Perception-inspired Tone Mapping

Dissertation

zur Erlangung des Grades des
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Naturwissenschaftlich-Technischen Fakultäten
der Universität des Saarlandes

vorgelegt von

Grzegorz Marek Krawczyk

Eingereicht am 9. Oktober 2007 in Saarbrücken

Grzegorz Marek Krawczyk
Max-Planck-Institut für Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken, Germany
krawczyk@mpi-inf.mpg.de

Betreuender Hochschullehrer – Supervisor:
Dr.-Ing. habil. Karol Myszkowski, MPI für Informatik, Saarbrücken, Germany

Gutachter – Reviewers:
Dr.-Ing. habil. Karol Myszkowski, MPI für Informatik, Saarbrücken, Germany
Prof. Dr. Hans-Peter Seidel, MPI für Informatik, Saarbrücken, Germany
Dr. Erik Reinhard, University of Bristol, United Kingdom

Dekan – Dean:
Prof. Dr. Thorsten Herfet, Universität des Saarlandes, Saarbrücken, Germany

Datum des Kolloqiums – Date of Defence:
30. November 2007

Vorsitzender – Chairman:
Prof. Dr. Joachim Weickert, Universität des Saarlandes, Saarbrücken, Germany

Akademischer Mitarbeiter – Academic Member of the Faculty:
Dr. Rafał Mantiuk, MPI für Informatik, Saarbrücken, Germany

Abstract

The display of high dynamic range images and video requires a tone mapping algorithm to depict their original appearance on existing display devices whose capabilities in terms of dynamic range are insufficient. The insightful application of knowledge about human visual system can assure high fidelity of depiction in such an algorithm.

In this thesis, we design new tone mapping models and improve existing algorithms by an informed use of human perception to provide a high fidelity depiction of high dynamic range. We develop a real-time tone mapping solution which reproduces the subjective appearance of dynamic HDR contents by accounting for perceptual effects that significantly contribute to the appearance of natural scenes. We design a computational model of lightness perception that can be applied to high quality tone mapping for static images to reproduce their original HDR appearance in terms of lightness. We identify common distortions typical to tone mapping which may hinder the comprehension of image contents, we design appropriate metrics to measure the perceived magnitude of these distortions and evaluate existing tone mapping algorithms accordingly. To compensate for observed distortions, we introduce a method which improves the tone mapping results beyond numerically optimized solution by using techniques strongly based on perception of contrasts. Presented solutions can be efficiently integrated in varied HDR applications including photography, playback of HDR video, image synthesis, light simulations, predictive rendering, and computer games.

Kurzfassung

Die Anzeige von Bildern und Videos mit hohem Kontrastumfang (HDR) erfordert einen Algorithmus für die Tonabbildung, um ihr ursprüngliches Aussehen auf vorhandenen Bildschirmen, deren Fähigkeiten in Kontrastumfang unzureichend sind, darzustellen. Die aufschlussreiche Anwendung des Wissens über das menschliche visuelle System kann die Wiedergabetreue eines solchen Algorithmus gewährleisten.

In dieser Doktorarbeit entwerfen wir neue Modelle für die Tonabbildung und verbessern vorhandene Algorithmen durch eine informative Anwendung von menschlicher Wahrnehmung um die Wiedergabetreue der HDR zu gewährleisten. Wir entwickeln eine Echtzeit-Tonabbildung Lösung, die das subjektive Aussehen von dynamischem HDR Inhalt reproduziert dadurch dass die Wahrnehmungseffekte, die erheblich zum Aussehen der natürlichen Szenen beitragen, berücksichtigt werden. Wir entwerfen ein Computermodell der menschliches Helligkeitsvorstellung welches wir in eine Tonabbildung anwenden, um damit das ursprüngliche HDR Aussehen von statischen Bildern in hoher Qualität zu reproduzieren. Weiterhin identifizieren wir die Verzerrungen, die bei Tonabbildungen typisch sind und das Verständnis des Bildinhalts hindern könnten. Wir entwerfen passende Metriken, um die wahrgenommenen Größe dieser Verzerrungen zu messen und vorhandene Algorithmen dementsprechend zu bewerten. Zur Kompensierung der Verzerrungen führen wir eine Methode vor, die das Tonabbildungsergebnis basierend auf eine Kontrastwahrnehmung über die numerisch optimierte Lösung hinaus verbessert. Die vorgestellten Lösungen können in vielseitigen HDR Anwendungen einschließlich Fotographie, Wiedergabe von HDR Videos, Bildsynthese, Globale Beleuchtung, und Computerspielen effizient integriert werden.

Summary

The high dynamic range (HDR) techniques overcome the legacy constraint of limited contrast and tonal range in digital images and video which are now adequate to accommodate the complete information about light in nature. The display of HDR contents, however, requires a tone mapping algorithm to depict their original appearance on existing display devices whose capabilities in terms of dynamic range are insufficient. Unfortunately, the tone mapping process inherently decreases the original quality of HDR contents. By taking the interdisciplinary approach in which we combine computer graphics and image processing with the knowledge of human visual perception, we design new tone mapping models, evaluate existing algorithms and improve their results to provide a high fidelity depiction of HDR appearance.

The subjective appearance of natural scenes is highly influenced by the perceptual effects caused in the early stages of human vision. These effects are stimulated by absolute luminance levels and are not present when observing standard displays. To account for this, we develop computational models that predict and simulate these perceptual effects and we embed their appearance in the tone mapped visual contents such that the depicted scenes are perceived by the human vision in the same way as in the natural conditions. We efficiently combine these models with a high quality tone mapping and achieve real-time performance.

The key perceptual dimension of image appearance related to tonal range is lightness. Therefore the high fidelity depiction of HDR contents requires that lightness is well reproduced during the tone mapping. To address this, we design a computational model of the modern lightness perception theory and apply it to obtain high quality tone mapping for static images. A comparison with the existing methods demonstrate that our model leads to a more accurate reproduction of appearance of HDR scenes.

The reduction of tonal range during tone mapping inherently distorts contrasts of original HDR data and a too strong distortion impedes the comprehension of image contents. By simulating the human perception of contrasts we design objective metrics that can measure the perceived magnitude of such distortions and we evaluate existing tone mapping operators accordingly. Our evaluation provides perceptually meaningful information and facilitates the choice of an appropriate tone mapping algorithm.

Finally, to overcome the observed distortions we introduce a method which improves the tone mapping results beyond numerically optimized solution by using techniques inspired by contrast illusions. We automatically identify image features which require restoration and insert into an image the so called countershading profiles which robustly enhance the perceived magnitude of contrasts with a sparing use of tonal range. We further develop a visual detection model which assures that our enhancement are not perceived as objectionable artifacts. Our new image processing tool generalizes the well-known unsharp masking.

Overall, the methods presented in this dissertation successfully improve and evaluate the fidelity of tone mapping by an insightful use of knowledge about human visual perception. Presented solutions can be efficiently integrated in varied HDR applications including photography, playback of HDR video, image synthesis, lighting simulation, predictive rendering, and computer games.

Zusammenfassung

Methoden für hohen Kontrastumfang (HDR) überwinden die Abwärtsbeschränkungen für Kontrast- und Tonumfang in Digitalbildern und Videos, die jetzt ausreichend sind um die kompletten Informationen über Licht in der Natur aufzunehmen. Die Anzeige des HDR Inhalts erfordert jedoch einen Algorithmus für Tonabbildung, um das ursprüngliche Aussehen auf vorhandenen Bildschirmen, deren Fähigkeiten im Kontrastumfang unzureichend sind, darzustellen. Leider verringert der Tonabbildungsprozess schon an sich die ursprängliche Qualität des HDR Inhalts. Um eine hohe Qualität der Wiedergabetreue der HDR zu gewährleisten, kombinieren wir Computergraphik und Bildverarbeitung mit dem Wissen der menschlichen Wahrnehmung. Damit entwerfen wir neue Tonabbildungsmodelle, bewerten vorhandene Algorithmen und verbessern die Ergebnisse von existierenden Algorithmen.

Wir entwickeln eine Echtzeit-Tonabbildungslösung, die das subjektive Aussehen des dynamischen HDR Inhalts unter Berücksichtigung der Wahrnehmungseffekte, die erheblich zum Aussehen der natürlichen Szenen beitragen, reproduziert. Das subjektive Erscheinungsbild der natürlichen Szenen wird stark durch die Effekte, die in den frühen Stadien des menschlichen Sehens verursacht werden, beeinflusst. Diese Wahrnehmungseffekte werden von absoluten Luminanzniveaus angeregt und sind bei der Beobachtung auf gewöhnlichen Bildschirmen nicht vorhanden. Um dieses zu berücksichtigen, entwickeln wir Berechnungsmodelle die diese Wahrnehmungseffekte voraussagen und simulieren. Wir lassen diese Effekte in die Tonabbildungsergebnisse einfließen, so dass die auf dem Bildschirm dargestellten Szenen genauso wie unter natürlichen Bedingungen wahrgenommen werden. Wir kombinieren diese Modelle mit einer hoch-qualitativen Tonabbildung und erzielen Echtzeitleistung.

Das wichtigste Wahrnehmungsmaß des Bildaussehens bezüglich Tonumfang ist die Helligkeit. Infolgedessen erfordert eine hohe Wiedergabetreue des HDR Inhalts eine gute Reproduktion der Helligkeitsvorstellung während der Tonabbildung. Dafür entwerfen wir ein Berechnungsmodell basierend auf der Theorie der Helligkeitswahrnehmung und wenden es für die Tonabbildung von statischen Bildern an. Ein Vergleich mit vorhandenen Methoden zeigt, dass unser Modell zu einer realistischeren Wiedergabe des Aussehens der HDR Szenen führt.

Die Tonumfangreduzierung während der Tonabbildung verzerrt Kontraste der ursprünglichen HDR Daten, und eine zu starke Verzerrung behindert das Verständnis des Bildinhalts. Daher entwerfen wir Metriken, die die wahrgenommene Größe dieser Verzerrungen messen indem sie die menschliche Wahrnehmung von Kontrasten simulieren. Dadurch sind wir in der Lage vorhandene Tonabbildungsalgorithmen entsprechend auszuwerten. Unsere Auswertung liefert wahrnehmungssinnvolle Informationen und erleichtert die Wahl einer passenden Tonabbildungsmethode.

Schließlich, zum abgleichen beobachteter Verzerrungen führen wir eine Methode vor, die das Tonabbildungsergebnis über die numerisch optimierte Lösung hinaus verbessert, indem sie eine starke optische Kontrasttäuschung ausnutzt. Wir identifizieren automatisch die Bildregionen welche eine Wiederherstellung erfordern und fügen so genannte *countershading* Profile ein. Diese Profile steigern robust die Wahrnehmung von Kontrasten und verbrauchen dabei sparsam den Kontrastumfang. Weiterhin entwickeln wir einen Erkennungsmodel das gewährleistet, dass unsere Kontrastwiederherstellungen als keine sichtbaren Artefakte wahrgenommen werden. Unser neues Bildverarbei-

tungswerkzeug generalisiert eine Standardmethode die als *unsharp masking* bekannt ist.

Die aufschlussreiche Anwendung des Wissens über das menschliche visuelle System in unseren Methoden, die in dieser Doktorarbeit dargestellt werden, erlaubt eine erfolgreiche Auswertung und Verbesserung der Wiedergabetreue des HDR Inhalts. Die dargestellten Lösungen können in vielseitigen HDR Anwendungen einschlielich Fotographie, Wiedergabe von HDR Videos, Bildsynthese, Globale Beleuchtung, und Computerspielen effizient integriert werden.

Acknowledgments

My first thanks go to my supervisor Dr.-Ing. habil. Karol Myszkowski for inspiring my interest in computer graphics and starting my career in science. Throughout my graduate studies Dr. Myszkowski provided me with his continuous support of my work, inspiration for new ideas, and gave me many possibilities to explore the opportunities in research. His contribution to my achievements is invaluable.

I would like to thank Prof. Dr. Hans-Peter Seidel for providing an excellent work environment at the Max-Planck Institute both on technological and research sides, and for his support for our projects in the novel field of High Dynamic Range imaging.

Dr. Erik Reinhard kindly agreed to serve as an external reviewer, which I am very grateful for. His pioneering work in the field of High Dynamic Range provided me with the inspiration for my first steps in this area.

Further, I would also like to thank Prof. Dr. Wolfgang Heidrich for hosting me at his group in the University of British Columbia in Vancouver and Helge Seetzen for inviting me to BrightSide Tech. also in Vancouver. During these visits I had a hands-on experience on prototypes of the HDR display and could evaluate the practical sides of my research. From the industry part, Prof. Dr. Bernd Höfflinger, Daniel Brosch (both at IMS-CHIPS), and Dr. Volker Gengenbach provided us with their HDR cameras and supported with their technical advises which let us generate the first HDR video contents for computer graphics purposes.

Many of the projects described in this dissertation would not have been feasible without the support and collaboration of my colleagues at the Max-Planck Institute. I would like to take the opportunity and name some of them here (in the alphabetical order): Thomas Annen, Kirill Dmitriev, Michael Goesele, Vlastimil Havran, Radosław Mantiuk, Rafał Mantiuk, Tina Scherbaum, Kaleigh Smith, Miłosław Smyk, Akiko Yoshida, and Dorota Zdrojewska. I am especially indebted to Lukas Ahrenberg, Robert Bargmann, Miguel Granados, and Andrei Linţu, who became close friends and made these four years of my life unforgettable.

The always pleasant assistance of Sabine Budde and Conny Liegl in administrative bureaucracy and with planning trips made my life so much easier in many situations and I thank for that very much. I also very often relied on the computer support from the side of Martin Fuchs and Carsten Stoll, and on the technical support of Michael Laise and Axel Köppel. I thank for their precious help.

Finally, I shall not forget the most important never-failing support I received. I thank my parents, Ewa and Marek Krawczyk, and my whole family who encouraged and supported me during my whole life. I am grateful to Eijenia Steganţova who became the closest friend during my graduate studies.

Contents

Chapter 1

Introduction

The recent advances in digital image processing known as High Dynamic Range (HDR) imaging bring a totally new visual experience to recording and displaying real-world equivalent images and video. The HDR techniques promise unconstrained capture of complete light information about scenes, high quality processing, and reproduction on various media with a high fidelity to the real-world appearance.

Historically, the 8-bit representation of visual data, which prevails in both digital capture and display devices, sets a hard limit on the range of tones that can be recorded, processed, and viewed. The choice of such representation seemed to be well motivated in 90s, when digital image processing proliferated, because it corresponded to the technical capabilities of devices at that time and offered comparable characteristics to analogue photography and video which could have served as a requirements' reference. Nowadays, practically all devices related to the main-stream digital image and video processing are manufactured according to the standards developed at that time despite significant technological advances in the field. While such a long-term standardization is advantageous in bridging varied technologies in video and media industry, this so called display-referred representation of visual contents significantly confines the visual experience of digital images and video compared to the real-world experience.

The HDR techniques abandon these legacy constraints and present a revolutionary approach to capturing, storing, processing and displaying of visual contents. Primarily, the quality of these processes is not any more driven by the capabilities of existing devices but is adequate to accommodate the complete information about light in nature. The intensity of tones in a usual natural scene may strongly vary between major image areas and very finely within details of materials and textures. The 8-bit representation of visual data often causes that very bright image areas are clipped to white, very dark ones to black, and subtle light changes are rounded due to quantization, thus in each of these cases a part of information about the original scene is lost. In contrast, the HDR representation imposes no limit on the tonal range and aims at no loss of fine details. Such a rich description of visual data permits an unprecedented visual experience of watching movies and photographs, playing computer games, or inspecting visualizations.

We are currently observing a rapid development of HDR technologies at all stages of

image and video processing pipeline. The HDR can now be captured both with the new types of imaging sensors and also using standard cameras and special software techniques. Recently, even off-the-shelf digital cameras are equipped with exposure bracketing feature which delivers HDR capture to amateur photographers and the required algorithms are implemented in most of image processing packages. Last years have also brought dedicated file formats and compression techniques for HDR with a notable example of OpenEXR which is now widely supported. However, the final stage of the pipe-line – the presentation is still in its most legacy form, despite the rapid growth of technical capabilities of displays including resolution, contrast range, and peak luminance levels. Even though the HDR displays present a significant step forwards and give an exciting foretaste of HDR experience, their current capabilities are not yet on a par with the real-world appearance. To bridge the gap between displays and the rest of the pipeline, prior to display the HDR contents need to be processed using the so called tone mapping algorithm to adjust their tonal range to the devices' capabilities.

This dissertation is dedicated to an in-depth analysis of the tone mapping problem. We approach the topic from an interdisciplinary point of view, because we observe that a successful design of a tone mapping algorithm needs to combine the knowledge of computer graphics and image processing with the substantial understanding of the human visual perception. While much research has been already done in the area, in this thesis we do not limit our interest to introducing yet another new algorithm. Rather, our aim is to select and apply the aspects of perception which should be considered in the context of displaying HDR content, to investigate and evaluate the perceptual quality of existing tone mapping solutions, and to seek further possibilities for quality improvement by exploiting knowledge of human visual system.

1.1   Problem Statement

The extensive range of tones available in high dynamic range images and video offers a high fidelity representation of natural scenes. Yet, the technical capabilities of existing display devices are insufficient to directly depict such rich visual contents. Therefore a tone mapping algorithm is required which prior to display reduces the tonal range of HDR data to match the devices' capabilities. While such a reduction inherently decreases the original quality of HDR contents, a successful tone mapping algorithm should strive to depict HDR images with high fidelity to their originals and at a minimal side effect on quality.

The high fidelity of depiction requires that the appearance of a tone mapped image matches closely the true appearance of the original HDR scene and that the ability of observes to comprehend image contents remains unaffected. A thorough understanding of human visual perception is necessary to both estimate the true appearance of HDR and to design an algorithm that reduces the tonal range while maintaining the appearance unchanged. In particular, one needs to identify which aspects of human visual perception have a significant contribution to the appearance and to build corresponding computational models. For real-time applications, additionally a balance between the complexity of the models and their accuracy has to be found.

The reduction of tonal range inherently distorts the original HDR data to some extent

and a too strong distortion impedes the comprehension of image contents. The quality of tone mapping operators could be measured by the degree to which the distortions have been avoided. For this, a metric has to be designed that is able to compare the perceptibility of corresponding image contents between two images with different tonal range – the tone mapping result and the original HDR. Further, to achieve the highest fidelity of HDR depiction the perceivable distortions detected by the metric should be restored in a post tone mapping step. Although, if one assumes that tone mapping results are obtained as the most optimal numerical solution, such a restoration would have to overcome the numerical limits.

In the scope of this dissertation we cover all of the said aspects of tone mapping. We first develop a real-time tone mapping solution which reproduces the subjective appearance of dynamic HDR contents by accounting for perceptual effects that significantly contribute to the appearance of natural scenes. We then design a computational model of lightness perception that can be applied to high quality tone mapping for static images to reproduce the original HDR appearance of tones. Next, we identify common distortions typical to tone mapping which may hinder the comprehension of image contents, we design appropriate metrics to measure the perceived magnitude of these distortions, and evaluate existing tone mapping algorithms accordingly. To compensate for observed distortions, we introduce a method which improves the tone mapping results beyond numerically optimized solution by using techniques strongly based on perception of contrasts. Presented solutions can be efficiently integrated in varied HDR applications including photography, playback of HDR video, image synthesis, lighting simulation, predictive rendering, and computer games.

1.2 Main Contributions

The fundamental ideas discussed in this dissertation have already been partially published in international journals and presented at conferences. They have been further summarized in the overview papers [Mantiuk et al. 2007a, Mantiuk et al. 2007b], in the books [Krawczyk et al. 2007b, Myszkowski et al. 2008], and at the tutorial [Myszkowski and Heidrich 2005]. Here, they are combined under the common concept of applying the knowledge of human visual perception to the processing of high dynamic range visual contents for standard displays. With respect to these publications, we revise presented methods and demonstrate improved results. The key contributions can be summarized as follows:

- Real-time tone mapping with simulation of perceptual effects. We design a real-time implementation of the photographic tone reproduction in graphics hardware and extend it at a minimal computational cost with selected perceptual effects which significantly influence the appearance of scenes. Such effects convey the subjective impression of night scenes and bright light sources which normally is not communicated on standard displays. We use this tone mapping algorithm in the HDR video player and in real-time realistic image synthesis. [Krawczyk et al. 2005c, Mantiuk et al. 2004, Dmitriev et al. 2004, Havran et al. 2005].

- Computational model of lightness perception. Based on a descriptive model of the anchoring theory of lightness perception we develop a computational model which aims at the accurate reproduction of HDR image appearance in terms

of lightness. We validate the model by simulating the appearance of known perceptual illusions and apply it to tone mapping for high fidelity reproduction of HDR. [Krawczyk et al. 2005b, Krawczyk et al. 2006, Krawczyk et al. 2007a].

- Objective evaluation of tone mapping operators. We identify contrast distortions that typically happen in tone mapping because of the dynamic range reduction. We design appropriate metrics that measure the perceived magnitude of these distortions, and evaluate existing tone mapping algorithms accordingly. Our evaluation facilitates the choice of an appropriate tone mapping algorithm under certain known requirements and permits to easily compare new algorithms to the state-of-the-art. [Smith et al. 2006].

- Contrast restoration by adaptive countershading. The results of our evaluation indicate that all existing tone mapping operators introduce a certain degree of contrast degradation. We observe that the perceived magnitude of contrast can be robustly increased with a sparing use of tonal range by the so called countershading profiles. We automatically identify the image features which require restoration and insert suitable profiles into the tone mapping result. A suprathreshold visual detection model assures that our enhancement do not introduce objectionable artifacts. [Krawczyk et al. 2007c].

1.3   Chapter Overview

This dissertation is structured as follows. In the next chapter, we give general information on image representations and introduce the field of High Dynamic Range. In Chapter 3 we explain the fundamentals of human visual perception which are relevant to the topics discussed in this thesis. Our real-time tone mapping method for dynamic HDR contents is presented in Chapter 4. The computational model of lightness perception is derived in Chapter 5 together with the demonstration of various applications. We evaluate existing tone mapping operators in Chapter 6 and design a contrast restoration method in Chapter 7. The dissertation is summarized in Chapter 8 with conclusions and outlook for future work. Additionally, in Appendix A we describe in details the calibration of standard and HDR cameras which is useful to capture HDR contents used throughout this thesis and in Appendix B we describe our Open Source software for working with HDR images and video.

Chapter 2

High Dynamic Range Imaging

We start this dissertation with an introduction to the field of high dynamic range (HDR) imaging. We explain here the difference between the standard digital image representation and the new high dynamic range imaging and indicate the advantages of the latter. We give an overview of capture techniques that can provide HDR images and HDR video which are used as input in the methods presented in the following chapters. We further explain in detail the process of tone mapping which is the main focus of this thesis. For a broader picture, we also briefly review the examples of applications in which high dynamic range imaging is particularly attractive.

2.1   Digital Images and Color Spaces

The topics discussed in this thesis focus around digital images and video. The digital image is a numerical data structure for representation of visual contents. It consists of usually rectangular matrix of image elements – pixels. Each pixel has an individually defined intensity. The intensity is usually described by three numbers to define color, but it can also be one number for monochrome images or more numbers for multi-spectral data. The numbers are called color components.

The way in which the color components determine the actual color is defined by the specific color space that is used. The most popular color space for digital images is sRGB [Stokes and Anderson 1996]. It defines color by three primaries: red, green, blue, and follows the additive mixing model [Hunt 1995]. The additive mixing model means that each number defines how much of each of the primary lights have to be emitted to create the desired color. The sRGB standard defines the spectral specification of these three primaries, which is the same as the recommendation for standard displays [ITU 1990], and the nonlinear transformation between the physical intensity of these primaries and the actual 8-bit number stored in the digital image – the gamma correction. This color space is matched to the so called standard display whose specifications are a reference for the manufactures and guarantee a similar appearance of visual contents on various media that follow the standard including also cameras, scanners, and printers. Unfortunately, the sRGB specification is tailored for displays and it is not capable of representing the complete light information in the scene. The rep-

name	formula	example	context
contrast ratio	$CR = 1 : (Y_{peak}/Y_{noise})$	1:500	displays
log exposure range	$D = \log_{10}(Y_{peak}) - \log_{10}(Y_{noise})$	2.7 orders	HDR imaging,
	$L = \log_2(Y_{peak}) - \log_2(Y_{noise})$	9 f-stops	photography
signal to noise ratio	$SNR = 20 \cdot \log_{10}(Y_{peak}/Y_{noise})$	53 [dB]	digital cameras

Table 2.1: Measures of dynamic range and their context of application. The example column illustrates the same dynamic range expressed in different units.

resentation of certain colors or brightness levels requires values that lay outside the specified 8-bit range.

The CIE XYZ is a special color space, which is based on direct measurements of the human eye, that can describe all apparent colors at all light levels. Although the tristimulus representation does not define all possible spectra, it does allow to recreate all possible perceptible colors because of *metamerism* (see [Hunt 1995]). Metamerism occurs when two color samples of different spectral power distribution appear to be of the same color. It happens because color is sensed by human vision with three types of photoreceptors that respond to a cumulative energy from a certain range of wavelengths (Section 3.1.2). The primaries of XYZ color space, however, do not correspond to these response functions. Instead, the Y component has been designed to correspond to luminance – the amount of luminous power perceived by a human eye, and X, Z primaries have been optimized for metameric matches. All components are described by non-negative real numbers.

The CIE XYZ is a generic color space and it serves as a basis from which many other color spaces are defined, but itself is not popular in practical use, because the majority of devices are based on RGB primaries. To combine the generality of XYZ space with the popularity of RGB, the non-linearity and 8-bit restriction of sRGB color space is dropped. The RGB intensities are in this case linearly related to luminance, but the representation of some colors requires negative values of the primaries which is not physically correct. Nevertheless, such linear RGB representation of digital images is particularly common in high dynamic range imaging which is described further in this chapter. Digital images and video with linear RGB representation can be captured using photometrically calibrated camera systems as explained in Appendix A, or obtained through color space conversions. These conversions are precisely specified by mathematical equations and can be found for instance in [Hunt 1995, Wyszecki and Stiles 2000].

2.2   Dynamic Range

In principle, the term *dynamic range* is used in engineering to define the ratio between the largest and the smallest quantity under consideration. With respect to images, the observed quantity is the luminance level and there are several measures of dynamic range in use depending on the applications. They are summarized in Table 2.1.

The *contrast ratio* is a measure used in display systems and defines the ratio between the luminance of the brightest color it can produce (white) and the darkest (black). In

case the luminance of black is zero, as for instance in HDR displays [Seetzen et al. 2004], the first controllable level above zero is considered as the darkest to avoid infinity. The ratio is usually normalized by the black level for clarity.

The *log exposure range* is a measure commonly adopted in high dynamic range imaging to measure the dynamic range of scenes. Here the considered ratio is between the brightest and the darkest parts of a scene given in luminance. The log exposure range is specified in orders of magnitude which permits the expression of such ratios in a concise form using the logarithmic base 10 and is usually truncated to one floating point position. It is also related to the measure of allowed exposure error in photography – exposure latitude. The *exposure latitude* is defined as the luminance range the film can capture minus the luminance range of the photographed scene and is expressed using logarithm base 2 with precision up to $^1/_3$. The choice of logarithm base is motivated by the scale of exposure settings, aperture closure (f-stops) and shutter speed (seconds), where one step double or halfs the amount of captured light. Thus the exposure latitude tells the photographers how large a mistake they can make in setting the exposure parameters while still obtaining a satisfactory image. This measure is mentioned here, because its units, *f-stop steps* or *f-stops* in short, are often perhaps incorrectly used in HDR photography to define the luminance range of a photographed scene alone.

The *signal to noise ratio* (SNR) is most often used to express the dynamic range of a digital camera. In this context, it is usually measured as the ratio of the intensity that just saturates the image sensor to the minimum intensity that can be observed above the noise level of the sensor. It is expressed in decibels [dB] using 20 times base-10 logarithm.

The actual procedure to measure dynamic range is not well defined and therefore the numbers vary. For instance, display manufacturers often measure the white level and the black level with a separate set of display parameters that are fine-tuned to achieve the highest possible number which is obviously overestimated and no displayed image can show such a contrast. On the other hand, HDR images often have very few pixels of extremely bright or dim value. An image can be low-pass filtered before the actual dynamic range measure is taken to assure a reliable estimation. Such filtering averages the minimum luminance thus gives a reliable noise floor, and smoothes single pixels with very high luminance thus gives a reasonable maximum amplitude estimate. Such a measurement is more stable compared to the non-blurred maximum and minimum luminance.

2.3 Low vs. High Dynamic Range

The term *low dynamic range* (LDR) refers in general to the 8-bit and 16-bit representations of visual contents, which are currently the most common standards in digital imaging. Such LDR representation is practically supported by all consumer products including digital cameras, scanners, displays, printers, storage formats and media. Importantly, the term does not, however, refer to the number of bits per se, but rather to the maximum dynamic range that such representation can accommodate and to its output oriented design.

The LDR contents do not actually store the measured scene colors captured by a camera, but their processed version which can be directly depicted on a typical display

	Camera Dynamic Range	
Standard (Low) Dynamic Range		**High Dynamic Range**
50 dB	*Camera Dynamic Range*	120 dB
1:200	*Display Contrast*	1:15,000
8-bit or 16-bit	*Quantization*	floating point or variable
display-referred	*Image Representation*	scene-referred
display-limited	*Fidelity*	as good as the eye can see
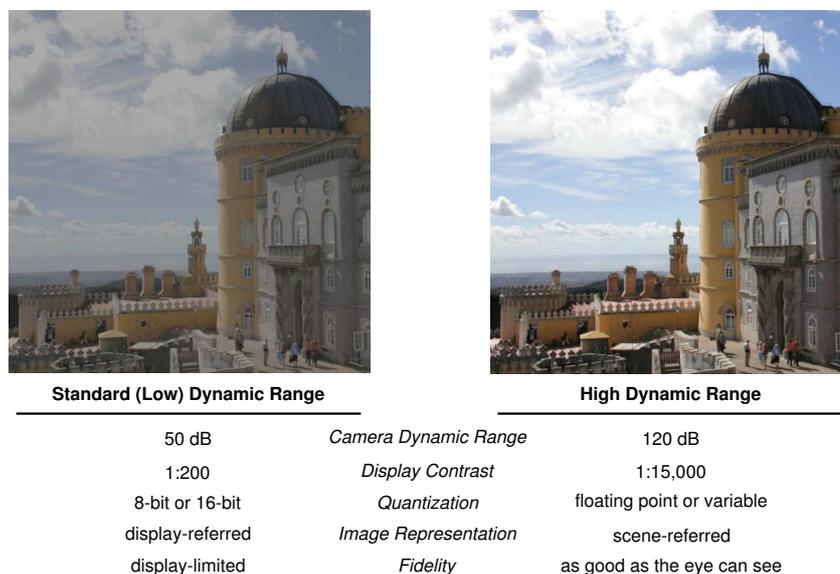
Figure 2.1: The advantages of HDR compared to LDR from the applications point of view. The quality of the LDR image have been reduced on purpose to illustrate a potential difference between the HDR and LDR visual contents. The given numbers serve as an example and are not meant to be a precise reference.

device. Such direct depiction will closely match the appearance of the photographed scene as long as a display follows the "standard display recommendations" [ITU 1990]. These specifications, developed in the 90s, are adjusted to the capabilities of the displays at that time and are also appropriate for other media such as prints and projectors. For the price of compatibility, these specifications are very restrictive and in principle limit both the maximum dynamic range and the color gamut of visual contents. Moreover, they are currently outdated by rapid advances in capture and display technology.

The main goal of *high dynamic range* (HDR) imaging is to abandon such legacy restrictions and to provide the precise representation of real world light intensities that define the entire scene appearance. Unlike the *display-referred* representation typical to LDR contents, the precision of such a *scene-referred* representation matches or surpasses capabilities of human vision and in principle corresponds to the original light values captured from a scene. In practice, the term high dynamic range is used with respect to the visual contents whose dynamic range is higher than that of LDR contents and whose intensities are linearly proportional to the original luminance or actually equal to it. The accommodate such a rich representation, data is stored in variable precision formats, often directly in floating point format. The perceptually best motivated representation of the HDR contents is the CIE XYZ color space, favorably photometrically calibrated [Mantiuk et al. 2007a].

From an applications point of view, the HDR technologies deliver more capture and display contrast, more precise quantization, and higher color fidelity. In photography, the true range of real world luminance permits scene captures that are free of under- and over-exposures. These qualities are summarized and simulated in Figure 2.1. There is, however, one caveat. The display-referred representation guarantees approximately

the same appearance of visual contents on any media as long as they follow the standards, because they have been stored according to the standards. The scene-referred representations are in most cases impossible to be directly depicted on even the most current devices and require that an appropriate rendering happens prior to or during the display. New recommendations can hardly be proposed in view of constantly improving display capabilities. A reasonable assumption, which could guarantee the same as display-referred representations, is that a display rendering algorithm should aim at the reproduction of the original appearance of a scene given the capabilities of the particular device. Such appearance reproduction for display purposes is the main focus of this dissertation.

2.4 Capture Techniques Capable of HDR

In recent years several new techniques have been developed that are capable of capturing images with a dynamic range of up to 8 orders of magnitude at video frame rates. In principle, there are two major approaches to capturing such a high dynamic range: to develop new HDR sensors or to expose LDR sensors to light at more than one exposure level and later recombine these exposures into one high dynamic range image by means of a software algorithm. With respect to the second approach, the variation of exposure level can be achieved in three ways. The exposure can change in time, meaning that for each video frame a sequence of images of the same scene is captured, each with a different exposure. The exposure can change in space, such that the sensitivity to light of pixels in a sensor changes spatially and pixels in one image are non-uniformly exposed to light. Alternatively, an optical element can split light onto several sensors with each having a different exposure setting. We summarize such software and hardware solutions to HDR capture in the following sections.

2.4.1 Temporal Exposure Change

This is probably the most straightforward and the most popular method to capture HDR with a single low dynamic range sensor. Although such a sensor captures at once only a limited range of luminance in the scene, its operating range can encompass the full range of luminance through the change of exposure parameters. Therefore a sequence of images, each exposed in such a way that a different range of luminance is captured, may together acquire the whole dynamic range of the scene, see Figure 2.2. Such captures can be merged into one HDR frame by a simple averaging of pixel values across the exposures, after accounting for a camera response and normalizing by the exposure change (for details on the algorithm refer to Appendix A). Theoretically, this approach allows to capture scenes of arbitrary dynamic range, with an adequate number of exposures per frame, and exploits the full resolution and capture quality of a camera.

HDR capture based on the temporal exposure change has, however, certain limitations especially in the context of video. Correct reconstruction of HDR from multiple images requires that each of the images capture exactly the same scene at a pixel level accuracy. This requirement cannot be practically fulfilled, because of camera motion and motion of objects in a scene, and pure merging techniques lead to motion artifacts and ghosting. To improve quality, such global and local displacements in images
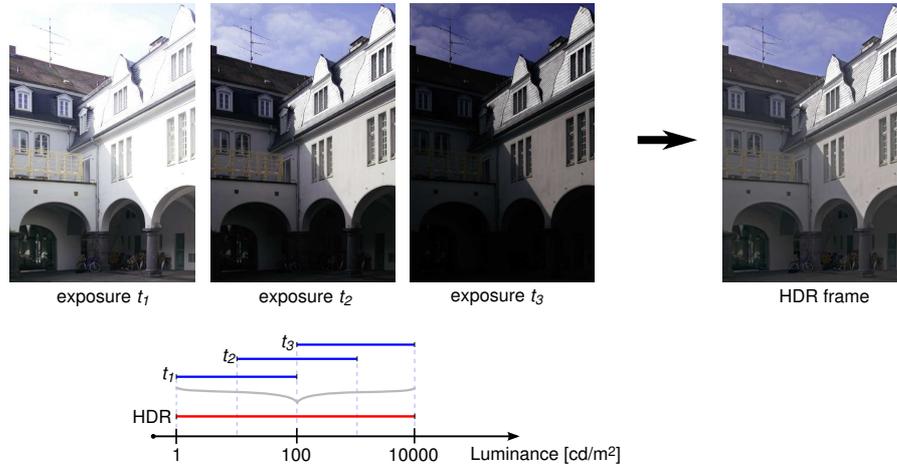
Figure 2.2: Three consecutive exposures captured at immediate time steps t_1, t_2, t_3 contain different luminance ranges of a scene. The HDR frame merged from these exposures contains the full range of luminance in this scene. HDR frame tone mapped for illustration using a lightness perception inspired technique [Krawczyk et al. 2005b].

within an HDR frame must be re-aligned using for instance optical flow estimation. Further, alignment of images that constitute one frame has to be temporally coherent with adjacent frames. A complete solution that captures two images per frame and allows for real-time performance with 25 fps HDR video capture is described in [Kang et al. 2003]. An alternative solution that captures a much larger dynamic range of about 140dB, but does not compensate for motion artifacts is available from [Uner and Gustavson 2007].

The temporal exposure change requires a fast camera, because the effective dynamic range depends on the amount of captures per frame. For instance a 200Hz camera is necessary to have a 25fps video with 8 captures per frame that can give an approximate dynamic range of 140dB [Uner and Gustavson 2007]. With such a short time per image capture, the camera sensor must have a sufficiently high sensitivity to light to be able to operate in low light conditions. Unfortunately, such a boosted sensitivity usually increases noise.

2.4.2   Spatial Exposure Change

To avoid potential artifacts from motion in the scene, the exposure parameters may also change within a single capture [Nayar and Mitsunaga 2000], as an alternative to the temporal exposure change. The spatial exposure change is usually achieved using a mask which has a per pixel variable optical density. The number of different optical densities can be flexibly chosen and they can create a regular or irregular pattern. Nayar and Mitsunaga [Nayar and Mitsunaga 2000] propose to use a mask with a regular pattern of four different exposures as shown in Figure 2.3. Such a mask can be then placed directly in front of a camera sensor or in the lens between primary and imaging elements.

scene capture without the mask

mask with optical densities
varying per pixel

scene capture through the mask
(varying pixel exposures)
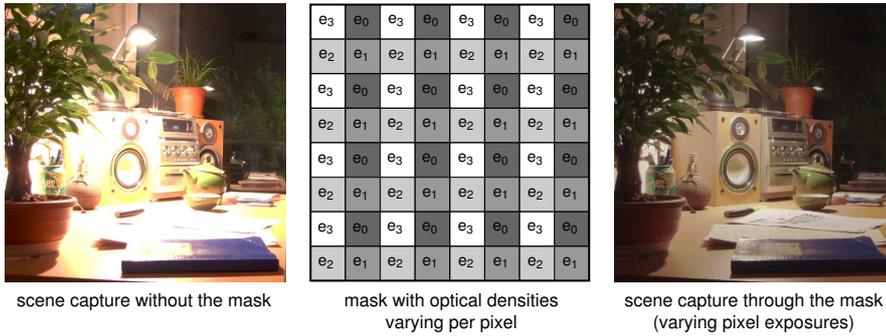
Figure 2.3: Single exposure using a standard image sensor cannot capture full dynamic range of the scene (left). The mask with per pixel varying optical densities $e_3 = 4e_2 = 16e_1 = 64e_0$ (middle) can be put in front of a sensor. Using such a mask at least one pixel per 4 is well exposed during the capture (right). The right image is best viewed in the electronic version of the thesis.

For the pattern shown in Figure 2.3, the full dynamic range can be recovered either by aggregation or by interpolation. The aggregation is performed over a small area which includes a capture of that area through each optical density, thus at several different exposures. The different exposures in the area are combined into one HDR pixel by means of a multi-exposure principle explained in the previous section, at the cost of a reduced resolution of the resulting HDR frame. To preserve the original resolution, HDR pixel values can also be interpolated from adjacent pixels in a similar manner as colors from the Bayer pattern. Depending on the luminance levels, aliasing and interpolation artifacts may appear.

The effective dynamic range in this approach depends on the number of different optical densities available in the pattern. A regular pattern of 4 densities, as shown in Figure 2.3, such that $e_3 = 4e_2 = 16e_1 = 64e_0$ gives a dynamic range of about 85dB for an 8-bit sensor [Nayar and Mitsunaga 2000]. The quantization step in the reconstructed HDR frame is non-uniform and increases for high luminance levels. The size of the step is, however, acceptable, because it follows the gamma curve.

An alternative implementation of spatial exposure change, Adaptive Dynamic Range Imaging (ADRI), utilizes an adaptive optical density mask instead of a fixed pattern element [Nayar and Branzoi 2003]. Such a mask adjusts its optical density per pixel informed by a feedback mechanism from the image sensor. Thus saturated pixels increase the density of corresponding pixels in the mask, and noisy pixels decrease. The feedback, however, introduces a delay which can appear as temporal over- or underexposure of moving high contrast edges. Such a delay, which is minimally one frame, may be longer if the mask with adapting optical densities has high latency.

Another variation of spatial exposure change is implemented in a sensor whose pixels are composed of more than one light sensing element each of which has a different sensitivity to light [Street August 1998]. This approach is, however, limited by the size of the sensing element per pixel, and practically only two elements are used. Although in such a configuration, one achieves only a minor improvement in the dynamic range, so far only this implementation is applied in commercial cameras (Fuji Super CCD).

2.4.3   Multiple Sensors with Beam Splitters

Following the multi-exposure approach to extending dynamic range, one can capture several exposures per video frame at once using beam splitters [Aggarwal and Ahuja 2004]. The idea, so called split aperture imaging, is to direct the light from the lens to more than one imaging sensor. Theoretically this allows to capture HDR without making any quality trade-offs and without motion artifacts. In practice, however, the effective dynamic range depends on the number of sensors used in the camera and such a solution may become rather costly when a larger dynamic range is desired. Further, splitting the light requires an increased sensitivity of the sensors.

2.4.4   Solid State Sensors

There are currently two major approaches to extend the dynamic range of an imaging sensor. One type of sensor collects charge generated by the photo current. The amount of charge collected per unit of time is linearly related to the irradiance on the chip (similar to a standard CCD chip [Janesick 2001]), the exposure time is however varying per pixel (sometimes called "locally auto-adaptive" [Lulé et al. 1999]). This can for instance be achieved by sequentially capturing multiple exposures with different exposure time settings or by stopping after some time the exposure of the pixels that would be overexposed during the next time step. A second type of sensor uses the logarithmic response of a component to compute the logarithm of the irradiance in the analog domain. Both types require a suitable analog-digital conversion and generate typically a non-linearly sampled signal encoded using 8–16 bits per pixel value. Several HDR video cameras based on these sensors are already commercially available. Such cameras allow to capture dynamic scenes with high contrast, and compared to software approaches, offer considerably wider dynamic range and quality independent of changes in the scene content as frame-to-frame coherence is not required. The properties of two of such cameras: HDRC VGAx from IMS-CHIPS [Hoefflinger 2007] and Lars III from Silicon Vision are studied in detail in Section A.4.

2.5   Tone Mapping

The contrast and brightness range in typical HDR images exceeds capabilities of current display devices or print. Thus these media are inadequate to directly reproduce the full range of captured light. Tone mapping is a technique for the purpose of reducing contrast and brightness in HDR images to enable their depiction on LDR devices. The process of tone mapping is performed by a tone mapping operator.

Particular implementations of a tone mapping operator are varied and strongly depend on a target application. A photographer, computer graphics artist or a general user will most probably like to simply obtain nice looking images. In such cases, one most often expects a good reproduction of appearance of an original HDR scene on a display device. In simulations or predictive rendering, the goals of tone mapping may be stated more precisely: to obtain a perceptual brightness match between HDR scene and tone mapped result, or to maintain equivalent object detection performance. In visualization or inspection applications often the most important is to preserve as much of fine detail

information in an image as possible. Such a plurality of objectives lead to a large number of different tone mapping operators.

Various tone mapping operators developed in recent years can be generalized as a transfer function which takes luminance or color channels of an HDR scene as input and processes it to output pixel intensities that can be displayed on LDR devices. The input HDR image can be calibrated so that its luminance is expressed in SI units cd/m^2 or it may contain relative values which are linearly related to luminance. The transfer function may be the same for all pixels in an image (global operator) or its shape may depend on the luminance of spatially local neighbors (local operator). In principle, all operators reduce the dynamic range of input data. Since most of the algorithms process only luminance, color images have to be converted to a color space that decouples luminance and chrominance, e.g. Yxy. After processing, the tone mapped intensities are used instead of the original luminance in the inverse transform to the original color space of the image.

2.5.1 Luminance Domain Operators

The most naïve approach to tone mapping is to "window" a part of luminance range in an HDR image. That is to map a selected range of luminance using a linear transfer function to a displayable range. Such an approach, however, renders dark parts of image black and saturates bright areas to white, thus removing the image details in the areas. A basic sigmoid function:

$$L = \frac{Y}{Y+1}, \tag{2.1}$$

maps the full range of scene luminance Y in the domain $[0, \inf)$ to displayable pixel intensities L in the range of $[0, 1)$. Such a function assures that no image areas are saturated or black, although contrast may be strongly compressed. Since the mapping in equation (2.1) is the same for all pixels, it is an example of a global tone mapping operator. Other global operators include logarithmic mapping [Drago et al. 2003], the sigmoid function derived from photographic process [Reinhard et al. 2002], a mapping inspired by the response of photoreceptors in the human eye [Reinhard and Devlin 2005], a function derived through histogram equalization [Ward et al. 1997]. The subtle difference in tone mapping result using these functions is illustrated in Figure 2.4. Usually, one obtains a good contrast mapping in the medium brightness levels and low contrast in the dark and bright areas of an image. Therefore, intuitively, the most interesting part of an image in terms of its contents should be mapped using the good contrast range. The appropriate medium brightness level for the mapping is in many cases automatically determined as a logarithmic average of luminance values in an image:

$$Y_A = \exp\left(\frac{\sum \log(Y + \varepsilon)}{N}\right) - \varepsilon, \tag{2.2}$$

where Y denotes luminance, N is the number of pixels in an image, and ε denotes a small constant representing the minimum luminance value. The Y_A value is then used to normalize image luminance prior to mapping with a transfer function. For example, in equation (2.1) such a normalization would map the luminance equal to Y_A to 0.5 intensity which is usually displayed as middle-gray (before the gamma correction).
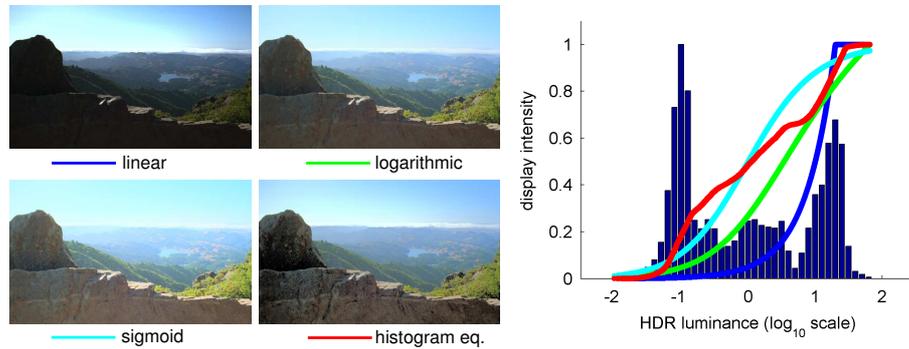
Figure 2.4: Comparison of global transfer functions. The plot illustrates how luminance values are mapped to the pixel intensities on a display. The steepness of the curve determines the contrast in a selected luminance range. Luminance values for which display intensities are close to 0 or 1 are not transferred. Source HDR image courtesy of Greg Ward.

The Y_A is often called the adapting luminance, because such a normalization is similar to the process of adaptation to light in human vision.

2.5.2   Local Adaptation

While global transfer functions are simple and efficient methods of tone mapping, the low contrast reproduction in dark and bright areas is a disadvantage. To obtain a good contrast reproduction in all areas of an image, the transfer function can be locally adjusted to a medium brightness in each area:

$$L = \frac{Y'}{Y'_L + 1},\qquad(2.3)$$

where Y' denotes HDR image luminance normalized by the globally adapting luminance $Y' = Y/Y_A$ and Y'_L is the locally adapting luminance. The value of globally adapting luminance Y_A is constant for the whole image, while the locally adapting luminance Y'_L is an average luminance in a predefined area centered around each tone mapped pixel. Practically, the Y'_L is computed by convolving the normalized image luminance Y' with a Gaussian kernel. The standard deviation of the kernel σ defines the size of an area influencing the local adaptation and usually corresponds in pixels to 1 degree of visual angle. The mechanism of local adaptation is again inspired by similar processes occurring in human eyes. Figure 2.5 illustrates the improvement in tone mapping result through introduction of local adaption.

The details are now well visible in dark and bright areas of the image. However, along high contrast edges one can notice a strong artifact visible as dark and bright outlines – the halo. The reason why such artifact appears is illustrated in Figure 2.6. Along a high contrast edge the area of local adaptation includes both high and low luminance, therefore the computed average in the area is inadequate for any of them. On the side of high luminance the local adaptation is more and more under-estimated as the tone mapped pixels are closer to the edge, therefore equation (2.3) gradually computes much

Logarithmic average of luminance in the HDR image.

Gaussian blur of the HDR image with kernel size ~1deg of visual angle.

uniform global adaptation map Y_A

global Y_A and local Y'_L adaptation
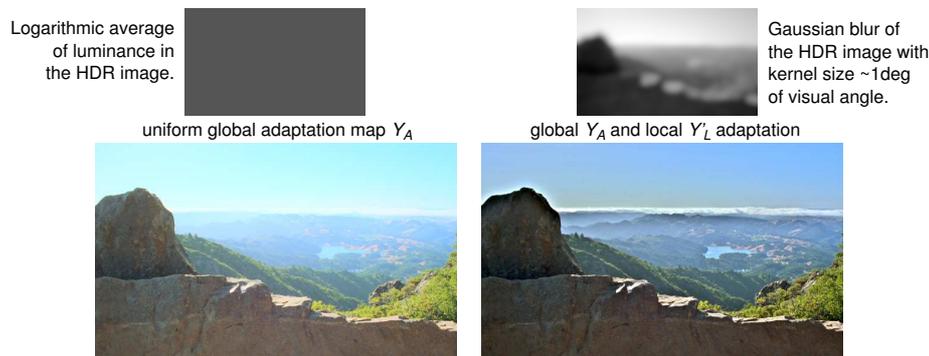
Figure 2.5: Tone mapping result with global, equation (2.1), and local adaptation, equation (2.3). The local adaptation (right) improves the reproduction of details in dark and bright image areas, but introduces halo artifacts along high contrast edges.
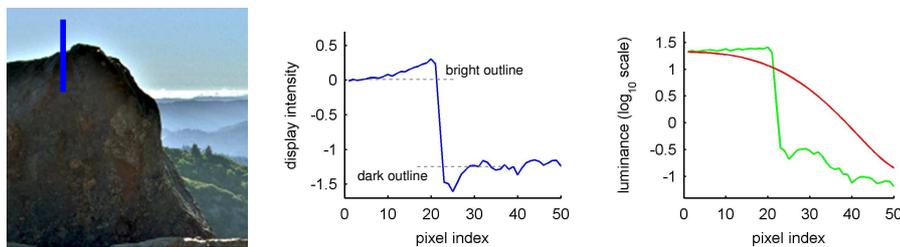
Figure 2.6: The halo artifact along a high contrast edge (left) and plots illustrating the marked scanline. Gaussian blur (under-) over-estimates the local adaptation (red) near a high contrast edge (green). Therefore the tone mapped image (blue) gets too bright (too dark) closer to such an edge.

higher intensities than appropriate. The reverse happens on the side of low luminance. A larger blur kernel spreads the artifact over a larger area, while a smaller blur kernel reduces the artifact but also reduces the reproduction of details.

2.5.3 Prevention of Halo Artifacts

Many image processing techniques have been researched to prevent the halo artifacts out of which the notable solutions are automatic dodging and burning [Reinhard et al. 2002] and the use of bilateral filtering instead of Gaussian blur [Durand and Dorsey 2002].

The automatic dodging and burning technique derives intuitively from the observation that a halo is caused by a too large adaptation area, Figure 2.6, but also a large area is desired for a good reproduction of details. Therefore, the size of the local adaptation area is adjusted individually for each pixel location such that it is as large as possible but does not introduce halo. The halo artifact appears as soon as both very high and very low luminance values exist in an adaptation area and significantly change the estimated local adaptation. Therefore, by progressively increasing the adaptation area

for each pixel, the following test can detect the appearance of halo:

$$|Y_L(x, y, \sigma_i) - Y_L(x, y, \sigma_{i+1})| < \varepsilon. \tag{2.4}$$

For each pixel, the size of the adaptation area, defined by the standard deviation of the Gaussian kernel σ_i, is progressively increased until the difference between the two successive estimates is larger than a predefined threshold ε. The result of the Gaussian blur for the largest σ_i that passed the test is then used for given pixel in equation (2.3). The example of estimated adaptation areas is illustrated in Figure 2.7. The whole process can be very efficiently implemented using the Gaussian pyramid structure as described in [Reinhard et al. 2002].

Figure 2.7: Estimated adaptation areas for pixels marked as blue cross. In each case, the green circle denotes the largest, thus the most optimal adaptation area. A slightly larger areas denoted as red circles would change the local adaptation estimate Y_L more than acceptable threshold in equation (2.4) and would introduce a halo artifact.

Bilateral filtering is an alternative technique to prevent halos [Durand and Dorsey 2002]. The reason for halos, Figure 2.6, can also be explained by the fact that the local adaptation for a pixel of high luminance is incorrectly influenced by pixels of low luminance. Therefore, excluding pixels of significantly different luminance from local adaptation estimation prevents the appearance of halo in a smilar way as in equation (2.4). The bilateral filter [Tomasi and Manduchi 1998] is a modification of the Gaussian filter which includes an appropriate penalizing function:

$$Y_L^p = \sum_{q \in N(P)} f_{\sigma_s}(\|p - q\|) \cdot Y^q \cdot g_{\sigma_r}(|Y^p - Y^q|). \tag{2.5}$$

In the above equation, p denotes the location of the tone mapped pixel, q denotes pixel locations in the neighborhood $N(p)$ of p. The first two terms of equation, $f_{\sigma_s} \cdot Y^q$, define Gaussian filtering with spatial σ_s. The last term, g_{σ_r}, excludes from the convolution those pixels whose luminance value differs from the tone mapped one by more than σ_r. Both f and g are Gaussian functions, and luminance is usually expressed in the logarithmic space for the purpose of such filtering. The bilateral filtering process is shown in Figure 2.8.

Compared to the automatic dodging and burning, the bilateral filter better reproduces details at the edges, because in most cases a relatively larger area is used for estimation of local adaptation. Although the exact computation of equation (2.5) is very expensive, a good approximation can be computed very efficiently [Durand and Dorsey 2002, Chen et al. 2007].
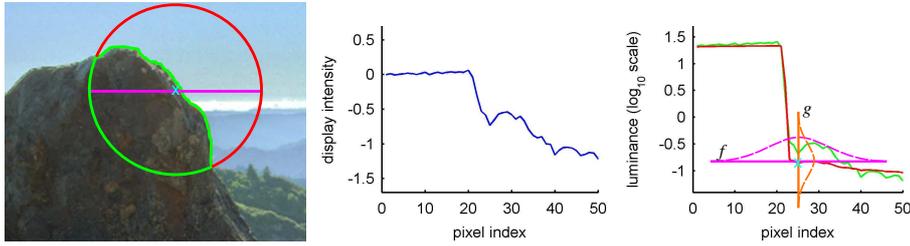
Figure 2.8: Bilateral filtering of a similar scanline as in Figure 2.6, here marked in magenta. The penalizing function *g* improves the estimation of the local adaptation (red) by excluding pixels in the neighborhood *f* (magenta) whose luminance value is outside the defined range (orange). Thus, the local adaptation for the pixel marked with a cross (left image) is estimated only from the pixels in the area outlined in green, while the Gaussian blur would also include pixels in the area outlined in red.

2.5.4 Contrast Domain Operators

The tone mapping methods discussed so far perform the dynamic range reducing operations directly on luminance or on color channel intensities. However, one can observe that an image with a wide range of luminance also contains a large range of contrasts. Therefore, as an alternative to luminance range compression, contrast magnitudes in the image can be reduced. Since the contrasts convey semantical information in images, such a control over contrast can be advantageous. For instance, small contrasts usually represent the reflectance properties of surfaces, like texture, medium contrasts often define the outlines of objects, and large contrasts represent changes in illumination. Particularly, large contrasts are in most cases the cause of a high dynamic range. By preserving small and medium contrasts, and reducing large contrasts, one can reduce the dynamic range of illumination and at the same time preserve good visibility of details from the original HDR image. Such a contrast based processing gives a better control over transferred image information than the luminance based operators. The latter, however, give a better control over brightness mapping. In fact, it is hard to impose a target luminance range for contrast based compression.

A typical contrast based tone mapping operator includes the following steps. First, the input luminance is converted to a contrast representation. The magnitudes of contrasts are then modulated using a transfer function for contrast – the tone mapping step. Next, the modulated contrast representation is integrated to recover the luminance information, and such luminance is then scaled to fit the available dynamic range. Finally, since the result of integration is calculated with an unknown offset, the image brightness of the tone mapping result is adjusted.

Contrast in tone mapping applications is most often measured as a logarithmic ratio of luminance:

$$C = \log \frac{Y^p}{Y^q}, \tag{2.6}$$

where Y^p and Y^q denote luminance of adjacent pixel location. The contrast representation of an image is computed as a gradient of $\log Y$, since the logarithm of division is equal to the difference of logarithms. For tone mapping, such a representation is often multi-resolution to measure contrasts between adjacent pixels (full resolution) and ad-

jacent areas in an HDR image (coarser resolutions). The contrasts are then modulated by a transfer function, for example [Fattal et al. 2002]:

$$T(C) = \frac{\alpha}{|C|} \cdot \left(\frac{|C|}{\alpha} \right)^{\beta}.$$
(2.7)

Given that $\beta \in (0, 1)$, such a function attenuates gradients that are stronger than α and amplifies smaller ones. Thus, if α is equal to medium contrasts in an image, equation (2.7) reduces the dynamic range caused by large differences in illumination and enhances fine scale details. More complex transfer functions are also possible including for instance contrast equalization [Mantiuk et al. 2006]. As the final step, the modulated contrast representation of an HDR image has to be integrated in order to obtain intensities in a tone mapped image. The integration step is performed by solving the Poisson equation and the image brightness adjustment step is left for manual setting by a user. The stages of the contrast domain tone mapping process are illustrated in Figure 2.9.

(a) HDR image, clipped   (b) contrast representation   (c) contrast transfer map   (d) tone mapping result
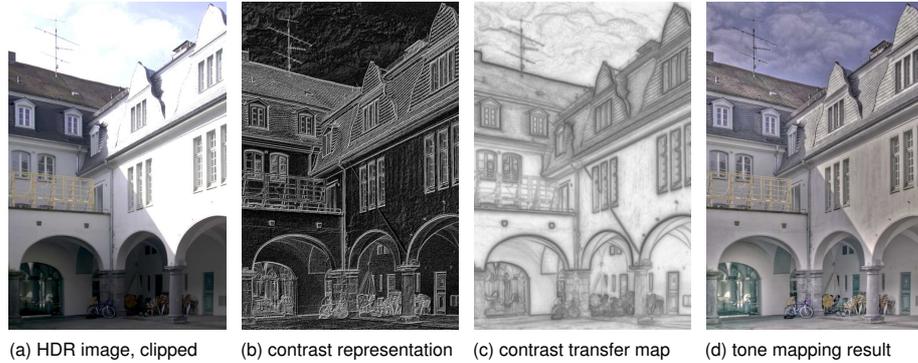
Figure 2.9: Contrast domain tone mapping. The HDR image (a) is transformed to a contrast representation (b) which is multiplied by a contrast transfer function (c). The contrast representation is then integrated to obtain a tone mapped image (d). In (b) white denotes strong local contrast and black no contrast. In (c) black denotes strong contrast attenuation and white marks no change in local contrast.

2.5.5   Summary

In the previous sections we have introduced the general ideas behind tone mapping algorithms. Many variations of such algorithms exist that differ in subtle details from each other and we refer the reader to [Reinhard et al. 2005] for detailed descriptions.

Remarkably, all tone mapping operators change the pixel intensities in a tone mapped image and their relations with spatial neighbors with respect to the original HDR. The nature of the algorithms is mostly inspired by typical image processing or computer vision approaches and with some equations adopted from known processes happening in early stages of human vision. This imposes certain mathematical properties of tone mapped images and assures that basic appearance properties, like adaptation to light, are preserved. However, as explained in Chapter 3, the appearance of images is largely

influenced by processes occuring in later parts of the visual system. Such later stages, the cognitive processes, have so far not been studied in the context of tone mapping. We argue that these process need to be considered to preserve the original appearance of HDR images during tone mapping, and in Chapter 5 we propose a tone mapping operator which is inspired by such a cognitive interpretation of scenes.

With a large number of available tone mapping algorithms whose results differ only subtly, an evaluation framework would facilitate the choice of an appropriate operator for a given application. While several psychophysical evaluations have already been conducted [Drago et al. 2002, Kuang et al. 2004, Yoshida et al. 2005, Ledda et al. 2005] which offer a form of ranking of available algorithms, the particular reasons why some operators are preferred over others is not well understood. Also, it is not easy to evaluate new algorithms and include them in such rankings. Therefore, in Chapter 6 we propose an objective evaluation of tone mapping operators which allows to understand the effect of tone mapping algorithms in terms of perceptual change of contrasts and brightness. We evaluate several state-of-the-art operators accordingly.

2.6 HDR Applications

High dynamic range imaging offers an unprecedented quality of capture and representation of visual contents. Such a complete visual information, including the true range of real world luminance, can improve the quality of many computer graphics applications and can enable simulations or measurements that before have not been possible.

Currently the most popular application is HDR photography in which the high dynamic range permits to capture photos free of under- and over-exposures with an unprecedented level of details. New tone reproduction algorithms provide the users with a better control over tones, contrast, and levels of detail in their photography.

The capture of physically accurate light measurements, enabled by HDR techniques, opens new possibilities in realistic image synthesis. For instance, a digital representation of light surrounding a certain scene can be captured and used to synthesize computer graphics objects with a high fidelity to the natural light conditions [Havran et al. 2005]. This permits the simulation of real-world light conditions in computer graphics visualizations [Dmitriev et al. 2004]. The true light information permits also to model the behavior of human visual system and to simulate the actual scene appearance in various illumination conditions. Precise light measurements, offered by HDR techniques, further increase the quality of acquisition of objects' appearance. The reflectance properties of their complex materials can be measured and used in realistic rendering of digitized objects to obtain high quality virtual reproductions [Gösele 2004].

Emerging HDR displays [Seetzen et al. 2004], whose capabilities outperform those of modern displays, offer interesting visualization possibilities. For instance, the ergonomics of LCD displays can be directly investigated in various illumination conditions by taking a display-in-display approach [Dmitriev et al. 2004]. The calculated light interaction with LCD panel can be directly visualized on the HDR display whose display range can easily accommodate such information. It can be envisaged that further ideas for applications will appear with the growing popularity of HDR imaging.

Chapter 3

Human Visual Perception

Visual information about our surroundings is available to us through the act of seeing. The ability to see involves capturing light with our eyes and interpreting it with our brain in such a way that we can understand what we see, remember it, react to it. Human eyes successfully operate in an incredibly varied range of light conditions from the darkest night to the brightest day and in all these conditions we consistently recognize known objects, faces, materials, our environment. From the moment the light enters our eyes, the full process of perception is facilitated by a complex system that consists of several stages. First, the light is focused in the eye with a lens onto the light sensitive back of the eye where photoreceptors transform captured photons into the neuronal signals. These signals are then transmitted to the brain where they are decomposed in a hierarchical way and processed in the visual cortex. Only such processed visual information appears to us as the images that we see and understand.

The human visual system is well adapted to its tasks, but is not perfect. The optics of an eye, like any solid optical element, have certain characteristic and limits. The photoreceptors are distributed on the retina with a finite resolution that can be reduced under circumstances to increase sensitivity but impairing the visual acuity. The channels that transmit signals to brain have a relatively low bandwidth compared to the amount of captured data and some information is sent to the brain in a faster but reduced form. On the other hand, cognitive interpretation of transmitted signals is subjective and highly influenced by our prior knowledge. In the end, observed scenes do not necessarily appear to us exactly the same as they actually are.

The investigation of true perception of scenes as performed by the human visual system is very worthwhile in the context of this dissertation. The knowledge of how human vision works may let us focus on those aspects that are important when depicting HDR scenes on devices with limited capabilities or when measuring the perceptual quality of such depictions. For instance, the ability to adapt to a change in light levels such that the perceived scenes appear with a roughly constant overall brightness is a fundamental process of the human visual system that makes tone mapping at all possible. Further analysis may indicate what effects typical to our visual system contribute to the overall subjective appearance of scenes and perhaps to what extent they should be accounted for during tone mapping. Likewise, the understanding of how the visual information is in fact interpreted by the human brain may help to prepare the depictions of HDR
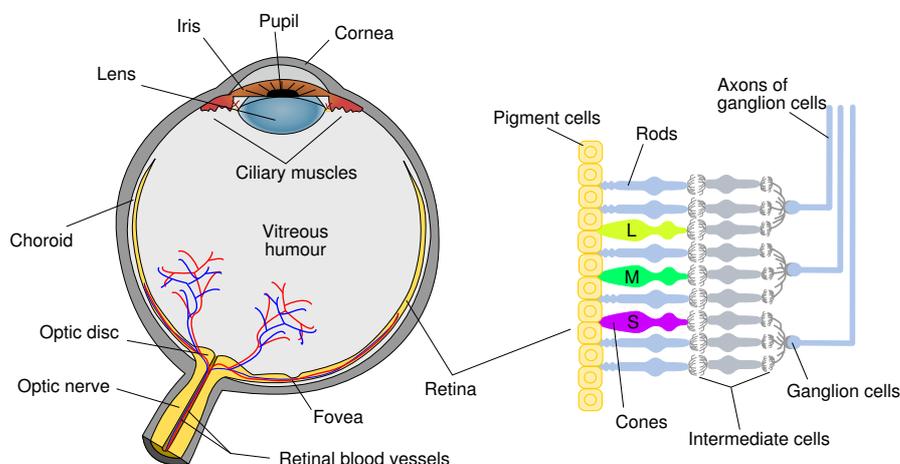
Figure 3.1: The general plan of the human eye and the retina with the names of their principal parts. Original image of the eye courtesy of Rhcastilhos, Wikimedia Commons.

in accordance with the mechanisms of perception such that visual communication is facilitated.

The aim of this chapter is to introduce the basic knowledge of human vision and visual perception, and to point the reader to those aspects that are particularly interesting in view of the topics addressed in this thesis. In the following sections, we give rather succinct descriptions of the terms that are necessary to understand the further chapters of this dissertation and refer the reader for detailed descriptions to related text books [Palmer 1999, Fairchild 1998, Wyszecki and Stiles 2000, Wandell 1995]. We further elaborate on each of the topics in the relevant chapters of the thesis where we also cite research papers.

3.1   The Eye

The act of seeing starts with the moment light enters our eyes and generates a visual stimulus. The physical construction of an eye (Figure 3.1) and the means by which the light is registered determines the physical appearance of images. The human eye is an organ of approximately spherical shape filled with a clear gel, the vitreous humour, which separates the lens and the retina. The eyeball is fixed in an eye socket and is moved with extrinsic muscles to aim at a point of fixation.

The visual process which happens in the human eye can be divided into the optical part and the sensory part. The optical part gathers the light entering our eyes with a lens which focuses it on the back of the eye – the retina. The retina contains pigment cells and underlying photoreceptors which actually respond to light. The photoreceptors transform captured photons into the neuronal signals and thus begin the sensory part. The neuronal signals of photoreceptors are passed through various intermediate cells and are aggregated by ganglion cells whose axons form an optic nerve which connects

to the brain. Each part of this process interferes to some extent with the original visual information. In the following sections we describe these parts in detail and discuss their impact on image appearance.

3.1.1 Optical System

The optical part of the visual system is embodied in the eyeball and in the main part consist of cornea, pupil, iris, and lens (refer to Figure 3.1). The cornea is the transparent front of the eye which is the most exposed part to light. Together with the lens it refracts light and whereas it has the major contribution to the eye's optical power its refraction parameter is fixed. The iris is the most recognized component of the eye with a characteristic pigment. It contains the muscles that allow to contract and extend the pupil. The pupil is a black opening in the eye that regulates the amount of light that enters it. The variable size of the pupil with respect to the level of light is the most apparent indication of the adaptation of human visual system to light. The adaptation to light is, however, primarily a sensory process and the actual contribution of variable pupil size is minor. The main task of the optical system of the eye is the accommodation. It is the process of adjusting the shape of the lens using the ciliary muscles to allow focusing on various distances. The accommodation guarantees that objects of interest form a sharp image on the retina.

The optical system of the eye, like any man-made optical lens, has its characteristics and limitations. The primary observable limitation is the scattered light in the eye as it passes through the cornea, lens, and vitreous humour [Wyszecki and Stiles 2000]. It can affect vision when for instance an eye is observing a bright light source against a dark background as illustrated in Figure 3.2. The *veiling glare* that is observed in such cases causes a decrease in contrast and increase in brightness in the vicinity of such bright areas. The amount of light scatter can be measured and described in the form of a point spread function (PSF) dependent on the pupil diameter [Wandell 1995]. The PSF gives the relative amount of light registered by photoreceptors at a specified angular distance on the retina from the place where the point light source is directly projected. As the pupil diameter increases, the amount of scattering also increases. An example of PSF function can be found in Section 4.2.5 where we use it to simulate such glare effects in tone mapping.

The visual effect of light scatter in the eye has an interesting perceptual implication. People so strongly associate the perception of glare as being caused by a strong source of light that painting such a visual effect around an area in an image causes this area to appear to be much brighter than it actually is. This phenomenon is visible in Figure 3.2.

3.1.2 Sensory Part

The sensory part of the visual system starts when the light focused by the optical part falls on the retina (Figure 3.1) and triggers a chain of neural events that eventually transform captured photons to a coded signal which is sent through the optic nerve to the brain for visual interpretation. The retina is a thin layer of neural cells which contains light sensitive photoreceptors and a non-photosensitive melanin pigment. The photons that hit photoreceptors are absorbed by the photosensitive pigment and elicit
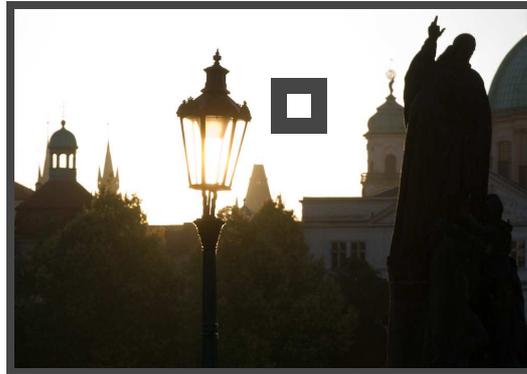
Figure 3.2: In the image above, the rising sun is hidden exactly behind the street lamp. When looking at this lamp, the light scatter causes that the middle parts of the lamp create much weaker contrast with the background than the rest and that the area around has a higher brightness. Additionally, the presence of glare makes the area inside the lamp appear luminous and brighter than the square beside given for reference and the rest of the sky, although their brightness is equal.

signal that passes through the neural cells in the retina. The photons which miss the photoreceptors are absorbed by the melanin pigment.

There are two kinds of photoreceptors: rods and cones containing the actual photosensitive pigment. Cones are less sensitive to light than the rod cells, but their response times are faster. There are three kinds of cones which together allow the perception of color and whose absorption properties define the spectrum of the visible light which is 350nm - 750nm. The three cone types have their peek response approximately to yellowish-green (L cones), green (M cones), and bluish-violet (S cones). While the human visual system can directly measure only three colors, according to the opponent color theory [Wyszecki and Stiles 2000] the full color gamut is visible by actually measuring the relative differences between the responses of different cone types. There is only one type of rod cell that absorbs a similar spectrum as the summary response of the cones, but being more sensitive to the blue part and almost not sensitive to the wavelengths above the red part. There are on average 6 million cone cells with the highest density around the central part of the retina – the fovea (Figure 3.1) providing high visual resolution of the object in focus. On average 100 million rod cells are densely located on the outer area of the fovea supporting the peripheral vision.

The actual neural processes that happen in the retina are not well understood and the explanations given in the literature are considered to be more or less speculative, although supported by profound evidence [Wyszecki and Stiles 2000]. In general, the ganglion cells (Figure 3.1) receive signals from bipolar and horizontal cells which themselves receive input from the photoreceptors. A ganglion cell receives input from several photoreceptors which creates a single receptive field. The receptive fields are organized in a center/surround manner such that the light falling on their center is activating the response and the light falling on the immediately surrounding region is inhibiting the response. Such an organization provides an efficient edge enhancement but also indicates that the brain receives encoded differences in signals rather than their absolute intensity levels registered by the photoreceptors. This leads to an important conclusion

which is that the human visual system interprets information given in the form of local intensity differences – contrasts.

Various properties of photoreceptors and their distribution on the retina have a visible impact on the appearance of scenes and we explain the relevant aspects in the following sections.

3.1.3 Vision Modes

Cones, which allow to distinguish colors, respond well in dim to bright light (10^{-1} to 10^{+8} cd/m^2). Rods, which are much more sensitive to light than cones, respond best in darkness and up to moderate light (10^{-6} to 10^{+1} cd/m^2), but are blinded by luminances above 10^{+2} cd/m^2 and do not output any usable signals at such illumination levels. These luminance characteristics of photoreceptors determine three vision modes of the human visual system: photopic, mesopic and scotopic, which are illustrated in Figure 3.3.
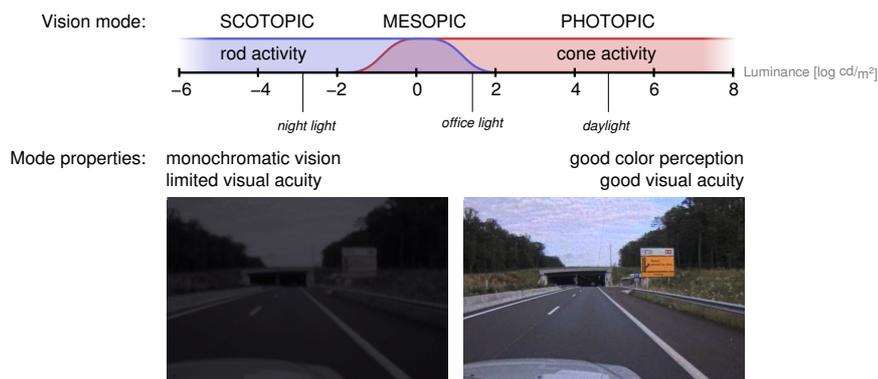
Figure 3.3: Vision modes of the human visual system with an illustration of usual scene appearance.

Photopic vision is active under well-lit conditions which usually occur in daylight. During photopic vision only cone photoreceptors are active, while rods are blinded by too strong illumination. Photopic mode is characterized by good color perception and sharp vision.

The scotopic vision mode refers to dim and dark illumination conditions and occurs mostly at night. The light is registered only by rods, because the sensitivity of cones is too low. Since cones are inactive, colors are not distinguishable. As the illumination decreases the neural signal is aggregated over larger groups of rods to increase the effective sensitivity of the visual system. Such aggregation, however, reduces the actual resolution of rods on the retina which is sensed as a decrease in the visual acuity.

The mesopic vision occurs under dim illumination and is a transition mode between photopic and scotopic vision. Both cones and rods are active resulting in a color vision with slightly stronger sensitivity to blue colors over the red ones. This is known as Purkinje shift.

The difference in vision modes has a strong impact on image appearance. As illustrated in Figure 3.3, the daylight vision is associated with good perception of colors and good acuity, while at night most things look colorless and details are not well visible. These appearance properties are very common and the scenes that depict the properties of scotopic vision are immediately thought to have low illumination levels. This is among others exploited in movies: the night scenes are post-processed to have realistic night vision qualities, because display devices cannot depict images at sufficiently low luminance levels such that these quality appear naturally.

3.1.4 Photoreceptor Response

The light captured by a photoreceptor is transformed into the neuronal signal of proportional strength. There is a significant evidence that the relation is non-linear [Hunt 1995] and the response of both rods and cones is usually approximated with a sigmoid function shown in Figure 3.4. Such a response has several important properties. When the intensity of the stimulus is low, the capture noise is suppressed by the lower non-linearity. When the intensity is very high, the response gradually reaches its maximum beyond which no signal increase is registered. The main part is approximately linear, so that the differences between intensity signals of moderate luminance are well transferred.
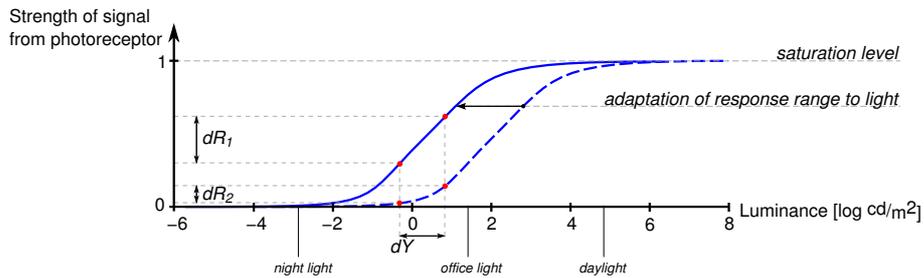
Figure 3.4: Sigmoid response to light of a photoreceptor. The luminance difference dY causes a much more pronounced signal difference dR_1 if the response range is well adapted to light conditions, compared to the difference dR_2 observed in an unadapted state.

Under constant ambient light conditions, the photoreceptors successfully register the luminance range of about 4 orders of magnitude. However, the range of luminance in nature can be as high as 10^8 during the day and as low as 10^{-4} in the night. To accommodate such a much wider range, the photoreceptors adapt their response range to the current ambient light level. This can be observed as a shift in the response curve with respect to the medium luminance range as illustrated in Figure 3.4. Such an adaptation is the fundamental ability of the eye which leads to an approximately similar sensory appearance of a scene independently of the absolute luminance level, except for changes in the vision modes.

The shape of the response function and the adjustment of the response range with respect to the medium luminance in the scene inspired several tone mapping operators, including the one discussed in Section 4.2.1.

3.1.5 Temporal Light and Dark Adaptation

The process of adjusting the response range of photoreceptors to match the illumination condition, mentioned in the previous section, is not immediate. Normally, the adaptation processes are mostly not noticed because the changes in the illumination during the course of day and night are very slow. Sudden changes, however, cause visible loss in the sensitivity as illustrated in Figure 3.5. For instance, when on a sunny day one immediately enters a dark theatre, the interior is at first dark and no details can be discerned – only after several seconds the silhouettes of objects start to appear.
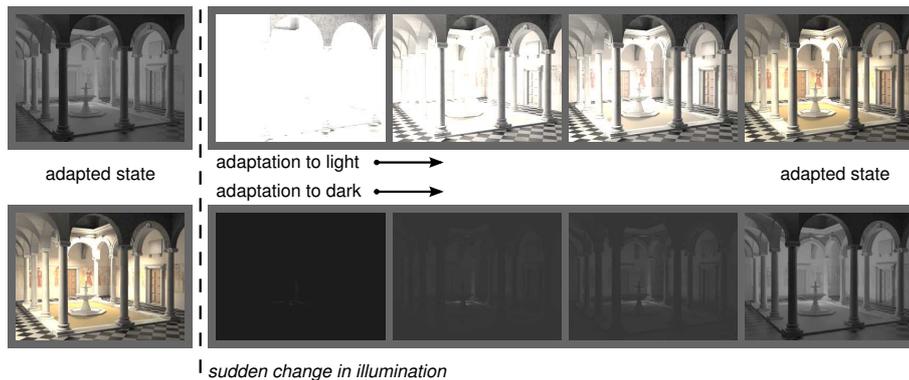
Figure 3.5: Visual experience in certain time intervals during the temporal adaptation to light and to dark caused by a sudden change in illumination. The visibility improves with time because the response range of photoreceptors adjusts to the medium illumination in the scene as illustrated in Figure 3.4.

The adaptation of photoreceptors response to light is a temporal process. In a simple form, the adaptation is accomplished by neural processes attributed to cell interconnections in the retina and by chemical processes of bleaching and regeneration of photosensitive pigment. The neural processes react very fast and are mostly accomplished after several seconds. The chemical processes account for significant changes in illumination, are much slower and visibly asymmetric. The adaptation from bright sunlight to complete darkness takes up to 30 minutes while the reverse process is fully accomplished in about 5 minutes.

The precise time course of adaptation can be measured with threshold sensitivity experiments. In such experiments, the subjects are first exposed to a certain ambient illumination for enough time to adapt to its intensity. Next, the illumination changes suddenly and experimenters measure the subjects' ability to detect a small luminance difference on a test stimulus. If we recall Figure 3.4, in the adapted state the difference in the luminance on the stimuli falls on the linear part of the photoreceptors response thus produces the strongest difference in output signal. As long as the adaptation process is not complete, the difference of output signal is compressed by the non-linear part and the stimulus has to be stronger to be perceived. In the adapted state the visibility of the luminance difference in the test stimuli is the strongest.

The measured time course of dark and light adaptation is shown in Figure 3.6. The plots start with a sudden change in illumination which results in high detection thresh-

(a) dark adaptation, rods and cones   (b) light adaptation – rods   (c) light adaptation – cones
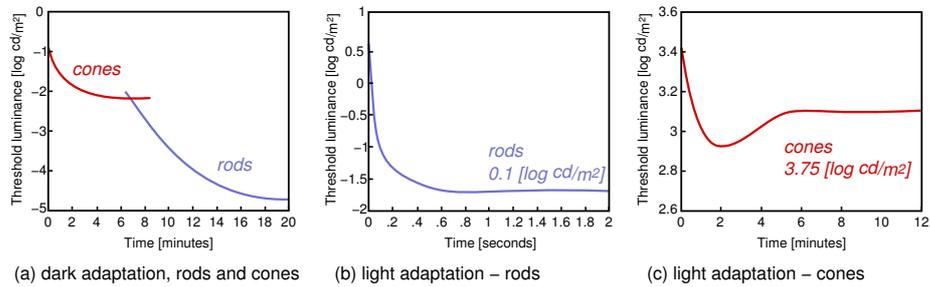
Figure 3.6: Time course of dark adaptation (a) and light adaptation (b,c) as a function of threshold sensitivity. Dark adaptation was to complete darkness, light adaptation to the specified luminance levels. Plots after [Ferwerda et al. 1996].

olds, thus low sensitivity. The sensitivity of both rods and cones progresses asymptotically. During dark adaptation, the process of cones is faster but cones soon reach their maximum sensitivity. The sensitivity level is for a moment constant because the rods still have not recovered from the strong illumination. With time, rods dominate vision and continue the adaptation process until maximum sensitivity is reached. The light adaptation in the scotopic range is extremely rapid and nearly 75% of the process is accomplished within first 400ms. The cone system adapts to light much slower and requires about 3 minutes to reach maximum sensitivity which then slightly decreases. Due to their asymptotic nature, the adaptation processes are often approximated with an exponential function as explained in Section 4.2.2.

3.1.6   Perceptual Implications

The optical and sensory parts of early vision have a definitive impact on the subjective appearance of scenes. The properties of rods and cones define clearly distinguishable vision modes, photopic and scotopic, which deliver a totally different appearance of the same scene depending on the absolute luminance of ambient illumination. The course of light and dark adaptation has a temporal but significant influence on the appearance of dynamic scenes in which the illumination changes suddenly.

Further, we are so strongly accustomed to the effects related to early vision that we tend to associate appropriate illumination levels to match the corresponding image appearance [Spencer et al. 1995]. Therefore simulation of scotopic vision conveys to the observers the message of low illumination conditions. The veiling glare in images indicates the presence of light sources and causes that the areas centered in the glare appear to be brighter than they actually are.

Existing display devices show images at the luminance levels largely corresponding to the mesopic vision. When, for instance, the low luminance levels of night scenes are transposed to the luminance range of a display their subjective appearance is not any more observed. Since the characteristics of these early vision processes have been measured, it seems reasonable to simulate them during the display to convey the subjective appearance of scenes with luminance levels outside the display range. We propose an appropriate solution in Chapter 4.

3.2 Visual Sensitivity

Contrast, which is a difference in luminance between adjacent or distant areas, is the primary form of visual information that is delivered to the brain. The visual sensitivity is therefore measured by our ability to respond to physical contrasts of different properties across different observation conditions.

Contrast is captured by the receptive fields and neural interconnections in the retina. According to the multi-resolution theory of vision, for interpretation the visual signal is split into several channels each dedicated to contrasts of a particular spatial and temporal frequency, and orientation. The sensitivity to the information in different channels varies and depends on the significance of their information in terms of recognition of natural scenes. Further, all contributing components are highly adaptive and the visual sensitivity changes as a function of light intensity.

The understanding of human response to contrast is interesting in the context of tone mapping algorithms, because from the technical side their essential goal is to reduce the contrast in the HDR contents. In the following sections we focus on the measured observations that characterize the effective human perception of contrast, while the underlying processes are fundamentally described in handbooks [Wandell 1995, Palmer 1999].

There are two main aspects of contrast perception. The ability of detecting signal on a uniform background – contrast detection (threshold aspect), and the ability to judge if one signal generates a stronger contrast than the other – contrast discrimination (suprathreshold aspect). The difference between the two is illustrated in Figure 3.7. In contrast detection (a), the difference in luminance dY between the background and the patch is measured. For the patch to be discernible it needs to be stronger than the visibility threshold. In contrast discrimination (b), the difference between contrasts of two patches with their backgrounds d^2Y is measured. It needs to be strong enough so that the difference in brightness between the two patches is discernible. The performance of contrast perception depends on the ambient illumination level, spatial frequency of the signal, and the presence of other signals in the area of interest.
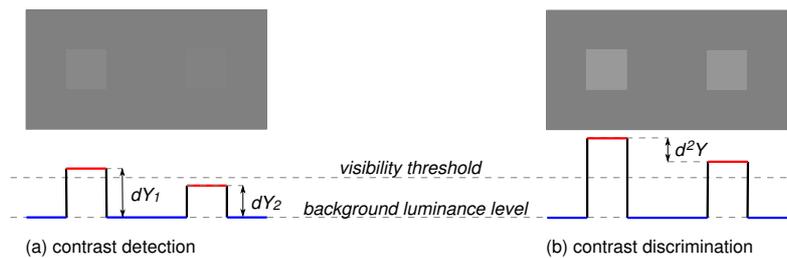
Figure 3.7: Test stimuli and luminance profiles illustrating contrast detection and contrast discrimination.

3.2.1 Luminance Masking

The most basic response to contrast is measured by the ability of detecting a luminance change dY on a uniform background of luminance Y (Figure 3.7). This directly corre-

sponds to the psychophysical measure of *just noticeable difference – jnd* for luminance. For a single stimulus in a form of patch shown in Figure 3.7, the smallest detectable luminance difference, the detection threshold, changes as a function of the background luminance. The effect is called luminance masking, because the existing luminance in the background masks the visibility of stimuli whose luminance is slightly lower or higher.

Initially, according to Weber's Law the relation of just noticeable difference in luminance with respect to the background luminance has been assumed to be constant. Currently, several more precise threshold versus intensity (tvi) functions are in use, including the one defined in the CIE standard [CIE 1981] which we use in the further chapters of this dissertation. The tvi function is plotted in Figure 3.8 for reference. In photopic vision the ratio of the visibility threshold to the background luminance is approximately constant and Weber's law gives a good prediction. In the mesopic and scotopic range, however, the detection thresholds do not decrease significantly with respect to the background.
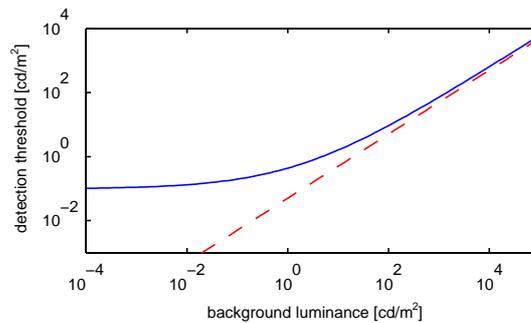
Figure 3.8: Visibility threshold as the function of background luminance (tvi) from the CIE standard [CIE 1981] (solid line) and Weber's Law for reference (dashed line).

3.2.2   Spatial Contrast Sensitivity

The human vision response to contrasts in complex images varies depending on the frequencies of their components. When observing periodic signals, which can be thought of as an approximation of natural images, the detection thresholds discussed in the previous section further depend on the spatial frequency of the signal. The effect is illustrated in Figure 3.9.

The pattern shown in Figure 3.9 measures the spatial contrast sensitivity function (CSF) for human vision. The CSF in principle increases the visibility thresholds given by the tvi function for low and high frequencies. It also indicates that our perception is best at detecting medium frequencies which usually define the outlines of objects in a scene. The sensitivity is expressed in terms of cycles per degree of visual angle, therefore the ability to perceive contrast in a pattern of certain frequency changes with the viewing distance.
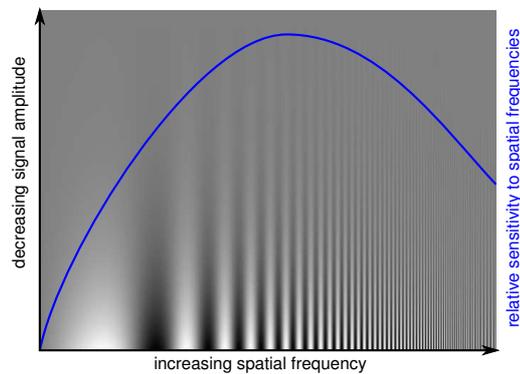
Figure 3.9: Spatial contrast sensitivity chart after [Campbell and Robson 1968]. The amplitude of signal decreases uniformly for each frequency, but the perceived signal disappears non-uniformly as approximately outlined by the relative sensitivity plot.

3.2.3 Contrast Masking

The ability of human vision to detect signals of certain frequencies is further impeded by the presence of other visible signals in the area of interest which have a similar frequency and spatial orientation. This is because existing contrasts mask the new contrast of the introduced signal. The effect can be observed in Figure 3.10.

test signal   masking signal (image) with the test signal
superimposed
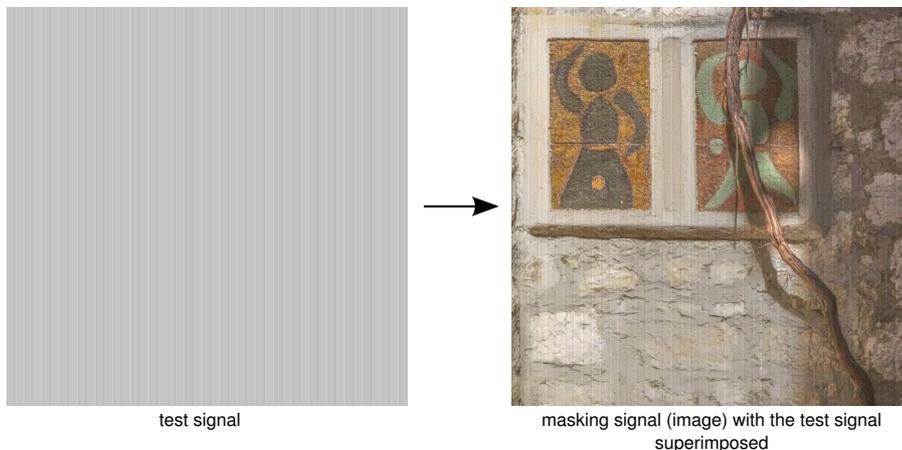
Figure 3.10: Contrast masking example. The visibility of the test signal in the image depends on the local image contents. It is hardly perceivable in the areas with high frequency textures or with patterns of similar orientation. The test signal consists of periodic countershading profiles introduced in Chapter 7.

Contrast masking is measured by finding a necessary amplitude of the test signal such that it is visible when super-imposed on the existing signal. If the existing signal is uniform, this is the same effect as the luminance masking therefore the visibility thresholds are equivalent. When the amplitude of contrasts of the existing signal increases, the initial visibility threshold changes as described by a threshold elevation function. The

threshold elevation function increases the detection threshold as a function of the local contrast of the same frequency and orientation and is usually modeled by a power function with a typical exponent between 0.65 - 1, mostly 0.7 [Daly 1993] (Figure 3.11). The function has two asymptotic regions, one with slope of zero and one with slope near 1. The zero slope occurs for low contrasts of the masking signal (the existing signal) that are not visible and therefore do not change the visibility threshold which in this case is the same as for the uniform background. As soon as the local sub-band contrast of masking signal is greater than the threshold contrast, the contrast of the test signal must be stronger to be visible.
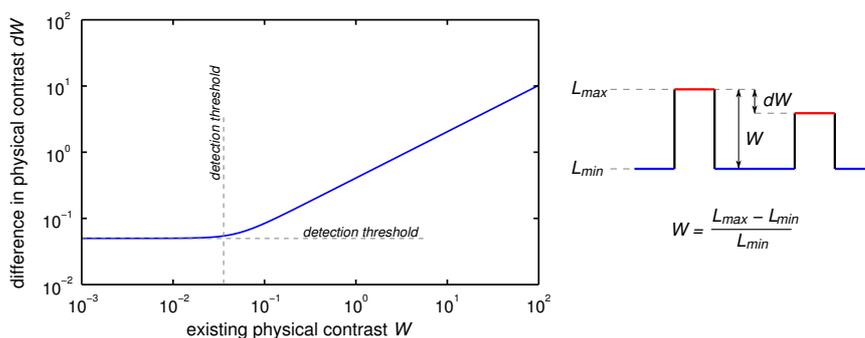
Figure 3.11: Necessary difference in the physical contrast dW, so that it is visible in the presence of the existing contrast W. Plot based on the threshold elevation function after [Daly 1993].

3.2.4   Visual Detection Models

The measurements of human visual sensitivity to various properties of physical contrasts are often used to design computational models of contrast detection and discrimination. Such models permit to estimate whether certain visual signals are visible to an average observer. These signals can be both useful information, which should be strong enough to be above the visibility threshold, and undesired information like compression artifacts whose magnitude should be kept below the predicted visibility level. In this dissertation we exploit both of these aspects and use such models to evaluate the quality of contrast reproduction (Chapter 6) and to predict the potential visibility of contrast enhancement as a halo artifact (Chapter 7).

3.2.5   Processing of Visual Information

While both color contrasts and luminance contrasts deliver information to the brain, the luminance is the primary source and colors are supplementary information. Figure 3.12 illustrates a color image and two versions of it: one contains only luminance contrasts and the other only color contrasts. The recognition of image contents is equally good in the luminance image as it is in the original image. In contrary, the contents of the color only image are recognized with a substantial difficulty and some information is missing.

(a) luminance contrasts only (b) luminance and color contrasts (original image) (c) color contrasts only
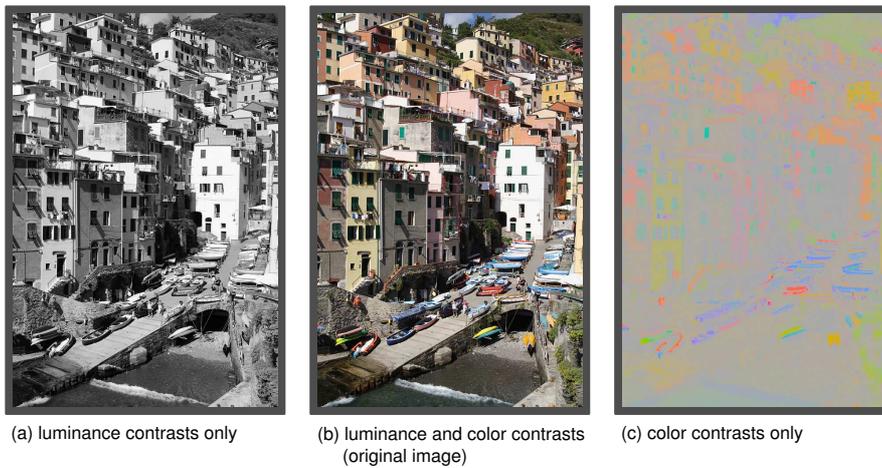
Figure 3.12: The original image (b) decomposed into the luminance contrasts alone (a) and the color contrasts (c).

The theoretical approach to processing of visual information distinguishes two visual pathways: magnocellular and parvocellular [Wandell 1995], also popularly known as the "where system" and the "what system" respectively [Livingstone 2002]. The "where system" is responsible for the perception of depth and motion, for the spatial organization of scene objects, and for the figure/ground segregation. It can be characterized by a fast response to changes and high contrast sensitivity, but it is color blind and its visual acuity is lower by a factor of 2 compared to the "what system". The "what system" is color selective and has a high visual acuity, but its sensitivity to contrast is low and it responds slower to changes. The "what system" is responsible for recognition of objects, including faces, and perception of colors. It can be subdivided into the "form system" which uses luminance and color to define shapes of objects, and the "color system" which identifies color of surfaces.

We have focused our discussion in the previous sections on luminance contrast alone, because it appears to be the major factor in the successful interpretation of scenes. Luminance contrast is the common component of both "where" and "what" systems and permits the perception of objects and their spatial organization in the scene. Since it is directly affected by the process of tone mapping, we argue that it requires closer perceptual investigation. Consequently, we evaluate existing tone mapping operators in terms of their good reproduction of luminance contrasts (Chapter 6) and develop a contrast enhancement technique that facilitates the perception of image features after tone mapping (Chapter 7).

3.2.6 Contrast Illusions

The visual information is sensed locally, through the receptive fields, and registered as contrasts due to the center-surround construction of such fields. Therefore luminance differences deliver useful information which is propagated over the uniform areas. Also, a contrast needs to be sufficiently strong so that it is above the visibility threshold and can be interpreted by the visual system. Too weak luminance differences
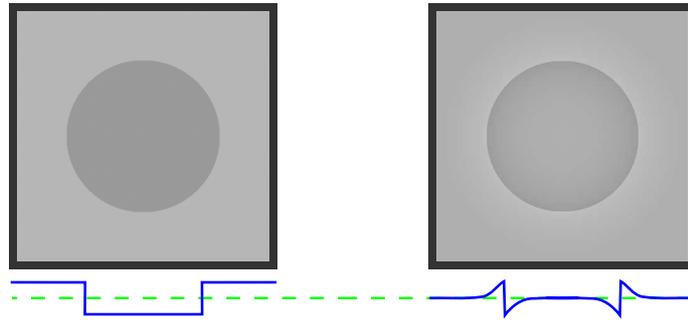
Figure 3.13: The appearance of the right image matches the left one, although the luminance profiles of these images differ (bottom plots). The gradual darkening and brightening at the borders of areas of equal luminance creates a perceived brightness difference between them – the Craik-O'Brien-Cornsweet illusion (right image).

are sensed as uniform areas.

The insensitivity to certain visual signals leads to strong contrast illusions. A carefully shaped luminance profile at an edge between two areas, like in Figure 3.13, causes change in the brightness of the whole areas and increases the perceived contrast between them [Dooley and Greenfield 1977]. Apparently, the gradual change of the luminance away from the edge towards the mean value is not well observed by the human visual system. The only information in the image, immediate contrast at the edge, defines the brightness relation between the two patches which is propagated over the whole area of the patches – hence the illusory brightness difference. Such a perceived contrast between image areas is strong and appears even for a consecutive combination of profiles or when an area is isolated from the area which contains the profile.

The appearance of illusion is not limited to simple uniform areas. Interestingly, it seems to be stimulated not only by physical aspects of the visual signal, but also by cognitive interpretation [Purves et al. 1999]. For instance, certain visual cues that the profile is caused by a difference in the illumination, possibly confirmed by the perspective information, strongly enhances the effect. In the example shown in Figure 3.14, a rough Cornsweet profile on the border of two pages creates the illusion but the same profile overlaid on an out-of-context area gives a very weak effect.

Such illusory contrast effects permit to influence the change in the brightness appearance of larger image areas only by modifying their borders. Thus an informed use of such Cornsweet profiles can be used for image enhancement with a very sparing use of dynamic range. We take advantage of such possibility and design the appropriate image processing tool in Chapter 7.

3.3   Image Appearance

The final appearance of visual contents is a product of the cognitive processes highly influenced by our understanding of scene components and their remembered appearance. After [Fairchild and Johnson 2003] one can give the following scene to interpret: a yellow house with the blue door viewed during sunset. A closer inspection of the
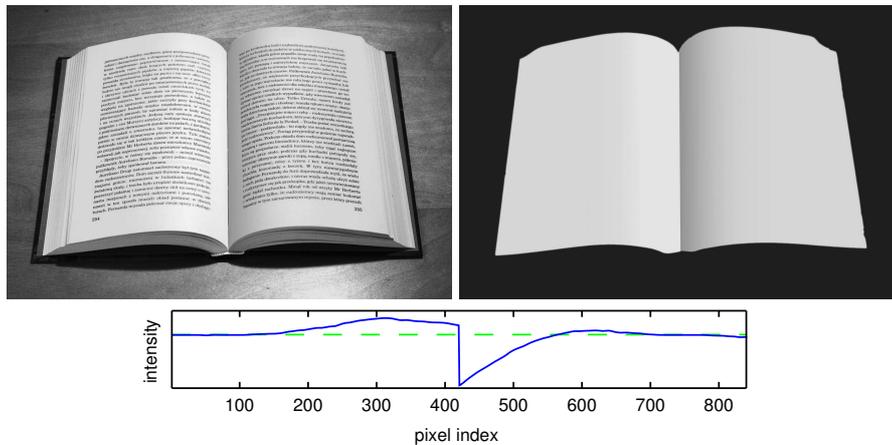
Figure 3.14: An open book illuminated from the right creates a profile similar to Cornsweet (bottom plot). Due to the perceptual illusion, the left page of the book appears to be brighter than the right page, although their luminance is identical apart from the profile shape. The plot illustrates the intensities of a scanline at the level of the page footer. The illusion appearing on the book is articulated by contextual information – the same luminance profile shown in the right image does not exhibit such a strong effect.

scene reveals that a nearby wall casts a shadow precisely on the door which means that the blue color is actually caused by the pure illumination from the sky. Since it is now apparent that the blue is a property of the illumination, the appearance of the door changes from blue to yellow – its true reflectance. This illustrates the ability of the human visual system to identify the illumination in the scene and to discard it during the interpretation of objects' appearance. This effect is known as lightness and color constancy and leads to a similar appearance of scenes independent of the illumination.

The appearance of objects is first determined through the interpretation of their visual stimuli. The stimuli can be perceived as: illuminant, illumination, surface, volume, or unrelated stimuli [Fairchild and Johnson 2003]. The illuminant attribute is assigned to objects which are perceived as being a source of light. The illumination appearance is attributed to the properties of prevailing illumination rather than objects and is mediated by illuminated objects that reflect light and cast shadows. In the presence of a physical and recognizable object the stimulus can be attributed as the property of its surface. The stimuli can also be interpreted as the cause of transparency and, in a special case, as unrelated information when the stimuli is observed in an out of context mode, for instance through an aperture. By determining such classes of visual stimuli, the human visual system tries to discard the illumination information in the observed scenes. A failure in the correct interpretation can lead to a visible difference in the image appearance as illustrated by the yellow house example.

The interpreted stimuli are further defined by five perceptual attributes: brightness, lightness, hue, colorfulness, and chroma. Brightness is the perceived luminance coming for the scene, discarding its contextual properties. Lightness is a contextual interpretation of brightness and is judged relative to the brightness of a similarly illuminated area that appears to be white. Hue is defined by the dominant wavelength of stimuli.

Chroma defines the color purity as a difference from the gray tone. Colorfulness is the subjective impression of color purity determined by the absolute luminance level (the Hunt effect [Fairchild and Johnson 2003]). Lightness and chroma can only be attributed to objects in the scene and are not a property of illumination, while colorfulness is an interaction of illumination with an object. Further, brightness and colorfulness are related to the absolute amount of energy emitted from observed objects.

In the context of reproducing the appearance of scenes on displays in changed observation conditions the following can be observed. The hue attribute is in general not affected by changes in illumination (excluding extreme cases of monochromatic illumination). On the other hand, the brightness and colorfulness cannot be reproduced at all if a scene is displayed with a different luminance range than the original, which is usually the case, because these attributes depend on the absolute luminance level. For a good reproduction of appearance, the effort should therefore be focused on a good match of lightness and chroma between the original scene and its displayed depiction. For the most part, lightness requires the correct estimation of luminance perceived as white, and chroma requires the correct estimation of lightness and the perceived saturation of the color [Fairchild and Johnson 2003]. The correct estimation of lightness is one of the topics of this dissertation and in the next section we briefly review existing lightness perception theories.

3.3.1   Perception of Lightness

Any luminance value can be perceived as literally any lightness value (shade of gray) depending on its context within the image. Initially, lightness has been assumed to be equivalent with reflectance which can be obtained by dividing luminance by the estimated illumination – a straightforward realization of lightness constancy. This assumption, however, has been undermined with empirical evidence.

The problem of lightness perception and lightness constancy has been studied extensively in the last two centuries for which a detailed account can be found in [Palmer 1999]. At first, the Gestalt theorists rejected the assumption that luminance per se is the stimulus for lightness. The most prominent theories follow Wallach's observation [Wallach 1948] that the perceived lightness depends on the ratio of the luminance at edges between neighboring image regions. This inspired the retinex theory [Land and McCann 1971], in which it is assumed that even for remote image regions such a ratio can be determined through the edge integration of luminance ratios along an arbitrary path connecting those regions.

Lightness can be well modeled by the retinex algorithm under the condition that the illumination changes slowly, which effectively means that sharp shadow borders cannot be properly processed. To overcome this problem, Gilchrist and his collaborators suggested that the human visual system performs an edge classification to distinguish illumination and reflectance edges [Gilchrist 1977]. This led to the concept of the decomposition of retinal images into the so called intrinsic images [Barrow and Tenenbaum 1978, Arend 1994] with reflection, illumination, depth and other information stored in independent image layers. The lightness perception theories based on intrinsic images can predict lightness constancy very successfully. However they define only relative lightness values for various scene regions. Their important shortcoming is the lack of a rule which would define the association between the predicted relative

lightness and the perceived white, grays and black across the whole scene. Furthermore, being developed for good lightness prediction, these theories fail to account for lightness constancy failures typical to human vision [Gilchrist et al. 1999].

The mapping of relative lightness to the perceived shades of gray is solved by anchoring. There are several rules of anchoring, each of which defines a method to assign one particular absolute lightness value (e.g. white, black, middle gray) to one relative lightness value – the so called anchor value. The remaining mapping can be immediately found through the known lightness ratios. In particular, the anchoring can be directly applied to the intrinsic image models, although initially it was not included, by mapping the maximum value in the reflectance layer to white.

The problem of lightness constancy failures and absolute lightness assignment, altogether, is addressed by the anchoring theory of lightness perception developed by Gilchrist et al. [Gilchrist et al. 1999]. One of the key arguments of the theory is that lightness mapping can differ even within a single image depending on the considered context of the image. In this theory, such ambiguity is accounted for by the concept of frameworks. Frameworks are image components which are grouped by the terms of Gestalt principles: mainly by common illumination, but also by proximity, similarity, co-planarity, good continuation, and common fate. An image is composed of multiple frameworks whose areas can overlap. The anchoring rule can give correct lightness estimates when considered only within one framework. The net lightness of a surface in an image can be found by estimating the influence of each of the frameworks on that surface and by calculating the weighted product of lightness mappings within each of the frameworks.

The main weakness of the lightness perception theories is that they are given in a descriptive form and lack computational models. To account for correct lightness reproduction in tone mapping, we formulate the computational model of the anchoring theory of lightness perception in Chapter 5. Our choice for this theory is motivated by its sound explanation of the particular appearance of many experimental scenes and its extensive experimental studies with human subjects.

Chapter 4

Real-time Tone Mapping for HDR Video

Low dynamic range (LDR) image and video contents, which are stored in the display-referred representation, are usually directly shown on a display. The parameters of a reference display and the preferable observation conditions are well defined in standards [ITU 1990] and guarantee that such in a sense oblivious depiction delivers good quality. While the general parameters of displays usually follow the standard, the observation conditions, however, may not match the reference. Users are able to compensate this mismatch by adjusting few parameters like contrast, brightness and saturation to improve the picture, however the range of adjustment is limited. Moreover, the excessive adjustment of these controls may not only lay beyond the capabilities of a display, but may also reveal artifacts in the image contents. This happens because the display-referred representation contains only sufficient image and video quality to produce good results under the assumed conditions. For instance, too strong contrast amplification would show contouring artifacts. On the other hand, strong increase of brightness would not reveal the details of dark picture parts as perhaps one would expect, because the brightness of these parts lays outside the dynamic range of a reference display and therefore their contents have not been stored in the stream.

Contrary to the display-referred contents, scene-referred HDR contents are not limited to the capabilities of typical displays. HDR video compression [Mantiuk et al. 2004], for example, stores as wide luminance range as the human eye can observe in a real-world scene. This in most cases largely exceeds capabilities of displays, therefore such scene-referred contents require processing (tone mapping) prior to display. Such processing is performed on the side of the target display and thus has several notable advantages with respect to the display-referred representation. First, the HDR contents can be processed in such a way that delivers the best quality on the actually used display under the actual observation conditions. Second, the limited range of brightness and contrast adjustments can be relaxed and moreover the quality of their effect is improved. For instance, the brightness correction of HDR data reveals contents of too bright or too dark picture parts because, unlike in the case of LDR data, the information there is not clipped. Finally, the ample amount of luminance information in HDR video permits to add new controls that take advantage of the available dynamic range

and can increase the realism of a picture. We observe, that for a range of luminance levels typical to certain scenes like nights or sunny days, the depiction of true luminance is not feasible on displays. Hence all such scenes appear in an almost similar way on the screen, although an average observer would expect to see bright saturated colors only on a sunny day, while subdued grayish tones with low acuity are common in the night and a veiling glare usually appears around bright lights. When these phenomena are ignored, well visible details appear unrealistic in dimly illuminated scenes, because the acuity of human vision is normally degraded in such conditions. On the other hand, perceptual effects like glare cannot be evoked because the maximum luminance of typical displays is not high enough. However, we are so used to the presence of such phenomena, that adding glare to an image can increase the subjective brightness of the tone mapped image [Spencer et al. 1995]. Therefore by simulating such perceptual phenomena, we can increase the realism of HDR contents by reducing the appearance mismatch between the real-world and display.

In our work, we focus on a real-time implementation of a high quality tone mapping operator and on the introduction of new controls that take advantage of the luminance range available in the stream. To match the real-world appearance of recorded HDR scenes, we enhance the tone mapping algorithm by incorporating the most significant perceptual effects that are related to the absolute luminance levels in the scene and to the optics of the eye (Section 3.1). Improving over previous work, we observe that these effects have much in common in terms of spatial analysis and show that making use of such similarities have a tremendous impact on the performance. Further, we add the functionality for convenient inspection of verbatim information in a selectable dynamic range. Such an inspection tool is necessary in cases when it is required to view the exact contents of the recorded scene, as for instance in forensic applications. We implement our approach in the graphics hardware as a stand-alone HDR image/video processing module and achieve a real-time performance. The computational overhead of our extensions to tone mapping is negligible. Although we primarily demonstrate the module in the context of HDR video playback, it can be as well applied to the final stage of a real-time renderer or to other stream of HDR contents like an input from a surveillance camera or data for visualization.

4.1   Previous Work

The tone mapping of HDR contents has been widely addressed in research and we have introduced most of the existing algorithms in Section 2.5. Simple algorithms, which are based on a tone reproduction curve, can be implemented very efficiently in the graphics hardware [Drago et al. 2003], but such methods fail in reproducing fine details in the HDR scenes. Most of the recent algorithms deliver a higher quality but at the cost of the increased complexity and only few are able to achieve interactive rates at 1Mpx resolution [Goodnight et al. 2003]. In contrast, our work is unique in a sense that we aim at real-time tone mapping performance wihout compromising the quality of HDR.

Certain perceptual effects, like the lack of visual acuity or color perception in night scenes, have already been accounted for in several tone mapping algorithms [Ferwerda et al. 1996, Ward et al. 1997, Durand and Dorsey 2000, Pattanaik et al. 2000]. These effects, however, have been discussed only in the context of global operators and have

been applied to the whole image with a uniform intensity. Such an approach can lead to the unrealistic depiction when a wide range of luminance is present in the scene and certain phenomena should be observed only in a part of the scene. Furthermore, the proposed solutions have been composed of multiple stages each involving complex processing such as convolutions. Although the implementation of individual perceptual effects on graphics hardware is intuitive, a naïve combination to include all of them does not even allow for an interactive performance on the graphics hardware currently available.

4.2 Computational Models

With many tone mapping algorithms available, we want to use a method that provides good, widely acknowledged results for static images. The parameters of the method should provide sufficient control to enable maintaining temporal coherence of picture during the HDR video playback. Furthermore, we require that a real-time performance is feasible for at least a reasonable approximation of such a method and that the trade off between the quality and performance can be adjusted to adapt to the capabilities of available graphics hardware. At the same time we want that the spatial analysis involved in tone mapping bear some similarities to the nature of perceptual effects that we plan to simulate. We have found that photographic tone reproduction [Reinhard et al. 2002] satisfies our requirements. In the following sections, we justify our choice by briefly explaining the tone mapping algorithm and each of the perceptual effects that we include, and by showing the apparent similarities in the spatial analysis of perceived images.

Throughout the tone mapping pipeline, we assume the RGB color model where each channel is described by a positive floating point number. For the proper estimation of the simulated perceptual effects, the pixel intensity values in the HDR contents should be calibrated to $\frac{cd}{m^2}$. Such calibrated contents can be obtained using the photometric calibration procedure outlined in Appendix A from both standard and HDR cameras described in Section 2.4. In our implementation, we consider the values to be in the range from 10^{-4} to 10^8, which is sufficient to describe the luminance intensities perceivable by human vision. The algorithm produces tone mapped RGB floating point values in the range $[0\!:\!1]$ which are then quantized to 8-bit values by an OpenGL driver.

4.2.1 Tone Mapping

The algorithm proposed by Reinhard et al. [Reinhard et al. 2002] operates on the luminance values which can be extracted from RGB intensities using the standard CIE XYZ transform (Section 2.1). The method is a global operator, sigmoid scaling function, combined with a local dodging & burning technique that allows to preserve fine details as described in Section 2.5.3. The results are driven by two parameters: the adapting luminance for the HDR scene and the key value. The adapting luminance ensures that the global scaling function provides the most efficient mapping of luminance to the display intensities for given illumination conditions in the HDR scene. The key value controls whether the tone mapped image appears relatively bright or relatively dark. While the general background for this tone mapping operator has been given in Section 2.5, here we focus on a precise definition of the used method.
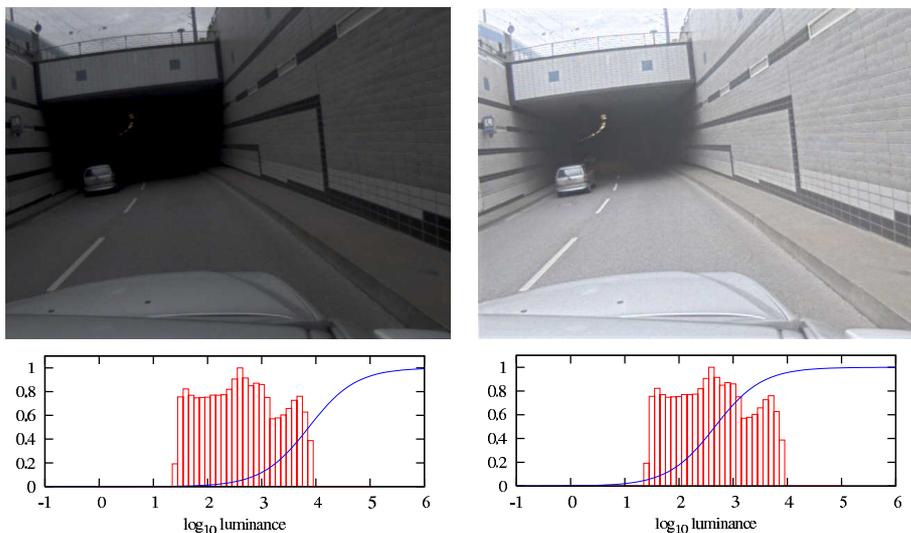
Figure 4.1: Tone mapping of an HDR image with a low key (left) and a high key (right). The curve on the histograms illustrates how the luminance is mapped to normalized pixel intensities.

In this algorithm, the source luminance values Y are first mapped to the relative luminance Y_r:

$$Y_r = \frac{\alpha \cdot Y}{\bar{Y}},$$

(4.1)

where $\bar{Y}$ is the logarithmic average of the luminance in the scene, which is an approximation of the *adapting luminance*, and α is the key value. The relative luminance values are then mapped to the displayable pixel intensities L using the following function:

$$L = \frac{Y_r}{1 + Y_r}.$$

(4.2)

The above formula maps all luminance values to the $[0:1]$ range in such way that the relative luminance $Y_r = 1$ is mapped to the pixel intensity $L = 0.5$. This property is used to map a desired luminance level of the scene to the middle intensity on the display. Mapping a higher luminance level to middle gray results in a subjectively dark image (low key) whereas mapping a lower luminance to middle gray will give a bright result (high key) (see Figure 4.1). The modulation of the key value in equation (4.1) with respect to the adapting luminance in the scene allows to simulate a relatively dark appearance of night scenes compared to bright day scenes. We explain our solution in Section 4.3.1.

The tone mapping function in equation (4.2) may lead to the loss of fine details in the scene with wide dynamic range due to the extensive contrast compression. Reinhard et al. [Reinhard et al. 2002] propose a solution to preserve local details by employing a spatially variant local adaptation value V in equation (4.2):

$$L(x,y) = \frac{Y_r(x,y)}{1 + V(x,y)}.$$

(4.3)

The local adaptation V equals to an average luminance in the surround of a pixel. The size of the surround, however, has to be carefully chosen. As explained in Section 2.5.2, larger areas guarantee good detail preservation, but a too large surround covering a high contrast edge will lead to well known inverse gradient artifacts, *halos*. To find an appropriate value of V for a pixel, the size of the surround is successively increased as long as it does not introduce any artifacts. For this purpose a Gaussian pyramid is constructed with successively increasing kernel:

$$g(x,y,s) = \frac{1}{\pi s^2} \cdot e^{-\frac{x^2+y^2}{s^2}}. \tag{4.4}$$

The spatial extent of the Gaussian kernel for the first scale is one pixel wide which is obtained with $s = (2\sqrt{2})^{-1}$. On each successive scale the spatial extent parameter s is 1.6 times larger. The Gaussian functions used to construct seven scales of the pyramid are plotted in Figure 4.2. As we later show, such a pyramid is very useful in introducing the perceptual effects to tone mapping.
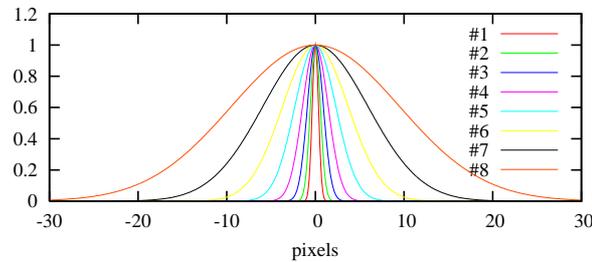
Figure 4.2: Plot of the Gaussian profiles used to construct the scales of the pyramid used for local dodging & burning in the tone mapping algorithm. The smallest scale is denoted as #1 and the largest #8. The plots are normalized by the maximum value for illustration purposes.

4.2.2 Temporal Luminance Adaptation

The luminance values in the HDR video can significantly change from frame to frame and cause unnatural brightness change in the tone mapping results. The human vision reacts to such changes through the temporal adaptation processes (Section 3.1.5). The time course of adaptation differs depending on whether we adapt to light or to darkness, and whether we perceive mainly using rods (during night) or cones (during a day). While several models have been introduced to computer graphics, it seems that it is not as important to faithfully model the process as to somehow account for it at all [Goodnight et al. 2003].

In the tone mapping algorithm chosen by us, the luminance adaptation can be modeled using the adapting luminance term in equation (4.1). Instead of using the actual adapting luminance $\bar{Y}$ for the displayed frame, a filtered value $\bar{Y}_a$ can be used. The value of $\bar{Y}_a$ changes according to the adaptation processes in human vision, eventually reaching the actual value if the adapting luminance is stable for some time. The process of adaptation can be modeled using an exponential decay function [Durand and Dorsey 2000]:

$$\bar{Y}_a^{new} = \bar{Y}_a + (\bar{Y} - \bar{Y}_a) \cdot (1 - e^{-\frac{T}{\tau}}), \tag{4.5}$$

where T is the discrete time step between the display of two frames, and τ is the time constant describing the speed of the adaptation process. The time constant is different for rods and for cones:

$$\tau_{rods} = 0.4\text{sec} \quad \tau_{cones} = 0.1\text{sec}, \tag{4.6}$$

thus the speed of the adaptation depends on the level of the illumination in the scene. The time required to reach the fully adapted state depends also whether the observer is adapting to light or dark conditions. The values in equation (4.6) describe the adaptation to light. For practical reasons the adaptation to dark is not simulated because the full process takes up to tens of minutes. Instead, we perform the adaptation symmetrically, neglecting the case of a longer adaptation to dark conditions.

4.2.3 Scotopic Vision

Human vision operates in three distinct adaptation conditions: scotopic, mesopic, and photopic (Section 3.1.3). The photopic and mesopic vision provide color vision, while in the scotopic range color discrimination is not possible because only rods are active. The cones start to loose their sensitivity at about $3.4\frac{cd}{m^2}$ and become completely insensitive at $0.03\frac{cd}{m^2}$ where the rods are dominant. We model the sensitivity of rods σ after [Hunt 1995] with the following function:

$$\sigma(Y) = \frac{0.04}{0.04 + Y}, \tag{4.7}$$

where Y denotes the luminance. The sensitivity value $\sigma = 1$ describes the perception using rods only (monochromatic vision) and $\sigma = 0$ perception using cones only (full color discrimination). The plot of equation (4.7) is shown in Figure 4.3.
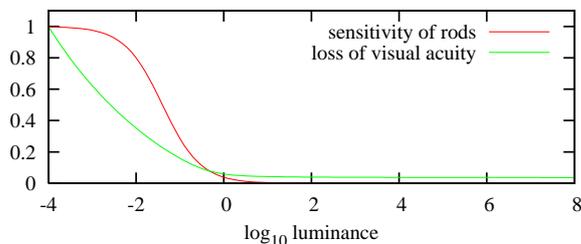
Figure 4.3: The influence of perceptual effects on vision depending on the luminance level. For details on rods sensitivity and visual acuity refer to Sections 4.2.3 and 4.2.4 respectively.

4.2.4 Visual Acuity

Perception of spatial details in human vision is not perfect and becomes limited with a decreasing illumination level. The performance of visual acuity is defined by the highest resolvable spatial frequency and has been investigated in [Shaler 1937]. [Ward et al. 1997] offer the following function fit to the data provided by Shaler:

$$RF(Y) = 17.25 \cdot \arctan(1.4\log_{10} Y + 0.35) + 25.72, \tag{4.8}$$

where *Y* denotes the luminance and *RF* is the highest resolvable spatial frequency in cycles per degree of the visual angle. The plot of this function is shown in Figure 4.4.
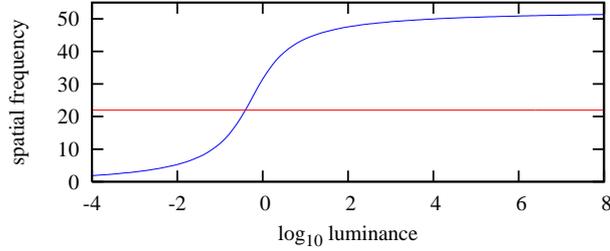
Figure 4.4: Plot of the highest resolvable spatial frequency for a given luminance level which illustrates the effect of loss of the visual acuity. Spatial frequency is given in cycles per degree of visual angle. The horizontal line marks the maximum displayable spatial frequency on a 15 inch LCD in typical viewing conditions.

To simulate the loss of visual acuity on a display device we need to map the visual degrees to pixels. Such a mapping depends on the size of the display, the resolution, and the viewing distance. For a typical observation of a 15 inch screen from half a meter at 1024×768 resolution we assume 45 pixels per 1 degree of the visual angle. It is important to note that the highest frequency possible to visualize in such conditions is 22 cycles per visual degree. Therefore, technically we can simulate the loss of visual acuity only for luminance below $0.5 \frac{cd}{m^2}$. The irresolvable details can be removed from an image by the convolution with the Gaussian kernel from equation (4.4) where *s* is calculated as follows [Ward et al. 1997]:

$$s_{acuity}(Y) = \frac{width}{fov} \cdot \frac{1}{1.86 \cdot RF(Y)}. \tag{4.9}$$

width denotes width in pixels and *fov* is the horizontal field of view in visual degrees. For typical observations the *width* to *fov* relation equals 45 pixels. We plot the profile of the kernel, according to equation (4.4), for several luminance values in Figure 4.5.

In Figure 4.3 we show the amount of lost visual acuity with respect to the luminance level. Apparently the loss of the visual acuity correlates with the increasing sensitivity of rods, and is therefore only present in monochromatic vision.
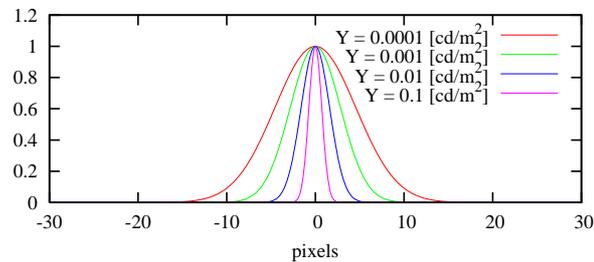
Figure 4.5: Plot of the profiles of the Gaussian kernels which can be used to simulate the loss of visual acuity at different luminance levels.

4.2.5   Veiling Luminance

Due to the scattering of light in the optical system of the eye, sources of relatively strong light cause the decrease of contrast in their vicinity – glare (Section 3.1.1). The amount of scattering for a given spatial frequency ρ under a given pupil aperture d is modeled by an ocular transfer function [Deeley et al. 1991]:

$$OTF(\rho,d) = \exp\left(-\frac{\rho}{20.9-2.1\cdot d}^{1.3-0.07\cdot d}\right),$$

$$d(\bar{Y}) = 4.9 - 3\tanh(0.4\log_{10}\bar{Y} + 1). \tag{4.10}$$

In a more practical manner the scattering can be represented in the spatial domain as a point spread function. In Figure 4.6 we show point spread functions for several adapting luminance levels, which were numerically found by applying the inverse Fourier transform to equation (4.10).
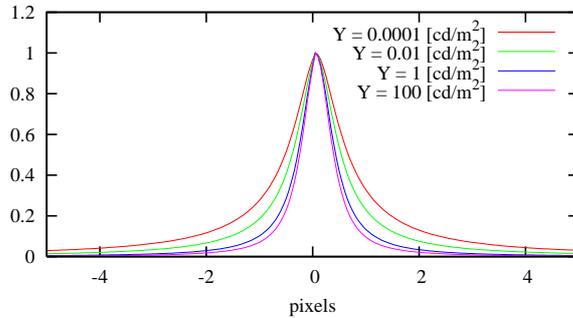
Figure 4.6: The point spread function illustrating scattering of light in the optical system of the eye for several adapting luminance levels.

Another model of the glare effect was introduced in computer graphics by Spencer et al. [Spencer et al. 1995]. They describe this phenomenon with four point spread functions linearly combined with three sets of coefficients for different adaptation conditions (scotopic, mesopic and photopic). Since their model is complex, and it is not obvious how to apply it in continuously changing luminance conditions, we decided to employ the model developed by Deeley at al. [Deeley et al. 1991], which describes the effect with one function that changes continuously for all adaptation levels.

4.2.6   Similarities in Spatial Analysis

Apparently, the visual acuity and the veiling luminance are based on the spatial analysis of an image modeled using the point spread functions. At the same time, a Gaussian pyramid is required to perform local tone mapping. Interestingly, convolution on particular scales corresponds to the convolution required to simulate visual acuity and glare at various luminance levels. This is an important observation which allows to model these effects by reusing the appropriate levels of the Gaussian pyramid without additional impact on the performance. The correspondence between the scales from the tone mapping (Figure 4.2) and the appropriate convolutions for visual acuity and veiling luminance are plotted in Figure 4.7.
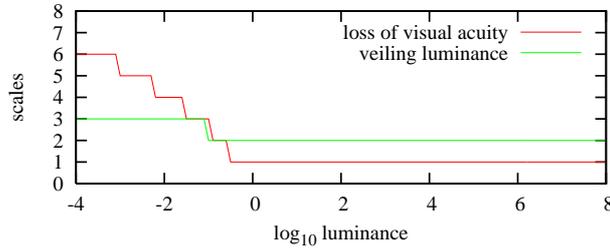
Figure 4.7: The correspondence between the scales from the tone mapping (Figure 4.2) and the appropriate convolutions for visual acuity (Figure 4.5) and veiling luminance (Figure 4.6).

The Gaussian pyramid constructed for the purpose of tone mapping contains only luminance values. This is sufficient to simulate the light scattering in the eye, but at first glance, visual acuity requires to perform the convolution on all three RGB channels. However, Figure 4.3 illustrates that the noticeable loss of visual acuity is present only in the scotopic vision where colors are not perceived. Since we simulate the loss of visual acuity combined with scotopic vision, we can simulate it using the luminance channel only.

4.3 Method

We present a method that successfully combines tone mapping with the effects mentioned in the previous section, which we implement in the graphics hardware for a real-time performance. We first show some of our improvements to the tone mapping method in terms of perceived brightness and luminance adaptation process and then explain technical details of our hardware implementation.

4.3.1 Key value

The key value, explained in Section 4.2.1, determines whether the tone mapped image appears relatively bright or dark, and in the original paper [Reinhard et al. 2002] is left as a user choice. In his follow-up paper [Reinhard 2002], Reinhard proposes a method of automatic estimation of the key value that is based on the relations between minimum, maximum and average luminance in the scene. Although the results are appealing, we feel this solution does not necessary correspond to the impressions of everyday perception. The critical changes in the absolute luminance values may not always affect the relation between the three values. This may lead to dark night scenes appearing too bright and very light too dark.

The key value, α in equation (4.1), takes values from $[0:1]$ range where 0.05 is the low key, 0.18 is a typical choice for moderate illumination, and 0.8 is the high key. We propose to calculate the key value based on the absolute luminance. Since the key value has been introduced in photography, there is no scientifically based experimental data which would provide an appropriate relation between the key value and the luminance, so the proper choice is a matter of experience. We therefore empirically specify key

values for several illumination conditions and interpolate the rest using the following formula:

$$\alpha(\bar{Y}) = 1.03 - \frac{2}{2 + \log_{10}(\bar{Y} + 1)}, \tag{4.11}$$

where α is the key value and $\bar{Y}$ is an approximation of the adapting luminance. The plot of this estimation is shown in Figure 4.8.
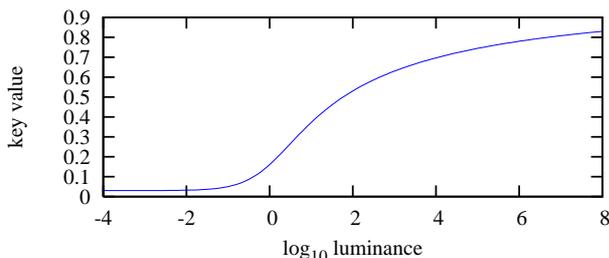
Figure 4.8: Key value related to the adapting luminance in a scene.

4.3.2   Temporal Luminance Adaptation

We model the temporal luminance adaptation based on equation (4.5). However, in our algorithm we do not perform separate computations for rods and cones, which makes it difficult to properly estimate the adaptation speed having two time constants τ_{rod} and τ_{cone} instead of one. To account for this, and still be able to correctly reproduce the speed of the adaptation, we interpolate the actual value of the time constant based on the sensitivity of rods (equation 4.7):

$$\tau(\bar{Y}) = \sigma(\bar{Y}) \cdot \tau_{rod} + (1 - \sigma(\bar{Y})) \cdot \tau_{cone}, \tag{4.12}$$

which we then use to process the adaptation value using equation (4.5).

4.3.3   Hardware Implementation

In order to perform tone mapping with perceptual effects, we need to compose three maps: a local adaptation map for the tone mapping, a map of visible spatial details to simulate visual acuity, and a map of light scattering in the eye for the glare effect. We will refer to these maps as *perceptual data*. Because different areas of these maps require different spatial processing, they cannot be constructed in one rendering pass. Instead, we render successive scales of the Gaussian pyramid and update the maps by filling in the areas for which the current scale has appropriate spatial processing. In the last step we use these three maps to compose the final tone mapped result.

Technically, we implement our tone mapping method as a stand-alone module, which can be added at the final rendering stage to any real-time HDR renderer or HDR video player. The only requirement is that the HDR frame is supplied to our module as a floating point texture, which can be efficiently realized using for instance *pixel buffers*. In addition to a texture which holds the HDR frame, our module requires the allocation of five textures for processing: two textures for storing adjacent scale levels, two for

Figure 4.9: Illustration of the rendering process for tone mapping which includes computing the local adaptation, visual acuity and glare. The input is an HDR frame with RGB channels and the output is a display. The blue boxes represent the texture data and yellow boxes represent rendering steps. The rendering steps marked by a gray rectangle are repeated for each scale to successively create the coarser scales of the Gaussian pyramid. After the rendering of each scale, the textures representing the perceptual data and the adjacent scales are swapped.

holding the previous and the current set of perceptual data (due to the updating process), and one intermediate texture for the convolutions. Since the three maps contain only luminance data, we can store them in a single texture in separate color channels.

The process of rendering the perceptual data is illustrated in Figure 4.9. We start with calculating the luminance from the HDR frame and mapping it to the relative luminance according to equation (4.1). We calculate the logarithmic average of the luminance $\bar{Y}$ in the frame using the down sampling approach described in [Goodnight et al. 2003], and apply the temporal adaptation process (equation 4.5). The map of relative luminance values constitutes the first scale of the Gaussian pyramid. At each scale of the Gaussian pyramid, we render the successive scale by convolving the previous scale with the appropriate Gaussian (equation 4.4). We perform the convolution in two rendering passes: one for the horizontal and one for the vertical convolution. To increase the performance we employ down-sampling, where the factor of down sampling is carefully chosen to approximate the kernel. Refer to Figure 4.10 for our choice of the scaling factors and the corresponding approximations of the Gaussian kernels from Figure 4.2. Having the current and the previous scales, we update the perceptual data on a per pixel basis in a separate rendering pass. The local adaptation is computed using the measure of the difference between the previous and the current scale as described in [Reinhard et al. 2002]. For the acuity map, we first estimate the proper scale for the luminance of the current pixel. If it falls between the previous and current scales, we interpolate the final value and update the map. In the other case the previous value is copied without change. The mapping from luminance to scale for visual acuity (Figure 4.7) is cached

in a look-up texture to skip redundant computations. We update the glare map in the same manner, with one difference: the appropriate scale for glare depends on the adapting luminance and is uniform for the whole frame so we supply it as a parameter to the fragment program. Before descending to the next scale of the Gaussian pyramid, the texture containing the current scale becomes the previous scale, and the texture with the current set of the perceptual data becomes the previous set.
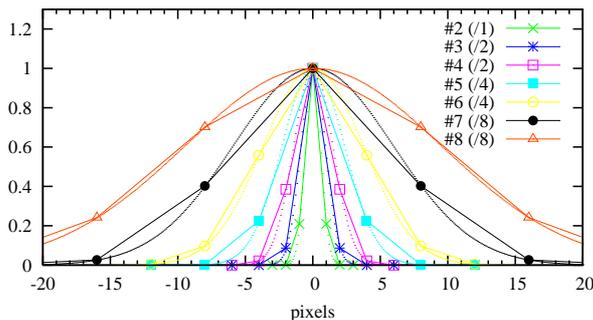
Figure 4.10: The effective approximation (solid lines) of the Gaussian kernels (dotted lines) from Figure 4.2 due to the down sampling. The values in parenthesis show the down sampling factor for each scale. For scale #1 we use the original image.

After descending to the lowest scale of the Gaussian pyramid, the perceptual data texture is complete. In the final rendering step, we tone map the HDR frame and apply the perceptual effects. For this, we use equation (4.3) from Section 4.2.1 in a slightly modified form to account for the loss of the visual acuity and the glare:

$$L(x,y) = \frac{Y_{acuity}(x,y) + Y_{glare}(x,y)}{1 + V(x,y)}, \tag{4.13}$$

where L is the final pixel intensity value, Y_{acuity} is the spatially processed luminance map that represents the visual acuity, Y_{glare} is the amount of additional light scattering in the eye, and V is the local adaptation map. Because the glare map in fact contains the relative luminance from the appropriate scale of the Gaussian pyramid, we estimate the additional amount of scattering in the following way to include only the contribution of the highest luminance:

$$Y_{glare} = Y_{gmap} \cdot \left(1 - \frac{0.9}{0.9 + Y_{gmap}} \right), \tag{4.14}$$

where Y_{gmap} denotes the glare map from the perceptual data.

We account for the last perceptual effect, the scotopic vision, while applying the final pixel intensity value to the RGB channels in the original HDR frame. Using the following formula, we calculate the tone mapped RGB values as a combination of the color information and the monochromatic intensity proportionally to the scotopic sensitivity:

$$\begin{bmatrix} R_L \\ G_L \\ B_L \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \cdot \frac{L \cdot (1 - \sigma(Y))}{Y} + \begin{bmatrix} 1.05 \\ 0.97 \\ 1.27 \end{bmatrix} \cdot L \cdot \sigma(Y), \tag{4.15}$$

where $\{R_L, G_L, B_L\}$ denotes the tone mapped intensities, $\{R,G,B\}$ are the original HDR values, Y is the luminance, L is the tone mapped luminance, and σ is the scotopic

Figure 4.11: The sample results of our method showing the simulated perceptual effects: glare (left) and scotopic vision with loss of visual acuity (right). The close-up in the right image inset shows the areas around the car in such way that their brightness match to illustrate the loss of visual acuity. The source HDR animation "Rendering with Natural Light" (left) courtesy of Paul Debevec.

sensitivity from equation (4.7). The constant coefficients in the monochromatic part account for the blue shift of the subjective hue of colors for the night scenes [Hunt 1995].

An alternative implementation of this tone mapping method, although without perceptual effects, was previously introduced in [Goodnight et al. 2003]. They propose a method to vectorize luminance which allows for efficient convolutions with large support kernels. However, we resigned from their approach due to the performance reasons – the real-time performance of this algorithm for a 512×512 frame is reached only when the computations are limited to two scales, which is not sufficient to introduce the perceptual effects. On the other hand, the down-sampling approach provides higher performance with sufficient accuracy of computations.

4.4 Results

We demonstrate our method in combination with an HDR video player. The player renders the compressed HDR video stream [Mantiuk et al. 2004] to a floating point texture, which is then processed as described in Section 4.3.3. The sample results of our method including the perceptual effects are shown in Figure 4.11. The left image depicts a computer generated scene in moderate lighting conditions with strong illumination coming from behind the trees in the background. Such a setup would evoke a glare effect in the real-world perception, which is not visible when pure local tone mapping is applied (left part). However, accounting for this perceptual phenomenon not only contributes to the realism of the rendered image but also increases a subjective impression of the dynamic range (right part). The right image shows a car driving scene in day light and at night. Two perceptual phenomena are typical to night illumination: the scotopic vision and the loss of visual acuity. Clearly, in the perceptual tone mapping of the night scene (right part), it is difficult to distinguish the colors and the overall brightness is low, which suggests the low illumination of the scene. The inset shows a close-up of the car with increased brightness for the night scene to illustrate

the simulated loss of visual acuity.

4.4.1 Dynamic Range Exploration Tool

Particularly in the context of HDR video [Mantiuk et al. 2004], in order to fully benefit from the HDR information encoded in such a video stream we additionally develop a convenient dynamic range exploration tool. The dynamic range exploration tool allows the user to view a selected range of luminance in a rectangular window displayed on top of the video (see Figure 4.12). The user can move the window interactively and choose which part of dynamic range should be linearly mapped to the display for closer inspection. When smaller range of luminances is chosen than the capabilities of the display, the tool linearly increases the contrast of contents in the exploration window.

On a technical level the dynamic range exploration tool simply by-passes the tone mapping process and displays linearly mapped part of the dynamic range in a selected window. Such a tool is, however, very convenient in the context of applications which require that the displayed data are exactly the same as in the original scene without any image processing. This is for instance required in medical and forensic applications.

Figure 4.12: Dynamic range exploration tool in form of a window that reveals verbatim details in HDR video contents.

4.4.2 Performance

We measured the performance of our method on a desktop PC with a Pentium4 2GHz processor and a NVIDIA GeForce 6800GT graphics card. We give the time-slice required for the tone mapping with our method at several frame resolutions in Table 4.1. In a configuration with an HDR video player, where additional time is required for the decompression of the HDR video stream, we were able to obtain the playback at 27Hz. It is important to note that the performance of our solution is scalable. If the time-slice required for our method is too long for a certain application, the number of rendered scales can be limited at the cost of local performance of the tone mapping and the accuracy of the visual acuity processing for very low illumination conditions.

The main bottleneck in the performance is caused by the amount of context switching required for the multi-pass rendering using the *pixel buffers* extension. The currently

	320x240	640x480	1024x768
8 scales	8ms (58Hz)	25ms (27Hz)	80ms (10Hz)
6 scales	7ms (62Hz)	21ms (30Hz)	66ms (12Hz)
4 scales	6ms (62Hz)	16ms (30Hz)	51ms (14Hz)

Table 4.1: Time-slice required for the display of an HDR frame using our method at several frame resolutions and several sizes of the Gaussian pyramid. In the parenthesis, we give the playback frame rate which we obtained with our method plugged to an HDR video player (note that the resolution also affects the frame decompression speed).

developed *frame buffer object* extension to OpenGL may provide an improvement, because it eliminates the need for such context switching between the rendering passes thus reducing the delays. Also, current OpenGL drivers do not implement linear interpolation of floating point textures during the up-sampling. Such an interpolation is crucial for the quality of the results and currently is implemented in the fragment program as an additional operation.

In relation to the previous tone mapping techniques which accounted for perceptual effects, our method has the following advantages: we employ a local tone mapping technique, the perceptual effects are applied locally depending on the luminance in a given area, and we make use of the apparent similarities in spatial analysis between the effects to provide a very efficient implementation. The importance of simulating the scotopic vision and the loss of visual acuity was noticed by Ferwerda et al. [Ferwerda et al. 1996]. However, they applied these effects only in the context of global tone mapping with uniform intensity over the whole image. This may lead to visible inaccuracies when a dark scene with an area of considerably brighter illumination is processed. In such an area, the loss of color would be unrealistic, and too low spatial frequencies would be removed there. This fact was noticed by Ward et al. [Ward et al. 1997] who proposed to apply the perceptual effects locally, still in combination with a global tone mapping method. Yet, in their work each of the effects has been treated separately and involved complex processing making it inapplicable to real-time processing. In the attempt to provide an interactive tone mapping solution, Durand et al. [Durand and Dorsey 2000] reverted to global application of the perceptual effects, which in fact had the same drawbacks as the [Ferwerda et al. 1996] model.

Our real-time implementation leads as well to several constraints. For instance only the tone mapping algorithms, which make use of the Gaussian pyramid, can be implemented in such an efficient combination with the perceptual effects. Therefore, our framework is not appropriate for several different approaches to tone mapping like decomposition into intrinsic images [Durand and Dorsey 2002] or contrast domain algorithms [Fattal et al. 2002]. Also, more complex functions for glare effect simulation may not benefit from our framework, if for instance their point spread functions cannot be approximated with the supplied Gaussian kernels.

4.5 Conclusions

In view of the increasing application of HDR images and video, we showed how to process such data in order to be able to render them on typical display devices with

a substantial dose of realism. We emphasize that it is not only necessary to reduce the contrast in such data, but it is also equally important to account for the perceptual effects which would appear in real-world observation conditions. Owing to the observation that the perceptual effects share similarities in the spatial analysis of the perceived image with the tone mapping algorithm, we were able to efficiently combine them into a stand-alone rendering module and reached real-time performance. The implementation of our method can be built upon any real-time rendering system which outputs HDR frames or any HDR video player. To demonstrate the importance of the account for perceptual effects, we plugged our method into an HDR video player leading to an enhanced realism of the displayed video. We improved the standard methods of simulation of the perceptual effects by applying them locally depending on the illumination in an area, and by providing smooth transition between different adaptation conditions. We envisage that in future the use of such a module will be standard in every real-time HDR renderer and HDR video player.

Chapter 5

Lightness Perception in Tone Mapping

When presenting tone mapped HDR images on display media, it is desirable to reproduce the appearance of corresponding real world (HDR) scenes. In Section 3.3, we have discussed five perceptual dimensions which define the appearance of a scene: brightness, lightness, colorfulness, chroma and hue. Intuitively, during the dynamic range compression of an HDR scene these dimensions should remain unchanged. While most of the tone mapping operators given in Section 2.5 do not change the hue, the change in brightness and colorfulness of a scene cannot be prevented because these qualities depend on the absolute amount of light energy. Consequently, the preservation of appearance requires careful reproduction of lightness and chroma during the tone mapping. Throughout this chapter we focus on analysis and reproduction of lightness.

Lightness is a perceptual quantity measured by the human visual system which describes the amount of light reflected from a surface normalized for the illumination level. Contrary to brightness, which describes a visual sensation according to which an area exhibits more or less light, the lightness of a surface is judged relative to the brightness of a similarly illuminated area that appears to be white. This leads to a similar appearance of perceived objects independently of the lighting and viewing conditions, which is known as lightness constancy [Palmer 1999]. The existence of lightness constancy enables the reproduction of appearance in tone mapping in which both lightning and viewing conditions change between the original and the reproduced scenes.

The lightness constancy achieved by the human visual system is not perfect and many of its failures appear in specific illumination conditions or even due to changes in the background over which an observed object is imposed [Gilchrist 1988]. It is well known that lightness constancy increases for scene regions that are projected over wider retinal regions [Rock 1983]. This effect is reinforced for objects whose perceived size is larger even for the same retinal size [Gilchrist and Cataliotti 1994]. The reproduction of HDR images on various media not only limits the luminance range but also introduces further constraints like a narrow field of view. Some failures of lightness constancy still appear in such conditions (simultaneous contrast for instance), but other effects, such as the Gelb illusion, are only observed if a scene covers the complete

field of view. The appearance of an HDR image cannot be correctly reproduced if a tone mapping operator does not take into account such phenomena. In Section 3.3.1, we have briefly reviewed several lightness perception theories which strive to explain how the human visual system perceives lightness in scenes. However, only recently the anchoring theory of lightness perception [Gilchrist et al. 1999] provides a sound explanation for an unprecedented number of perceptual experiments on lightness constancy and lightness constancy failures.

In this chapter, we investigate in detail the anchoring theory of lightness perception [Gilchrist et al. 1999] in the context of tone mapping. The principal concept of this theory is the perception of complex scenes in terms of groups of consistent areas – frameworks. Such areas, following the Gestalt theorists, are defined by regions of common illumination. The key aspect of image perception is the estimation of lightness within each framework through anchoring to the luminance perceived as white, followed by the computation of the global lightness. We derive a computational model for automatic decomposition of HDR images into frameworks which is based on the heuristics defined in the theory. We use the model in a tone mapping operator which predicts lightness perception of the real world scenes and aims at its accurate reproduction on low dynamic range displays. Furthermore, we observe that a decomposition into frameworks opens new grounds for local image analysis in view of human perception.

5.1   Previous Work

A number of tone mapping operators is to a certain extent influenced by theories of perception of brightness and lightness. Initially, the algorithms were based on the power-law relationship between the brightness and the corresponding luminance, as proposed in [Stevens and Stevens 1960]. The main objective was to preserve a constant relationship between the brightness of a scene perceived on a display and its real counterpart for any lighting condition. Implementations of this approach were presented in [Tumblin and Rushmeier 1993] (Stevens law) and in [Drago et al. 2003] (Weber-Fechner law). Further attempts in lightness reproduction lead to direct application of the Retinex theory [Land and McCann 1971] to tone mapping. Jobson et al. [Jobson et al. 1997] proposed a multi-resolution Retinex algorithm for luminance compression, which unfortunately leads to halo artifacts for HDR images along high contrast edges. Inspired by the lightness perception model based on contrast integration, Fattal et al. proposed a successful gradient domain tone mapping operator [Fattal et al. 2002]. The concept of intrinsic images [Barrow and Tenenbaum 1978, Arend 1994] to separate the illumination and reflectance (detail) layers inspired many algorithms. The idea was first implemented in [Tumblin et al. 1999] where it was assumed that these layers were explicitly provided which is the case only for synthetic images. Later, several methods for an automatic layer separation have been introduced. The LCIS operator [Tumblin and Turk 1999] separates the image into large scale features (presumably illumination) and fine details. A much better separation has been achieved using the bilateral filter [Durand and Dorsey 2002].

Evidently, perception theories have been inspiring tone mapping algorithms to a certain extent. Although the early operators are based on simple theories which do not account well for lightness in complex scenes, the new algorithms build upon intrinsic images

and contrast models which are very advanced theories. The separation into illumination and reflectance presents a convenient image processing tool for detail preserving dynamic range compression. However the theory remains vague about how to map luminance to lightness in such separated layers so that the scene appearance is well preserved. Moreover, the intrinsic image models fail to address the apparent failures in lightness constancy [Gilchrist et al. 1999]. On the other hand, the algorithms inspired by Retinex and contrast theories focus on optimizing brightness relations and leave the aspect of lightness mapping to the user. The resulting images look as if enhanced through numerical optimizations rather than having the original appearance reproduced. Overall, many questions in the area of appearance reproduction remain open, motivating us to develop a computational model of the currently most advanced theory of lightness perception and to examine it with analysis in the context of natural scenes.

5.2 Anchoring Theory Of Lightness Perception

The anchoring theory of lightness perception by [Gilchrist et al. 1999] is qualitatively different from other recent lightness models and is based on a combination of global and local anchoring of lightness values. In the following sections we explain the main concepts of this theory. First, we discuss the estimation of lightness within the simple scenes (background/patch stimuli) using the anchoring rules. Next, we explain how to extend the anchoring of lightness to complex scenes using frameworks.

5.2.1 Anchoring Rule

In order to relate luminance values to lightness, it is necessary to define at least one mapping between the luminance value and the value on the scale of perceived gray shades – *the anchor*. The anchor cannot be defined once for absolute luminance values, because each luminance level can be perceived as any shade of gray depending on the observation conditions. It therefore must be tied to a measure of relative luminance. Two such measures are commonly used for anchoring: the average luminance rule and the highest luminance rule. Once the anchor is defined for the scene, the lightness value for each luminance value can be estimated by the luminance ratio between the value and the anchor. This mapping is referred to as *scaling*. Although usually a veridical scaling is assumed, the compression or expansion of the range is possible if necessary.

The average luminance rule derives from the adaptation-level theory [Helson 1964] and states that the average luminance in the visual field is perceived as middle gray. Thus the relative luminance values in a scene should be anchored by their average value to middle gray. This assumption was later commonly adopted in tone mapping techniques [Ferwerda et al. 1996, Tumblin et al. 1999, Pattanaik et al. 2000, Reinhard et al. 2002].

The highest luminance rule initially defined the anchor as a mapping of the highest luminance in the visual field to a lightness value perceived as white. However, the evident perception of self-luminous surfaces (lighter than white) leads to an extended definition. According to [Li and Gilchrist 1999] there is a tendency of the highest luminance to appear white and a tendency of the largest area to appear white. When the highest luminance covers the largest area, the highest luminance becomes a stable

anchor and is mapped to white. However, if the darker area becomes larger, the highest luminance starts to be perceived as self-luminous and the anchor becomes a weighted average of the luminance proportionally to the occupying area.

The experimental evaluation of the average luminance rule versus the highest luminance rule was presented in [Li and Gilchrist 1999]. In this study, the visual field of the observers was limited to a large acrylic hemisphere with one half painted matte black and the other half painted middle-gray. The experiment was conducted in isolated conditions to prevent the uncontrolled influence of other stimuli. Li and Gilchrist reported that the middle-gray half was seen by the observers as fully white, while the black half was seen as dark gray. Additionally, when the black area was increased and became considerably larger than the middle-gray area, the perceptual effect of self-luminosity for the middle-gray part was reported. Other findings, based on Mondrians [Palmer 1999], which are more complex stimuli, agree with these conclusions [Gilchrist and Cataliotti 1994]. The experimental evidence decisively favors the highest luminance rule over the average luminance rule.

5.2.2   Complex Images

The anchoring rule, described in the previous section, cannot be applied directly to complex images. Instead, [Gilchrist et al. 1999] introduce the concept of decomposition of an image into segments, *frameworks*, in which the anchoring rule can be applied directly. In their theory, following the Gestalt theorists, frameworks are defined by regions of common illumination. For instance, all objects being under the same shadow would constitute a framework. Additionally, proximity is also considered as a grouping factor. A real-world image is usually composed of multiple frameworks.

Framework regions can be organized in an adjacent or a hierarchical way and their areas may overlap. Additionally, the whole scene constitutes an additional *global framework* with its global anchor. The lightness of a target in a scene is computed according to the anchoring rule in each framework. A target in a complex scene that belongs to more than one framework, may have different lightness values when anchored within different frameworks. According to the model, the net lightness of a given target is predicted as a weighted average of its lightness values in each of the frameworks in proportion to their strength and with a certain constant level of influence of the global framework. The strength of a framework is mainly determined by its size and the variety of luminance values it contains – *articulation*. Frameworks with lower variance or smaller sizes have less influence on the net lightness.

5.3   Computational Model

The anchoring theory has been presented without a formal model. On a technical level this requires the development of a method for automatic segmentation of an image into frameworks, to build heuristics estimating the influence of each framework on total lightness, and to estimate the anchors within the frameworks. Furthermore, the algorithm must perform accurately when used with natural scenes.

The presented model takes an image with relative luminance values as an input. Such values can be computed from RGB channels of an HDR image according to the CIE

XYZ luminous efficiency functions. We first segment the input image into the overlapping frameworks and define the probabilities with which pixels belong to each framework. Next, we estimate the anchor in each framework and finally calculate the net lightness in the scene.

5.3.1 Decomposition into Frameworks

We define the framework as a probability map over the whole image in which each pixel is assigned a probability of being part of that framework. The framework, in order to be valid, must be defined by a number of pixels which belong to this framework with a very high probability, for instance above 95%. In principle, the task of segmentation into frameworks is to identify luminance values which potentially represent a common illumination and to assign to each pixel for each framework the probability that the pixel is under considered illumination. Due to the lack of explicit information about the distribution of illumination, we assume here that the contrast range typical to everyday situations is wide enough to allow us to identify such separate illumination areas based on luminance. It is for example possible to identify shadowed areas on a sunny day, dim interior of a room with a window view, street light illumination in a night scene, and similar.

Initially, we have experimented with several segmentation algorithms in order to find a plausible decomposition into frameworks. The mean shift segmentation [Comaniciu and Meer 2002] has produced the most appropriate results. However, segmentation algorithms in general assign a pixel to only one segment and therefore do not implement the notion of probability of belonging to a segment. A border between two frameworks which might occur on a smooth gradient in the image is in such a situation impossible to represent correctly. We therefore decide to tailor a custom decomposition method.

As mentioned before, our method is based on the luminance intensities in the HDR image. We start with the standard K-means clustering algorithm to find the centroids that provide an appropriate segmentation of the HDR image into frameworks. We operate on a histogram in the $\log_{10}$ of luminance. We initialize the K-means algorithm with values ranging from the minimum to maximum luminance in the HDR image with a luminance step equal to one order of magnitude and we execute the iterations until the algorithm converges. Upon convergence, we remove centroids representing empty segments.

Given the centroid values, we initially assign the probability values based on the difference between the pixel value and the centroid. We model such an attribution to the centroid with a Gaussian function:

$$P_i(x,y) = e^{\frac{-(C_i - Y(x,y))^2}{2\sigma^2}}, \tag{5.1}$$

where P_i represents the probability map for framework i, C_i is the centroid for that framework, Y denotes the luminance of the HDR image (both C_i and Y are in the $\log_{10}$ space), and the variance σ equals to the maximum distance between adjacent centroids. The attribution values are normalized to correctly represent the probabilities.

Often at this stage, pairs of centroids may represent similar frameworks – all pixels in an image belong with a similar probability to both of them. We iteratively merge

the centroids whose probability maps differ by less than 20% on average, and the new
centroid value is equal to the weighted average of both proportionally to their area:

$$C_{i,j} = \frac{C_i \cdot S_i + C_j \cdot S_j}{S_i + S_j}, \tag{5.2}$$

where C_i and C_j are the values of the too similar centroids, and S_i and S_j denote the
number of pixels clustered to these centroids. After the merge, probability values are
recalculated according to (5.1), and the iteration is repeated until no centroids need to
be merged.

As the next step, we spatially process the probability map of each framework to include
the proximity aspect of Gestalt grouping factors. The spatial processing smoothes local
variations in the probability values which may appear due to textured surfaces. High
local variations, however, cannot be smoothed because they may define the outline of
objects or frameworks. The bilateral filter [Tomasi and Manduchi 1998] is an appro-
priate image processing tool for this purpose. We filter the probability map of each
framework with a bilateral filter in which the range variance is set to 0.2 and the spatial
variance to 17 pixels.

(a) luminance channel of an HDR   (b) intensity based frameworks   (c) spatially processed frameworks
image
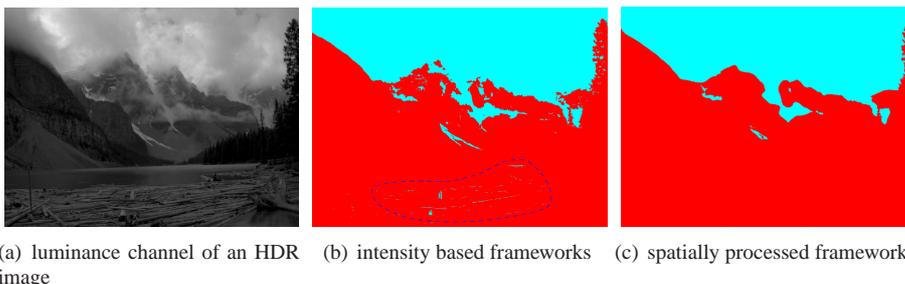
Figure 5.1: The decomposition of an HDR image into frameworks before and after
spatial processing. Notice the artifacts in the intensity based frameworks decomposi-
tion (marked area) which are corrected after the spatial processing. The HDR image
appears dark because it has been exposed for details in the clouds. The original image
courtesy of Greg Ward.

(a) initial centroids   (b) probabilities for initial cen- (c) valid centroids with probabili-
                            troids after K-means convergence   ties
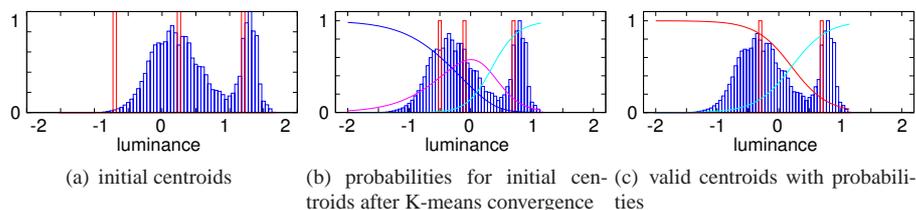
Figure 5.2: The histogram of the HDR image from Figure 5.1(a) illustrating the esti-
mation of centroids which provide an appropriate decomposition into frameworks. In
the middle and right histograms the probability distributions are shown for each frame-
work. The maxima of the probabilities do not always match the centroids due to the
normalization.

We demonstrate the decomposition procedure on an example HDR image shown in
Figure 5.1 with details of decomposition in Figure 5.2. First, we converge the initial

segmentation to identify luminance values that would represent the most accurate decomposition into frameworks, Figure 5.2(a). We then calculate the probabilities and merge one centroid which does not represent a valid framework, Figures 5.2(b), 5.2(c). The final centroids define the probability maps based only on the luminance property and contain several incorrect assignments visible in Figure 5.1(b). The reflections on the logs at the bottom of the image are incorrectly assigned to a framework which mainly contains clouds. We refine the probability maps to include also spatial interactions as shown in Figure 5.1(c).

5.3.2 Strength of Frameworks

Apart from the illumination conditions, the strength with which a framework influences the lightness of a given surface also depends on the articulation of the framework and its relative size. If a framework is highly articulated, the pixels tend to be strongly anchored within this framework. Also, large frameworks have a higher influence on the lightness than small ones. Strength of a framework is defined as the product of the articulation and size factors.

We estimate the articulation factor independently for each framework based on the mean contrast calculated as a standard deviation of logarithmic luminance in the framework. The articulation factor A_i is the mean contrast in the framework divided by the mean contrast in the image. Thus the frameworks which contain a larger part of the contrasts in the image have a stronger influence on the lightness.

Similarly, a larger framework will have a tendency to have a higher influence on the lightness of surfaces, while a relatively small framework will have a rather limited impact. We estimate the size factor in the following way:

$$X_i = 1 - e^{\frac{-(S_i)^2}{2 \cdot 0.03^2}}, \tag{5.3}$$

where X_i denotes the size factor of the framework i, and S_i represents the normalized relative area of a framework. Here we attenuate the influence of frameworks with a size below 10% of the total image area. The 10% value is chosen arbitrarily and can be modified if necessary. Although, it is not as important to derive a precise number as to include the factor at all. The lack of the factor may lead to an excessive influence of unimportant frameworks on the net lightness estimation.

We apply the strength factor to the frameworks by multiplying their probability maps P_i by their respective articulation factor A_i and size factor S_i. We then normalize the probability maps again and obtain the final result of the decomposition into frameworks. Most of the time, all frameworks in an image will have a similar articulation. Sometimes however, a uniform area like a background may constitute a framework due to its unique illumination. Articulation prevents such a background framework to play an important role in the computation of the net lightness by minimizing the local anchoring of pixels to this framework in favor of other frameworks. In an extreme situation, when all frameworks have minimum articulation, the framework with the highest anchor is assigned a maximum articulation, thus imposing the global anchoring in the image.

5.3.3   Estimation of Anchor

After the HDR image is decomposed into frameworks, we estimate the anchor within each framework. Since we employ the highest luminance rule, we need to find the luminance value that would be perceived as white in case a given framework would be observed as stand-alone.

(a) anchoring for framework #1   (b) anchoring for framework #2   (c) net lightness
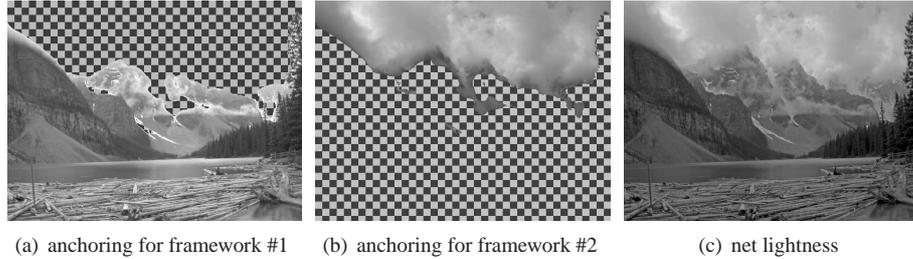
Figure 5.3: Local anchoring in the frameworks and the net lightness calculated as described in Section 5.3.4.

Although we apply the highest luminance rule, we cannot directly use the highest luminance in the framework as an anchor. As discussed in Section 5.2.1, there is a relation between what is perceived as white and its area relatively to the surround. This implies that a spatial filtering is required prior to the estimation of the local anchor. Our procedure is to filter the area of a framework with a medium sized Gaussian kernel to suppress plausible small areas with high luminance. We thus eliminate potential self-luminous areas, allowing us to take the highest luminance of the rest of the pixels as the anchor. In Figure 5.3 we show the two frameworks identified in the example HDR image with their lightness computed according to the local anchor.

Alternatively, the self-luminance areas could be identified using image processing tools which remove highlights. Numerous algorithms are available as for instance [Schluens and Koschan 2000, Wesolkowski et al. 2001]. However, we have noticed throughout our experiments that such algorithms can at most remove specular highlights in HDR images. The direct light sources, which are also self-luminous, or larger highlights are left in the image. On average these methods performed less robustly. Therefore we excluded them from consideration.

Ultimately, it would be interesting to know which luminance in each framework is assumed to be white by an average human observer. Interestingly, the perceptual evidence of preferred brightness adjustment in images may be helpful here. In an experiment, [Yoshida et al. 2006] asked subjects to adjust the contrast and brightness to match their preference in a number of LDR images (an LDR image has only one framework). While the preferred brightness and contrast adjustments are totally different between subjects, they observed that after the adjustment all subjects align histograms along a very similar luminance value depicting white. Apparently, the subjects performed the anchoring to white. Therefore, if one can build a model which predicts the preferred brightness of images, indirectly this could also be used as the model to predict anchoring to white. In the recent evaluation of such models [Krawczyk et al. 2007a], the preferred brightness in a set of 33 images is best predicted by a combined anchoring to white, middle gray, and black, which in effect maps to 60% reflectance. The predictions based on image processing methods directly applying anchoring to white

were less accurate. Since the anchoring to white has been confirmed in the original experiment, the lower accuracy is most probably caused by imprecise estimation of luminance perceived as white.

5.3.4 Net Lightness

Given the decomposed frameworks and estimated local anchors we compute the net lightness of the pixels by merging the frameworks. We process each framework individually. We sum the original luminance values of the HDR image normalized by the locally estimated anchor value and proportionally to the probability map:

$$L(x,y) = 30\% \cdot \sum_i (Y - W_i) \cdot P_i(x,y) + 70\% \cdot (Y - W_0), \tag{5.4}$$

where L denotes the final lightness value, Y the original luminance of the HDR image, W_i the local anchor of framework i, W_0 the anchor in the global framework (all these values are in the $\log_{10}$ space), and P_i is the probability map. The 30% and 70% coefficients for local and global anchor influence respectively are arbitrarily suggested in [Gilchrist et al. 1999] and can be modified if necessary. In Figure 5.3 we illustrate how the net lightness has been computed for the sample HDR image. A comparison of the net lightness result to the original HDR image in Figure 5.1 illustrates an improved perception of image contents in the processed image.

5.4 Model Analysis

The main focus of this chapter is the computational model of the lightness perception theory applied to the tone mapping of HDR images. A thorough verification of the presented model would require a psychophysical experiment which is beyond the scope of this thesis. Instead, we test our computational model by simulating two experiments related to the perception of lightness. The first one analyzes the accuracy of the decomposition into frameworks for natural scenes and the second experiment is a simulation of the Gelb illusion using various lightness mapping algorithms for HDR images (tone mapping operators).

5.4.1 Frameworks within Multi-Illuminant Scenes

According to the anchoring theory of lightness perception, successfully identified frameworks should define the areas in which the lightness is perceived homogeneously [Gilchrist et al. 1999]. The evidence for such lightness perception can be obtained through a distribution of probe disks of constant known luminance value across an image. The disks should have the same lightness within a framework independently of the ratio of their luminance to the background luminance on which they are placed.

Such an experiment has recently been presented by Gilchrist and Radonjic [Gilchrist and Radonjic 2005]. In Figure 5.4 we provide an HDR reproduction of this experiment using an image similar to the original material. We decompose the HDR image into frameworks using our computational model and place several probe disks of constant

(a) probe discs in natural scene   (b) probe discs   (c) identified frameworks
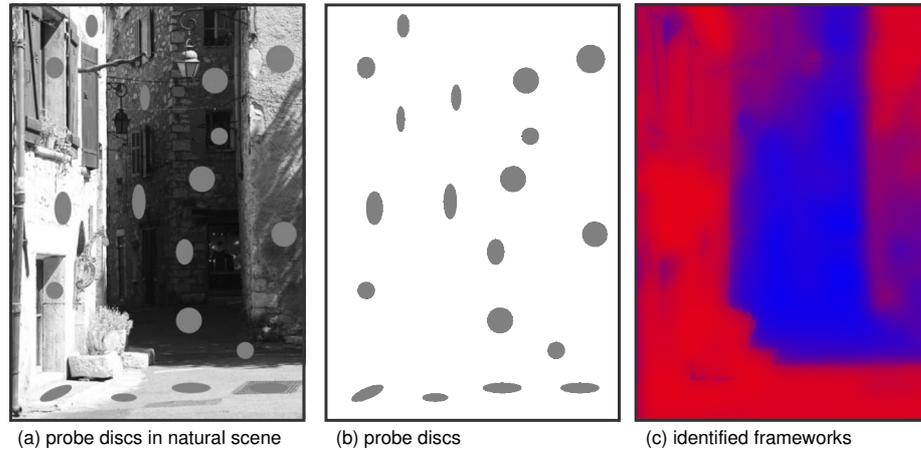
Figure 5.4: The probe discs of constant luminance (middle) are inserted to a multi-illuminant natural scene (left). The perceived brightness of the probes changes in the context the scene, but is constant within the two identified frameworks (right) and is independent of the local background luminance. The contrast ratios of the probes with the background range in the shadow framework from 1:2 to 1:9 (disc is an increment) and in the light framework from 9:1 to 2:1 (disc is a decrement).

luminance value in various areas of the image. The contrast ratios between the probe disks and the background in both shadow and light frameworks range from 1:2 to 1:9. The lightness of the probes is perceived consistently within the area of the frameworks independently of the background. These leads to the same conclusions as in the original experiment: lightness is determined by the frameworks and the influence of local contrasts is minimal. Our contribution here is not to confirm the theory, but to provide an automated method for obtaining an appropriate decomposition. In this sense, the reproduction of the experiment serves as the evidence that frameworks areas are accurately identified using our computational model.

5.4.2 Anchoring in the Gelb Illusion

The Gelb Effect is a well known illusion which provides a good example of lightness constancy failure [Gilchrist et al. 1999]. In the illusion, one observes perceptual darkening of a surface despite its constant reflectance and constant illumination. The failure is caused by the appearance of new brighter surfaces in the scene. The illusion can be reproduced in a darkroom with low ambient light using several patches of gray paper with a different reflectance. A single beam of light should first illuminate only the darkest paper, which will appear to be white. Placing a bit brighter paper beside the existing one causes perceptual darkening of the darkest paper which has initially appeared to be white. Each time a brighter paper is added to the scene, it becomes white and all others immediately become darker. This perceptual illusion can by definition be attributed to the anchoring in general and to the highest luminance rule in particular. It can neither be explained with the contrast theories because the papers do not have to be placed adjacent to each other [Gilchrist et al. 1999], nor with intrinsic image models because the illumination does not change. Furthermore, if the scene occupies only a part of the

visual field, like a tone mapped image observed on a display, the perceptual darkening will not appear because other visible surfaces may serve as a white reference. Therefore it needs to be reproduced during tone mapping to preserve the appearance of the original scene.

(a) scene setup for Gelb illusion   (b) decomposed frameworks   (c) illumination layer
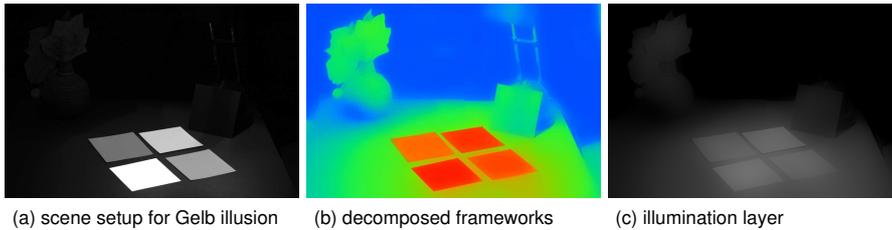
Figure 5.5: Photograph of the scene in which the Gelb effect can be observed (left). The middle image shows the decomposition into frameworks obtained from our model – red, green, and blue define the distinct frameworks, yellow marks the shared influence of the red and the green frameworks. The right image shows the illumination layer obtained with the bilateral filtering which is used in the intrinsic images model.

We have performed a study of this experiment to validate the results of our algorithm. For comparison, we chose two other methods whose principle goal includes the reproduction of appearance of the original image. The photographic tone reproduction algorithm presented by Reinhard et al.[Reinhard et al. 2002] is based on a sigmoid function and follows the rule of anchoring to middle-gray. The fast bilateral filtering presented by Durand and Dorsey[Durand and Dorsey 2002] is inspired by the theory of intrinsic images. We will refer to the first one as the *middle-gray anchoring* and to the latter as the *intrinsic images model*. In the study, we used four HDR captures of exactly the same scene setup, showing from one to four patches with progressively increasing maximum reflectance. The relative reflectance of the patches was respectively equal to $39\%, 56\%, 72\%$, and 100% with the reference to the brightest one. The area, in which we showed the patches, was illuminated from the top and in our conditions the Gelb illusion was reproduced. A photograph of the setup with all the patches visible is shown in Figure 5.5.

The results of tone mapping of the four HDR images are shown in Figure 5.6 and the respective reproductions of lightness of the patches are plotted in Figure 5.7. All tone mapping methods reveal the objects placed outside of the main illumination that are not visible in a standard photograph in Figure 5.5. The intrinsic images model maps the lightness of the patches in each of the images to an approximately constant value and maintains the overall brightness of the scene background constant. This is in accordance with the lightness constancy rule, but contrary to what was observed in the real setup. The middle-gray anchoring reproduces the perceptual darkening of the patches, however the brightest one is mapped to white only when all four patches are visible. Further, each brighter patch causes the darkening of scene background which was originally not observed. The lightness perception model presented in this chapter reproduces both the Gelb illusion on the patches and holds the lightness constancy of the objects in the scene background.
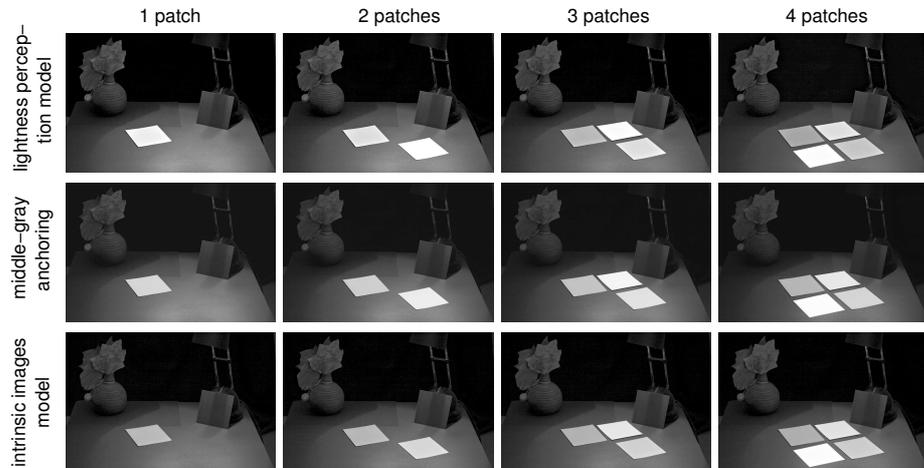
Figure 5.6: Simulation of the Gelb Effect by three tone mapping methods. The mapping of lightness by each of the tone mapping is plot in Figure 5.7. The *intrinsic images model* refers to [Durand and Dorsey 2002] operator, the *middle-gray anchoring* to [Reinhard et al. 2002], and *lightness perception model* to the operator presented in this chapter.

Analysis

The decomposition of the scene into frameworks (shown in Figure 5.5) in the lightness perception model permits the processing of patches and the rest of the scene separately. The estimation of local anchors using the highest luminance rule estimates the appearance of patches in accordance with the observations in the original conditions. The net lightness calculation with the influence of a global anchor maintains the brightness relation between the frameworks.

The lightness constancy of the intrinsic images model can be explained as follows. In the illumination layer (shown in Figure 5.5), obtained by processing the original HDR image with the bilateral filter, the brightness of each patch is approximately equal while the actual differences are in the reflectance layer. The tone mapping reduces the dynamic range of the illumination layer and overlays the unmodified reflectance layer. Since the intensities in the illumination layer do not significantly change between the four images, the lightness mapping is constant. Therefore, neither the average luminance rule nor the highest luminance rule applied to the illumination layer could reproduce the Gelb illusion. The application of the highest luminance rule to the reflectance layer or to the final tone mapping result could reproduce the darkening of the patches, however it would also cause the undesired darkening of other image parts.

The middle-gray anchoring reproduces the darkening of the patches because the addition of a new brighter patch causes change in the average luminance of the scene. When the average luminance increases in a new image, the patches, which have constant luminance, are mapped to darker gray shades. Unfortunately, such a global connection causes the overall darkening of the scene which is not expected, and the brightest patch is mapped to white only when all four patches are present in the scene.
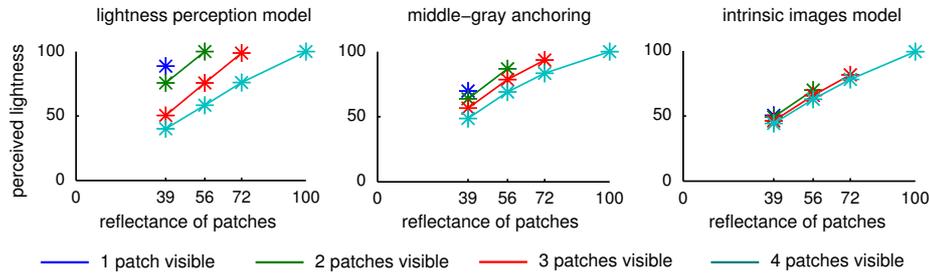
Figure 5.7: The plots illustrate how the mapping of luminance of the patches to lightness changes between the four images in case of each of the three tone mapping operators. On the scale of lightness, the value 100 maps to white and 0 to black. Refer to the corresponding images in Figure 5.6.

5.5 Applications

The computational model of lightness perception theory opens new possibilities in processing of HDR images. As a direct application we present a tone mapping operator which aims at the reproduction of lightness as closely as possible to the lightness perceived in a natural scene. Besides, decomposition into frameworks defined by homogeneous illumination gives an interesting possibility to perceptually supported image processing.

5.5.1 Tone Mapping

Based on the computational model of lightness perception, we derive a tone mapping algorithm for contrast reduction in HDR images. The algorithm takes as input an HDR image defined by floating point RGB values that are linearly related to luminance and produces a displayable LDR image as a result. The contrast reduction process is based on the luminance channel.

The main technical goal of tone mapping is to reduce the contrast of the original HDR image. While through the net lightness computation the dynamic range in the image is reduced, it may still exceed the capabilities of the display device. Hence, an additional dynamic range reduction may be necessary to achieve good results. For the purpose of tone mapping we use a modified version of net lightness computation (5.4) which includes a dynamic range reduction (scaling) by a factor D_i:

$$L(x,y) = 70\% \cdot \sum_i \frac{(Y - W_i)}{D_i} \cdot P_i(x,y) + 30\% \cdot (Y - W_0). \tag{5.5}$$

The value of D_i can be set individually for each framework in such way that it scales down the dynamic range of a framework if it exceeds the capabilities of the target display device. Also the influence of the global framework has been limited, because it counteracts the luminance range compression goal.

We first calculate the luminance from the RGB colors using the CIE Yxy color space and segment the input scene into the frameworks. Next, we estimate the anchor in each framework, i.e. the luminance value perceived as white. We then compute the local

pixel lightness within each framework. Finally, we calculate the net lightness of each pixel using equation (5.5). We recover the color information with an inverse CIE Yxy transform, using the computed lightness L instead of luminance channel Y. The result is suitable to be viewed on an LDR display device.

To illustrate the performance of our tone mapping algorithm, we have selected several HDR images which contain various distinct illumination features. We present our results in Figure 5.8, where each row contains the source HDR image, the tone mapping result, and a map of framework areas. The decomposition into frameworks obtained using our model correlates with the intuitive impression of which areas have common illumination: daylight, shadow, desk lamp, room interior, exterior, etc. The separation performs well even in presence of occluders like the grating structure of the window pane. The extracted frameworks are plausible despite the lack of semantic information, which might seem to be necessary to perform a successful decomposition. In the results shown here, the scaling D_i has not been necessary, however we have sometimes reduced the influence of the global framework down to 10% for images with a particularly high dynamic range. Interestingly, the images tend to be decomposed only into two or three frameworks, although there is no restriction on the number.

One important issue is that Gilchrist's model generally assumes approximately diffuse surfaces and if self-luminous areas exist, they occupy a limited field of view. In our application we use this theory for complex scenes beyond what has originally been tested but we do not observe any problems invalidating our approach. To our knowledge, perceptual models of lightness perception able to deal with natural scenes, which contain large self-luminous surfaces, do not exist.

The evaluation of aesthetic properties of this tone mapping can be done with recently presented methodology [Ledda et al. 2005, Yoshida et al. 2005]. In Figure 5.9, we provide analysis of how the luminance values are mapped to the lightness levels in three different tone mapping techniques: the presented lightness perception model, the global version of photographic tone reproduction [Reinhard et al. 2002] which is a sigmoid mapping function, and the fast bilateral filtering [Durand and Dorsey 2002] which is inspired by the intrinsic images model. The technical quality of a tone mapping algorithm can be measured by the efficiency in use of the available limited dynamic range on a display device. In our example, the global operator performs a strictly monotonic reproduction, thus leads to the loss of fine details as explained in Section 2.5. The local adaptation using the bilateral filtering enables a more efficient use of the available dynamic range. The preservation of details can be observed as a deviation from the monotonic mapping of luminance. However, the tone mapping using frameworks permits to break this monotonicity and perform partially independent mapping of luminance in two distinct image areas, resulting in an even more efficient use of the dynamic range. The mapping within the frameworks is not uniform because of the varying influence of the global framework. We further evaluate the properties of the lightness perception tone mapping in Chapter 6, where we analyze it in terms of communication of contrast in images.

5.5.2   Local Image Processing

Image processing algorithms are usually applied with uniform parameter settings over the whole image. When the algorithm is localized, the parameters for the method are

source HDR image (linear mapping)	lightness perception tone mapping	decomposition into frameworks
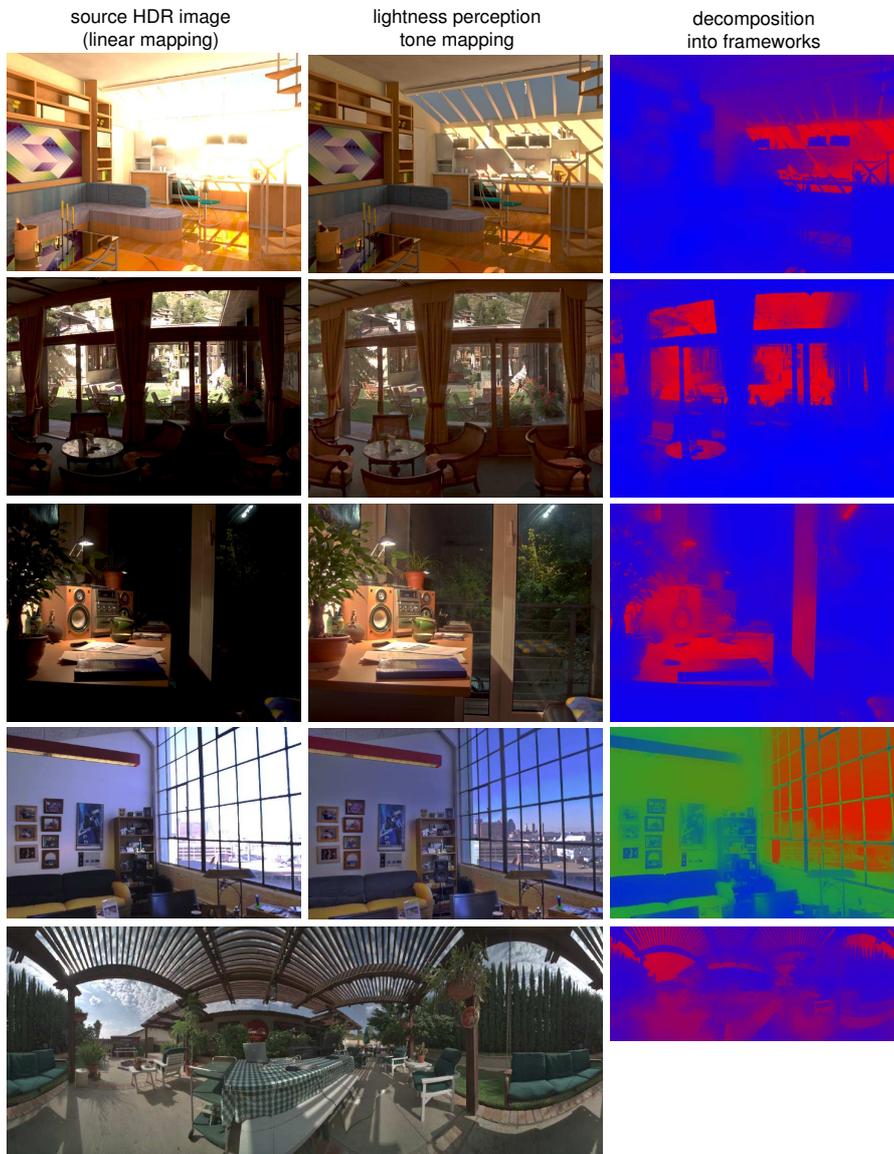
Figure 5.8: Results of lightness perception tone mapping. The left column (apart from the last row) contains the source HDR image shown with the linear mapping. The middle column contains results of the presented tone mapping method. The right column depicts the decomposition into frameworks. Red, blue and green colors depict the distinct frameworks. Higher saturation of the color illustrates stronger anchoring within the framework and the intermediate colors depict the influence of more frameworks on an image area. The HDR images in the 2nd and 5th row from the top courtesy of SpheronVR, and the HDR image in the 4th row courtesy of Byong Mok Oh.
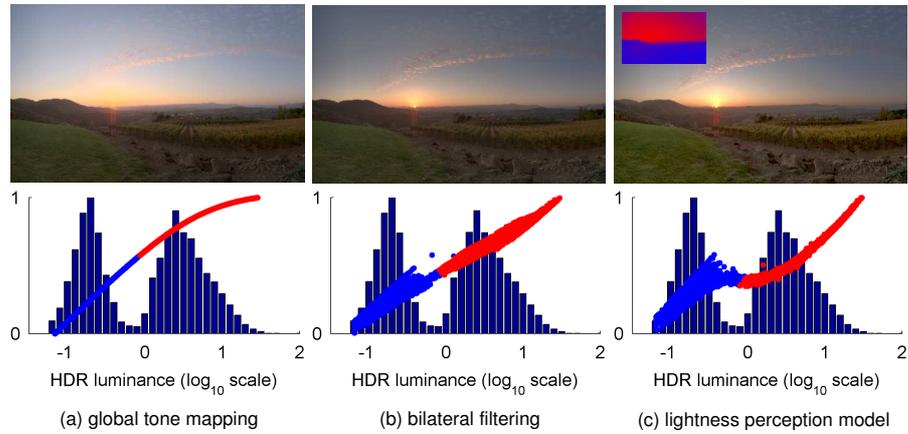
Figure 5.9: Comparison of three tone mapping operators: global sigmoid function, local tone mapping with bilateral filter, and lightness perception tone mapping. Color of the mapping functions in the plots correspond to the marked areas of frameworks (inset in image (c)). HDR image courtesy of SpheronVR.

set based on some constant local neighborhood. However, it may often be desired to vary the parameters of such algorithms between the areas of an image. In this case it often has to be done manually by the user. With the use of frameworks decomposition, it is now possible to identify the areas that are perceived homogeneously in an image. For the purpose of automated image processing, this permits to estimate the most appropriate parameters for a given algorithm individually for each framework.

In digital photography, it often happens that an image contains two different sources of illumination – for instance daylight from a cloudy sky and warm indoor tungsten light as in Figure 5.10. Such an image requires a white balance correction. However, correcting for the daylight will result in an increased orange cast in the tungsten light. The decomposition into frameworks allows the identification of such separate areas and enables different white balance correction in each of them. Again, frameworks represented as probability maps guarantee proper blending of edges where differently processed areas merge. One can envisage further possibilities in which our decomposition into frameworks reduces the required amount of manual interaction.

5.5.3   Performance

The estimation of lightness in a 4Mpx image using our computational model takes below a minute on a modern PC. The timing mainly depends on the number of decomposed frameworks, since the majority of computations is spent on the decomposition stage. Once the frameworks are known, the estimation of anchor and net lightness computation consists of simple operations. The K-means algorithm operates on a histogram and is therefore independent of the image resolution. The only bottleneck is the spatial processing using the bilateral filter, although we use an efficient approach presented by Durand and Dorsey [Durand and Dorsey 2002].

(a) global white balance   (b) white balance within frameworks   (c) frameworks
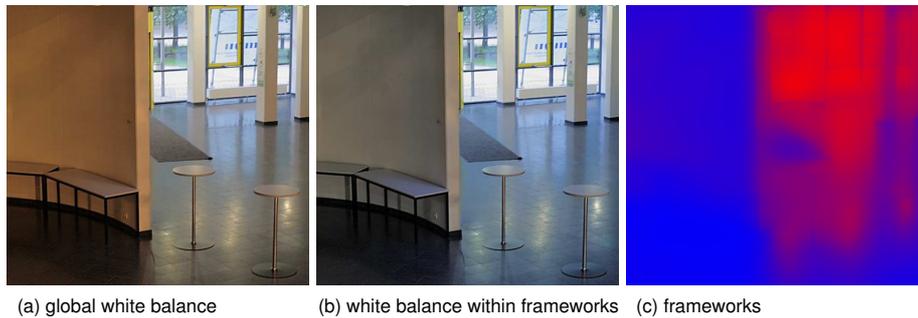
Figure 5.10: The daylight from a cloudy sky dominates the white balance in image (a) and causes the orange color cast in the interior illuminated by a tungsten light. The frameworks (c) can be used to separate such areas of different illumination and to perform an independent white balance in each of them (b).

5.6 Conclusions

We have presented a computational model of the anchoring theory of lightness perception. The model provides a practical implementation of the key concepts of this theory and aims at an accurate estimation of lightness in real world scenes captured as HDR images. We leveraged the theory to handle complex images by developing an automatic method for image decomposition into frameworks. Through the estimation of local anchors we formalized the mapping of the luminance values to lightness. We examined the accuracy of our model by reproducing the results of two perceptual experiments that were initially conducted to prove the accuracy of the theory.

We have demonstrated a novel tone mapping operator which aims at the accurate reproduction of lightness perception of real world scenes on low dynamic range displays. The strength of our operator is especially evident for difficult shots of real world scenes, which involve distinct regions with significantly different luminance levels. Moreover, the decomposition of an image into frameworks gives additional potential for automated image processing fine tuned to the perceptual aspects of the HVS.

Chapter 6

Objective Evaluation of Tone Mapping

Existing tone mapping algorithms can be generalized as a transfer function in the form of a "black box" which converts scene luminances to displayable pixel intensities. While the universal goal of such a transfer function is to reduce the original dynamic range and at the same time preserve the original appearance of an HDR image, a particular realization of it can be variable and depends on the objectives of a target application. In many cases one may wish to simply obtain nice looking images that resemble the original HDRs, but the requirements may also be more precise: perceptual brightness match, good visibility of details, equivalent object detection performance in tone mapped and corresponding HDR image, etc. In view of the technical limitations and constrained observation conditions for standard displays, such requirements can only be met at the cost of other image properties. For instance, if an available dynamic range is assigned to enable good visibility of details (local contrasts), there is no dynamic range left to depict global contrast variations in the scene. The trade-off between these conflicting goals is often balanced through an optimization process, but sometimes the design of an algorithm is focused on the requirements and is oblivious to the side-effects. In the end, the overall impact of image processing operations on the perceived image quality or fidelity to the real world appearance is not thoroughly understood.

Recent psychophysical studies attempt to evaluate tone mapping operators in terms of subjects' preference or fidelity of the real world scene depiction [Drago et al. 2002, Kuang et al. 2004, Ledda et al. 2005, Yoshida et al. 2005]. In such studies each operator is treated as a "black box" and its performance is compared on the whole with respect to other operators, without an attempt at understanding the reasons for subjects' judgments. While some studies of tone mapping operators go further and take into account the reproduction of overall brightness, global contrast or details (local contrast) in dark and bright image regions [Ledda et al. 2005, Yoshida et al. 2005], they remain focused on comparing the operator performance for each of these tasks. These studies, however, provide no deeper analysis of how the pixels of an HDR image have been transformed by tone mapping and in what way the outcome of such a transformation depends on image content.

In this chapter, instead of subjective analysis, we focus on developing objective metrics which could help in understanding how particular image characteristics, such as contrast or brightness, are distorted by tone mapping and determining the impact of such distortions on perceived image quality. We identify that major distortions in tone mapped images with respect to their HDR originals come from global and local contrast modulations. We create relevant metrics which evaluate the magnitude of *Global Contrast Change* and *Detail Visibility Change* along a perceptually meaningful scale and perform a corresponding study of 8 tone mapping algorithms. Besides the evaluation, the output of these metrics can be used as a feedback for perceptual enhancements [Smith et al. 2006].

6.1 Related Work

A number of perception-based visible difference (fidelity) metrics for image pairs have been developed, mostly for image compression and color reproduction applications (refer to [Winkler 2005] for a recent survey of such metrics). State of the art fidelity metrics such as the Visible Differences Predictor (VDP) [Daly 1993] or the Sarnoff Visual Discrimination Model (VDM) [Lubin 1995] include many important characteristics of the HVS, such as eye optic imperfections, luminance masking, the contrast sensitivity function (CSF), and pattern masking, making them very general metrics. However, such complex metrics may perform worse than simpler metrics specialized for the task of detecting well-defined distortion types, such as blocking artifacts that arise in image compression [Winkler 2005]. The majority of existing fidelity metrics are based on HVS models developed through threshold psychophysical experiments, the goal of which is to determine the magnitude of a simple stimulus so that it becomes just noticeable. Such metrics successfully detect the presence of perceivable image distortions, but perform poorly in estimating the magnitude of suprathreshold distortions and predicting their distraction to the human observer [Chandler and Hemami 2003]. With its spatial features for estimating imperceptible texture details, the iCAM model [Fairchild and Johnson 2003] is an exception. However, since the magnitude of perceptual responses to local contrast is not available, it can not be used to determine the change in detail visibility.

In this work, we are mostly concerned with one well defined suprathreshold distortion: contrast compression due to tone mapping. While much work has been done in the subjective evaluation of different tone mapping operators [Ledda et al. 2005, Yoshida et al. 2005], to our knowledge, we present the first feature-based characterization and objective perceptual measure of tone mapping distortion. Since fidelity metrics dealing with image pairs of significantly different dynamic ranges have not so far been proposed, and since we have found existing models to be ill-suited for our purposes, we present custom fidelity metrics for comparing perceived contrast differences between an original HDR image and its tone mapped LDR counterpart.

6.2 Distortion Metrics

All successful tone mapping operators balance the tradeoff between accurate reproduction of the luminance range and preservation of details. One can argue that the

photographic tone reproduction operator [Reinhard et al. 2002] best reproduces global contrast, while the *gradient domain compression* [Fattal et al. 2002] operator best preserves details. However, the accuracy of such statements may depend on the particular HDR image, and as concluded by evaluations of tone mapping operators [Yoshida et al. 2005, Ledda et al. 2005], it is difficult for one tone mapping operator to be well-suited to all types of images. Regardless of technique, each tone mapping operator introduces a degree of distortion into the resulting LDR tone mapped image. Drawing conclusions from previous evaluations and our own observations, we identify two major contrast distortions resulting from tone mapping:

Global Contrast Change the ratio between lightest and darkest areas of the HDR is reduced in the LDR,

Detail Visibility Change (textures and contours) the high frequency contrasts of the HDR image become less prominent, disappear, or become exaggerated in the LDR.

A significant Global Contrast Change is undesirable not only for esthetic reasons, but also because of changes in image understandability, despite good detail visibility. Certain specialized tone mapping operators assign a wider dynamic range to detailed regions to preserve textures and contours, which results in a narrower dynamic range available for global luminance changes, decreasing the ratio between lightest and darkest areas. Detail Visibility Change occurs either because a region becomes entirely saturated or because an area is mapped to very few or very low brightness levels. The second case is especially interesting from the perceptual point of view, because the physical contrasts still exist in the LDR image, however the details are invisible to the human observer.

Our goal is to determine the apparent distortion in detail visibility and global contrast change which were introduced during the tone mapping of HDR image. We focus on the luminance compression aspect of the operators. Instead of analyzing particular algorithms one by one, we consider tone mapping as an unknown transformation applied to the luminance of an HDR image, resulting in an LDR image. To do so, we use knowledge of human perception to compare a real world or synthetic scene, captured as an HDR image, to its LDR tone mapping as depicted on a given display device. The output of our metric consists of a single value representing the global contrast change factor and a map representing the magnitude of change in detail visibility. The units of the detail visibility map are Just Noticeable Differences (JND), which allows for an informed use of this information for potential perceptually based corrections [Smith et al. 2006].

To compare images of significantly different dynamic ranges we compare the luminance of an HDR image, denoted as Y, to the luminance shown on a display device, denoted as L. To accurately predict the displayed luminance, we assume that sufficient characteristics of the display device are known so that we can calculate the luminance value in cd/m^2 of each LDR image pixel. For an sRGB monitor, this requires black and white levels increased by an ambient illumination level. Similarly, a photometrically calibrated HDR image is desirable.

We transform the gamma corrected intensity values[1] y of the LDR image to display luminance values L. Given the display black L_{black} and white L_{white} levels in cd/m^2

---

[1]image luminance is calculated from the RGB channels according to the [ITU 1990] standard.

and assuming sRGB response, the transformation is the following:

$$L = L_{black} + \text{sRGB}^{-1}(y) \cdot (L_{white} - L_{black}). \tag{6.1}$$

If the absolute luminance values of an HDR image are unknown, we align the relative HDR values Y to the LDR values L according to the average logarithmic luminance, a method often used as an adaptation estimate in tone mapping [Drago et al. 2003, Reinhard et al. 2002].

6.2.1   Global Contrast Change

The change in ratio between brightest and darkest points of an image is a traditional definition of global contrast change that is necessarily adjusted by tone mapping, and so would not be considered a distortion. Particularly, since tone mapping algorithms most often use the whole display dynamic range, above definition always results in a constant global contrast. Yet images resulting from different tone mapping operators can create starkly different impressions of global contrast, meaning that such a naïve measure is not appropriate. Contrary to this definition and others, such as one using the multi-resolution definition given by [Matkovic et al. 2005], we consider global contrast change to be a characteristic defined by the shape of the tone mapping function, thus removing the emphasis on extreme lights and darks which have less impact on the impression of global contrast. Our definition of global contrast change is more closely related to image comprehension, which according to Gestalt theorists, involves the cognitive task of separating the image into recognizable objects, most importantly, the separation of foreground objects from the background [Livingstone 2002]. As such, a decrease in global contrast may make comprehension of the LDR image more difficult, indicating a loss in visual communication efficacy.

While it is sensible to analyze tone mapping functions to obtain a global contrast estimate, these functions are either unknown or not well-defined, as in the case of *gradient domain compression*. However, we argue that a general approximation of the tone mapping function is sufficient for estimating global contrast. In our metric, we approximate the tone mapping function using linear regression in the logarithmic domain:

$$\log_{10} L \approx TM(\log_{10} Y) = \mathscr{C} \cdot \log_{10} Y + \mathscr{B}, \tag{6.2}$$

where $\mathscr{C}$ and $\mathscr{B}$ are estimated coefficients, and Y and L are the luminances of the HDR and LDR images. The meaning of logarithm in equation (6.2) is two-fold. First, the logarithm of luminance provides a crude approximation of brightness and the calculated values of the coefficients reflect the brightness mapping. Further, the linear regression estimates a general tendency of the mapping rather than being prone to detail enhancing procedures which do not influence global contrast relations. Second, if we exponentiate the equation (6.2), we obtain a standard contrast scaling equation in image processing [Pratt 1991]:

$$L \approx TM(Y) = Y^{\mathscr{C}} \cdot 10^{\mathscr{B}}, \tag{6.3}$$

where $\mathscr{C}$ adjusts contrast and $10^{\mathscr{B}}$ adjusts brightness. Summarizing, equations (6.2,6.3) estimate the shape of the tone mapping curve and relate the contrast in LDR image L to its original HDR Y. Therefore the coefficient $\mathscr{C}$ obtained through linear regression denotes the Global Contrast Change, such that $\mathscr{C} < 1$ indicates a decrease in the global

contrast in the LDR image, whereas $\mathscr{C} > 1$ indicates an increase with respect to the original HDR.

The result of applying our measure of Global Contrast Change to two tone mappings (one global and one local) is shown in Figure 6.1. While both methods make use of the entire available dynamic range, the shapes of their mapping functions differ: the global mapping function is well-defined, as opposed to the non-uniform and scattered local mapping function. Higher global contrast is obtained with the global tone mapping method, whereas the detail preserving local method exhibits a smaller ratio between bright and dark areas (the function approximation is nearly flat).
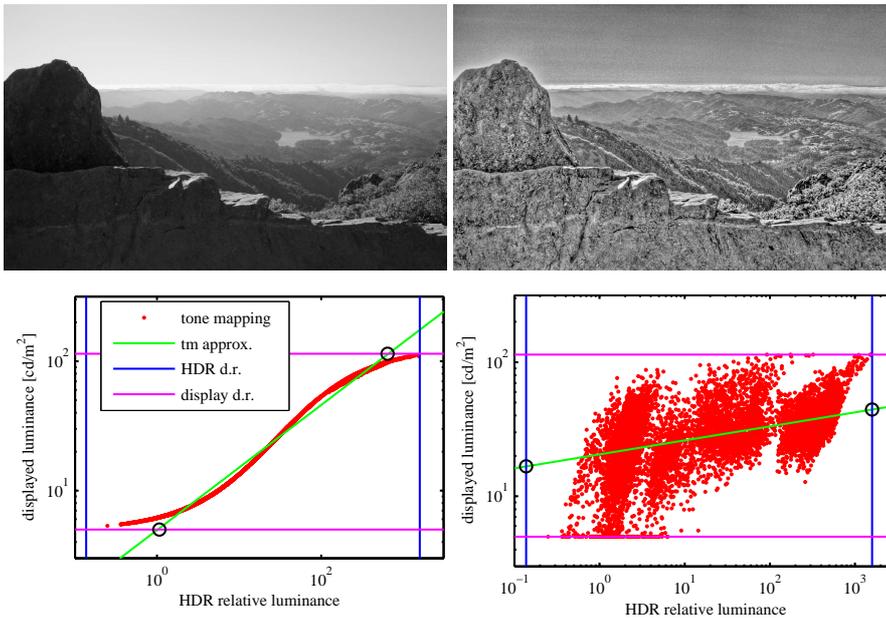
Figure 6.1: Global Contrast estimation for global [Reinhard et al. 2002] (left) and local [Fattal et al. 2002] (right) tone mapping. Each plot shows pixel-by-pixel mapping between HDR and LDR, linear brightness mapping estimation, and dynamic ranges (d.r.) of LDR and HDR. Global Contrast Change for global tone mapping is $\mathscr{C} = 0.49$ and for local $\mathscr{C} = 0.10$.

6.2.2 Detail Visibility Change

Details of textures and contours can be described as the high frequency contrast between a pixel and its adapting field. Visibility, the response of the HVS to the magnitude of such contrasts, is not linear and depends on the adaptation level. Contrast visibility can be analyzed in terms of contrast detection and contrast discrimination. We use contrast detection for identifying visible details in both the HDR and LDR images, and we use contrast discrimination for identifying the magnitude of visible difference in detail contrast between the HDR and LDR images.

We start by identifying high frequency contrasts that presumably create texture and contour details in the image. For each pixel Y_i we estimate the adapting luminance Y_i^{sp}

in its neighboring area and calculate the contrast expressed as a logarithmic ratio of luminance values:

$$G(Y_i, Y_i^{sp}) = \log_{10} \frac{\max(Y_i, Y_i^{sp})}{\min(Y_i, Y_i^{sp})}. \tag{6.4}$$

We simulate the adaptation to low spatial frequencies in an image and we take special care to prevent the influence of significantly different luminance values on an adaptation level. We obtain the adaptation map Y^{sp} by processing the HDR image with a low pass bilateral filter in the logarithmic domain. Such a filter removes high frequencies while preserving high contrast edges. The adaptation map is refined by eliminating frequencies above 20 cycles per pixel and preserving edges of logarithmic contrast ratio higher than 0.25. We calculate the high frequency contrasts of the LDR image in the same way. It is important to note that the particular choice of the bilateral filter for estimating the adaptation map is not critical. Other algorithms known from tone mapping can be used as well, as long as they do not introduce artifacts at high contrast edges.

To estimate the Detail Visibility Change between two images of significantly different dynamic range, knowledge of the hypothetical HVS response to given physical contrasts under given adaptation conditions is required. A reasonable prediction for a full range of contrast values is given by the following transducer function that is derived and approximated by Mantiuk et al. [Mantiuk et al. 2006]:

$$T(G) = 54.09288 \cdot G^{0.41850}, \tag{6.5}$$

with the following properties:

$$T(0) = 0 \quad \text{and} \quad T(G_{threshold}) = 1. \tag{6.6}$$

The transducer function estimates the HVS response to physical contrast in Just Noticeable Difference (JND) units. Thus for a given contrast threshold, $G_{threshold}$, a transducer value equals 1 JND. It is important to note that this measure holds for suprathreshold measurements, since it not only estimates the detection, but also the magnitude of change.

The approximation given by Equation (6.5) has been derived with the assumption of 1% contrast detection threshold[2], i.e. $G_{threshold} = \log_{10}(1.01)$. Although such an assumption is often made in image processing for LDR, the detection threshold depends on an adapting luminance level and is described by the Threshold Versus Intensity (TVI) function [CIE 1981]. The TVI function shows that this threshold varies in the luminance range of displays and the dynamic range in HDR is often high enough to make this 1% assumption for the detection threshold inaccurate. We therefore derive a scaling factor $t(Y^{sp})$ for the transducer function (6.5) which adjusts its properties (6.6) to match the TVI function given an adapting luminance:

$$t(Y^{sp}) = \frac{\log_{10} 1.01}{\log_{10} \frac{Y^{sp} + \text{tvi}(Y^{sp})}{Y^{sp}}}. \tag{6.7}$$

Such a scaling factor is appropriate because the approximation of the transducer function (6.5) was derived with starting conditions from (6.6), and since the influence of the

---

[2]While equation (6.5) gives a good approximation of the response to contrast in a wide range of physical contrasts, it actually has a slightly larger fitting error for near-threshold values. Thus equation (6.5) does not precisely satisfy equation (6.6) for 1% detection threshold. For detailed derivation see [Mantiuk et al. 2006]

threshold is multiplicative [Mantiuk et al. 2006]. Figure 6.2 illustrates the magnitude of change in the HVS response depending on the adapting luminance. The response changes by a factor of almost 1 order of magnitude within the visible range of luminance on a display. In practice, the scaling factor reduces the response to contrast in the dark areas of an image.
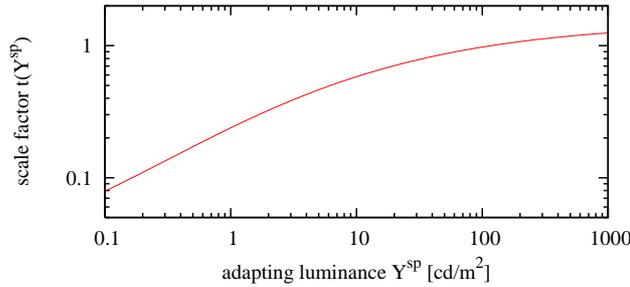

Figure 6.2: Plot of the scale factor from equation (6.7). Luminance range of a typical LCD display is 2 to 200 cd/m^2.

Given the scaled transducer function, we can estimate the hypothetical response of the HVS to the high frequency contrasts measured with equation (6.4):

$$T^*(Y_i, Y^{sp}) = T(G(Y_i, Y_i^{sp})) \cdot t(Y_i^{sp}). \tag{6.8}$$

The response T^* is expressed in JND units, which means that a detail Y_i is visible under given luminance conditions only if $T^* > 1$. Given this relation, we are able to estimate the details of a displayed LDR image and the details of an HDR image which would be visible to a human observer. Furthermore, since the transducer function is a suprathreshold measure, we are able to estimate change by comparing the magnitude of detail visibility in a displayed LDR image to its HDR version (spatial arguments are omitted for brevity):

$$\Delta \overline{T^*}(Y_i, L_i) = \begin{cases} 1 & \text{for} \quad \overline{T^*}(Y_i) > 1 > \overline{T^*}(L_i), \\ 0 & \text{for} \quad \|\overline{T^*}(Y_i) - \overline{T^*}(L_i)\| < 1, \\ \overline{T^*}(Y_i) - \overline{T^*}(L_i) & \text{otherwise.} \end{cases} \tag{6.9}$$

For practical reasons, we consider the average detail visibility measure over its neighboring pixels, denoted as $\overline{T^*}$, because we are interested in general detail visibility in a certain small area. As shown in equation 6.9, we consider three cases of detail visibility change. When a response to high frequency contrast in the HDR image is attenuated from above 1 JND to below 1 JND in the tone mapped image, the change is 1 JND. When the difference in response is below 1 JND, the change is deemed invisible and is set to 0. In all other cases, the magnitude of Detail Visibility Change is set to the difference in responses T^*. We illustrate the performance of this measure in Figure 6.3.

6.3 Analysis of Tone Mapping Algorithms

We analyzed the performance of 8 tone mapping methods in terms of Global Contrast Change and Detail Visibility Change using the presented metrics. The analysis was performed on a set of 18 HDR images with an average dynamic range of approximately 4
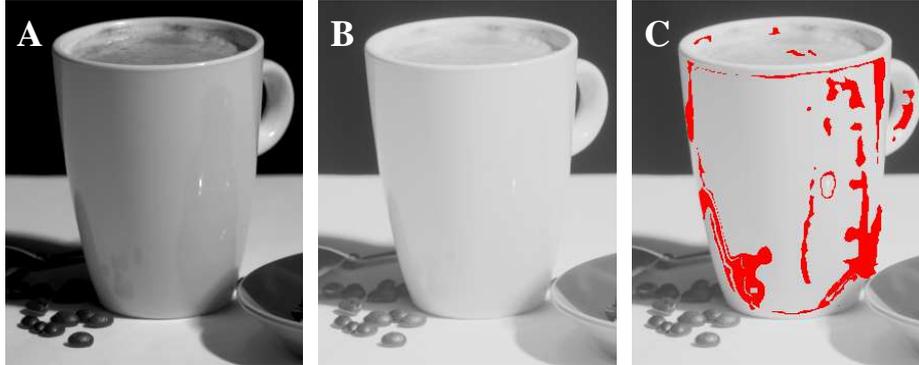
Figure 6.3: Detail Visibility. HDR image (A) contains subtle reflection on a surface of the cup. A global tone mapping (B) reveals the coffee beans in the shadow but the reflection details become indiscernible. The areas of the image with lost details are predicted by our metric (C), where red color marks $\Delta\overline{T^*} > 1$.

orders of magnitude and a resolution between 0.5 and 4 megapixels. The set contained a variety of scenes with differing lighting conditions and included panoramic images. We tested the following global (spatially uniform) tone mapping algorithms: *gamma correction* ($\gamma = 2.2$), *adaptive logarithmic mapping* [Drago et al. 2003], *photographic tone reproduction (global)* [Reinhard et al. 2002], *photoreceptor* [Reinhard and Devlin 2005]; and the following local (detail preserving algorithms): *gradient domain compression* [Fattal et al. 2002], *bilateral filtering* [Durand and Dorsey 2002], *lightness perception* (Chapter 5), *photographic tone reproduction (local)* [Reinhard et al. 2002]. The tone mapped LDR images were obtained either from the authors of these methods or by using publicly available implementations: *pfstmo*, Appendix B. Tone mapping parameters were fine tuned whenever default values did not produce satisfactory images.

In practice, the contrast detection component of our Detail Visibility Change metric required calibration to correctly estimate the visibility of subtle details in extreme dark and light regions. We introduced a scaling factor to equation 6.7 to increase the predicted response of the HVS to contrasts, and found that a value of 1.89 led to satisfactory predictions in our set of test images. The display characteristics corresponded to a typical consumer LCD with an sRGB response, black level at $2.5cd/m^2$, and white level at $210cd/m^2$ measured in office illumination conditions.

We measure the Global Contrast Change according to the description in Section 6.2.1 and the results of our analysis are summarized in Figure 6.4. There is an apparent advantage of the *photographic tone reproduction (local & global)* methods in conveying the global contrast impression almost without any change. These methods were also among the top rated in other studies [Ledda et al. 2005, Yoshida et al. 2005]. In contrast the *gradient domain compression* causes a severe decrease in the global contrast. Other local methods perform moderately. Particularly, in case of the *lightness perception* model the decrease of global contrast is caused by the optimization of difference in luminance between the frameworks. The superior performance of the global methods is traded for less efficient reproduction of details as observed in the further analysis.

We analyze the Detail Visibility Change for two cases that are part of equation (6.9):
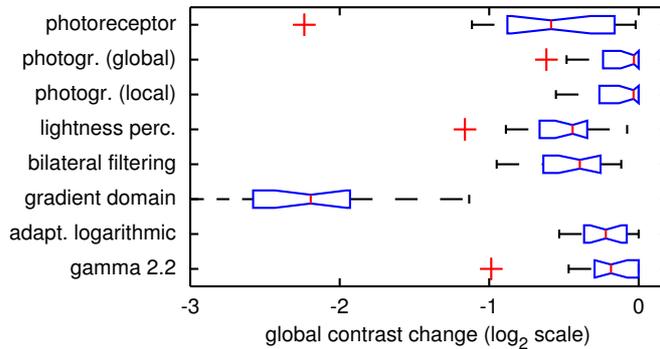
Figure 6.4: The influence of various tone mapping operators on the change of the global contrast $\mathscr{C}$ from equation (6.2). The negative values denote the decrease in global contrast and 0 means no change. The red bars show the median, whiskers denote 25th and 75th percentile of data, and the red crosses are outliers.

the loss of detail visibility and the change in the magnitude of the detail visibility. The loss of detail visibility refers to the $\overline{T^*}(Y_i) > 1 > \overline{T^*}(L_i)$ case in equation (6.9) and describes the situation in which details have been visible in the HDR image but are not perceivable in the tone mapped image. To measure the change in the magnitude of detail visibility we analyze the areas in which the details are visible both in the HDR and in the tone mapped image and the analysis refers to the $\overline{T^*}(Y_i) - \overline{T^*}(L_i)$ case in equation (6.9). The average decrease and increase of the visibility are calculated separately. Following [Yoshida et al. 2005], we further split the analysis into the dark and bright image areas. To segment these areas, we assign 33% of the darkest pixels in an image to the dark area, and 33% of the brightest pixels to the bright area. The results are summarized in Figures 6.5 and 6.6. The results of the increase in detail visibility are not shown because they can be only observed for the *gradient domain compression*.

The analysis of Figure 6.6 indicates that the dynamic range compression and the change in luminance levels lead to a decreased perception of details in case of all operators. The magnitude of change, however, is in most cases below 1 JND. This means that the loss of detail visibility, largely observed in Figure 6.5, is unlikely caused by the stark luminance range compression, but rather even a minor compression causes the magnitudes of details to drop below the visibility threshold. This would suggest that a minimal correction is sufficient to restore the visibility. The detail preserving tools implemented in local tone mapping methods seem to perform well in bright image areas, however the dark image areas are often not well reproduced with the exception of the *gradient domain compression* and the *adaptive logarithmic mapping*. Notably, the *adaptive logarithmic mapping*, which is a global operator, preserves details exceptionally well in dark image areas. This advantage comes at the cost of a slightly higher loss of details in bright areas. The *lightness perception* tone mapping performs on par with other local methods, being slightly advantageous in the bright image areas. The *gradient domain compression* is particularly interesting, because the results of this detail preserving method indicate both the increase and decrease in detail visibility while at the same time the visibility of any details is not lost. Such behavior indicates good performance of the contrast transfer function which attenuates large contrasts and increases the small ones as explained in Section 2.5.4.
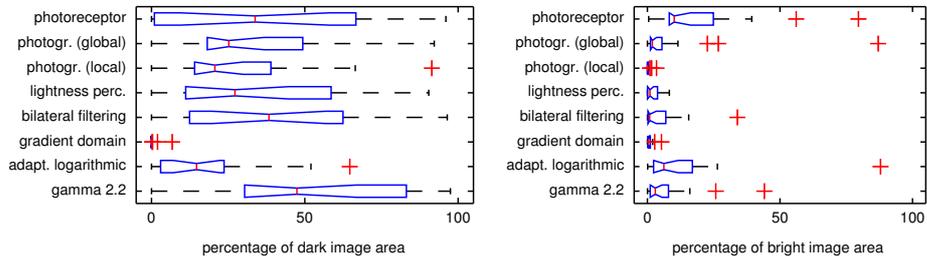
Figure 6.5: The influence of various tone mapping operators on the loss of the details visibility. The analysis are split into dark (left) and bright (right) image areas. The percentage denotes the part of the dark/bright image area in which details have been visible in the HDR image but are not perceivable in the tone mapped image.
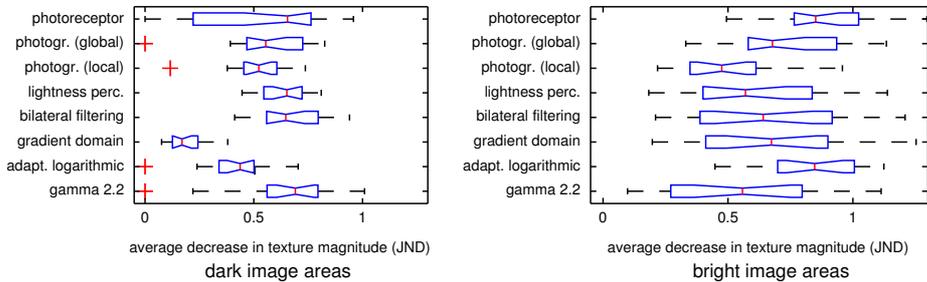
Figure 6.6: The average decrease of the magnitude of details visibility caused by the analyzed tone mapping operators. The analysis are split into dark (left) and bright (right) image areas. The average is calculated over the parts where details are visible both in the HDR and in the tone mapped image. 0 denotes no change in visibility and 1 JND denotes a visible change.

Overall, the better performance of the global tone mapping operators in the analysis of Global Contrast Change is not surprising. However, the performance of the algorithms in terms of Detail Visibility Change is very unstable across the test images and there is no obvious winner of the evaluation. Interestingly, the enhancements required to improve the results do not necessarily need to be strong. While the discovery of a new universal operator seems unlikely, our analysis motivates the development of enhancement algorithms that could restore the missing information in tone mapped images based on their HDR originals. Such enhancements can be obtained using colors [Smith et al. 2006] or carefully shaped countershading profiles as explained in Chapter 7.

6.4 Conclusions

Based on experience and conclusions from previous work we identified two major distortions introduced to luminance while tone mapping: Global Contrast Change and Detail Visibility Change. To our knowledge, we present the first objective perceptual metric for the measure of contrast distortions between an HDR image and its LDR depiction. To construct these metrics, we extended the transducer function to handle HDR luminance levels. We analyzed selected tone mapping operators using our met-

rics and we provided an indicative characterization of these operators in terms of global contrast and detail preservation in dark and light regions. Since only luminance values are evaluated by our distortion metrics, their application is most suitable for the luminance-based subset of tone mapping operators. Our techniques for distortion detection and magnitude evaluation can be used with other methods of perceived contrast enhancement [Calabria and Fairchild 2003, Smith et al. 2006], including luminance manipulation at contrasting edges – an enhancement method exploited in Chapter 7.

Chapter 7

Restoration of Lost Contrast

Successful comprehension of observed images and scenes depends on our ability to distinguish their features. Human vision identifies scene features through the apparent contrasts that they create within their context. Well visible contrasts facilitate the recognition of objects in a scene, identification of their texture, understanding of their spatial distribution, and the ability to judge brightness between adjacent and distant areas. Together, these features directly influence people's assessment of overall image quality [Janssen 2001]. Therefore, a well pronounced rendition of perceived contrasts is an important goal of computer graphics algorithms which process visual information. Unfortunately, often this goal is not achieved due to either technical limitations or poor input data. In tone mapping for instance, the insufficient capabilities of displays require reduction of the dynamic range in images, which inevitably leads to attenuation of contrasts and loss of visual information as shown in Chapter 6. In rendering on the other hand, poor design of illumination or bad shading algorithms produce low contrast images in which comprehension of scene content is strongly confined [Luft et al. 2006].

In this chapter we are concerned with the problem of communicating contrasts in images that suffered from contrast degradation with respect to their original. In case of a tone mapped image, the original is its source High Dynamic Range (HDR) version. Such HDR images can be captured with HDR cameras, using multi-exposure techniques, or obtained in many rendering application in particular in realistic image synthesis and lighting simulation as explained in Chapter 2. Even if rendering leads to low dynamic range images, e.g. non-photorealistic rendering, contrasts from the depth map can be used for similar purposes [Luft et al. 2006]. Unlike in typical contrast enhancement tools such as histogram equalization or contrast equalization, we do not want to change the general appearance of processed images. Nor de we try to restore the physical contrasts in the image, especially given that most often it is not possible due to the dynamic range restrictions. Instead, we propose to enhance the perceived contrasts through a gradual modulation of brightness in the vicinity of the contrasting edge inspired by a family of known perceptual illusions [Kingdom and Moulden 1988]: Craik, O'Brien, Cornsweet. These illusions, which we have briefly introduced in Section 3.2.6, address several models of gradual darkening and lightening of areas towards their common edge to which we in general refer as *countershading profiles*. Our approach has particular advantages in that the contrast enhancement can

be achieved within the available dynamic range, and the modifications do not change the general appearance of an image because they are limited to areas along the edges of the enhanced features. Furthermore, the perceived contrast may be larger than would be normally achievable on a target display. Similar techniques have since long time been used by artists to obtain better contrasts in paintings, as explored by Livingstone [Livingstone 2002].

We present an image processing tool that creates countershading profiles for an image to enhance perceived contrast of features degraded with respect to the original. Our tool can be considered as a generalization of *unsharp masking* – an image enhancement technique which in certain cases also creates countershading profiles by overlaying the difference of an image and its blurred version. The development of a new algorithm is motivated by the disadvantages of the traditional unsharp masking which cannot be applied to automatically correct individual image features. To deliver the automatic correction with respect to a reference image, we combine the countershading algorithm with a multi-resolution contrast metric. The metric measures local contrast of features at different scales, compares the processed image to its reference, and drives the spatial extent and the strength of countershading profiles. We first demonstrate how to match the physical amplitude of a reference contrast with the amplitude at the profiled edge, and later we adjust the amplitude according to findings in psychophysics to reduce the perceptual difference between them. Finally, excessive countershading profiles may become visible as halo artifacts and degrade the image quality, which in most cases is unacceptable and in fact reduces the strength of the contrast enhancement. We employ the visual detection model to estimate the maximum amplitude of a countershading profile that is not objectionable in a given area based on the luminance threshold, contrast sensitivity and the contrast masking effects.

We start with a review of unsharp masking and contrast enhancement techniques recently used in computer graphics in Section 7.1, and we summarize relevant findings in psychophysics in Section 7.2. Next, in Section 7.3 we present a new algorithm to create the countershading profiles. In Section 7.4 we introduce the visual detection model used to adjust the adaptive countershading to prevent undesired halo artifacts, and draft the implementation in Section 7.5. Finally, we illustrate and discuss possible applications in Section 7.6.

7.1   Previous Work

Unsharp masking [Pratt 1991] is the technique in which a Gaussian blurred image Y_σ is subtracted from its original luminance Y to create an unsharp mask that is added to the original image with a coefficient c:

$$\mathscr{Y} = Y + c \cdot (Y - Y_\sigma), \tag{7.1}$$

where $\mathscr{Y}$ is the enhanced image and σ determines the spatial extent of the Gaussian kernel. The magnitude of the correction c needs to be adjusted by the user and all pixels in the image are corrected with the same coefficient. However, the enhancement happens in two dissimilar ways: through the countershading and through the reintroduction of features. The highest quality of correction is gained only for image features whose scale is similar to or larger than the size of the Gaussian kernel [Neycenssac 1993], because they obtain valid countershading profiles. Small kernels, however, lead
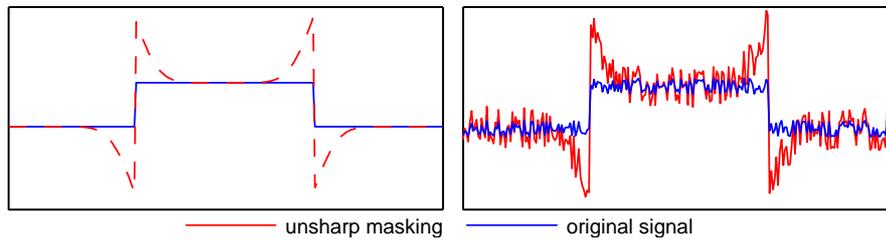
Figure 7.1: Countershading using unsharp masking gives correct results when the kernel size is adjusted to the size of the feature (left). If unsharp masking is used to enhance the contrast lost on a step edge with details, the filter models the countershading profile on the edge but also strongly amplifies all the features of a smaller scale (right).

to sharpening effects at the edges of larger features [Neycenssac 1993] and have limited capabilities to enhance contrast [Kingdom and Moulden 1988]. All features smaller than the kernel size are reintroduced with a varied strength which is influenced by their scale and the difference from the local average, as illustrated in Figure 7.1. The noise amplification caused by such a reintroduction of the small scale features and the sharpening artifacts at the high contrast edges can be minimized with the adaptive unsharp masking [Polesel et al. 2000, Ramponi et al. 1996]. Psychophysical findings, which show that the uniform physical correction is perceived as stronger in the dark parts of an image than in light areas, motivated the non-linear adaptive unsharp masking [Ramponi et al. 1996]. In spite of the numerous improvements to this technique, we are not aware of any method for an automatic enhancement using individually adjusted kernel sizes and profile magnitudes to create the countershading profiles that are appropriate for enhanced features without distorting other parts of the restored image.

The influence of weak contrasts on a limited comprehension of the spatial distribution of objects in a scene has been studied by Luft et al. [Luft et al. 2006]. They show that unsharp masking using the depth map of a scene strongly enhances the cognition of spatial distribution of objects. Their results are very good because depth maps extract precisely the edges which outline objects in a scene and correction of these edges improves the perception of the spatial organization. The intensity of countershading, however, depends only on the depth relations of objects behind, and therefore unnaturally looking dark outlines may appear over the objects further behind in the scene. The visual model presented here limits the countershading strength based on the actual image contents to prevent the visible degradation of images, thus limiting such artifacts.

The loss of communicated information is also typical for tone mapping, where the contrasts are explicitly reduced in an HDR image to fit into the dynamic range of a display or print. In Chapter 6 we show that, despite the different approaches to tone mapping, each algorithm suffers from a certain amount of contrast degradation leaving space for improvements towards the reference HDR. To better communicate lost contrast information, fine details can be corrected with opposite colors guided by a single-resolution local contrast metric, and the largest contrast can be restored with a segmentation based countershading technique adjusted by a single global contrast measure [Smith et al. 2006]. Unfortunately, in such an approach all features of the intermediate size remain
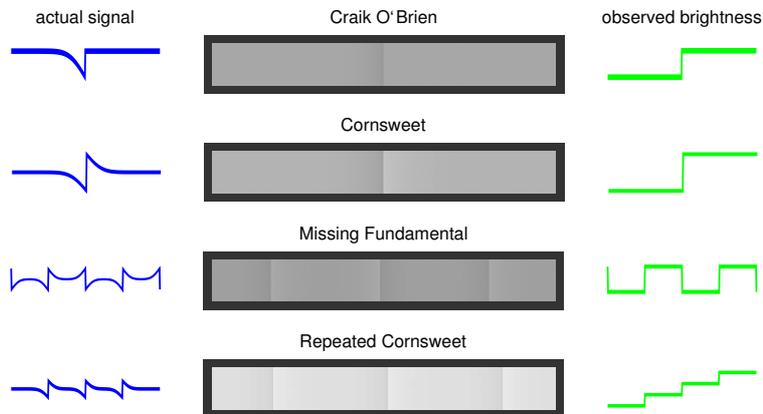
Figure 7.2: Different countershading profiles and their influence on the perceived brightness. Plots after [Kingdom and Moulden 1988].

uncorrected and the countershading is applied to only one arbitrary edge in the image. We propose to strongly couple the countershading with a multi-resolution local contrast metric and automatically correct features at various scales in a consistent manner with the individually adjusted profiles. Further, we provide a perception model which counteracts the objectionable halo artifacts.

A comprehensive model of the human visual system is embedded in the Visual Differences Predictor [Daly 1993], which detects the differences between the reference and distorted images. Such a visual model accounts for luminance masking, spatial contrast sensitivity, and contrast masking in spatial frequency and orientation bands. However, it is computationally expensive and therefore is often simplified in computer graphics applications. Predicting visible rendering artifacts [Ramasubramanian et al. 1999], for instance, is successfully done with a simpler model which ignores the orientation bands. We derive a similar detection model to prevent the countershading profiles from appearing as the halo artifacts. While these models are more focused on the near-threshold noise detection, in our context the supra-threshold effects of lower frequencies are of more interest.

7.2   Perceptual Background of Countershading

In Section 3.2.6, we have introduced the illusion of perceived brightness difference between two adjacent surfaces of equal intensity. This difference is induced by a carefully shaped brightness profile at the border between these surfaces – this particular example of countershading is the Cornsweet profile. We have observed that the perceived difference appears both on simple uniform patches, Figure 3.13, and in natural images Figure 3.14. The latter figure illustrates also that contextual information, like shape or perspective, articulate the effect of the illusion.

While the Craik-O'Brien-Cornsweet illusion describes only several cases, Kingdom et al. [Kingdom and Moulden 1988] summarize a family of border profiles and their influence on the brightness of adjacent areas. As shown in Figure 7.2, practically any
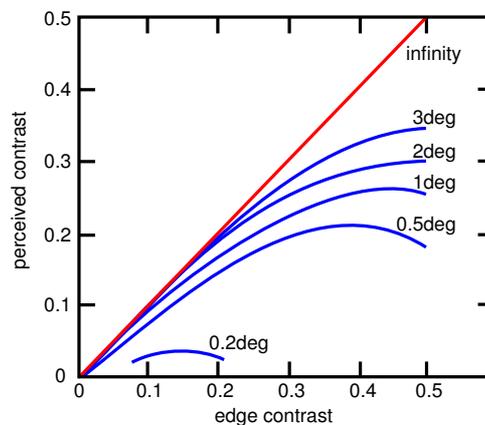
Figure 7.3: Apparent contrast of the Cornsweet profile as a function of the peak contrast of the profile and its width in degrees of visual angle. The straight line (red) denotes the actual step edge (Cornsweet profile of infinite size). Plot after [Dooley and Greenfield 1977, Fig.4], contrasts given in Michelson measure.

form of countershading and the combination of them leads to a magnification of the perceived contrast. Such profiles can be modeled using the Gaussian function in which the amplitude and the standard deviation determine the intensity of countershading. The modeling algorithm is technically similar to unsharp masking, and leads to alike profiles (compare with Figure 7.1).

The informed use of such an illusion to enhance images requires, however, that the perceptual properties of the profiles are well understood. Dooley and Greenfield [Dooley and Greenfield 1977] determined the relation that gives the amplitude of a countershading profile that is required to obtain a perceptual contrast match with a step edge with respect to a simple stimulus. Additionally, Burr [Burr 1987] observed the increase of perceived contrast when a countershading profile is added to an existing step edge. It has also been found that the spatial extent of a profile determines the maximum possible enhancement and has to be appropriate for the magnitude of the corrected contrast. For instance, a Cornsweet profile of 1 visual degree can simulate an edge of up to 0.25 Michelson contrast (i.e. strong supra-threshold contrast), but further amplification of this profile leads in fact to a decreased illusion as illustrated in Figure 7.3. As soon as the low frequency of the profile can be independently detected, the profile is clearly distinguished at an edge and the increase in the amplitude has no effect on the perceived contrast [Burr 1987]. This suggests that the contrast enhancement using the countershading profiles should be guided by a visual detection model. Finally, the illusions created by the spatially larger profiles are not affected by an additive noise [Burr 1987], thus the countershading profiles applied to the differently textured areas give consistent effects.

The strength of the perceived contrast enhancement due to countershading is influenced by visual cues. In particular, a contextual hint that the countershading profile results from a difference in the illumination of two surfaces, possibly confirmed by the perspective information, almost doubles the strength of the effect [Purves et al. 1999]. This is confirmed by the success of the Luft et al. [Luft et al. 2006] approach, in which objects separated by different depths are likely to be differently illuminated as well.

In contrary, the confidence that a profile is a feature of the surface reflectance significantly reduces the illusion. These observations, especially related to the larger scale contrasts, cannot be explained by the receptive field properties of the lower order visual neurons, or by the fact that both the step edge and the countershading profile have almost the same frequency characteristics when normalized by the contrast sensitivity function [Kingdom and Moulden 1988].

While the early explanations of the Cornsweet effect are based on threshold sensitivity, the illusion is clearly supra-threshold and in fact a consistent theory explaining all experimental findings has not been found so far. Our decision to use the modified supra-threshold sensitivity [Dooley and Greenfield 1977] is motivated by the fact that this model explains well the results of the experiments which measure the apparent contrast, including the supra-threshold effects, of up to 0.7 in Michelson measure. Clearly, the visual cues strongly articulate the effect [Purves et al. 1999], but even if an appropriate model was available, it would require a robust decomposition into illumination and reflectance which practically is only possible in image synthesis.

7.3   Image Processing for Countershading

We develop a method that creates the countershading profiles to enhance the perceived contrasts of edges in the restored (input) image that are less pronounced than the corresponding contrasts in the specified reference image. We identify such edges in the restored image by comparing it to the reference image using the multi-resolution local contrast metric. Guided by the metric, we create the profiles from the sub-band components such that the profiles are individually adjusted to the corrected features without distorting information that has been well preserved from the reference.

7.3.1   Multi-resolution Local Contrast Metric

We use a metric which measures the physical local contrast at several frequency bands in a similar manner to Peli [Peli 1990]. We decompose an image into a Gaussian pyramid, in which each lower level is filtered by a 5×5 kernel and its resolution is halved as described in [Burt and Adelson 1983]. Such a decomposition splits the image frequencies into octaves which corresponds to the frequency separation observed for the human visual system [Peli 1990]. Thus, on the highest level we measure the contrast of fine details, and on the lowest level the contrasts between the major areas in the image. The lowest level we consider is 4 pixels long in the smaller dimension, and we ignore the base band. For each pixel at each pyramid level l, we calculate the local sub-band contrast C_l using the formula:

$$C_l = \frac{|Y - Y_{mean}|}{Y_{mean}}, \qquad (7.2)$$

where Y is the luminance of a pixel at the pyramid level l and Y_{mean} is the local mean luminance. In practice Y_{mean} is taken from the corresponding pixel at the lower pyramid level. The final output of the metric is the pyramid that contains the ratios of the

corresponding local contrasts between the input image and its reference:

$$R_l = \min(\frac{C_l^{inp}}{C_l^{ref}}, 1).\tag{7.3}$$

In this equation, the input and reference image pair can for instance be a tone mapping result and its original HDR image. The ratio R_l is limited to the maximum value of 1 because the detail amplification with respect to the reference is not considered.

7.3.2 Adaptive Countershading

We develop a method which selectively adds the countershading profiles to the restored image guided by the sub-band local contrast ratio (7.3) from the metric. We start with an observation that the addition of successive levels of the full resolution Difference of Gaussians up to a certain level (here the example for 3 levels):

$$U = (Y - Y_{\sigma(1)}) + (Y_{\sigma(1)} - Y_{\sigma(2)}) + (Y_{\sigma(2)} - Y_{\sigma(3)}),\tag{7.4}$$

gives the same result as unsharp mask, equation (7.1), for this level:

$$U = (Y - Y_{\sigma(3)}),$$

where $\sigma(l) = 2^{l-1}/\sqrt{2}$ denotes the Gaussian blur at level l and Y is the luminance of the reference image. When the terms of such a sum are further multiplied by the output of the multi-resolution metric which locally adjusts the amplitudes of the sub-band components:

$$P = \sum_{l=1}^{N} (1 - \uparrow R_l) \times (\log Y_{\sigma(l-1)}^{ref} - \log Y_{\sigma(l)}^{ref}),\tag{7.5}$$

we obtain the countershading profiles P which are adjusted to match the contrasts in the reference image. In equation (7.5) l denotes the level of the Gaussian pyramid with N being the lowest, R_l are the contrast ratios from the metric at the selected level, operator ($\uparrow$) denotes upsampling to the full resolution, operator $\times$ is the element-wise multiplication, and $Y_{\sigma(0)}^{ref}$ is the luminance of the reference image. The difference $(\log Y_{\sigma(l-1)}^{ref} - \log Y_{\sigma(l)}^{ref})$ is a sub-band component of the countershading profile at the level l. The luminance Y and the countershading profiles P are calculated in the logarithmic space. Such a coarse approximation of brightness prevents too strong darkening which would happen in the linear space. The sub-band components are not taken from the contrast metric, but are stored in the full resolution in order to preserve the phase information which would be lost by the resolution reduction. The contrasts in the input image are restored by adding the countershading profiles P to the luminance of the input image in the logarithmic space.

Equation (7.5) has several good properties. The uncontrolled amplification of features does not happen because the algorithm is adjusted to the reference image. The counter-shading profiles are created from the reference image, because certain features might have been lost in the input image, thus both detail enhancement and detail reintroduction are solved using one framework. The sharp edges of large scale features are detected by the contrast metric at the top level and at all the levels down to the scale corresponding to the size of these features. Therefore the countershading for such
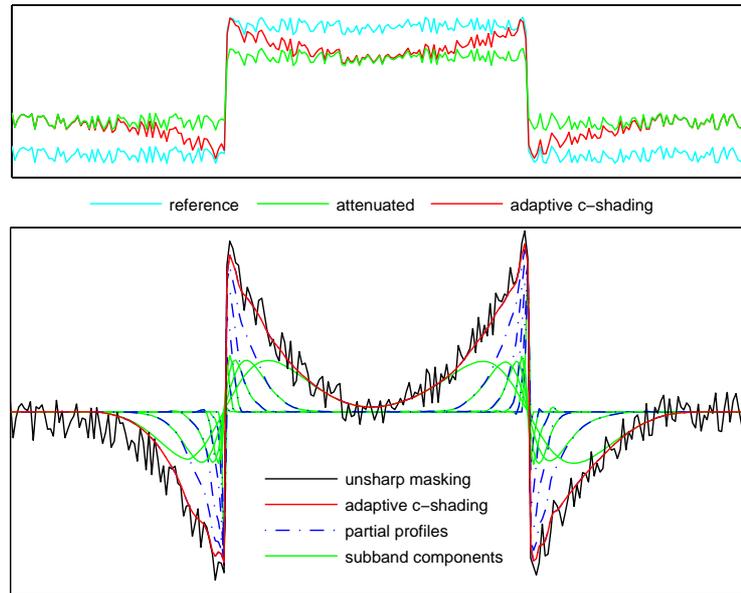
Figure 7.4: Countershading profile for a step edge with details (top), where the step edge is attenuated while the details are preserved with respect to the reference. Adaptive countershading (bottom) recovers the smooth profile which prevents artifacts. The unsharp mask profile is distorted by the high frequency contents of the reference edge and exaggerates details during enhancement as shown in Figure 7.1. Unlike in our method, unsharp masking also requires manual adjustment of the spatial extent and the amplitude of the profile.

edges is progressively composed from the sharp components and the components with a larger spatial extent. This is illustrated in Figure 7.4 along with a comparison to the traditional unsharp masking.

At this stage, the multi-resolution contrast metric assures that the physical contrast of the features in the image restored with the countershading profiles are equal to their physical contrast in the reference image.

7.3.3   Saturation of Profiles

Countershading profiles may increase or decrease luminance values beyond the available dynamic range and cause the saturation to black or white, removal of details, and clearly reveal the presence of a profile. Therefore, the parts of the profile that correct beyond the available range have to be attenuated, as shown in Figure 7.5. The attenuation is performed successively starting from the lowest frequency sub-bands, and separately for the darkening and the lightening parts of the profiles. Each sub-band component is attenuated so that the restored image plus the countershading profiles does not exceed the dynamic range. We motivate our bottom-up approach with the fact that the saturation is mostly caused by the much larger amplitudes of the low-frequency components. Although the strength of the contrast enhancement is reduced

Figure 7.5: Countershading may exceed the dynamic range limit and cause clipping: the profile is then distorted and fine details are removed (left). Successive attenuation of a profile retains all details and as much of the profile as possible (right).
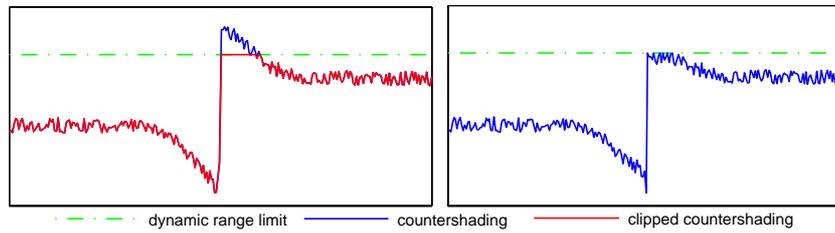
in such case, the asymmetric profiles still increase the perceived contrast, as shown in Figure 7.2, and the degradation of the restored image is prevented.

7.3.4 Natural Image Statistics

One aspect evident in the analysis of Cornsweet [Dooley and Greenfield 1977], is that strong contrasts cannot be corrected with small profiles. However, according to findings in natural image statistics [Bovik et al. 2000], the average amplitudes of frequencies in images tend to decay as a power function, being large for low frequencies and small for high frequencies. Such a phenomenon is known as the power law for the amplitude of frequencies. This observation assures that in the context of natural images we are highly unlikely to encounter the necessity to correct a very high contrast with a small profile.

7.4 Perception of Countershading Profiles

The countershading profiles modulate physical contrasts at edges in an image in order to increase the perceived contrasts between features. However, as soon as the low frequency of a profile can be independently detected, the whole profile is distinguishable at an edge, and the increase in the profile amplitude has no further effect on the perceived contrast [Burr 1987]. To counteract such situations, we develop a visual detection model which assures that the sub-band components of profiles remain below the objectionable amplitude.

We use a model which explains the behavior of the Cornsweet illusion with a good accordance to the perceptual experiments which match the apparent contrast of a profile with the contrast of a step edge [Dooley and Greenfield 1977]. The model is based on a spatial contrast sensitivity function (CSF), but its sensitivity to the frequencies varies with the amplitude of a profile, as shown in Figure 7.6. It therefore estimates the amplitude thresholds above which the components of the profile become individually visible and render a much weaker Cornsweet illusion with objectionable halo artifacts. This model, however, analyzes single Cornsweet profiles on a uniform 2D background and it may be too conservative for natural images. The contrast masking effect, explained in Section 3.2.3, suggests that in areas which already contain features of certain spatial and orientation characteristics, the acceptable amplitude of the profile may be higher
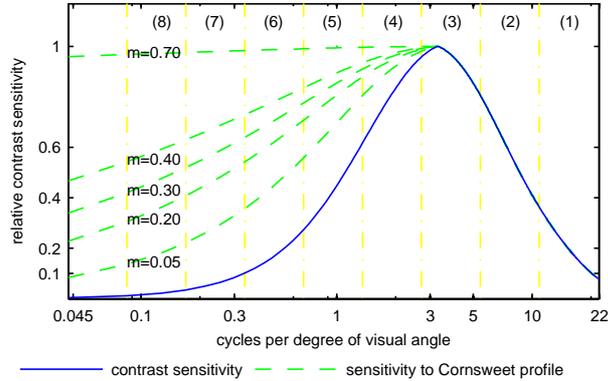
Figure 7.6: Contrast sensitivity function for the threshold effects and the supra-threshold model of tolerance to the magnitude of sub-band components of the Cornsweet profile added to an existing step edge of Michelson contrast m. Vertical lines denote frequency ranges of sub-bands at pyramid levels numbered in the top.

if the profile is composed of signals with similar characteristics. Such a selectivity of independent visual channels fits well to our multi-resolution contrast analysis which uses filter banks motivated by the visual channels in human perception. We therefore improve the model by accounting for this strong effect in human perception.

An important insight from [Dooley and Greenfield 1977], shown in Figure 7.6, is that the sensitivity to the higher frequencies in the Cornsweet profile and in the step edge is similar which justifies our approach to equal the contrast at the profile edge to the contrast of the feature edge.

We take a standard approach to modeling visual detection models which we have briefly introduced in Section 3.2.4. In such a model we combine three effects typical to human vision: luminance masking, spatial contrast sensitivity and contrast masking. For the sub-band component at level l of our pyramid representation, the maximum amplitude of a profile ΔY expressed in luminance is calculated as follows:

$$\Delta Y = \frac{tvi(Y_{mean})}{csf_l} \cdot masking(C_l, Y_{mean}), \tag{7.6}$$

where Y_{mean} is the local mean luminance in the sub-band (considered as the adapting luminance) and C_l is the sub-band contrast at the level l. tvi is the threshold versus intensity function [CIE 1981], csf_l is the relative loss of contrast sensitivity for the spatial frequency band l [Daly 1993], and *masking* describes the contrast masking at the given contrast C_l. The ΔY is calculated for each pixel at the sub-band level l.

The threshold versus intensity function tvi describes the luminance masking effect, explained in Section 3.2.1, by giving the minimum luminance change which is visible at the adapting luminance level. While tvi describes the thresholds for a patch shown on a uniform background, the human visual response to complex images varies depending on the frequencies of the components. Our sensitivity to the contrast at a given spatial frequency is described by the contrast sensitivity function (CSF), which we give in Section 3.2.2. In practice, the CSF function increases the luminance thresholds estimated by tvi for very high and low frequencies.

In our model, however, instead of using the typical CSF which describes the detection thresholds, we use the function given in [Dooley and Greenfield 1977, Fig.6] which determines the tolerable amplitudes of the countershading profiles and also accounts for the increase in sensitivity to low frequencies when a profile is added to an existing edge of Michelson contrast m. Due to the lack of an equation, we provide a fit based on the normalized CSF [Daly 1993]:

$$csf_l(m) = csf_l^{0.74-0.83 \cdot m^{0.35}} \tag{7.7}$$

This function replaces csf_l in equation (7.6) for levels l with frequencies lower than the frequency of peak sensitivity $\approx 5cpd$. Since in the original publication this relation is expressed using Michelson contrast, for compatibility we recalculate here our contrast measure C. The plot is given in Figure 7.6.

Signals added to textured areas are harder to perceive than if added to uniform areas. In Section 3.2.3, we have shown that the existing contrasts in an area mask the contrast of the introduced signal. Contrast masking elevates the detection threshold as a function of the local sub-band contrast C_l in the corrected image:

$$masking(C_l, Y_{mean}) = \max\left(1, \left(\frac{C_l}{T_C(Y_{mean})}\right)^{0.7}\right), \tag{7.8}$$

where T_C is the threshold contrast for the local mean luminance level Y_{mean}, and

$$T_C(Y_{mean}) = \frac{tvi(Y_{mean})}{Y_{mean}}.$$

Contrast masking is modeled by a power function with a typical exponent 0.7 [Daly 1993], which increases the thresholds as soon as the local sub-band contrast is greater than the threshold contrast. Contrast masking is normally considered within the frequency band and the orientation band [Daly 1993]. We ignore the orientation bands due to the high computational costs. In case of the low frequencies, the introduced profile in our case has the same orientation as the existing signal (corrected edge) which gives a strong masking effect.

The maximum tolerable amplitude of the profile ΔY from equation (7.6) sets the limit for the amplitude of the sub-band component of the countershading profile at the given location.

7.5   Implementation

The adaptive countershading algorithm restores the degraded image Y^{inp} with respect to its reference Y^{ref} and outputs an enhanced version of Y^{inp}. The algorithm operates only on luminance values. To process a color image, the RGB channels are converted to Yxy color space and reverted back to RGB with an enhanced luminance channel Y. We clip the colors that are mapped out of the sRGB gamut after processing. Nonstandard references, such as depth maps, may be directly used instead of contrast ratios R. However, such ratios have to be manually scaled to reasonably guide the strength of the contrast enhancement.

The algorithm outline is given in Figure 7.7. The process is fully automatic given the reference image Y^{ref} or arbitrary data passed as the contrast ratios R. Initially, the

---

Adaptive Countershading(Y^{inp}, Y^{ref})

  1 $C^{inp} = \text{contrasts\_pyramid}(Y^{inp})$
  2 $C^{ref} = \text{contrasts\_pyramid}(Y^{ref})$

  3 $P_C = \text{profile\_components}(\log_{10}(Y^{ref}))$
  4 $P = 0$ // countershading profile

  5 $n = \log_2(\min(width, height))$
  6 **for** $l := n..1$
  7    $R(l) = C^{inp}(l)/C^{ref}(l)$
  8    $A_R = 1 - \min(1, R(l))$
  9    $A_S = \text{saturation\_limit}(\log_{10}(Y^{inp}) + P, P_C(l))$
10    $A_V = \text{visual\_model\_limit}(Y^{inp}, C^{inp}(l), P_C(l))$
11    $P{+}= P_C(l) \cdot \text{upsample}(\min(A_R, A_S, A_V))$

12 $RESULT = 10^{(\log_{10}(Y^{inp}) + P)}$

Figure 7.7: Pseudocode implementation of adaptive countershading for contrast restoration in a tone mapped HDR image.

"contrasts_pyramid" function measures local contrasts in the input images at different resolution levels as described in Section 7.3.1. Then the "profile_components" function calculates the sub-band components of the countershading profiles which are used in equation (7.5). The main loop in lines 6-11 builds the countershading profiles P from the sub-band components P_C. For each spatial location at the resolution level l, we calculate the desired and the maximum allowed amplitudes of the sub-band component $P_C(l)$ of the profile. A_R is the desired amplitude of the profile component which would match the original contrast. A_S is the maximum amplitude of the profile component which does not saturate the input image enhanced by already calculated profiles P, and A_V is the maximum amplitude of the profile that cannot be detected in the image, equation (7.6). In line 11, the countershading profile P is enhanced with the new sub-band component at the magnitude required to match the original contrast A_R, but not larger than A_S and A_V. The resolutions of A_R, A_S, A_V correspond to the resolution of the local contrast metric at the given level, and the profile components are in the full resolution of the restored image. The result of the process is the input image Y^{inp} modulated by the calculated countershading profiles P.

The visual model, embedded in the "visual_model_limit" function, assumes an sRGB display and requires that the image frequencies are calculated in cycles per degree of visual angle which depend on the screen resolution and the viewing distance. The results in Section 7.6 are obtained for the resolution 1280×1024 viewed from the distance of $1.5\times$ the screen height. An enhancement of a 1Mpx image requires about a minute on a modern PC. The bottleneck of the algorithm are the convolutions, three are calculated per pyramid level: to measure the contrasts in Y^{ref}, contrasts in Y^{inp}, and to calculate the components of P. A linear filter is used for upsampling.
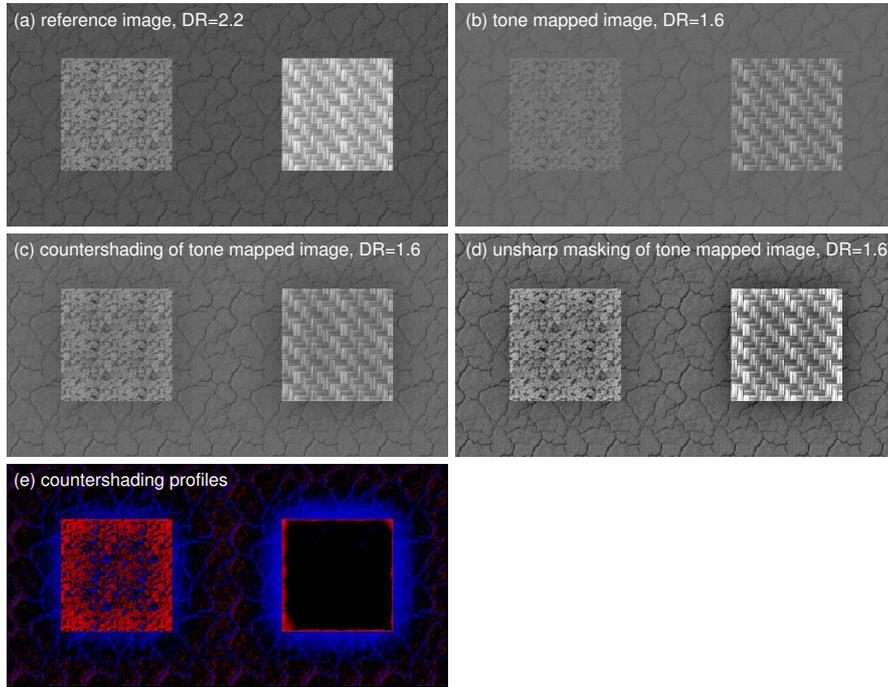
Figure 7.8: Test pattern for the contrast restoration by adaptive countershading. *DR* describes the dynamic range in $\log_{10}$ units of luminance. In the image (e), the blue countershading profiles darken the image and red lighten, their intensity corresponds to the profile magnitude. The right patch in (e) obtained no lightening because of the dynamic range limit. Refer to Section 7.6 for details.

7.6 Results and Applications

We first demonstrate adaptive countershading on a test pattern, Figure 7.8. The reference image (a) contains a textured background and two textured patches. After the tone mapping (b), the texture of the right patch has been preserved, while the textures of the background and the left patch have been attenuated. Also, the contrast between both patches and background has been attenuated. Thus the left patch illustrates global tone mapping and the right one local. The goal of the correction is to restore the contrast between the patches and the background, to restore the visibility of the textures in the background and the left patch, to assure that the texture of the right patch is not emphasized and that objectionable halo artifacts do not appear. The image (a) spans the full dynamic range and in the images (b,c,d) the dynamic range is artificially limited for demonstration purposes. The countershading (c) visibly enhances the contrasts compared to the tone mapped image (b). The texture details of the background and the left patch are restored to almost the same level as in the reference image (a). The contrast and the brightness of the right patch have also improved, although it cannot match the reference due to the dynamic range restrictions. The details of the right patch remain unchanged, which is confirmed in the map (e). Unsharp masking with the spatial extent and the magnitude manually adjusted for correction of the patch to background contrast is shown in image (d). The image is visibly enhanced, however, when compared

Figure 7.9: Image tone mapped using the contrast equalization [Mantiuk et al. 2006] (top/left) and restored by adaptive countershading (bottom/left). The restored image better communicates the brightness relations and the depth in the image. (top/right) shows unsharp masking with parameters set manually to equal dominant countershading profile. Although overall enhancement of unsharp masking is impressive, the changes are hardly controllable and modify the style of the image.

to the reference (a), the background and the right patch details have clearly too strong magnitude. The undesired halo is well visible in the right patch where also some areas became saturated.

7.6.1   Post Tone Mapping Restoration

In Chapter 6 we observe that tone mapping inevitably leads to the attenuation of contrasts and loss of visual information with respect to the original HDR, because the insufficient capabilities of displays require the reduction of dynamic range in HDR images. In Figure 7.9 an HDR image has been tone mapped with a contrast equalization technique [Mantiuk et al. 2006] to reveal the details. Unfortunately, the result does not depict any more the strong brightness difference between the clouds and the building which is very apparent in the original image. This has been detected by the multi-resolution contrast metric and corrected with the appropriate countershading profiles to reintroduce the brightness difference. After the enhancement, the overall appearance of the tone mapped image including the fine details is not changed. Such a correction is not possible with unsharp masking, although the size and the magnitude of the blur in the mask has been manually adjusted according to the metric data. The reason is that the larger kernel, which is required for this correction, amplifies details so strong that the countershading effect disappears. Another example is shown in Figure 7.10, where an HDR image has been tone mapped with the logarithmic mapping [Drago et al.

tone mapped image

countershading of tone mapped image

countershading profiles

Figure 7.10: Image tone mapped using logarithmic mapping [Drago et al. 2003] (top) and restored using adaptive countershading (middle). Subtle changes to the image bring back the contrast at the horizon and the details, but do not change the style of the image. Image courtesy of SpheronVR.

2003]. After using this global operator, some cloud details in the sky are not visible any more, the area around the sun becomes almost isoluminant, and much contrast has been lost in the horizon area. This is automatically restored with the adaptive countershading and the style of the particular tone mapping algorithm is not changed. In both examples the halo artifacts are not disturbing even though a stronger correction was allowed by the visual detection model in Figure 7.9 because of the masking by the clouds. Figure 7.11 illustrates tone mapping of an HDR image with two different techniques. The global operator (a) preserves well brightness relations between lit and shadowed image areas but looses the texture due to a large dynamic range compression. The local operator (b) equalizes all contrasts in the image so that all information is preserved from the original HDR. However, since both lit and shadowed areas are very detailed, there is not enough dynamic range left to depict the shadow boundaries. The brightness relations between image parts are not pronounced. The adaptive countershading automatically restores missing image features in both tone mapping results, rendering a rich image which preserves the original look of each of the operators.

7.6.2 Adaptive Depth Sharpening

Unsharp masking using the depth map of a scene strongly enhances cognition of the spatial distribution of objects [Luft et al. 2006]. We obtain a similar enhancement using

(a) global tone mapping

(b) contrast equalization tone mapping
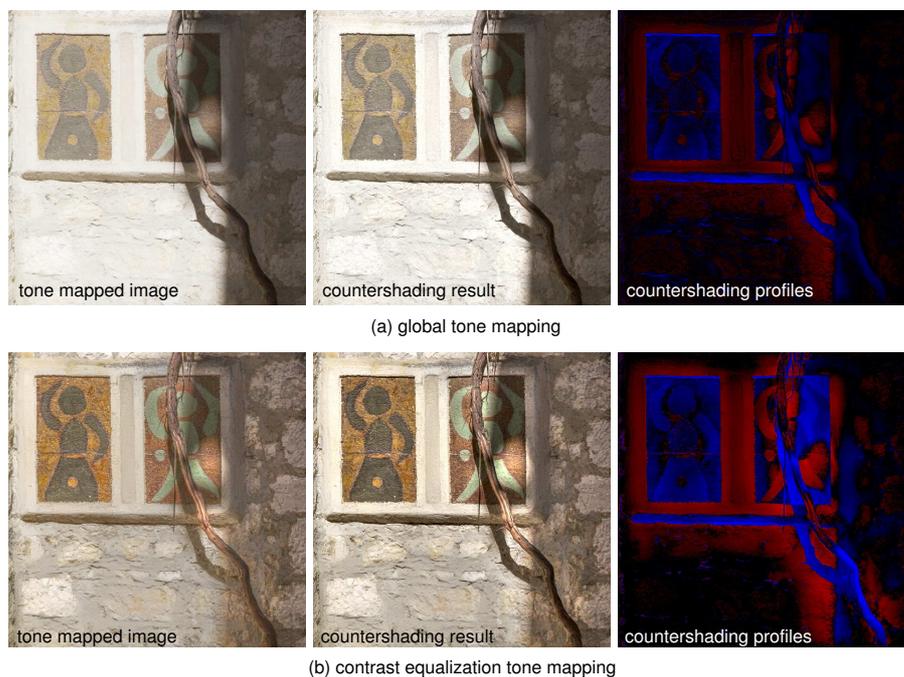
Figure 7.11: An HDR image tone mapped with two different techniques: global operator [Reinhard and Devlin 2005] and contrast equalization [Mantiuk et al. 2006]. Adaptive countershading automatically restores the visibility of texture details in the globally tone mapped image. The shadow boundaries, which became weak after the contrast equalization tone mapping, are automatically enhanced by adaptive countershading so that the brightness relations between the image areas can be well recognized. Note that the particular style of the tone mapping operator remains unchanged.

the adaptive countershading by measuring the relation of a depth map of an image to a uniform map in place of the contrast ratios R in equation (7.3) and by using the depth map instead of the reference luminance in equation (7.5). In our approach, Figure 7.12, the intensity of countershading does not only depend on the depth relations of objects, but is also guided by the visual model which prevents the appearance of unnaturally looking dark outlines over objects further behind in the scene. The visual model limits the countershading strength based on the actual image contents and prevents visible degradations.

7.7 Conclusions

Based on findings from psychophysics, we have explained how to enhance contrast in images using the Craik-O'Brien-Cornsweet illusion in a controlled way by employing the multi-resolution local contrast metric to guide the strength of enhancement and the visual detection model to prevent the appearance of objectionable artifacts. Countershading in most cases cannot be expected to restore the original contrast of the reference, however the enhancement is well visible when profiles are well adjusted and are

Figure 7.12: Poor design of illumination in the scene results in a "flat" look of the image (a). Countershading using depth information (b) enhances cognition of the spatial distribution of objects in the scene (c). The visual models limits the appearance of countershading as halo artifacts. Unnaturally looking dark outlines may appear over objects further behind in the scene if only depth relations are considered, image (d) from [Luft et al. 2006]. Image and depth data from [Scharstein and Szeliski 2003].

masked by image contents.

We have presented an image processing tool to create countershading profiles which are individually and automatically adjusted to enhance selected image features that require such correction when compared to the reference. The same framework is also able to reintroduce lost contrast information. We have demonstrated how it can be used to enhance images using their HDR originals or the depth information as the reference. Comparing to the results of traditional unsharp masking, the enhanced images better communicate information through contrast while the overall appearance is not distorted and the enhancement is achieved within the available dynamic range.

This research direction can be furthered by evaluating the achieved corrections in a perceptual experiment. Such an experiment could measure the actually perceived strength of the countershading enhancement in complex images for stimuli of different scales and given a variety of contrast references.

Chapter 8

Summary

In the following we summarize the contributions of this thesis, draw conclusions, and we end with an outlook on future work.

8.1 Conclusions

The continuing interest of this dissertation was to approach various aspects of tone mapping with a strong emphasis on human visual perception. Through this interdisciplinary point of view, the several methods presented in this thesis successfully improve and evaluate the fidelity of tone mapping.

By approaching the human visual system as a black box we have identified the perceptual effects which significantly contribute to the appearance of scenes and included them in real-time tone mapping with a minimal overhead. These effects are simulated according to known behavior of human visual system with respect to the absolute luminance levels in a scene. This leads to an increased level of realism in the depiction of dynamic HDR contents particularly in applications for HDR video playback or real-time realistic image synthesis. Such effects convey the subjective impression of appearance of night scenes and bright light sources which normally is not communicated on standard displays.

The appearance of natural images is influenced by both sensory and cognitive processes. The knowledge acquired from perception theories lets us design a computational model of the anchoring theory to obtain accurate reproduction of HDR image appearance in terms of lightness. We demonstrated the application of the model to tone mapping, including difficult examples that are not well handled by other algorithms, and we validated the fidelity of its reproduction by successfully simulating the appearance of known perceptual illusions.

Psychophysical models of contrast perception let us investigate the quality of tone mapping in terms of communicating original HDR contents. Such an objective evaluation gives a perceptually meaningful ranking without the burden involved in evaluations with human subjects, and furthermore permits the study of underlying reasons for better visual performance of some algorithms over others. The output of our metric can

be further used to guide the restoration processes.

The perceptual illusions of contrast inspired us to exploit the possibility of strong perceived contrast enhancement within the available dynamic range. Our adaptive countershading can automatically fix any imperfections of an arbitrary tone mapping result through the use of such illusions even if the numerical result is well optimized. Our technique generalizes unsharp masking, a common enhancement tool in photography, by enabling a selective enhancement of image features of various sizes with no manual intervention. Finally, the known characteristics of human perception of contrasts let us build a supra-threshold visual detection model which assures that our enhancements do not introduce objectionable visual artifacts.

The results of presented methods show that the merge of image processing with knowledge of human visual perception can deliver an improved fidelity when depicting HDR contents on displays with limited capabilities. Although the contrast and luminance range of consumer displays grows rapidly, their match to the real-world seem to be still distant and a certain degree of tone mapping is necessary. In that sense, methods presented in this thesis may have long lasting application in the fields of computer graphics, digital photography, video, and cinema. While many aspects of perception have been addressed in this thesis, our work motivates further research in the area. Together with this dissertation, we provide an Open Source software for working with HDR images and video, and we hope it will promote the HDR techniques and facilitate further developments.

8.2   Future Work

Future work in the area of HDR depiction will always be to research techniques that produce results with an increased fidelity to the real-world. One particularly interesting direction to pursue is the appearance of color. The correct reproduction of color is in general neglected under the assumption that it is not significantly influenced by dynamic range reduction. However, certain findings in psychophysics, including the Hunt effect, indicate that this is not entirely true. Also our evaluation and restoration techniques could be more accurate if extended to consider visual information represented by color.

The majority of the methods presented in this thesis is developed for static images. Further investigation could be made, to verify if adaptive countershading and lightness perception model are applicable to time sequences. Particularly, the concept of frameworks gives a unique possibility for a correct simulation of time-dependent local adaptation. A naïve approach to simulate the effect of local adaptation is to smooth the changes of individual pixel values over time, thus simulating the luminance adaptation of the photo receptors. For moving objects that have a significantly different luminance level than the background, this may lead to ghosting effects. In fact, the HVS performs a tracking of moving objects of interest with smooth-pursuit eye movements, therefore the retinal image of these objects is unchanged despite their movement on the display. With the help of frameworks we could follow the objects and perform the local adaptation correctly.

Finally, we recognize that certain techniques presented in the thesis could be improved. Particularly, we would like to design a more robust estimator of luminance perceived as

a white surface within the framework for the lightness perception model. Furthermore, several psychophysical evaluations which are beyond the scope of this thesis would be beneficial to the presented techniques. For instance, the evaluation framework uses psychophysical models obtained for a simple stimulus to compare contents of complex images. Although we have not noticed any incorrectness, a psychophysical validation of the models in the new context is an important next step. Similarly, the actual strength of countershading corrections could be verified in a study with human subjects in which one would compare the original HDR to a restored tone mapping result. Finally, we have presented only a limited number of enhancements made possible by our adaptive countershading technique and we believe that this restoration framework could support more applications.

Bibliography

[Aggarwal and Ahuja 2004] AGGARWAL, M., AND AHUJA, N. 2004. Split aperture imaging for high dynamic range. *Int. J. Comput. Vision 58*, 1, 7–17.

[Arend 1994] AREND, L. 1994. *Lightness, Brightness, and Transparency*. Hillsdale, NJ: Lawrence Erlbaum Associates, ch. Intrinsic image models of human color perception, 159–213.

[Barrow and Tenenbaum 1978] BARROW, H., AND TENENBAUM, J. 1978. Recovering intrinsic scene characteristics from images. In *Computer Vision Systems*. Academic Press, 3–26.

[Bovik et al. 2000] BOVIK, A. C., GIBSON, J. D., AND BOVIK, A., Eds. 2000. *Handbook of Image and Video Processing*. Academic Press, Inc., Orlando, FL, USA.

[Burr 1987] BURR, D. 1987. Implications of the Craik-O'Brien illusion for brightness perception. *Vision Resarch 27*, 11, 1903–1913.

[Burt and Adelson 1983] BURT, P. J., AND ADELSON, E. H. 1983. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications COM-31*, 4 (April), 532–540.

[Calabria and Fairchild 2003] CALABRIA, A., AND FAIRCHILD, M. 2003. Perceived image contrast and observer preference: The effects of lightness, chroma, and sharpness manipulations on contrast perception. *The Journal of Imaging Science and Technology 47*, 479–493.

[Campbell and Robson 1968] CAMPBELL, F. W., AND ROBSON, J. G. 1968. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology 197*, 551–556.

[Chandler and Hemami 2003] CHANDLER, D., AND HEMAMI, S. 2003. Suprathreshold image compression based on contrast allocation and global precedence. In *Human Vision and Electronic Imaging VIII*, SPIE, volume 5007, SPIE, 73–86.

[Chen et al. 2007]   CHEN, J., PARIS, S., AND DURAND, F. 2007. Real-time edge-aware image processing with the bilateral grid. *ACM Trans. on Graphics (Siggraph'07)* (Aug.).

[CIE 1981]   CIE. 1981. *An Analytical Model for Describing the Influence of Lighting Parameters Upon Visual Performance*, vol. 1. Technical Foundations, CIE 19/2.1. International Organization for Standardization.

[Comaniciu and Meer 2002]   COMANICIU, D., AND MEER, P. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 5.

[Daly 1993]   DALY, S. 1993. *Digital Images and Human Vision*. MIT Press, ch. 14: The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, 179–206. ISBN: 0-262-23171-9.

[Debevec and Malik 1997]   DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97 Conference Proceedings*, Addison Wesley, T. Whitted, Ed., Annual Conference Series, ACM SIGGRAPH, 369–378. ISBN 0-89791-896-7.

[Deeley et al. 1991]   DEELEY, R., DRASDO, N., AND CHARMAN, W. N. 1991. A simple parametric model of the human ocular modulation transfer function. *Ophthalmology and Physiological Optics 11*, 91–93.

[Dmitriev et al. 2004]   DMITRIEV, K., ANNEN, T., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2004. A CAVE System for interactive modeling of global illumination in car interior. In *ACM Symposium on Virtual Reality Software and Technology (VRST 2004)*, ACM Press, New York, NY, USA, R. Lau and G. Baciu, Eds., 137–145.

[Dooley and Greenfield 1977]   DOOLEY, R. P., AND GREENFIELD, M. I. 1977. Measurements of edge-induced visual contrast and a spatial-frequency interaction of the Cornsweet illusion. *Journal of the Optical Society of America 67*.

[Drago et al. 2002]   DRAGO, F., MARTENS, W. L., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2002. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Tech. Rep. MPI-I-2002-4-002, Max-Planck-Institut für Informatik, Im Stadtwald 66123 Saarbrücken, Germany, October.

[Drago et al. 2003]   DRAGO, F., MYSZKOWSKI, K., ANNEN, T., AND CHIBA, N. 2003. Adaptive logarithmic mapping for displaying high contrast scenes. In *Proc. of Eu-*

rographics, P. Brunet and D. Fellner, Eds., EG, 419–426.

[Durand and Dorsey 2000] DURAND, F., AND DORSEY, J. 2000. Interactive tone mapping. *11th Eurographics Workshop on Rendering*, 219–230.

[Durand and Dorsey 2002] DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proc. of ACM SIGGRAPH 2002*, Computer Graphics Proceedings, Annual Conference Series, ACM.

[Fairchild and Johnson 2003] FAIRCHILD, M., AND JOHNSON, G. 2003. Image appearance modeling. In *Human Vision and Electronic Imaging VIII*, SPIE, volume 5007, SPIE, 149–160.

[Fairchild 1998] FAIRCHILD, M. D. 1998. *Color Appearance Models*. Addison-Wesley. ISBN 0-201-63464-3.

[Fattal et al. 2002] FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. In *Proc. of ACM SIGGRAPH 2002*, ACM, 249–256.

[Ferwerda et al. 1996] FERWERDA, J. A., PATTANAIK, S., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *SIGGRAPH 96 Conference Proceedings*, Addison Wesley, H. Rushmeier, Ed., Annual Conference Series, ACM SIGGRAPH, 249–258.

[Gilchrist and Cataliotti 1994] GILCHRIST, A., AND CATALIOTTI, J. 1994. Anchoring of surface lightness with multpile illumination levels. *Investigative Ophthamalmology and Visual Science 35*.

[Gilchrist and Radonjic 2005] GILCHRIST, A., AND RADONJIC, A., 2005. Probe disks reveal framework effects within multi-illuminant scenes. Eurpoean Conference on Visual Perception, A Coruña, Spain.

[Gilchrist et al. 1999] GILCHRIST, A., KOSSYFIDIS, C., BONATO, F., AGOSTINI, T., CATALIOTTI, J., LI, X., SPEHAR, B., ANNAN, V., AND ECONOMOU, E. 1999. An anchoring theory of lightness perception. *Psychological Review 106*, 4, 795–834.

[Gilchrist 1977] GILCHRIST, A. L. 1977. Perceived lightness depends on perceived spatial arrangement. *Science 195*, 185–187.

[Gilchrist 1988] GILCHRIST, A. 1988. Lightness contrast and failures of constancy: A common explanation. *Perception & Psychophysics 43*, 415–424.

[Goodnight et al. 2003]   GOODNIGHT, N., WANG, R., WOOLLEY, C., AND HUMPHREYS, G. 2003. Interactive time-dependent tone mapping using programmable graphics hardware. In *Proceedings of the 14th Eurographics workshop on Rendering*, Eurographics Association, EG, 26–37.

[Gösele 2004]   GÖSELE, M. 2004. *New Acquisition Techniques for Real Objects and Light Sources in Computer Graphics*. PhD thesis, Universität des Saarlandes.

[Grossberg and Nayar 2003]   GROSSBERG, M. D., AND NAYAR, S. K. 2003. High Dynamic Range from Multiple Images: Which Exposures to Combine? In *Proc. ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*.

[Havran et al. 2005]   HAVRAN, V., SMYK, M., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Interactive system for dynamic scene lighting using captured video environment maps. In *Rendering Techniques 2005: Eurographics Symposium on Rendering*, Eurographics, Konstanz, Germany, K. Bala and P. Dutré, Eds., EG, 31–42,311.

[Helson 1964]   HELSON, H. 1964. *Adaptation-level theory*. New York: Harper & Row.

[Hoefflinger 2007]   HOEFFLINGER, B., Ed. 2007. *High-Dynamic-Range (HDR) Vision*, vol. 26 of *Springer Series in Advanced Microelectronics*. Springer.

[Hunt 1995]   HUNT, R. 1995. *The Reproduction of Colour in Photography, Printing and Television: 5th Edition*. Fountain Press.

[ITU 1990]   ITU-R RECOMMENDATION BT.709. 1990. *Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange*. Geneva.

[Janesick 2001]   JANESICK, J. R. 2001. *Scientific Charge-Coupled Devices*. SPIE.

[Janssen 2001]   JANSSEN, R. 2001. *Computational Image Quality*. SPIE Press, Bellingham, WA 97227-0010, USA. ISBN 0-8194-4132-5.

[Jobson et al. 1997]   JOBSON, D. J., RAHMAN, Z., AND WOODELL, G. A. 1997. A multi-scale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing: Special Issue on Color Processing 6*, 7, 965–976.

[Kang et al. 2003]   KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2003. High dynamic range video. *ACM Transactions on Graphics (Proceedings of SIG-GRAPH 2003) 22(3)*, 319–325.

[Kingdom and Moulden 1988] KINGDOM, F., AND MOULDEN, B. 1988. Border effects on brightness: a review of findings, models and issues. *Spatial Vision 3*, 4, 225 – 262.

[Krawczyk et al. 2005a]   KRAWCZYK, G., GOESELE, M., AND SEIDEL, H.-P. 2005. Photometric calibration of High Dynamic Range cameras. Research Report MPI-I-2005-4-005, Max-Planck-Institut für Informatik, Stuhlsatzen-hausweg 85, 66123 Saarbrücken, Germany, April.

[Krawczyk et al. 2005b]   KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Lightness perception in tone reproduction for High Dynamic Range images. In *The European Association for Computer Graphics 26th Annual Conference : EUROGRAPHICS 2005*, Blackwell, Dublin, Ireland, M. Alexa and J. Marks, Eds., vol. 24 of *Computer Graphics Forum*, 635–645.

[Krawczyk et al. 2005c]   KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Perceptual effects in real-time tone mapping. In *SCCG '05: Proceedings of the 21st Spring Conference on Computer Graphics*, ACM, Budmerice, Slovakia, 195–202. 1st Best Paper Award.

[Krawczyk et al. 2006]   KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Computational model of lightness perception in High Dynamic Range imaging. In *Human Vision and Electronic Imaging X, IS&T/SPIE's 18th Annual Symposium on Electronic Imaging (2006)*, SPIE, San Jose, CA, USA, B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds., vol. 6057 of *SPIE*, 1–12.

[Krawczyk et al. 2007a]   KRAWCZYK, G., MANTIUK, R., ZDROJEWSKA, D., AND SEIDEL, H.-P. 2007. Brightness adjustment for HDR and tone mapped images. In *The 15th Pacific Conference on Computer Graphics and Applications*, IEEE, Computer Graphics and Applications.

[Krawczyk et al. 2007b]   KRAWCZYK, G., MYSZKOWSKI, K., AND BROSCH, D. 2007. *High-Dynamic-Range (HDR) Vision*, vol. 26 of *Springer Series in Advanced Microelectronics*. Springer, ch. 11. HDR Tone Mapping, 147–178.

[Krawczyk et al. 2007c]   KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2007. Contrast restoration by adaptive countershading. In *The European Association for Computer Graphics Annual Conference EUROGRAPHICS 2007*, Blackwell, vol. 26 of *Computer Graphics Forum*, EG. 2nd Best Paper Award.

[Kuang et al. 2004]   KUANG, J., YAMAGUCHI, H., JOHNSON, G. M., AND FAIRCHILD, M. D. 2004. Testing hdr image rendering algorithms. In *Proc. of IS&T/SID 12th Color Imaging Conference*, SPIE, 315–320.

[Land and McCann 1971]   LAND, E., AND MCCANN, J. 1971. Lightness and retinex theory. *Journal of the Optical Society of America 61*, 1, 1–11.

[Ledda et al. 2005]   LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. In *Proc. of ACM SIGGRAPH 2005*, ACM.

[Li and Gilchrist 1999]   LI, X., AND GILCHRIST, A. 1999. Relative area and relative luminance combine to anchor surface lightness values. *Perception & Psychophysics 61*, 771–785.

[Livingstone 2002]   LIVINGSTONE, M. 2002. *Vision and Art: The Biology of Seeing*. Harry N. Abrams.

[Lubin 1995]   LUBIN, J. 1995. *Vision Models for Target Detection and Recognition*. World Scientific Publishing Company, Inc., ch. A Visual Discrimination Model for Imaging System Design and Evaluation, 245–283.

[Luft et al. 2006]   LUFT, T., COLDITZ, C., AND DEUSSEN, O. 2006. Image enhancement by unsharp masking the depth buffer. *ACM Transactions on Graphics 25*, 1206–1213.

[Lulé et al. 1999]   LULÉ, T., KELLER, H., WAGNER, M., AND BÖHM, M. 1999. LARS II - A High Dynamic Range Image Sensor with a-Si:H Photo Conversion Layer. In *1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*.

[Mantiuk et al. 2004]   MANTIUK, R., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2004. Perception-motivated High Dynamic Range video encoding. *ACM Transactions on Graphics 23*, 3 (July), 733–741. (Proc. ACM SIGGRAPH '04).

[Mantiuk et al. 2006]   MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception 3*, 3, pp. 286 – 308.

[Mantiuk et al. 2007a]   MANTIUK, R., KRAWCZYK, G., MANTIUK, R., AND SEIDEL, H.-P. 2007. High Dynamic Range imaging pipeline: Perception-motivated representation of visual content. In *Human Vision and Electronic Imaging XII*, SPIE, San Jose, USA, B. E. Ro-

gowitz, T. N. Pappas, and S. J. Daly, Eds., vol. 6492 of *Proceedings of SPIE*.

[Mantiuk et al. 2007b] MANTIUK, R., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2007. High Dynamic Range image and video compression - fidelity matching human visual performance. In *Proc. of IEEE International Conference on Image Processing (ICIP 2006)*, IEEE.

[Matkovic et al. 2005] MATKOVIC, K., NEUMANN, L., NEUMANN, A., PSIK, T., AND PURGATHOFER, W. 2005. Global contrast factor-a new approach to image contrast. In *Computational Aesthetics in Graphics, Visualization and Imaging 2005*, Eurographics Association, EG, 159–168.

[Mitsunaga and Nayar 1999] MITSUNAGA, T., AND NAYAR, S. 1999. Radiometric Self Calibration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 374–380.

[Myszkowski and Heidrich 2005] MYSZKOWSKI, K., AND HEIDRICH, W., 2005. High Dynamic Range techniques in graphics: from acquisition to display. Tutorial 7 at the Annual Conference EUROGRAPHICS 2005. Speakers: Michael Goesele, Wolfgang Heidrich, Bernd Höfflinger, Grzegorz Krawczyk, Karol Myszkowski, Matthew Trentacoste.

[Myszkowski et al. 2008] MYSZKOWSKI, K., MANTIUK, R., AND KRAWCZYK, G. 2008. *High Dynamic Range Video*. Synthesis Lectures on Computer Graphics and Animation. Morgan & Claypool Publishers. http://www.morganclaypool.com/toc/cgr/1/1.

[Nayar and Branzoi 2003] NAYAR, S., AND BRANZOI, V. 2003. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proc. of IEEE International Conference on Computer Vision (ICCV 2003)*, IEEE, 1168–1175.

[Nayar and Mitsunaga 2000] NAYAR, S., AND MITSUNAGA, T. 2000. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 472–479.

[Neycenssac 1993] NEYCENSSAC, F. 1993. Contrast enhancement using the laplacian-of-a-gaussian filter. *CVGIP: Graph. Models Image Process. 55*, 6, 447–463.

[Palmer 1999] PALMER, S. 1999. *Vision Science: Photons to Phenomenology*. The MIT Press, ch. 3.3 Surface-Based Color Processing.

[Pattanaik et al. 2000]   PATTANAIK, S. N., TUMBLIN, J. E., YEE, H., AND GREENBERG, D. P. 2000. Time-dependent visual adaptation for fast realistic image display. In *Proc. of ACM SIGGRAPH 2000*, ACM Press / ACM SIG-GRAPH / Addison Wesley Longman, Annual Conference Series, SPIE, 47–54. ISBN 1-58113-208-5.

[Peli 1990]   PELI, E. 1990. Contrast in complex images. *Journal of the Optical Society of America A 7* (Oct.), 2032–2040.

[Polesel et al. 2000]   POLESEL, A., RAMPONI, G., AND MATHEWS, V. 2000. Image enhancement via adaptive unsharp masking. *IEEE Transactions on Image Processing 9*, 505–510.

[Pratt 1991]   PRATT, W. K. 1991. *Digital image processing (2nd ed.)*. John Wiley & Sons, Inc., New York, USA.

[Purves et al. 1999]   PURVES, D., SHIMPI, A., AND LOTTO, B. R. 1999. An empirical explanation of the Cornsweet effect. *J. Neurosci. 19*, 19, 8542–8551.

[Ramasubramanian et al. 1999]   RAMASUBRAMANIAN, M., PATTANAIK, S. N., AND GREENBERG, D. P. 1999. A perceptually based physical error metric for realistic image synthesis. In *Proc. of ACM SIGGRAPH 1999*, ACM Press, ACM, 73–82.

[Ramponi et al. 1996]   RAMPONI, G., STROBEL, N., MITRA, S. K., AND YU, T.-H. 1996. Nonlinear unsharp masking methods for image contrast enhancement. *Journal of Electronic Imaging 5* (July), 353–366.

[Reinhard and Devlin 2005]   REINHARD, E., AND DEVLIN, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1, 13–24.

[Reinhard et al. 2002]   REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics 21*, 3, 267–276.

[Reinhard et al. 2005]   REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann.

[Reinhard 2002]   REINHARD, E. 2002. Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools: JGT 7*, 1, 45–52.

[Robertson et al. 2003]   ROBERTSON, M. A., BORMAN, S., AND STEVENSON, R. L. 2003. Estimation-theoretic Approach to

Dynamic Range Enhancement Using Multiple Exposures. *Journal of Electronic Imaging 12*, 2 (April), 219–228.

[Rock 1983]   ROCK, I. 1983. *The logic of perception*. MIT Press.

[Scharstein and Szeliski 2003]   SCHARSTEIN, D., AND SZELISKI, R. 2003. High-accuracy stereo depth maps using structured light. *CVPR*, 195.

[Schluens and Koschan 2000]   SCHLUENS, K., AND KOSCHAN, A. 2000. Global and local highlight analysis in color images. In *Proc. 1st Int. Conf. on Color in Graphics and Image Processing (CGIP)*, 300–304.

[Seetzen et al. 2004]   SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. In *Proc. of ACM SIGGRAPH 2004*.

[Seger et al. 1993]   SEGER, U., GRAF, H.-G., AND LANDGRAF, M. E. 1993. Vision Assistance in Scenes with Extreme Contrast. *IEEE Micro 12*, 1, 50–56.

[Shaler 1937]   SHALER, S. 1937. The relation between visual acuity and illumination. *Journal of General Physiology 21*, 165–188.

[Smith et al. 2006]   SMITH, K., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Beyond tone mapping: Enhanced depiction of tone mapped HDR images. In *EUROGRAPHICS 2006 (EG'06)*, Blackwell, Vienna, Austria, E. Gröller and L. Szirmay-Kalos, Eds., vol. 25 of *Computer Graphics Forum*, Eurographics, 427–438.

[Spencer et al. 1995]   SPENCER, G., SHIRLEY, P., ZIMMERMAN, K., AND GREENBERG, D. 1995. Physically-based glare effects for digital images. In *Proceedings of ACM SIGGRAPH 95*, ACM, 325–334.

[Stevens and Stevens 1960]   STEVENS, S., AND STEVENS, J. 1960. Brightness function: parametric effects of adaptation and contrast. *Journal of the Optical Society of America 50*, 11 (Nov.), 1139A.

[Stokes and Anderson 1996]   STOKES, M., AND ANDERSON, M., 1996. A standard default color space for the Internet - sRGB. http://www.w3.org/Graphics/Color/sRGB.

[Street August 1998]   STREET, R. August 1998. High dynamic range segmented pixel sensor array. Tech. rep., U.S. Patent 5789737.

[Tomasi and Manduchi 1998]   TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering of gray and colored images. In *Proc. of IEEE International Conference on Computer Vision*, IEEE, 836–846.

[Tumblin and Rushmeier 1993]   TUMBLIN, J., AND RUSHMEIER, H. E. 1993. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications 13*, 6 (Nov.), 42–48.

[Tumblin and Turk 1999]   TUMBLIN, J., AND TURK, G. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *Siggraph 1999, Computer Graphics Proceedings*, Addison Wesley, Los Angeles, A. Rockwood, Ed., Annual Conference Series, ACM, 83–90.

[Tumblin et al. 1999]   TUMBLIN, J., HODGINS, J. K., AND GUENTER, B. K. 1999. Two methods for display of high contrast images. *ACM Transactions on Graphics 18*, 1 (January), 56–94. ISSN 0730-0301.

[Uner and Gustavson 2007]   UNER, J., AND GUSTAVSON, S. 2007. High dynamic range video for photometric measurement of illumination. In *Human Vision and Electronic Imaging XII*, SPIE, volume 6501, SPIE.

[Wallach 1948]   WALLACH, H. 1948. Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology 38*, 310–324.

[Wandell 1995]   WANDELL, B. A. 1995. *Foundations of Vision*. Sinauer Associates, Sunderland, Massachusetts.

[Ward et al. 1997]   WARD, G., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics 3*, 4, 291–306.

[Wesolkowski et al. 2001]   WESOLKOWSKI, S., TOMINAGA, S., AND DONY, R. D. 2001. Shading- and highlight-invariant color image segmentation using the MPC algorithm. In *Proceedings of the Conference on Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts VI (CI-01)*, vol. 4300 of *SPIE Proceedings Series*, 229–240.

[Winkler 2005]   WINKLER, S. 2005. *Digital Video Quality: Vision Models and Metrics*. John Wiley & Sons, Ltd, West Sussex, England.

[Wyszecki and Stiles 2000]   WYSZECKI, G., AND STILES, W. S. 2000. *Color Science: Concepts and Methods, Quantitative Data and Formula*, second ed. John Wiley & Sons, Inc.

[Yoshida et al. 2005]   YOSHIDA, A., BLANZ, V., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Perceptual evaluation of tone

mapping operators with real-world scenes. In *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)*, SPIE, San Jose, USA, vol. 5666 of *SPIE Proceedings Series*, IS&T/SPIE, 192–203.

[Yoshida et al. 2006]   YOSHIDA, A., MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Analysis of reproducing real-world appearance on displays of varying dynamic range. In *EUROGRAPHICS 2006 (EG'06)*, Blackwell, Vienna, Austria, E. Gröller and L. Szirmay-Kalos, Eds., vol. 25 of *Computer Graphics Forum*, Eurographics, 415–426.

Appendix A

Photometric Calibration of HDR Cameras

Ideally, in a photometrically calibrated system the pixel value output by a camera would directly inform about the amount of light that this camera was exposed to. However, in view of display-referred representation it has become important to obtain a visually pleasant image directly from a camera rather than such a photometric image. With the advance of high dynamic range imaging, however, the shift of emphasis in requirements can be observed. Many applications such as HDR video, capture of environment maps for realistic rendering, image-based measurements require photometrically calibrated images with absolute luminance values per pixel. For instance, the visually lossless HDR video compression [Mantiuk et al. 2004] is based on a model of human vision performance in observing differences in absolute luminance. An incorrect estimation of such performance due to the uncalibrated input may result in visible artifacts or less efficient compression. The capture technologies, however, especially in the context of HDR, are very versatile and a simple solution to obtain the photometric output from all types of cameras is not possible.

In this chapter we explain how to perform the absolute photometric calibration of HDR cameras and we validate the accuracy of two HDR video cameras for applications requiring such calibration. For camera response estimation, we adapt an existing technique by Robertson et al. [Robertson et al. 2003] to the specific requirements of HDR camera systems [Krawczyk et al. 2005a]. We determine the absolute photometric calibration to obtain camera output in luminance units. We compare the measurements obtained with the absolute photometric calibration to measurements performed with a luminance meter and discuss the achieved accuracy in the light of possible applications.

A.1 Camera Response to Light

An image or a frame of a video is recorded by capturing the irradiance at the camera sensor. At each pixel of the sensor, photons collected by a light sensitive area are transformed to an analog signal (electric charge) which is in turn read and quantized by a controller. Such a quantized signal is further processed to reduce noise, interpolate

i – image index

j – pixel position index

t_i – exposure time of image i

y_{ij} – pixel value of input image i at position j

$I(\cdot)$ – camera response function

x_j – estimated irradiance at pixel position j

$w(\cdot)$ – weighting function from certainty model

m – pixel value from a set of possible camera output values

Table A.1: Symbols and notation in formulas for response estimation.

color information from the Bayer pattern, or enhance image quality, and is finally output from a camera. The camera response to irradiance, or light, describes the relation between incoming light and produced output value. The details of the capture process are often unknown thus the camera response is conveniently analyzed as a black box, which jointly describes the sensor response and built-in signal processing. In principle, the estimation of a camera response can be thought of as reading out the camera values for each single light quantity, although this is practically not feasible.

The camera response to light can be inversed to retrieve original irradiance value. Directly, the inverse model produces values that are only proportional (linearly related) to the true irradiance. The scale factor in this linear relation depends on the exposure settings and has to be estimated by additional measurements.

The HDR cameras have a non-linear and sometimes non-continuous response to light and their output range exceeds 8 bit. Our choice of the framework for response estimation explained in the following section is motivated by its generality and the lack of restricting assumptions on the form of the response.

A.2   Mathematical Framework for Response Estimation

The camera response is estimated from a set of input images based on the expectation maximization approach [Robertson et al. 2003]. The input images capture exactly the same scene, with correspondence at the pixel level, but the exposure parameters are different for each image. The exposure parameters have to be known and the camera response is observed as a change in the output pixel values with respect to a known change in irradiance. For the sake of clarity, in this section we assume that the only parameter is the exposure time, but in general case it is necessary to know how many times more or less energy has been captured during each exposure. Since the exposure time is proportional to the amount of light captured in an image sensor, it serves well as the required factor. In the mathematical formulas below, we obey the notation given in Table A.1 and consider only images with one channel.

There are two unknowns in the estimation process. The primary unknown, the camera response function I, models the relation between the camera output values and the irradiance at the camera sensor, or luminance in the scene. The camera output values for a scene are provided as input images, but the irradiance x coming from the scene

is the second unknown. The estimation process starts with an initial guess on the camera response function, which for instance can be a linear response, and consists of two steps that are iterated. First, the irradiance from the scene is computed from the input images based on the currently estimated camera response. Second, the camera response is refined to minimize the error in mapping pixel values from all input images to the computed irradiance. The process is terminated when the iteration step does not improve the camera response any more. We explain the details of the process below.

Estimation of Irradiance

Assuming that the camera response function I is correct, the pixel values in the input images are mapped to the relative irradiance by using the inverse function I^{-1}. Such relative irradiance is proportional to the true irradiance from the scene by a factor influenced by the exposure parameters (e.g. exposure time), and the mapping is called linearization of camera output. The relative irradiance is further normalized by the exposure time t_i to estimate the amount of energy captured per unit of time in the input images i at pixel position j:

$$x_{ij} = \frac{I^{-1}(y_{ij})}{t_i}. \tag{A.1}$$

Each of the x_i images contains a part of the full range of irradiance values coming from the scene. This range is determined by the exposure settings and is limited by the dynamic range of the camera sensor. The complete irradiance at the sensor is estimated from the weighted average of this partial captures:

$$x_j = \frac{\sum_i w_{ij} \cdot x_{ij}}{\sum_i w_{ij}}. \tag{A.2}$$

The weights w_{ij} are determined for camera output values by the certainty model discussed later in this section. Importantly, the weights for the maximum and minimum camera output values are equal to 0, because the captured irradiance is bound to be incorrect in the pixels for which the sensor has been saturated or captured no energy.

Refinement of Camera Response

Assuming that the irradiance at the sensor x_j is correct, one can recapture the camera output values y'_{ij} in each of the input images i by using the camera response:

$$y'_{ij} = I(t_i \cdot x_j). \tag{A.3}$$

In the ideal case when the camera response I is perfectly estimated, the y'_{ij} is equal to y_{ij}. During the estimation process, however, the camera response function needs to be optimized for each camera output value m by averaging the recaptured irradiance x_j for all pixels in the input images y_{ij} that are equal to m:

$$E_m = \{(i, j) : y_{ij} = m\}, \tag{A.4}$$

$$I^{-1}(m) = \frac{1}{Card(E_m)} \sum_{i,j \in E_m} t_i \cdot x_j. \tag{A.5}$$

Certainty model

The presence of noise in the capture process is conveniently neglected in the capture model in equations (A.1, A.3). A complete capture model would require characterization of possible sources of noise and incorporation of appropriate noise terms to the equation. This would require further measurements and analysis of particular capture technology in the camera, thus is not practical. Instead, the noise term can be accounted for by an intuitive measure of confidence in the accuracy of captured irradiance. In typical 8-bit cameras, for instance, one would expect high noise in the low camera output values, quantization errors in the high values, and good accuracy in the middle range. An appropriate certainty model can be defined by the following Gaussian function:

$$w(m) = \exp\left(-4 \cdot \frac{(m-127.5)^2}{127.5^2}\right). \tag{A.6}$$

The certainty model can be further extended with knowledge about the capture process. Normally, longer exposure times, which allow to capture more energy, tend to exhibit less random noise than short ones. Therefore an improved certainty model for input images y_{ij} can be formulated as follows:

$$w_{ij} = w(y_{ij}) \cdot t_i^2. \tag{A.7}$$

Such weighting function minimizes the influence of noise on the estimation of irradiance in equation (A.2). This happens apart from noise reducing properties of the image averaging process itself.

Minimization of Objective Function

After the initial assumption on the camera response I, which is usually linear, the response is refined by interactively computing equations (A.2) and (A.5). At the end of every iteration, the quality of estimated camera response is measured with the following objective function:

$$O = \sum_{i,j} w(y_{ij}) \cdot (I^{-1}(y_{ij}) - t_i \cdot x_j)^2. \tag{A.8}$$

The objective function measures the error in the estimated irradiance for input images y_{ij} when compared to the simulated capture of the true irradiance x_j. The certainty model requires that the camera output values in the range of high confidence give more accurate irradiance estimates. The estimation process is terminated as soon as the objective function O falls below predetermined threshold.

The estimation process requires an additional constraint, because two dependent unknowns are calculated simultaneously. Precisely, the values of x_j depend on the mapping of I and the equations are satisfied by infinitely many solutions to I which differ by a scale factor. Convergence to one solution is enforced, in each iteration, through normalization of the inverse camera response I^{-1} by the irradiance causing the medium camera output value $I^{-1}(m_{med})$.

A.3 Procedure for Photometric Calibration

In the following sections we outline a step-by-step procedure for photometric calibration of HDR cameras.

Scene Setup for Calibration

The response estimation algorithm requires that each camera output value is observed in more than one input image. Moreover, frequent observations of the value reduce the impact of noise. Therefore, an ideal scene for calibration is static, contains a range of luminance wider than the expected dynamic range of the camera, and smoothly changing illumination which gives a uniform histogram of output values. Additionally, neutral colors in the scene can minimize the possible impact of color processing in a color camera.

When calibrating HDR cameras, a static scene with a sufficiently wide dynamic range may not be feasible to create. In such a case, it is advisable to prepare several scenes, each covering a separate but partially overlapping luminance range, and stitch them together into a single image.

Capture of Images for Calibration

Input images for the calibration process capture exactly the same scene with varying exposure parameters. A steady tripod and remote control of a camera are essential requirements. A slight out-of-focus reduces edge aliasing due to sensor resolution and limits potential sharpening in a camera, thus makes the estimation process more stable.

HDR cameras often do not offer any adjustment of exposure parameters or available adjustments are not bound to have a linear influence on captured energy. The aperture value cannot be changed to adjust the exposure, because it modifies the depth-of-field, vignetting, and diffraction pattern, thus practically changes the scene between input images. Instead, the optical filters, such as neutral density (ND) filters, can be mount in front of the lens to limit the amount of irradiance at the sensor at a constant exposure time. The ND filters are characterized by their optical density which defines the amount of light attenuation in logarithmic scale. In the response estimation framework, such optical density can be used to calculate a simulated exposure time of captured images:

$$t_i = t_0 \cdot 10^{D_i}, \tag{A.9}$$

where t_i is simulated exposure time of image i captured through an optical filter of density D_i calculated with respect to the true exposure time t_0. If t_0 is not known from the camera specifications, it can be assumed equal to 1. One should make sure that the optical filters are spatially uniform and equally reduce the intensity of all captured wavelengths.

Following the analysis in [Grossberg and Nayar 2003], it can be suggested to acquire two images that are exposed similarly and one that is considerably different. Additionally, when calibrating a video camera one may capture a larger number of frames for each of the exposures. Such a superfluous number of input images will reduce the influence of image noise on the response estimation.

Figure A.1: Cameras used in our experiment: HDRC VGAx (lower left), Silicon Vision Lars III (center), Jenoptik C14 (lower right), and Minolta LS-100 luminance meter (top).

Absolute Photometric Calibration

The images of the calibration scene are input to the estimation framework from Section A.2 to obtain a camera response. For an RGB or multi-spectral camera, the camera response has to be estimated for each color channel separately. Here, we assume that a camera captures monochromatic images with spectral efficiency corresponding to luminance. In case of an RGB camera, an approximation of luminance Y can be calculated from color channels using RGB to XYZ color transform.

The relative luminance values obtained from the estimated response curve are linearly proportional to the absolute luminance with a scale factor dependent on the exposure parameters and the lens system. Absolute calibration is based on the acquisition of a scene containing patches with known luminance Y. The scale factor f is determined by minimizing relative error between known and captured luminance values:

$$Y = f \cdot I^{-1}(m). \tag{A.10}$$

A.4   Example Calibration of HDR Video Cameras

We demonstrate photometric calibration of two HDR video cameras: the Silicon Vision Lars III camera and the HDRC VGAx camera. For comparison purposes we also include the Jenoptik C14 – a high-end, CCD based LDR camera (see Figure A.1). The Lars III sensor is an example of a locally auto-adaptive image sensor [Lulé et al. 1999]: the exposure is terminated for each individual pixel after one out of 12 possible exposure times (usually powers of 2). For every pixel, the camera returns the amount of charge collected until the exposure was terminated as a 12-bit value and a 4-bit timestamp. The HDRC sensor is a logarithmic-type sensor [Seger et al. 1993] and the camera outputs 10-bit values per pixel [Hoefflinger 2007].
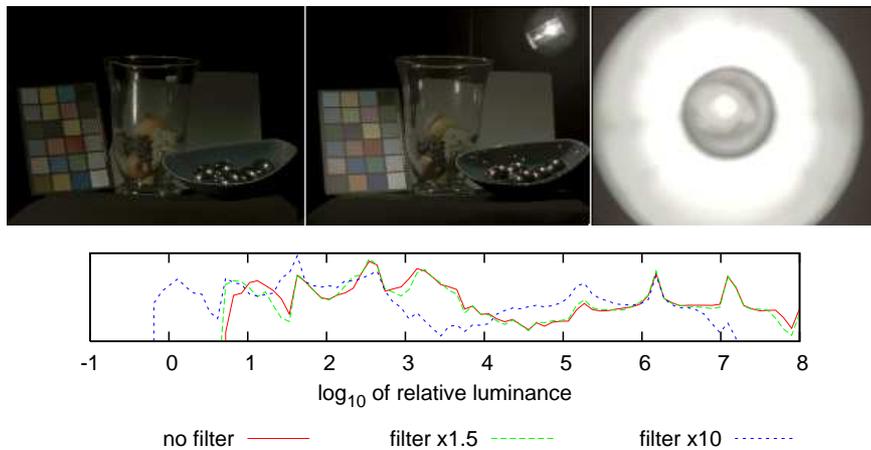
Figure A.2: Three scene setups for the estimation of response curves (tone mapped for presentation). The histogram shows the luminance distribution in the stitched images for acquisition without filter, and using ND filters with ×1.5 and ×10 optical density. This setup covers 8 orders of luminance magnitude.

Estimation of Camera Response

To cover the expected dynamic range of calibrated cameras, we acquire three scene setups with varied luminance characteristic (see Figure A.2): a scene with moderate illumination, the same scene with a strong light source, and a light source with reflector shining directly towards the cameras. Stitching these three images together yields an input for the response estimation algorithm covering a dynamic range of more than 8 orders of magnitude. Each scene setup has been captured without any filter and with a ×1.5 ND filter and a ×10 ND filter. The response of C14 camera was estimated using a series of 13 differently exposed images of a GretagMacbeth ColorChecker.

The estimated responses of the three cameras are shown in Figure A.3. The certainty functions have been modeled using equation (A.6) such that maximum confidence is assigned to the middle of operational luminance range and limits to zero at the camera output levels dominated by noise. A single response curve has been estimated for the monochromatic Lars III camera and separate curves have been determined for the three color channels of the other cameras. As we had access to the raw sensor values of the HDRC camera before Bayer interpolation, we estimated the response curve for each channel directly from corresponding pixels in order to avoid possible interpolation artifacts.

Figure A.3 shows that the response curves of the two HDR cameras both cover a considerably wider range of luminance than the high-end LDR camera that covers a range of about 3.5 orders of magnitude. The different shapes of the HDR response curves are caused by their respective sensor technology and the encoding. The logarithmic HDRC VGAx camera has the highest dynamic range (more than 8 orders of magnitude), but an offset in the A/D conversion makes the lower third of the 10-bit range unusable. The multiple exposure values of the locally auto-adaptive Lars III camera are well visible as discontinuities in the response curve. Note that the luminance range is covered continuously and gaps are only caused by the encoding. The camera covers a dynamic
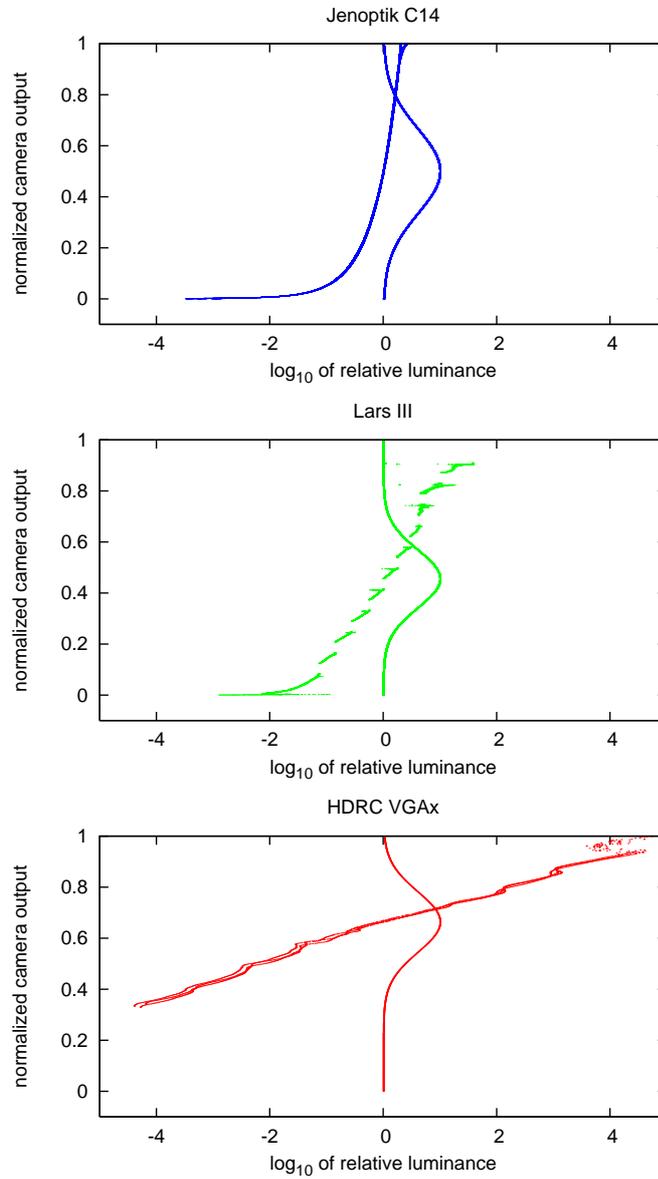
Figure A.3: The estimated response curves and corresponding weighting functions from the certainty model (value 1.0 represents the full confidence in capture accuracy, 0.0 represents no confidence). The peaks of the weighting functions are centered at the middle of the operational range of each camera.
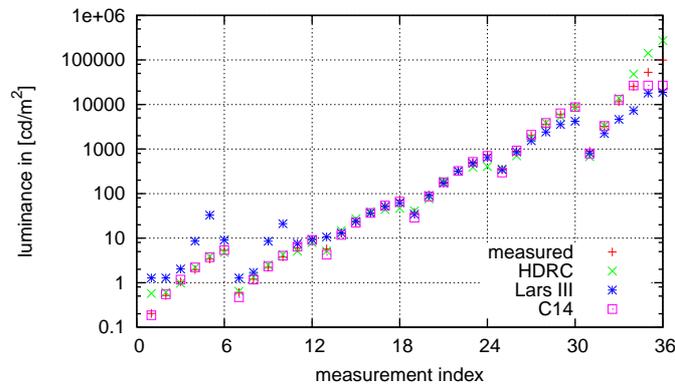
Figure A.4: The results of absolute calibration. The estimated response curves were fitted to the measurements of 6 gray patches of GretagMacbeth ColorChecker chart under 6 different illumination conditions.

range of about 5 orders of magnitude. Noise at the switching points between exposure times is well visible.

Results of Photometric Calibration

The inverse of the estimated responses convert the camera output values into relative luminance values. To perform an absolute calibration, we acquired a GretagMacbeth ColorChecker chart under 6 different illumination conditions. The luminance of the gray patches was measured using a Minolta LS-100 luminance meter yielding a total of 36 samples and an optimal scale factor was determined for each camera. The accuracy of the absolute calibration for the 36 patches can be seen in Figure A.4. The calibrated camera luminance values are well aligned to the measured values proving that the response curve recovery was accurate. To quantify the quality of the absolute calibration, we calculated the average relative error for these data points. For the HDRC camera, relative error in the luminance range of 1–$10,000\,[cd/m^2]$ is 13% while the relative error for the Lars III camera in the luminance range of 10–$1,000\,[cd/m^2]$ amounts to 9.5%. Note that these results can be obtained with a single acquisition. Using multiple exposures, the C14 camera is capable of an average relative error of below 7% in the range 0.1–$25,000\,[cd/m^2]$, thus giving the most accurate results.

A.5 Quality of Luminance Measurement

The described procedure for photometric calibration of HDR cameras proved to be successful, however the accuracy obtained for example HDR cameras is not very high. Although one should not expect to match the measurement quality of a luminance meter, still the relative error of the LDR camera is lower than of HDR cameras. Besides, both HDR cameras keep the error below 10% only in the range of luminance that is much narrower than their operational range. The low accuracy in low illumination is mostly caused by noise in the camera and can be hardly improved in the calibration

process. On the other hand, the low accuracy in high luminance range can be affected by the calibration process: a very bright scene was required to observe high camera output values. The only possibility to get a bright enough scene was to directly capture a light source, but the intensity of the light source might not have been stable during the capture and an additional noise have been introduced to the estimation process.

To improve the results, we fit the estimated response to an a priori function appropriate for the given HDR sensor. Thus, for the HDRC camera we fit the parameters of a logarithmic function $y_j = a * \log(x_j) + b$ and for the decoded values[1] of the Lars III camera we fit a linear function $y_j = a * x_j + b$. We compare the relative errors achieved by the pure response estimation including absolute calibration and the function fit in Figure A.5. The average relative error is equal to about 6% for the HDRC camera and luminance values above $1\,[cd/m^2]$. For the Lars III camera it is also about 6% for luminance values above $10\,[cd/m^2]$. Especially for high luminance values above $10,000\,[cd/m^2]$, the calibration via function fitting provides more accurate results. In addition, the fitting approach allows to extrapolate the camera response for values beyond the range of the calibration scene. To verify this, we acquired an extremely bright patch $(194,600\,[cd/m^2])$ and compared the measurement of the light meter to the calibrated response of the HDR cameras. Only the readout from the HDRC camera derived via function fitting is reliable while the HDRC response curve seems to be bogus in that luminance range. The Lars III camera reached the saturation level and yielded arbitrary results. Likewise, this patch could not be recorded with the available settings of the LDR camera.
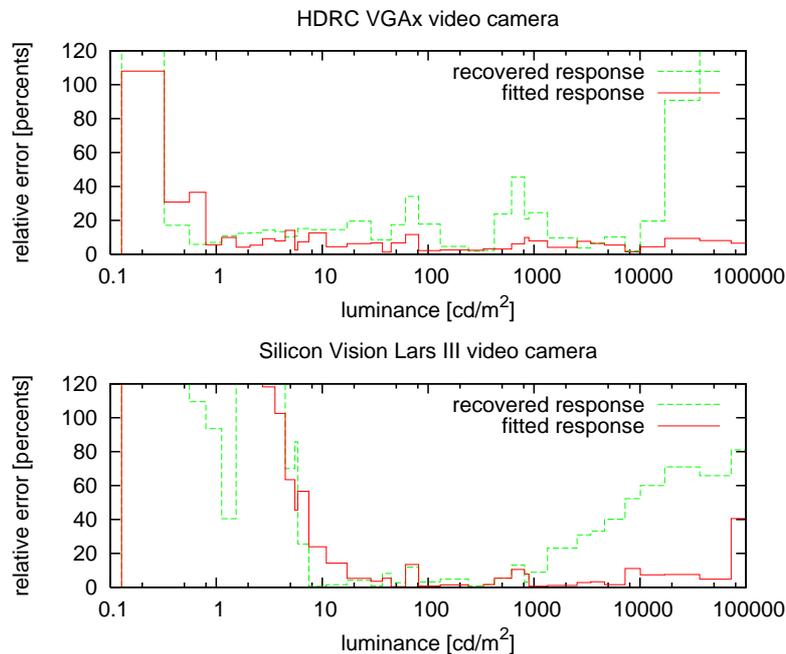

Figure A.5: Comparison of the relative errors in luminance measurement achieved by the pure response estimation including absolute calibration and by the function fit.

---

[1] according to the data sheet, the 16-bit output value of Lars III camera is in fact a composite of a 12-bit mantissa m and a 4-bit exponent value e; i.e. $y_j = m \cdot 2^e$.

A.6 Alternative Response Estimation Methods

In principle, three different approaches can be used to estimate the response of 8-bit cameras ([Reinhard et al. 2005] provides a good survey). The method of Robertson et al. [Robertson et al. 2003] has been selected, because of its unconstrained applicability to varied types of sensors in cameras. For completeness, we briefly discuss here the remaining two methods in view of possible application to photometric calibration of HDR cameras.

The algorithm developed by Debevec and Malik [Debevec and Malik 1997] is based on the concept that a particular pixel exposure is defined as a product of the irradiance at the film and the exposure time, transferred by the camera response function. This concept is embedded in an objective function which is minimized to determine the camera response curve. The objective function is additionally constrained by the assumption that the response curve is smooth, which is essential for the minimization process. Whereas this assumption is generally true for LDR cameras based on CCD technology, the response curve is normally not smooth in locally autoadaptive HDR sensors. Furthermore, the process of recovering the response curve is based on solving a set of linear equations. While the size of the matrix representing these linear equations is reasonable for 8-bit data, memory problems may occur for arbitrary precision data typical to HDR acquisition so that extensive sub-sampling is required.

The method proposed by Mitsunaga and Nayar [Mitsunaga and Nayar 1999] computes a radiometric response function approximated using a high-order polynomial without precise knowledge of the exposures used. The refinement of the exposure times during the estimation process is major advantage, however the process itself is limited to computation of the order of the polynomial and its coefficients. The authors state that it is possible to represent virtually any response curve using a polynomial. This fact is true for LDR cameras based on a CCD sensor, however it is not possible to approximate the logarithmic response of some CMOS sensors in this manner. Polynomial approximation also assumes that the response curve is continuous, which depends on the encoding.

A.7 Discussion

The ability to capture HDR data has a strong impact on various applications, because the acquisition of dynamic sequences that can contain both very bright and dark luminance (such as sun and deep shadows) at the same moment is unprecedented. Photometrically calibrated HDR contents offer further benefits. Perceptually enabled algorithms employed in compression or tone mapping can appropriately simulate the behavior of human visual system. Dynamic environment maps can be captured in real time to faithfully convey the illumination conditions of the real world to rendering algorithms. The results of global illumination solutions can than be directly compared to the real world measurements.

We presented estimation approach for photometric calibration of HDR cameras extended with a function fit. Although the relative error achieved by the function fitting approach is lower, the response estimation algorithm is useful to obtain the exact shape

of the camera response and to give confidence that the chosen a priori function is correct. It can also help to understand the behavior of the sensor, especially if the encoding is unknown. The low precision of the measurements in the luminance range below $10\,[cd/m^2]$ is a clear limitation which can be explained by the high noise level in the sensors. The quality of a high-end CCD camera such as the Jenoptik C14 combined with traditional HDR recovery algorithms still cannot be achieved consistently over the whole dynamic range of the HDR cameras.

The function fitting approach has strong advantages in the quality of the results and the ability to extrapolate from the calibration data. The confidence in extrapolated measurements is however limited and the error cannot be predicted because the exact shape of the response function in this range is unknown. Finally, the accuracy of the photometric calibration is not the only important quality measure. Depending on the application, other issues such as the quantization of the luminance values might have an important influence on the quality of the measurements and need to be further investigated.

Appendix B

Software

To facilitate the work with HDR we have developed a set of software tools that provide a wide range of image and video processing functionality. The tools share a common design pattern based on system pipes which permits to combine them in form of filters in a processing pipeline, similar to the *netpbm* toolkit. Such a pipeline starts with an input program that reads a list of images and forwards the data in a uniform manner to the next tool. The subsequent tools can perform certain image processing operations including cropping, rotating, and tone mapping. The last tool in the pipeline usually stores the processed content.

The communication in the pipeline is facilitated by a generic protocol *pfs* whose implementation is offered as a C++ library. The protocol is also straightforward to implement in other languages. The tools exchange data using the pipes commonly supported by many operating systems. Such a design eases the implementation of new tools and permits to transparently combine programs written in various programing languages including MATLAB®and GNU Octave scripts, Perl, Python and many others. This is very advantageous in rapid prototyping often required in research. The design principles, including the choice of data representation in the pipeline, are published in [Mantiuk et al. 2007a].

The main package of the software is *pfstools* and it is currently extended with *pfstmo* and *pfscalibration*. The whole software is Open Source and can be compiled on several operating systems. It is supported by an active news-group that gathers users and developers. The HDR software code has been adopted to several new 3rd party projects. More details and download options can be found on the project page:
http://www.mpi-inf.mpg.de/resources/pfstools/

B.1   pfstools

pfstools is the main package of the software. It implements the generic communication protocol in the stand-alone library *libpfs*, and contains numerous basic image processing tools including an HDR capable viewer. *pfstools* supports many HDR and standard file formats including: Radiance RGBE, OpenEXR, Tiff, LogLuv, PFM, PPM, RAW formats of digital cameras, and practically all 8-bit formats through ImageMagick®.

Project page:
http://www.mpi-inf.mpg.de/resources/pfstools/

B.2   pfscalibration

pfscalibration package provides an implementation of the [Robertson et al. 2003] method
for the recovery of the response curve of arbitrary cameras. Tools provided in this pack-
age can be used for photometric calibration of both off-the-shelf digital cameras and
HDR cameras as described in Appendix A, and for the recovery of high dynamic range
images from the set of low dynamic range exposures as explained in Section 2.4.1.

Project page:
http://www.mpi-inf.mpg.de/resources/hdr/calibration/pfs.html

B.3   pfstmo

pfstmo package contains implementations of the state-of-the-art tone mapping opera-
tors, including those described in Section 2.5. The implementations are suitable for
convenient processing of both static images and animations, and in the most part have
been used throughout this dissertation.

Project page:
http://www.mpi-inf.mpg.de/resources/tmo/

Curriculum Vitae – Lebenslauf

Curriculum Vitae

1979			born in Szczecin, Poland
1986	–	1994	Primary School, Szczecin, Poland
1994	–	1998	High School Degree, Szczecin, Poland
1998	–	2003	M.Sc. in Computer Science, Szczecin University of Technology, Poland
2003	–	2007	Ph.D. in Computer Graphics, Max-Planck-Institut für Informatik, Saarbrücken, Germany

Lebenslauf

1979			geboren in Stettin, Polen
1986	–	1994	Grundschule, Stettin, Polen
1994	–	1998	Oberschule mit Abiturabschluss, Stettin, Polen
1998	–	2003	Diplom in Informatik, Technische Universität Stettin, Polen
2003	–	2007	Promotion in Computergrafik, Max-Planck-Institut für Informatik, Saarbrücken, Deutschland