KRZYSZTOF TEMPLIN

Depth, Shading, and Stylization in Stereoscopic Cinematography

A thesis for obtaining the title of Doctor of Engineering
of the Faculties of Natural Science and Technology
of Saarland University

September 2015
Saarbrücken, Germany

UNIVERSITÄT
DES
SAARLANDES

Supervisors
PD. Dr.-Ing. Karol Myszkowski
Prof. Dr. Hans-Peter Seidel

Dean
Prof. Dr. Markus Bläser

Colloquium

Date
26 February, AD 2016

Chair
Prof. Dr.-Ing. Philipp Slusallek

Reviewers
Prof. Dr. Hans-Peter Seidel
PD. Dr.-Ing. Karol Myszkowski
Prof. Dr. Wojciech Matusik

Academic assistant
Dr. Marek Vinkler

Abstract

Due to the constantly increasing focus of the entertainment industry on stereoscopic imaging, techniques and tools that enable precise control over the depth impression and help to overcome limitations of the current stereoscopic hardware are gaining in importance. In this dissertation, we address selected problems encountered during stereoscopic content production, with a particular focus on stereoscopic cinema. First, we consider abrupt changes of depth, such as those induced by cuts in films. We derive a model predicting the time the visual system needs to adapt to such changes and propose how to employ this model for film cut optimization. Second, we tackle the issue of discrepancies between the two views of a stereoscopic image due to view-dependent shading of glossy materials. The suggested solution eliminates discomfort caused by non-matching specular highlights while preserving the perception of gloss. Last, we deal with the problem of film grain management in stereoscopic productions and propose a new method for film grain application that reconciles visual comfort with the idea of medium-scene separation.

Kurzfassung

Aufgrund der ständig steigenden Beachtung der stereoskopische Abbildung durch die Unterhaltungsindustrie, gewinnen Techniken und Werkzeuge an Bedeutung, die eine präzise Steuerung der Tiefenwahrnehmung ermöglichen und Einschränkungen der gegenwärtigen stereoskopischen Geräte überwinden. In dieser Dissertation adressieren wir ausgewählte Probleme, die während der Erzeugung von stereoskopischen Inhalten auftreten, mit besonderem Schwerpunkt auf der stereoskopischen Kinematographie. Zuerst betrachten wir abrupte Tiefenänderungen, wie sie durch Filmschnitte hervergerufen werden. Wir leiten ein Modell her, das die Zeit vorhersagt, die für das menschliche Sehsystem notwendig ist, um sich an solche Änderungen der Tiefe zu adaptieren, und schlagen vor wie dieses Modell für Schnittoptimierung angewendet werden kann. Danach gehen wir das Problem der Unstimmigkeiten zwischen den zwei Ansichten eines stereoskopischen Bildes, infolge der sichtabhängigen Schattierung von glänzenden Materialien, an. Die vorgeschlagene Lösung eliminiert das visuelle Unbehagen, welches von nicht zusammenpassenden Glanzlichtern verursacht wird, indessen bewahrt sie die Glanzwahrnehmung. Zuletzt behandeln wir das Problem des Filmkornsmanagements in stereoskopischen Produktionen und schlagen eine neue Methode für das Hinzufügen vom Filmkorn vor, die die visuelle Behaglichkeit mit der Idee der Medium-Szenen-Trennung in Einklang bringt.

Summary

Stereoscopic 3D is a very compelling illusion, that allows for the depiction of objects and scenes with a unique sense of depth. Currently, it is receiving renewed attention, which is particularly evident in cinematography. However, stereoscopic imaging is plagued with numerous issues, ranging from technical limitations of the capturing, processing, and display equipment, issues with the visual comfort of the spectators, to problems with realism of depiction and artistic considerations. In this dissertation, we address three problems encountered in stereoscopic content production.

First, we consider abrupt changes of depth of the point of interest, such as those induced by cuts in films. For regular, non-3D films, the spectator changes the gaze direction using mostly saccades. However, stereoscopic films involve also changes in depth, which require changes of the vergence angle of the eyes, and these new movements are by nature much slower. At the same time, there is a clear tendency to decrease the length of shots in modern, non-3D films, possibly to better control the attention of spectators and increase their engagement, with the result, that little time is left for the visual system to adapt. Thus, if the filmmakers want to maintain this fast-paced editing style in stereoscopic productions, they need to be very careful about how they construct their shots. In an attempt to facilitate this process, we derive a model that predicts how much time is needed to adapt the vergence angle after a change in stereoscopic depth of the stimulus. Then, we propose to use this model as the cost function of the optimization procedure, that given a cut and a set of points of interest, minimizes the average vergence angle adaptation time.

Second, we tackle the issue of stereoscopic depiction of specular highlights on glossy surfaces. Since the highlights are view-dependent reflections of the light-sources in the scene, they have their own parallax and, consequently, their own position in depth, which differs from the position of the object they appear on. Moreover, the highlights often have inconsistent shape or topology across views, which makes them difficult to fuse. View-independent shading is a simple approach that improves visual comfort; however, it decreases the realism of the depiction. We propose an intermediate solution – *highlight microdisparity* – which removes major discrepancies in the highlights, but preserves their distinct

placement in depth. Our technique can be generalized to multiple reflections or other view-dependent effects, such as those observed in refractive media.

Last, we deal with the problem of stereoscopic film grain. Due to technical or artistic reasons, films often contain considerable amounts of film grain, and techniques for matching, adding, or removing it, play a significant role in the post-production process. Intuitively, grain should be treated independently in each view of a stereoscopic image. However, the visual system can deal only with limited amounts of uncorrelated grain. A state-of-the-art solution projects the grain onto the geometry of the scene, but this approach has certain drawbacks of perceptual and aesthetic nature, and it is unclear how to deal with "fuzzy surfaces", such as out-of-focus areas, sky, or light-scattering media. We propose a new grain placement method, which ensures that the grain can be still fused by the visual system, but does not have disadvantages of the on-surface approach.

A common theme in our techniques for specular reflections and film grain is that we reach an acceptable compromise between two state-of-the-art approaches. In order to maintain certain perceptual properties, we are less conservative about the treatment of the surfaces as having a unique stereoscopic depth, but the effect is kept within limits to avoid difficulties in binocular fusion.

Zusammenfassung

Das stereoskopische 3D ist eine sehr faszinierende Illusion, die die Darstellung von Objekten und Szenen mit einem einzigartigen Eindruck von Tiefe ermöglicht. Derzeitig bekommt es erneute Beachtung, dies wird besonders in der Kinematographie offensichtlich. Dennoch wird die stereoskopische Darstellung von zahlreichen Problemen geplagt, die sich von technischen Einschränkkungen der Aufnahme-, Bearbeitungs- und Ausgabegeräte, über Schwierigkeiten mit der visuellen Behaglichkeit der Betrachter, bis zur Problematik der realistischen Abbildung und artistischen Aspekten erstreckt. In dieser Dissertation adressieren wir drei Probleme, denen man bei der Produkion von stereoskopischen Inhalten begegnet.

Zuerst beschäftigen wir uns mit abrupten Änderungen der Tiefe des betrachtetenden Punktes, wie sie die durch Filmschnitte hervorgerufen werden. In herkömmlichen, nicht-3D Filmen, ändert der Betrachter seine Blickrichtung meistens durch Sakkaden. Jedoch beinhalten stereoskopische Filme auch Änderungen der Tiefe, die Änderungen des Vergenzwinkels der Augen erforderlich machen und diese neue Augenbewegungen sind ihrer Natur nach viel langsamer. Gleichzeitig gibt es eine eindeutige Tendenz die Länge der Einstellungen in modernen, nicht-3D Filme zu verkürzen, möglicherweise um die Beachtung der Zuschauer besser zu kontrollieren und um ihr Engagement zu erhöhen, mit der Auswirkung, dass wenig Zeit für das visuelle System bleibt um sich anzupassen. Falls Filmemacher diese schnelle Schnittfolgen in stereoskopischen Produktionen beibehalten wollen, müssen sie sehr sorgfältig sein wie sie ihre Einstellungen konstruieren. Um diesen Prozess zu erleichtern, leiten wir ein Modell her, das vorhersagt, wie viel Zeit notwendig ist um den Vergenzwinkel nach einer Änderung der Tiefe eines Stimulus anzupassen. Danach schlagen wir vor dieses Modell als Kostenfunktion eines Optimierungsverfahrens, das für einen beliebigen Schnitt und eine Menge Punkte des Interesses die durchschnittliche Vergenzwinkelanpassungszeit minimiert, zu verwenden.

Darauf fassen wir das Problem der stereoskopischen Darstellung von Glanzlichter auf glänzenden Materialien an. Da Glanzlichter sichtabhängige Reflexionen von Lichtquellen in der Szene sind, haben sie eine eigene Parallaxe und infolgedessen eine eigene 3D-Lage, die sich von der Lage des Objekts auf dem

sie erscheinen unterscheidet. Überdies haben Glanzlichter in verschiedenen Ansichten eines stereoskopischen Bildes oft nicht zusammenpassende Formen oder Topologien, wodurch die binokulare Fusion schwierig wird. Sichtunabhängige Schattierung ist eine einfache Vorgehensweise, die die visuelle Behaglichkeit erhöht, jedoch vermindert sie den Realismus der Darstellung. Wir schlagen eine vorläufige Methode vor – die Glanzlichtmikrodisparität (Eng. *highlight microdisparity*) – die die größeren Unstimmigkeiten der Glanzlichter aufhebt, aber ihre unterschiedliche 3D-Lage bewahrt. Unsere Methode lässt sich für den Fall der mehrfachen Reflexionen oder anderen sichtabhängigen Effekten wie beispielsweise jene, die in refraktive Medien beobachtet werden können, verallgemeinern.

Zuletzt behandeln wir die Problematik des stereoskopischen Filmkorns. Aus technischen oder artistischen Gründen beinhalten Filme häufig eine erhebliche Menge an Filmkorn und Verfahren für dessen Anpassung, Auftrag, oder Entfernung spielen eine wichtige Rolle in der Nachbearbeitung. Intuitiv sollte das Filmkorn in beiden Ansichten eines stereoskopischen Bildes unabhängig behandelt werden, jedoch kann das visuelle System nur eine begrenzte Menge unkorreliertes Filmkorn handhaben. Eine fachübliche Lösung projiziert das Filmkorn auf die Geometrie der Szene, aber eine solche Vorgehensweise hat gewisse Nachteile perzeptueller und ästhetischer Natur, und es ist unklar, wie "verschwommene Oberfläche", wie beispielsweise Unschärfen, der Himmel oder lichtstreuende Medien behandelt werden sollen. Wir schlagen eine neue Methode für die Positionsbestimmung des Filmkorns vor, die sicherstellt, dass das Filmkorn vom visuellen System fusioniert werden kann, aber nicht die Nachteile der projektiven Vorgehensweise hat.

Ein gemeinsames Motiv unserer Methoden für Glanzlichter und Filmkorn ist ein akzeptabler Kompromiss zwischen zwei fachüblichen Vorgehensweisen. Um bestimmte perzeptuelle Eigenschaften zu erhalten, sind wir weniger strikt im Bezug auf die Eindeutigkeit der Tiefe von Oberflächen in der Szene, dennoch halten wir den Effekt in Grenzen, um Probleme mit der binokularen Fusion zu vermeiden.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Karol Myszkowski for his continuous support and encouragement, for always believing in me, and for never saying 'no' to my unconventional research ideas. I would like to thank my co-supervisor Hans-Peter Seidel, the director of the Computer Graphics Group at MPI Informatik for creating such a great research environment and letting me be a part of it. I am indebted to Wojciech Matusik for giving me the opportunity to work at MIT CSAIL and co-supervising my research on the eye vergence model, which became a significant part of my dissertation.

I would like to thank all my collaborators and colleagues from MPI Informatik and MIT CSAIL. I am especially grateful to Piotr Didyk, who collaborated with me on nearly all my scientific projects. Without his advice and our innumerable discussions my doctorate certainly would not have been as successful. I would like to thank Tobias Ritschel who helped me kick-start my doctorate and provided the programming framework that I used to generate some of the results presented here. I am grateful to Aude Oliva for providing access to the eye-tracker, and to Lavanya Sharan and Zoya Bylinskii for showing me how to use it. Zoya, Silke Jansen, and Christian Richardt helped me by proof-reading parts of my disseration; of course, any remaining errors are only mine. Finally, I would like to thank all the anonymous heroes who took part in my perceptual experiments.

Of equal importance was the moral support provided throughout my doctoral studies by my family and friends. I am especially grateful to my parents, Barbara and Lech, my sister Ela, and Fr. Johannes Kreier, who were always there whenever I needed them.

Contents

Chapter 1

Introduction

When we observe a real-world scene, each of our eyes sees it from a slightly different vantage point, which means that relative distances between images of objects when projected onto the retina are different in each eye. The human visual system uses these differences as one of the sources of information about the depth in the scene, and the depth impression resulting from this process is called *binocular stereopsis*. The principle of stereoscopic imaging – the idea that one can present two separate images dichoptically (i. e., simultaneously, one to each eye) to evoke the illusion of three-dimensionality – has been known at least since the first half of the 19th century. However, to this day stereoscopic imaging remains a very problematic medium, and the aim of our work is to advance the state of the art in several selected aspects. We start this chapter with an overview of the history of stereoscopic imaging (Sec. 1.1), including its use in cinematography (Sec. 1.2) and other areas (Sec. 1.3). Next, we briefly discuss the technical aspects of stereoscopic imaging (Sec. 1.4) and review stereoscopic display technologies (Sec. 1.5). After this short introduction, we summarize the novel contributions of this dissertation (Sec. 1.6).

1.1 History of Stereoscopy

As noted by Brewster [1856, p. 6], the fact that "the pictures of bodies seen by both eyes are formed by the union of two dissimilar pictures formed by each" was known and published already by ancient mathematicians, such as Euclid. Brewster goes on to note that the subject of binocular vision was treated in detail in the second century A.D. by a Greek physician Galen, and discusses works on the subject by Leonardo da Vinci and François d'Aguilon, among others. However, the idea that one can present two separate images dichoptically, that is one to each eye, in order to obtain an illusion of three-dimensionality (a "relief"), surfaced only in the 1830s: The first stereoscope was constructed by Charles Wheatstone,

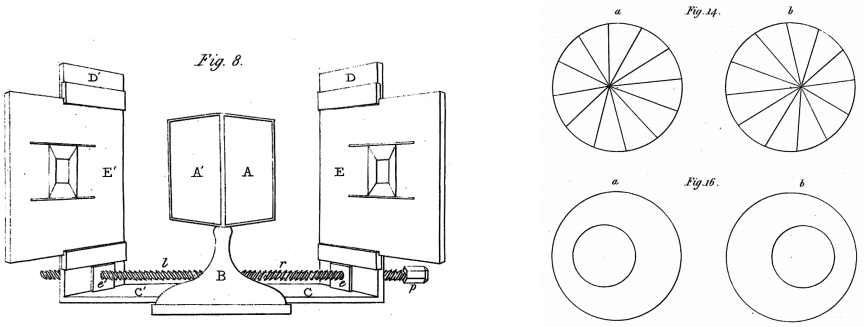


Figure 1.1. A schematic drawing of the Wheatstone stereoscope, likely the first stereoscope ever constructed *(left)* and simple abstract stereoscopic images used by Wheatstone in his experiments *(right)*. Drawings: Wheatstone [1838]

who presented his invention to the Royal Society in 1838[1] [Le Conte Stevens 1882]. In its center, the stereoscope had two mirrors at the right angle, reflecting the lines of sight to the left and the right side of the instrument, where simple line drawings were placed (Fig. 1.1). More complicated figures were not used to demonstrate, that the impression of a "relief" is not due to skillful shading but solely due to differences between the images presented to the left and the right eye.

This design was improved by Brewster, who in 1849 went on to construct his own, more compact stereoscope (Fig. 1.2, top-left) that utilized lenses instead of mirrors [Le Conte Stevens 1882]. Brewster's stereoscope was presented during The Great Exhibition in 1851 and notably admired by the British queen Victoria [Stafford et al. 2001, p. 357]. His design became quite successful, and according to Brewster [1856, p. 36] over half a million exemplars had been sold by 1856.

Endless variations and improvements of this design have been proposed, including, but not limited to, the Holmes stereoscope, the more recent View-Master stereoscope, or cardboard stereoscopes for smart phones (Fig. 1.2).

1.2 Stereoscopic Cinema

Stereoscopy is of course not limited to static imagery, but can be naturally applied to sequences of images to produce stereoscopic films and animations. The illusion of stereoscopic 3D is a very compelling one and as argued by Mendiburu [2009, p. 3], it reduces the effort involved in the suspension of disbelief and thus significantly increases the immersion experience. It also seems to affect the

[1] Brewster [1856, p. 19] writes that Elliot, who in 1834 or earlier decided to build a very simple stereoscope, was first, but he did not accomplish the task until the year 1839. His instrument was a simple wooden box, which did not contain any optical elements, and the union of the images was achieved through free fusion. It was therefore more of a case for a hand-drawn stereogram, rather than a real instrument.

Figure 1.2. Clockwise from top-left: Hand-held Brewster's stereoscope with lens. An American version of the stereoscope, designed by Holmes. The View-Master stereoscope. Google Cardboard stereoscope with a Nexus 5 smartphone. Pictures: Le Conte Stevens [1882], www.captainbluehen.com, www.google.com/get/cardboard

emotions of the viewer in a more profound way, as Sandrew [2012] suggests, that during stereoscopic watching visual information affects lower order emotional areas of the brain more directly than for a regular film, and evokes a more primal, subconscious response. Despite these advantages, stereoscopy – unlike other improvements to the cinematic art, such as artificial lighting, sound, or color – seemingly has failed to make the leap from novelty to standard practice [Higgins 2012]. Interestingly, its popularity comes and goes in regular, thirty-year intervals: after the dawn in the 1920s, and the crazes of the 1950s and 1980s, it is once again gaining attention [Tachi 2013]. Both film theorists and practitioners often point to the fact, that with honorable exceptions such as Hitchcock's *Dial M for Murder*, early stereoscopic films were characterized by gratuitous effects and gimmickry. However, this seems to be changing, as stereoscopic depth is nowadays used with more restraint, and it is more often employed to support film narration, e. g., it is set to match emotional states of the characters or to underline relations between them [Neuman 2009, Atkinson 2011, Higgins 2012]. On the technical side, early stereoscopic productions suffered from the imperfections of the technology, such as misalignment of the acquisition and projection systems, leading to visual fatigue experienced by the audience [Lipton 1982, p. 12]. Recent technological advances have made full digital intermediate or even entirely digital pipeline feasible options, and now filmmakers have tools enabling production of content of unprecedented quality. Some authors argue, that due to this fact, the stereoscopic cinematography is finally here to stay [Seymour 2008]. Although current display systems are far from perfect, and numerous problems (such as the conflicts of depth cues or the need of special eye-wear) remain largely unsolved, this prediction seems to be confirmed by the numbers. The count of "3D-capable" theaters around the world has been steadily increasing and every year more stereoscopic films are released [Acuna 2013].

1.3 Other Applications

In this dissertation we focus mostly on the use of stereoscopic imaging in modern cinematography, and, to a lesser extent, in computer games. It is worth mentioning, however, that stereoscopy has found its use also in other areas, such as fine arts, science, or medicine. Preparing a good stereogram requires a fairly accurate reproduction of the geometry in both half-images, since geometric inaccuracies are much more apparent in stereoscopic images. Thus, photography seems to be a great medium for execution of stereograms, and stereoscopy probably would not have become so popular if not for the invention of photography [Le Conte Stevens 1882]. Nevertheless, there have also been attempts in painting stereoscopic images by hand, such as the abstract works of Oskar Fischinger in the late 1940s [Zone 2014, p. 173], Salvador Dalí's paintings of the 1970s [Seckel 2004, p. 34], and numerous stereoscopic adaptations of comic books by Ray Zone [Barnes 2012]. Ferragallo [1995] explored the idea of stereoscopic tiling in archi-

tecture, and Biegon [2005] experimented with "twin-reliefs" – stereoscopic artworks, in which the half-images are relief sculptures rather then two-dimensional reproductions. Besides the fine arts, stereoscopic imaging is commonly used as a visualization tool in various sciences. To give a few examples, in geology stereoscopic photographs are used to document both landscapes, as well as smaller-scale objects such as fossils [Allaby 2013]. In chemistry, stereoscopic images are used when the three-dimensional aspects of a compound are important [Gerig 1974]. A notable example from medical sciences is the monumental, 10-volume Edinburgh Stereoscopic Atlas of Anatomy by Cunningham et al. [1911]. The stereopsis is in itself an extensively studied aspect of the human vision [Julesz 1971]. An important tool in such research are random dot stereograms, where the depth impression is evoked solely by the binocular disparity, in isolation from any other sources of depth information [Julesz 1964]. Stereoscopic imaging has been successfully exploited in treating stereo-blindness [Barry and Sacks 2010], and stereoscopic computer games for treatment of amblyopia and strabismus have been developed[2].

1.4 Basics of Stereoscopic Imaging

The principle of stereoscopic imaging is to draw, capture, or render two images of the same scene from two different viewing positions, and then show them simultaneously to the observer, one image to each eye (dichoptic presentation). The two images are called *half-images* or *views*, and when taken together, they are referred to as a *stereo pair*, a *stereo image*, or a *stereogram*. Due to the change of the vantage point, any given object may assume a different position within each half-image, and this difference of coordinates is called *disparity*. The half-images forming a stereo pair on their own are conventional images that can be watched separately as any regular, monoscopic image. The "3D effect" appears only when they are combined using appropriate display equipment, with the depth impression being the result of our visual system interpreting the disparities in the stereo pair. In a sense, stereoscopy "tricks" the visual system into fusing retinal images of two distinct objects instead of one, as is the case under normal viewing conditions.

Most stereoscopic images are pairs of rectilinear projections of a (virtual or real) scene formed by two (virtual or real) cameras. The viewing directions of the cameras can be set to converge at a certain point (toe-in arrangement) or they can be kept parallel (see Fig. 1.3). The toe-in arrangement, however, tends to introduce two non-matching keystone distortions, resulting in depth plane curvature and vertical disparities of objects in the stereo pair (see Fig. 1.4). Vertical disparities are hard to fuse, and thus this arrangement is geometrically inferior to the parallel one, in which objects having equal distances from the plane of projection (common for both cameras) have equal and purely horizontal

[2] http://www.seevividly.com/

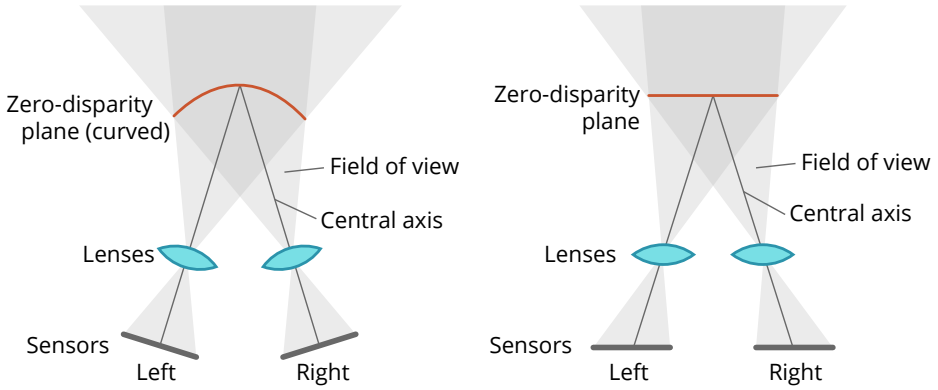


Figure 1.3. Comparison of the toe-in stereo-camera set-up *(left)* and parallel stereo-camera set-up *(right)*. In the former configuration the cameras are directed towards the point of interest, which results in depth plane curvature, as indicated by the solid red line. In the latter configuration, sensors (projection planes) of both cameras are co-planar, and points with equal distances have equal disparities.

disparities in the stereo image [Lipton 2010]. Hereafter, we will always assume a parallel camera set-up.

Stereoscopic imaging can be seen as a three-step transformation: first, scene-space coordinates of an object are transformed into its sensor coordinates; next, the sensor coordinates are re-scaled to display coordinates; and finally, the display coordinates determine the placement of the object in the three-dimensional image space. Thus, the apparent distance to any individual object in a stereogram depends on the geometries of the camera setup and the display setup. As illustrated on the left in Figure 1.5, the parameters defining the camera setup are: (i) interaxial distance b, i. e., the distance between the cameras, (ii) the convergence distance c_0, i. e., the distance to the intersection of central axes, which is controlled by symmetric, horizontal shifts of sensors, and (iii) the field of view, as determined by the sensor width w_c and the focal length f. For the object at distance c from the convergence plane (thus at distance $c_0 + c$ from the cameras), the transformation to camera disparity p_c is given by the equation:

$$p_c = \frac{bc}{c + c_0} \cdot \frac{f}{c_0}. \tag{1.1}$$

On the display side, as shown on the right in Figure 1.5, the three parameters determining the depth impression are (iv) interocular distance e, i. e., the distance between the observer's eyes, (v) distance to the display d_0, (vi) and display size w_d. The conversion from camera disparity p_c to display disparity p_d is performed by scaling it with the ratio $w_d \cdot w_c^{-1}$. Given display disparity p_d, perceived distance d from the display (thus the distance $d_0 + d$ from the observer) is given by the

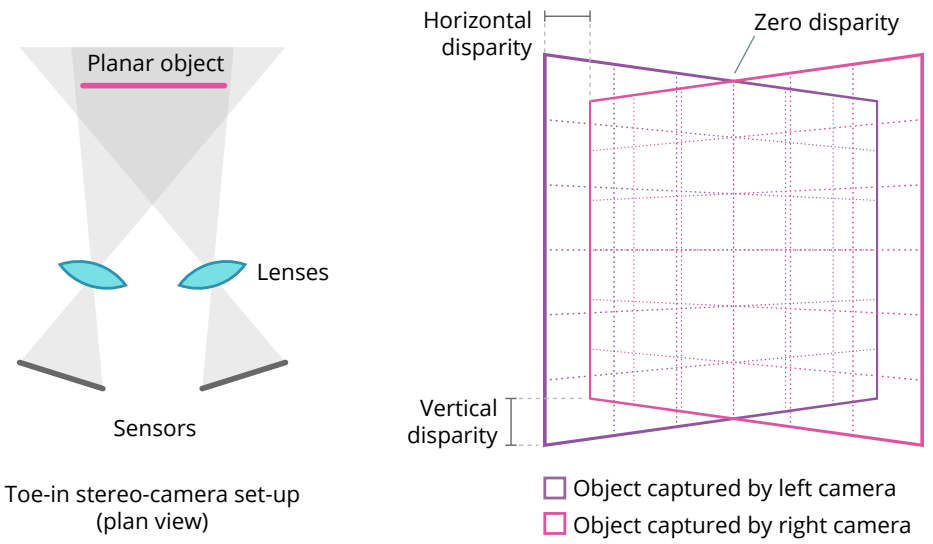


Figure 1.4. Keystone distortion in stereoscopic images. Toe-in stereo-camera set-up *(left)* results in vertical disparities and unequal horizontal disparities for equidistant points, causing perceived depth plane curvature *(right)*.

equation:

$$d = \frac{d_0 p_\mathrm{d}}{e - p_\mathrm{d}}. \tag{1.2}$$

Equations 1.1–1.2 are general and remain valid for negative c and d, corresponding to objects in front of the convergence plane and in front of the display, respectively.

The particular situation, known as the *orthostereo condition*, in which the camera convergence distance and the field of view correspond to the distance and the visual angle subtended by the display, and the camera interaxial distance is equal to the interocular distance of the observer provides an exact, one-to-one re-creation of the captured scene in terms of the binocular stereopsis [Phillips 2010, p. 404][3]. Adjusting any of the three camera parameters while keeping the display setup unchanged introduces depth distortions to the perceived image, which may become objectionable when too extreme. Analogously, distortions appear when the presentation conditions are changed (e. g., different display device is used) without adapting the content accordingly. In particular, since the relation between disparity and depth impression is not linear (Equation 1.2), shape distortions are also introduced by simple re-scaling of the the display size. For a thorough analysis of the relation of the camera and display parameters to

[3] Phillips assumes inifinite convergence distance of cameras which is corrected by a horizontal translation during display. The two formulations are equivalent; however, setting the proper convergence during acquisition is preferred, because it avoids loss of data at the sides of the image [see Woods et al. 1993, p. 2].

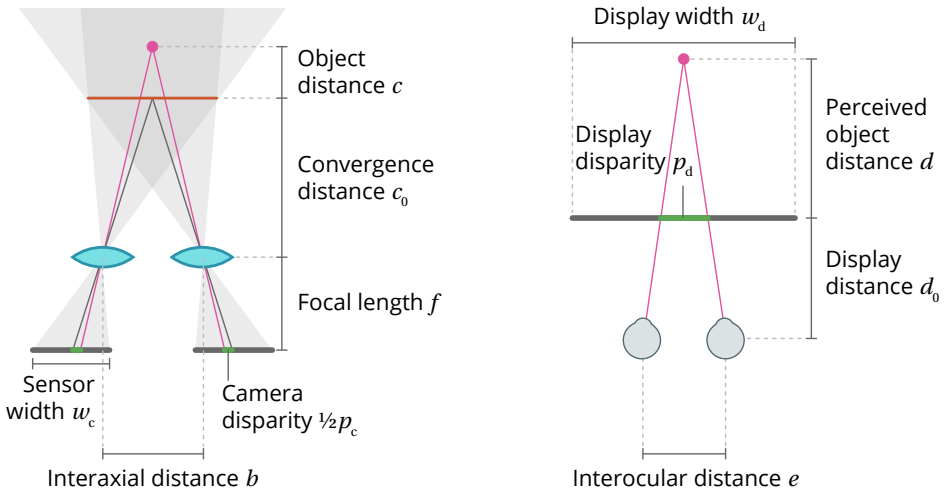


Figure 1.5. An illustration of the camera setup geometry *(left)* and the display setup geometry *(right)*. The conversion between camera disparity p_c and display disparity p_d is determined by the equation $p_d = p_c \cdot w_d \cdot w_c^{-1}$.

the depth perception we refer the reader to the work by Woods et al. [1993]. The special-case problem of adapting stereoscopic broadcasts of field sports to the viewing conditions was addressed in our recent work with Calagari et al. [2014].

A note on disparity So far we have been using the term 'disparity' mostly to denote the difference between coordinates of homologous points within the *stereoscopic image*, with positive distances by convention corresponding to points behind the display plane, and negative distances corresponding to points in front of the display plane (Fig. 1.6, left). For clarity, we sometimes refer to this quantity using the term *display disparity* (or *camera disparity* when talking about the image as captured by the sensors). This should be distinguished from the closely-related term of *binocular disparity* used in vision science, denoting the difference between angular coordinates of homologous retinal projections, with positive distances corresponding to points outside the isovergence circle and negative distances – inside the isovergence circle (Fig. 1.6, right). Although using the angular measure is more correct when discussing binocular perception, in many practical applications binocular disparities are well approximated by display disparities, and it is common in cinematography to use the latter measure [Mendiburu 2009, pp. 84–86]. We follow this convention in Chapter 3, which deals with eye vergence response modeling.

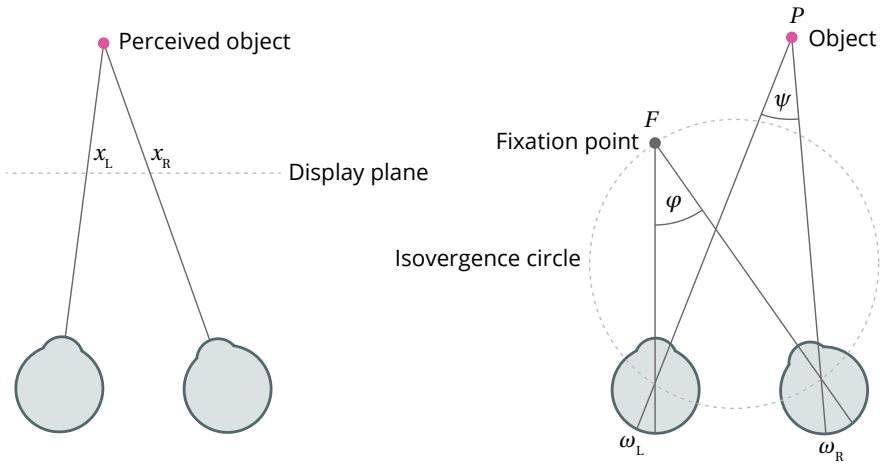


Figure 1.6. Two types of disparity. In stereoscopic imaging *(left)*, the display disparity of an object is the difference between its coordinates in the right and the left half-image ($x_R - x_L$). In vision science *(right)*, the binocular disparity of object P is the angular difference between its retinal coordinates in the right and left eye ($\omega_R - \omega_L$). It can be also expressed as $\varphi - \psi$, where φ is the vergence angle and ψ is the binocular subtense of P. The isovergence circle, passing through fixation point F and the nodal points of the eyes, is the locus of zero binocular disparity.

1.5 Types of Display Systems[4]

The simplest way to display a stereo image is to show the two half-images side by side, and let the observer change his/her eye vergence angle, so that the half-images overlap (so-called *free fusion*). The half-images can be shown in the left-right order (divergent free fusion) or in the right-left order (convergent free fusion). This technique requires no special equipment, but many people are not capable of freely adjusting the vergence angle. In particular, some people are able to perform only divergent free fusion or only convergent free fusion.

Although lens-based and mirror-based stereoscopes are still in use, especially in optometry and vision science, many other techniques have been developed, ranging from very simple ones, to ones involving sophisticated engineering. A common feature of these systems is the reliance on multiplexing of the half-images, so that they can be displayed in the same physical location. This allows for a compact design and more convenient viewing, but often at the price of reduction of the visual quality of the signal, e. g., inferior color reproduction, brightness, or resolution. An important issue of the systems based on multiplexing is cross-talk, i. e., leakage of some portion of the signal intended for one eye to the other eye [Woods 2012]. If the amount of cross-talk is significant, it can hinder the binocular fusion and in extreme cases spoil the 3D effect completely.

[4]The review of stereoscopic image presentation methods is partially based on our work published elsewhere [Templin 2016].

Since low cross-talk is critical for the appreciation of our techniques, we present all results in the left-right, side-by-side format, suitable for divergent free fusion. Additionally, digital stereograms are available for download on the corresponding project websites (see Sec. 1.6 for the URLs).

Anaglyph images, first described by Rollmann in 1853, are probably the most wide-spread multiplexing-based technique of presenting stereo images. Here, the half-images are reproduced using different color palettes and superimposed using additive color mixing. The resulting image is viewed using glasses with two color gels, each transmitting the light from the corresponding half-image, while blocking the light from the other one. Various color combinations are used, such as green-magenta or red-blue, with the red-cyan combination being the most popular. Besides the glasses, this method does not require any specialized equipment, such as a dedicated printer or display, which makes it very inexpensive and convenient. On the downside, the colors of the image need to be modified, thus this method suffers from poor color reproduction. Moreover, each view is transformed in a different way, so the outcome may exhibit significant discrepancies in color, which in turn cause unpleasant retinal rivalry. Finally, since the filters usually do not separate the two half-images perfectly, cross-talk may appear.

Several algorithms have been proposed to enhance the quality of anaglyph images. A trade-off between the color reproduction and viewing comfort can be made, by using various color conversion matrices [Hainich and Bimber 2011, pp. 384–387]. If the spectral absorption curves of the gels and the spectral distributions of the display primaries are known, the task of generating an anaglyph image that matches as closely as possible the input stereo pair can be posed as an optimization problem [Dubois 2001]. By performing a few color matching tasks, the color transformation can be adjusted to reduce cross-talk [Sanftmann and Weiskopf 2011]. Proprietary systems have been developed, such as ColorCode 3-D[5], combining advanced image processing techniques with careful color gel selection.

A significant improvement in the color reproduction can be achieved by light polarization. The half-images are displayed using polarized light, and corresponding polarization filters in the glasses block out the light of non-matching polarization. In the theatrical environment this requires a polarization-preserving screen. In yet another approach, developed by Infitec and used in Dolby 3D systems, each half-image is displayed using different sets of primaries with slightly different wavelengths. Each filter transmits all corresponding primaries, but in narrow, disjoint sub-bands to prevent cross-talk. This approach reduces the color gamut to some extent, it eliminates, however, the need for polarization-preserving screens [Jorke and Fritz 2006, Hainich and Bimber 2011, pp. 386–387].

The systems described so far exploit various passive light filtering techniques, but also an active approach is possible. Systems using so-called shutter glasses,

[5] http://www.colorcode3d.com/

interleave the half-images temporally: the display shows the left and the right view in alternation, and synchronized glasses transmit the signal only to the corresponding eye, while blocking out the other eye. Such systems require displays of sufficiently high frame rate to prevent noticeable flickering, since the effective frame rate for each eye is halved. A popular example of a consumer-grade, shutter-based system is Nvidia 3D Vision[6]

Finally, so-called *autostereoscopic* displays eliminate the eye-wear completely. By covering the image surface with lenslet arrays or parallax barriers one can ensure, that only selected pixels are visible to each eye. This technique can be generalized to *automultiscopic* displays, which reproduce more than two views, by trading the spatial resolution of the image for increased angular resolution. Alternatively, one can obtain full spatial resolution by sacrificing temporal resolution thanks to time multiplexing [Hainich and Bimber 2011, pp. 395–401]. Automultiscopic displays are more robust to changes of the observer's position and provide limited reproduction of the head motion parallax. However, their resolution is still quite limited, and possibilities to improve the user experience by manipulating the content have been investigated [Zwicker et al. 2006, Didyk et al. 2013, Du et al. 2014]. Implementing software resolution enhancement techniques [Templin et al. 2011] could be potentially helpful in this regard, too. To support correct accommodation, a multiscopic display needs to reach angular resolution at which the pupil of the eye is covered by a signal from multiple views, however, the brute-force approach of simply increasing the resolution incurs very high hardware costs. An alternative solution is to employ multi-focal displays, effectively placing several conventional displays at different depths by means of mirrors and beam splitters [Akeley et al. 2004]. One such design has been successfully used in vision research to show influence of the accommodation-vergence conflict on the visual comfort [Hoffman et al. 2008]. For a more extensive and detailed survey of advanced displays, including various volumetric designs we refer to the work by Masia et al. [2013].

1.6 Our Contributions

Despite the significant improvements in display devices as well as in image generation, capture, and post-processing techniques, many consumers as well as film makers are still skeptical about the quality of current stereoscopic content and the future of the technology itself. These concerns are usually related to naturalness, effortlessness, and overall appearance. In general, it is not sufficient to produce two good images in place of one to arrive at a good stereoscopic effect, which imposes many restrictions on the production and post-production process [Zilly et al. 2011]. Although the presence of binocular disparity in a certain way brings the percept closer to reality, it might also accentuate the lack of other depth cues. For instance, commonly used display systems do not reproduce the

[6] http://www.nvidia.com/object/3d-vision-main

effects of eye accommodation or head motion parallax, which leads to visual discomfort and geometry distortions, respectively. In particular, any conscious or subconscious head movements meant to adjust the vantage point have no effect, which is likely to be important in the context of view-dependent shading. Moreover, in cinematography the orthostereo condition is rather an exception than the rule, and one has to deal with various undesired effects, such as apparent up- and down-scaling of objects or so-called card-boarding. To ensure a pleasant viewing experience, geometry-bending tricks such as horizontal image translation, multirigging, or non-linear disparity mapping are routinely used [Mendiburu 2009, pp. 83, 109–110, 129–137]. Finally, due to the added dimension, virtually any visual effect needs to be applied with much greater attention to detail to avoid visibility of artifacts. Thus, there are numerous issues that need to be addressed during creation of a stereoscopic film, ranging from rather technical, comfort-related ones, such as distribution of disparities or coherence of the stereo views, to more artistic ones, such as perception of shapes. In this dissertation we do not limit ourselves to a certain class of issues, but cover both technical as well as aesthetic aspects, and propose novel solutions to selected problems one encounters in stereoscopic cinematography. Since a large number of stereoscopic productions are not shot natively but are post-converted[7], we pay special attention to the applicability of the presented techniques to the context of 2D-to-3D conversion. Although our methods can be used with the default parameter settings and minimal user intervention, film making is a creative process, in which one rarely settles for fully automatic procedures. Thus, our main intent was to provide film makers with new tools that lend themselves to interactive use, rather than with "one-click solutions". Our work is based on three articles we previously presented at conferences and published in international journals [Templin et al. 2014a, 2012, 2014b], the scope and contributions of which are outlined below. We omit results of our work with Ritschel et al. [2012] and with Calagari et al. [2014] as less relevant to the dissertation topic.

1.6.1 Eye Vergence Model [Templin et al., Siggraph 2014]

Sudden temporal depth changes, such as cuts that are introduced by video edits, are usually not encountered in the real world. Moreover, the visual system is constantly forced to adapt the vergence angle to new display disparities in spite of conflicting accommodation requirements. Thus, rapid depth changes are potentially very challenging for the audience and can significantly degrade the quality of stereoscopic content. They may lead to confusion, reduced understanding of the scene, and overall attractiveness of the content. Often the problem cannot be solved by matching the depth around the transition, as this might lead to objectionable flattening of the scene. The novel contribution of this line of our work is a series of eye-tracking experiments we conducted to better understand this

[7] http://www.realorfake3d.com/

limitation of the human visual system. The data we obtained allowed us to derive and evaluate a model describing adaptation of vergence to disparity changes on a stereoscopic display. Besides computing user-specific models, we also estimated parameters of an average observer model, the unique characteristic of which is its data-driven foundation and expression with an analytic formula. Our model enables a range of strategies for visualizing and controlling (in particular minimizing) the adaptation time of the audience. Additional materials related to this part are available on-line at http://resources.mpi-inf.mpg.de/VergenceModel/.

1.6.2 Specular Highlights Disparity [Templin et al., Siggraph 2012]

Human stereo perception of glossy materials is substantially different from the perception of diffuse surfaces: a single point on a diffuse surface appears the same for both eyes, whereas for specular surfaces its appearance differs. Since highlights are blurry reflections of light sources they have depth themselves, which is different from the depth of the reflecting surface. We call this difference in depth impression "highlight disparity". Due to artistic motivation, for technical reasons, or because of incomplete data, stereoscopic highlights are often treated as a view-independent effect and thus placed on the surface, without any disparity. However, it has been shown that lack of disparity decreases the perceived glossiness and authenticity of a material. We try to remedy this contradiction by introducing a novel technique for depiction of glossy materials, which improves over simple on-surface highlights, and avoids problems of geometrically correct reflections. The proposed approach is computationally simple, can be easily integrated in an existing (GPU) shading system, and allows for local and interactive artistic control. We evaluate our contribution in a subsequent perceptual study and briefly discuss an extension to refractive/reflective objects with multiple ray-tracing events [Dąbała et al. 2014]. Additional materials are available on-line at http://resources.mpi-inf.mpg.de/HighlightMicrodisparity/.

1.6.3 Stereoscopic Film Grain [Templin et al., Pacific Graphics 2014]

Independent management of film grain in each view of a stereoscopic video can lead to visual discomfort. The existing alternative is to project the grain onto the scene geometry. Such grain, however, looks unnatural, changes object perception, and emphasizes inaccuracies in depth arising during 2D-to-3D conversion. We propose an advanced, novel method of grain positioning that scatters the grain in the scene space. In a series of perceptual experiments, we estimate the optimal parameter values for the proposed method, analyze the user preference distribution among the proposed and the two existing methods, and show influence of the method on the object perception. See http://resources.mpi-inf.mpg.de/FilmGrain/ for additional materials.

Chapter 2

Related Work

In this chapter we review previous work relevant to the topic of this dissertation, starting from the existing basic research pertaining to the human visual system and perception in order to provide the background to our own work and to motivate the design choices of the presented techniques and perceptual experiments. We shall also give an overview of related research in image processing and computer graphics. In the first, more physiology-oriented section, we give an overview of results related to the functioning of the eyes during stereoscopic viewing of a depth-changing stimulus (Sec. 2.1). Specifically, we discuss the properties and models of eye vergence movements, coupling of vergence and accommodation, and relation of temporal changes in the content to binocular fusion performance and visual comfort. We refer the reader to a survey by Meesters et al. [2004] for an in-depth discussion of other aspects of stereoscopic display perception. Next, we proceed to findings in binocular perception of lustrous (glossy) surfaces to underline the importance of binocular cues for the correct material perception and to justify the functioning of the proposed specular highlight rendering technique (Sec. 2.2). Then, we provide perceptual background on the binocular vision of stimuli which show structural similarity to film grain. This way we are able to motivate our choice of the grain representation structure, which enables its comfortable viewing as a distinct volumetric structure, while facilitating efficient and simple implementation of the compositing algorithm (Sec. 2.3). Next, we shift our focus to the works addressing the problem of analysis and processing of the scene's depth structure (Sec. 2.4), and we also look into the topic of glossy and view-independent materials in rendering (Sec. 2.5). In the last section of this chapter, we discuss the uses of noise in computer graphics and review works on modeling of film grain. We also touch upon stereoscopic stylization and point volumes in data visualization (Sec. 2.6).

2.1 Eye Vergence

Eye vergence is triggered by the depth changes of a fixation target, and can be performed with high accuracy both in the real world and stereoscopic display observation conditions. It is mostly driven by retinal disparity, with other factors, such as blur or proximity cues affecting it to a lesser extent [Horwood and Riddell 2008]. Vergence is a relatively slow process when compared to other eye movements, e. g., saccades (below 60 ms), and requires about 195–750 ms for convergence and 240–1000 ms for divergence [Semmlow and Wetzel 1979, Krishnan et al. 1977]. Vergence latency also seems to demonstrate an asymmetric behavior (180–250 ms for convergence and 190–210 ms for divergence). However, there has been some controversy whether convergence latency is greater or less than divergence latency [Krishnan et al. 1973, Semmlow and Wetzel 1979]. Alvarez et al. [2005] found that divergence dynamics are dependent on the initial fixation distance.

Vergence is a two-stage process, where at first the fast transient (a.k.a. phasic) mechanism (reacts even for brief 200 ms flashes) brings the vergence in the proximity of the target depth, and then the slower sustained (a.k.a. tonic) mechanism is responsible for the precise verging on the target, as well as further tracking of slower depth changes. Semmlow et al. [1986] found that for less dynamic depth changes, with the ramp velocity below 2 deg/s, only the sustained mechanism is active, above 9 deg/s the transient mechanism dominates, and otherwise both mechanisms are active. For small depth changes within Panum's fusional area, the motoric vergence is not activated, and sensoric fusion of images on the retina is sufficient. Vergence adaptation (similar to luminance adaptation) has been observed, in which the sustained mechanism supports a given eye vergence angle [Hung 1992].

There is a large body of research on measurements of vergence in response to pulse, step, ramp, and sinusoidal disparity stimuli. For us, the step-like changes are the most relevant. Most experiments used physical targets or passively-shifted screens [Hung et al. 1997]. Simple stimuli such as vertical lines were used to eliminate other cues that could affect vergence. Special care was taken to suppress accommodation by using pinhole apertures for blur-free viewing. A wide range of disparities ±35 deg have been considered [Erkelens et al. 1989], but a typical range was below ±10 deg with relatively large step amplitudes.

In our work we focus on the disparity steps within the smaller range of ±2.5 deg, since stimuli outside this range are likely to cause visual fatigue in stereoscopic display conditions. By using an off-the-shelf stereoscopic display in our measurements, we ensure that the conditions are possibly similar to the ones in expected applications, where accommodation conflict may affect the vergence [Vienne et al. 2014]. For similar reasons, we validate our model using real-world images, to account for the influence of pictorial cues. In addition to the step magnitude, the initial display disparity is important in our measurements, both for convergence and divergence.

Vergence vs. Accommodation When the depth of the stimulus is changed, not only the vergence distance, but also the accommodation (focusing) distance needs to be adapted. While the vergence control system is driven mostly by retinal disparity, and the accommodation control system is primarily retinal blur-driven, both systems are coupled via reflexive cross-link interactions – accommodative convergence and convergence accommodation, described by AC/A and CA/C ratios, respectively [Lambooij et al. 2009, Fig. 1]. The AC/A ratio quantifies the change in vergence due to accommodation in the absence of retinal disparity, whereas the CA/C ratio – the change in accommodation caused by vergence in the absence of retinal blur. Since the accommodation distance for typical stereoscopic displays is always constant, and thus usually inconsistent with the vergence distance, stereoscopic viewing requires unnatural decoupling of the vergence and accommodation systems. When the display disparity changes and the sensoric fusion is not possible anymore, the vergence system adapts the vergence angle to reduce the retinal disparity, and at the same time drives the accommodation away from the screen (convergence accommodation). If the resulting retinal blur is sufficiently large, the accommodation system reacts to counteract the loss of sharp vision, thereby driving the vergence back towards the display (accommodative convergence) [Lambooij et al. 2009]. In contrast to stereoscopic displays, real world objects away from the fixation point appear blurred, which postpones diplopia, since the limits of fusion increase for lower spatial frequencies.

Existing research demonstrates unstable behavior of the visual system under stereoscopic conditions as compared to real-target conditions [Okuyama 1998, Ukai and Kato 2002]. Hoffman et al. [2008] constructed a multi-plane display, that allowed them to separately control focal and vergence distances, and proved the vergence-accommodation conflict to be one of the sources of visual fatigue induced by stereoscopic displays. Additionally, they showed that this conflict hinders binocular fusion performance. Shibata et al. [2011] provided estimates of the range of vergence distances around the screen that ensure comfortable viewing experience (*comfort zone*). A rule of thumb frequently used in stereo acquisition is that the comfort zone corresponds to the disparity range of ca. 70 arcmin around the screen [Zilly et al. 2011]. Since it is a rather conservative estimate, we allow display disparities within a wider range of ca. ± 2.5 deg. This approximately corresponds to the comfort zone in desktop viewing conditions given by Shibata et al. [2011, Fig. 23].

Vergence Modeling Schor [1979] and Hung et al. [1986] proposed sophisticated models of the eye vergence dynamics, which employ the concepts of control engineering (a negative feedback loop) to simulate the transient and sustained mechanisms. Also extended models, handling accommodation-vergence cross-linking have been proposed [Hung and Semmlow 1980, Schor 1992], and a validation against measurement data has been performed. However, disparity steps

interesting for us have not been treated extensively enough for our purposes, and the stereoscopic display conditions were not considered. Although some work focused on such conditions, the main goal was to investigate developmental plasticity in children exposed to stereoscopic games [Rushton and Riddell 1999] or to study the change in the post-task measures of AC/A and CA/C [Eadie et al. 2000]. In this dissertation, we propose a simple data-driven model of the vergence response that is tuned to step-like disparity changes under stereoscopic display conditions. We consider vergence dynamics as a function of the initial and target display disparities, and our goal is minimization of the vergence adaptation time at scene cuts through disparity editing.

Temporal Changes vs. Comfort Yano et al. [2004] reported that visual discomfort was induced if images were moved in depth according to a step pulse function, even if the images were displayed within the depth of focus. In a related work by Tam et al. [2012], influence of disparity and velocity on visual comfort was investigated, and a significant interaction between velocity and disparity was shown. The negative effect of object velocity on visual comfort was apparent even when the objects were displayed within the generally accepted visual comfort zone of less than 1 deg of horizontal disparity. Results obtained by Lambooij et al. [2011] show that rapidly moving objects and changing screen disparity indeed have a significant effect on visual comfort; however, their dominant role was not confirmed. Li et al. [2014] compared different types of motion and found that in-depth motion generally induces more visual discomfort than planar motion.

Several metrics of visual comfort for stereoscopic videos taking motion into account have been proposed [Cho and Kang 2012, Jung et al. 2012, Du et al. 2013], and their common feature is penalization of fast in-depth motion. Although these metrics could be used to inform stereoscopic content production, e. g., optimization of the camera parameters, they consider only continuous motion and it is unclear how they could be applied to discrete disparity steps at edit points.

2.2 Binocular Perception of Gloss

Binocular rivalry seems to be a key component in the distinct appearance of lustrous surfaces, for which the resulting luminance does not agree in both eye's images, even when accounting for binocular disparity and registering the two images [Dove 1850, Brewster 1861, Paille et al. 2001]. As described by Kirschmann [1895], a relation between luster and the disparity (parallax) of highlights exists. Blake [1985] derived equations that allow a machine – but maybe also the human visual system – to infer the shape of an object from disparities of specular highlights. Later, the perception of highlight disparity was analyzed in a matching experiment in which the participants were asked to adjust a rendered highlight's disparity on a convex or concave surface to obtain maximal realism [Blake and

Bülthoff 1990]. They found that most of the time highlights were correctly placed behind the convex surface, but with a bias towards the surface. We hypothesize that this bias originated from the fact, that looking at highlights (also real ones) with significant disparity causes visual discomfort. For the concave surface, the participants claimed that the most realistic gloss impression was obtained when highlights appeared on the surface or behind it. One could draw the conclusion that whereas on-surface highlights are not realistic and reduce perceived glossiness [Wendt et al. 2008], rendering them physically is also not the best option. Hurlbert et al. [1991] observed that for convex surfaces glossiness perception was not affected by the magnitude of the highlight disparity, but highlights had to be placed behind the surface. For concave surfaces perceived glossiness increased with the highlight disparity irrespectively of its sign, which means that highlight placement behind the surface is acceptable, although usually it is not physically correct [Blake and Brelstaff 1988]. Similar convex-concave asymmetry in the binocular perception of glossy surfaces was recently reported by Kerrigan and Adams [2013]. Our highlight microdisparity technique relies on these findings, as it always places highlights behind the surface, while avoiding excessively large disparities so that best viewing comfort and realistic appearance is achieved.

Obein et al. [2004] investigated the relation between gloss sensation and specular gloss value in the context of monocular and binocular vision. They reported that binocular factors play the most important role in the judgment of high-gloss values, while for medium-gloss surfaces the gloss sensitivity is similar as for monocular vision. Clearly, for higher gloss values the distinctness of the reflected image with high spatial frequency content enables better localization of relative highlight positions for each eye, which facilitates stereo matching [Hess et al. 1999]. We conform to these observations and focus on highly glossy surfaces, where low disparities lead to realistic and comfortable gloss depiction. See the review by Chadwick and Kentridge [2015] for more information on gloss perception, including a discussion of the importance of binocular vision for glossy appearance of surfaces, and the very recent work on key stereoscopic characteristics of the specular reflection by Muryy et al. [2014].

2.3 Random Dot Layers and Volumes

The perception of film grain as a stereoscopic structure shows a number of analogies to depth perception in random-dot stereograms (RDSs) [Julesz 1964], where binocular correspondence between dots is found without any explicit prior reference to a specific object. In both cases, such correspondence can be found only through local pooling over the dot patterns, as each dot, when considered independently, could be matched to a large number of its counterparts in the other eye. Lankheet and Lennie [1996] investigated various factors that can affect the human visual system's sensitivity to binocular correlation detection, which is required for depth recovery in the stereoscopic dot structure. They considered

the dot life time as short as 26 ms and did not observe any improvement in the correlation sensitivity when the dots have been displayed for longer times. This suggests that binocular correlation processing well integrates location-varying information in successive frames for dynamic RDSs. Moreover, such time-varying fresh patterns of dots, which represent consistently the same disparity relationships, reduce a chance for a false disparity match in the neuronal receptive field, as it is unlikely that at the next frame the new dot pattern will support again the same false match [Cumming and DeAngelis 2001, p. 217]. Also, the overall dot density does not seem to affect in any significant way the correlation performance, at least when the dot density is beyond 40 dots/deg^2 [Lankheet and Lennie 1996, Fig. 5]. All these observations apply to our film grain approach, where a new dot pattern is generated for each frame with the dot life time of at least 20 ms (assuming the framerate 50 fps or less), and a typical dot density falling into the range 75–550 dots/deg^2 (estimated by counting the local extrema of the grain pattern).

The problem of stereo-transparency perceived in surfaces defined solely by disparity in RDSs has been investigated [Akerstrom and Todd 1988, Tsirlin et al. 2008, 2010], where one of the key issues is the visibility of distinct transparent layers. Tsirlin et al. [2008, Fig. 9] found that even three layers cannot be visually separated for the dot density higher than 8–10 dots/deg^2 per layer. Moreover, the visual separability of the layers is significantly deteriorated when the number of layers increases or dot patterns overlap between layers [Tsirlin et al. 2010], and when the inter-layer disparity drops below 1.9 arcmin [Tsirlin et al. 2008]. Since the density of grain dots is relatively high, the layered grain representation composed of several layers becomes a simple alternative to a full volumetric structure. We pursue this design option in Sec. 5.1, as the layering approach enables simple real-time GPU implementation, which is important in the context of computer games.

Relatively little is known about the perception of stereoscopic volumes of dots. Recently, Goutcher et al. [2012] investigated the sensitivity of the human visual system to changes in the range and distribution of disparity-defined volumes of dots, and observed that for many ranges dots drawn from the Gaussian distribution could not be distinguished from an entirely uniform distribution. They concluded that the visual system uses an impoverished representation of the structure of stereoscopic volumes. This means that using more sophisticated distributions is not likely to have much visual impact. Therefore, in this work we always assume the uniform dot density allocation, and all our efforts to improve the appearance of stereoscopic grain are focused on modulating the thickness of its volumetric structure.

2.4 Stereoscopic Depth Processing

Stereoscopic scene analysis is an active research area, and a number of systems have been proposed that provide stereoscopic content creators with useful feedback and let them manipulate the reproduction of the scene depth. Here we give four examples of such tools, and we refer the reader to work by Smolic et al. [2011] for an overview of the state of the art in stereoscopic video post-production and processing.

Masaoka et al. [2006] proposed a system that based on the capture and viewing conditions estimates the distortion of the stereoscopic depth, and helps to detect "puppet-theater effect" (unnatural miniaturization), "cardboard effect" (unnatural flatness), or excessive disparities in the scene.

Wang and Sawchuk [2008] developed a general framework for disparity manipulation, that works in three stages: first, disparity maps are generated from the input sequence, next, the user is provided with several tools that let him or her manipulate the disparities, and last, new stereoscopic images are synthesized, based on the modified disparity maps and the initial sequence.

Lang et al. [2010] identified temporal changes of disparity as an important factor in stereoscopic film making: as they report, stereoscopic film makers often employ a continuous modification of the depth at scene transitions to ensure that the salient elements are at similar depths. In their work, they proposed non-linear disparity mapping operators, that can be used as a post-process for adjusting the depth distribution within the scene, and they show how one can gradually interpolate between different remapping operators to compensate for the sharp disparity jumps at cuts. Nevertheless, as noted by Lang et al., depth discontinuities can be also exploited as a storytelling element or a visual effect, and are used to evoke emotional responses [see also Mendiburu 2009, p. 154]. Our model could be used to inform all such transition-related modifications, by providing the actual times necessary to adapt to a given disparity change.

Koppal et al. [2011] describe a tool, that given rough takes or still images of the scene, provides a visualization of the stereoscopic depth perception in the target viewing conditions to inform the final capture. The tool provides also a "box widget" functionality, that enables post-capture adjustment of the camera parameters (field of view, camera position, etc.) using view interpolation. Additionally, parameter coupling and parameter cross-fading at cut points are possible. This enables, for instance, cross-fading the horizontal image translation, so that the salient objects are at zero disparity at the moment of transition, which is a simpler alternative to the approach considered by Lang et al.

Heinzle et al. [2011] proposed a computational stereoscopic camera system with programmable control loop, in which the convergence and the interaxial distance of the cameras, along with other camera parameters such as focus, zoom, or exposure time, can be adjusted automatically. In contrast to their work, which focuses on optimizing the stereoscopic parameters *within* one shot, we are mainly concerned with transitions *between* two neighboring shots. Although

their system could be extended to handle multiple cameras communicating with each other, this would help only to a limited extent because it is not fully known at the time of shooting how the final edit (order of shots, times of transitions, etc.) will look like. In our work we mainly target the editing and post-production stages of the film-making process.

Oskam et al. [2011] introduced a system that controls the stereo camera in interactive virtual environments, e. g., games. The system smoothly corrects the camera parameters in such a way that temporal non-linearities of the depth transformation are minimized. It handles abrupt changes in the depth distribution of the scene by adjusting the convergence and interaxial distance of the cameras to keep the scene or a particular salient object within a prescribed depth range. Alternatively, the system can keep a series of points in the scene as close as possible to defined depths. Unlike our work, Oskam et al. focus on interactive applications, and do not consider vergence angle adaptation times.

Bernhard et al. [2014] showed how binocular fusion times can be reduced by means of active manipulation of the convergence plane. The object of interest is brought back to the zero-disparity plane once the change in gaze has been detected, but before the vergence adaptation is complete. In contrast to Bernhard et al.'s active approach, we propose a cut optimization process that keeps the disparities constant during the vergence adaptation. The improvement in our case comes from a more informed choice of the initial and target disparities. Nevertheless, both approaches could be potentially combined.

2.5 Rendering of Gloss

With the current stereo equipment, disparity is more limited than in real world scenes, and thus it requires special processing [Lang et al. 2010, Didyk et al. 2011]. However, such manipulations deal with diffuse surface disparity and ignore the disparity of reflections. Stereoscopic highlight processing has been used in film production [Robertson 2009], however we are not aware of any technical publication analyzing the problem or describing an automatic solution. Attempts have been made to achieve a perceptual normalization of gloss in monocular images [Pellacini et al. 2000, Wills et al. 2009]. In a recent work, da Graça et al. [2014] studied the effect of stereoscopy on the perception of computer-generated metal-flake paints.

2.6 Noise, Grain, and Points

Adding noise can help hide banding artifacts [Daly and Feng 2003] or enhance the perceived sharpness [Johnson and Fairchild 2000, Kurihara et al. 2009] of the image. The human visual system tends to naturally mask repetitive signals through adaptation processes that lead to increasing contrast detection threshold

for such signals. This way the effective noise visibility is reduced while the salience of novel image content is enhanced [Fairchild and Johnson 2005].

Procedural noise is an important tool to add visually rich appearance to synthetic images [Lagae et al. 2010]. Stephenson and Saunders [2007] describe the synthesis of film grain based upon its noise-power spectrum. De Stefano et al. [2006] proposed a method based on a causal auto-regressive model to generate plausible-looking grain patterns given input samples of existing grain. Gomila et al. [2013] described a low-complexity system in which the grain in the input video is modeled and its metadata is transmitted together with the de-grained signal. Based on the metadata, the grain is re-synthesized and added back to the video at the receiver end. In our work we focus on *adding* grain, and thus, we assume that the grain pattern is already given.

Adding film grain to images can be seen as an NPAR-style operation. There are a number of papers dealing with the problem of stylizing stereoscopic imagery [Northam et al. 2012, Stavrakis and Gelautz 2005, 2004, Kim et al. 2013]; however, they focus on minimizing conflicts between the left and the right eye, and no effort is made to separate the stylization and the objects in depth. In the context of grain application, these algorithms are therefore analogous to on-surface grain. In stereoscopic line drawing, Lee et al. [2013] found that brush stroke texture stylization enhances the depth impression with respect to plain lines.

Stereoscopically displayed volumetric point clouds are common data representation in immersive virtual reality systems developed for medical imaging, scientific visualization, and volumetric rendering applications. Wang et al. [2010] observe that by increasing the point density or size, the ability to explore 3D environments might be deteriorated due to occlusions. Our goals are quite different as stereoscopic grain is not intended as a means to convey any specific information, but rather to accentuate the rich stereoscopic appearance of the scene.

Chapter 3

Eye Vergence Model

In this chapter, we are concerned with rapid changes of the depth in stereoscopic content and the related adaptation of the vergence angle of the observer's eyes. Humans have a good understanding of the environment they observe and move through, a so-called "mental image", which enhances their capabilities in focusing on different objects [Finke 1989]. When watching a film, however, the observed scene is merely a sequence of disconnected shots shown on a flat screen, and it is easy to get confused or lose track of the point of interest. Each shot usually uses a different vantage point than the preceding shot or shows a completely different environment. The new viewpoint is not known in advance and the time of transition from one shot to another is also unexpected. In the case of stereoscopic films, the task of following the action is even more challenging because of the added dimension. A large and unpredictable change in disparity results in a loss of binocular fusion, and a confusing double image is seen (diplopia) until the observer has adapted his or her gaze. Such adaptation, however, in addition to saccadic eye movements, requires also much slower vergence movements. Furthermore, the vergence system is interconnected with the accommodation system, but their goals are in conflict, since the verging distance is different from the focusing distance (see Sec. 2.1). This conflict has been identified as one of the sources of observer discomfort and fatigue in stereoscopic viewing, and it has been proven to hinder the performance in a stimulus identification task [Hoffman et al. 2008, Lambooij et al. 2009].

The Hollywood style of combining shots has developed into a set of formal conventions that obey the dynamics of visual attention and control the continuity of space, time, and action. In early films, shots were combined using cuts, dissolves, and fades to mark the structure of the film as defined by scenes and acts. Nowadays, however, almost 99 percent of all edits are cuts. An extensive analysis by Cutting et al. [2011] shows that average shot duration has declined from ca. 10 s in the 1930s to ca. 3.5 s after 2000. In the extreme case of the 1985 film *Rocky IV*, the average shot length is as short as 2.2 s. Cutting et al. suggest that decreased shot length might help control the attention of the viewers and

increase their engagement. However, such an accumulation of cuts combined with rich stereoscopic content challenges the visual system by forcing frequent adjustments of the vergence angle over a possibly wide range of depths. This requires a different approach to editing, e. g., some ultra-short "MTV-style" shots need to be replaced by more slow-paced edits. However, films are often released simultaneously in the regular and stereoscopic formats, and one should not expect that directors, cinematographers, and editors will entirely give up on their artistic visions and style because of the limitations of the medium [see Neuman 2009]. To this end, various stereoscopic post-production techniques have been developed to make cuts natural and effortless for viewers (see Sec. 2.4). Such manipulations range from simple depth adjustments to sophisticated transitions, where multiple sequences with gradually changing depth are combined [Owens 2013]. Since performed manually, all these manipulations are time-consuming and expensive. Owens [2013] reported that the editing of transitions was one of the most challenging tasks in the post-production of the concert film *U2 3D*.

To address the problem of rapid depth changes, we propose to relate the cut quality to vergence-angle adaptation time. We present a series of experiments with human observers, in which vergence responses were measured using consumer stereoscopic equipment and a high-frame-rate eye tracker. The measurements allowed to derive a model, that given the initial and target disparities, describes the vergence-angle adaptation curve. This model enables prediction of the adaptation time after cuts, and, in consequence, its visualization and minimization. We demonstrate the impact of the minimization on the visual quality of stereoscopic content in a separate experiment. In summary, we make the following contributions:

- measurements of vergence response to instantaneous disparity changes defined by initial and target disparities;
- derivation and evaluation of a model relating a disparity change to the vergence curve, along with average observer parameters; and
- design of an interactive tool for visualization and minimization of adaptation times.

3.1 Model Derivation

In this section, we experimentally derive and evaluate a model of eye-vergence response to step-like changes in disparity. We also estimate model parameters for an average observer. The collected data is useful in a number of applications, as discussed in Sec. 3.2.

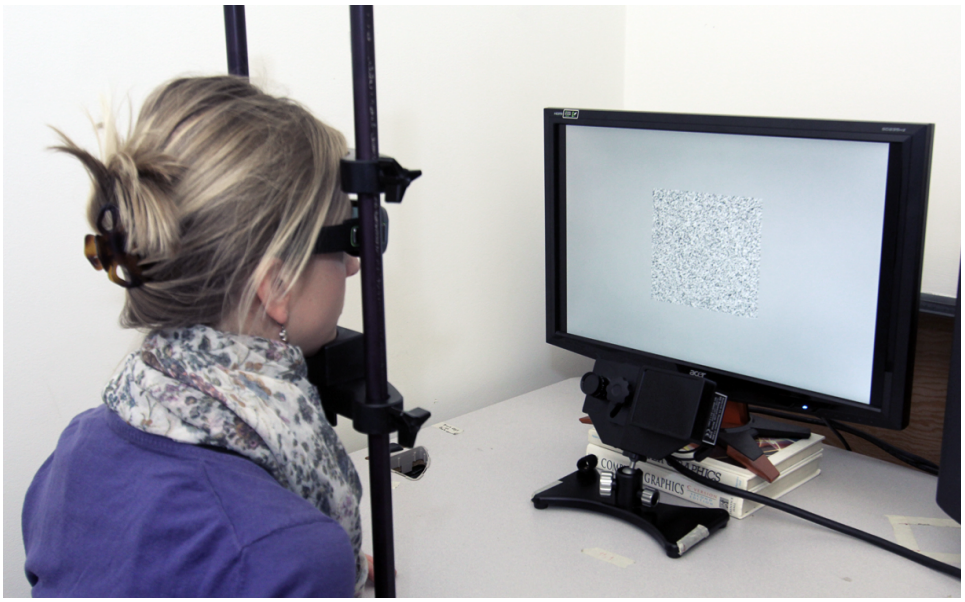


Figure 3.1. Our experimental setup. The stimuli were displayed using an Nvidia 3D Vision 2 kit (active shutter glasses) and an Acer GD235HZ 23.6-inch screen with the native resolution of 1920×1080. Both eyes were tracked with an EyeLink 1000 Plus eye tracker with a desktop mount. The temporal resolution of the tracker was 500 samples per eye per second. A chinrest was used to stabilize the subject's head at the viewing distance of 55 cm. Photo: P. Didyk

Participants Sixteen subjects (8 F, 8 M) took part in our experiment. They were members of the computer graphics or the computer vision group, between 21 and 35 years old. All had normal or corrected-to-normal vision, and all passed a test for stereo-blindness.

Equipment Stimuli were presented using an Nvidia 3D Vision 2 kit and an Acer GD235HZ 23.6-inch screen with the native resolution of 1920 × 1080. In order to measure the vergence responses, both eyes were tracked using an EyeLink 1000 Plus eye tracker with a desktop mount. The tracker records 1000 samples per second (500 per eye), allowing for fine-scale analysis of the vergence response. The spatial accuracy according to the eye-tracker manufacturer is up to 0.25–0.5°. A chinrest was used to stabilize the subject's head, and the viewing distance was fixed to 55 cm. Our experimental setup is shown in Figure 3.1.

Stimulus The stimulus in our experiment was a low-pass filtered white-noise patch changing its disparity in discrete steps over time. The patch was presented centrally on the screen, on a neutral grey background, and it subtended ca. 11 degrees of visual angle. A single trial consisted of a sequence of disparities $d_1, d_2, \ldots, d_n$, chosen from a fixed set D. The ordering of the disparities was randomized to avoid learning effect, but only Eulerian paths were used, i. e., $d_1 = d_n$,

and every possible transition appeared exactly once. Since prediction has been shown to have influence on vergence response (periodic disparity changes can be followed by vergence without typical latency [Hung 1998]), the time between the onsets of consecutive stimuli was set randomly between 1.25 s and 2.5 s.

Task Each session of the experiment started with a calibration procedure, as described in the eye-tracker manual. Next, every participant had to perform m trials, and the task was to simply observe the patch. The participants were encouraged to take breaks whenever they felt tired, and after each break the eye tracker was re-calibrated. The entire session took approximately 40 minutes.

Data Analysis After each session, binary output of the eye tracker was converted to a plain-text version using the converter tool provided by the manufacturer. Next, the data was processed using a custom parser to extract gaze coordinates and times of disparity changes, and read into MATLAB R2012a. The times of stimulus onsets were marked in the output files with timestamps – a functionality provided by the tracker's API, which enabled easy synchronization of the gaze data with stimuli. For each transition, we extracted the one-second segment following it, smoothed using a small box filter, and converted it to vergence angles. The angles were approximated by the difference between the x-coordinates of the two gaze positions expressed in pixels (for our experimental setup the approximation error is negligible). Missing or unreliable samples (due to, e. g., blinks, saccades, or tracking errors) were interpolated linearly, and the segments that required interpolation of more than 50% samples were excluded. Data for transitions of one type was grouped, and an asymmetric sigmoid curve was fitted to the average. Next, for each type of a transition, the time to reach 95% of the required vergence change was determined, and two surfaces were fitted to the obtained data points. Since we were interested in the relative gaze positions, the significance of drift was low. Moreover, adaptation times were determined by the 95%-of-change position, which is not very sensitive to shifts, scaling, etc. Based on these premises, we believe the precision was sufficient for our purposes.

3.1.1 Pilot Experiment

In order to gain insight into the relation of the vergence response to the initial and target disparities, as well as to estimate the number of trials m necessary for the response curves to converge, we conducted a pilot study. In it, one subject (S7) performed $m = 30$ trials, with $d_i = 0, \pm 30, \pm 60, \pm 90$ px, and the cut-off frequency of the low-pass filter $f = 20$ cpd. This resulted in $30 \cdot 7 \cdot 6 = 1260$ measured transitions. The results are presented in Figs. 3.2 and 3.3.

Discussion The signal converged quickly, giving relatively smooth data after ca. 5 repetitions, and little could be gained after ca. 10 repetitions. The vergence response can be modeled very well by sigmoid functions of the form $v = ae^{be^{ct}} + d$,

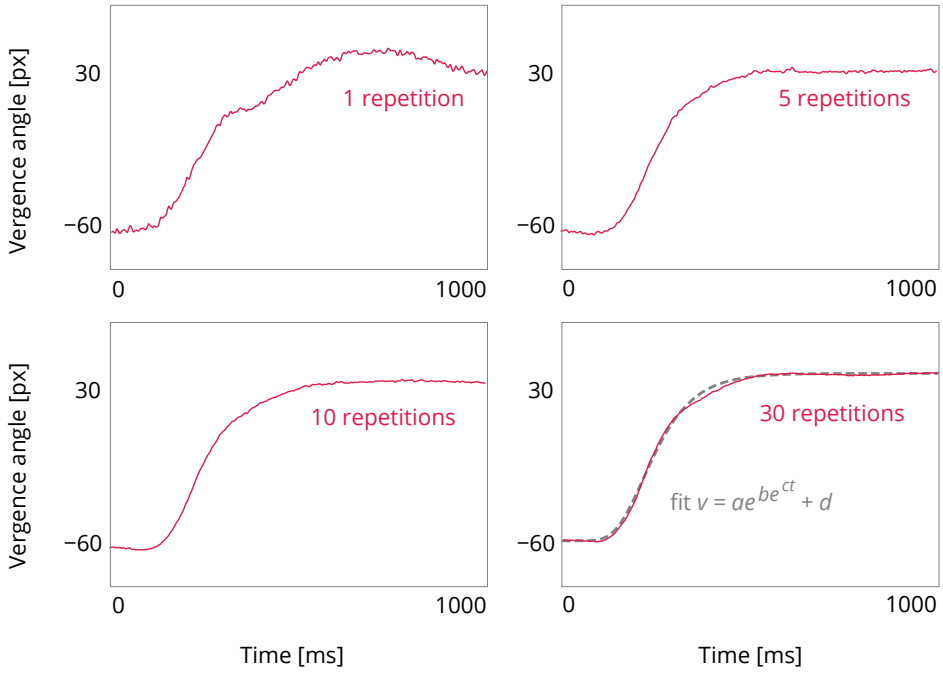


Figure 3.2. The results of the pilot experiment. Averaged responses of subject S7 to a −60 px → 30 px step, after 1, 5, 10, and 30 repetitions. The curve after 30 repetitions is shown together with a fit of a Gompertz function.

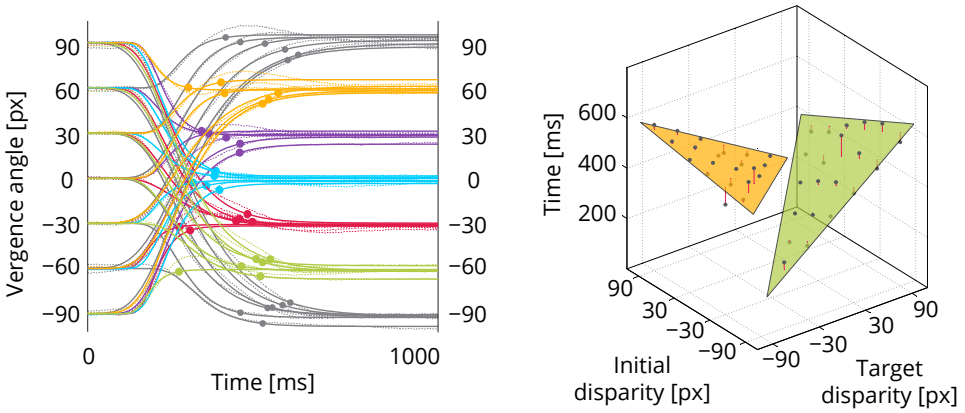


Figure 3.3. The results of the pilot experiment (continued). In the left panel average vergence responses to all 42 possible disparity steps for subject S7 are shown (dashed lines), together with fitted curves (solid lines), and points where the curves reach 95% of vergence change (solid circles). In the right panel, we plotted the transition time against the initial and target disparity. These points are almost perfectly modeled by two planes – the standard deviation of the error is approximately 27 ms. The two planes represent divergence (green) and convergence (yellow). We leave a gap between the planes, where times begin to increase due to Panum's fusional area and tolerance of the visual system to vergence errors. The diagonal is a singularity, where no transition is present, because the initial and target disparities are equal.

known as the Gompertz curves. The 95%-point does not depend on parameters a and d, and can be obtained using the following formula: $p_{95} = \ln(\ln(0.95)/b)/c$. The obtained data points can be modeled almost perfectly using two planes, with the mean error close to 0, and a standard deviation of ca. 27 ms. In light of these findings, we decided to limit the disparity values used in the main experiment to $d_i = \pm 30, \pm 90$ px, and the number of repetitions m to 10.

3.1.2 Main Experiment

The aim of the main experiment was twofold: to confirm that vergence times can be well modeled using two planes, as suggested by the pilot experiment, and, if so, to estimate parameters of the average-observer model, useful in practical applications. In this experiment $n = 16$ subjects performed $m = 10$ trials (except subjects S6, S9, and S10 for whom $m = 5$), with the cut-off frequency $f = 10$ cpd. The range of disparities for subject S9 was reduced to 2/3, due to reported problems with fusion. The results are presented in Figs. 3.4 and 3.5.

Discussion The average standard deviation of error after fitting the planes to the obtained data equals 36 ms. This indicates a very good fit, and justifies our assumption that the vergence adaptation time can be modeled using planes. In particular, this means that the data from subject S9, who saw rescaled disparities,

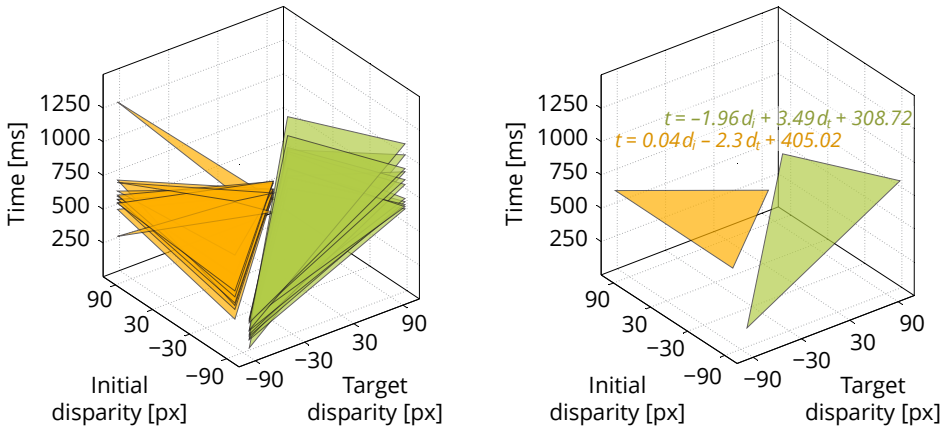


Figure 3.4. The results of the main experiment. In the left panel we present fits for all subjects, after exclusion of 4 outliers (subjects S1, S6, S8, and S14). These subjects were excluded due to serious difficulties with correct fusion of the stimuli. For completeness, we provide their data in the supplemental materials. The right panel shows the average of all fits from the left panel, along with equations of the planes. These planes describe transition times for the average observer.

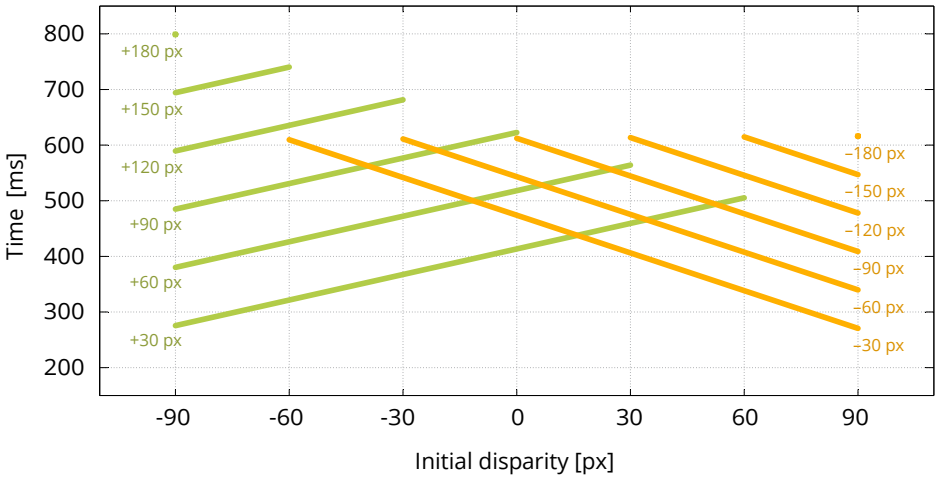


Figure 3.5. The results of the main experiment (continued). Here, we show a number of diagonal sections of the average planes from Fig. 3.4. Each line represents disparity steps of the same magnitude and direction, but with different initial disparities. See the text for a detailed discussion.

could be easily included in the average model.

As expected, our measurements show that given the initial disparity and direction, steps with larger magnitude lead to longer vergence adaptation times. An interesting finding is that the adaptation time depends also on the step direction and initial disparity. Given the initial disparity (Fig. 3.5, right, abscissae) and step magnitude (one yellow and one green line per magnitude), steps *towards* the screen are generally faster: To the right of the graph, yellow lines (convergent steps) have lower times than the corresponding green lines (divergent steps). To the left, this is reversed. Note, that corresponding yellow and green lines intersect near the point of zero initial disparity (screen plane). We hypothesize that it might be related to the accommodation-vergence coupling, which attracts vergence towards the screen plane, where the A/V conflict disappears.

Additionally, given the step magnitude and direction (Fig. 3.5, either one yellow or one green line), with decreasing initial disparity, convergent steps get slower whereas divergent steps get faster. This effect could be convincingly explained by the scale of the A/V conflict which increases with disparity magnitude. At *negative* initial disparities, divergent steps work towards resolving the conflict, whereas convergent steps work towards increasing it. With *positive* initial disparities, the roles are reversed. The larger the magnitude of the initial disparity, the more stress is put on the visual system, and the demand to resolve (or not to increase) the conflict is higher. Thus, the larger discrepancy between convergent and divergent steps. These effects should be taken into account while optimizing stereoscopic content, as simple minimization of *disparity difference* will not necessary lead to shorter adaptation times.

Another interesting finding is that with fixed *target disparity*, adaptation times for convergent steps are hardly dependent on the step magnitude. This phenomenon, at first unintuitive, could be explained by the A/V coupling as well: larger step magnitudes, which should intuitively contribute to longer adaptation times, may be offset by varying initial stress exerted by the A/V conflict on the visual system.

In our experiment, we considered only a computer display observed at a relatively short distance. On the one hand, at larger viewing distances the depth of field increases, thereby reducing the importance of the A/V coupling, the hypothesized cause of the observed variation in vergence adaptation time. On the other hand, discomfort induced by step-like motion in depth has been observed even for disparities within the DOF [Yano et al. 2004]. Answering the question, if similar effect of initial disparity on the adaptation time can be observed in other viewing conditions, e. g., in cinema, requires further investigation.

3.1.3 Evaluation

The obtained model was derived using simple stimuli (flat, white-noise patterns). On the one hand, this approach has several advantages: the exact disparity is known, regardless of fixation points; the measurements can be repeated easily;

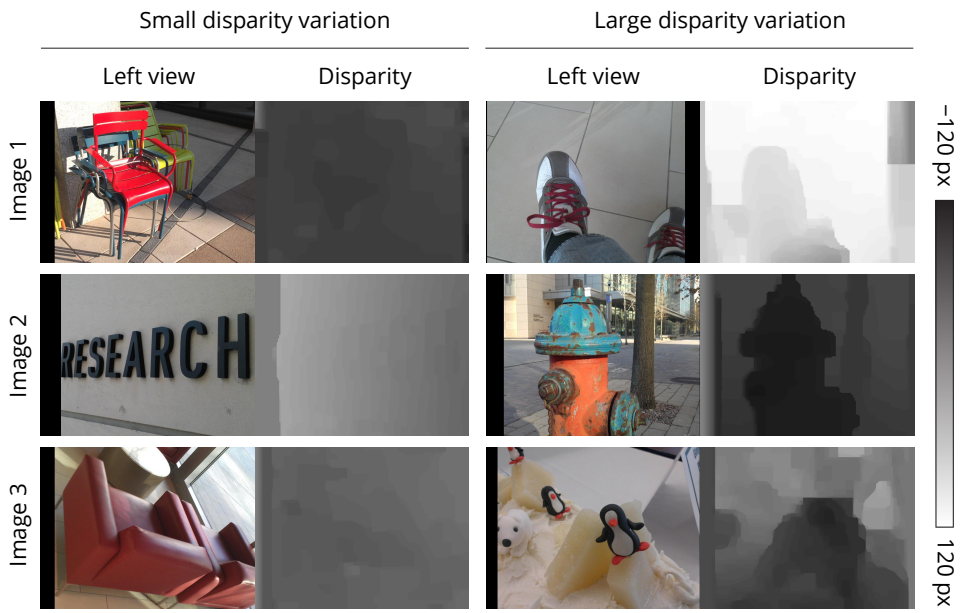


Figure 3.6. The two groups of stimuli used in the evaluation, one with larger, and one with smaller disparity variation across pictures. The black bars on the sides are floating stereoscopic windows added to avoid frame violation or large disparity steps at the edge (the shoe example).

and the learning effect is reduced, since the subject has no memory related to the spatial arrangement of objects in case of repeated images. On the other hand, it is unclear how well the model predicts the response to cuts between natural images: the presence of complex luminance patterns or high-level processes related to scene understanding may very well influence the transition times. Therefore, we conducted a validation experiment, to see to what extent the model can be generalized.

Participants and Stimuli Four participants from the original sixteen (S3, S7, S11, and S16) took part in the validation of the model. Six stereoscopic photographs taken with an LG Optimus 3D P725 smartphone were used (see Fig. 3.6). They were divided into two groups of three, one with smaller and the other with larger disparity changes across pictures. The disparities in the pictures were estimated using the SIFT flow algorithm [Liu et al. 2011]. In a single trial, a 6.5-minute random sequence composed of the three photographs from one of the groups was shown. As previously, a single appearance of a picture lasted between 1.25 s and 2.5 s (chosen randomly), and there were no breaks between appearances. The task was to simply observe the pictures, and the participants were asked to perform one trial for each group.

Data Analysis and Results After cleaning and segmenting the tracking data, a semiautomatic procedure was employed to group segments of the same type, enabling averaging of measurements. In the first, automatic step segments where a saccade occurred at the time of the cut, or within the first 100 ms after the cut, were discarded. Then, initial disparity was estimated using the disparity map and the fixation coordinates just before the cut (initial fixation). The target disparity was chosen using the following heuristic: whenever the duration of the first fixation was shorter than 300 ms, the second fixation was used; otherwise, the initial fixation was assumed to be also the target fixation.

In the second, manual step, all segments were briefly reviewed to correct filtering and target fixation errors. The false negatives were the cases when the saccade near the cut was small enough not to change significantly the vergence response. The false positives were the non-typical cases, including, but not limited to, eye-tracker errors, clearly incorrect vergence response indicating lack of fusion, segments with unusually large saccade-to-fixation ratio, erratic saccades indicating partial fixations, etc. In the end, 718 out of 3028 segments were discarded. We provide all annotated segments along with a custom viewer/editor as additional materials[1], and encourage the reader to inspect the data we used in this evaluation.

In the end, segments with the same initial/target disparities were grouped; groups with 5 or more members were averaged and compared against the model prediction for the respective subject. The results of the experiment are presented in Fig. 3.7.

Discussion Although our prediction slightly overestimated the time of transition for photographs, our model correlated well with the actual time, as indicated by the relatively low standard deviation of the error. The study proves that our model is a good predictor (up to an additive constant) of transition time for natural images. We hypothesize that improved performance was due to the presence of higher-order cues, absent in white-noise stimuli, where the sole depth cue was the binocular disparity. It is also possible that the adaptation was facilitated to some extent by the learning effect due to the small number of images.

3.2 Applications

In this section, we propose a set of tools for aiding in the production of stereoscopic content, that utilize our model to minimize vergence adaptation times. We also analyze the impact of the minimization on visual quality in one of the proposed tools using an object-recognition experiment.

[1] Available at http://resources.mpi-inf.mpg.de/VergenceModel/.

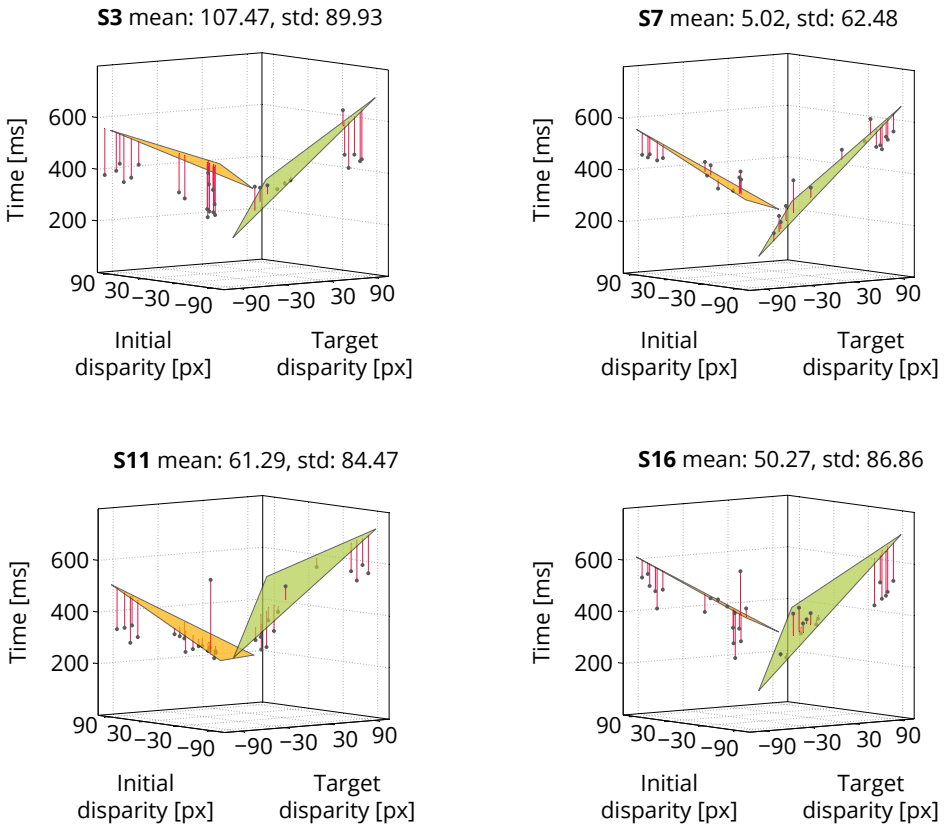


Figure 3.7. The results for subjects S3, S7, S11, and S16 (top to bottom, left to right). The planes show model predictions, whereas the solid circles represent the observed data. The mean and standard deviation of the error for subjects S16, S7, S3, and S11 are respectively 50±87 ms, 5±62 ms, 107±90 ms, and 61±84 ms.

3.2.1 Production Tools

Transition Time Visualization A straightforward application of the model is a visualization tool providing stereographers and VFX artists with an interactive analysis of transition times. In order to evaluate stereoscopic transition and estimate the transition time, we first need to determine the pairs of disparity values between which the transitions occur. A naïve approach would be to measure the transition time between corresponding pixels in both sequences; however, it is not very useful, as in most cases people change the fixation point immediately after the transition, and no change in vergence happens (see the data browser provided in supplemental materials). Therefore, the fixation points in both sequences need to be precisely determined.

Such data can be obtained from various sources, e. g., it is possible to use eye-tracker data. This does not require many subjects, as it has been shown that eye scan-paths form highly repetitive patterns between different spectators for the same video sequences [Wang et al. 2012]. Moreover, skilled directors are capable of precisely guiding and predicting viewers' attention. Such prediction is further facilitated by the tendency of increasing object motion in modern movies [Cutting et al. 2011] and by the fact that typical 2D-movie cuts trigger saccades towards the screen center [Mital et al. 2011, Wang et al. 2012]. Thus, the information about fixation points for our methods can be very reliably provided by the directors. Besides, Carmi and Itti [2006] observed that the saccades immediately after the cut are driven mostly by the bottom-up factors and can be predicted relatively well by existing saliency models. Once the fixation points before and after the cut are known, the corresponding disparity values need to be determined. This can be obtained directly from the rendering pipeline for animated movies, using user input in the case of 2D-to-3D conversion, or using disparity estimation techniques for natural scenes when the depth map is not available. Once the fixation points along with disparity values are known, transition times can be directly calculated from the model. Since computing model predictions is inexpensive, it can be used to provide real-time preview of transition times.

Camera Parameters Optimization Apart from predicting transition times and visualizing them for editing purposes, one can automate the process of stereoscopic content preparation. An optimization problem for cuts can be defined, and our model can serve as the core of the cost function.

As discussed in Sec. 2.4, stereoscopic content can be optimized by manipulating various parameters. These can be changed for the entire sequence (e. g., from cut to cut), or selectively around the cuts, with smooth blending back to original parameters [Lang et al. 2010, Koppal et al. 2011]. There is a wide range of manipulations that can be used to adjust stereoscopic content. They range from very simple ones, like changing camera separation and convergence (i. e., the plane of zero parallax), to more complicated ones, such as depth remapping. All such manipulations can be easily integrated and used with our model.

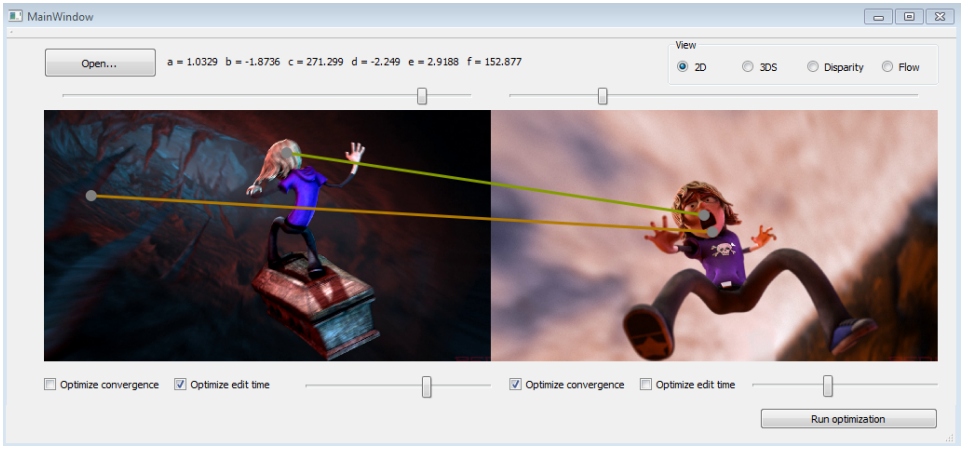


Figure 3.8. The design of our cut optimization tool. The colors of lines connecting the points of interest before and after the cut visualize the corresponding vergence adaptiation times (green – less, red – more). Pictures from *Dracula 4D* courtesy of Red Star 3D, www.redstar3d.com

Cut Positioning If the two sequences between which the cut occurs overlap in time, it is also possible to find the best moment for the cut. To this end, we optimize not only stereoscopic parameters, but also the position of the cut. This can be performed efficiently by simply iterating over all possible cut positions, in addition to all horizontal shifts of the left/right views. The optimal cut can be chosen automatically or can be shown to the editor as a suggestion. The design of a tool performing these tasks is shown in Fig. 3.8 and in the supplemental video.

3.2.2 Impact on Visual Quality

Visual quality can be defined in many ways, using various objective and subjective criteria. In the following experiment, we focus on the time necessary to recognize the 3D arrangement of objects after a cut. We assume shorter recognition times to be an indicator of higher quality. We measured the time needed to recognize object arrangement, and showed that this time closely matches our model. In practice, this means that when cuts are optimized using the proposed production tools, the time necessary to recognize objects in the scene is minimized.

Methods The equipment and viewing conditions were the same as in other experiments, but no eye tracker was used. As stimuli, we used two shots corresponding to a cut in the 3D version of the *Big Buck Bunny* animation. We modified them by placing two small dark-gray circles between the eyes of the character, with approximately the same disparity as the character (see Fig. 3.9, left, inset). Two 3D arrangements of circles for each shot were considered: one with the upper, and one with the lower circle closer to the observer. The disparity difference between the circles was 2 px. The convergence in the shots was modified

so that the average disparity of the circles was equal to d_i before and d_t after the cut. Seven pairs of disparity steps were used: $-75 \rightarrow -105/90$, $-60 \rightarrow -90/-30$, $-30 \rightarrow -90/60$, $0 \rightarrow -30/30$, $30 \rightarrow -60/90$, $60 \rightarrow 30/90$, and $75 \rightarrow -90/105$ px. For each initial disparity, both a convergent and a divergent step was possible, which prevented anticipatory eye movements in subjects. In order to determine the arrangement recognition time for all 14 steps, we performed 14 independent QUEST threshold estimation procedures [Watson and Pelli 1983], each estimating the time of 75% correctness. A single trial of each procedure had the following structure: First, the first shot was shown for 2 s. Next, the second shot was shown for a period between 0.1 and 1.5 s (controlled by QUEST). The arrangement of the circles was chosen randomly in every trial. After the screen was blanked, the subject was asked to indicate if the arrangement was the same in both shots: If the same circle (i. e., upper or lower) was closer to the observer both before and after the cut, the subject had to press the Y key, and the N key otherwise. Such a task definition ensured that the subject actually performed the vergence transition $d_i \rightarrow d_t$. All 14 procedures were performed in parallel, randomly interleaved. A session of the experiment lasted 20 min (average standard deviation in a QUEST instance 73 ms). Subjects S3, S11, S12, S15, and S16 took part in the experiment. S11 participated in three sessions, S16 in two, and the remaining three in one session.

Results The data obtained using the above procedure was fitted with two planes minimizing the RMSE. The planes obtained from all subjects were averaged (first within subjects, then between subjects), and compared to their average model. The results are presented in Fig. 3.9. A corrective constant shift of 83 ms minimizes the RMSE, and yields a low prediction error of 42 ms. This correlation implies that optimizing camera convergence using our model instead of disparity distance as the cost function will produce cuts with shorter recognition times. Similar improvement can be expected when optimizing other camera parameters or cut positions. This illustrates the practical importance of our model for S3D games and films.

3.3 Summary

We proposed a new model which predicts the time a human observer needs to adapt vergence to rapid disparity changes. We first presented measurements of transition times for simple stimuli, and demonstrated that these times are valid also for complex scenes. The experiment revealed interesting facts about the operation of the visual system during observation of a depth-changing stimulus, and provides stereoscopic content creators with valuable knowledge. Additionally, we proposed a set of tools for the editing of stereoscopic content to minimize the vergence-angle adaptation time after cuts. An important property of the proposed optimization techniques is that the manipulations are applied only

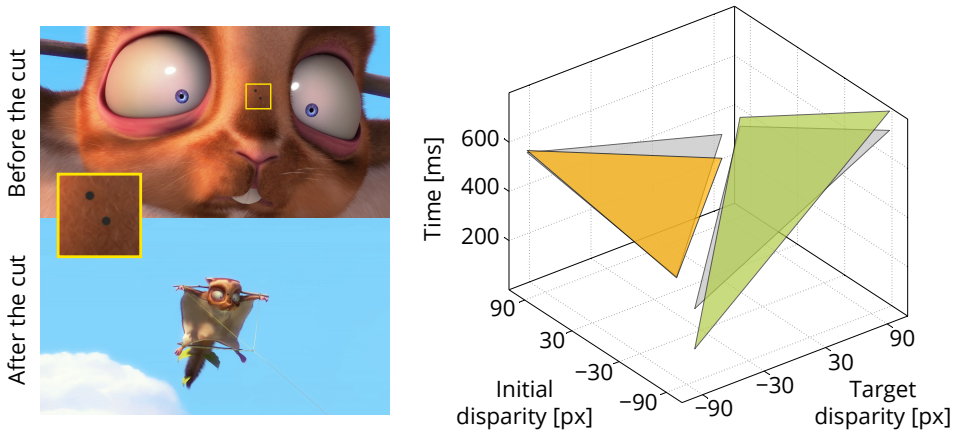


Figure 3.9. *Left:* Stimuli used in the object recognition experiment. *Right:* The results of the experiment; the gray planes represent the obtained data after corrective shift of 83 ms, and the yellow/green planes represent the average model of the subjects, predicting the data with the RMSE equal 42 ms. Pictures from *Big Buck Bunny* CC-BY Blender Foundation and Janus B. Kristensen

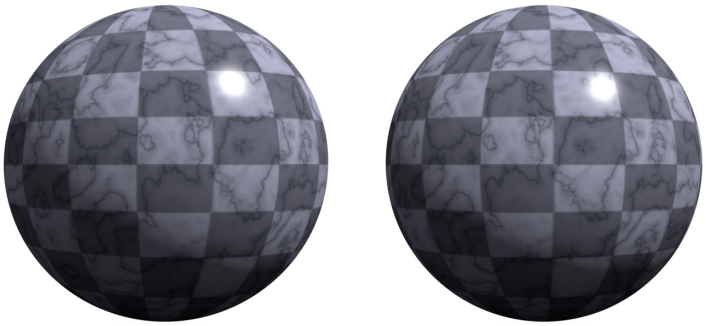
locally around cuts, which has limited effect on the depth impression created by the artist. To our knowledge, this is the first work that proposes to automatically edit stereoscopic cuts taking into account varying performance of the human visual system in adapting to rapid disparity changes. Finally, we demonstrated the impact of minimizing adaptation times on the visual quality of S3D content as measured by a subject's performance in the 3D object recognition task. An interesting avenue for future work would be an extensive user study quantifying how shorter transition times influence visual fatigue.

Chapter 4

Specular Highlights Disparity

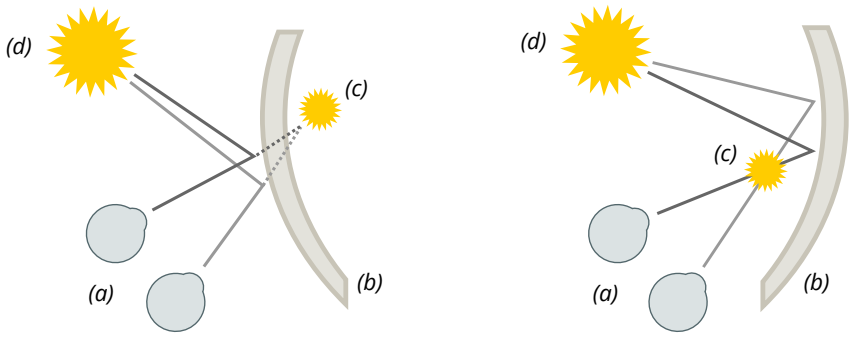
The possible range and variation of depth in a stereoscopic image are limited by viewing comfort considerations, and a trade-off between comfort and depth impression can be made by using disparity manipulation techniques, such as depth compression. These techniques assume, that the disparity is well defined by the scene's geometry. While this assumption is valid for solid, diffuse surfaces, it does not hold for materials with view-dependent shading. Specular reflections, which are (possibly blurry) images of the light sources in the scene, have their own depth, different from the surface on which they appear, and they are a potential source of excessive disparities. This phenomenon is illustrated in Fig. 4.1 and explained in Fig. 4.2. Additionally, depending on the geometry of the reflecting surface, highlights can change shape, disappear, or produce vertical disparities across the views of the stereoscopic image. One solution to this problem is to assume a common (cyclopean) eye position for both views when shading the surface. By doing so, we remove the highlight disparity and avoid shape discrepancies, however, such highlights seem to be "painted" on the surface. This is a significant shortcoming, as it is known, that highlight disparity is an important factor in the material perception (see Sec. 2.2 and Sec. 2.5). Nevertheless, on-surface highlights are quite common, presumably for three main reasons: because artists consider them to be more pleasant to watch; because of performance, e. g., in games that cannot afford to shade twice [Sousa et al. 2012, p. 163]; and because the necessary information is missing, e. g., in 2D-to-3D conversion. We address this problem by introducing a technique called *highlight microdisparity*, which avoids issues of geometrically-correct shading while preserving correct material perception (Fig. 4.3). Our contributions are as follows:

- a problem analysis of highlight stereo depiction,
- a simple and safe alternative approach to highlight rendering that improves over on-surface and physical highlights,
- a perceptual study validating our approach.



Left eye Right eye

Figure 4.1. Highlights have a different disparity than the objects they appear on. Note the shift of the highlight relative to the checker-board texture of the sphere.



Highlight seen behind the surface Highlight seen in front of the surface

Figure 4.2. An observer *(a)* looks at a specular surface *(b)* and sees a highlight *(c)* which is a (possibly blurred) image of a light source *(d)*. Depending on the geometry of the reflecting object, the highlight appears behind *(left)* or in front of the surface *(right)*. Note, that it is not placed *on* the surface.

Figure 4.3. Physically-based highlights often cause binocular conflicts *(left)*. On-surface highlights look less glossy and less authentic *(middle)*. When rendered using the proposed technique, they are detached from the surface but do not introduce conflicts *(right)*.

4.1 Highlight Microdisparity

Our technique assumes a shading model composed of a view-independent diffuse and a view-dependent specular components, such as Phong model. First, the diffuse component is rendered with usual disparity into a stereo pair. Next, the specular component is rendered into another stereo image pair assuming a common, cyclopean eye position for shading. Note, that the only effect of this change is modification of the equations defining the luminance of the reflections, and no image at the cyclopean position is actually rendered. Consequently, such highlights would appear *on* the surface of the rendered object, because they are calculated in a view-independent manner, thus, they have no disparity relative to the object surface. Finally, the specular stereo pair is warped horizontally to reintroduce disparity between diffuse and specular shading, and both stereo pairs – diffuse and specular – are combined. The warping is performed independently in each image of the specular stereo pair and it displaces the left half-image to the left and the right half-image to the right. Horizontal warping avoids unpleasant vertical disparities found in physical stereo highlights [Blake 1985] and maintains consistent shape between the half-images. Next, we describe in more detail how the amount of warping is controlled locally using a combination of four parameters: basic warping w_b, curvature weighting w_c, edge detector w_e, and artistic control w_a.

Basic warping The warping constant w_b is chosen to be large enough to make highlights visibly detached from the surface but small enough not to introduce objectionable artifacts and keep the highlight disparity low. Because the required amount of shift is small, the mismatch between geometry and highlights should not become apparent in monocular images. Our results were rendered using 2×2 super-sampling, and the warping was performed before scaling down the image. In every shown example, the shift of the highlights after down-sampling is not greater than 2 pixels for each view – 4 pixels in total – which in desktop display conditions corresponds to ca. 6 arcmin.

Curvature, edges and creases The highlight disparity depends on the surface curvature: for high curvature, the disparity decreases because the surface reflects larger portions of the environment and thus the reflections undergo compression [Blake 1985, Fleming et al. 2004]. Applying constant disparity without accounting for the curvature would lead to objectionable results in the regions where the highlights should be placed almost on the surface. The purpose of the curvature factor w_c is to suppress the warping for highly curved surfaces, and it is proportional to the magnitude of the second derivative of the surface depth in the horizontal direction, calculated in the image space. This approach is inspired by the method for enhanced surface depiction introduced by Vergne et al. [2009]. First, for every pixel **p**, we determine the first derivative in horizontal

direction using the normal vector $n(\mathbf{p})$: $g_x(\mathbf{p}) = -n_x(\mathbf{p})/n_z(\mathbf{p})$, where the z-axis points in the depth direction. Next, we approximate the second derivative by $h_x(\mathbf{p}) = \big(g_x(\mathbf{p}_+) - g_x(\mathbf{p}_-)\big)/2$, where $\mathbf{p}_+$ and $\mathbf{p}_-$ are the horizontal neighbors of $\mathbf{p}$. Finally, we set:

$$w_c(\mathbf{p}) = \begin{cases} 1 & \text{if } |h_x(\mathbf{p})| = 0, \\ 0 & \text{if } |h_x(\mathbf{p})| \geq c_{\max}, \\ \big(c_{\max} - |h_x(\mathbf{p})|\big)/c_{\max} & \text{otherwise.} \end{cases}$$

In our experiments $c_{\max} = 0.03$ was used.

Another factor limiting the warping procedure are edges and creases, because the highlights should not move over them. We detect edges by convolving the image depth map with a 3×3 Laplacian kernel, and thresholding the outcome. Thus, w_e equals 0 when an edge has been detected and 1 otherwise. The detection of creases is handled implicitly by the curvature weighting component, since the second derivative has a large magnitude in their vicinity.

Artistic control Spatially localized artistic control can be introduced by defining m sparse specular disparity constraints (h_1, ϵ_1), …, (h_m, ϵ_m) at surface locations $\mathbf{p}_1$, …, $\mathbf{p}_m$. Gaussian radial basis functions are used to propagate the constraints to arbitrary spatial locations $\mathbf{p}$:

$$s(\mathbf{p}) = \sum_{i=1}^{m} e^{-\epsilon_i r_i^2} h_i \quad \text{with} \quad r_i = |\mathbf{p} - \mathbf{p}_i|.$$

The s function is evaluated independently for every pixel. The parameters ϵ_i control the range of the constrains, whereas h_i – their strength and direction (an increase of the highlight disparity for positive h_i, and a decrease for negative h_i). We set $w_a(\mathbf{p}) = 2^{s(\mathbf{p})}$, to approximately linearize the strength of the effect. Fig. 4.11 illustrates how the four parameters influence the result. An example of manual changes to highlights is also given in Fig. 4.8. In all remaining pictures we assume $w_a = 1$.

Warping by gathering Having computed the four factors we combine them into a single warping coefficient $w \in \mathbb{R}^2 \to \mathbb{R}$, $w(\mathbf{p}) = w_b\, w_c(\mathbf{p})\, w_e(\mathbf{p})\, w_a(\mathbf{p})$, that defines the warping map as

$$\bar{w}(x, y) = \pm \min_{i \geq 0} \{w(x \pm i, y) + i\}$$

for left and right views respectively. The map $\bar{w}(\mathbf{p})$ can be computed from w by checking a few pixels in the neighborhood of $\mathbf{p}$. Finally, the warped specular image is defined as:

$$I_S^W(x, y) = I_S(x + \bar{w}(x, y), y).$$

4.2 Results

The proposed approach was implemented using a GPU, achieving interactive rendering rates for the full pipeline on a consumer PC. In Sec. 4.2.1 we show the results for three use cases (full rendering, performance-critical rendering, and 2D-to-3D) and we present the outcome of a perceptual study in Sec. 4.2.2. Please note, that all stereo images in this chapter serve only as a preview of the effect. Refer to additional materials for high-quality stereo pairs[1].

4.2.1 Use cases

Full rendering Here, the full scene information is available and the resources are sufficient to compute physically-based highlights, however, our method is used to minimize distractive effects of geometrically correct reflections, such as excessive horizontal disparities, vertical disparities, and binocular rivalry. The results of our approach in full rendering are shown in Figs. 4.3 and 4.4–4.7. The ability to locally and interactively control our approach is demonstrated in Fig. 4.8 and the accompanying video.

Performance-critical rendering Here, the full scene information is available as well, but the computational resources are limited, and we cannot render two images fast enough. A good example of such a situation are computer games, where the stereo image is produced using image-based warping techniques [Sousa et al. 2012]. Typically, the highlights in such a case are warped together with the geometry, and appear on the surface. Our technique can be used to warp the highlights independently, and an example usage in a game-like environment is shown in Fig. 4.9.

2D-to-3D In this use case, our technique is an additional step in a 2D-to-3D pipeline, which increases realism of the obtained results. First, the depth information in the picture is recovered, and the highlights are separated. Next, the diffuse and specular layer of the image are warped according to the depth map to produce a stereo pair. Then, small disparity is added to the specular layer, and both layers are combined. The result of this approach is presented in Fig. 4.10. The diffuse and specular layers were taken from [Tan and Ikeuchi 2005], the depth map was painted manually, and constant normal field was assumed.

4.2.2 Perceptual study

To verify our findings we conducted a perceptual study, where Figs. 4.3, 4.4–4.7 were presented to 10 naïve subjects (7 F, 3 M) using a Zalman M240W polarized display. Three images of each scene with physical, on-surface, and our highlights were shown on a neutral grey background next to each other. The placement of

[1] Available at http://resources.mpi-inf.mpg.de/HighlightMicrodisparity/.

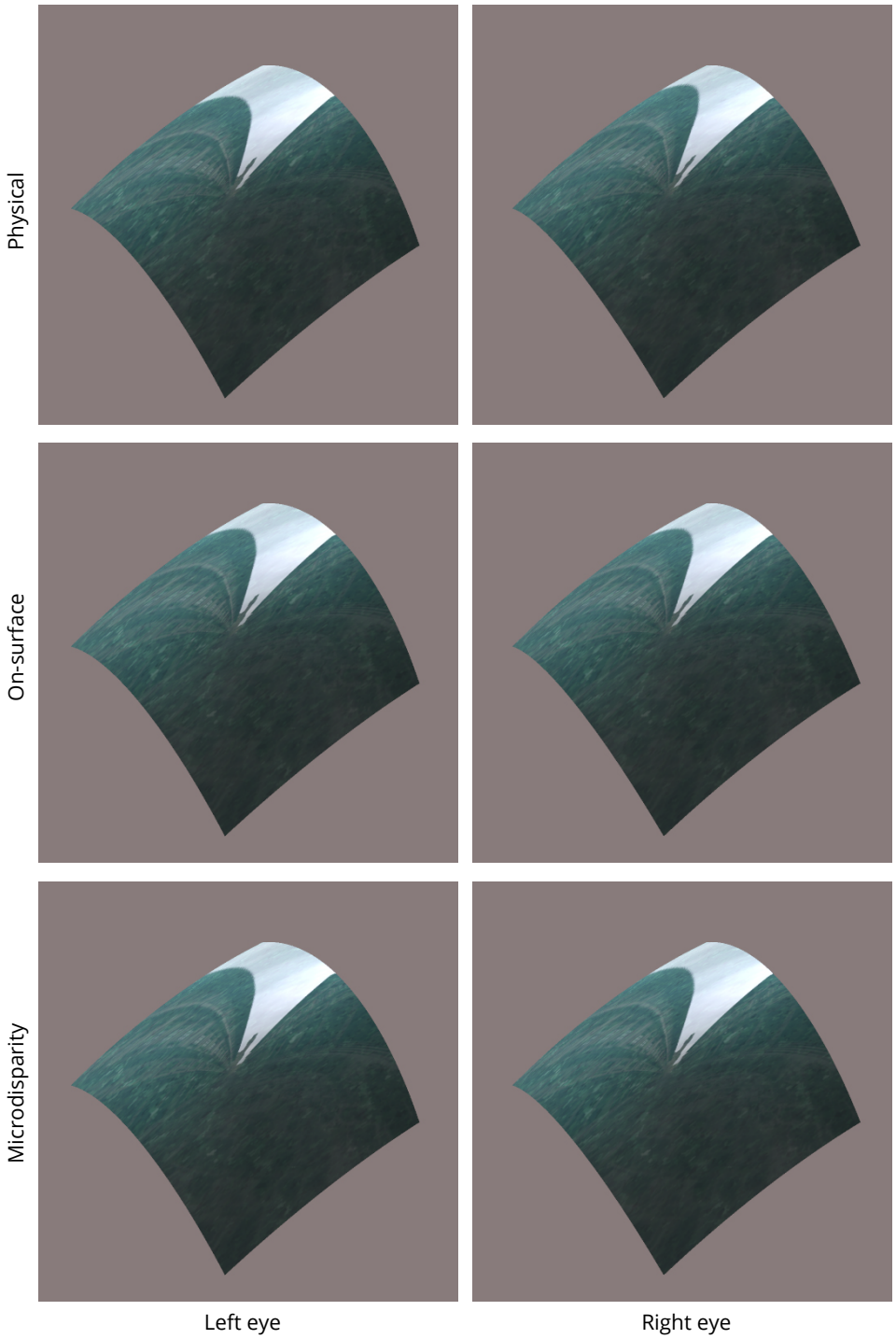


Figure 4.4. Results of using our method. From top to bottom: physical highlights, on-surface highlights, highlights with microdisparity.

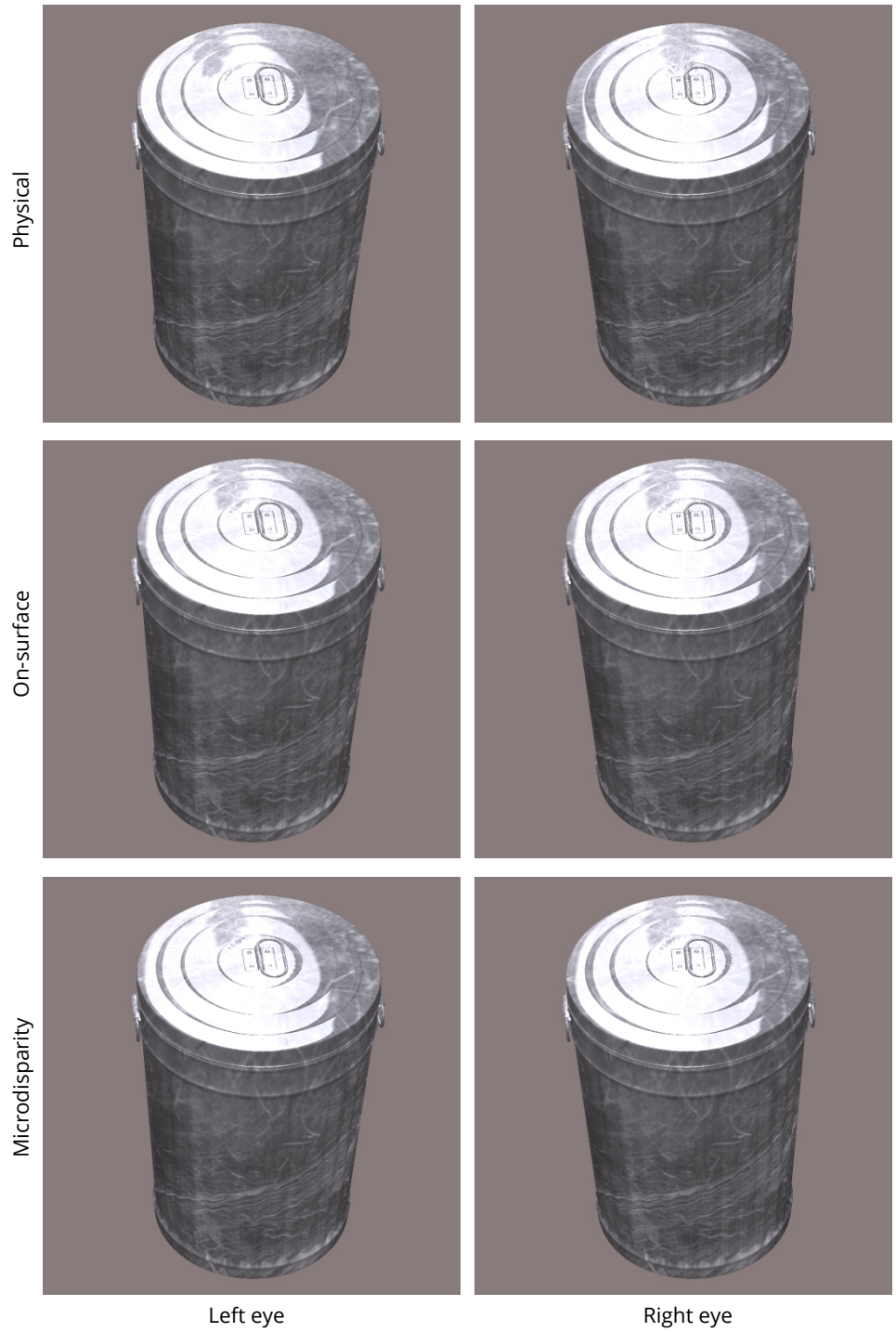


Figure 4.5. Results of using our method – continued. From top to bottom: physical highlights, on-surface highlights, highlights with microdisparity. Mesh: ClayOgre (www.blendswap.com)

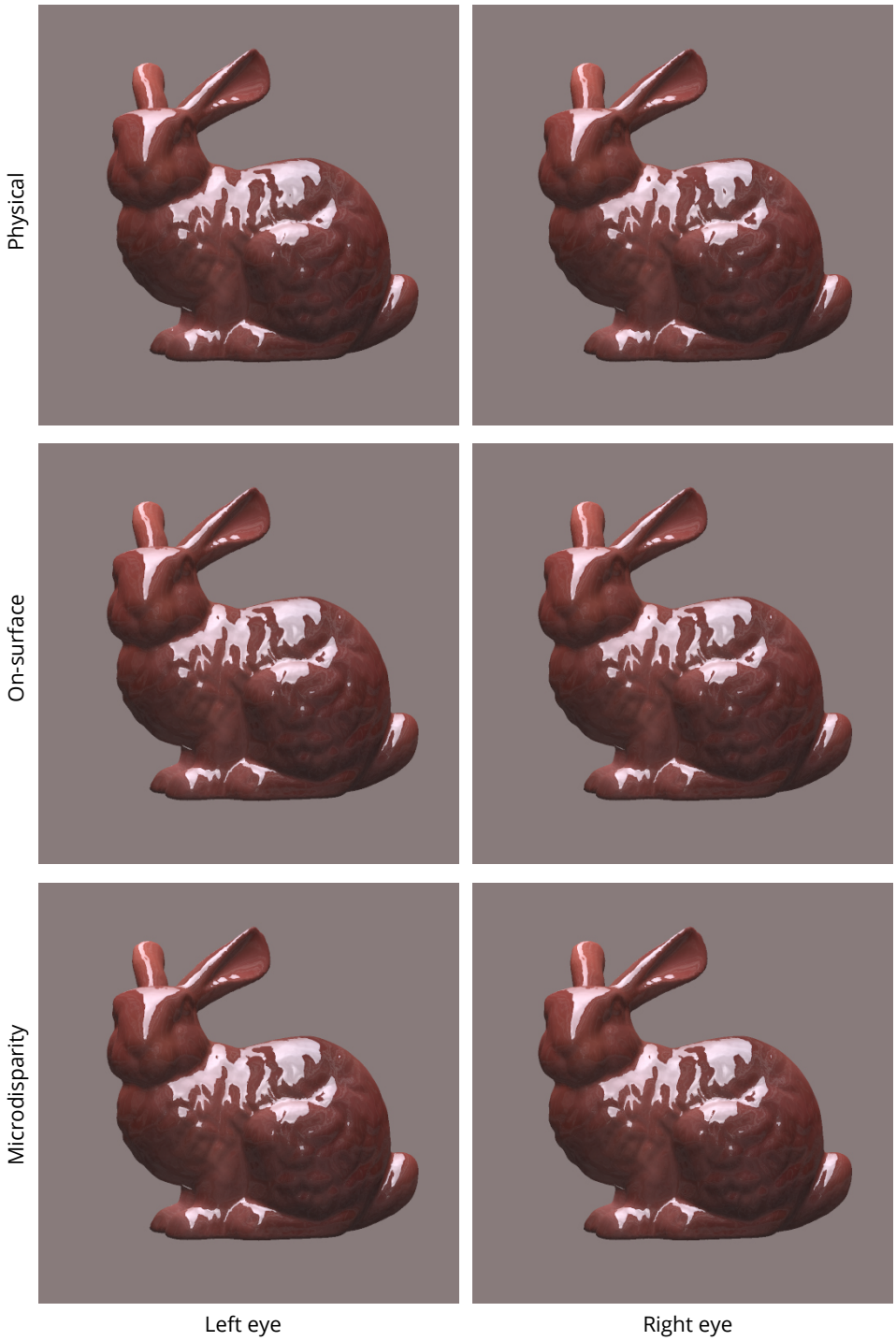


Figure 4.6. Results of using our method – continued. From top to bottom: physical highlights, on-surface highlights, highlights with microdisparity. Mesh: Stanford Repository

Figure 4.7. Results of using our method – continued. From top to bottom: physical highlights, on-surface highlights, highlights with microdisparity. Mesh: Georia Tech Models Archive

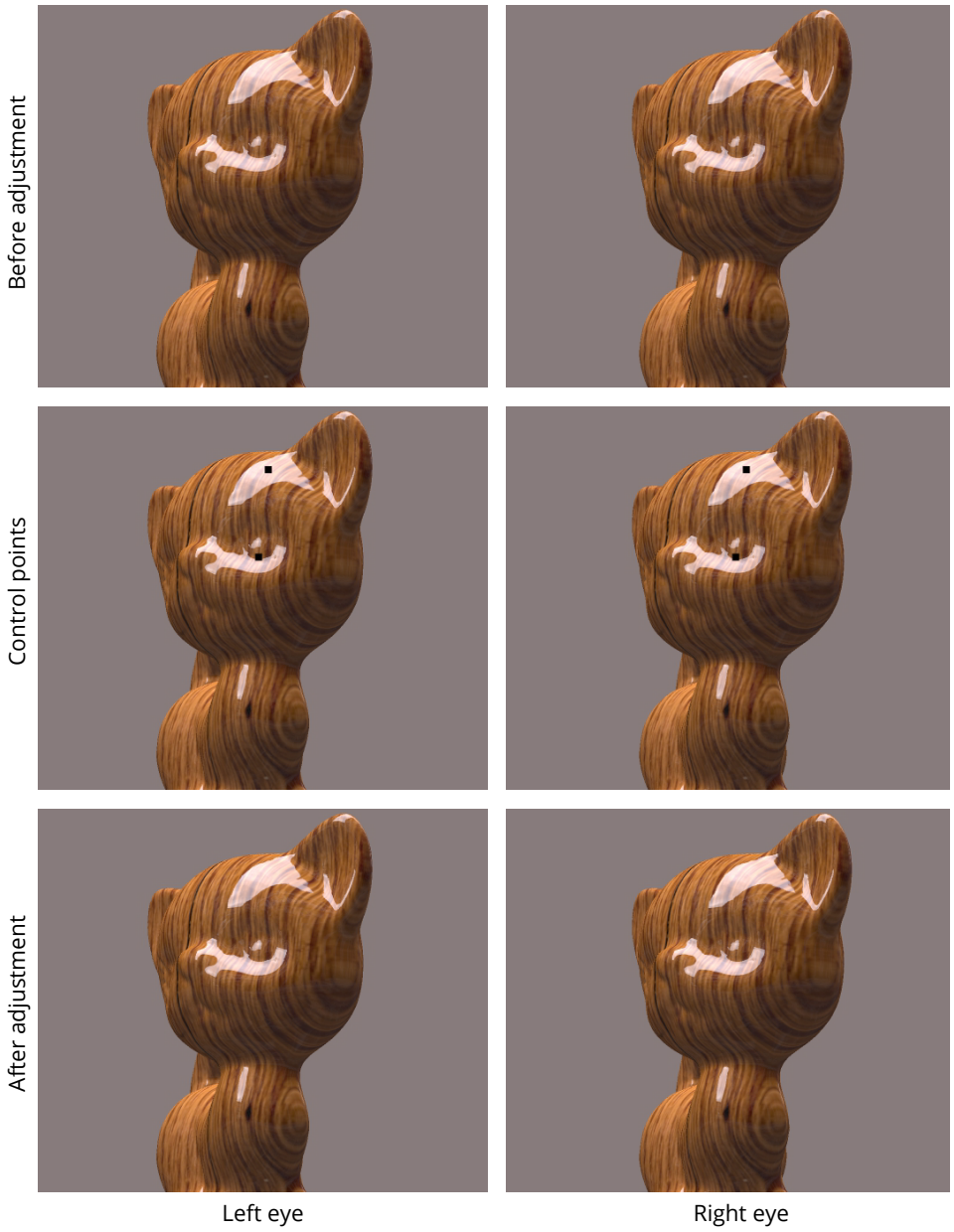


Figure 4.8. Highlight disparity *(top)* is locally adjusted near the ear and the eye *(middle)* to obtain an improved result *(bottom)*. Mesh: AIM@SHAPE

Figure 4.9. Performance-critical applications like games produce stereo images using image-based warping *(top)*. Our approach can improve highlight depiction by warping them differently *(bottom)*. Scene: T. Ritschel

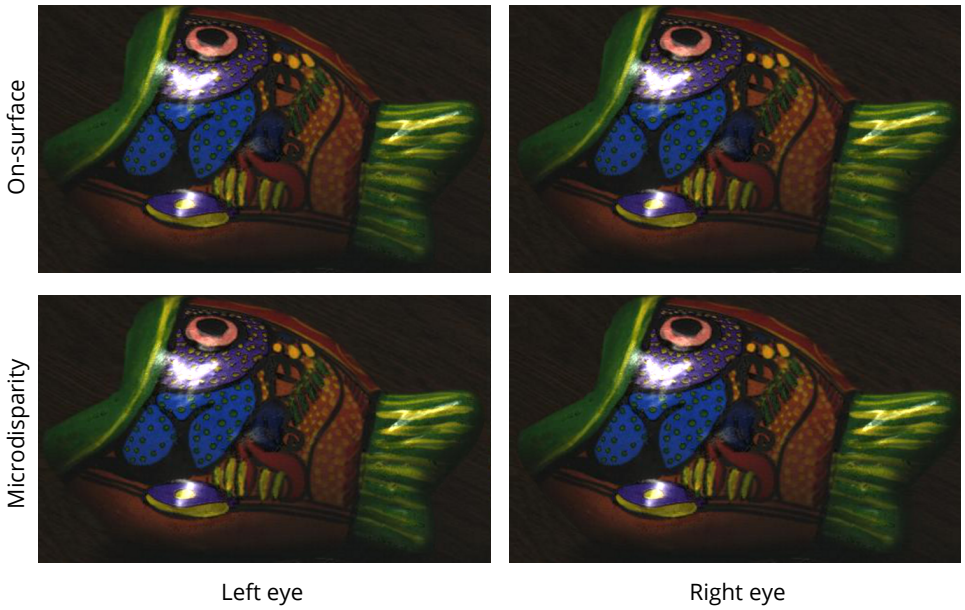


Figure 4.10. A stereo pair is generated from a single image by warping it according to the depth map *(left)*. In addition to the basic warp, a small disparity is added to the highlights to enhance the picture *(right)*.

the versions were randomized for each test. In three sessions we asked subjects to indicate the most *unrealistic* highlight depiction, then the most *uncomfortable* impression, and finally the *preferred* one. We chose this negative formulation, as our main goal is to reduce artifacts while retaining material and stereo perception.

The results of the study are shown in Tab. 4.1. Surprisingly, many subjects judged physical simulation as unrealistic. However, as noted by Wendt et al. [2008], a physically more correct rendering does not have to *appear* more realistic. While 38 % of the participants found on-surface highlights unrealistic, more subjects consider this technique superior to the costly, physical highlights, explaining its success in practical applications. In the last session 40 % were in favor of our technique, while 34 % and 26 % preferred on-surface and physical highlights, respectively. We obtained $\kappa < 0$ (Fleiss' kappa) in all three sessions which suggests poor agreement between the subjects. The advantage of our technique found in the first two sessions was statistically significant, however it was not significant in the last session.

4.3 Discussion

Both negative formulation of the first two questions and the order of sessions might have biased the third experiment to our advantage. However, the construction of the study made our non-expert subjects more aware of possible issues

	Unrealistic			Discomfort			Preference		
Fig.	Phys.	Flat	Ours	Phys.	Flat	Ours	Phys.	Flat	Ours
4.3	60 %	30 %	10 %	90 %	0 %	10 %	10 %	40 %	50 %
4.4	50 %	20 %	30 %	70 %	20 %	10 %	10 %	30 %	60 %
4.5	70 %	30 %	0 %	60 %	30 %	10 %	30 %	20 %	50 %
4.6	40 %	50 %	10 %	60 %	30 %	10 %	50 %	30 %	20 %
4.7	30 %	60 %	10 %	60 %	40 %	0 %	30 %	50 %	20 %
Avg.	50 %	38 %	12 %	68 %	24 %	8 %	26 %	34 %	40 %
$p <$	.0005	.0074		.0001	.0385		.1482	.3715	

Table 4.1. The results of the user study (see text for discussion).

with stereo gloss depiction. Surprisingly, in the preference session we did not find a significant effect. One may notice that our method works better when highlights are well-defined and isolated (Figs. 4.3, 4.4, 4.5), rather than complex or of lower sharpness (Figs. 4.6 and 4.7). In the latter cases the conflict between left and right views is smaller, and the resulting discomfort is perhaps more tolerable.

Including curvature when using our model proved to be useful. It is visible how constant highlight disparity would be objectionable on an object with varying curvature (Fig. 4.11a). However, when the highlights appear mostly in high-curvature regions, the need for our technique is perhaps less obvious, because the conflicting highlights are of smaller size, and the disparity reintroduced by our method is attenuated by the curvature term (see Fig. 4.6).

We ignored the dependency of disparity visibility on surface glossiness, i. e., sharpness of highlights. For less glossy surfaces the amount of disparity introduced to highlights may be too small to be detectable, and thus the advantage over on-surface highlights negligible. On the other hand, for very blurry highlights even physical disparity can be hard to spot, and the overall difference between the three methods is not substantial in such cases. Fig. 4.7 is a good illustration to this issue: the highlights are blurry, hence the disparity added by our method cannot be easily spotted. However, it is also hard to find a region in which physically-based highlights cause strong binocular conflicts.

Another simplification is the lack of distinction between highlights and mirror reflections. In the case of reflections, the observer can distinguish shapes of the reflected objects, however disparity added by our algorithm lacks appropriate variation, and the cardboard effect appears (see Fig. 4.12). Those cases would need to be handled separately, either manually or by specialized algorithms. Finally, it is possible that in some cases dealing with less pronounced highlight disparities than in our tests, physical computation would be superior to our method. Nevertheless, our method can be considered a safe automatic replacement for physical computation.

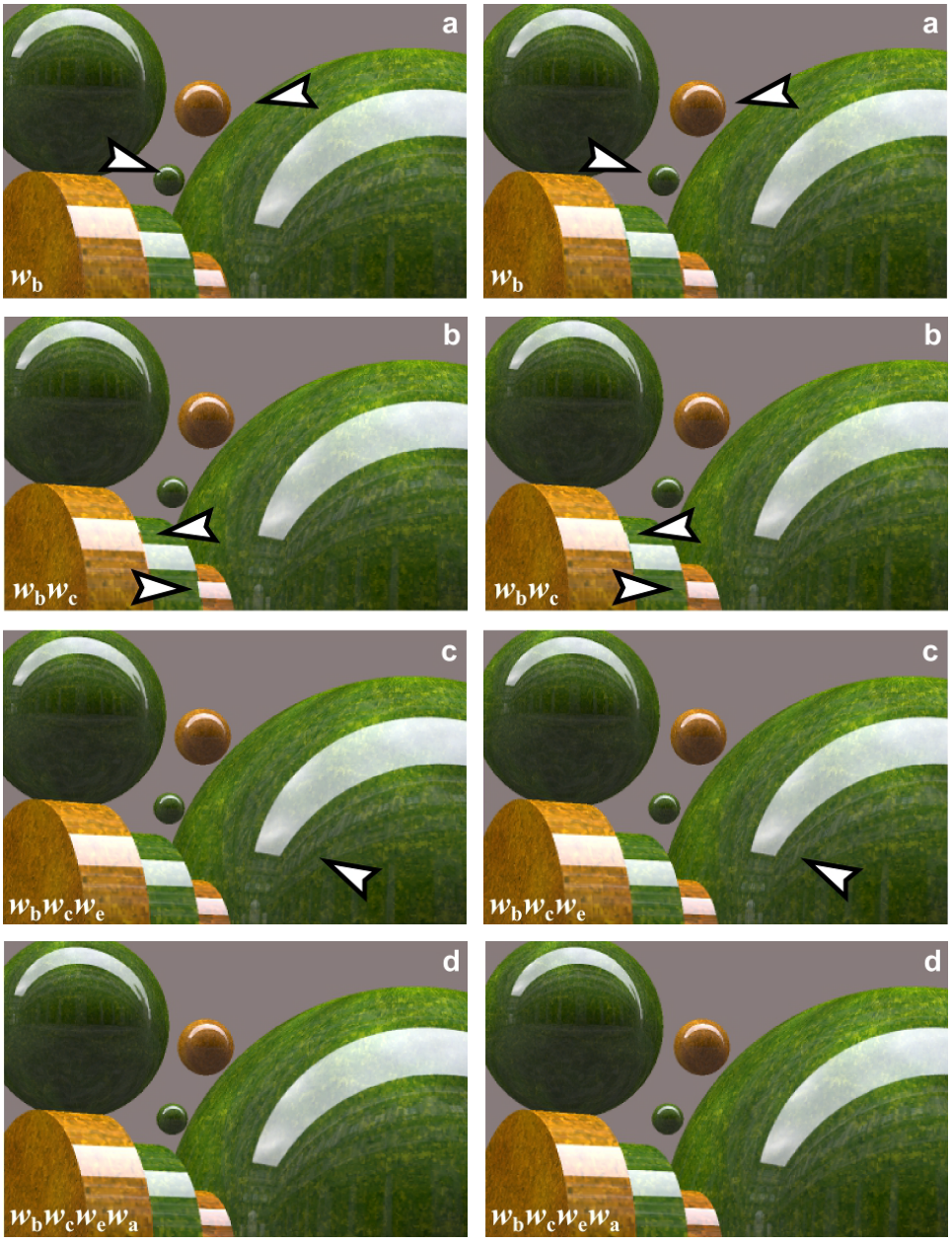
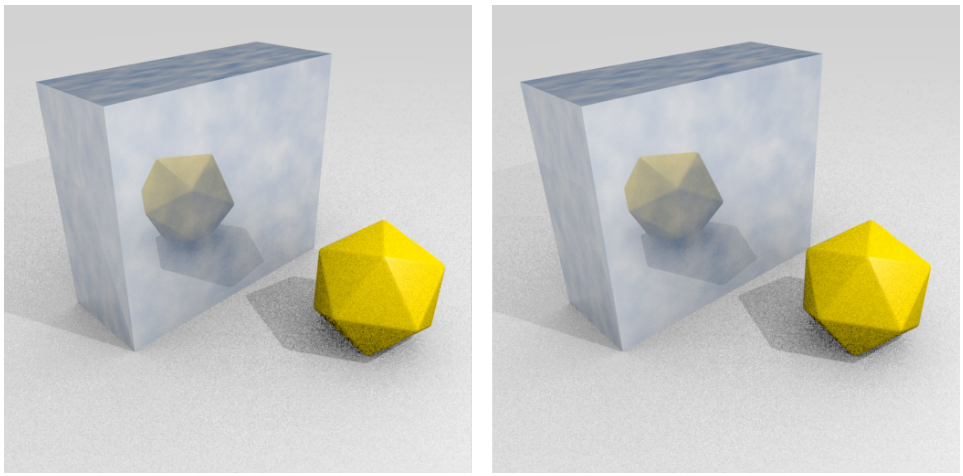


Figure 4.11. Without w_c, the highlight disparity on the smaller spheres is too large *(a)*. Missing w_e leads to jagged and/or floating highlights at the edges *(b)*. The artist may decide to adjust highlight disparity locally *(c)* to optimize the result *(d)*.



Left eye Right eye

Figure 4.12. A failure case. In a planar mirror the reflected scene is clearly visible, but the reflection disparity is constant. Note the unnatural flatness of the polyhedron reflection (cardboard effect).

Temporal aspects Sakano and Ando [2010] investigated stereopsis with simultaneous head motion and observed only weak glossiness enhancement with respect to the binocular cue only. However, their stimuli did not contain any highlight disparity and were composed of flat, differently oriented facets. Wendt et al. [2010] considered smoothly curved surfaces and physically-correct highlight disparity, which in combination with surface motion resulted in more reliable judgment of gloss strength than stereopsis for static stimuli. We relegate as future work more systematic study on the impact of camera and object motion, or dynamic lighting changes on the performance of our technique. As can be seen in the accompanying video, our method has very good temporal consistency.

4.4 Generalization to Multiple Layers

The described technique decomposes the scene into a diffuse and a specular layer. This approach is oblivious to the fact, that at any point in the image there may be actually multiple specular layers with their own distinct disparities. In a subsequent work with Dąbała et al. [2014] we extended our method to handle the general case of rendering scenes with multiple reflections and/or refractions. This extension introduces two main novelties: (1) disparity and rivalry estimation for each node of the ray tree of the scene, and (2) local manipulations of the stereo camera separation for each pixel and render-tree node.

The technique assumes Whitted-style [1980] ray-tracing, in which only perfect (single-direction) reflections and refractions are accounted for. This allows to encode the image as a binary ray tree, in which the root contains the diffuse

radiance at the first hit point, the left child – the radiance after the first reflection, the right child of the left child – the radiance after a reflection followed by a refraction, etc. The image is effectively decomposed into layers, each representing some level of indirection. In practice, we considered paths with at most two reflections/refractions. For each layer, corresponding points for the left and the right view are matched. This is accomplished using a computer-vision-based correspondence algorithm, simulating the matching performed by the human visual system. After the matching has been performed, binocular rivalry is also estimated by comparing luminance patterns of the corresponding image patches.

The data obtained this way is combined into a cost function, penalizing comfort zone violations, large disparities between the layers, and binocular rivalry. Next, the cost function is used to drive the camera interaxial distance optimization, which is performed locally for each pixel of the image. To prevent large deviations from the original image, a data term is also included in the cost function. Finally, the locally-optimal camera parameters are smoothed to ensure spatial coherence of the resulting image. The entire pipeline is illustrated in the Fig. 4.13.

4.5 Summary

In this chapter we presented a method of rendering highlights in stereoscopic 3D, that helps to preserve appearance of glossy materials. The technique is easy to include in an existing shading system and can be computed efficiently. Our approach provides a good alternative to physical and on-surface highlights.

In the follow-up work [see Dąbała et al. 2014] an alternative approach was explored in which the highlights are not manipulated in the image-space, but are indirectly influenced by local adjustments of the stereoscopic camera parameters. This new method is also capable of handling scenes with multiple layers of semitransparent, reflective and/or refractive objects.

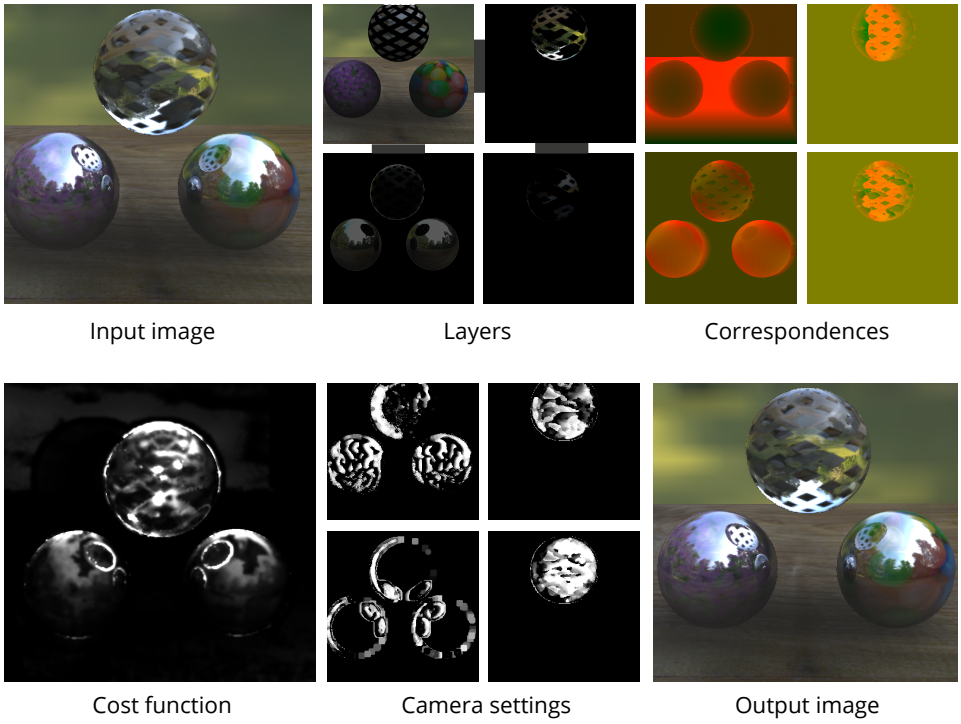


Figure 4.13. The overview of the pipeline proposed by Dąbała et al. [2014]. The input synthetic image is represented in a layered manner, with each layer corresponding to a node in the ray tree. For each layer stereo correspondences are estimated, and used to calculate the cost function. The cost function drives the per-layer and per-pixel optimization of the stereo camera settings. Spatially smoothed settings are eventually used to synthesize the optimal image. Pictures: Dąbała et al. [2014]

Chapter 5

Stereoscopic Film Grain

Analog photographs and films often feature a random high-frequency texture, commonly called *film grain*. It is a by-product of the photographic process, in which crystals of silver salts that were exposed to light are transformed into larger groups of metallic silver or dye clouds, creating an image of visible granularity, as shown in Figure 5.1. Film grain is often considered an artifact and is removed in post-production. This, however, is not always an easy task, since there is no simple way of discriminating between random noise and fine details of the photographed objects. Moreover, grain is sometimes intentionally preserved or even added to evoke certain mood, stylize, or imitate the look of old movies (see Fig. 5.2). The fundamental requirement in such cases is to retain a uniform look of grain between various regions of the film. For example, when computer-generated elements are inserted into a scene, matching film grain has to be added. This allows to seamlessly integrate different types of content, without creating a clear distinction between them, which would be perceived as an artifact. For similar reasons, grain has to be added also to fully synthetic shots, because of a possible mismatch with the previous, real shot. Even if the objects in the scene are real and are merely to be processed (e. g., resized), the grain has to be removed, and added back afterwards [Seymour 2011a]. Thus, grain management, i. e., the set of techniques for removing, adding, and matching the grain, is a significant part of the film post-production process.

A noisy pattern similar to grain can also appear in digital photography (Fig. 5.1, bottom), however it is often recognized as less appealing than analog grain. On the other hand, pictures taken with the sensor set to lower sensitivities can look too clean. Therefore, film grain can be added to mask digital noise or compensate for a "too synthetic" look [Kurihara et al. 2008]. The idea of adding grain is not limited to photography or film, but also appears in computer games [Giant Bomb 2012]; e. g., the best-selling game *Limbo*, which uses strong film grain as a means of stylization (see Fig. 5.3). Refer to Sec. 2.6 for review of research related to noise, grain, and point volumes in graphics, image processing, and scientific visualization.

Figure 5.1. Grain in analog black-and-white photograph *(top)*, in analog color photograph *(middle)*, and sensor noise in a digital photograph *(bottom)*. Photos: Halicki (CC BY 3.0), RX-Guru (CC BY-SA 3.0), Sean Molin (CC BY-NC-ND 2.0)

Figure 5.2. Some films use heavy grain as a means of stylization. From top to bottom: *Saving Private Ryan*, *300*, and *Planet Terror*. Frames: DreamWorks SKG, Warner Bros., Dimension Films

Figure 5.3. The issue of grain is not limited to films and appears in other media, such as computer games (*Limbo* in the above example). Artwork: Playdead

Grain in stereoscopic 3D Grain application has been identified as a significant problem in the 3D film post-production process: for example, the VFX company Pixomondo spent weeks on R&D just to address the issue of grain in the Oscar-winning film *Hugo* [Seymour 2011b]. The difficulty is due to the interplay between the left and the right half-image. If the same grain pattern is added to both views, it is fused by the observer and has the depth of the screen plane. This leads to an unpleasant shower-door effect, and causes double vision if the distance between the screen plane and the scene is large. Another option is to add two uncorrelated grain patterns, in agreement with what happens when two cameras are used during the capture. However, only limited amounts of uncorrelated grain can be tolerated [Lankheet and Lennie 1996], because presence of many unmatched features impedes fusion and leads to visual discomfort. Binocular rivalry may occur, and cause a characteristic "shiny look" (Fig. 5.4, top row). The last option is to project grain on the surface of the objects, i. e., display it at the same depth as the object it occludes. This technique does not have disadvantages of the two previous ones, and is a natural choice especially in imagery created in the process of 2D-to-3D conversion, since the grain can be displaced together with the objects and does not require much additional attention. However, this approach creates impression that the grain belongs to the objects' texture, and emphasizes any imperfections of depth (Fig. 5.4, middle row). Conversion from 2D does not preclude usage of uncorrelated grain, however it is not an easy task to remove all existing grain, and thus some portion of it may remain on the surface. The industry standard is to use uncorrelated grain, projected grain, or combination of both [Seymour 2011b, 2012, Ridanovic 2011]. Winter and Gandolph [Winter and Gandolph 2013] build on the idea of projected grain, and propose how to handle grain in the case of uncertain depth values in the stereoscopic content.

Our contribution We propose a new approach to adding grain, in which the input grain pattern is decomposed into particles and distributed in depth (Fig. 5.4, bottom row). We draw inspiration from the way other film artifacts are treated during 2D-to-3D conversion: lens flares, or scratches and bigger dust particles found in old films are usually placed somewhere between the objects and the observer. To our knowledge, however, it has not been proposed so far to treat film grain in the same way.

We motivate our choice by the idea of medium-scene separation: There is a distinction between the mental image of a depicted object and its depiction, and one cannot see both at the same time [Gombrich 2000]. Projecting grain on the surface of objects violates this distinction in a certain way – instead of a stereoscopic *grainy depiction* of an object we obtain a stereoscopic depiction of a *grainy object*. By detaching the grain from the objects, we make an effort to restore, at least partially, the medium-scene separation which is disrupted when moving from two-dimensional imagery to stereoscopic 3D. Additional benefit of our approach is that we avoid emphasizing potential stereoscopic artifacts, such

as unnatural flatness or depth map errors. Lastly, our approach can improve the stereoscopic composition of a scene: In traditional films, one should avoid "visual clutter", as it leads to a feeling of uneasiness in the audience. In S3D this rule is reversed – if there are too few objects in the scene, stereoscopic depth cues will be too sparse, and the overall look will be less intense [Mendiburu 2009]. Since our grain introduces additional details in depth, it can help solve this problem. See Sec. 2.3 for additional background in stereoscopic perception of point volumes.

5.1 Stereoscopic Grain

The input to our algorithm are the left and right half-images L, R, together with the film grain pattern G to be applied, and the dense correspondence map $d\colon \mathbb{N}^2 \to \mathbb{R}$ between L and R. For any pixel position $\mathbf{p}$ in R, $[\mathbf{p}_x + d(\mathbf{p}), \mathbf{p}_y]$ is the corresponding position in L. The grain is applied to the image using an application operator $\oplus$, which is typically a weighted addition, with the weights dependent on the pixel intensities in the input image.

The output is a modified grain pattern G', such, that the stereo pair $(L \oplus G,\ R \oplus G')$, gives impression of grain floating in space. To achieve this goal, the grain pattern needs to be re-interpreted as a collection of shapes in 3D space, that appears exactly as G when seen by the left eye. Based on that interpretation G' is determined. We proceed in two steps: (1) the grain pattern is segmented into individual particles, that are afterwards assigned to n different layers; (2) these layers are then appropriately stacked in depth, with increasing distance from the surface of the objects.

Grain segmentation In this step every pixel of the grain pattern G is assigned to one of the layers $G_1, G_2, \ldots, G_n$. For any pixel $\mathbf{p}$, $G_i(\mathbf{p})$ equals $G(\mathbf{p})$ if $\mathbf{p}$ has been assigned to layer i, and 0 otherwise. The assignments are made using a similar approach to *watershed by flooding* introduced by Beucher and Lantuejoul [1979]. First, we detect local luminance minima and maxima in G using a 3×3 min- and max-filter, and assign them to layers by random. Next, the assignments are propagated iteratively. In each iteration, pixels that have been already assigned to layers propagate their assignments to their immediate unassigned neighbors. If at any iteration two or more pixels try to propagate to the same pixel, the one with the closest luminance value has the precedence. This process is illustrated in Figure 5.5. Since the spread between grains is usually in the order of several pixels, only few iterations are needed to assign all pixels to layers. An exemple result of this algorithm is shown in Fig. 5.6.

Layer stacking Now, the layers can be distributed in depth. The baseline distribution is obtained by putting

$$G'_{\mathrm{b}}(\mathbf{p}) = \bigoplus_{i=1}^{n} G_i(\mathbf{p}_x + d(\mathbf{p}) - \frac{i}{n} \cdot \alpha, \mathbf{p}_y),$$

Figure 5.4. Film grain overlay in stereoscopic 3D. Grain that is added independently in each eye is hard to fuse and causes discomfort. In extreme cases, binocular rivalry appears, and the image looks "shiny" *(top row)*. Projecting grain on the surface does not ensure medium-scene separation *(middle row)*. Our technique ensures that grain is separated from the scene, but is easy on eyes *(bottom row)*. Use uncrossed (parallel) free fusion to see the examples. Scene: Blender Foundation, www.bigbuckbunny.org

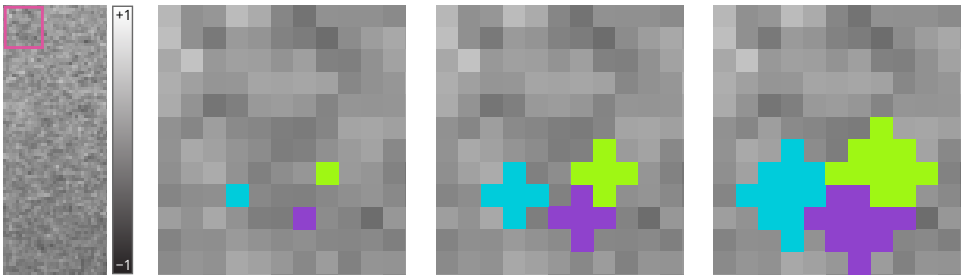


Figure 5.5. Initial steps of grain decomposition. (a) Local extrema are assigned to random layers, as indicated by colors. For illustration purposes we ignored the existence of other extrema. (b–c) In subsequent iterations, assignments are propagated to neighboring pixels. (c) In case of two or more pixels propagating their assignments to the same location, the one with the closest luminance value takes the precedence.

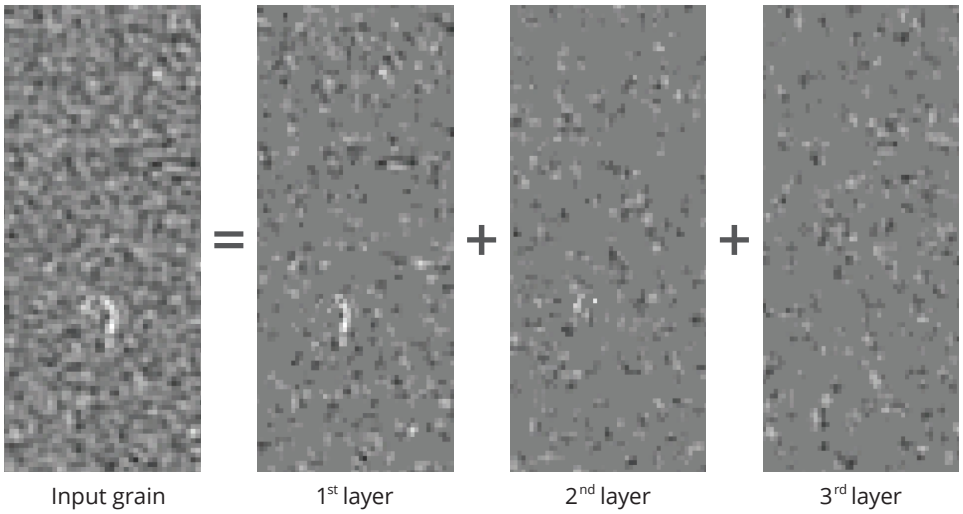


Figure 5.6. Each pixel of the grain pattern is assigned to one of the layers.

where n is the number of layers, and α is a parameter defining the thickness of the "grain cloud": When $\alpha = 0$, the grain is placed on the surface of the objects, for $\alpha < 0$ it appears to be inside, and for $\alpha > 0$ it surrounds them. The greater α the thicker the cloud around the objects.

Using constant α results in a regular distribution of film grain. In some cases, however, this may not be the best solution, therefore, we allow replacing α with a smoothly varying activity map $A\colon \mathbb{N}^2 \to [\alpha_{\min}, \alpha_{\max}]$, which maps a position in the image to a desired grain-cloud thickness. The $\alpha_{\min}$ and $\alpha_{\max}$ parameters are derived experimentally in Sec. 5.3 A thick grain cloud can obscure small depth details in the original scene, due to the *disparity masking* phenomenon [Howard and Rogers 2002, Chapter 19.6.3d], where the perception of a disparity corrugation is affected by another, superimposed signal. Therefore, it is necessary to modulate α value taking into account the scene geometry, and use a smaller value in regions with a high disparity variation. Bigger values of α may be used in flat regions to maximize depth impression, and counteract objectionable flatness (e. g., cardboarding effect or lack of details). An important observation is that masking affects mostly signals of similar spatial frequencies. As grain adds mostly high frequency disparity corrugations, A needs to account only for those. Additionally, A does not need to account for very high spatial frequencies (above 5 cpd) because those have a negligible effect on disparity perception [Tyler 1975]. As a result, we first need to separate the signal that should be considered by the function A. We do it using a simplified version of the binocular disparity model presented by Didyk et al. [2011]. The vergence angles are computed separately for each location in the scene assuming that the observer verges on it perfectly. Thus, the correspondence map d is converted to a vergence map v, operating in visual angles instead of pixel shifts. Here, we follow terminology from perception literature, where *disparity* is defined as difference of *vergence angles* [Howard and Rogers 2002, Fig. 19.1]. Then, the relevant disparity signal is separated by a band-pass filter with cut-off frequencies φ_{L} and φ_{H}. One could consider a full frequency decomposition to multiple, narrow frequency bands, as it was done in the original disparity model. However, we found that our solution is sufficient and more practical. It is also motivated by the fact that the human visual system has only a limited number of visual channels that are tuned to different disparity frequencies. Although the individual channel bandwidth has not been clearly established, the existing estimate suggest the range of 2–3 octaves [Howard and Rogers 2002, Chapter 19.6.3d]. We found $\varphi_{\mathrm{L}} = 0.625$ and $\varphi_{\mathrm{H}} = 5$ cpd (3 octaves) to give good results.

The band-limited vergence map contains signal whose perception may be affected by the additional grain disparity. At this point we are not interested in exact disparity values, but rather in regions where the thickness of the grain cloud needs to be attenuated due to high vergence variations in the original image. Therefore, we apply thresholding at the amplitude θ (we used 2 arc min), and apply low-pass filter with a cut-off frequency $0.5 \cdot \varphi_{\mathrm{L}}$. We denote the result as $\hat{v}$,

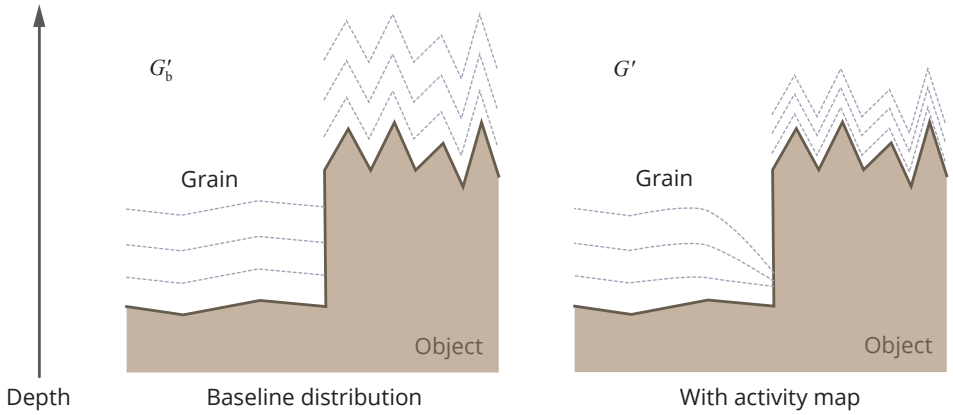


Figure 5.7. A thick grain cloud suspended above the object may mask small depth details of the geometry *(left)*. Using activity map *(right)* the thickness of the grain cloud is attenuated in the regions with high disparity variance; hence, the small depth details stay visible.

and use it to modulate the activity map:

$$A = \alpha_{\max} - \hat{v} \cdot (\alpha_{\max} - \alpha_{\min}).$$

Finally, the resulting grain distribution is defined as

$$G'(\mathbf{p}) = \bigoplus_{i=1}^{n} G_i(\mathbf{p}_x + d(\mathbf{p}) - \frac{i}{n} \cdot A(\mathbf{p}), \mathbf{p}_y),$$

and is illustrated in Fig. 5.7. See Fig. 5.8 for a comparison of pictures with and without the activity map.

Compositing In order to add our grain layer to existing footage, they are both combined using addition in gamma-corrected space as the grain application operator. This guarantees that the grain is approximately equally visible everywhere in the picture.

5.2 Results

We applied our method to two rendered sequences – BIRD and SINTEL (Fig. 5.4 and Fig. 5.10) and one video sequence – BALLET (Fig. 5.11, right side). Each sequence simulates a different type of grain: large and clearly visible grain of an old, black-and-white film (SINTEL), less pronounced grain of a more recent film (BIRD), and fine grain of a modern (yet grainy) film (BALLET). In the SINTEL and BIRD sequences, we used freely available scans of 35 mm film[1]. To mimic

[1] Downloaded from http://7dblue.wordpress.com/tools-downloads/.

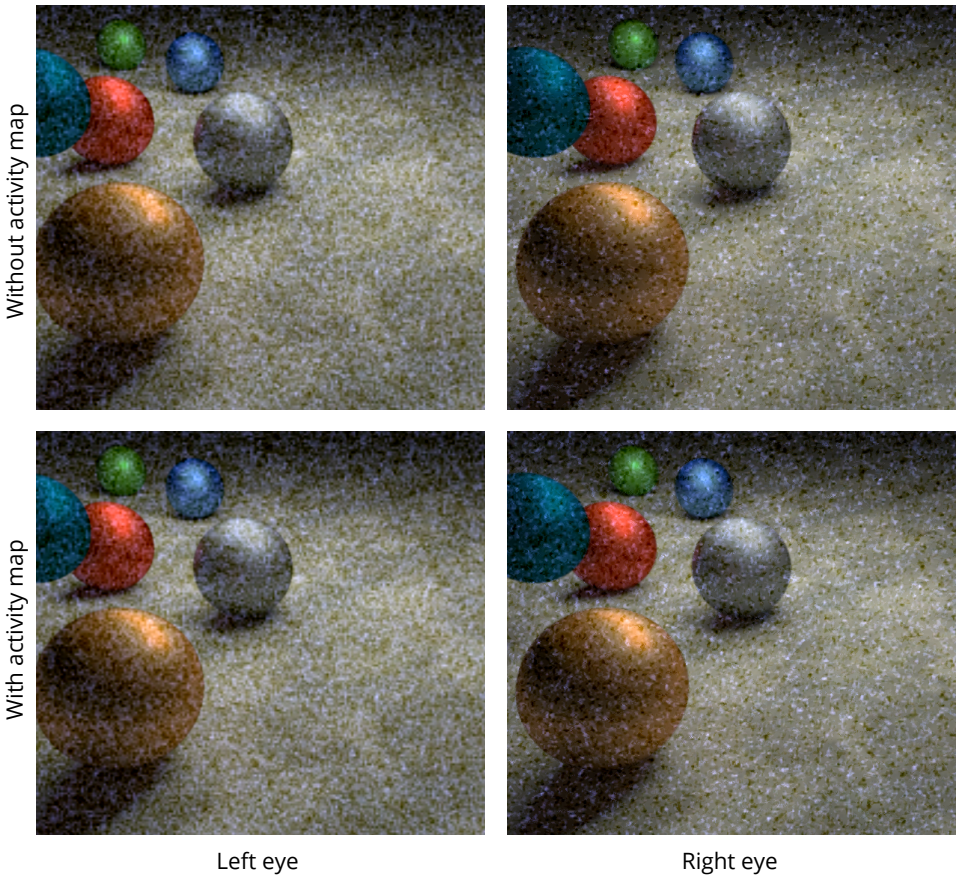


Figure 5.8. In the regions with high disparity variation, a thick grain cloud may attenuate perceived distances in depth *(top)*. The activity map detects such regions and reduces the thickness accordingly *(bottom)*. Note, how the distances between the balls are better preserved. The thickness of the grain cloud on the right side remains unchanged.

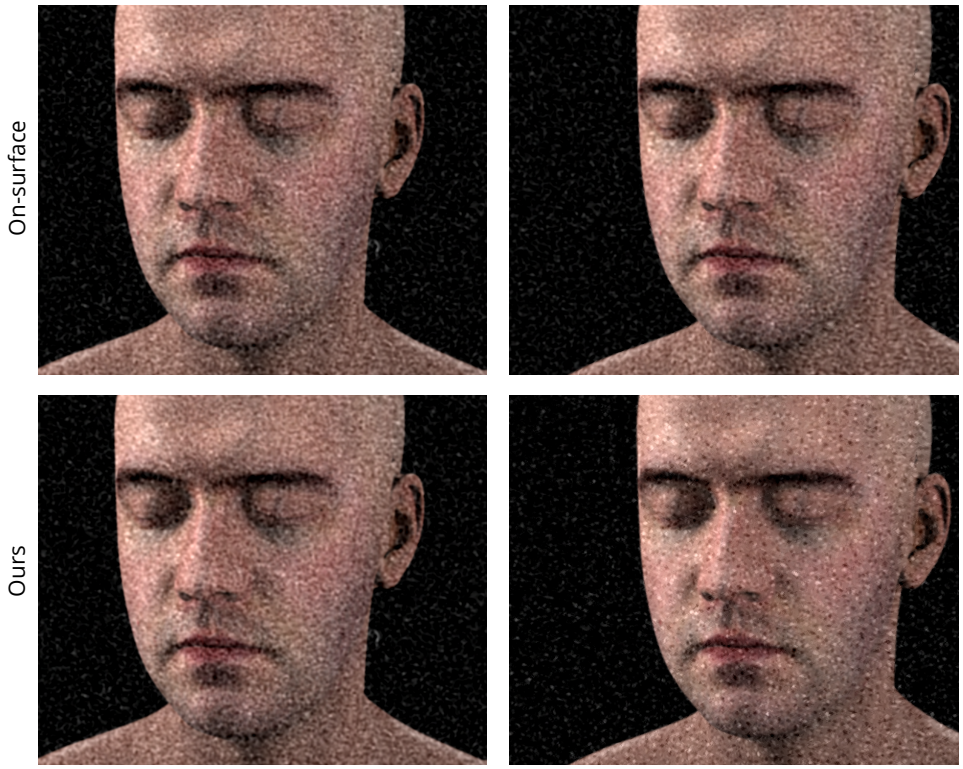


Figure 5.9. When the depth buffer is not detailed enough, on-surface grain contributes to undesired cardboarding effect *(top)*. Our floating grain solves this problem *(bottom)*. See study in Sec. 5.5 for details. Note, that the grain in this figure was exaggerated for the purpose of illustration. Model: Lee Perry-Smith (www.ir-ltd.net)

different sizes of film stock, two differently sized crops of the grain images were used. For the BALLET sequence, the grain was generated in Adobe After Effects CS4. We believe that with this variability, we exhausted the range of useful sizes of grain. For example, it is unlikely to find bigger grain in films or games than the one used in the SINTEL sequence. On the other hand, the grain in BALLET sequence is barely visible. Additionally, we also generated an image of a face, where the depth map was artificially compressed to mimic a typical artifact of 2D-to-3D compression, and compared on-surface grain to ours (Fig. 5.9). We used $n = 5$ layers and we set the volume parameters to $\alpha_{\min} = 5.3\,\text{px}$ (ca. 8 arc min) and $\alpha_{\max} = 9.6\,\text{px}$ (ca. 14.4 arc min). These parameters were derived in a perceptual study described in Sec. 5.3. The figures *serve as an illustration only* (in particular Fig. 5.9 features *exaggerated* grain). We refer user to the supplemental materials[2], where the resulting videos are provided. Please note, that it is very important to use a stereo system with minimal cross-talk levels, because the stereoscopic grain

[2] Available at http://resources.mpi-inf.mpg.de/FilmGrain/.

Figure 5.10. Results of our algorithm for sequence SINTEL. We compare uncorrelated grain *(first row)* with on-surface grain *(second row)*, and our floating grain *(third row)*. The images are supposed to be viewed using uncrossed (parallel) free fusion, and are provided as an illustration only. Please refer to the supplemental materials for the full video sequences. Copyright: Blender Foundation (www.sintel.org)

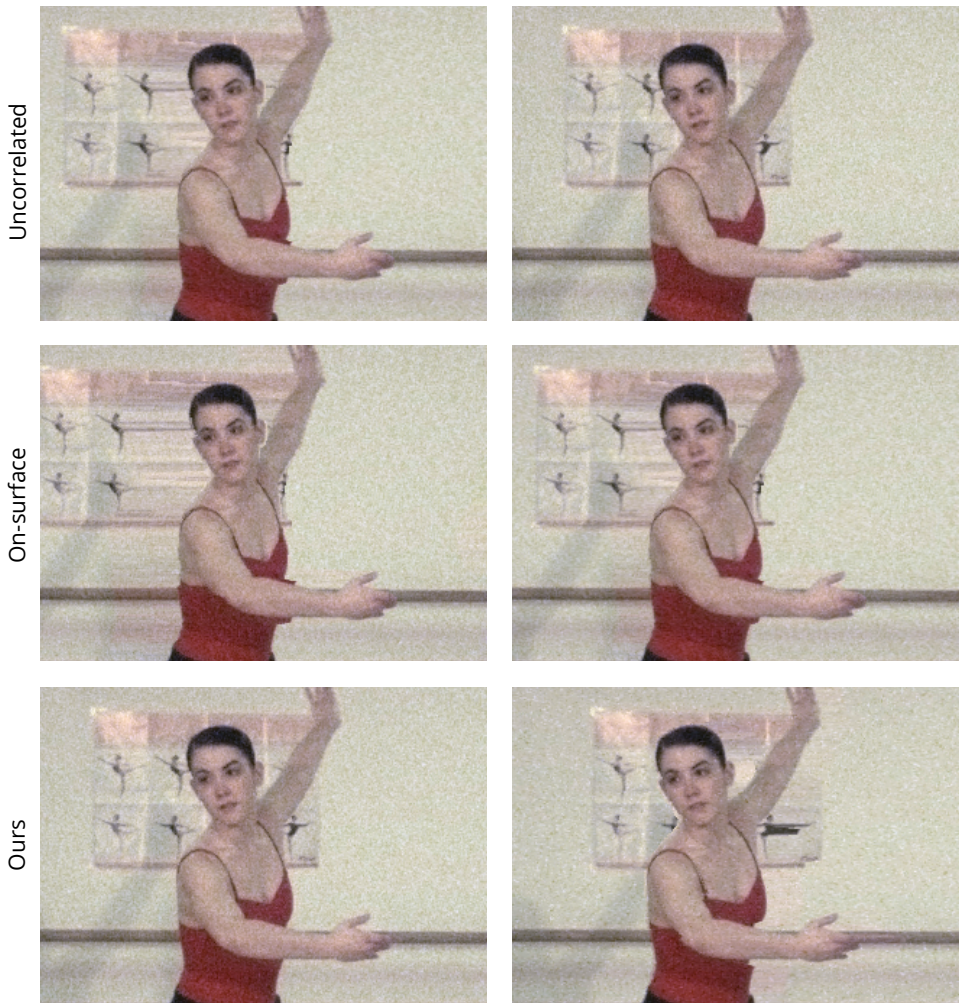


Figure 5.11. Results of our algorithm for sequence BALLET. We compare uncorrelated grain *(first row)* with on-surface grain *(second row)*, and our floating grain *(third row)*. The images are supposed to be viewed using uncrossed (parallel) free fusion, and are provided as an illustration only. Please refer to the supplemental materials for the full video sequences. Video: Microsoft Research

effect can be easily destroyed by ghosting. For similar reasons, videos should be watched at full resolution (no subsampling). Therefore, we discourage use of anaglyph glasses or systems that reduce resolution, e. g., row-interleaved displays, and recommend shutter glasses or dual-projector systems.

5.3 Parameters Estimation

Our method for stereoscopic grain has two free parameters $\alpha_{\min}$ and $\alpha_{\max}$, which are responsible for controlling thickness of the grain volume. Although both of them could be set by a skillful artist, in this section, we present a procedure that was used to obtain good values that can be used independently of the content.

Subjects Thirteen subjects (7 F, 6 M) took part in the experiment. All had basic background in computer graphics or computer vision, however, they were naïve with respect to the goal of the study, and their knowledge in stereoscopic graphics was limited. They had normal or corrected-to-normal vision, and were screened for stereo-blindness.

Equipment We used an Asus VG278HE 27-inch display (1920×1080 pixels), along with NVIDIA 3D Vision 2 active shutter glasses. The screen was observed from a distance of 50 cm. Measurements were performed in controlled, office-lighting conditions. The stimuli were displayed on a neutral grey background.

Task Because the participants were not familiar with different solutions for stereoscopic grain, the first part of the experiment was a training part. The subjects were shown the BIRD sequence, and they could switch between different kinds of grain (i. e., on-surface, uncorrelated, and our volumetric grain). They were also free to manipulate thickness of the volumetric grain using a slider and pause the sequence. Pausing was allowed only in the training session, in other experiments this option was disabled. Afterwards, they were asked to adjust the thickness of the volumetric grain so that the volume appearance is clear. In order to check whether they can distinguish among different kinds of grain after this short introductory session, they were shown the three different methods in random order (volumetric grain with their own settings), and were asked to assign them to their names. Ten participants did not have problems with identifying the methods, and they took part in the main experiment.

In order to estimate the two parameters ($\alpha_{\min}$ and $\alpha_{\max}$) we designed a two-step process. To estimate $\alpha_{\min}$, the participants were asked to adjust the thickness of our grain in the sequences BIRD, SINTEL, and BALLET (presented in random order), so that it had a just noticeable volume. At this point the attenuation map was disabled. Next, the map was enabled and the participants could adjust $\alpha_{\max}$ to their liking. Table 5.1 (second and third column) presents the total and by-scene averages of the two parameters.

scene	average $\alpha_{\min}$	average $\alpha_{\max}$	on-surface	ours
BIRD	5.2 ± 0.8 px	8.0 ± 1.4 px	7/10	8/10
SINTEL	4.4 ± 0.4 px	9.0 ± 1.9 px	9/10	9/10
BALLET	6.3 ± 1.1 px	12.1 ± 2.2 px	5/10	7/10
average	5.3 px	9.6 px	21/30	24/30

Table 5.1. The results of the parameter-estimation study and the preference study. The second and the third columns show average values of α-parameters by scene. The indicated errors are standard errors of the mean. The last two columns show how many times the given method was preferred over uncorrelated grain. Both results are significant, with p-values in one-sided sign test 0.02 and 0.0007, respectively. The result for ours vs. on-surface (16/30, not shown) is not significant.

5.4 Preference Study

In order to evaluate our technique, we conducted a preference study the day after the parameter estimation study (Sec. 5.3), in which the same 10 subjects participated. The apparatus and viewing conditions were the same as in the parameter estimation experiment.

Stimuli The sequences BIRD, SINTEL, and BALLET were used as the stimuli. Each sequence was processed using the three grain application methods, i. e., uncorrelated, on-surface, and ours. The grand average values of $\alpha_{\min}$ and $\alpha_{\max}$ obtained in the parameter estimation study were used for our method.

Task In a single trial, the subject was presented one of the sequences, and could freely switch between three versions (labeled A, B, and C) corresponding to different grain application methods. The subject was asked by the experimenter to indicate the version he/she preferred the most, and confirm the choice by pressing the Enter key. Then, the indicated version was removed, and the same question was repeated for the remaining two versions. Order of sequences, and order of methods for each sequence was randomized. The results of this study are presented in Table. 5.1 (fourth and fifth column).

5.5 Shape Naturalness

In the third study we analyzed the influence of our technique on shape perception, and its ability to mask artifacts of the depth map. The subjects and viewing condition were the same as in the two other studies.

Stimulus In this experiment we used the FACE sequence in two versions: with on-surface grain and our grain. The depth buffer in this sequence had been

remapped to enforce insufficient variation in depth, that often arises in the process of 2D-to-3D conversion.

Task This experiment consisted of a single trial. In it, the subject was shown the two versions of the sequence side-by-side (labeled A and B) in a random order. Next, the subject was asked by the experimenter to indicate in which version the face appeared more natural in terms of the 3D shape. Eight out of ten subjects found the face in the sequence processed using our method, as having more natural shape. The result is significant with $p < 0.055$ in the one-sided sign test.

5.6 Additional Results

An interesting case are "surfaces" with ill-defined depth, such as sky, participating media, or out-of-focus areas. To compare the performance of the two depth-dependent methods, i. e., on-surface and our grain, we generated additional four sequences: SQUIRREL, where the sky constitutes a large portion of the image, CANDLE, containing a semi-transparent smoke volume, and STONES and FLOOR with depth-of-field effects. In the case of the sky we assumed an arbitrary constant depth at some distance behind the character. To determine the depth for the smoke, we used a stereo correspondence algorithm [Hosni et al. 2013]. In the out-of-focus areas we used the depth of the corresponding non-blurred sequence. Using the *Match Grain* effect in Adobe After Effects CS6 (at the default settings and 16 px sample size) we closely matched selected frames from the feature films *Saving Private Ryan*, *300*, and *Planet Terror*. The results are presented in Figs. 5.12–5.15. See Fig. 5.2 for the reference frames from the films and refer to the supplemental material for the resulting full-resolution animations.

5.7 Discussion

On-surface grain adds to luminance patterns on the objects, thus influencing their depth perception [Didyk et al. 2012]. In several cases this may be undesired: First, infinite planes (e. g., sky or very distant backgrounds) and areas of undefined depth (e. g., out-of-focus backgrounds, smoke) look unnatural when their originally fuzzy depth becomes strictly defined (see Figs. 5.12–5.15). Second, when there is not enough depth variation, some objects may seem too flat (see Fig. 5.9). Last, when the depth buffer is not perfect, the errors are more evident (see Fig. 5.16). Our solution avoids all these problems: the sky and out-of-focus areas seem to have volume, and unnatural flatness and errors are masked.

We hypothesize that our approach might actually cause some detail hallucination. Stereoscopic grain introduces additional disparity signal to the original scene disparity map, which stimulates disparity-selective neurons that otherwise might not be activated. Additionally, there are a number of effects related to

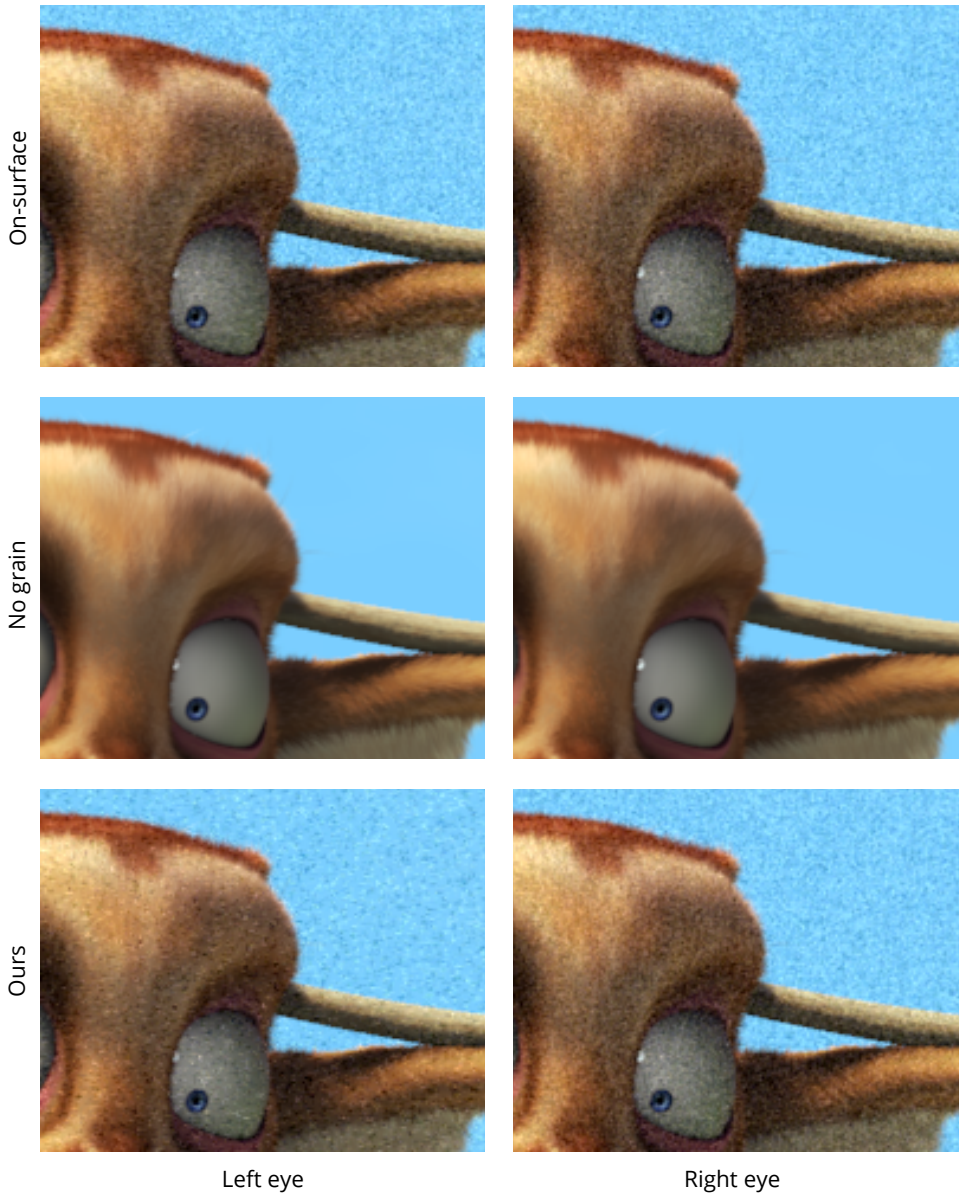


Figure 5.12. Sequence with an ill-defined "surface" – sky example. Grain matches the film *300*. In these cases applying one-layer projected grain changes the fuzzy perception of the objects, which is maintained by our method. See the supplemental material for the full-resolution sequences. Scene: Blender Foundation (www.bigbuckbunny.org)

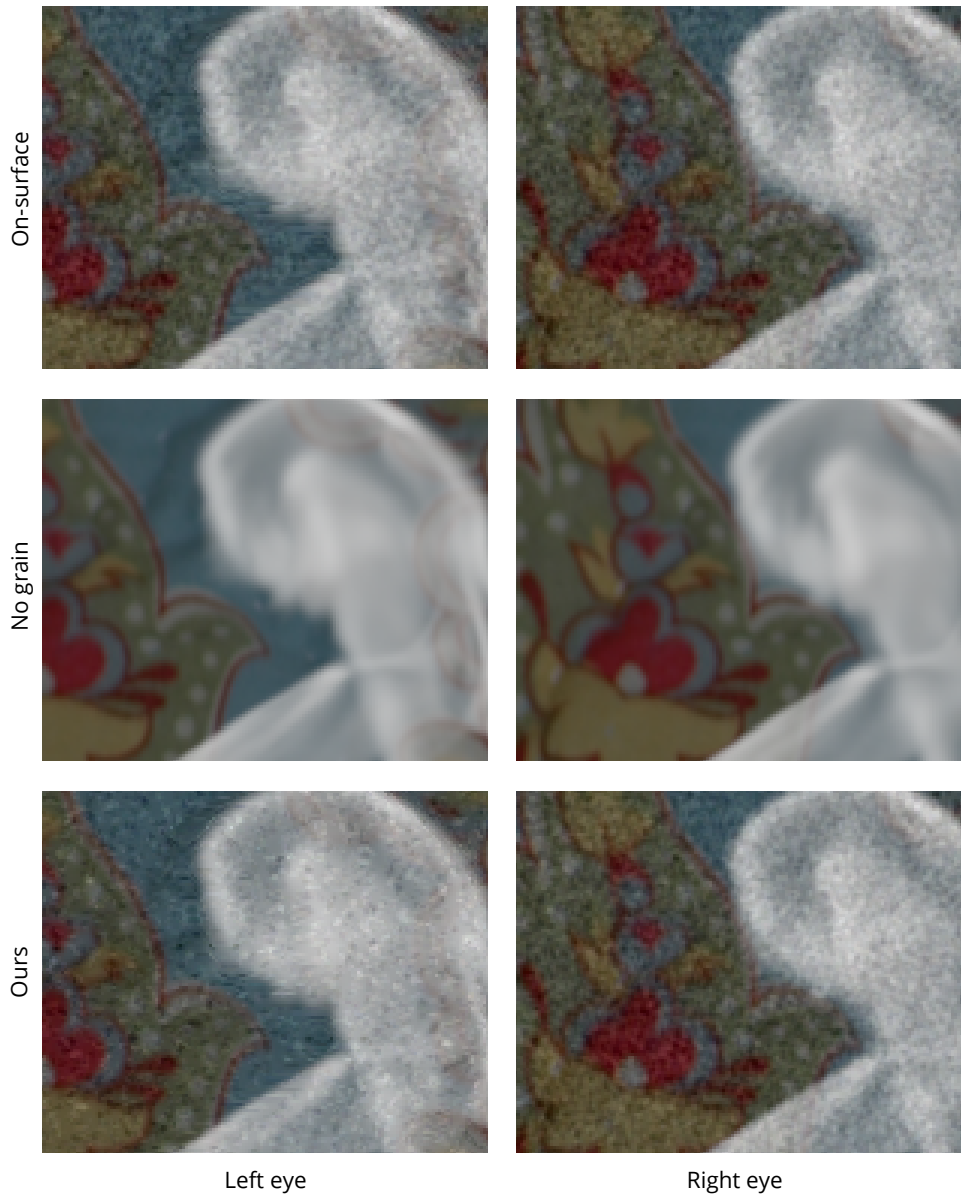


Figure 5.13. Sequence with an ill-defined "surface" – smoke example. Grain matches the film *Planet Terror*.

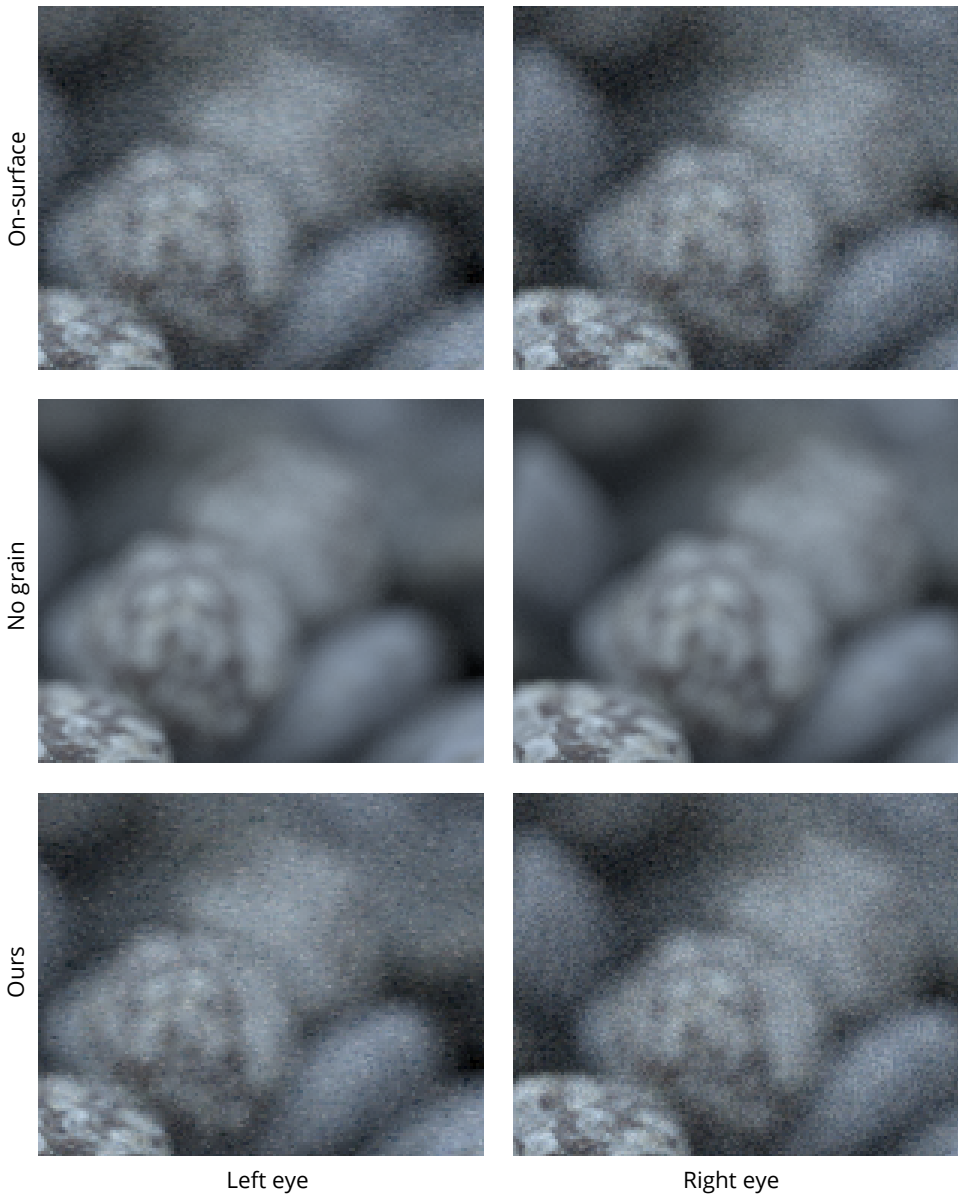


Figure 5.14. Sequence with an ill-defined "surface" – out-of-focus example. Grain matches the film *Saving Private Ryan*. Mesh: StevenColemanDesigns (www.blendswap.com)

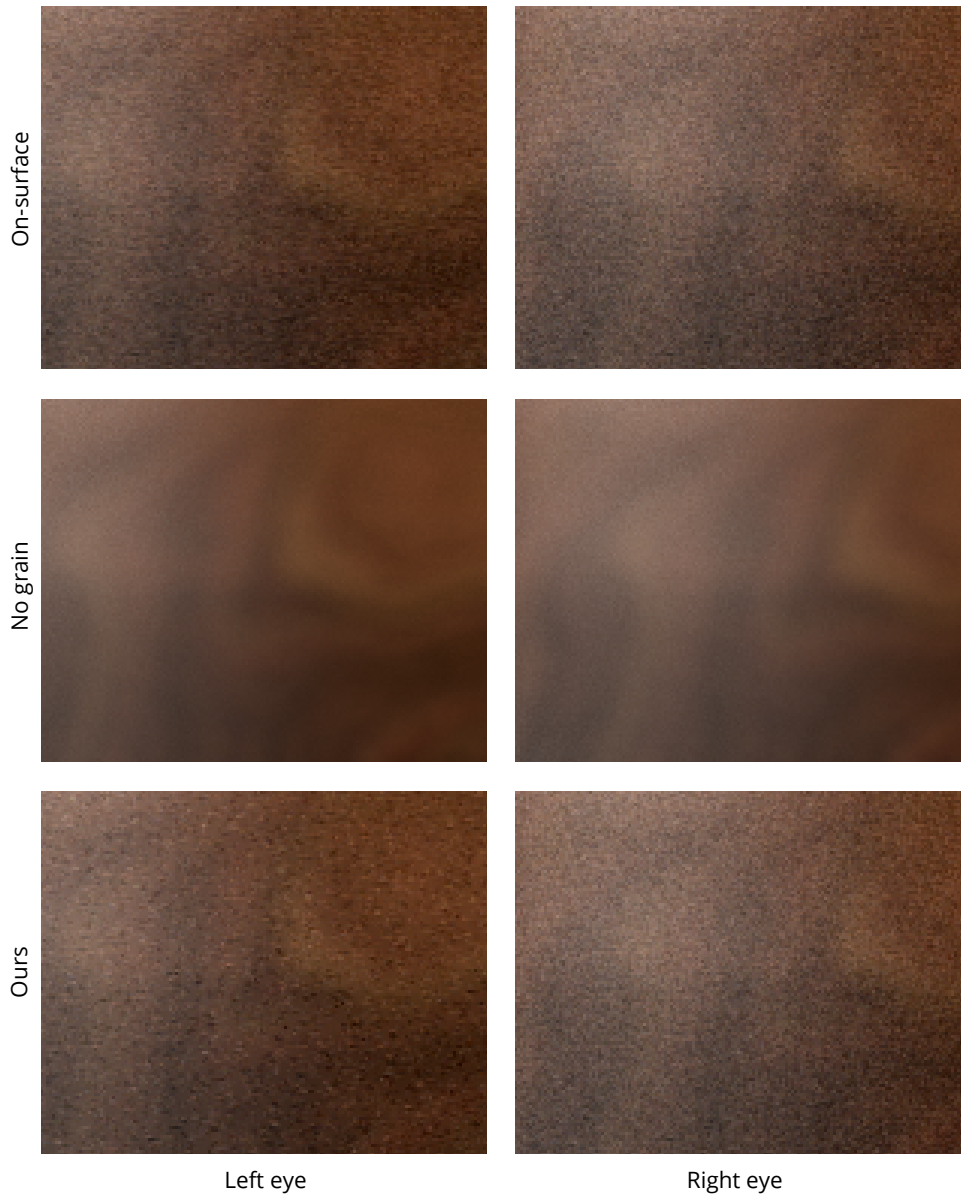


Figure 5.15. Sequence with an ill-defined "surface" – out-of-focus example. Grain matches the film *Saving Private Ryan*.

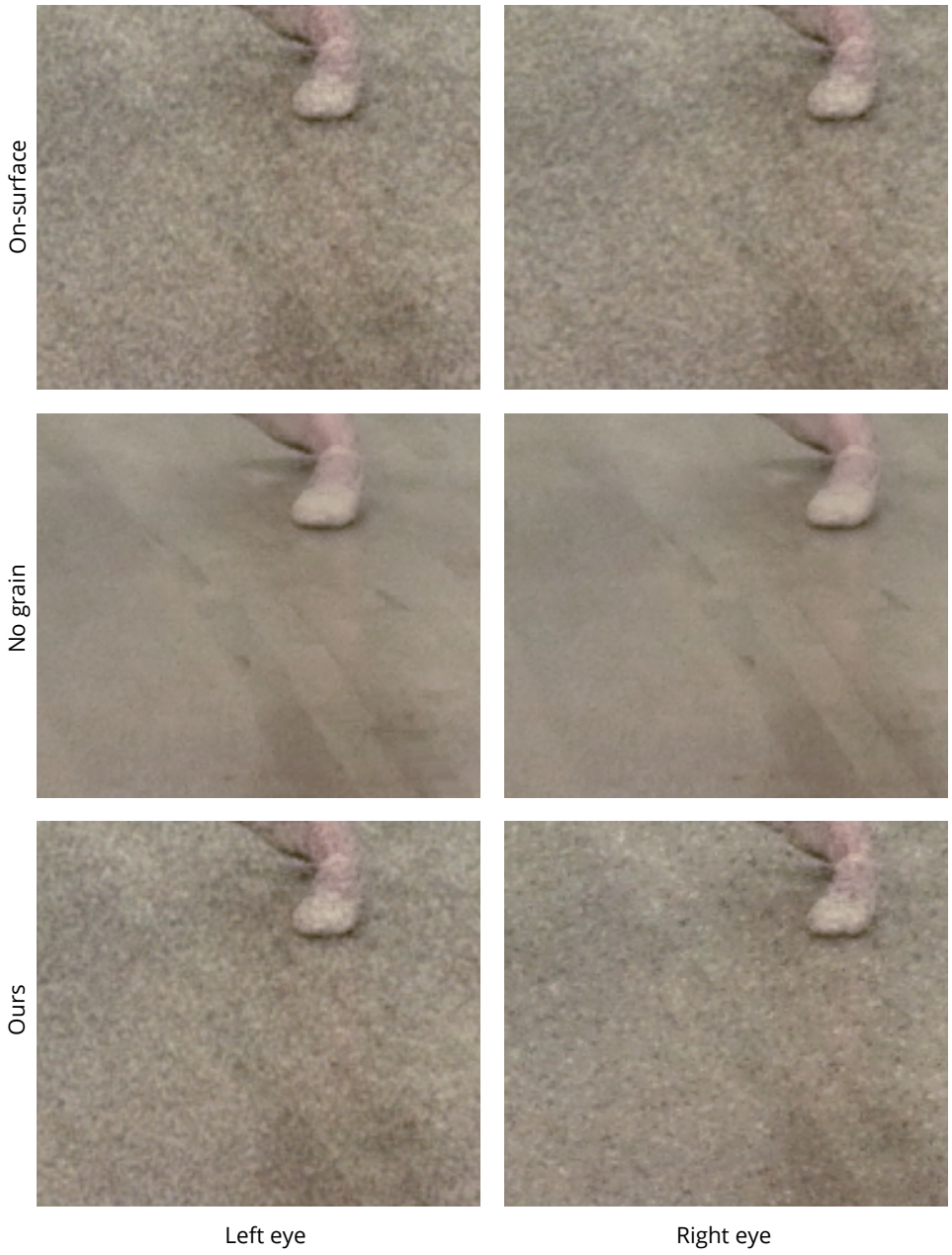


Figure 5.16. A fragment of the BALLET sequence. With on-surface grain the artifacts of the 2D-to-3D conversion are emphasized, whereas with our grain they are masked.
Video: Microsoft Research

layered RDSs, including attraction between near layers and depth repulsion for layers more distant than 4–6 arc min [Stevenson et al. 1991].

The number of layers that we used in the experiments ensured that the grain was perceived as a volume rather than separate layers. We did not explore further the influence of the quantity and relative placement of the layers, since the visual system seems to be insensitive to the specifics of the dot distribution within a volume (Sec. 2.3). For the same reasons, random distribution of dots within the given volume would yield visually equivalent results. Complete randomization of the grain placement would produce excessive pixel disparities, and would consequently break the binocular fusion.

Since the image-space size of the grain is not modulated, the apparent size of the particles may depend on the distance to the observer, with the more distant particles appearing larger. However, this was a deliberate design choice, as we wanted to modify only one view of the stereoscopic pair in order to maintain backward-compatibility (the other view is identical to the 2D version).

The approach we took in the parameter estimation study, with the estimation preceded by a training part, may have biased the results in favor of our method, because otherwise some subjects would have not noticed the differences between the strategies. However, we feel that it was a justified choice, because at least basic knowledge in stereoscopic 3D and film production, as well as attention to detail is required to appreciate this subtle effect.

The results of the study in Sec. 5.3 showed that 77% of subjects were able to discern different grain placement methods. The study in Sec. 5.4 showed that both the on-surface and our method are preferred over the uncorrelated grain (p-values in one-sided sign test 0.02 and 0.0007, respectively). Although the difference we found between the on-surface and our grain was not statistically significant, being on par with the industry standard can by no means be considered a failure, because in the end it is a matter of taste which tool to use, and the decision should be left to the artist. Additionally, the study in Sec. 5.5 demonstrated that there *are* cases, when using our method instead of the on-surface one is beneficial: when there are artifacts in the depth map, they are less obvious when our method is used. The main goal of the industry has always been increasing the picture quality. However, despite the technological advances in film-making, grain can be clearly seen even in very recent mainstream stereoscopic 3D films (e. g., *Transformers: Age of Extinction*). Furthermore, intentional lowering of the quality is a very common technique among designers (e. g., "grunge" typefaces) and artists (e. g., "low bit" aesthetic in music). It is unclear if the industry will eventually enforce completely grain-free S3D production in the future, however, we predict that film grain will continue to appear in films at least as a means of stylization (e. g., *Hugo*).

5.8 Summary

In this chapter we introduced a new technique of film grain application, based on distributing the grain in front of the objects in the scene. Our approach is especially practical in stylized content, where the visibility of the grain is usually quite high, and uncorrelated grain would put too much stress on the visual system of the observer. This is in line with the results of our preference study. The advantage of our technique over the projection approach is twofold: it does not emphasize artifacts of the reconstructed depth map (too flat or erroneous depth) and do not change the appearance of the objects in the scene. Moreover, it can handle well the cases when the depth is not strictly defined in the scene (sky, smoke, out of focus areas, etc.).

Chapter 6

Conclusion and Future Work

Despite its relatively long development history, stereoscopic imaging technology still suffers from major shortcomings preventing its wider adaptation. In particular, many people are skeptical about the idea of stereoscopic cinema and invariably choose screenings in the traditional, non-stereoscopic format. In this dissertation, we analyzed and proposed novel solutions to three important problems of stereoscopic film production in the hope that our efforts will contribute in making stereoscopic screenings a more enjoyable experience.

First, we addressed the issue of sudden temporal depth changes, such as those introduced by video edits. We modeled eye vergence response to unexpected, step-like disparity changes, and described how our model can be used for film cut visualization and optimization. Our model provides additional insight into the dynamics of the observer's eye vergence movements, and thus is likely to help in the process of film editing and facilitate building fast-paced film narrations. Using a protocol based on the subjective assessment of fusion instead of eye tracking, Mu et al. [2015] independently derived a model predicting fusion times after disparity changes, and, similarly to us, found that fusion times do not rely solely on the step magnitude. An interesting avenue for future research would be to analyze how shorter adaptation times relate to the overall comfort of the observer. From our measurements it follows that decreasing adaptation time often comes at the price of increased distance of the stimulus from the display plane, which in turn increases the cue conflict (e. g., the accommodation-vergence conflict) induced by the stereoscopic image. In this context, it should be investigated if (and to what extent) the benefits of decreased "cognitive load" outweigh the drawbacks of increased "physiological load".

Second, we presented a new approach to stereoscopic rendering of materials with view-dependent shading that can contribute to the visual discomfort experienced by the observers. In the follow-up work, this technique has been generalized to allow for the rendering of scenes with semi-transparent, refractive and/or reflective objects [Dąbała et al. 2014]. Exploring interactions of our method with distribution effects, such as depth-of-field and motion blur, remains

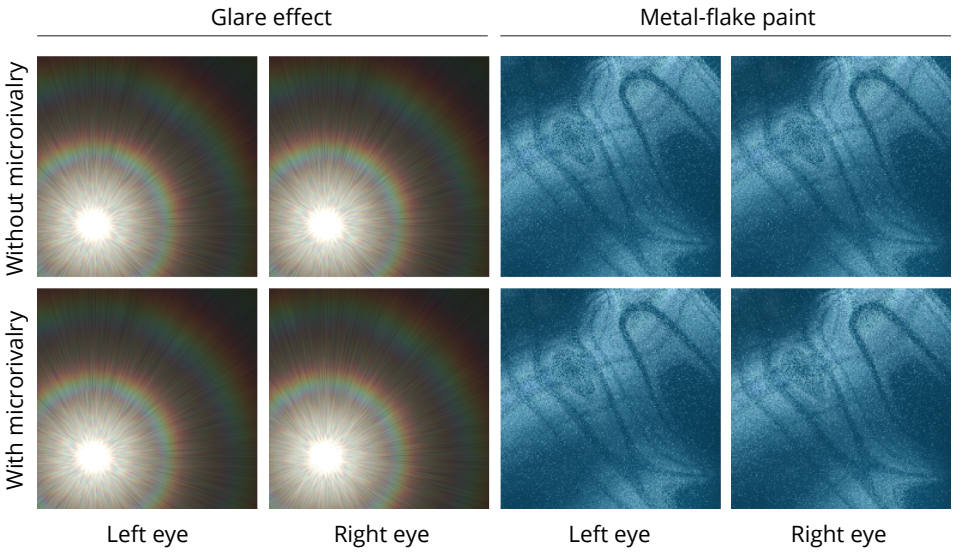
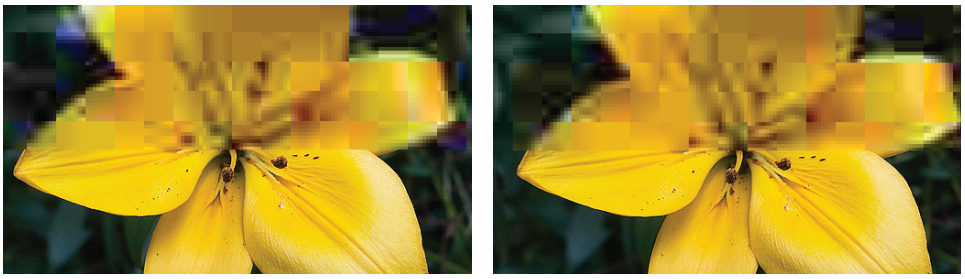


Figure 6.1. The glare effect *(left)* appears independently in each eye, and the sparkle patterns in metal-flake paints *(right)* are view-dependent. Versions including these subtle binocular conflicts *(bottom)* look more realistic compared to ordinary images *(top)*.

a future direction of work. Other possibilities, besides accounting for the interplay with other depth cues, include modification of highlights to achieve a certain material appearance and manipulation (exaggeration) of specular disparity to amplify the depth impression or obtain a stylized depiction of materials. The proposed method of introducing small highlight disparities can be seen as an instance of a more general class of techniques which we term *microrivalry*. There are a number of phenomena that lead to subtle differences in binocular images, such as thin-film interference, metal-flake paints, or glare effects [Ďurikovič and Martens 2003, Ritschel et al. 2009]. Taking them into account during rendering stereoscopic images can increase their realism (see Fig. 6.1). The role of stereoscopy in the perception of metal-flake paints has been recently investigated by da Graça et al. [2014] as part of a larger project with the aim of developing a series of mathematical models describing the interaction of various materials with light. It has been also proposed to use subtle inter-view differences in order to introduce certain "visual richness" to stereoscopic tone-mapped HDR images [Yang et al. 2012]. Chapiro et al. [2015] recently showed how to manipulate surface normals in order to enhance the depth impression in flat images or stereoscopic 3D images with shallow depth.

Third, we proposed a new way of handling film grain in stereoscopic videos, which finds a good balance between artistic and practical constraints. In our work we considered only film grain, however, our approach could be extended to handle other forms of visual noise. We propose to apply similar methods to 3D images and videos where JPEG/MPEG compression artifacts are clearly visible. As



Left eye Right eye

Figure 6.2. Upper part of the stereoscopic picture has been downsampled and compressed using low-quality settings of the JPEG format. Because the left and right channels have been encoded independently, it is very problematic to fuse them. Orignal photograph: JJ Harrison, CC-BY-SA.

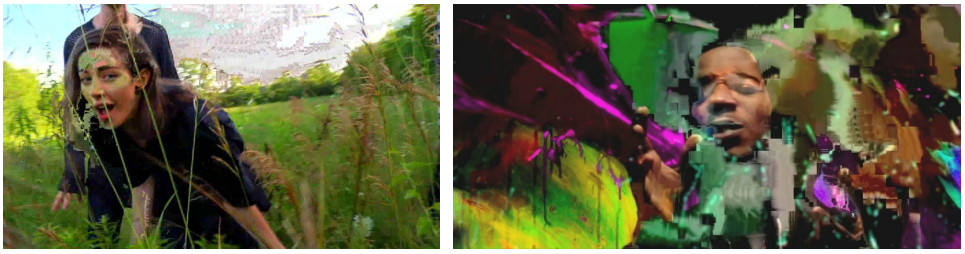


Figure 6.3. Stills from 2009 music videos for Chairlift's *Evident Utensil* by Ray Tintori and Bob Weisz *(left)* and Kanye West's *Welcome to Heartbreak* directed by Nabil Elderkin *(right)*. Both videos use coding errors as a means of stylization. To our knowledge, the possibilities of using such techniques in stereoscopic 3D have not been explored yet. Pictures: Dead Video / Live Video Festival (vimeo.com/4641545), nabil elderkin (vimeo.com/4578366)

previously, the basic idea is that the medium should be separated from the scene, and thus we do not want the artifacts to be visible on the surface of the objects. Independent processing of the left and right channels accomplishes this goal, but only partially. If the level of compression is considerable, the observer may find it hard to fuse the stereo image pair (see Fig. 6.2). Analogously, the JPEG/MPEG artifacts should be placed somewhere between the objects and the spectator. This way the scene will look "submerged" in the medium rather than simply "cut out" of it. One can question the need of such techniques: since network bandwidths and the processing power of computers are constantly increasing, such artifacts are becoming less of a problem, at least for desktop viewing conditions. However, 3D-capable hand-held devices are gaining popularity, and we foresee that compression artifacts, at least for some time, can still be an issue in this segment of less powerful devices. Additionally, compression artifacts can be introduced on purpose, as a means of stylization or artistic expression. Unlikely as it sounds, the aesthetic value of JPEG artifacts has already been acknowledged [Ruff and Simpson 2009]. A related idea of intentionally introducing coding errors

(or "glitches") to a data stream is a well-established practice in the visual arts [Menkman 2011]. Examples of videos stylized using coding errors are shown in Fig. 6.3. To our knowledge, the possibility of applying such a stylization in the context of stereoscopic graphics has not been explored yet.

In summuary, our work covers various stages of the process of stereoscopic film-making – from the spatial arrangement of the scene, to the objects' shading, to the final stylization of the entire video sequence. Since in the past few years stereoscopic film making has gained the attention of industry on an unprecedented scale, our research is particularly timely. In addition to the possible avenues for future work outlined above, a whole new range of research opportunities is becoming possible with the emergence of the new trend of capturing and presenting films at frame rates higher than the standard rate of twenty-four frames per second [Quesnel et al. 2013]. The influence of the increased visual clarity of high-frame-rate video on stereoscopic perception appears to be one of the more important lines of future research for stereoscopic cinema.

Bibliography – Own Work

Calagari, K., Templin, K., Elgamal, T., Diab, K., Didyk, P., Matusik, W., and Hefeeda, M. M. 2014. Anahita: A system for 3D video streaming with depth customization. In *Proc. ACM International Conference on Multimedia*, pages 337–346, Florida, USA.

Dąbała, Ł., Kellnhofer, P., Ritschel, T., Didyk, P., Templin, K., Myszkowski, K., Rokita, P., and Seidel, H.-P. 2014. Manipulating refractive and reflective binocular disparity. *Computer Graphics Forum (Proc. Eurographics)*, 33(2), 53–62.

Ritschel, T., Templin, K., Myszkowski, K., and Seidel, H.-P. 2012. Virtual passepartouts. In *Proc. Symposium on Non-Photorealistic Animation and Rendering (NPAR)*, pages 57–63, France.

Templin, K. 2016. Stereo and anaglyph images. In *Encyclopedia of Color Science and Technology*, chapter Computer Graphics. Springer. To appear.

Templin, K., Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., and Seidel, H.-P. 2011. Apparent resolution enhancement for animations. In *Proc. Spring Conference on Computer Graphics (SCCG)*, pages 85–92, Slovakia.

Templin, K., Didyk, P., Ritschel, T., Myszkowski, K., and Seidel, H.-P. 2012. Highlight microdisparity for improved gloss depiction. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 31(4), 92:1–5.

Templin, K., Didyk, P., Myszkowski, K., Hefeeda, M. M., Seidel, H.-P., and Matusik, W. 2014a. Modeling and optimizing eye vergence response to stereoscopic cuts. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 33(4), 145:1–8.

Templin, K., Didyk, P., Myszkowski, K., and Seidel, H.-P. 2014b. Perceptually-motivated stereoscopic film grain. *Computer Graphics Forum (Proc. Pacific Graphics)*, 33(7), 349–358.

Bibliography

Acuna, K. 2013. 3 signs that 3D movies are the way of the future. *Business Insider*. http://www.businessinsider.com/3d-movies-have-a-future-in-hollywood-2013-1. Accessed 06.05.2015.

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

Akerstrom, R. A. and Todd, J. T. 1988. The perception of stereoscopic transparency. *Attention, Perception, and Psychophysics*, 44(5), 421–432.

Allaby, M. 2013. *A Dictionary of Geology and Earth Sciences*. Oxford University Press.

Alvarez, T. L., Semmlow, J. L., and Pedrono, C. 2005. Divergence eye movements are dependent on initial stimulus position. *Vision Research*, 45(14), 1847–1855.

Atkinson, S. 2011. Stereoscopic-3D storytelling – rethinking the conventions, grammar and aesthetics of a new medium. *Journal of Media Practice*, 12(2), 139–156.

Barnes, M. 2012. Ray Zone, the '3D King of Hollywood,' Dies at 65. *The Hollywood Reporter*. www.hollywoodreporter.com/news/ray-zone-3d-king-hollywood-batman-391266. Accessed 04.05.2015.

Barry, S. and Sacks, O. 2010. *Fixing My Gaze: A Scientist's Journey Into Seeing in Three Dimensions*. Basic Books.

Bernhard, M., Dellmour, C., Hecher, M., Stavrakis, E., and Wimmer, M. 2014. The effects of fast disparity adjustments in gaze-controlled stereoscopic applications. In *Proc. Symposium on Eye Tracking Research & Applications (ETRA)*, pages 111–118.

Beucher, S. and Lantuejoul, C. 1979. Use of watersheds in contour detection. In *Int. Workshop on Image Processing, Real-time Edge and Motion Detection*, pages 2:1–12.

Biegon, G. 2005. Stereoscopic synergy: Twin-relief sculpture and painting. *Leonardo*, 38(2), 92–100.

Blake, A. 1985. Specular stereo. In *Proc. International Joint Conference on Artificial Intelligence*, volume 2, pages 973–976.

Blake, A. and Brelstaff, G. 1988. Geometry from specularities. In *Proc. International Conference on Computer Vision*, pages 394–403.

Blake, A. and Bülthoff, H. 1990. Does the brain know the physics of specular reflection? *Nature*, 343(6254), 165–168.

Brewster, D. 1861. On binocular lustre. *Reports of British Association*, 2, 29–31.

Brewster, D. 1856. *The Stereoscope, its History, Theory, and Construction, with its Applications to the Fine and Useful Arts and to Education*. John Murray, London.

Carmi, R. and Itti, L. 2006. Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46(26), 4333–4345.

Chadwick, A. and Kentridge, R. 2015. The perception of gloss: A review. *Vision Research*, 109, Part B, 221–235.

Chapiro, A., O'Sullivan, C., Jarosz, W., Gross, M., and Smolic, A. 2015. Stereo from shading. In *Proc. EGSR (Experimental Ideas & Implementations)*, pages 119–125.

Cho, S.-H. and Kang, H.-B. 2012. Subjective evaluation of visual discomfort caused from stereoscopic 3D video using perceptual importance map. In *TENCON 2012 – 2012 IEEE Region 10 Conference*, pages 1–6.

Cumming, B. and DeAngelis, G. 2001. The physiology of stereopsis. *Annual Review of Neuroscience*, 24, 203–238.

Cunningham, D., Waterston, D., Neres, F., and Cryer, M. 1911. *The Edinburgh Stereoscopic Atlas of Anatomy*. Keystone View Co. & Imperial Publishing Co.

Cutting, J., Brunick, K., Delong, J., Iricinschi, C., and Candan, A. 2011. Quicker, faster, darker: Changes in hollywood film over 75 years. *i-PERCEPTION*, 2(6), 569–576.

Daly, S. and Feng, X. 2003. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Color Imaging VIII: Processing, Hardcopy, and Applications*, volume 5008 of *SPIE*, pages 455–466.

De Stefano, A., Collis, B., and White, P. 2006. Synthesising and reducing film grain. *Journal of Visual Communication and Image Representation*, 17(1), 163–182.

Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., Seidel, H.-P., and Matusik, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Transactions on Graphics*, 31(6), 184:1–10.

Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., and Seidel, H.-P. 2011. A perceptual model for disparity. *ACM Transactions on Graphics*, 30(4), 96:1–9.

Didyk, P., Sitthi-Amorn, P., Freeman, W. T., Durand, F., and Matusik, W. 2013. Joint view expansion and filtering for automultiscopic 3D displays. *ACM Transactions on Graphics*, 32(6), 221:1–8.

Dąbała, Ł., Kellnhofer, P., Ritschel, T., Didyk, P., Templin, K., Myszkowski, K., Rokita, P., and Seidel, H.-P. 2014. Manipulating refractive and reflective binocular disparity. *Computer Graphics Forum*, 33(2), 53–62.

Dove, H. 1850. Über die Ursachen des Glanzes und der Irradiation, abgeleitet aus chromatischen Versuchen mit dem Stereoskop. *Annalen der Physik*, 159 (5), 169–183.

Du, S.-P., Masia, B., Hu, S.-M., and Gutierrez, D. 2013. A metric of visual comfort for stereoscopic motion. *ACM Transactions on Graphics*, 32(6), 222:1–9.

Du, S.-P., Didyk, P., Durand, F., Hu, S.-M., and Matusik, W. 2014. Improving visual quality of view transitions in automultiscopic displays. *ACM Transactions on Graphics*, 33(6), 192:1–9.

Dubois, E. 2001. A projection method to generate anaglyph stereo images. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1661–1664.

Ďurikovič, R. and Martens, W. 2003. Simulation of sparkling and depth effect in paints. In *Proc. Spring Conference on Computer Graphics (SCCG)*, pages 207–213.

Eadie, A. S., Gray, L. S., Carlin, P., and Mon-Williams, M. 2000. Modelling adaptation effects in vergence and accommodation after exposure to a simulated virtual reality stimulus. *Ophthalmic and Physiological Optics*, 20(3), 242–251.

Erkelens, C. J., Van der Steen, J., Steinman, R. M., and Collewijn, H. 1989. Ocular vergence under natural conditions. II. Gaze-shifts between real targets differing in distance and direction. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 236(1285), 441–446.

Fairchild, M. D. and Johnson, G. M. 2005. On the salience of novel stimuli: Adaptation and image noise. In *IS&T/SID 13th Color Imaging Conference*, pages 333–338.

Ferragallo, R. 1995. Stereoscopic architectural surfaces. *Stereo World*, 22(1), 20–21.

Finke, R. 1989. *Principles of Mental Imagery*. MIT Press.

Fleming, R. W., Torralba, A., and Adelson, E. H. 2004. Specular reflections and the perception of shape. *Journal of Vision*, 4(9), 798–820.

Gerig, J. 1974. *Introductory Organic Chemistry*. Elsevier.

Giant Bomb. 2012. Film grain. http://www.giantbomb.com/film-grain/92-487/. Accessed 13.06.2014.

Gombrich, E. H. 2000. *Art and illusion*. Princeton University Press.

Gomila, C., Llach, J., and Cooper, J. 2013. Film grain simulation method. US Patent 8,447,127.

Goutcher, R., O'Kane, L., and Wilcox, L. M. 2012. Representation of stereoscopic volumes. *Journal of Vision*, 12(9), 221.

da Graça, F., Paljic, A., Lafon-Pham, D., and Callet, P. 2014. Stereoscopy for visual simulation of materials of complex appearance. In *Proc. SPIE. Vol. 9011. Stereoscopic Displays and Applications XXV*, pages 90110X:1–12.

Hainich, R. R. and Bimber, O. 2011. *Displays: Fundamentals and Applications*. A K Peters/CRC Press.

Heinzle, S., Greisen, P., Gallup, D., Chen, C., Saner, D., Smolic, A., Burg, A., Matusik, W., and Gross, M. H. 2011. Computational stereo camera system with programmable control loop. *ACM Transactions on Graphics*, 30(4), 94:1–10.

Hess, R., Kingdom, F., and Ziegler, L. 1999. On the relationship between the spatial channels for luminance and disparity processing. *Vision Research*, 39 (3), 559–568.

Higgins, S. 2012. 3D in depth: 'Coraline', 'Hugo', and a sustainable aesthetic. *Film History: An International Journal*, 24(2), 196–209.

Hoffman, D., Girshick, A., Akeley, K., and Banks, M. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3), 1–30.

Horwood, A. M. and Riddell, P. M. 2008. The use of cues to convergence and accommodation in naïve, uninstructed participants. *Vision Research*, 48(15), 1613–1624.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C., and Gelautz, M. 2013. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(2), 504–511.

Howard, I. P. and Rogers, B. J. 2002. *Seeing in Depth*, volume 2: Depth Perception. I. Porteous.

Hung, G. K. 1992. Adaptation model of accommodation and vergence. *Ophthalmic and Physiological Optics*, 12(3), 319–326.

Hung, G. K. 1998. Dynamic model of the vergence eye movement system: Simulations using Matlab/Simulink. *Computer Methods and Programs in Biomedicine*, 55(1), 59–68.

Hung, G. K. and Semmlow, J. L. 1980. Static behavior of accommodation and vergence: Computer simulation of an interactive dual-feedback system. *IEEE Transactions on Biomedical Engineering*, BME-27(8), 439–447.

Hung, G. K., Semmlow, J. L., and Ciuffreda, K. J. 1986. A dual-mode dynamic model of the vergence eye movement system. *IEEE Transactions on Biomedical Engineering*, BME-33(11), 1021–1028.

Hung, G. K., Zhu, H., and Ciuffreda, K. J. 1997. Convergence and divergence exhibit different response characteristics to symmetric stimuli. *Vision Research*, 37(9), 1197–1205.

Hurlbert, A., Cumming, B., and Parker, A. 1991. Recognition and perceptual use of specular reflections. *Investigative Ophthalmology & Visual Science*, 32(4), Supplement.

Johnson, G. M. and Fairchild, M. D. 2000. Sharpness rules. In *IS&T/SID 8th Color Imaging Conference*, pages 24–30.

Jorke, H. and Fritz, M. 2006. Stereo projection using interference filters. In *Proc. SPIE. Vol. 6055. Stereoscopic Displays and Virtual Reality Systems XIII*, pages 60550G:1–8.

Julesz, B. 1964. Binocular depth perception without familiarity cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis. *Science*, 145(3630), 356–362.

Julesz, B. 1971. *Foundations of Cyclopean Perception*. The University of Chicago Press.

Jung, Y. J., Lee, S.-i., Sohn, H., Park, H. W., and Ro, Y. M. 2012. Visual comfort assessment metric based on salient object motion information in stereoscopic video. *Journal of Electronic Imaging*, 21(1), 011008:1–16.

Kerrigan, I. S. and Adams, W. J. 2013. Highlights, disparity, and perceived gloss with convex and concave surfaces. *Journal of Vision*, 13(1), 9:1–10.

Kim, Y., Lee, Y., Kang, H., and Lee, S. 2013. Stereoscopic 3D line drawing. *ACM Transactions on Graphics*, 32(4), 57:1–13.

Kirschmann, A. 1895. Der Metallglanz und die Parallaxe des indirecten Sehens. *Philosophische Studien*, 11, 147–189.

Koppal, S. J., Zitnick, C. L., Cohen, M., Kang, S. B., Ressler, B., and Colburn, A. 2011. A viewer-centric editor for 3d movies. *IEEE Computer Graphics and Applications*, 31(1), 20–35.

Krishnan, V., Farazian, F., and Stark, L. 1973. An analysis of latencies and prediction in the fusional vergence system. *American Journal of Optometry and Archives of American Academy of Optometry*, 50(12), 933–939.

Krishnan, V., Farazian, F., and Stark, L. 1977. Dynamic measures of vergence accommodation. *American Journal of Optometrics and Physiological Optics*, 54, 470–473.

Kurihara, T., Manabe, Y., Aoki, N., and Kobayashi, H. 2008. Digital image improvement by adding noise: An example by a professional photographer. In *Image Quality and System Performance V*, volume 6808 of *SPIE*, pages 1–10.

Kurihara, T., Aoki, N., and Kobayashi, H. 2009. Analysis of sharpness increase by image noise. In *Human Vision and electronic Imaging XIV*, volume 7240 of *SPIE*, pages 724014:1–9.

Lagae, A., Lefebvre, S., Cook, R., DeRose, T., Drettakis, G., Ebert, D. S., Lewis, J. P., Perlin, K., and Zwicker, M. 2010. State of the art in procedural noise functions. In *EG 2010 - State of the Art Reports*.

Lambooij, M., IJsselsteijn, W., and Heynderickx, I. 2011. Visual discomfort of 3D TV: Assessment methods and modeling. *Displays*, 32(4), 209–218.

Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3), 030201:1–14.

Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., and Gross, M. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Transactions on Graphics*, 29(4), 75:1–10.

Lankheet, M. J. and Lennie, P. 1996. Spatio-temporal requirements for binocular correlation in stereopsis. *Vision Research*, 36(4), 527–538.

Le Conte Stevens, W. 1882. The stereoscope: Its history. *Popular Science Monthly*, 21, 37–53.

Lee, Y., Kim, Y., Kang, H., and Lee, S. 2013. Binocular depth perception of stereoscopic 3D line drawings. In *Proc. ACM Symposium on Applied Perception*, pages 31–34.

Li, J., Barkowsky, M., and Le Callet, P. 2014. Visual discomfort of stereoscopic 3D videos: Influence of 3D motion. *Displays*, 35(1), 49–57.

Lipton, L. 1982. *Foundations of the Stereoscopic Cinema*. Van Nostrand Reinhold.

Lipton, L. 2010. How 3D works. In Okun, J. A. and Zwerman, S., editors, *The VES Handbook of Visual Effects*, chapter Stereoscopic 3D, pages 387–396. Focal Press.

Liu, C., Yuen, J., and Torralba, A. 2011. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 978–994.

Masaoka, K., Hanazato, A., Emoto, M., Yamanoue, H., Nojiri, Y., and Okano, F. 2006. Spatial distortion prediction system for stereoscopic images. *Journal of Electronic Imaging*, 15(1), 013002:1–12.

Masia, B., Wetzstein, G., Didyk, P., and Gutierrez, D. 2013. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computer & Graphics*, 37, 1012–1038.

Meesters, L., IJsselsteijn, W., and Seuntiens, P. 2004. A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3), 381–391.

Mendiburu, B. 2009. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press.

Menkman, R. 2011. *The Glitch Moment(um)*. Institute of Network Cultures.

Mital, P., Smith, T., Hill, R., and Henderson, J. 2011. Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1), 5–24.

Mu, T.-J., Sun, J.-J., Martin, R., and Hu, S.-M. 2015. A response time model for abrupt changes in binocular disparity. *The Visual Computer*, 31(5), 675–687.

Muryy, A. A., Fleming, R. W., and Welchman, A. E. 2014. Key characteristics of specular stereo. *Journal of Vision*, 14(14), 14:1–26.

Neuman, R. 2009. Concurrent monoscopic and stereoscopic animated film production. In *ACM SIGGRAPH 2009 Talks*, page 38.

Northam, L., Asente, P., and Kaplan, C. S. 2012. Consistent stylization and painterly rendering of stereoscopic 3D images. In *Proc. NPAR*, pages 47–56.

Obein, G., Knoblauch, K., and Vienot, F. 2004. Difference scaling of gloss: nonlinearity, binocularity, and constancy. *Journal of Vision*, 4(9), 711–720.

Okuyama, F. 1998. Human visual accommodation and vergence eye movement while viewing stereoscopic display and actual target. In *Proc. 20th Annual International Conference of the IEEE*, volume 2, pages 552–555.

Oskam, T., Hornung, A., Bowles, H., Mitchell, K., and Gross, M. H. 2011. OSCAM – optimized stereoscopic camera control for interactive 3D. *ACM Transactions on Graphics*, 30(6), 189:1–8.

Owens, C. 2013. Invited talk. 2nd Toronto International Stereoscopic 3D Conference.

Paille, D., Monot, A., Dumont-Becle, P., and Kemeny, A. 2001. Luminance binocular disparity for 3D surface simulation. In *Proc. SPIE*, volume 4299, page 622.

Pellacini, F., Ferwerda, J., and Greenberg, D. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proc. SIGGRAPH*, pages 55–64.

Phillips, S. 2010. Stereoscopic design. In Okun, J. A. and Zwerman, S., editors, *The VES Handbook of Visual Effects*, chapter Stereoscopic 3D, pages 396–405. Focal Press.

Quesnel, D., Lantin, M., Goldman, A., and Arden, S. 2013. An exploration into the creation of variable frame rate (VFR) stereoscopic 3D narrative productions. Emily Carr University of Art and Design. http://research.ecuad.ca/s3dcentre/projects/hfr-high-frame-rate-research/. Accessed 27.04.2015.

Ridanovic, I. 2011. Interview with stereographer Daniele Siragusano. http://www.hdhead.com/?p=279. Accessed 13.06.2014.

Ritschel, T., Ihrke, M., Frisvad, J. R., Coppens, J., Myszkowski, K., and Seidel, H.-P. 2009. Temporal glare: Real-time dynamic simulation of the scattering in the human eye. *Computer Graphics Forum*, 28(2), 183–192.

Robertson, B. 2009. Monsters of the deep. *Computer Graphics World*, 32(3).

Rollmann, W. 1853. Zwei neue stereoskopische Methoden. *Annalen der Physik*, 166, 186–187.

Ruff, T. and Simpson, B. 2009. *JPEGS*. Aperture.

Rushton, S. K. and Riddell, P. M. 1999. Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. *Applied Ergonomics*, 30(1), 69–78.

Sakano, Y. and Ando, H. 2010. Effects of head motion and stereo viewing on perceived glossiness. *Journal of Vision*, 10(9), 15:1–14.

Sandrew, B. B. 2012. 3D movies and the primal brain. *Innovation in Advanced Digital Imaging*. http://bsandrew.blogspot.com. Accessed 19.02.2015.

Sanftmann, H. and Weiskopf, D. 2011. Anaglyph stereo without ghosting. *Computer Graphics Forum*, 30, 1251–1259.

Schor, C. M. 1979. The relationship between fusional vergence eye movements and fixation disparity. *Vision Research*, 19(12), 1359–1367.

Schor, C. M. 1992. A dynamic model of cross-coupling between accommodation and convergence: Simulations of step and frequency responses. *Optometry and Vision Science*, 69(4), 258–269.

Seckel, A. 2004. *Masters of Deception: Escher, Dalí & the Artists of Optical Illusion*. Sterling Publishing Company.

Semmlow, J. and Wetzel, P. 1979. Dynamic contributions of the components of binocular vergence. *Journal of the Optical Society of America*, 69(5), 639–645.

Semmlow, J., Hung, G., and Ciuffreda, K. 1986. Quantitative assessment of disparity vergence components. *Investigative Ophthalmology and Visual Science*, 27(4), 558–564.

Seymour, M. 2008. Art of digital 3D stereoscopic film. *fxguide*. http://www.fxguide.com/featured/art_of_digital_3d_stereoscopic_film/. Accessed 06.05.2015.

Seymour, M. 2011a. Case study: How to make a Captain America wimp. *fxguide*. http://www.fxguide.com/featured/case-study-how-to-make-a-captain-america-wimp/. Accessed 21.08.2015.

Seymour, M. 2011b. Hugo: A study of modern inventive visual effects. *fxguide*. http://www.fxguide.com/featured/hugo-a-study-of-modern-inventive-visual-effects/. Accessed 21.08.2015.

Seymour, M. 2012. Art of stereo conversion: 2D to 3D – 2012. *fxguide*. http://www.fxguide.com/featured/art-of-stereo-conversion-2d-to-3d-2012/. Accessed 21.08.2015.

Shibata, T., Kim, J., Hoffman, D. M., and Banks, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8), 11:1–29.

Smolic, A., Kauff, P., Knorr, S., Hornung, A., Kunter, M., Müller, M., and Lang, M. 2011. Three-dimensional video postproduction and processing. *Proceedings of the IEEE*, 99(4), 607–625.

Sousa, T., Kasyan, N., and Schulz, N. 2012. CryENGINE. In Engel, W., editor, *GPU Pro 3*. CRC Press.

Stafford, B., Terpak, F., and Poggi, I. 2001. *Devices of Wonder: From the World in a Box to Images on a Screen*. Getty Research Institute.

Stavrakis, E. and Gelautz, M. 2004. Image-based stereoscopic painterly rendering. In *Proc. EGSR*, pages 53–60.

Stavrakis, E. and Gelautz, M. 2005. Stereoscopic painting with varying levels of detail. In *Stereoscopic Displays and Virtual Reality Systems XII*, volume 5664 of *SPIE*, pages 450–459.

Stephenson, I. and Saunders, A. 2007. Simulating film grain using the noise-power spectrum. In *Proc. EG UK Theory and Practice of Computer Graphics*, pages 69–72.

Stevenson, S. B., Cormack, L. K., and Schor, C. M. 1991. Depth attraction and repulsion in random dot stereograms. *Vision Research*, 31(5), 805–813.

Tachi, S. 2013. From 3D to VR and further to telexistence. In *Proc. 23rd International Conference on Artificial Reality and Telexistence (ICAT)*, pages 1–10.

Tam, W. J., Speranza, F., Vázquez, C., Renaud, R., and Hur, N. 2012. Visual comfort: stereoscopic objects moving in the horizontal and mid-sagittal planes. In *Proc. SPIE*, page 8288:13.

Tan, R. and Ikeuchi, K. 2005. Separating reflection components of textured surfaces using a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2), 178–193.

Tsirlin, I., Allison, R. S., and Wilcox, L. M. 2008. Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, 8(5), 5:1–10.

Tsirlin, I., Wilcox, L. M., and Allison, R. S. 2010. Perceptual artifacts in random-dot stereograms. *Perception*, 39(3), 349–355.

Tyler, C. W. 1975. Spatial organization of binocular disparity sensitivity. *Vision Research*, 15(5), 583–590.

Ukai, K. and Kato, Y. 2002. The use of video refraction to measure the dynamic properties of the near triad in observers of a 3-D display. *Ophthalmic and Physiological Optics*, 22(5), 385–388.

Vergne, R., Pacanowski, R., Barla, P., Granier, X., and Schlick, C. 2009. Light warping for enhanced surface depiction. *ACM Transactions on Graphics*, 28(3), 25:1–8.

Vienne, C., Sorin, L., Blondé, L., Huynh-Thu, Q., and Mamassian, P. 2014. Effect of the accommodation-vergence conflict on vergence eye movements. *Vision Research*, 100, 124–133.

Wang, C. and Sawchuk, A. A. 2008. Disparity manipulation for stereo images and video. In *Proc. SPIE Vol. 6803. Stereoscopic Displays and Applications XIX*, pages 68031E:1–12.

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., and Heeger, D. J. 2012. Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, 12(1), 16:1–27.

Wang, N., Paljic, A., and Fuchs, P. 2010. A study of perception of volumetric rendering for immersive scientific visualization. In *20th International Conference on Artificial Reality and Teleexistence (ICAT'2010)*, pages 145–152.

Watson, A. B. and Pelli, D. G. 1983. QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics*, 33(2), 113–120.

Wendt, G., Faul, F., and Mausfeld, R. 2008. Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision*, 8(1), 14:1–10.

Wendt, G., Faul, F., Ekroll, V., and Mausfeld, R. 2010. Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision*, 10(9), 7:1–17.

Wheatstone, C. 1838. Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society*, 128, 371–395.

Whitted, T. 1980. An improved illumination model for shaded display. *Communications of the ACM*, 23(6), 343–349.

Wills, J., Agarwal, S., Kriegman, D., and Belongie, S. 2009. Toward a perceptual space for gloss. *ACM Transactions on Graphics*, 28(4), 103:1–15.

Winter, M. and Gandolph, D. 2013. Film grain for stereoscopic or multi-view images. US Patent App. 13/762,479.

Woods, A., Docherty, T., and Koch, R. 1993. Image distortions in stereoscopic video systems. In *Proc. SPIE 1915. Stereoscopic Displays and Applications IV*, pages 36–48.

Woods, A. J. 2012. Crosstalk in stereoscopic displays: a review. *Journal of Electronic Imaging*, 21(4), 040902:1–21.

Yang, X., Zhang, L., Wong, T.-T., and Heng, P.-A. 2012. Binocular tone mapping. *ACM Transactions on Graphics*, 31(4), 93:1–10.

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

Zilly, F., Kluger, J., and Kauff, P. 2011. Production rules for stereo acquisition. *Proceedings of the IEEE*, 99(4), 590–606.

Zone, R. 2014. *Stereoscopic Cinema and the Origins of 3-D Films, 1838–1952*. University Press of Kentucky.

Zwicker, M., Matusik, W., Durand, F., Pfister, H., and Forlines, C. 2006. Antialiasing for automultiscopic 3D displays. In *Proc. EGSR*, pages 73–82.