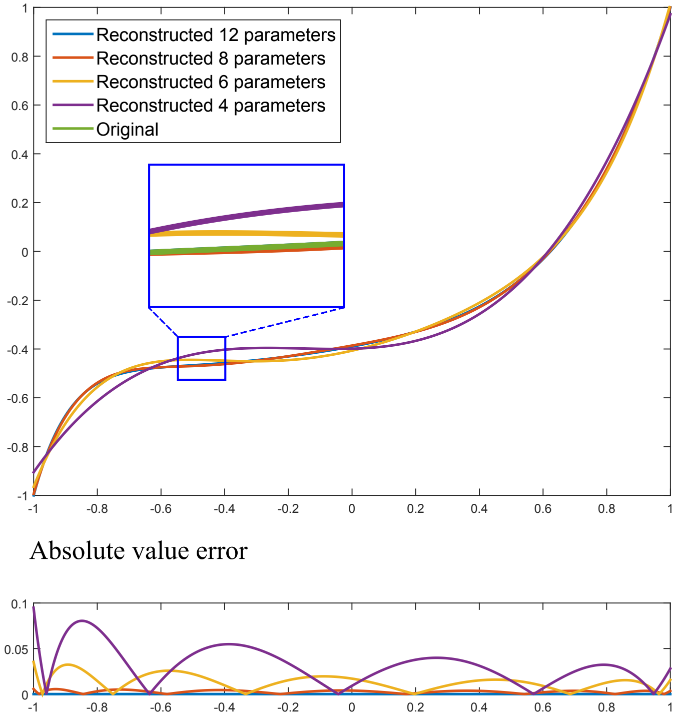


Figure 6.5: *A function is approximated with different numbers of parameters (top). The absolute value of the error between the original function and the reconstructed representation is shown in a larger scale (bottom).*

6.3.2 Coding

As discussed previously, an approximated but visually lossless representation of a lumipath $\bar{g}^{p,f}$ can be specified by a $N_{p,f}$-tuple of Chebyshev coefficients $\left(c_0^p, \dots, c_{N_{p,f}}^{p,f}\right)$.

In practice, these coefficients are highly correlated over space and time which allows for further compression of the data. In this section we present a suitable coding approach, which quantizes Chebyshev coefficients and reorganizes them into monochrome video sequences (Fig. 6.6). H.264 is then used as a standard video coding method [Wiegand et al., 2003] for efficient compression. As our results show in the next section, we achieve very reasonable bitrates with this approach, which shows that CDR is a suitable solution for the distribution challenge. Our method for CDR video coding still leaves room for improvement, however. Better use of the nature of the provided data could provide improved data rates (see Sec. 6.5 for further discussion).

We represent all lumipaths in an image-like format, which then allows the application of a video codec. We compute the maximum degree $N := \max_{p,f} N_{p,f}$ and set $c_k^{p,f} := 0$ for $k > N_{p,f}$, which leads to a representation $\bar{g}^{p,f}(x) = \sum_{k=0}^{N} c_k^{p,f} \psi_k(x)$ of the function described in Equation 6.6, but with a fixed parameter N. Each lumipath $\bar{g}^{p,f}$ is now specified by an N-tuple

$$\mathbf{c}^{p,f} := \left(c_1^{p,f}, \ldots, c_N^{p,f}\right). \tag{6.7}$$

To get an image-like representation, we represent the tuples $\mathbf{c}^{p,f}$ of all pixels of a frame by *coefficient* matrices $\mathtt{C}_\mathtt{k}^\mathtt{f} \in \mathbb{R}^{\mathrm{h} \times \mathrm{w}}$ for k from 1 to N, which by construction have the same pixel resolution $\mathrm{h} \times \mathrm{w}$ as $\mathtt{I}_\alpha^f$ and $\mathtt{I}_\beta^f$. We uniformly quantize all entries of all matrices $\mathtt{C}_\mathtt{k}^\mathtt{f}$ to 8-bit integers [Sayood, 2000] obtaining N matrices $\bar{\mathtt{C}}_k^f$. A bitdepth of 8 is used because it corresponds to the maximum bit depth for images which are supported for compression by the main profile of H.264. Fig. 6.6 shows the first 8 coefficient images for a frame of sequence *Gunman*. It can be observed that the energy and variance in the coefficient images drops rapidly with increasing coefficient index. Most of the information is concentrated within the first coefficients. Coefficients often have uniform values over large image regions. We further observed very smooth behavior over time. Thus, the information content of such coefficient images and videos is relatively limited in practice as compared to the images and videos themselves, making them very well compressible.

A compressed representation of all lumipaths is obtained by storing (i) one integer value which represents the degree N, (ii) two floating point values representing the minimum and maximum value used for 8-bit quantization, and (iii) an encoded representation of the image sequences $\bar{\mathtt{c}}_k^1, \ldots, \bar{\mathtt{c}}_k^F$ for $k = 1, \ldots, N$ which is obtained with H.264. Fig. 6.7 shows the individual bitrates for each of the coefficient image sequences of a CDR video example. As suggested by Fig. 6.6, we can observe that the bitrate rapidly drops for coefficients with higher index values.

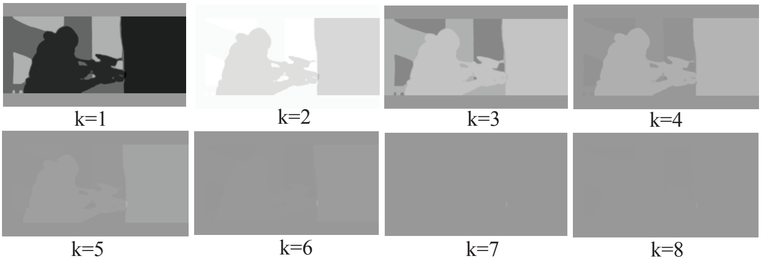


Figure 6.6: *Coefficient images $\bar{\mathtt{c}}_k^1$ of sequence Gunman.*

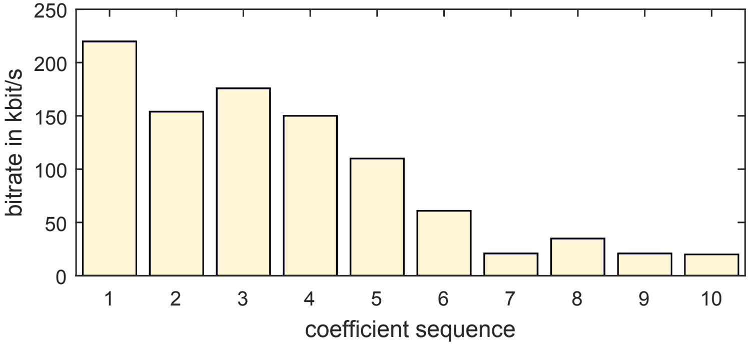


Figure 6.7: *Bit rates of individual coefficient sequences of sequence Gunman for a quantization parameter of 30.*

6.4 Results

We tested our system on a number of video sequences obtained from three short feature films: *Tears of Steel*[1], *Big Buck Bunny*[2]and *Lucid Dreams of Gabriel* [3]. As it is impossible to visualize HDR imagery with traditional SDR displays, in this work all results are presented with tonemapped images. It is important to note that these representations do not show the full extent of our method. We used Adaptive Logarithmic Mapping [Drago et al., 2003] to tonemap HDR frames for presentation as it is easily implemented and only requires a single input parameter.

[1] Tears of Steel - Old Man, Pannel, Gunman, Rockets scenes, copyright Blender Foundation (www.mango.blender.org).

[2] Big Buck Bunny - Bunny, Bird and Peach scenes, copyright Blender Foundation (www.bigbuckbunny.org).

[3] Lucid Dreams of Gabriel - Car scene, copyright Disney Research, ETH Zurich (www.disneyresearch.com/luciddreamsofgabriel).

6.4.1 Evaluation

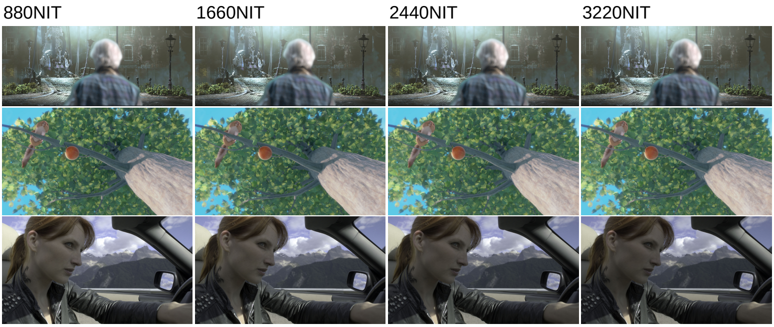


Figure 6.8: *Display Adaptive Tone Mapping can be used to generate content for displays with different luminance levels, but does not allow for precise artistic control of content.*

Our system allows precise local control of luminance when grading for any display in the dynamic range hull. A comparison of a sample grading produced by the authors of this paper and automatic methods can be seen in Fig. 6.11, top. It is interesting to note that content creators may intend to maintain a particular luminance contrast in their scenes, which could be lost through global tone mapping operations. Notice the loss of contrast between the background and foreground in the *Old Man* and *Pannel* scenes, and the excessively dark man in the *Gunman* scene when the views are interpolated linearly. In contrast, automatic methods such as Display Adaptive Tone Mapping [Mantiuk et al., 2008] can preserve the appearance of the scene across multiple dynamic ranges (Fig. 6.8), but they do not allow the artistic freedom enabled by the localized editing of lumipaths as we do in our method. Such methods also do not account for the efficient approximation and distribution of the generated content to a multitude of different target displays. Another fundamental difference is that display adaptive tone mapping only utilizes single source grading for deriving any intermediate gradings. Our method in contrast uses the maximum and minimum grading

To showcase some possibilities of artistic gradings that can be achieved using our method, we point the reader to Fig. 6.11, bottom. Notice that the *Rockets* scene can be graded to either emphasize the details near the rocket motor, or create a bloom effect to convey the brightness of the flames. The *Bird* and *Peach* scenes are graded to give greater emphasis on either the main object or the background of the scene. In the *Car* scene, grading can be used to create the sensation of a cloud above the scene, or that of a sunny day.

Gradings as presented above can be efficiently encoded using the method presented in Section 6.3. When using H.264 video coding, the FFmpeg library was employed and a group of pictures size of 24 was used for all sequences, while the quantization parameter (QP) was varied to control the loss of quality. Fig. 6.9 shows the data rates for the lumipaths of five sample scenes, with an average of 1.52 *Mbit/s* for the highest quality setting, which corresponds to approximately 13% of the current business-to-consumer distribution bandwidth used for 1080i50 television signals.

6.4.2 Perceptual Validation

We performed a perceptual experiment to test whether the distortions introduced due to lossy coding of the lumipath information would lead to visually noticeable artifacts. In our user study, we showed video content to our subjects on a SIM2 HDR display [Sim2, 2015]. The CDR video, which was created with minimum and maximum grades at 100

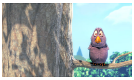
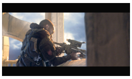
QP	 Bird	 Car	 Gunman	 Peach	 Rockets
30	0.768	2.903	1.110	1.318	1.477
40	0.400	1.411	0.787	0.397	0.871
50	0.300	0.659	0.534	0.221	0.552

Figure 6.9: *This figure shows the total bitrates of the lumipaths for five sequences. The bitrates are expressed in Mbit/s and obtained by encoding with different quantization parameters. These sequences are used for the perceptual experiment.*

and 4000 nits, respectively, was sampled over the continuous dynamic range at 700, 1500, and 3000 nits. We performed a 2 alternate forced-choice procedure (2AFC) on a set of 5 videos (*Bird, Car, Gunman, Peach and Rockets*). After a short training session where compression artifacts were explicitly pointed out, participants were tasked with selecting the video with better quality. The comparison was always performed between a reference uncompressed video sample with either itself (in which case the choice was at chance), or the same sample compressed using a quality parameter $QP \in \{30, 40, 50\}$. 16 volunteers participated (5F, 11M), aged 25 to 36 with normal or corrected-to-normal vision.

We performed *ANOVA* analysis on the results of the experiment and found a number of interesting interactions. Answers for $QP = 50$ and $QP = 40$ were found to be significantly different from the reference ($\sigma \ll 0.001$ and $\sigma = 0.043$, respectively). In addition, both $QP = 30$ and $QP = 40$ were found significantly different from $QP = 50$ ($\sigma \ll 0.001$ and $\sigma = 0.012$). No difference was found between the reference and $QP = 30$. These results suggest that

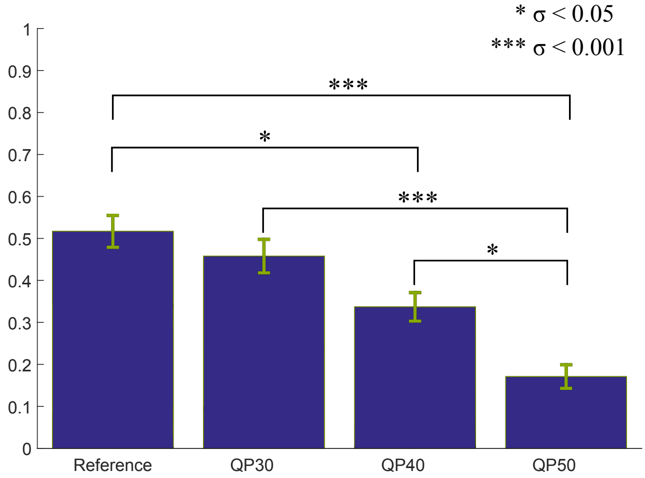


Figure 6.10: *This figure shows the results of our user study, averaged over all sequences and brightness levels. Our supplementary material contains the raw data and full statistical analysis for this experiment. The values on the Y-axis represent the ratio at which the reference was considered to have better quality, with a value of 0 meaning the reference was preferred in every trial.*

participants were unable to see the difference in quality between the uncompressed material and the $QP = 30$ version, but could clearly distinguish it from $QP = 40$ and $QP = 50$. The values presented above are shown in Fig. 6.10.

Participants were also more likely to see differences in the *Peach* sequence than the *Car*, *Gunman* and *Rockets* sequences ($\sigma < 0.05$). This could be explained by the fact that the video shown in *Peach* was computer generated and had a very clear image edge separating the slow-moving object of interest from a flat, motionless background - making compression artifacts stand out particularly strongly. No significant interactions were found for the *brightness* parameter.

These results indicate that the lumigraph data of CDR video can be compressed in a visually lossless way at QP30 to about 13% (1.52 *Mbit/s*) of the corresponding video bitrate on average.

6.5 Discussion

Our method is not without limitations. In this work we presented a formulation and implementation of a novel content creation and distribution paradigm. While we demonstrated that our method results in a low bandwidth overhead and allows full artistic freedom, many of the components that we presented could be engineered for better performance. For example, a more sophisticated human visual system model could replace our current model for better predicting the threshold luminances, which could help reduce the number of polynomial basis coefficients while still maintaining a visually lossless representation (Section 6.3). While the editable lumipaths give the user full control over luminance during dynamic range mapping, our current implementation does not allow a similar control over chrominance, although the colors in the maximum and minimum gradings can be adjusted without any limitation. Formulating a representation for chrominance that is analogous to lumipaths, as well as extending our dynamic range mapping interface to support such a representation is left as future work. Further, more dedicated approaches for compression of lumipath image sequences (e.g. inspired by depth coding approaches [Merkle et al., 2013]) could reduce bitrates even further.

6.6 Conclusion

We presented CDR video, a novel representation of pixel-level luminance as a function of display dynamic range. We introduced dynamic range mapping as a new approach for content creation targetting displays with different dynamic ranges. An efficient approximation of CDR video through a polynomial series approximation was presented, as well as coding that consumes only 13% of current standard business-to-consumer distribution bandwidth. Together, these components form an end-to-end solution for content production and distribution for the wide variety of emerging HDR displays.

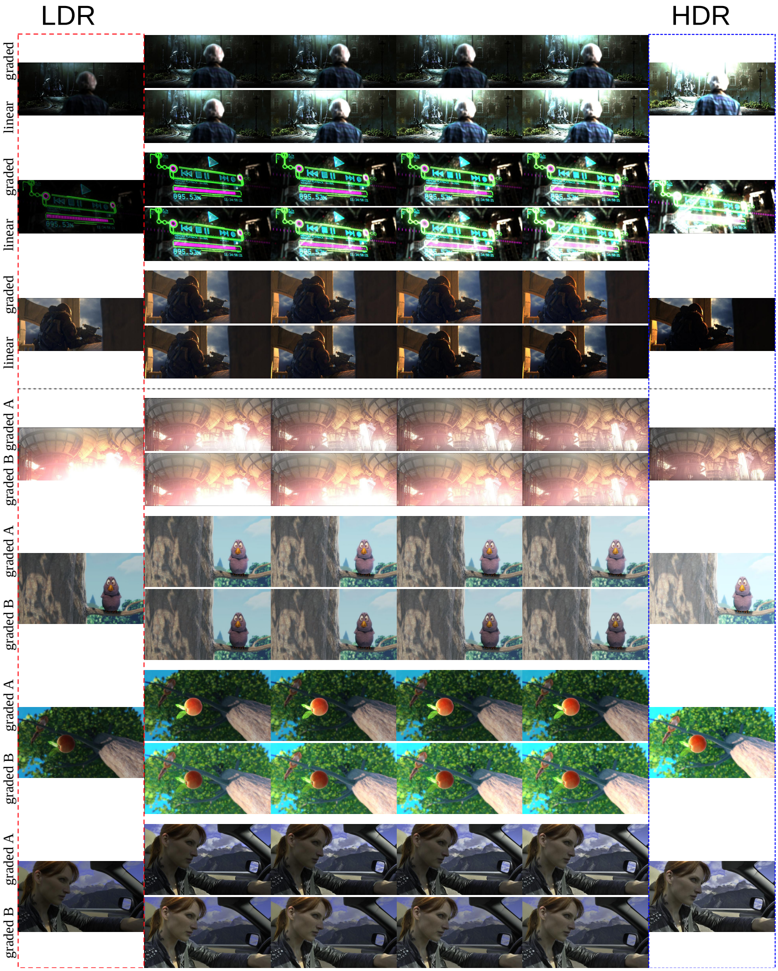


Figure 6.11: *(Top) shows a comparison of results color graded with our method, as compared to a naive linear interpolation between the SDR and HDR graded versions. Notice that our system allows for local control of the grading at each point of the dynamic range hull. (Bottom) Two gradings obtained using our system are shown in contrast to showcase different artistic possibilities that can be achieved using our system.*

CHAPTER 7

Conclusion

In this thesis we explored novel computational frameworks for high-quality content creation for novel displays. The display technologies used in the works presented here, namely stereo 3D displays and high dynamic range displays, are engineered to provide a more realistic viewing experience than that of traditional displays, but also suffer from technical limitations. In our work, perceptual techniques are applied to measure these limitations and computational and mathematical tools are generated to circumvent existing problems in attempts to maximize the perceptual benefits of novel display technologies. The overarching problem discussed in this work can be described as an attempt to reproduce the great amount of variation present in natural scenes on displays that are limited by their technology.

7.1 Summary of Contributions

The following sections recapitulate the main contributions present in this thesis.

7.1.1 Contributions on Perception of Cardboarding

In the case of S3D displays discussed in Chapters 3, 4 and 5, the inherent dichotomy between the tridimensional scene being shown and the flat display causes unpleasant viewing experiences. To avoid these, scenes are often flattened to present a shallower depth profile, but this action results in a perceptual artifact arising from the unnatural flatness of the presentation. In Chapter 3, this artifact, cardboarding, is explored in a series of perceptual experiments. The result is a better understanding of the cardboarding artifact. In particular, all our experiments point toward approximate compression thresholds of 80%, when depth compression is generally imperceptible and 30%, the point at which any additional compression is usually clearly noticed. In addition, the results of our "Rating" experiment could be used directly as a cost function for the compression of objects in S3D post-production pipelines.

7.1.2 Contributions on Autostereo Content Creation

Chapter 4 is concerned with a type of S3D display on which cardboarding is particularly prevalent, namely autostereoscopic displays. By their design, autostereoscopic displays must present only a shallow depth profile, otherwise content will be aliased. In our work we began by analyzing aliasing from a perceptual viewpoint. Knowing that in practice autostereoscopic displays are often used by content creators in a way that allows some aliasing, we devised a study to measure the practical subjective consequences of aliasing. We found that this artifact can be tolerated in small quantities, and described a subjectively acceptable limit for displayed depth on an autostereo screen. Knowing the limitations of autostereo screens is not enough to generate good quality autostereo content, however. Due to inherent difficulties in obtaining multiple views of a scene for display on multiview screens (in fact the number of views is not even known in advance unless a specific display model is targeted!), we employed a modified version of the Image Domain Warping algorithm [Lang et al., 2010] to perform stereo-to-multiview conversion. Before this final step, however, we use a novel optimization-based technique that remaps the depth of the input stereo 3D content to the perceptually acceptable base of the targeted autostereoscopic display in a two step procedure: first the overall depth range is brought down in a saliency based global mapping operation, followed by a localized optimization on salient regions that makes select objects on the scene appear more round, thus helping prevent cardboarding. We validated our system through an additional step of perceptual testing in which subjects found scenes enhanced by our method to be “more 3D” but did not notice significant visual artifacts.

7.1.3 Contributions on Stereo from Shading

Following our work on depth remapping, we decided to consider alternative ways of avoiding cardboarding as shown in Chapter 5. Armed with the knowledge that binocular disparity is only one of many 3D cues, attempting to use different cues to enhance 3D perception was a natural next step. Importantly, other depth cues could avoid the discomfort generated by the vergence-accomodation conflict. Most depth cues, however, are not easily modifiable in a post-production pipeline. For example, texture gradients, interpositions, perspective and locations relative to the horizon are all inherent parts of a scene. Motion parallax is impossible on standard 3D screens, and require motion from the user when displayed in multiview. In fact, some cues, such as relative size are inherent parts of the cognitive understanding humans have of the world. Inspired by the work of Puerta [Medina Puerta, 1989], we explored binocular lighting variations as a mean to generate a sensation of depth. As scene lighting can be modified more easily than most other 3D cues, such an effect could be achieved in a post-production setting. In order to avoid extreme rivalry artifacts caused by cast shadows and specular highlights, we restricted our method to diffuse shading. In order to tackle varying lighting conditions present in most scenes, we devised a novel algorithm that transfers the edits, which we call "shading stereo", to the targeted objects on the scene by rotating their normals instead of moving light sources. We proceeded to perform a series of exploratory user studies to make the use of shading stereo practically possible. We measured the acceptable range of rotation of normals in order to avoid perceptually disturbing visual rivalry artifacts. We proceeded to repeat the same experiment with the opposite edits performed for each binocular view, which resulted in significantly deteriorating quality, demonstrating that shading stereo is a binocular depth cue. Finally, we performed a validation experiment where cardboarded scenes augmented with shading stereo are compared to their standard unenhanced variants, and found that shading stereo provides a significant improvement in perceived roundness. As a last step, we demonstrate that our shading stereo method can be applied to live-action scenes if scene re-lighting is possible.

7.1.4 Contributions on HDR Content Creation

In our following work, we considered the challenges incurred by a different kind of emerging display technology - namely high dynamic range display. A great variety of displays with brightness contrast capabilities beyond that of traditional displays will be available soon, and providing good content for these displays is a significant challenge for contemporary content creators. In particular, we identify two main challenges. The *production* challenge is the difficulty of color grading content for a variety of dynamic ranges. Existing post-production pipelines often involve several grading steps for cinema, home and HDR display, but it is clear that adding even more color grading ranges would be impractical and undesirable. The *distribution* challenge is related to transmitting the correct grading, once generated, to the users. Assuming users will have a variety of display devices available with diverse dynamic ranges, the bandwidth required to encode several variations of color grading would be prohibitive. In the work presented in Chapter 6 we demonstrate a novel end-to-end pipeline for the generation and distribution of HDR content. In a first step, we implement an original user interface inspired by existing commercial color-grading tools, that allows the user to grade content simultaneously for a broad range of dynamic ranges. With such a grading, we generate a novel data structure we call “Continuous Dynamic Range Video”, which is in essence a continuous per-pixel function relating the maximum brightness of the target display to the selected pixel brightness at that level. Following, this enormous amount of per-pixel data is represented using a truncated Chebyshev Series. In our work we demonstrate that this representation converges rapidly and is made perceptually indistinguishable with the original content by bounding any allowed error using a simple model of the human visual system. In a final step this information, now discrete, is compressed using a standard video codec. The resulting information requires only 13% of the standard commercial video bandwidth in our test cases, but can represent the exact desired pixel color for any display contained within the CDR. In a final step we test our system for perceptual artifacts caused by the video codec, and find that the content is visually transparent with the uncompressed reference unless extreme compression settings are used.

7.2 Future Work

Fine grained options for improvements on the presented work and limitations of the individual methods presented in Chapters 3-6 are discussed within each chapter. In this section we will outline possible high-level directions for future work in the field of perceptual enhancements for novel displays.

Immersive display is a field with substantial interest from both research and industry in the past years. Although original systems like the CAVE automatic virtual environment [Cruz-Neira et al., 1993] are likely too complex for widespread adoption, modern immersive display often attempts to provide a wider than normal field of view. An example can be seen in some commercially available systems such as the IMAX [I. Corporation, 2010], that uses specialized hardware, or alternatively systems that work through the use of additional projection devices [Jones et al., 2013] or additional illumination [Weffers-Albu et al., 2011] together with standard displays as a means of showing content to the peripheral vision of observers. For such modes of display, the production and distribution challenges remain quite real. Content for IMAX displays requires higher resolutions than those of most available professional content, which incurs significant manual upscaling labor. Acquiring peripheral views for display around a main screen can prove equally difficult, as modern film sets utilize props and lighting immediately outside of the visible region. Methods for optimized capture or generation of such content could provide significant help with the adoption of these display modalities.

Virtual reality comprises another type of immersive display that places the screen on a wearable headset, such as the commercially available Oculus Rift [2014] or recently developed research prototypes [Huang et al., 2015]. In this case, perceptual limitations remain significant and could benefit from approaches similar to those outlined in this thesis. In particular, *VR sickness* [Kolasinski, 1995; LaViola Jr, 2000] is considered a major impairment to the use of wearable displays and could be caused by a strong presence of the vergence-accomodation conflict. Perceptual models of confortable VR and computational methods that change the viewed images to alleviate the problem could both prove to be useful tools to make VR use mainstream.

High frame rate is often considered a desirable trait for high quality depictions of motion [McDonnell et al., 2007]. At the same time, it is widely accepted in the film industry that the current cinematic standard of 24 frames per second is often too low to provide a realistic impression of human movement. While partial models of human perception of video frame rates have been presented [Watson, 2013], no well accepted practical system exists for selecting perceptually meaningful frame rates for video content. Further progress in understanding the response of the human visual system to various types of temporally varying stimuli could help guide content creators towards making meaningful choices, while practical systems for high-quality video upscaling could bring legacy content to a higher temporal resolution without reducing the quality of the assets.

7.3 Conclusions

The content creation algorithms presented in this thesis can be applied as components in a cinematic post-production pipeline. Perceptual modeling and validation as used in this thesis are important guiding tools for technological efforts in computer graphics and image processing.

Stereoscopic 3D is likely to remain important for a number of applications in the near future, in particular those related to virtual and augmented reality. In these cases, depth limitations and cardboarding are likely to remain a problem. Although recent advances in multi-plane display [Narain et al., 2015] are a promising direction for obtaining fully accommodative displays, the acceptable depth of field remains quite shallow (similarly to autostereoscopic displays) [Wetzstein et al., 2011b; Wetzstein et al., 2012c]. These and additional limitations suggest that perceptual enhancements like efficient use of depth as presented in Chapter 4 or unconventional depth cues as in Chapter 5 will still be necessary for the next generation of 3D display technologies.

HDR content production and distribution is likely to increase significantly in the near future. Popular streaming services such as Amazon and Netflix have recently announced support for HDR broadcast. This makes the production and distribution challenges particularly important. A variety of technologies that address these problems are likely to appear in the following months, with our work described in Chapter 6 being the earliest effort. While our solution requires significantly less bandwidth than a naive solution consisting of transmitting additional views, it could be further improved with the use of dedicated compression methods that capitalize on the properties of polynomial series approximations.

References

[Abrash, 2014] Michael Abrash. What VR could, should, and almost certainly will be within two years. Presentation given at Steam Dev Days event, 2014.

[Aydın et al., 2008] Tunç O. Aydın, Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Dynamic range independent image quality assessment. In *ACM Trans. Graph. (Proc. of SIGGRAPH)*, volume 27(3), 2008.

[Aydın et al., 2014] Tunç O. Aydın, Nikolce Stefanoski, Simone Croci, Markus Seidel, Gross, and Aljoscha Smolic. Temporally coherent local tone mapping of hdr video. In *ACM Trans. Graph. (Proc. of SIGGRAPH Asia)*, volume 33(6), 2014.

[Bailey et al., 2003] Mike Bailey, Thomas Rebotier, and David Kirsh. Quantifying the relative roles of shadows, stereopsis, and focal accommodation in 3D visualization. In *IASTED VIIP*, pages 992–997, 2003.

[Boitard et al., 2014] Ronan Boitard, Rémi Cozot, Dominique Thoreau, and Kadi Bouatouch. Zonal brightness coherency for video tone mapping. *Signal Processing: Image Communication*, 29(2):229 – 246, 2014.

[Bowman et al., 2007] Doug Bowman, Ryan P McMahan, et al. Virtual reality: how much immersion is enough? *Computer*, 40(7):36–43, 2007.

[Brewster, 1856] David Brewster. *The Stereoscope; Its History, Theory and Construction, with Its Application to the Fine and Useful Arts and to Education, Etc.* John Murray, 1856.

[Broucke, 1973] R Broucke. Algorithm: ten subroutines for the manipulation of chebyshev series. *Communications of the ACM*, 16(4):254–256, 1973.

[Canon Inc., 2015] Canon Inc. DP-V2410 4k reference display, 2015.

[Chai et al., 2000] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum. Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 307–318. ACM Press/Addison-Wesley Publishing Co., 2000.

[Chapiro et al., 2014a] Alexandre Chapiro, Olga Diamanti, Steven Poulakos, Carol O'Sullivan, Aljoscha Smolic, and Markus Gross. Perceptual evaluation of cardboarding in 3D content visualization. In *Proc. ACM SAP*, 2014.

[Chapiro et al., 2014b] Alexandre Chapiro, Simon Heinzle, Tunc Aydin, Steven Poulakos, Matthias Zwicker, Aljoscha Smolic, and Markus Gross. Optimizing stereo-to-multiview conversion for autostereoscopic displays. *Comp. Graph. Forum*, 33(2), 2014.

[Chapiro et al., 2015a] Alexandre Chapiro, Tunc Aydin, Nikolce Stefanoski, Simone Croci, Aljoscha Smolic, and Markus Gross. Art-directable continuous dynamic range video. *Computers & Graphics*, 2015.

[Chapiro et al., 2015b] Alexandre Chapiro, Carol O'Sullivan, Wojciech Jarosz, Markus Gross, and Aljoscha Smolic. Stereo from shading. In *Proceedings of EGSR (Experimental Ideas & Implementations)*, 2015.

[Cruz-Neira et al., 1993] Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142. ACM, 1993.

[Dabala et al., 2014] Lukasz Dabala, Petr Kellnhofer, Tobias Ritschel, Piotr Didyk, Krzysztof Templin, Karol Myszkowski, P Rokita, and Hans-Peter Seidel. Manipulating refractive and reflective binocular disparity. *Comp. Graph. Forum*, 33(2):53–62, 2014.

[Davis, 1975] Philip J Davis. *Interpolation and approximation*. Courier Corporation, 1975.

[Debevec and Malik, 2008] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM, 2008.

[Dehaene, 2003] Stanislas Dehaene. The neural basis of the weber–fechner law: a logarithmic mental number line. *Trends in cognitive sciences*, 7(4):145–147, 2003.

[Didyk et al., 2011] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. A perceptual model for disparity. *ACM Trans. Graph.*, 2011.

[Didyk et al., 2012a] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. Apparent stereo: The cornsweet illusion can enhance perceived depth. In *Proc. IS&T SPIE HVEI*, pages 1–12, 2012.

[Didyk et al., 2012b] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, Hans-Peter Seidel, and Wojciech Matusik. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph.*, 31(6), 2012.

[Didyk et al., 2013] Piotr Didyk, Pitchaya Sitthi-Amorn, William Freeman, Frédo Durand, and Wojciech Matusik. Joint view expansion and filtering for automultiscopic 3D displays. *ACM Transactions on Graphics (TOG)*, 32(6):221, 2013.

[Dolby Laboratories, 2015] Inc. Dolby Laboratories. Dolby vision white paper. `http://www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-white-paper.pdf`, 2015.

[Drago et al., 2003] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22(3):419–426, 2003.

[Du et al., 2013] Song-Pei Du, Belen Masia, Shi-Min Hu, and Diego Gutierrez. A metric of visual comfort for stereoscopic motion. *ACM Trans. Graph.*, 32(6), 2013.

[Durand and Dorsey, 2002] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.*, 21(3):257–266, 2002.

[Durand et al., 2005] Frédo Durand, Nicolas Holzschuch, Cyril Soler, Eric Chan, and François X Sillion. A frequency analysis of light transport. *ACM Transactions on Graphics (TOG)*, 24(3):1115–1126, 2005.

[Eilertsen et al., 2013] Gabriel Eilertsen, Robert Wanat, Rafal K. Mantiuk, and Jonas Unger. Evaluation of tone mapping operators for HDR-video. *Computer Graphics Forum*, 32(7):275–284, 2013.

[Epstein and Rogers, 1995] W. Epstein and S. Rogers. *Perception of Space and Motion, Chapter 3: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth.* Academic Press, 1995.

[Fattal et al., 2002] Raanan Fattal, Dani Lischinski, and Michael Werman. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21(3):249–256, 2002.

[Fechner, 1858] GT Fechner. Über ein wichtiges psychophysiches grundgesetz und dessen beziehung zur schäzung der sterngrössen. abk. k. *Ges. Wissensch., Math.-Phys. K*, 1(4), 1858.

[Ferwerda and Luka, 2009] James Ferwerda and Stefan Luka. A high resolution, high dynamic range display for vision research. 9(8):346, 2009.

[Ferwerda and others, 2001] James Ferwerda et al. Elements of early vision for computer graphics. *Computer Graphics and Applications, IEEE*, 21(5):22–33, 2001.

[Ferwerda et al., 1996] James A. Ferwerda, Sumanta N. Pattanaik, Peter Shirley, and Donald P. Greenberg. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH 1996, pages 249–258, 1996.

[Gabriel, 1908] Lippmann Gabriel. La photographie intégrale. *Comptes-Rendus, Académie des Sciences*, 146:446–551, 1908.

[Gaebler et al., 2014] Michael Gaebler, Felix Biessmann, Jan-Peter Lamke, Klaus-Robert Müller, Henrik Walter, and Stefan Hetzer. Stereoscopic depth increases intersubject correlations of brain networks. *Neuroimage*, 100:427–434, 2014.

[Gil et al., 2007] Amparo Gil, Javier Segura, and Nico M Temme. *Numerical methods for special functions*. Siam, 2007.

[Greisen et al., 2013] Pierre Greisen, Marian Runo, Patrice Guillet, Simon Heinzle, Aljoscha Smolic, Hubert Kaeslin, and Markus Gross. Evaluation and fpga implementation of sparse linear solvers for video processing applications. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(8):1402–1407, 2013.

[Guarnieri et al., 2008] Gabriele Guarnieri, Luigi Albani, and Giovanni Ramponi. Image-splitting techniques for a dual-layer high dynamic range lcd display. *J. Electronic Imaging*, 17(4), 2008.

[Hajsharif et al., 2014] Saghi Hajsharif, Joel Kronander, and Jonas Unger. Hdr reconstruction for alternating gain (iso) sensor readout. In *Eurographics, Strasbourg, France, April 7-11, 2014*, 2014.

[Hanhart et al., 2014] Philippe Hanhart, Pavel Korshunov, Touradj Ebrahimi, Yvonne Thomas, and Hans Hoffmann. Subjective Quality Evaluation Of High Dynamic Range Video And Display For Future TV. In *International Broadcasting Convention (IBC)*, 2014.

[Held and Banks, 2008] Robert T Held and Martin 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.

[Hoffman et al., 2008] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33, 2008.

[Howard and Rogers, 2002] Ian P. Howard and Brian J. Rogers. *Seeing in Depth: Volume 2: Depth perception*. Oxford University Press, NY, USA, 2002.

[Huang et al., 2015] Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. ¡ light field stereoscope. In *ACM SIGGRAPH 2015 Emerging Technologies*, page 24. ACM, 2015.

[Huber et al., 2015] Rafael Huber, Benjamin Scheibehenne, Alexandre Chapiro, Seth Frey, and Robert Sumner. The influence of visual salience on video consumption behavior a survival analysis approach. In *Proceedings of ACM Web Science*, 2015.

[I. Corporation, 2010] I. Corporation. IMAX: a motion picture film format and a set of cinema projection standards, 2010.

[Ichihara et al., 2007] Shigeru Ichihara, Norimichi Kitagawa, and Hiromi Akutsu. Contrast and depth perception: Effects of texture contrast and area contrast. *Perception*, 36(5):686, 2007.

[ISO/IEC MPEG, 2015] ISO/IEC MPEG. Call for Evidence (CfE) for HDR and WCG Video Coding. `http://mpeg.chiariglione.org`, 2015.

[Ives, 1903] Frederic E Ives. Parallax stereogram and process of making same., April 14 1903. US Patent 725,567.

[Jain and Konrad, 2007] Ashish Jain and Janusz Konrad. Crosstalk in automultiscopic 3-d displays: Blessing in disguise? In *Electronic Imaging 2007*, pages 649012–649012. International Society for Optics and Photonics, 2007.

[Jones et al., 2001] G. Jones, D. Lee, N. Holliman, and D. Ezra. Controlling perceived depth in stereoscopic images. In *Proc. SPIE*, volume 4297, 2001.

[Jones et al., 2013] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 869–878. ACM, 2013.

[Junyent et al., 2015] Marc Junyent, Pablo Beltran, Miquel Farre, Jordi Pont-Tuset, Alexandre Chapiro, and Aljoscha Smolic. Video content and structure description based on keyframes, clusters and storyboards. In *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, 2015.

[Kersten and Mamassian, 2014] Daniel Kersten and Pascal Mamassian. Cast shadow illusions. *Oxford Compendium Of Visual Illusions*, 2014. (to appear).

[Kim et al., 2009] Min H Kim, Tim Weyrich, and Jan Kautz. Modeling human color perception under extended luminance levels. In *ACM Transactions on Graphics (TOG)*, volume 28, page 27. ACM, 2009.

[Kim et al., 2014] Joohwan Kim, Paul V Johnson, and Martin S Banks. Stereoscopic 3D display with color interlacing improves perceived depth. *Optics express*, 22(26):31924–31934, 2014.

[Kolasinski, 1995] Eugenia M Kolasinski. Simulator sickness in virtual environments. Technical report, DTIC Document, 1995.

[Lambooij et al., 2009a] M. Lambooij, W. A. IJsselsteijn, M. Fortuin, and I. Heynderickx. Visual discomfort and visual fatigue of stereoscopic displays: A review. *J. Imaging Science and Technology*, 53(5), 2009.

[Lambooij et al., 2009b] Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):30201–1, 2009.

[Lang et al., 2010] Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.*, 29(3), 2010.

[Lang et al., 2012] Manuel Lang, Oliver Wang, Tunc Aydin, Aljoscha Smolic, and Markus H Gross. Practical temporal consistency for image-based graphics applications. *ACM Trans. Graph.*, 31(4):34, 2012.

[Langer and Bülthoff, 1999] Michael S Langer and Heinrich H Bülthoff. Depth discrimination from shading under diffuse lighting. *Perception*, 29(6):649–660, 1999.

[LaViola Jr, 2000] Joseph J LaViola Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000.

[Lueder, 2011] Ernst Lueder. *3D Displays*, volume 32. John Wiley & Sons, 2011.

[Lukk, 2014] Howard Lukk. Emma. http://www.emmathemovie.com, 2014.

[Mantiuk et al., 2006a] RafałMantiuk, Alexander Efremov, Karol Myszkowski, and Hans-Peter Seidel. Backward compatible high dynamic range MPEG video compression. *ACM Trans. Graph.*, 25(3):713–723, 2006.

[Mantiuk et al., 2006b] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.*, 3(3):286–308, 2006.

[Mantiuk et al., 2008] RafałMantiuk, Scott Daly, and Louis Kerofsky. Display adaptive tone mapping. *ACM Trans. Graph.*, 27(3):68:1–68:10, 2008.

[Mantiuk et al., 2011] Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4):40:1–40:14, 2011.

[Masaoka et al., 2006] Kenichiro Masaoka, A. Hanazato, Masaki Emoto, Hirokazu Yamanoue, Yuji Nojiri, and Fumio Okano. Spatial distortion prediction system for stereoscopic images. *J. Electronic Imaging*, 15(1), 2006.

[Masia et al., 2013a] Belen Masia, Gordon Wetzstein, Carlos Aliaga, Ramesh Raskar, and Diego Gutierrez. Display adaptive 3D content remapping. *Computers & Graphics*, 37(8):983–996, 2013.

[Masia et al., 2013b] Belen Masia, Gordon Wetzstein, Piotr Didyk, and Diego Gutierrez. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*, 37(8):1012–1038, 2013.

[McDonnell et al., 2007] Rachel McDonnell, Fiona Newell, and Carol O'Sullivan. Smooth movers: perceptually guided human motion simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 259–269. Eurographics Association, 2007.

[Medina Puerta, 1989] Antonio Medina Puerta. The power of shadows: shadow stereopsis. *JOSA A*, 6(2):309–311, 1989.

[Meesters et al., 2004] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. SeuntiÃ«ns. A survey of perceptual evaluations and requirements of three-dimensional tv. *IEEE Trans. Circuits and Systems for Video Technology*, 14(3), 2004.

[Merkle et al., 2013] Philipp Merkle, Karsten Mueller, and Thomas Wiegand. Coding of depth signals for 3D video using wedgelet block segmentation with residual adaptation. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, 2013.

[Motion Picture Association of America, 2014] Motion Picture Association of America. Theatrical market statistics. http://www.mpaa.org/wp-content/uploads/2015/03/MPAA-Theatrical-Market-Statistics-2014.pdf, 2014.

[Narain et al., 2015] Rahul Narain, Rachel A Albert, Abdullah Bulbul, Gregory J Ward, Martin S Banks, and James F O'Brien. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Transactions on Graphics (TOG)*, 34(4):59, 2015.

[OCULUS VR, 2014] OCULUS VR. Oculus Rift development kit 2. https://www.oculus.com/dk2/, 2014.

[Ofcom, 2015] Ofcom. The communications market report. `http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr15/CMR_UK_2015.pdf`, 2015.

[Pattanaik et al., 2000] Sumanta N. Pattanaik, Jack Tumblin, Hector Yee, and Donald P. Greenberg. Time-dependent visual adaptation for fast realistic image display. In *Proc. of Conf. on Computer Graphics and Interactive Techniques*, SIGGRAPH 2000, pages 47–54, 2000.

[Perazzi et al., 2012] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 733–740. IEEE, 2012.

[Ramachandra et al., 2011] Vikas Ramachandra, Keigo Hirakawa, Matthias Zwicker, and Truong Nguyen. Spatioangular prefiltering for multiview 3D displays. *Visualization and Computer Graphics, IEEE Transactions on*, 17(5):642–654, 2011.

[Ramanarayanan et al., 2007] Ganesh Ramanarayanan, James Ferwerda, Bruce Walter, and Kavita Bala. Visual equivalence: towards a new standard for image fidelity. *ACM Trans. Graph.*, 26(3), 2007.

[Ranieri et al., 2012] Nicola Ranieri, Simon Heinzle, Quinn Smithwick, Daniel Reetz, Lanny S Smoot, Wojciech Matusik, and Markus Gross. Multi-layered automultiscopic displays. In *Computer Graphics Forum*, volume 31, pages 2135–2143. Wiley Online Library, 2012.

[Reinhard and Devlin, 2005] Erik Reinhard and Kate Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11:13–24, 2005.

[Reinhard et al., 2002] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267–276, 2002.

[Reinhard et al., 2010] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[Rusinkiewicz et al., 2006] Szymon Rusinkiewicz, Michael Burns, and Doug DeCarlo. Exaggerated shading for depicting shape and detail. In *ACM Trans. Graph.*, volume 25, pages 1199–1205. ACM, 2006.

[Samsung Electronics, 2015] Samsung Electronics. Samsung UE65JS9500 SUHD, 2015.

[Sayood, 2000] Khalid Sayood. *Introduction to Data Compression (2Nd Ed.)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000.

[Scher et al., 2013] Steven Scher, Jing Liu, Rajan Vaish, Prabath Gunawardane, and James Davis. 3D+2DTV: 3D displays with no ghosting for viewers without glasses. *ACM Trans. on Graph.*, 32(3), 2013.

[Schriber, 2014] Sasha A. Schriber. Lucid dreams of Gabriel. http://www.disneyresearch.com/luciddreamsofgabriel, 2014.

[Seetzen et al., 2004] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. High dynamic range display systems. *ACM Trans. Graph.*, 23(3):760–768, 2004.

[Shibata et al., 2011a] Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *J. of Vision*, 11(8), 2011.

[Shibata et al., 2011b] Takashi Shibata, Joohwan Kim, David M Hoffman, and Martin S Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision*, 11(8):11, 2011.

[Siegel and Nagata, 2000] M. W. Siegel and S. Nagata. Just enough reality: Comfortable 3-d viewing via microstereopsis. *IEEE Transs on Circuits and Systems for Video Technology*, 10(3), 2000.

[Sim2, 2015] Sim2. Model HDR47ES4MB. `http://www.sim2.com`, 2015. Accessed: 01 June, 2015.

[Singh and Shin, 2013] Darryl SK Singh and Jung Shin. Real-time handling of existing content sources on a multi-layer display. In *IS&T/SPIE Electronic Imaging*, pages 86480I–86480I. International Society for Optics and Photonics, 2013.

[Smolic et al., 2008] Aljoscha Smolic, Karsten Müller, Kristina Dix, Philipp Merkle, Peter Kauff, and Thomas Wiegand. Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 2448–2451. IEEE, 2008.

[Smolic et al., 2014] Aljoscha Smolic, Oliver Wang, Manuel Lang, Nikolce Stefanoski, Miquel Farre, Pierre Greisen, Simon Heinzle, Michael Schaffner, Alexandre Chapiro, Alexander Sorkine-Hornung, and Markus Gross. Image domain warping for advanced 3d video applications. IEEE COMSOC MMTC E-Letter, 2014.

[Šoltészová et al., 2011] Veronika Šoltészová, Daniel Patel, and Ivan Viola. Chromatic shadows for improved perception. In *Proc. ACM/EG NPAR*, pages 105–116, 2011.

[Sony Corporation, 2015a] Sony Corporation. BVM-X300. `http://www.sony.co.uk/pro/product/broadcast-products-professional-monitors-oled-monitors/bvm-x300/specifications/#specifications`, 2015.

[Sony Corporation, 2015b] Sony Corporation. s9005b series. `http://www.sony.co.uk/electronics/televisions/s9005b-series/specifications`, 2015.

[Stefanoski et al., 2013] Nikolce Stefanoski, Oliver Wang, Michael Lang, Pierre Greisen, Simon Heinzle, and Aljoscha Smolic. Automatic view synthesis by image-domain-warping. *Image Processing, IEEE Transactions on*, 22(9):3329–3341, 2013.

[Sternberg, 2008] Robert Sternberg. *Cognitive psychology*. Cengage Learning, 2008.

[Templin et al., 2012] Krzysztof Templin, Piotr Didyk, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Highlight microdisparity for improved gloss depiction. *ACM Trans. Graph.*, 31(4):92, 2012.

[Tocci et al., 2011] Michael D Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen. A versatile hdr video production system. In *ACM Transactions on Graphics (TOG)*, volume 30, page 41. ACM, 2011.

[Tompkin et al., 2013] James Tompkin, Simon Heinzle, Jan Kautz, and Wojciech Matusik. Content-adaptive lenticular prints. *ACM Transactions on Graphics (TOG)*, 32(4):133, 2013.

[Tong et al., 2006] Frank Tong, Ming Meng, and Randolph Blake. Neural bases of binocular rivalry. *Trends in cognitive sciences*, 10(11):502–511, 2006.

[Touze et al., 2013] D. Touze, Y. Olivier, D. Thoreau, and C. Serre. High dynamic range video distribution using existing video codecs. In *Picture Coding Symposium (PCS), 2013*, pages 349–352, Dec 2013.

[Valyus and Asher, 1966] NA Valyus and Harry Asher. *Stereoscopy*. 1966.

[Vangorp et al., 2014] Peter Vangorp, Rafat K. Mantiuk, Bartosz Bazyluk, Karol Myszkowski, Radoslaw Mantiuk, Simon J. Watt, and Hans-Peter Seidel. Depth from hdr: Depth induction or increased realism? In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, pages 71–78, 2014.

[Vizio, 2015] Vizio. Reference series. `http://www.vizio.com/r65b2.html`, 2015.

[von der Pahlen et al., 2014] Javier von der Pahlen, Jorge Jimenez, Etienne Danvoye, Paul Debevec, Graham Fyffe, and Oleg Alexander. Digital Ira and beyond: Creating real-time photoreal digital actors. In *ACM SIGGRAPH 2014 Courses*, SIGGRAPH '14, page 1:384, 2014.

[Wanat et al., 2012] Robert Wanat, Josselin Petit, and Rafal Mantiuk. Physical and perceptual limitations of a projector-based high dynamic range display. In *Theory and Practice of Computer Graphics, Rutherford, United Kingdom, 2012. Proceedings*, pages 9–16, 2012.

[Watson, 2013] Andrew B Watson. High frame rates and human vision: A view through the window of visibility. *SMPTE Motion Imaging Journal*, 122(2):18–32, 2013.

[Weffers-Albu et al., 2011] A Weffers-Albu, S de Waele, W Hoogenstraaten, and C Kwisthout. Immersive TV viewing with advanced Ambilight. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 753–754. IEEE, 2011.

[Wetzstein et al., 2011a] Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar. Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays. In *ACM Transactions on Graphics (ToG)*, volume 30, page 95. ACM, 2011.

[Wetzstein et al., 2011b] Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar. Layered 3d: tomographic image synthesis for attenuation-based light field and high dynamic range displays. In *ACM Transactions on Graphics (ToG)*, volume 30, page 95. ACM, 2011.

[Wetzstein et al., 2012a] Gordon Wetzstein, Douglas Lanman, Diego Gutierrez, and Matthew Hirsch. Computational displays: combining optical fabrication, computational processing, and perceptual tricks to build the displays of the future. In *ACM SIGGRAPH 2012 Courses*, page 4. ACM, 2012.

[Wetzstein et al., 2012b] Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph.*, 31(4):80, 2012.

[Wetzstein et al., 2012c] Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph.*, 31(4):80, 2012.

[Wiegand et al., 2003] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the h.264/avc video coding standard. *IEEE Trans. Cir. and Sys. for Video Technol.*, 13(7):560–576, July 2003.

[Wildeboer et al., 2010] Meindert Onno Wildeboer, Norishige Fukushima, Tomohiro Yendo, Mehrdad Panahpour Tehrani, Toshiaki Fujii, and Masayuki Tanimoto. A semi-automatic multi-view depth estimation method. In *Visual Communications and Image Processing 2010*, pages 77442B–77442B. International Society for Optics and Photonics, 2010.

[Woods et al., 1993] A. J. Woods, T. Docherty, and R. Koch. Image distortions in stereoscopic video systems. *Proc SPIE*, 1915, 1993.

[Yamanoue et al., 2000] H. Yamanoue, M. Okui, and I. Yuyama. A study on the relationship between shooting conditions and cardboard effect of stereoscopic images. *IEEE Trans.Circuits and Systems for Video Technology*, 10(3), Apr 2000.

[Yamanoue et al., 2006] H. Yamanoue, M. Okui, and F. Okano. Geometrical analysis of puppet-theater and cardboard effects in stereoscopic hdtv images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(6), 2006.

[Yang et al., 2012] Xuan S Yang, Linling Zhang, Tien-Tsin Wong, and Pheng-Ann Heng. Binocular tone mapping. *ACM Trans. Graph.*, 31(4):93, 2012.

[Zhang and Ferwerda, 2010] Dan Zhang and James Ferwerda. A low-cost, color-calibrated reflective high dynamic range display. 10(7):397, 2010.

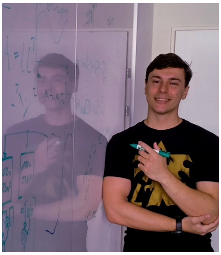
[Zhang et al., 2011] Yang Zhang, E. Reinhard, and David Bull. Perception-based high dynamic range video compression with optimal bit-depth transformation. In *IEEE International Conference on Image Processing (ICIP)*, pages 1321–1324, Sept 2011.

[Zilly et al., 2011] F. Zilly, J. Kluger, and P. Kauff. Production rules for stereo acquisition. *Proc. of the IEEE*, 99(4), April 2011.

[Zilly et al., 2014] Frederik Zilly, Christian Riechert, Marcus Müller, Peter Eisert, Thomas Sikora, and Peter Kauff. Real-time generation of multi-view video plus depth content using mixed narrow and wide baseline. *Journal of Visual Communication and Image Representation*, 25(4):632–648, 2014.

[Zund et al., 2015] Fabio Zund, Pascal Berard, Alexandre Chapiro, Stefan Schmid, Mattia Ryffel, Amit Bermano, Markus Gross, and Robert Sumner. Unfolding the 8-bit era. In *Proceedings of the European Conference on Visual Media Production*, 2015.

[Zwicker et al., 2006] Matthias Zwicker, Wojciech Matusik, Frédo Durand, Hanspeter Pfister, and Clifton Forlines. Antialiasing for automultiscopic 3D displays. In *ACM SIGGRAPH 2006 Sketches*, page 107. ACM, 2006.



Curriculum Vitae

Alexandre Chapiro

Education

Federal Institute of Technology Zurich (ETHZ) **Zurich, Switzerland**
PhD in Computer Science *2011-2015*
Advisor: Prof. Markus Gross. Co-Advisor: Dr. Aljoscha Smolic.

Institute of Pure and Applied Mathematics (IMPA) **Rio de Janeiro, Brazil**
Masters' Degree in Applied Mathematics. *2010–2011*
Advisor: Prof. Paulo Cezar Pinto Carvalho. Co-Advisor: Prof. Luiz Velho.

Federal University of Juiz de Fora (UFJF) **Juiz de Fora, Brazil**
Undergraduate Degree in Mathematics *2007–2009*

Experience

Research

CGL lab **ETHZ**
PhD researcher *Oct.2011–Oct.2015*

Visgraf lab **IMPA**
MSc researcher *Jan.2010–Aug.2011*

GCG lab **UFJF**
Undergraduate researcher *Aug.2007–Dec.2009*

Professional

DRZ lab **Disney Research**
Research assistant at Disney Research Zurich. *Oct.2011–ongoing*
Mentored by Senior Research Scientist Aljoscha Smolic.

European project **Reverie**
Participated in a project co-funded by the European Commission. *Oct.2011–Oct.2013*
Contract FP7-ICT-287723 REVERIE.

Teaching

Math. Foundations of Computer Graphics and Vision **ETHZ**
Teaching assistant, Feb.2014–Jun.2014 and Feb.2015–Jun.2015
Graduate class on mathematical techniques in visual computing.

Informatik 1 **ETHZ**
Teaching assistant, Feb.2012–Jun.2012 and Feb.2013–Jun.2013
Undergradute class on the fundaments of informatics for engineering students.

Academic Services

Reviewing for conferences:: SIGGRAPH 2015, SIGGRAPH Asia 2015, ICIP 2015, Pacific Graphics 2014, 3DV 2012

Reviewing for journals:: IEEE TVCG 2015, Computer Graphics Forum 2014

Volunteering for conferences:: Eurographics 2015

Talks and Presentations

Max Planck Institut 2015: invited talk: Perceptual Enhancements for 3D Displays

EGSR 2015: Stereo from Shading

SAP 2015: Perceptual Evaluation of Cardboarding in 3D Content Visualization

Eurographics 2015: Optimizing Stereo-to-Multiview Conv. for Autostereo. Displays

Languages

Portuguese: Fluent	*Native speaker*
Russian: Fluent	*Native speaker*
English: Fluent	*TOEFL iBT 117/120, Cambridge FCE and CAE exams with A grades*
Spanish: Fluent	*3 years of school in Spain*
French: Advanced	*Alliance Française DELF diplome - 2005*
German: Beginner	*Approximately A2 level*

Funding

European Commission Program FP7, REVERIE	*Oct.2011–Oct.2013*
Master program fellowship, CNPq	*Mar.2010–Jul.2011*
Research fellowship for undergraduate students, FAPEMIG	*Jan.2009–Dec.2009*
Research fellowship for undergraduate students, CNPq	*Jan.2008–Dec.2008*
OBMEP mathematics olympics medalist fellowship 2005/2006, CNPq	*Mar.2007–Jan.2008*
	Mar.2006–Mar.2007
Research fellowship for high-school students, CNPq	*Jan.2005–Dec.2005*

- Obs.: CNPq is the Brazilian national funding agency. FAPEMIG is the funding agency for the state of Minas Gerais, Brazil.

Academic Honors

- Youngest PhD to ever graduate from the Computer Graphics Laboratory at ETH Zurich at 25 years old.
- Finished 2-year Master's program at the National Institute of Pure and Applied Mathematics in 18 months.
- Finished 4-year undergraduate program at the Federal University of Juiz de Fora in Mathematics in 3 years with the highest GPA in the graduating class.
- Ranked 1st in admission examinations, Federal University of Juiz de Fora, 2007.

- Gold Medalist in OBMEP/2005 (Brazilian Mathematics Olympiad for Public School Students). 9th place among approximately 3.8 million students.
- Gold Medalist in OBMEP/2006 (Brazilian Mathematics Olympiad for Public School Students). 95th place among approximately 5.3 million students.
- Gold Medalist in OBA/2006 (Brazilian Astronomy Olympiad).

Publications

*Obs.: * Patent application in progress.*

Papers

Unfolding the 8-bit Era:
Zund, F., Berard, P., Chapiro, A., Schmid, S., Ryffel, A., Bermano, A., Gross, M., Sumner, R.
European Conference on Visual Media Production, London-UK, 2015.

Art-Directable Continuous Dynamic Range Video*:
Chapiro, A., Aydin, T., Stefanoski, N., Croci, S., Smolic, A., Gross, M.
Compters & Graphics, Elsevier, 2015.

Video Content and Structure Description Based on Keyframes, Clusters and Storyboards*:
Junyent, M., Beltran, P., Farre, M., Pont-Tuset, J., Chapiro, A., Smolic, A.
IEEE International Workshop on Multimedia Signal Processing, Xiamen-China, 2015.

Stereo from Shading:
Chapiro, A., O'Sullivan, C., Jarosz, W., Gross, M., Smolic, A.
Eurographics Symposium on Rendering, Darmstadt-Germany, 2015. (E&I track)

Perceptual Evaluation of Cardboarding in 3D Content Visualization:
Chapiro, A., Diamanti, O., Poulakos, S., O'Sullivan, C., Smolic, A., Gross, M.
ACM Symposium on Applied Perception, Vancouver-Canada, 2014. (Short paper)

Optimizing Stereo-to-Multiview Conversion for Autostereoscopic Displays*:
Chapiro, A., Heinzle, S., Aydin, T., Poulakos, S., Zwicker, M., Smolic, A., Gross, M.
Eurographics, Strasbourg-France, 2014.

Towards Mobile HDR Video:
Castro, T.K., Chapiro, A., Cicconet, M., Velho, L.
Eurographics, Llandudno-UK, 2011. (Extended abstract)

Detection of High Frequency Regions in Multiresolution:
Mota, V.F., Perez, E.A., Castro, T.K., Chapiro, A., Vieira, M.B.
ICIP - International Conference on Image Processing, Cairo-Egypt, 2009.

High Frequency Assessment from Multiresolution Analysis:
Castro, T.K. , Perez, E. A. , Mota, V. F. , Chapiro, A. , Vieira, M. B. , Freire, W. P.
ICCS - International Conference on Computational Science, Baton Rouge-USA, 2009.

Book chapters

Discrete Wavelets on Edges:
Chapiro, A. , Knop, T., Mota, V., Perez, E., Bernardes, M., Freire W. P.
InTech open publisher, 2011.

Posters

**The Influence of Visual Salience on Video Consumption Behavior
A Survival Analysis Approach***:
Huber, R., Scheibehenne, B., Chapiro, A., Frey, S., Sumner, R.
ACM Web Science, Oxford-United Kingdom, 2015.

Filter Based Deghosting for Exposure Fusion Video:
Chapiro, A., Cicconet, M., Velho, L.
SIGGRAPH 2011, Vancouver-Canada, 2011. Student Research Competition Semi-Finalist.

Towards Mobile HDR Video:
Castro, T.K., Chapiro, A., Cicconet, M., Velho, L.
ICCP - International Conference on Computational Photography, Pittsburg-USA, 2011.

Mountain's Pass Theorem:
Chapiro, A., Pereira, F.
EMED III - Minas-Gerais Meeting of Partial Diferential Equations, Itajuba-Brazil, 2009.

Other

Image Domain Warping for Advanced 3D Video Applications:
Smolic A., Wang, O., Lang, M., Stefanoski, N., Farre, M., Greisen, P., Heinzle, S., Schaffner, M.,Chapiro, A., Sorkine-Hornung, A., Gross, M.
IEEE COMSOC MMTC E - Letter, 2014. (Invited letter)

Computational Photography: Castro, T.K., Chapiro, A., Velho, L.
IMPA, Rio de Janeiro-Brazil, 2011. (Technical report)

Theses

Perceptual Enhancements for Novel Displays:
Alexandre Chapiro. Advisor: Prof. Markus Gross. Co-advisor: Dr. Aljoscha Smolic.
ETHZ, Zurich-Switzerland, 2015. (Final PhD thesis)

Improving Mobile Video:
Alexandre Chapiro. Advisor: Prof. Paulo Cezar Pinto Carvalho. Co-advisor: Prof. Luiz Velho.
IMPA, Rio de Janeiro-Brazil, 2011. (Final MSc thesis)

An Introduction to Degree Theory:
Alexandre Chapiro. Advisor: Prof. Luiz Fernando Oliveira Faria.
UFJF, Juiz de Fora-Brazil, 2009. (Final BSc thesis)