
Evaluation and Enhancement of HDR Image Appearance on Displays of Varying Dynamic Range

Akiko Yoshida

**Max-Planck-Institut für Informatik
Saarbrücken, Germany**

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

Eingereicht am 12. 10, 2008 in Saarbrücken.

Betreuender Hochschullehrer — Supervisors

Dr.-Ing. habil. Karol Myszkowski, MPI Informatik, Saarbrücken, Germany

Vorsitzenden — Chairperson

Prof. Dr.-Ing. Philipp Slusallek, Universität des Saarlandes, Saarbrücken, Germany

Gutachter — Reviewers

Dr.-Ing. habil. Karol Myszkowski, MPI Informatik, Saarbrücken, Germany

Prof. Dr. Hans-Peter Seidel, MPI Informatik, Saarbrücken, Germany

Mitarbeiter — Minute writer

Dr. Makoto Okabe, MPI Informatik, Saarbrücken, Germany

Dekan — Dean

Prof. Dr. Joachim Weickert, Universität des Saarlandes, Saarbrücken, Germany

Datum des Kolloquiums — Date of Defense

16. 12, 2008

Akiko Yoshida
Max-Planck-Institut für Informatik
Campus E1.4
66123 Saarbrücken, Germany
yoshida@mpi-inf.mpg.de

Abstract

The increasingly widespread availability of high dynamic range (HDR) technology has led to active study of the characteristics of the human visual system (HVS) in terms of brightness, lightness, contrast, and color perception and the application of the results of these studies to computer graphics. Because the development of HDR technology gives us display devices with much broader dynamic range for both high and low luminances, it is especially important to revise the models of HVS for the luminance ranges which are not covered by classical psychophysics, but required by the new HDR technology.

In this dissertation, we focus on the evaluation and enhancement of the appearance of HDR images as reproduced on low dynamic range (LDR) media. First, we conducted a psychophysical experiment on seven tone mapping operators (TMOs) to assess how tone mapped images are perceived differently by human observers and to find out which attributes of image appearance account for these differences. The results show qualitative differences in TMOs, however, it also turned out that it was hard to choose the consistently best algorithm in terms of the fidelity of tone mapped images to real-world scenes. Based on this result, we conducted another series of psychophysical experiments with a generic easy to understand TMO. This experiment focused on three parameters - brightness, contrast, and color saturation - in order to measure user preference for, and fidelity of, tone mapped HDR images across several types of emulated limited dynamic range displays. The results of this study provide novel guidance for creating more advanced TMO designs.

To evaluate the enhancement of image appearance, we conducted two series of experiments on contrast and brightness enhancement. A non-linear change in perceived contrast with respect to given physical contrast and a different adaptation luminance levels were measured by contrast scaling and contrast discrimination threshold experiments on HDR displays. These results lead to a model of just noticeable difference (JND) unit which provides uniformly changing perceived contrast in complex images. Finally, brightness enhancement caused by the glare illusion was investigated. We employed two profiles to evoke the glare illusion: a point spread function (PSF) of the human eye and a Gaussian kernel. The outcome of this study shows that the glare illusion increases the perceived luminance (brightness) by 20 – 35% for both convolution methods of a PSF and a Gaussian kernel. This means that faithful simulation of the human eye optics, which has been proposed before, is not necessary to achieve a strong brightness enhancement of the glare illusion because the Gaussian kernel, which has no theoretical justification in human perception, evokes the brightness enhancement at the same

or higher strength than the PSF.

Kurzfassung

Mit der zunehmenden Verbreitung von Bildverarbeitung mit hohem Dynamikumfang (HDR), wurden auch die Eigenschaften des menschlichen Sehens in Bezug auf Helligkeit, Kontrast und Farbwahrnehmung eingehend studiert und im Bereich der Computergrafik angewandt. Die Entwicklung der HDR-Technologie hat den Dynamikumfang von Anzeigegeräten sowohl für hohe als auch für niedrige Leuchtdichten deutlich erweitert. Deshalb ist es wichtig, die aktuellen, psychophysischen Modelle des menschlichen Sehens auf die Helligkeitsbereiche auszuweiten, die neuerdings von HDR-Technologien angeboten werden.

In dieser Dissertation konzentrieren wir uns auf die Evaluation und Verbesserung der Darstellung von HDR-Bildern auf Geräten mit niedrigem Dynamikumfang (LDR). Wir haben psychophysische Experimente mit sieben verschiedenen Arten der Dynamikkompression durchgeführt, um herauszufinden, wie unterschiedlich dynamikkomprimierte Bilder von Menschen wahrgenommen werden und welche Bildeigenschaften für die Unterschiede maßgeblich sind. Unsere Experimente haben nicht nur gezeigt, daß es durchaus qualitative Unterschiede zwischen verschiedenen Dynamikkompressionsverfahren (TMO) gibt, sondern auch, daß es kein durchgehend bestes Verfahren in puncto Wiedergabetreue gibt. Darauf aufbauend, haben wir eine zweite Reihe von Studien mit einem generischen, einfach zu verstehenden Dynamikkompressionsverfahren durchgeführt. Unser Hauptaugenmerk lag hierbei darauf, wie die drei Parameter Helligkeit, Kontrast und Farbsättigung die allgemeine Präferenz und Wiedergabetreue von HDR Bildern auf verschiedenen, emulierten Typen von LDR Displays beeinflussen. Das Resultat dieser Studien ist eine neue Herangehensweise an die Entwicklung von Dynamikkompressionsverfahren.

Zur Verbesserung der wahrgenommenen Bilddarstellung, haben wir zwei Experimente zur Helligkeits- und Kontrasterhöhung durchgeführt. Mit Hilfe von Kontrastskalierung und Versuchen zur Kontrastunterscheidungsschwelle auf HDR Displays konnten wir messen, daß der wahrgenommene Kontrast in einem nicht-linearen Verhältnis zum realen Kontrast steht und daß sich die Helligkeitsadaptation des Betrachters mit dem Kontrast verändert. Unsere Ergebnisse resultieren in einem Modell zum gerade noch wahrnehmbaren Unterschied (JND), das einen sich gleichmäßig verändernden, wahrgenommenen Kontrast in komplexen Bildern ermöglicht. Außerdem haben wir die Helligkeitserhöhung durch die Blendillusion untersucht. Um eine Blendillusion hervorzurufen haben wir zwei verschiedene Faltungen auf Bilder angewendet: die Point Spread Function (PSF) des menschlichen Auges und einen Gauss-Kernel. Diese Studie zeigt, daß beide Faltungen die wahrgenommene Helligkeit um 20-35% erhöhen. Obwohl ange-

nommen wird, daß die Blendillusion durch optische Eigenschaften hervorgerufen wird, ist also eine wirklichkeitsgetreue Simulation der menschlichen Optik nicht nötig um eine starke Helligkeitserhöhung zu erreichen, da der Gauss-Kernel eine ähnliche oder gar stärkere Helligkeitserhöhung bewirkt als die PSF. Im Gegensatz zur PSF hat der Gauss-Kernel aber keinen theoretischen Hintergrund in der menschlichen Wahrnehmung.

Summary

As the need for high dynamic range (HDR) technology has increased, knowledge of the human visual system (HVS) has been actively introduced in computer graphics research. Many characteristics of the HVS in terms of brightness, lightness, contrast, and color perception have been studied by psychophysical methods and used in computer graphics applications. However, because new HDR technology provides a much broader dynamic range on display devices for both high and low luminances, it is especially important to revise the models of HVS for the luminance ranges which are not covered by classical psychophysics but required by the new HDR technology.

How to capture or generate an HDR image has been an actively researched topic for many years. An HDR scene can be captured either by new imaging sensors or by conventional LDR cameras with software support, or it can be produced by using 3D renderers. In parallel with the development of HDR image generating techniques, a variety of compression algorithms and HDR file formats have been introduced. However, even if an HDR image is created, LDR displays cannot display it. To solve this problem, many tone mapping operators (TMOs) have been proposed to adjust the dynamic range of an HDR image to that of LDR display devices as a software approach. From the point of hardware approach, there exist dual-modulation HDR displays which can reproduce the display dynamic range between 0.01 and 3,000 cd/m^2 .

Although a variety of TMOs have been developed, no systematic perceptual evaluation exists to reveal their strength or weakness. We conducted a psychophysical experiment based on a direct comparison between the appearance of real-world scenes and tone mapped images of these scenes. The primary goal of this psychophysical experiment was to assess how tone mapped images are perceived differently by determining find out which attributes of image appearance account for the differences between TMOs. This experiment employed five attributes of image appearance: overall brightness, overall contrast, detail reproductions in bright and dark regions, and naturalness. The results show qualitative differences in TMOs, however, it also turns out that it is hard to choose one of the existing algorithms which will consistently perform the best in terms of the fidelity of tone mapped images.

Building on this result, we conducted another series of psychophysical experiments which employed a generic, easy to understand TMO with three parameters: brightness, contrast, and color saturation. These parameters were controllable by human observers, and the goal was to find: a) the most preferred image without any reference (*preference* task) and b) the closest image to the real-world scenes

(*fidelity* task). In addition, several types of limited dynamic range of displays were emulated in the experiments. The results show that this generic TMO is strongly affected by two factors: anchor (reference) white and contrast. It is also shown that the parameters can be automatically estimated based on the characteristics of an image for providing a reasonable “best-guessing” result. Additionally, the outcome from emulating several types of limited dynamic range displays indicates that the best resulting image depends on the purpose of the TMO: the best-looking (preference task) or the best fidelity task. These results for a simple and fundamental TMO will be applicable for more complicated TMO designs.

We also studied the enhancement of contrast and brightness in HDR images. A usual way to scale contrast in image processing changes contrast equally in the whole image. While this contrast scaling method is suitable for LDR displays, it leads to a non-uniform perceived change in contrast in HDR because of the lower contrast sensitivity of the human eyes for low luminances. Based on this non-uniformity of the perceived contrast change, we conducted two psychophysical experiments: contrast scaling and contrast discrimination threshold experiments for a complex image. The results of these two experiments were converted to just noticeable difference (JND) units to construct a model over perceived contrast, physical contrast, and adaptation luminance in order to provide uniformly changing perceived contrast in complex images.

Finally, we measured the brightness enhancement which is caused by the glare illusion. The glare illusion causes an object in an image to look much brighter if it is surrounded by smooth gradient profiles. To evoke the glare illusion, we employed two different convolution methods: a point spread function (PSF) of the human eye and a Gaussian kernel. A brightness matching experiment was conducted for the upper and lower border of the glare illusion by using a modified version of the increment/decrement method. The outcome of this study shows that the glare illusion increases the perceived luminance (brightness) by 20 – 35% for both of the convolution methods PSF and Gaussian kernel. This leads to the conclusion that, although it is believed the glare illusion is related to some optical system, faithful simulation of the human eye is not necessary to achieve the strong brightness enhancement caused by the glare illusion because the Gaussian kernel, which has no theoretical justification, in terms of human perception evokes the brightness enhancement at same or higher strength as the PSF.

Zusammenfassung

Mit der zunehmenden Verwendung von Bildverarbeitung mit hohem Dynamikumfang (HDR) wurden viele Eigenschaften des menschlichen Sehens in Bezug auf Helligkeit, Kontrast und Farbwahrnehmung in Studien erforscht und in der Computergrafik angewendet. Da jedoch die Entwicklung auf dem Gebiet der HDR-Technologie Geräte mit einem deutlich höheren Dynamikbereich, sowohl bei niedrigen als auch hohen Leuchtdichten hervorgebracht hat, müssen die bestehenden Modelle des menschlichen Sehens auf die Helligkeitsbereiche hin untersucht werden, die von der HDR-Technologie ermöglicht werden, von klassischen psychophysischen Studien aber nicht abgedeckt sind.

Das Aufnehmen oder Erzeugen eines HDR Bildes ist in den letzten Jahren stark in den Mittelpunkt der Forschung gerückt. Mittlerweile kann eine HDR-Szene entweder mit Hilfe von speziellen Bildsensoren direkt oder mit herkömmlichen Kameras mit geringerem Dynamikumfang (LDR) und dazu gehörender Software aufgenommen werden, oder sie kann mit einem 3D Renderer produziert werden. Parallel dazu wurden auch eine Vielzahl von Kompressionsverfahren und Dateiformaten für HDR entwickelt. Dennoch kann ein HDR-Bild nicht einfach auf einem LDR Display dargestellt werden. Um dieses Problem zu lösen wurden einige Dynamikkompressionsverfahren vorgestellt, die den Dynamikumfang eines HDR Bildes so reduzieren, daß es auf einem LDR Display dargestellt werden kann. Auf dem Gebiet der Display-Entwicklung gibt es mittlerweile HDR Displays mit doppelter Intensitätsmodulation und einem Kontrastumfang von $0,01 - 3,000 \text{ cd/m}^2$.

Trotz der Vielfalt an Dynamikkompressionsverfahren (TMO), gibt es keine Wahrnehmungsbasierten Studien zu den Vor- und Nachteilen der einzelnen Verfahren. Deshalb haben wir eine psychophysische Studie durchgeführt, die auf dem direkten Vergleich zwischen echten Szenen und tone-mapped Bildern der gleichen Szene basiert. Hauptaugenmerk lag hierbei darauf, wie tone-mapped Bilder im Vergleich wahrgenommen werden und welche Bildattribute für die Unterschiede zwischen verschiedenen tone-mapping Verfahren maßgeblich sind. Fünf Bildattribute wurden hierfür herangezogen: Gesamthelligkeit, Gesamtkontrast, Detailwiedergabe in hellen und dunklen Bereichen und Natürlichkeit des Bildes. Die Ergebnisse zeigen qualitative Unterschiede zwischen tone-mapping Verfahren auf, aber auch, daß kein einzelner TMO stets die wirklichkeitstreuesten Bilder liefert.

Darauf aufbauend haben wir eine weitere Studie mit einem generischen, einfach zu verstehenden TMO mit 3 Parametern durchgeführt: Helligkeit, Kontrast und Farbsättigung. Diese Parameter konnten von den Benutzern verändert werden um 1. ohne Referenzbild das bevorzugteste Bild und 2. das wirklichkeitstreueste Bild verglichen mit der echten Szene zu finden. Zusätzlich haben wir in der Studie ver-

schiedene Typen von LDR Displays emuliert. Die Ergebnisse zeigen, daß dieser generische TMO stark von zwei Faktoren beeinflusst wird: dem Referenz-Weiß und dem Kontrast. Wir konnten außerdem zeigen, daß die Parameter basierend auf Bildcharakteristika automatisch bestimmt werden können. Die Emulation verschiedener LDR Displays hat gezeigt, daß die Ergebnisbilder vom Anwendungszweck des TMO abhängig sind, je nachdem ob das schönste Bild oder das wirklichkeitsgetreuste Bild gefordert ist. Unsere Erkenntnisse anhand eines einfachen und grundlegenden TMO sind auch auf das Design von komplizierteren TMO anwendbar.

Wir haben ebenfalls studiert, wie Kontrast und Helligkeit in HDR Bildern verbessert werden können. In der Bildverarbeitung wird der Kontrast gewöhnlich einfach skaliert, was eine konstante Änderung des Kontrasts im gesamten Bild bewirkt. Während diese Art der Kontrasterhöhung für LDR Displays geeignet ist, bewirkt sie in HDR eine uneinheitliche Änderung des wahrgenommenen Kontrasts, weil das menschliche Auge in dunkleren Bereichen weniger empfindlich für Kontraste ist. Basierend auf dieser Beobachtung haben wir zwei psychophysische Studien zur Kontrastskalierung und zur Kontrastunterscheidungsschwelle durchgeführt. Die Ergebnisse der beiden Experimente wurden in Einheiten gerade noch wahrnehmbaren Unterschieds konvertiert um wahrgenommenen Kontrast, physikalischen Kontrast und Adaption zu modellieren. Das Modell ermöglicht letztendlich einen sich gleichmäßig verändernden Kontrast in komplexen Bildern.

Abschließend haben wir noch die Erhöhung der wahrgenommenen Helligkeit, wie sie durch die Blendillusion hervorgerufen wird. Die Blendillusion, die ein Objekt heller im Bild erscheinen lässt, wird hervorgerufen, wenn das Objekt von einem glatten Helligkeitsgradienten umgeben ist. Um diese optische Täuschung herbeizuführen, haben wir zwei verschiedene Faltungsmethoden verwendet: die Point Spread Function (PSF) des menschlichen Auges und einen Gauss-Kernel. In einer Studie mussten Probanden die Helligkeit von Bildern ohne Blendillusion an ein Referenzbild mit der Illusion anpassen. Dabei hat sich gezeigt, daß die wahrgenommene Helligkeit durch die Blendillusion um 20-35% erhöht wird, und zwar sowohl bei Verwendung der PSF als auch mittels Gauss-Kernel. Obwohl bisher angenommen wurde, daß die menschliche Optik für die Blendillusion verantwortlich ist, lässt sich daraus schließen, daß eine genaue Simulation des menschlichen Auges nicht nötig ist, um eine starke Erhöhung der Helligkeit zu bewirken. Der Gauss-Kernel, der keinen theoretischen Hintergrund in der menschlichen Wahrnehmung hat, bewirkt eine ebenso starke, wenn nicht gar stärkere Erhöhung der wahrgenommenen Helligkeit.

Acknowledgments

First of all, I would like to express my biggest gratitude to my supervisor Dr.-Ing. habil. Karol Myszkowski for his valuable comments, insightful suggestions, continuous support, and endless encouragement. Since I was an undergraduate student in Japan and was enrolled in the computer graphics lecture given by him, he has been inspiring me and keeping me interested in researching in computer graphics, especially in high dynamic range imaging and human visual perception fields. It would not have been possible to write this dissertation if I had not met him 10 years ago.

I would also like to thank Prof. Dr. Hans-Peter Seidel for providing us a great research environment at the Max-Planck-Institut für Informatik (MPII), which is, I believe, one of the best places to work on research in the world, and for his everlasting encouragement and big supports to our projects.

I especially wish to thank Dr. Volker Blanz, Dr. Rafał Mantiuk, Dr. Grzegorz Krawczyk, and Matthias Ihrke, who are the co-authors of my previous publications, to thank Dr. Michael Neff and Tunç Ozan Aydin for their lots of help and comments on my publications and this dissertation, and to thank Christian Fuchs for his help to prepare German translated versions of the abstract and long summary in this dissertation. In addition, I am also fortunate to have quite a number of people who are full of volunteering mind. I am deeply indebted to them for their participation in our psychophysical experiments. The projects on this dissertation would not have been completed without their contributions.

The precious work done by our assistants, Sabine Budde, Conny Liegl, Sonja Lienard, and Svetlana Borodina, helped me a lot for administrative bureaucracy, for business trip planning, and even for some private arrangements to make my life very smooth at MPII and in Germany. I would kindly like to appreciate to their help with their warm hearts. I also wish to thank Thomas Schultz, who is the library administrator of computer graphics group, Christine Kiesel, who is the librarian of MPII, Martin Fuchs, Carsten Stoll, and Martin Sunkel, who are the computer administrators of our group, and the people, who are working for the helpdesk of MPII, for their almost 24.7 support.

Finally, I am sincerely grateful to all of my former and present colleagues at MPII. It was such an excellent place, and I was such a lucky person to meet such nice people here. I have shared plenty of time with them not only for research but for lots of fun. These six years since studying in the Master course till finishing Ph.D. are unforgettable ever.

Contents

1	Introduction	1
1.1	Main Contributions	3
1.2	Chapter Overview	5
2	The Human Visual System (HVS)	7
2.1	The Eye	7
2.2	Visual Adaptation	9
2.3	Visual Sensitivity to Contrast	10
2.3.1	What is Contrast?	10
2.3.2	Contrast Detection and Discrimination Thresholds	13
2.3.3	Contrast Sensitivity Function (CSF)	13
2.3.4	Threshold versus Intensity (TVI) Function	15
2.3.5	Weber's Law	16
2.4	Visual Sensitivity to Brightness	17
2.4.1	What is Brightness?	17
2.4.2	Simultaneous Contrast	17
2.4.3	Anchoring Theory and Gelb Effect	18
2.4.4	Mach Bands and Craik-O'Brien-Cornsweet Illusion	18
2.5	Psychometric Scaling	19
2.5.1	Two-Alternative Forced Choice (2AFC) for 1D Scaling	20
2.5.2	Threshold Measurement	21
2.5.3	Rating Experiments and Multivariate Statistics	21
3	High Dynamic Range Imaging (HDRI)	25
3.1	HDR Image Acquisition	26
3.1.1	HDR Imaging Sensors	26
3.1.2	Multi-Exposure Techniques	27
3.1.3	3D Renderers for HDR Images	29
3.2	Tone Mapping Operators (TMOs)	30
3.2.1	Overview	30

3.2.2	Luminance Domain TMOs	30
3.2.3	Contrast Domain TMOs	32
3.3	Dual-Modulation HDR Displays	34
4	Testing Tone Mapping Operators with Human-Perceived Reality	37
4.1	Introduction	37
4.2	Previous Work	38
4.3	Perceptual Evaluation	39
4.3.1	HDR Image Acquisition	40
4.3.2	Tone Mapping Operators	40
4.3.3	Experimental Procedure	42
4.4	Results	45
4.4.1	Main Effects	45
4.4.2	Mahalanobis Distances	52
4.5	Discussion	57
4.6	Summary	60
5	Analysis of Reproducing Real-World Appearance on HDR Displays	63
5.1	Introduction	63
5.2	Previous Work	64
5.3	Experiments	66
5.3.1	Introduction	66
5.3.2	Subjects	66
5.3.3	Stimuli and Apparatus	66
5.3.4	Generic Tone Mapping	72
5.3.5	Experimental Procedure	73
5.4	Results	75
5.5	Discussion	77
5.5.1	Contrast and Brightness Preference	77
5.5.2	Improved Tone Mapping Algorithm	81
5.5.3	Image and Subject Influence on TMO Parameters	83
5.5.4	Choosing Default TMO Parameters	86
5.5.5	Influence of a Display	87
5.5.6	Preferred Display Device	89
5.6	Summary	90
6	Perception-Based Contrast Enhancement Model for Complex Images in HDR	93
6.1	Introduction	93
6.2	Previous Work	94
6.3	Contrast in Complex Images	95

6.4	Experiments on Measuring Contrast Scaling of a Complex Image in JND	97
6.4.1	Introduction	97
6.4.2	Stimuli and Apparatus	97
6.4.3	Experiment I: Contrast Scaling	99
6.4.4	Experiment II: Contrast Discrimination Threshold	100
6.5	Results	101
6.5.1	Experiment I: Contrast Scaling	101
6.5.2	Experiment II: Contrast Discrimination Threshold	103
6.6	A Model of Uniform Contrast Enhancement for Complex Images	103
6.7	Discussion	105
6.8	Application	107
6.8.1	Uniform Contrast Scaling	107
6.8.2	Global Contrast Scaling	107
6.8.3	Local Contrast Scaling	109
6.9	Summary and Future Work	110
7	Brightness of the Glare Illusion	111
7.1	Introduction	111
7.2	Previous Work	112
7.3	Preliminary Experiments	115
7.4	Experimental Setup for Measuring Brightness of the Glare Illusion	115
7.4.1	Apparatus for All Experiments	115
7.4.2	Methods for Generating the Glare Illusion	116
7.4.3	Stimuli for Experiment I	117
7.4.4	Stimuli for Experiments II, III, and IV	118
7.5	Experimental Procedure for Measuring Brightness of the Glare Illusion	118
7.6	Results	123
7.6.1	Experiment I: Measuring Brightness of Glare Illusion at $L_{bg} = 150 \text{ cd/m}^2$	123
7.6.2	Experiments II – IV: Measuring Brightness of Glare Illusion at $L_{bg} = 50, 100, 200 \text{ cd/m}^2$	127
7.7	Discussion	132
7.8	Summary	133
8	Conclusions and Future Work	135
8.1	Conclusions	135
8.2	Future Work	137
9	Curriculum Vitae	153

Chapter 1

Introduction

Significant progress in developing image and video recording devices has seen the advent of a variety of techniques such as down-sizing, increasing resolution, compression algorithms, and image stabilizers. However, most of the today's common recording devices capture only a part of the scene's luminance range that the human eyes can perceive at once. If an image of a scene whose luminance range is broader than the capability of a camera is taken by such cameras, darker and brighter areas in an image are under- and over-saturated respectively, and details in those regions are not very visible in the image. Such images that contain much broader dynamic range than conventional recording and display devices can handle are called *high dynamic range (HDR)* images. On the other hand, conventional recording, display devices, and images encoded in JPEG and MPEG formats are referred as *low dynamic range (LDR)*.

With the increasing the need for HDR technology, the development speed of HDR technology has been surprisingly rapid. How to capture or produce an HDR image has been actively researched so that an HDR scene can be now captured either with the new imaging sensors or by LDR cameras with some software support, or it can be produced by using 3D renderers. In addition to the existence of many ways to produce HDR scenes, there are also many compression techniques and HDR file formats available, for example, OpenEXR is a widely used HDR file format.

Yet, even if an HDR image is created, it cannot be displayed as an HDR image on an LDR display device. As shown in Figure 1.1, the capabilities of LDR displays (CRT and LCD) are very limited compared to the real-world luminance ranges. To conquer this problem, we can take both software and hardware approaches. In the software approach, a number of *tone mapping operators (TMOs)* have been

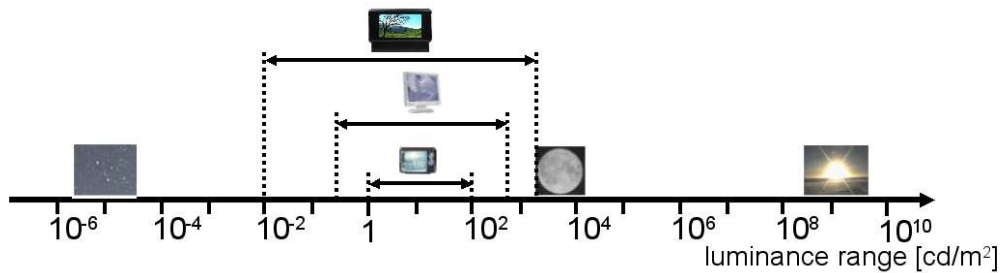


Figure 1.1: Comparison of the luminance ranges of the natural environments and the capabilities of display devices (CRT, LCD, and HDR from bottom to top).

presented to adjust the dynamic range of an HDR image to fit that of an LDR display device. On the other hand, new HDR display techniques have been introduced. The basic idea behind these devices is to mount a second light modulator to enable even stronger backlight, while still maintaining proper reproduction of black levels. Figure 1.1 shows a comparison of luminance ranges reproduced by LDR and HDR displays. Note that HDR display technology expand not only the upper limit of the dynamic range of the display but also its lower limit. The newest HDR display has a luminance range between 0.01 to 3,000 cd/m^2 while modern LCD and plasma displays are capable of 1.0 to 500 or even 1,000 cd/m^2 which is much more than has been possible until recently using CRT displays.

As increasing the need of HDR technology, another issue has started to play an important role in computer graphics: human visual perception. A variety of characteristics in *human visual system (HVS)* have been studied on brightness, lightness, contrast, and color perception by psychophysical studies, and they are well applied into computer graphics applications.

This dissertation is dedicated to evaluation and enhancement of the appearance of HDR images. We first conducted a perceptual evaluation of TMOs which are not directly compared to each other but compared to their corresponding real-world views. There had been conducted a number of psychophysical experiments comparing image-to-image, however, there existed no systematic perceptual evaluation to reveal the strength or weakness of the TMOs. It turns out that it is hard to choose one of the existing algorithms to perform consistently the best in terms of the fidelity of tone mapped images.

Based on the outcome above, we employed a generic TMO whose parameters are the three most important factors: brightness, contrast, and color saturation instead of using complicated TMO algorithms. These parameters were controlled by users

for both image preference and fidelity experiments. In addition, an HDR display emulated several types of display devices with limited dynamic range.

We also took into account the enhancement of contrast and brightness of an image. For contrast perception, there have been presented many psychophysical studies, however, they were conducted on very simple patterns of stimuli and did not consider the cases of HDR. Additionally, it was already known that the perceived contrast changes are not same for all adaptation luminances. We conducted subjective experiments to observe a relationship of perceived contrast change, given physical contrast, and different adaptation luminance in complex images so that the outcome of the experiment proposes a model to keep a uniform perceived contrast change for different adaptation levels.

Finally, brightness boosts caused by the glare illusion based on a point spread function of the human eye and a simple Gaussian kernel was investigated. There are several types of glare effects such as disability glare and veiling glare. The glare illusion is the one which increases the brightness of an object in an image when the object is surrounded by smooth gradient profiles. A number of rendering methods have been proposed for disability glare. These methods are based on some characteristics of HVS which are roughly equivalent to designing a point spread function of the eye optics; therefore, although they were meant to render disability glare, they are actually used to model the glare illusion. A simple Gaussian convolution, which has no strong justification as a point spread function of the eye, was also employed to compare its performance of the glare illusion to that of the optic-based algorithms.

1.1 Main Contributions

The ideas discussed in this dissertation have been already published in international journals and presented at conferences. In this dissertation, they are combined under the common concept of evaluation and enhancement of image appearance in HDR images by using displays with varying dynamic range from LDR to HDR. With respect to these publications, we revise presented methods and demonstrate improved results. The key contributions can be summarized in four parts.

- **Perceptual evaluation of TMOs with human-perceived reality** [[Yoshida et al. 2007a](#)]. We conducted a psychophysical experiment based on a direct comparison between the appearance of real-world scenes and HDR images of these scenes displayed on an LDR monitor. In our experiment, two HDR

scenes were tone mapped by seven existing tone mapping operators. The primary interest of this psychophysical experiment is to assess the differences in how tone mapped images are perceived by human observers and to find out which attributes of image appearance account for these differences when tone mapped images are compared directly with their corresponding real-world scenes rather than with each other. The human subjects rate image naturalness, overall contrast, overall brightness, and detail reproduction in dark and bright image regions with respect to the corresponding real-world scene (Chapter 4).

- **Analysis of reproducing real-world appearance on HDR displays** [Yoshida et al. 2006]. We proposed a novel approach to the tone mapping problem, in which the tone mapping parameters are determined based on the data from subjective experiments, rather than an image processing algorithm or a visual model. A series of experiments was conducted in which the subjects adjusted three generic TMO parameters: brightness, contrast and color saturation. They are to find a) the most preferred image without a reference image (preference task) and b) the closest image to the real-world scene which the subjects are confronted with (fidelity task). We analyze subjects' choice of parameters to provide more intuitive control over the parameters of a TMO. Unlike most of the researched TMOs that focus on rendering for standard low dynamic range monitors, we consider a broad range of potential displays, each offering different dynamic range and brightness. We simulate capabilities of such displays on an HDR display. This allows us to address the question of how tone mapping needs to be adjusted to accommodate displays with drastically different dynamic ranges (Chapter 5).
- **Perception-based contrast enhancement model for complex images in HDR** [Yoshida et al. 2008b]. Contrast in image processing is typically scaled using a power function (gamma) where its exponent specifies the amount of the physical contrast change. While the exponent is normally constant for the whole image, we observe that such scaling leads to perceptual nonuniformity in the context of HDR images. This effect is mostly due to lower contrast sensitivity of the human eyes for the low luminance levels. Such levels can be reproduced by an HDR display while they can not be reproduced by standard display technology. We conducted two perceptual experiments on a complex image: contrast scaling and contrast discrimination threshold, then we derived a model which relates changes of physical and perceived contrasts at different adaptation luminances. We used the model to adjust the exponent value such that we obtain better perceptual uniformity of global and local contrast scaling in complex images (Chapter 6).

- **Brightness of the glare illusion** [Yoshida et al. 2008a]. We measured the perceived luminance of the glare illusion in a psychophysical experiment. To evoke the illusion, an image is convolved with either a point spread function (PSF) of the eye or a Gaussian kernel. It is found that 1) the Gaussian kernel evokes an illusion of the same or higher strength than that produced by the PSF while being computationally much less expensive, 2) the glare illusion can raise the perceived luminance by 20 – 35%, 3) some convolution kernels can produce undesirable Mach-band effects and thereby reduce the brightness boost of the glare illusion. The reported results have practical implications for glare rendering in computer graphics (Chapter 7).

1.2 Chapter Overview

This dissertation is structured as follows. Chapter 2 reviews the human visual system (HVS) and several characteristics of HVS: human visual sensitivities on contrast and brightness and temporal visual adaptation. In Chapter 3, we give an overview of the high dynamic range imaging (HDRI) and its related techniques. Perceptual evaluation of recent TMOs is described in Chapter 4, and an analysis of reproducing real-world appearance on HDR displays is reported in Chapter 5. Then, two enhancement models are presented in Chapters 6 and 7 as perception-based contrast enhancement model for complex images in Chapter 6 and brightness of the glare illusion in Chapter 7. Finally, Chapter 8 concludes this dissertation and gives an outlook for future work.

Chapter 2

The Human Visual System (HVS)

2.1 The Eye

The human visual system (HVS) is a part of the nervous systems which makes it possible to see external objects by interpreting information via incoming lights into the human eye. The human eye has a very complex structure consisting of a number of small organic parts including the cornea, pupil, iris, lens, and retina as its main parts (see Figure 2.1). The cornea provides most of the eye's refractive power, the iris controls the entry of an incoming light into the eye, and the variable opening within the iris is called the pupil. The pupil determines the amount of light which can reach the retina. The lens changes its shape during the act of accommodation in order to provide focal control. Incoming light which goes through the cornea and lens are projected onto the retina which contains photoreceptor cells and neural tissues. The retina is an important component of the eye for considering sensitivity in HVS. Two major photoreceptor cells of the retina are called *rods* and *cones* (see Figure 2.2), which cover the wavelengths of lights between 400 to 700 nm. The rods are sensitive at low illumination levels (scotopic vision) whose peak of the sensitivity is at 498 nm while the cones are sensitive at high illumination levels (photopic vision) with the highest sensitivities at 420 nm for the short wavelength (blue), at 534 nm for the medium wavelength (green), and at 564 nm for the long wavelength (red) [Ferwerda 2001]. Vision models of scotopic, mesopic, and photopic visions are illustrated with the luminance range of the natural environments in Figure 2.3. Once incoming lights reach the retina, the projected light is transferred to the brain via optic nerve to interpret information.

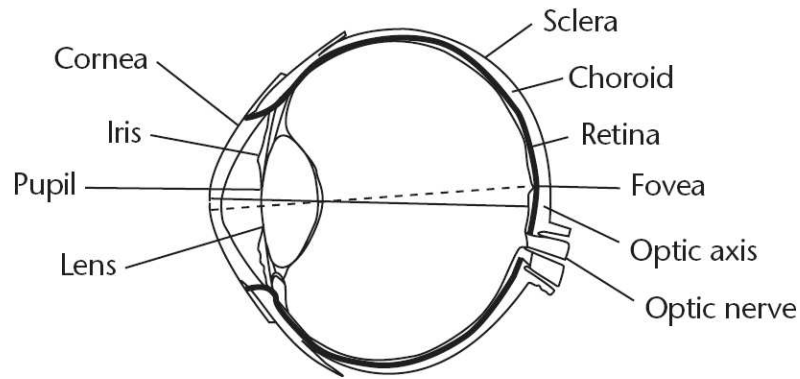


Figure 2.1: The structure of the human eye. After [Atkinson 1988].

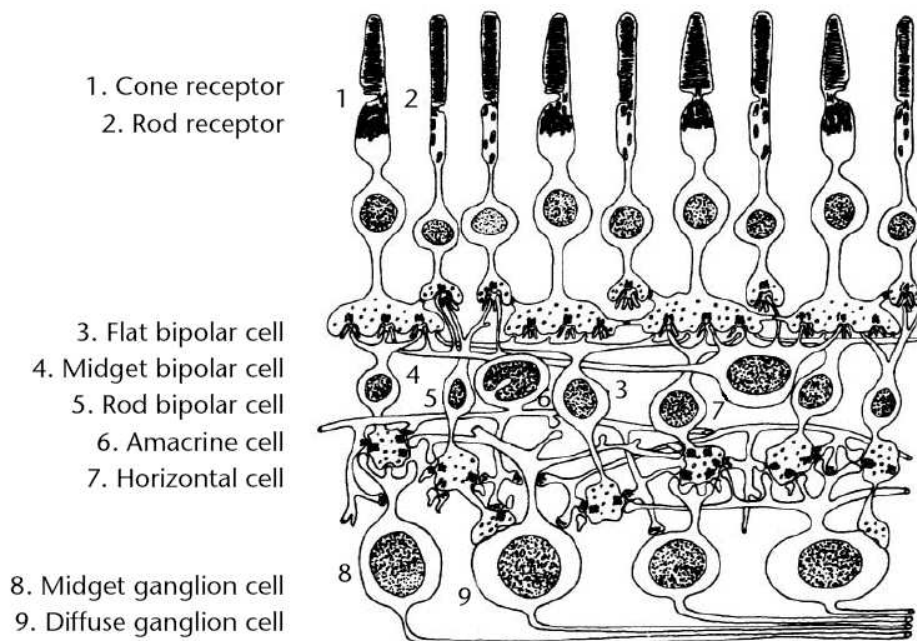


Figure 2.2: The structure of the retina. Two major photoreceptors are numbered as 1) cones and 2) rods. After [Atkinson 1988].

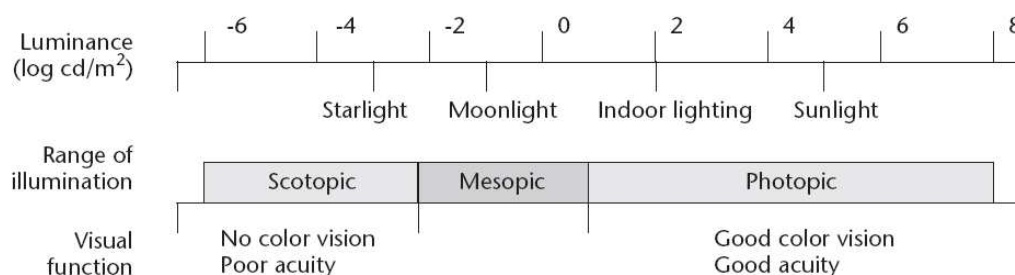


Figure 2.3: The luminance range for scotopic, mesopic, and photopic vision. After [Spillman et al. 1990].

This whole mechanism of the optic system in the human eye and information retrieving process in the brain is called HVS. In this chapter, we firstly review several fundamental characteristics of HVS: *visual adaptation* (Section 2.2) and visual sensitivity to *contrast* (Section 2.3) and *brightness* (Section 2.4). Refer to [Graham 1965, Spillman et al. 1990, Wandell 1995, Barten 1999, Ferwerda 2001] for more details on the other characteristics of HVS. Next, we give an overview of *psychometric scaling*, which are often used in HVS research (Section 2.5).

2.2 Visual Adaptation

As shown in Figure 2.3, the luminance dynamic range of the natural environments is quite broad. HVS are processed on this broad range of luminances by adaptation. Adaptation is achieved through the coordinated action of mechanical, photochemical, and neural processes in HVS [Ferwerda 2001]. For example, the human eyes cannot be adapted to a quick change of illumination instantaneously. Detection threshold of time for HVS can be measured by such an experiment that a subject is seated in a room under a certain ambient illumination for long enough time to be adapted to its intensity. Next, the illumination is changed suddenly, and then the ability of the subject for detecting a small luminance difference on a stimuli is examined. The length of time until he/she starts being able to detect the luminance difference is measured. Figure 2.4 shows the time course of *dark adaptation*, which HVS adjusts from bright to dark illumination level. The crossing point of the rod and cone curves is known as *Purkinje break* which indicates the transition from detection by the cones to detection by the rods [Kling and Riggs 1971].

The inverse case of dark adaptation, which HVS adjust from dark to bright illumination level, is called *light adaptation*. Figure 2.5 shows the time course of

light adaptation for the rods and cones. As shown in the figure, light adaptation is more rapid than dark adaptation. This characteristic of HVS for temporal visual adaptation is applied to design time-dependent tone mapping operators [Ferwerda et al. 1996, Pattanaik et al. 2000, Durand and Dorsey 2000, Reinhard et al. 2005].

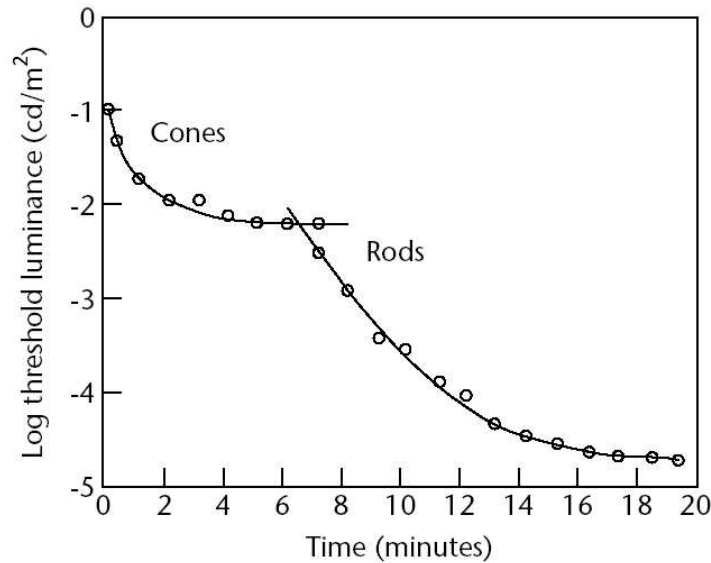


Figure 2.4: The time course of dark adaptation for the rods and cones. Image after [Ferwerda 2001]. Original data was measured by Hecht in [Murchison 1934].

2.3 Visual Sensitivity to Contrast

2.3.1 What is Contrast?

Contrast is the difference in visual properties which makes the representation of an object distinguishable from the others or from the background. There have been presented a number of ways to define contrast for the case of simple stimuli which contain two extreme intensities such as sinewave patterns or Gabor patches. The simplest way to calculate contrast is to take their ratio (*Simple Contrast*) as

$$C_s = \frac{L_{\max}}{L_{\min}}, \quad (2.1)$$

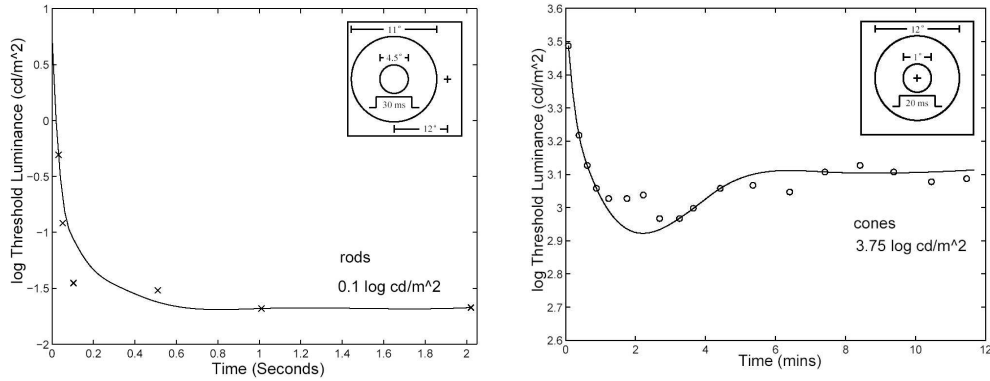


Figure 2.5: The time course of light adaptation for the rods (left) and cones (right). Images after Adelson [Adelson 1982] (left) and Baker [Baker 1949] (right).

where L_{\max} and L_{\min} are the maximum and minimum luminances. Simple contrast is often used for photography to specify the ratio between bright and dark areas in an image, but C_s is not practical to use for psychophysics. C_s is sometimes calculated in logarithmic domain: $\log C_s = \log \frac{L_{\max}}{L_{\min}}$ (*Logarithmic Ratio*) which actually denotes the logarithmic difference between L_{\max} and L_{\min} . *Signal to Noise Ratio (SNR)* is similar to the Logarithmic Ratio:

$$SNR = 20 \cdot \log_{10} \frac{L_{\max}}{L_{\min}} \quad (2.2)$$

which is given in the unit of decibels (*dB*).

In psychophysics experiment, *Weber's fraction* and *Michelson contrast* are often used contrast definitions. They are formulated as

$$C_w = \frac{L_{\max} - L_{\min}}{L_{\min}} \quad (2.3)$$

$$C_m = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} \quad (2.4)$$

for Weber's fraction C_w and Michelson contrast C_m , respectively. The ranges of values in contrast are -1.0 to ∞ for Weber's fraction and 0 to 1.0 for Michelson contrast. Weber's fraction is commonly used for measuring the local contrast of a single stimulus of uniform luminance on a uniform background while Michelson contrast is used for periodic patterns such as sigmoidal gratings. Weber's fraction is reviewed in Section 2.3.5.

All of the definitions presented above can be applied only for simple patterns. On the other hand, defining contrast in complex images is not as easy as the above definitions. Peli proposed a way to define contrast in complex images by using quasi-local definition of contrast [Peli 1990]. This definition is based on an idea that, because human contrast sensitivity highly depends on spatial frequency, especially near thresholds, contrast for each spatial frequency band is calculated separately to address the variation of contrast across an image. Peli's contrast is given in a 2D array for each band of spatial frequencies as

$$C_p(x, y) = \frac{b(x, y)}{l(x, y)} \quad (2.5)$$

where $b(x, y)$ is the band-pass filtered version of an original image, and $l(x, y)$ is the low-pass filtered version which includes all energy below the band (Note that $l(x, y) > 0$). x and y represent the coordinate of a pixel. In this method, Gaussian pyramid L and Laplacian pyramid B [Adelson and Burt 1981] are constructed for a given image. The band-pass image $b(x, y)$ is taken from the pyramid B , and the low-pass image $l(x, y)$ is taken from two levels below in the pyramid L (see Figure 2.6). Refer to [Peli 1990] for the details of his contrast definition and its applications.

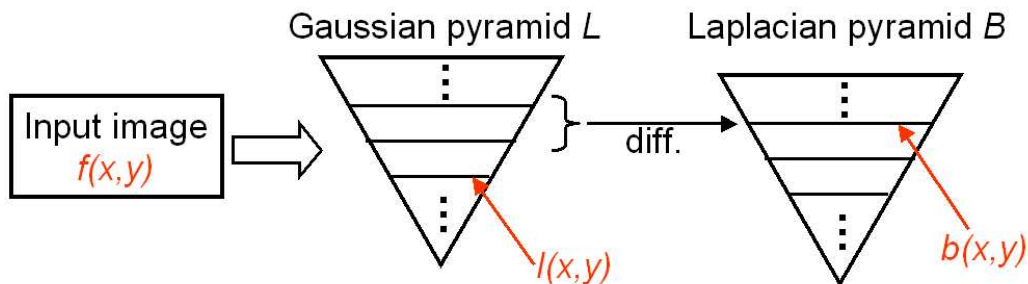


Figure 2.6: An illustration of the contrast definition in [Peli 1990].

The definitions of contrast reviewed above deal with *physical* contrast. However, contrast has another side: *apparent* contrast which denotes the perceived phenomena of contrast by HVS. Even if the same physical contrast stimuli are given, they can be sometimes judged to be stronger or weaker depending on such factors as image contents, adaptation luminance, and display devices. Because it is difficult to manipulate physical contrast due to display device limitations, considering apparent contrast based on HVS can be beneficial to enhance the perceived contrast.

In the following sections, we review several factors of contrast perception: *contrast detection and discrimination thresholds*, *contrast sensitivity function (CSF)*, *threshold versus intensity (TVI) function*, and *Weber's law*.

2.3.2 Contrast Detection and Discrimination Thresholds

Before describing perceptual theories of contrast, it is important to clarify the difference between *contrast detection* and *contrast discrimination* thresholds. Both of them are often used in psychophysical studies to measure perceptual characteristics of the human eye [Barten 1999]. The *contrast detection* threshold measures the smallest visible contrast of a given stimulus on a uniform background (see Figure 2.7 left). On the other hand, *contrast discrimination* threshold is a measurement of the smallest visible difference at a given stimulus with given pedestal contrast (see Figure 2.7 right). *Contrast detection* is a special case of *contrast discrimination* when its pedestal contrast (green part of the discrimination plot in Figure 2.7) is zero. For measuring contrast detection and discrimination thresholds, there are a number of psychophysical methods such as increment/decrement method, staircase method, Parameter Estimation by Sequential Testing (PEST) [Taylor and Creelman 1967], and QUEST [Watson and Pelli 1983].



Figure 2.7: Illustration of contrast detection and discrimination thresholds.

When a stimulus is 1) significantly above the detection or discrimination threshold or 2) very close or equal to the threshold, they are called 1) *suprathreshold* and 2) *subthreshold* or *threshold*, respectively. *Contrast detection* measures a performance of the human eye at *subthreshold* while *contrast discrimination* measurement deals with *suprathreshold* characteristics. *Contrast detection* thresholds have been modeled such as *Contrast Sensitivity Function (CSF)*, *Threshold versus Intensity (TVI) function*, and *Weber's law*, which are briefly reviewed in the following sections.

2.3.3 Contrast Sensitivity Function (CSF)

Contrast sensitivity function (CSF) is one of the well-known HVS characteristics in contrast perception. It describes the sensitivity of the human eye as a function of spatial frequencies. As shown in Figure 2.8, although the amplitude of

signals are decreasing uniformly, the perceived signals disappear non-uniformly. The sensitivity of the human eyes for spatial frequency is band-pass type, i.e., our eyes are the most sensitive at the medium spatial frequencies around 4 – 10 cycles per degree of visual angle. From the concept of image perception on the display, this means that the *CSF* depends on the viewing distance. Additionally, it is also affected by adaptation luminances (see Figure 2.9). Several models of the *CSF* have been used in computer graphics field [[Mannos and Sakrison 1974](#), [Daly 1993](#)].

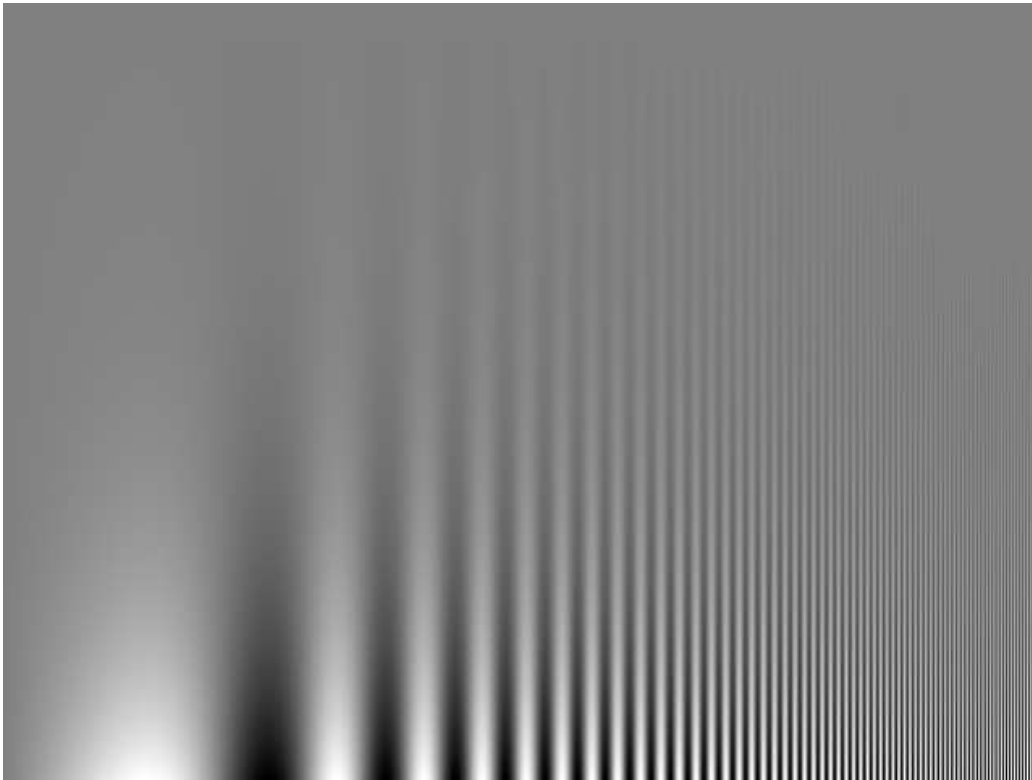


Figure 2.8: Contrast sensitivity function (CSF) over spatial frequencies (horizontal axis). The amplitude of signals decreases uniformly along the vertical axis for all frequencies, but the signals disappear non-uniformly for the human eyes. Our eyes are the most sensitive at the medium spatial frequencies. After [[Campbell and Robson 1968](#)].

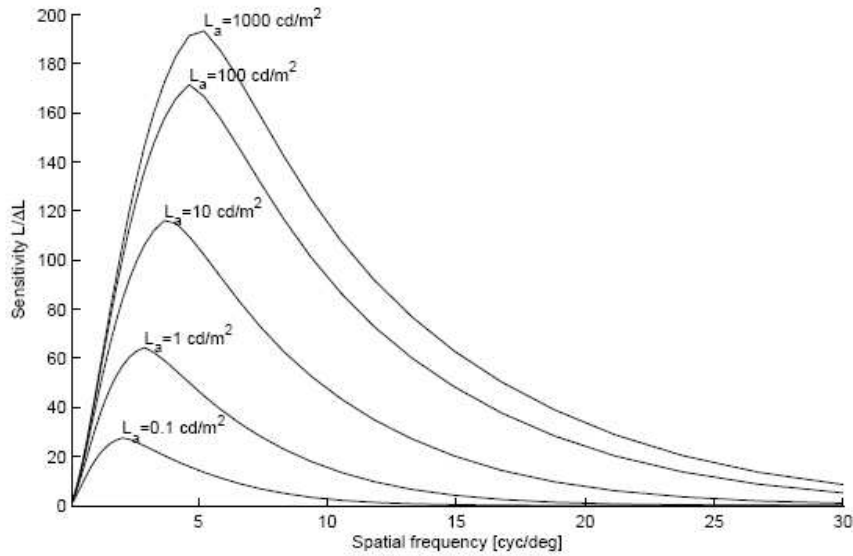


Figure 2.9: Family of contrast sensitivity functions (CSFs) with different adaptation luminances. Those CSFs are plotted based on [Daly 1993].

2.3.4 Threshold versus Intensity (TVI) Function

A measurement of visual adaptation can be obtained by the detection threshold method and then given as *Threshold versus Intensity (TVI)* function. Such an experiment is conducted as follows: a subject has been seated in front of a big dark screen for long enough time to be adapted to the illumination of the screen. In each trial, a disk of light at the center of the screen flashes for a few hundred milliseconds. The subject must answer whether the disk appeared or not. If the answer is yes, the intensity of the disk is decreased or vice versa. By repeating those steps, the detection thresholds against the corresponding background luminance are given in the *TVI* function (see Figure 2.10). The formulation of the *TVI* function in computer graphics has been modeled in several ways [Daly 1993, Ferwerda et al. 1996, Ashikhmin 2002]. The relation between CSF and TVI can be given by considering the maximum spatial frequency for a given adaptation luminance level (refer to [Mantiuk et al. 2006]).

Based on the *TVI* function, *Contrast versus Intensity (CVI)* function can be introduced as $cvi(L_{\text{adapt}}) = \frac{tvi(L_{\text{adapt}})}{L_{\text{adapt}}}$ which indicates *contrast detection* thresholds at a given adaptation luminance L_{adapt} .

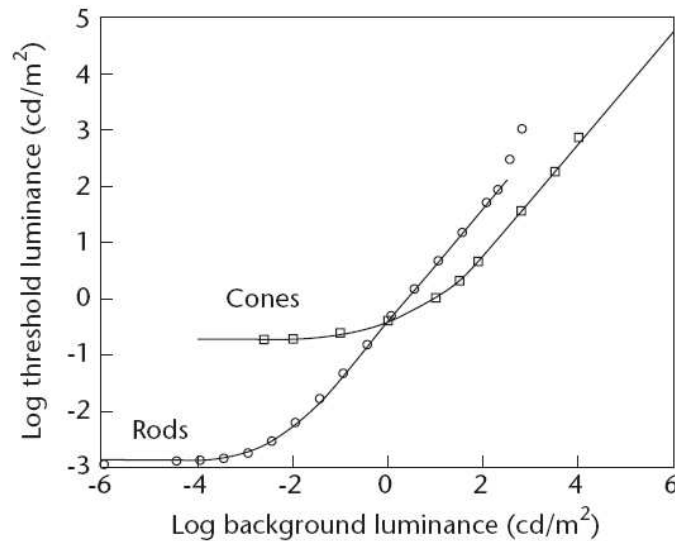


Figure 2.10: Threshold versus intensity (TVI) functions for the rods and cones. After [Ferwerda 2001].

2.3.5 Weber's Law

It is observed that the curve of the *TVI* function for rods is almost flat below $-4 \log \text{cd}/\text{m}^2$ as seen in Figure 2.10. After this point, the curve becomes nonlinear, then it becomes linear after $-2 \log \text{cd}/\text{m}^2$. For the cones, the curve is almost flat below $0 \log \text{cd}/\text{m}^2$, nonlinear below $2 - 3 \log \text{cd}/\text{m}^2$, and then linear for the rest. This linear relationship is called *Weber's law* which can be formulated as

$$\Delta L = kL \quad (2.6)$$

where L is a luminance value and k is a constant factor which is experimentally defined [Kling and Riggs 1971]. Weber's law describes the change in stimulus luminance that can just be discriminated (ΔL) is a constant fraction (k) of the starting luminance of the stimulus (L). It indicates that HVS have constant contrast sensitivity because the increase in thresholds with background luminance is corresponding to luminance with constant contrast. As discussed in Section 2.3.1, this law can be also used as one of the definitions of contrast when $L = L_{\min}$ and $\Delta L = L_{\max} - L_{\min}$ in Equation (2.3). Note that Weber's law holds for the luminance range greater than $500 \text{cd}/\text{m}^2$. Within the luminance range of the conventional displays ($1 - 500 \text{cd}/\text{m}^2$), Weber's law is not an accurate model of luminance masking.

2.4 Visual Sensitivity to Brightness

2.4.1 What is Brightness?

The term “brightness” (or “darkness” as its opponent) is often used to describe the sensation of light for subjective or relative measurement while the absolute measurement of light is given by luminance in the unit of cd/m^2 . It is very important to point out that brightness perception depends on many factors such as spatial and temporal distribution of light in the visual field. A number of effects can occur in brightness perception, for example, *simultaneous contrast*, *Gelb effect*, *Mach bands*, and *Craik-O’Brien-Cornsweet illusion*. They are briefly reviewed in the following sections.

2.4.2 Simultaneous Contrast

An object is perceived brighter or darker according to whether it reflects a higher or lower percentage of the incident light under natural environment. However, an object of moderate reflectance appears relatively brighter or darker according to whether spatially adjacent areas are considerably darker or brighter than the object itself. This dependence on adjacent areas for brightness or darkness is called *simultaneous contrast*. Figure 2.11 illustrates a classical example of simultaneous contrast. All of the inner squares have the same luminance, but their brightness varies according to their surroundings.

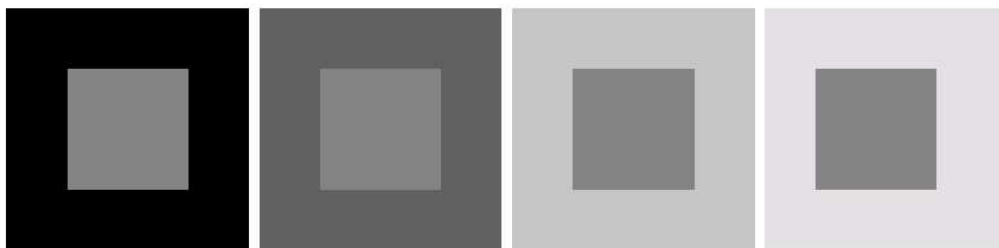


Figure 2.11: An example of simultaneous contrast. The inner gray squares can be perceived differently in brightness, although all of them are the same in luminance.

2.4.3 Anchoring Theory and Gelb Effect

Another example of how HVS depend on relative luminance is the *Gelb effect*. If an object of low reflectance (Object A) in a dark environment is illuminated by a light source which is not directly visible for a subject, Object A looks medium gray or white. Then, if another object of higher reflectance (Object B) is placed next to Object A, Object A now looks black while Object B looks white. The perceived blackness of Object A depends on the relative reflectance of Object B. The *Gelb effect* supports the *Anchoring Theory* presented by Gilchrist et al. [Gilchrist et al. 1999]. Krawczyk et al. extended the *Anchoring Theory* of lightness perception for complex images and applied as a tone mapping operator [Krawczyk et al. 2005].

2.4.4 Mach Bands and Craik-O'Brien-Cornsweet Illusion

Brightness perception also causes border contrast effects such as *Mach bands* and *Craik-O'Brien-Cornsweet illusion*. The illusion of *Mach bands* is illustrated in Figure 2.12. Mach bands are observed as over- and under-shoots of brightness if two uniform objects of high and low reflectance are connected by a gradient. The ramp should be neither too shallow nor too steep to make *Mach bands* visible. In the meaningful range of gradients to produce *Mach bands*, it is known that the steeper gradient causes stronger effect of *Mach bands* [Lotto et al. 1999].

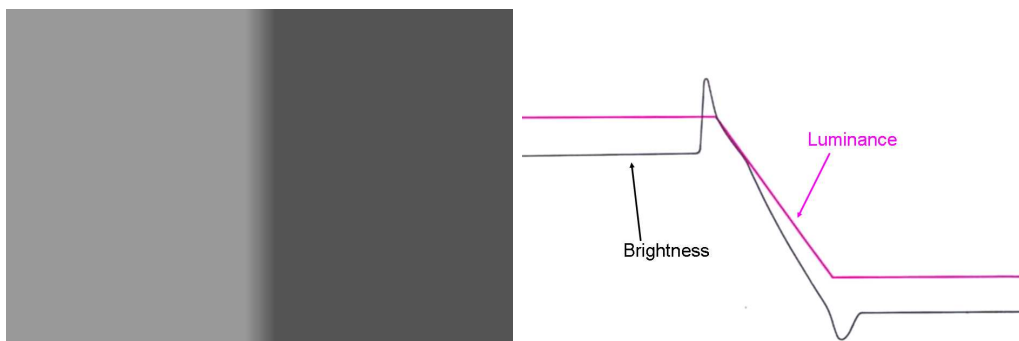


Figure 2.12: Left: An example of Mach bands. Right: Profiles of luminance and brightness of Mach bands. The actual profile in luminance is shown in the pink line while it is perceived as in the black line (brightness).

Similar to *Mach bands*, *Craik-O'Brien-Cornsweet illusion* is also a well-known border contrast effect [O'Brien 1959, Craik 1966]. If the parts of over- and under-shoots are given in an object of a uniform reflectance, the part with over-shoots

starts looking brighter, and the other side looks darker (see Figure 2.13). While *Mach bands* appear only in the areas which are close to the given gradient, *Craik-O'Brien-Cornsweet illusion* affects the perception of large areas. Purves et al. presented that this illusion still has a strong effect in 3D objects [Purves et al. 1999], and it was extended for 3D unsharp masking model by Ritschel et al. [Ritschel et al. 2008].

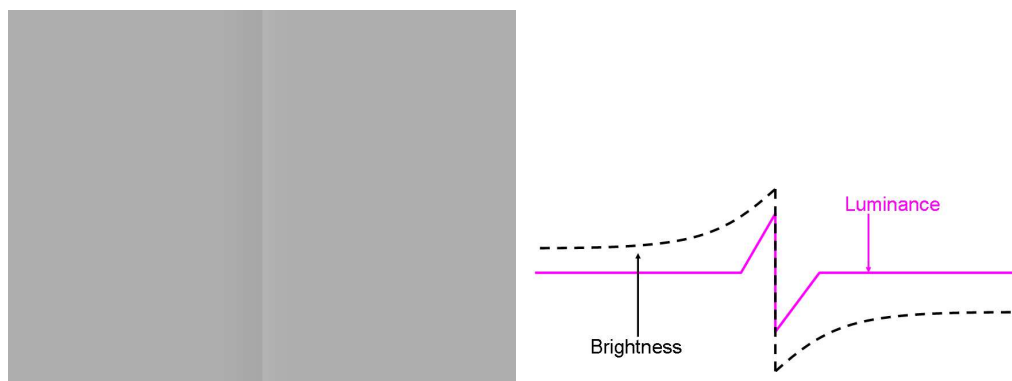


Figure 2.13: Left: An example of Craik-O'Brien-Cornsweet illusion. Right: Profiles of luminance and brightness of Craik-O'Brien-Cornsweet illusion. Pink line indicates the actual luminance profile while it is perceived as in the dashed black line.

2.5 Psychometric Scaling

Psychophysics, which is a part of experimental psychology, is a way to study the characteristics of HVS. It is the scientific study of the relation between physical stimulus and sensation which is observed by human subjects. Here, psychophysics can be a bridge between HVS and computer graphics because in computer graphics it is important to determine what factors of an image contribute to visual experience and to assess what method produces effective visual experience if several different methods are proposed for the same purpose.

Two important changes have recently occurred in psychophysics: development of the theory of signal detection and the refinements of methods for directly scaling sensory magnitude [Gescheider 1997]. These two improvements lead a wide applicability of psychophysics. As broadening the capability of psychophysics in research, there have been a number of experimental methods, background theories of statistics, and ways of analyzing or scaling data depending on the purpose of an

experiment. In this section, we briefly review the experimental methods and analyzes which are used in this dissertation. Refer to [Gescheider 1997, Engeldrum 2000, Coolican 2004] for the other theories and applications in psychophysics, and [Torgerson 1958] for scaling methods.

2.5.1 Two-Alternative Forced Choice (2AFC) for 1D Scaling

If a limited number of stimuli are given and one wants to know the scaling or ranking between those stimuli, *two-alternative forced choice (2AFC)* (sometimes called *pairwise comparison*) or *ranking method* are commonly used. Assume n stimuli are prepared. All possible combinations of n stimuli are presented to a subject, and he/she must answer a question which stimulus has stronger (or weaker for some cases) intensity for each pair of stimuli, for example, “which stimulus is brighter?”. In the end of the experiment with a large enough number of human subjects, a $n \times n$ square matrix M is constructed to determine the number of times which each stimulus was chosen. Each element $m_{i,j}$ at the i th row and j th column in the matrix M denotes the observed number of times which stimulus j was chosen when it was compared to stimulus i . The diagonal cells of M are left vacant. Next, the elements of the matrix M are normalized between 0 – 1. Then, based on the normalized matrix, a basic transformation matrix X is constructed such that the element $x_{i,j}$ is the unit normal deviate corresponding to each element of the normalized matrix and can be obtained by referring to a table of areas under the unit normal curve. The diagonal cells of X are filled with zeros. Finally, the elements of each column in the matrix X are summed up to obtain the scalings between given stimuli. Resulting scalings are given by distances between the stimuli.

This analysis is called Thurstone’s law of comparative judgment [Thurstone 1927, Thurstone 1967]. This is very simple method to achieve scalings between stimuli, but a big drawback is that the number of comparisons is given as $n(n - 1)/2$ which exponentially grows by increasing n . There have been presented several ways to reduce the number of comparison in 2AFC experiment (see [Torgerson 1958, Chapters 8 – 9] for the details). An alternative approach is to use *ranking method* which gives part or all of the stimuli to a subject instead of presenting one pair after another and then asks them to order the given stimuli. It can be also analyzed in the same way by using Thurstone’s law. 2AFC and Thurstone’s law are used in Chapter 6 for giving a scaling of perceived contrast with respect to given physical contrast.

2.5.2 Threshold Measurement

If one wants to measure detection or discrimination thresholds, probably the simplest approach is the *increment/decrement* method. In the experiment by increment/decrement method, a pair of the reference and target stimuli are presented to a human subject. The target stimulus is set either at the same intensity of the reference stimulus (Case 1) or at the level which is significantly different from the reference (Case2). Then, a subject is asked to start changing the intensity of the target stimulus until he/she starts seeing the difference (for Case1) or starts seeing the stimuli same (for Case 2).

The increment/decrement method is simple, however, its accuracy is sometimes doubtful. In the *Staircase* method, a pair of the reference and the target stimuli are presented as same as increment/decrement method. The target's intensity is increased whenever the different between the reference and target is not discriminated or decreased when there is no difference perceived.

For both increment/decrement method and staircase method, a subject is allowed to change the intensity of the target stimulus. In *Parameter Estimation by Sequential Testing (PEST)*, the intensity of the target stimulus is changed by the experimental program [[Taylor and Creelman 1967](#)]. Again, a pair of the reference and target stimuli are presented to a subject. The target stimulus is set significantly different from the reference. At each step, a subject must answer a question "do you see difference?". If the answer is yes, the intensity of the target stimulus is jumped close the the reference (commonly the width of the first jump is same as the difference between the reference and the intensity where the target started). An experiment is basically conducted by repeating these steps. Every time a subject answers in the different way as the previous time, the direction of changing the target's intensity is inversed and the width of a jump is reduced to its half size, while the target's intensity is changed to the same direction with the same width of a jump as far as being answered yes. One trial can be finished if the response of a subject start being constant enough. *QUEST*, a refinement of PEST, has also been presented [[Watson and Pelli 1983](#)]. PEST is employed in Chapter 6 for measuring contrast detection thresholds.

2.5.3 Rating Experiments and Multivariate Statistics

If one wants to know some ratings by human observers for each stimulus, it is necessary to ask them to score each stimulus for a given property. This approach is, of course, much harder than just asking yes/no questions, but it provides much

more possibilities in its results. There are a variety of ways to analyze data in rating experiments depending on how many variables exist and what the purpose of an experiment is. There are briefly four categories of the experimental purposes such that one wants to know 1) the degree of relationship among variables, 2) significance of group differences, 3) prediction of group membership, or 4) structure of given variables. For each group of the purposes, there are multiple possibilities to analyze data. It depends on many factors which are, for example, the number of *dependent variables (DVs)*, that are measured and depend on subjects' response and the number of *independent variables (IVs)*, that are manipulated or controlled.

For 1) the degree of relationship among variables, we can apply bivariate r , multiple R , hierarchical multiple R , canonical R , or multi-way frequency analysis. For 2) significance of group differences, there exist t -test, analysis of variance (ANOVA), analysis of covariance (ANCOVA), multivariate analysis of variance (MANOVA), Hotelling's T square, multivariate analysis of covariance (MANCOVA), and profile analysis. For 3) prediction of group membership, we may choose between one-way, hierarchical one-way, factorial, and hierarchical factorial discriminant functions. For 4) structure of given variables, principle component analysis (PCA) or factor analysis are recommended. Refer to [Tabachnick 1989] how to choose among multivariate statistical techniques.

In this dissertation, we use statistical analyzes to know significance of group differences (Purpose 2) in Chapters 4, 5, and 7. For this purpose, one important concept is *null hypothesis* which assumes that all of the population means of the given data are equal as

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (2.7)$$

where μ_i is each population mean and k is the number of populations. A null hypothesis is commonly used to obtain the reverse of what an experiment is actually believed.

To examine a null hypothesis, a *significance* test is used, for example, the simplest way is t -test for the case with one DV and one IV. If there is one DV with either one or multiple IVs, ANOVA is applied to a set of data in one-way or factorial way. If there are multiple DVs with either one or multiple IVs, MANOVA can be used. A null hypothesis is tested by comparing two estimates of variance with the given population mean and the population size. In, for example, ANOVA and MANOVA, *degrees of freedom* are calculated, statistic F value is manipulated, then finally a significance level is shown in a p -value which is the probability to accept a null hypothesis. It is usually concerned the difference between IVs over a given DV is significant if p is below 0.05, i.e., a null hypothesis is rejected with a probability over 95%. A significance level of $p < 0.01$ is sometimes used for very strict case of analyzes. Refer to [Tabachnick 1989] for more details of significance

tests and the details of the other categories of multivariate statistics. In addition to examining significances, it is also possible to construct a model over given IVs and DVs by using *multidimensional scaling* [[Borg and Groenen 1997](#)].

Chapter 3

High Dynamic Range Imaging (HDRI)

As shown in Figure 1.1 in Chapter 1, neither conventional display devices nor conventional cameras are capable of covering the luminance dynamic range that the human eyes perceive. If we take a *high dynamic range (HDR)* shot by a *low dynamic range (LDR)* camera, the areas which are out of the dynamic range of the camera are over- or under-saturated as seen in Figure 3.1. We can use short exposure to capture details in bright parts of a scene, but, on the other hand, the dark parts are completely invisible with a short exposure, or vice versa for long exposures.

As increasing the need of HDR techniques, there exist three categories of the solutions to capture HDR scenes: 1) shooting an HDR scene by using highest-end imaging sensors, 2) employing multi-exposure techniques with a series of LDR images, or 3) using 3D rendering programs. In Solution 1, there have been introduced several types of new imaging sensors, however, their high costs now prevent them from the wide-spread use. Solution 2 is probably the easiest and the most practical way to produce an HDR image of the natural environments nowadays, and there have been introduced several methods of multi-exposure techniques. Solution 3 provides more practical algorithms and applications to make HDR scenes by using 3D renderers. However, some of them require us to have enough experience and even the sense of art.

Even if an HDR image is created and stored in some HDR formats such as floating point TIFF, LogLov, OpenEXR, Radiance RGBE, or PFS, it is obvious that LDR display devices cannot display HDR images on them as HDR images. To conquer

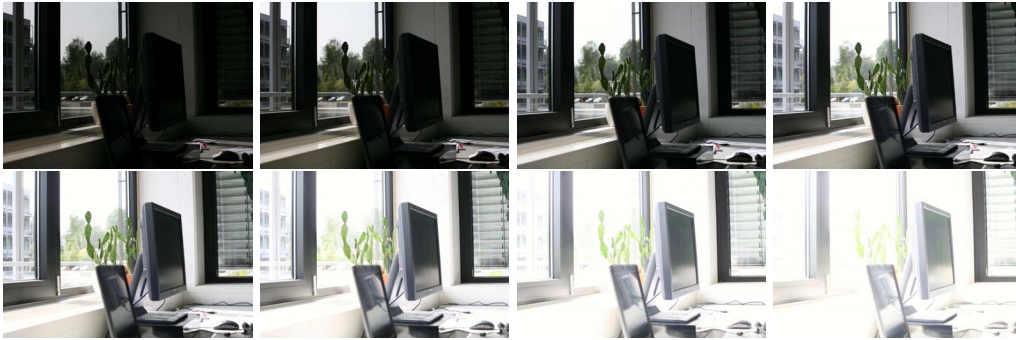


Figure 3.1: A series of low dynamic range (LDR) images with different exposures. Because the scene contains much broader luminance range than the dynamic range which the camera can take in one shot, the areas out of the dynamic range of the camera are either over- or under-saturated in each image. Images taken by the Canon EF 50mm lens mounted on the Canon EOS 5D.

this problem, a number of *Tone Mapping Operators (TMOs)* have been introduced for compressing the dynamic range of an HDR image to fit that of the LDR display devices. Using TMOs is a software approach to deal with HDR images on LDR displays. From the point of hardware view, there exist HDR displays by using projector- and LED-based dual-modulation technologies introduced by Seetzen et al. [Seetzen et al. 2004]. In the following sections, we review HDR images acquisition, tone mapping operators (TMOs), and dual-modulation HDR display technologies.

3.1 HDR Image Acquisition

3.1.1 HDR Imaging Sensors

In this section, we discuss two prominent examples of new imaging sensors which can take an HDR scene in one shot: *Digital Pixel Sensors (DPS)* and *high dynamic range CMOS (HDRC)*. DPS mount more transistors to pixels so that each pixel has its own analog-to-digital converter and its own logic circuits. Since DPS are capable of 10,000 frames per second, we can run them at higher frame rate than the actual image generation. After taking pictures with several exposures for each capture, they are combined into an HDR image at the lowest levels in each pixel signal. Another new imaging sensor is *high dynamic range CMOS (HDRC)*, a CMOS-based sensor with per-pixel readout logic. HDRC can cap-

ture each frame with up to four differently exposes captures to create an HDR image [Bloch 2007]. Simply speaking, both approaches depend on how many transistors can be mounted on a chip. The cameras with such new imaging sensors are already available in consumers' market, however, a big drawback of those highest-end sensors is their cost.

3.1.2 Multi-Exposure Techniques

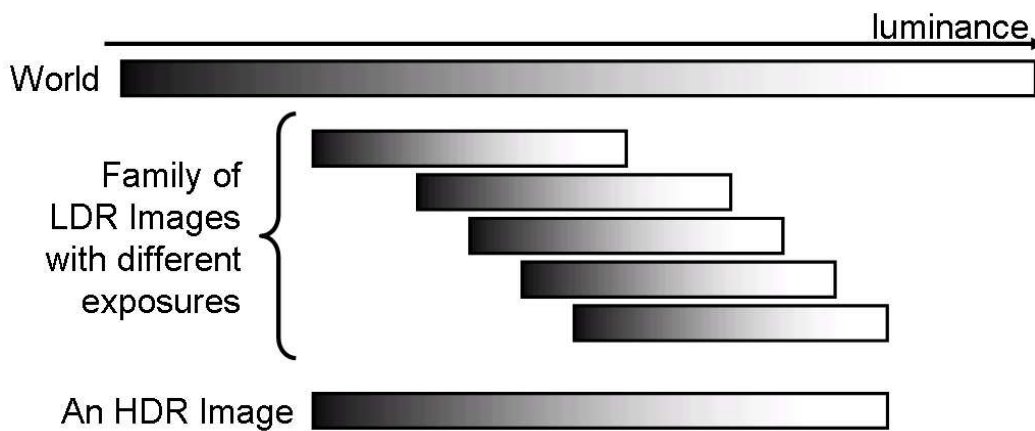


Figure 3.2: A principle idea of multi-exposure techniques. A series of LDR images are taken with different exposures to cover broader dynamic range than that of an LDR camera, then they are combined into an HDR image.

A basic idea of *multi-exposure techniques* is that an HDR image is reconstructed from a series of LDR images with different exposures taken by the same camera from the same position (refer to Figure 3.2). Several multi-exposure algorithms have been introduced such as [Mann and Picard 1995, Debevec and Malik 1997, Robertson et al. 1999, Nayar and Mitsunaga 2000, Ward 2003]. Each algorithm has a different strategy for recovering the response curve of a camera which is needed to express pixel values for each LDR image in the same intensity space. The method of Mann and Picard uses a relatively dark pixel of a taken image as a reference, then a nonlinear response curve is recovered with respect to the selected dark pixel [Mann and Picard 1995]. Debevec and Malik employed a physical property of imaging systems for recovering camera response curve and pixel weighting function [Debevec and Malik 1997]. Robertson et al.'s method [Robertson et al. 1999] is similar to Debevec and Malik method, but it takes all pixels of an image while Debevec and Malik method deals with randomly selected pixels. Robertson et al.'s method also produces a weighting function based on the

fact that longer exposures tend to produce a better signal-to-noise ratio. Nayar and Mitsunaga's method uses a flexible parametric model and roughly estimates the ratios of exposures instead of requiring precise estimates of exposure times [Nayar and Mitsunaga 2000]. For all methods, once the response curve of a camera and its weighting function are reconstructed, the recovered functions are applied to combine a series of LDR images taken with different amounts of exposure. An example of the response curve and its weighting function is shown in Figure 3.4 which is extracted from a series of LDR images in Figure 3.3 using Robertson et al.'s method.

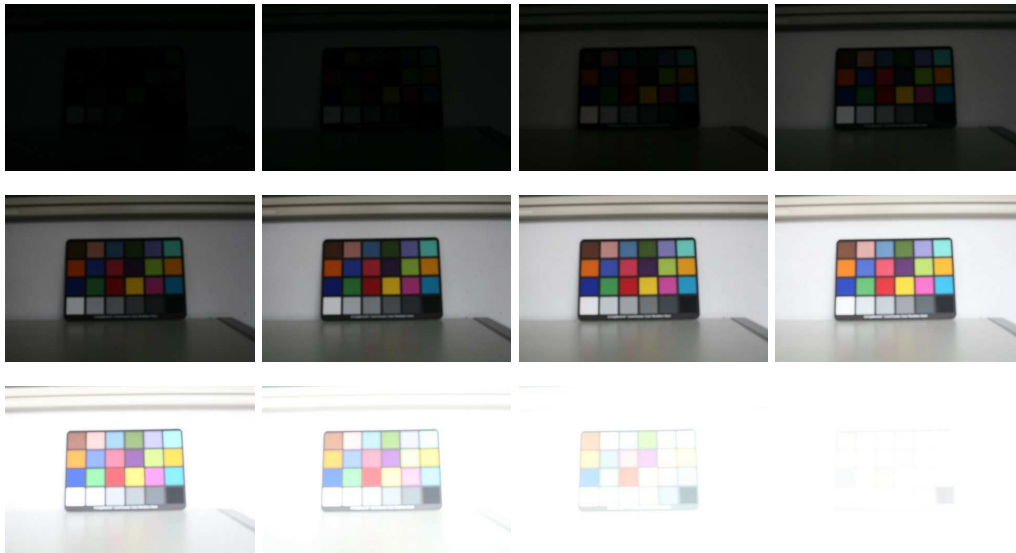


Figure 3.3: A series of LDR images with different exposures. They were used to extract the response curve and weighting function of the camera (see Figure 3.4). Images taken by the Canon EF 50mm lens mounted on the Canon EOS 5D.

The four algorithms reviewed above have a strict requirement that all LDR images must be taken from the exactly same position and orientation (e.g., using a tripod), and, in addition, no single movement of objects in an image is allowed. Ward presented a method to create an HDR image by combining a series of LDR images taken by a hand-held camera [Ward 2003]. His method arbitrarily selects one of the given LDR images as a reference, then it outputs a set of integer offsets for the rest of images. After calculating offsets, an HDR image is created by using a known recovering function for a camera response curve.

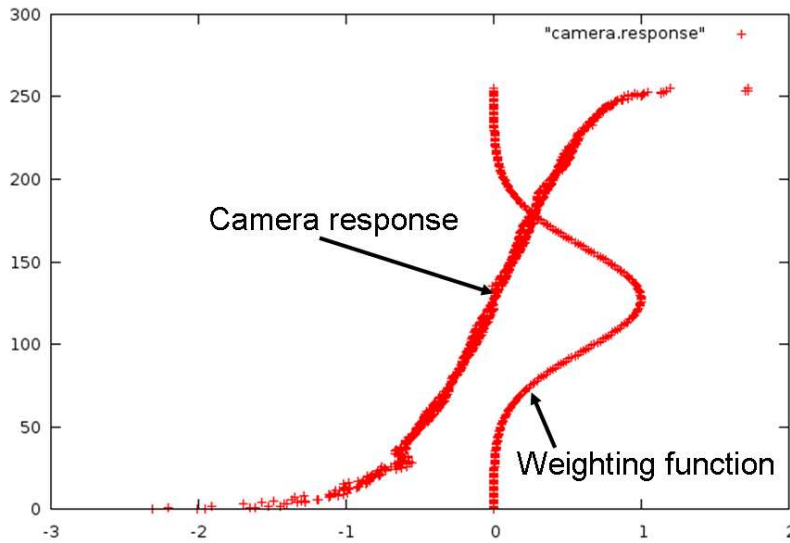


Figure 3.4: An example of weighting function and camera response curve extracted from Figure 3.3 by Robertson et al.’s method for the Canon EF 50mm lens mounted on the Canon EOS 5D.

3.1.3 3D Renderers for HDR Images

There are a number of 3D rendering methods to produce HDR scenes in *classical photorealistic rendering*, *physically based rendering*, and *image-based lighting*. The classical photorealistic rendering approaches such as *scanline* or *raytracing* use the lighting model of direct illumination. We can place fake objects to make the final result as HDR, for example, negative lights, shadow mapping, textures, and ramps. These approaches are the simplest and fastest methods to render HDR scenes, however, they require the users to have experience and artistic skills.

On the other hand, *global illumination* simulates the behavior of lights strictly according to the physics laws. In this category of the physically based rendering approaches, one may find *radiosity* and *Monte Carlo* method as its applications. Physically based rendering methods can produce more visually accurate results than the classical approaches, however, they still require some talent and experience for the users.

Using another option, an image-based lighting method such as [Debevec 1998], is more intuitive and has a huge potential to make use of HDR imagery. Image-based lighting firstly records environmental lighting characteristics in an image of a dome or a sphere, then the recorded characteristics of the surroundings are taken into account with the rendered scene by using global illumination technique. The

details and usage of these applications are summarized in [Bloch 2007] [Reinhard et al. 2007, Chapter 13].

3.2 Tone Mapping Operators (TMOs)

3.2.1 Overview

There were several early research papers about photographic tone reproduction. Jones and Condit researched the relationship of brightness and camera exposure with a number of exterior pictures and proposed prediction methods of a meter constant of an image [Jones and Condit 1941]. Bartleson and Breneman showed theoretical predictions of the optimum tone reproduction functions for prints and transparencies and proved it by experimental data [Bartleson and Breneman 1967]. Additionally, Miller et al. proposed an idea of tone mapping in the field of lighting engineering [Miller and Hoffman 1984].

In 1993, the idea of tone reproduction was firstly introduced by Tumblin and Rushmeier into the computer graphics community [Tumblin and Rushmeier 1993]. The main goal of tone reproduction is to adjust the dynamic range of an image to the range that can be displayed on physical devices when the luminance range of the images does not fit that of the physical devices. The dynamic range must be compressed for very bright images, and it must be expanded for very dark images. However, not only should a *Tone Mapping Operator (TMO)* adjust the dynamic range, but it should also preserve details and global contrast of an image. It is a hard and contradicting requirement. Usually, if overall contrast is simply preserved, details of an image may be lost or vice versa. How to preserve them at the same time is always a difficult problem to solve for TMOs.

There have been presented a number of TMOs which are reviewed in the following sections in terms of different scopes of TMO domains (refer to [Devlin et al. 2002] [Reinhard et al. 2007, Chapters 6 – 8] for the details), and Figure 3.5 shows an example of tone mapped images by different TMOs. Some of them are used in the perceptual evaluation of TMOs with respect to their corresponding real-world scenes in Chapter 4.

3.2.2 Luminance Domain TMOs

The simplest way to adjust the dynamic range of an HDR image to that of the display devices is *linear* TMO which scales luminance range of an HDR image

linearly. Linear TMO is also employed in logarithmic domain (*logarithmic linear* TMO). Without gamma correction, most of conventional displays follow approximately log-linear response.

The first TMO in computer graphics field, Tumblin and Rushmeier TMO, employs the characteristics of HVS for transforming the luminance of the real-world to that of a display device in order to preserve brightness [Tumblin and Rushmeier 1993]. This method works only on grayscale images, then it is revised to a new TMO which incorporates a linear scaling factor based on adaptation luminance [Tumblin et al. 1999].

Ward Larson et al. introduced *histogram adjustment* TMO [Ward Larson et al. 1997] based on [Ward 1994, Ferwerda et al. 1996]. This TMO leads to a monotonic tone reconstruction curve, and then it applies the curve globally to all pixels in an image.

Pattanaik et al. extended an early work [Tumblin and Rushmeier 1993] and presented a TMO which employs the idea of temporal visual adaptation of HVS [Pattanaik et al. 2000]. This method is constructed by two models: the visual adaptation model and the visual appearance model to take into account threshold visibility, color appearance, visual acuity, and sensitivity change over time for average luminance.

Photographic TMO is also a luminance domain operator [Reinhard et al. 2002]. This method is inspired by photographic film development and its printing process. The luminance of an HDR image is firstly adjusted by using the same function to all pixels, then photographic “dodging and burning” technique is applied to enhance the quality of an image. This method is extended to [Reinhard 2003] to be operated automatically.

Ashikhmin introduced a multi-pass approaching TMO to deal with two characteristics of HVS: signaling absolute brightness and local contrast [Ashikhmin 2002]. This TMO firstly calculates the local adaptation luminance, then applies the capacity function which is based on the linearly approximated TVI function. The final tone mapped image is produced by multiplying a detailed image given by the ratio of pixel luminance to the corresponding local adaptation.

Adaptive logarithmic mapping is based on logarithmic compression of luminance [Drago et al. 2003]. A family of different logarithmic functions are applied to preserve details of an image, for example, \log_{10} is applied for the high luminance levels. In the end, Perlin and Hoffert’s bias power function is used to interpolate between the images modified by different logarithmic bases [Perlin and Hoffert 1989].

Lightness perception TMO was presented by Krawczyk et al. [Krawczyk et al. 2005]. This method de-composites an HDR image into several areas of consistent luminance and local lightness values. It is inspired by *an anchoring theory* [Gilchrist et al. 1999] and emphasizes the importance of luminance by using it.

Photoreceptor-based TMO provides a sigmoidal compression for the dynamic range of an image [Reinhard and Devlin 2005]. This method is based on the behavior of photoreceptors and employs an idea of “semi-saturation” by linearly interpolating the geometric average and each pixel’s luminance by applying user-set parameters.

3.2.3 Contrast Domain TMOs

Gradient domain TMO identifies gradients at various scales in order to reduce the magnitudes of large luminance gradients while preserving small changes for high frequency details of an image [Fattal et al. 2002]. This TMO compresses low frequency details strongly but keeps high frequency details uncompressed. They also employ the gradient attenuation function, which is applied for each band of frequencies, in order to avoid halo artifacts.

A fast bilateral filtering considers two different spatial frequency layers of an image: a base layer and a detail layer [Durand and Dorsey 2002]. The base layer preserves high contrast edges and removes high spatial frequency details of lower contrast. The detail layer stores the difference between the original image and the base layer. The final tone mapped image is produced by summing up these two layers. This method is extended to *Trilateral filtering* which applies bilateral filtering twice in different domains [Choudhury and Tumblin 2003].

Mantiuk et al. presented two TMOs in their perceptual framework for contrast processing [Mantiuk et al. 2006]. Their *contrast mapping* method is similar to that of gradient domain method [Fattal et al. 2002] but is entirely based on the perceptual characteristic of human eyes. This method succeeds to avoid low-frequency artifacts in an image. Additionally, their *contrast equalization* method equalizes a histogram of contrast responses in order to produce a sharp and visually appealing tone-mapped image.

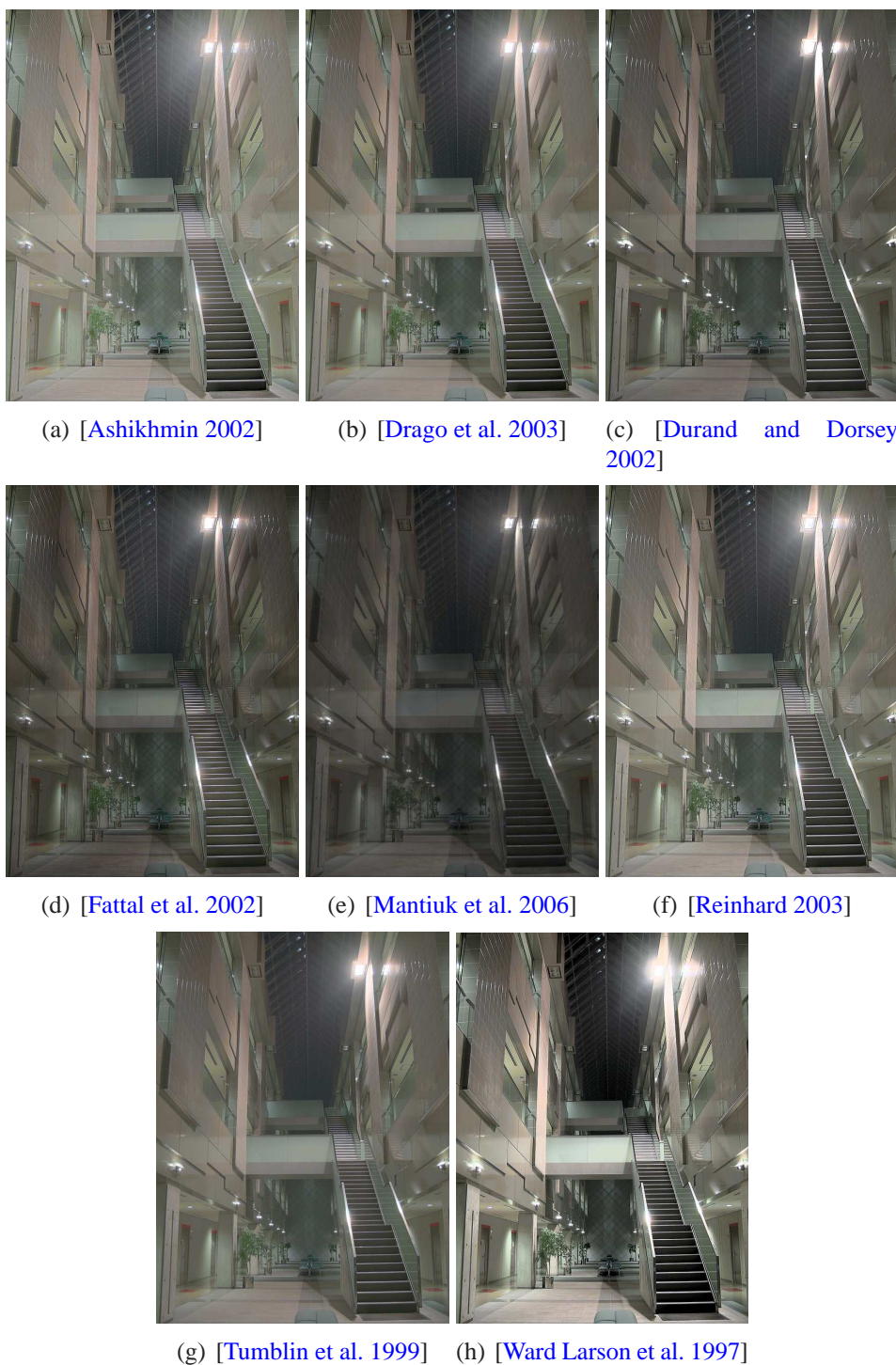


Figure 3.5: An example of tone mapped images by different TMOs.

3.3 Dual-Modulation HDR Displays

As a hardware approach to deal with HDR images, Seetzen et al. presented two techniques of HDR displays: *projector-based* and *LED-based* HDR displays [Seetzen et al. 2004]. Their basic idea is to employ a second light modulator which, through multiplicative effect with the originally used modulator, enables to achieve strong improvement of the admissible contrast between the darker and brighter pixels. This is, in particular, apparent for black levels. Even very strong light source used as backlight can be attenuated for almost perfect black levels by such dual modulators (see Figure 3.6).

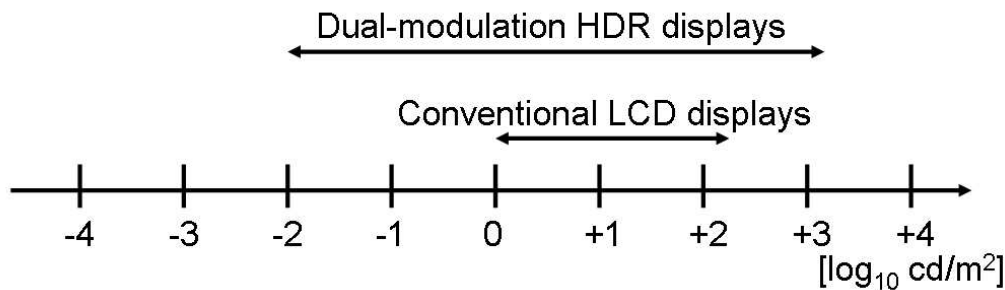


Figure 3.6: Dynamic range of conventional LDR displays and dual-modulation HDR displays. Note that HDR displays not only increase their maximum possible luminance level but also produce extremely dark states.

The *projector-based* HDR display contains a Digital Light Projector (DLP) with a Digital Mirror Device, an LCD panel, and the optics to combine DLP and LCD (see Figure 3.7). Its dynamic range in luminance depends on an LCD panel. Their prototype which mounts 15" XGA color LCD (Sharp LQ150X1DG0) has luminance range between 1 and 2700 cd/m^2 . Because it is difficult to keep the projector and the LCD panel always aligned perfectly at the pixel level, the projector intentionally introduce a blurred image which is then compensated in the LCD sandwich. The pixel values of the LCD are chosen to compensate for those effects. The *projector-based* HDR display can cover much broader dynamic range in luminance compared to today's conventional display, however, it has several drawbacks such as optical length requirement, its high power consumption, cost, thermal management, and video bandwidth.

Another technique, the *LED-based* HDR display, employs an array of white LEDs as a backlight instead of a projector. Each of the 760 LEDs for 18" LCD panel can be individually controlled with 1024 addressable steps. Its minimum possible

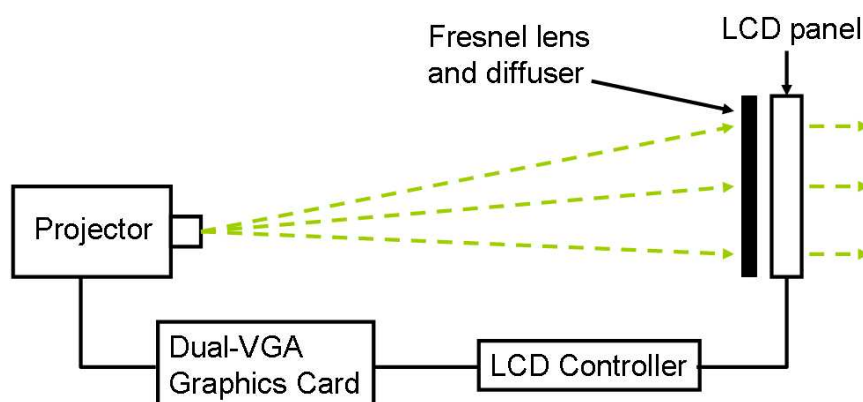


Figure 3.7: Scheme of a projector-based HDR display.

luminance value is below 0.03 cd/m^2 and it is $8,500 \text{ cd/m}^2$ at the maximum. The rendering algorithm on the *LED-based* HDR display is basically similar to that of the *projector-based* display, but it employs a better supported *point spread function (PSF)* for LEDs which are arranged on a hexagonal grid instead of a rectangular grid (refer to [Seetzen et al. 2004] for their details). This prototype with 18" LCD panel is now extended to the 37" HDTV resolution (see Figure 3.9) whose luminance range is between 0.01 and $3,000 \text{ cd/m}^2$ (confirmed by our measurement with MINOLTA LS-100 light meter¹).

¹<http://www.konicaminolta.com/instruments/products/light/luminance-meter/ls100-ls110/index.html>

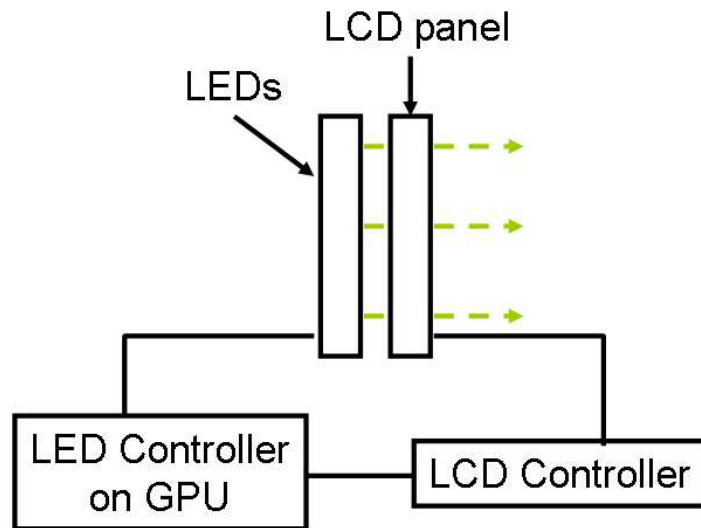


Figure 3.8: Scheme of an LED-based HDR display.



Figure 3.9: A 37" LED-based HDR display.

Chapter 4

Testing Tone Mapping Operators with Human-Perceived Reality

4.1 Introduction

As the need of high dynamic range (HDR) images has greatly increased, how to produce and display HDR images has been one of the important discussions. To represent HDR images on low dynamic range (LDR) display devices, a number of successful tone mapping operators (TMOs) have been presented (refer to Section 3.2). They are useful not only for HDR photography but also for lighting simulation in realistic rendering and global illumination techniques, which provide real-world luminance ranges.

Because a variety of tone mapping operators have been proposed, only a systematic perceptual evaluation can reveal the strengths and weaknesses of the wide range of approaches presented in recent years. We conducted a psychophysical experiment of a direct comparison between the appearance of real-world scenes and tone mapped images of these scenes. The primary interest of this experiment is to investigate the differences in perception of tone mapped images when they are directly compared with their corresponding real-world views and indirectly compared with each other. To our knowledge, this work was the first direct comparison of tone mapped images with corresponding real-world scenes at the time it was performed.

In this chapter, Section 4.2 reviews an overview of previous work, Section 4.3 describes our psychophysical experiment, Sections 4.4 and 4.5 show its results

and discuss about them, and then Section 4.6 summarizes this project. This work is a follow-up extension of Yoshida's Master thesis [Yoshida 2004]. Our main contribution is to demonstrate that qualitative differences between TMOs have a systematic effects on the perception by human's eye which can be a basis for selecting an appropriate TMO for a given purpose and also for an assessment of further approaches in TMOs [Yoshida et al. 2005, Yoshida et al. 2007a].

4.2 Previous Work

Image comparison techniques can be roughly classified into two major categories: *subjective* and *objective* methods. Subjective methods obtain data from human observers and the data are usually analyzed by statistical techniques [Newsham et al. 2002, Nijenhuis et al. 1997, Watson et al. 2001] while objective methods are based on theoretical models. The work with the similar goals to ours about perceptual evaluation of tone reproductions has been recently published in [Kuang et al. 2004, Kuang et al. 2005, Ledda et al. 2005].

Kuang et al. conducted two paired comparison experiments of eight TMOs [Kuang et al. 2004]. Tone mapped images were displayed on an LDR monitor without reference for observing overall rendering performance and grayscale tone mapping performance respectively. They also conducted a similar experiment with six TMOs (two poorly scored ones in [Kuang et al. 2004] were eliminated) [Kuang et al. 2005]. They asked subjects to examine overall preference and several image attributes (highlight details, shadow details, overall contrast, sharpness, colorfulness, and artifacts) over tone mapped color images displayed on an LCD monitor. As in [Kuang et al. 2004], they did not provide reference for subjects.

Ledda et al. also conducted psychophysical experiments of six TMOs. They asked subjects to make paired comparison of tone mapped images displayed on an LDR monitor with an HDR image as their reference [Ledda et al. 2005] on a Brightside HDR display [Seetzen et al. 2004]. In their experiment, the subjects actually never saw the real-world scenes measured by an HDR display. They measured overall similarity and detail reproduction of tone mapped images. First, they studied overall similarity of TMOs and clarified that their selected TMOs can be ranked. They asked subjects to compare a pair of tone mapped images and choose the one which appeared closer to the reference displayed on an HDR monitor. Following the first experiment, detail reproduction of TMOs in both bright and dark regions of an image were studied in the second experiment. It showed a different result from that of Kuang et al. [Kuang et al. 2004] for *fast bilateral filtering* [Durand and Dorsey 2002]. As Ledda et al. wrote, *bilateral filtering* generates images with

higher contrast and more detail visibility than in the reference images. Therefore, it had poor scores in the experiment with reference on an HDR monitor [Ledda et al. 2005] while it performed quite well without references [Kuang et al. 2004].

This observation is also confirmed by Yoshida et al. [Yoshida et al. 2006]. The experiments over the same HDR monitor as that of Ledda et al. were conducted with and without references of the corresponding real-world view. The result shows that subjects behaved differently with and without reference. They enhanced contrast proportionally to the dynamic range of an HDR display, even more than that of an original image, if they had no reference. However, if they had reference, they adjusted contrast almost same as that of an original image and kept it approximately on the same level even for different dynamic ranges of a display [Yoshida et al. 2006].

In parallel to our experiments, Ashikhmin and Goral [Ashikhmin and Goral 2006] and Čadík et al. [Čadík et al. 2006] conducted subjective experiments to evaluate TMOs with respect to their corresponding real-world views. They are discussed in details and compared to our results in Section 4.5.

4.3 Perceptual Evaluation

Our experiment is different from those which are reviewed in the previous section in several ways. First, we provide corresponding real-world views of a tone mapped image for subjects as reference. HDR monitors can produce much wider dynamic range of luminance than that of today's common display devices, but they still have their own limitations concerning the representation of high contrast in high frequency areas of an HDR image and their luminance range can not match that of the real-world [Seetzen et al. 2004]. Additionally, as shown in [Yoshida et al. 2006], subjects may behave in different ways for observing their preference or archiving the fidelity of an image and an HDR monitor can even provide more contrast than an original image. Because we want to measure how close tone mapped images are to the real-world view, we do not select an HDR monitor as reference.

Second, we select the ranking method using a slider. The paired comparison method is very simple and powerful for observing interval scales of given algorithms (TMOs in this case) along a given dimension. Thurstone's Law of comparative judgement [Thurstone 1927] is the most common used analysis for paired comparison. However, the paired comparison analyzed by Thurstone's Law has two problems; A paired comparison experiment needs $n(n - 1)/2$ experimen-

tal trials for n stimuli. The total number of trials increase very rapidly as the number of stimuli. Additionally, Thurstone's Law relies on the assumption of uni-dimensional scaling. It is quite useful if we want to compare the performance of one specific attribute, but in our experiment, we want to have more insight in multi-dimensional scaling than in uni-dimension. Therefore, we select the rating-scale method instead of the paired comparison. The details for those methods and analysis can be found in a book by Bartleson and Grum [[Bartleson and Grum 1984](#)].

4.3.1 HDR Image Acquisition

Figure 4.1 shows the scenes for our experiment. To acquire HDR images for our perceptual evaluation, we used a multi-exposure technique (refer to Section 3.1 about HDR image acquisition). Kodak Professional DCS560 digital camera and Canon EF 24mm and 14mm were used to shoot 15 LDR images with different shutter speeds ranging from 1/2,000 to 8.0 seconds. Both lenses have big enough field of view to cover the view of human eyes. Because those images were saved in a raw format of Kodak, they were converted to 36-bit TIFF format by using the program `raw2image` included in the Kodak DCS Photo Desk Package. To recover the camera response curve function, we employed Robertson et al.'s method [[Robertson et al. 1999](#)]. The HDR images were constructed with the recovered response curve and saved in the Radiance RGBE format [[Ward 1991](#)].

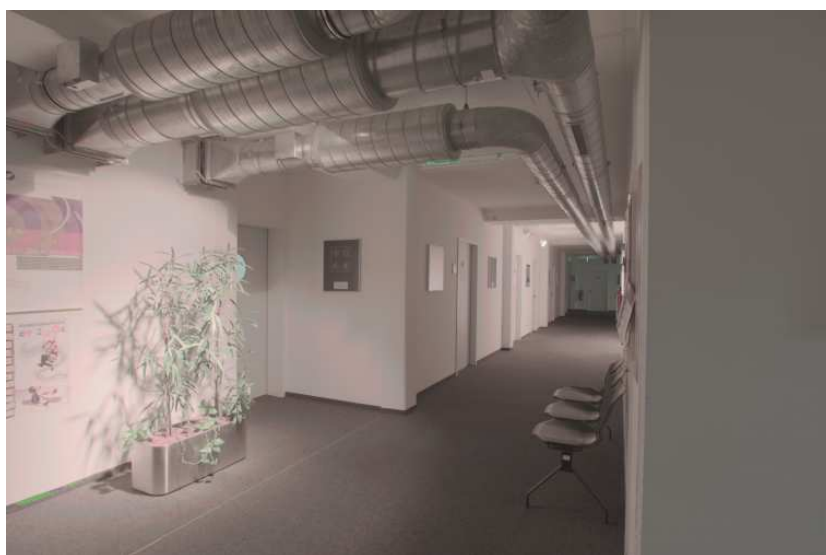
4.3.2 Tone Mapping Operators

Apart from the domain categorization of TMOs (see Section 3.2), there is a classical way to categorize TMOs either if it applies the same transformation onto all pixels (*global* TMO) or it adapts its scales to different areas of an image (*local* TMO). By this global/local categorization, it is obvious that global TMOs are much simpler and more easily implemented than local TMOs because they need to prepare only one function to apply for every pixel, however, global TMOs tend to lose the visibilities of details. On the other hand, local TMOs tend to lose an impression of global contrast compared to global TMOs. Additionally, one well-known problem which easily appears with local TMOs is halo artifacts (inverse gradient) which can manifest as a dark aura around a very bright region.

In this chapter, we mainly use the global/local grouping. Seven commonly used TMOs are chosen for our perceptual evaluation as shown in the following list with labels of "global" or "local" by this categorization:



(a) Scene 1. 13,630 and 0.021 cd/m^2 for the maximum and minimum luminances respectively.



(b) Scene 2. 506 and 0.019 cd/m^2 for the maximum and minimum luminances respectively.

Figure 4.1: View of the scenes for our experiment. Both are tone mapped by [Drago et al. 2003] for their presentation.

- Log-linear TMO (global)
- Histogram adjustment TMO (global) [Ward Larson et al. 1997]
- Pattanaik et al. (global) [Pattanaik et al. 2000]
- Ashikhmin (local) [Ashikhmin 2002]
- Fast bilateral filtering (local) [Durand and Dorsey 2002]
- Photographic TMO (local) [Reinhard et al. 2002]
- Adaptive logarithmic mapping (global) [Drago et al. 2003].

Refer to Section 3.2 for the characteristics of TMOs used in our experiment.

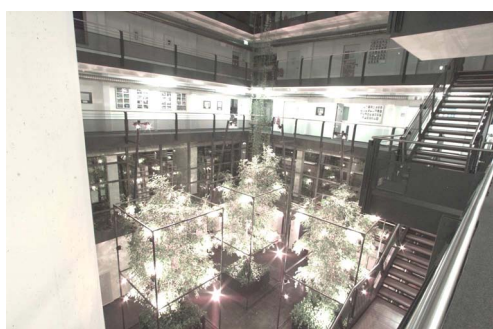
4.3.3 Experimental Procedure

Prior to the main experiment, we conducted a pilot study with experienced subjects to fine tune the parameter settings for each tone mapping operator. We asked subjects to choose the best image for each tone reproduction method from a selection of multiple tone mapped images. Additional post-processing, such as gamma correction, was performed according to the suggestions in each respective paper of the TMOs. All of the images used in our experiment are shown in Figures 4.2 – 4.3. All of the HDR images and our implementations of the TMOs used in the experiment are available at our website¹.

The main part of the experiment was performed with the participation of 14 subjects. They were graduate students and researchers of the Computer Graphics group in the Max-Planck-Institut für Informatik. Two of them are female, and the rest are male. The range of their age is 24 – 34, and all of them were naïve for the goal of the experiment and TMOs. Additionally, every participant reported normal or corrected to normal vision.

Each subject was asked to stay in the same position as each of the HDR images had been taken (Figure 4.1) and view seven images one after another for each of the two scenes. An sRGB-calibrated monitor (DELL UltraSharp 1800FP) displaying images of resolution $1,280 \times 1,024$ at 60.0 Hz was used. The subjects compared each image with its corresponding real-world view and gave ratings for image appearance and realism. Image appearance attributes judged in our experiment are overall brightness, contrast, detail reproductions in dark regions, and detail reproductions in bright regions. The subjects rated how well each of those attributes was reproduced in tone mapped images with respect to the real-world

¹http://www.mpi-sb.mpg.de/resources/tmo_and_real/



(a) Linear



(b) Bilateral filtering



(c) Pattanaik



(d) Ashikhmin



(e) Ward

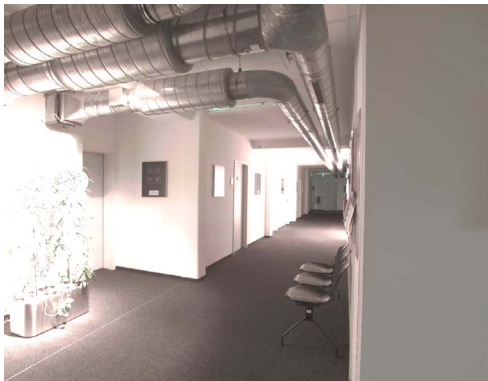


(f) Reinhard

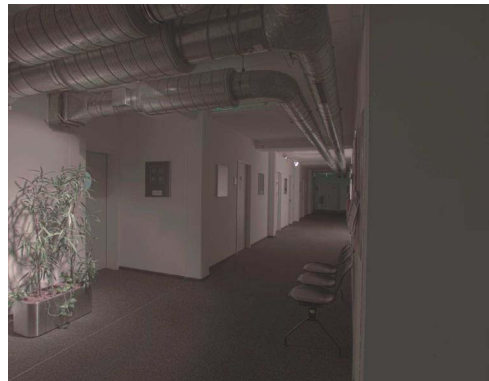


(g) Drago

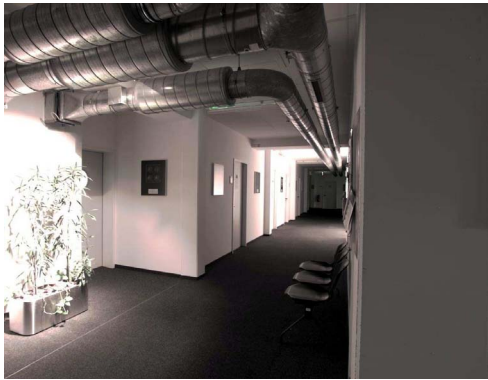
Figure 4.2: The tone mapped images for Scene 1.



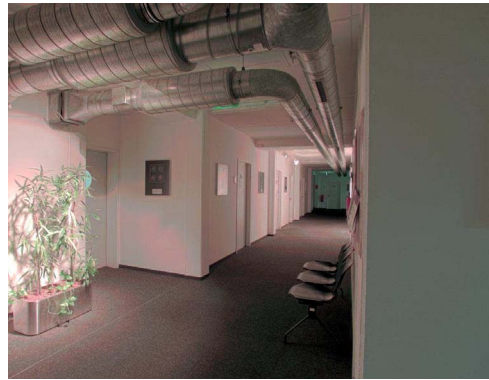
(a) Linear



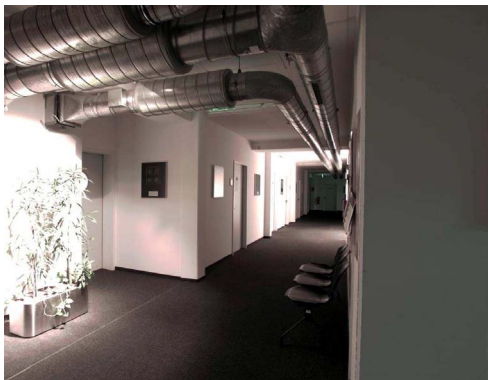
(b) Bilateral filtering



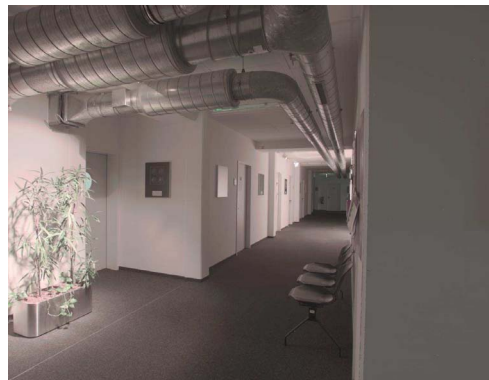
(c) Pattanaik



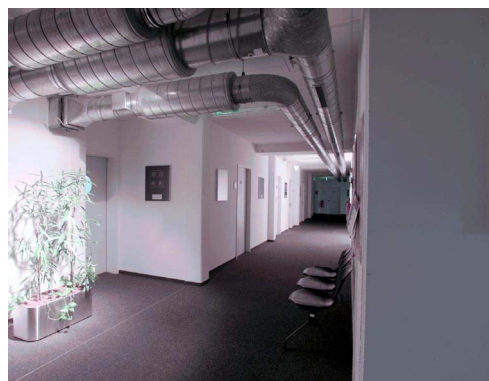
(d) Ashikhmin



(e) Ward



(f) Reinhard



(g) Drago

Figure 4.3: The tone mapped images for Scene 2.

view. Additionally, the subjects were asked to rate image naturalness in terms of reproducing the overall appearance of the real-world views. All of the ratings were done by moving scroll bars. The subjects were allowed to move back and forth among images for one scene (see Figure 4.4 for screenshots of our experiment and its user interface part). The whole procedure for one participant took approximately 20 to 30 minutes.

4.4 Results

Our experiment is a seven (TMOs) \times two (scenes) within-subjects design (see Tabachnick [Tabachnick 1989] for the details of the design types in multivariate analysis). This experiment has two *independent variables* (IVs): the TMOs and the scenes and five *dependent variables* (DVs): overall brightness, overall contrast, detail reproductions in dark regions, detail reproductions in bright regions, and naturalness (see Section 2.5 about DVs and IVs). Our primary interest is whether the images produced by different TMOs are perceived differently when they are compared to their corresponding real-world views. For analyzing a set of data, the Statistics Toolbox of MATLAB was used [MathWorks, Inc.].

As preliminary data processing, all obtained scores were normalized over each of the attributes and each of the subjects in order to scale the standard deviations 1.0 as $x_i \rightarrow \frac{x_i - \mu_x}{\sigma_x}$ where x_i is a score and μ_x and σ_x are respectively the mean and the standard deviation over an attribute of a subject.

4.4.1 Main Effects

Because two scenes were used in the experiment, we examined how much influence comes from the difference of two scenes. The *main effect* of the scenes was first tested by using *analysis of variance* (ANOVA). The *degrees of freedom*, F distribution and a probability value p , which is derived from F , are used to determine whether there is a statistically significant difference between populations of samples (see Section 2.5 about significance test). The higher p value, the more we can believe that the populations of samples are not statistically different (refer to Tabachnick [Tabachnick 1989] for more details of the main effect and ANOVA). In our experiment, the scene difference is not statistically significant and small enough to be ignored ($p \gg 0.05$ for all attributes). It follows our goal to investigate the tone mapping performance for indoor architectural scenes.

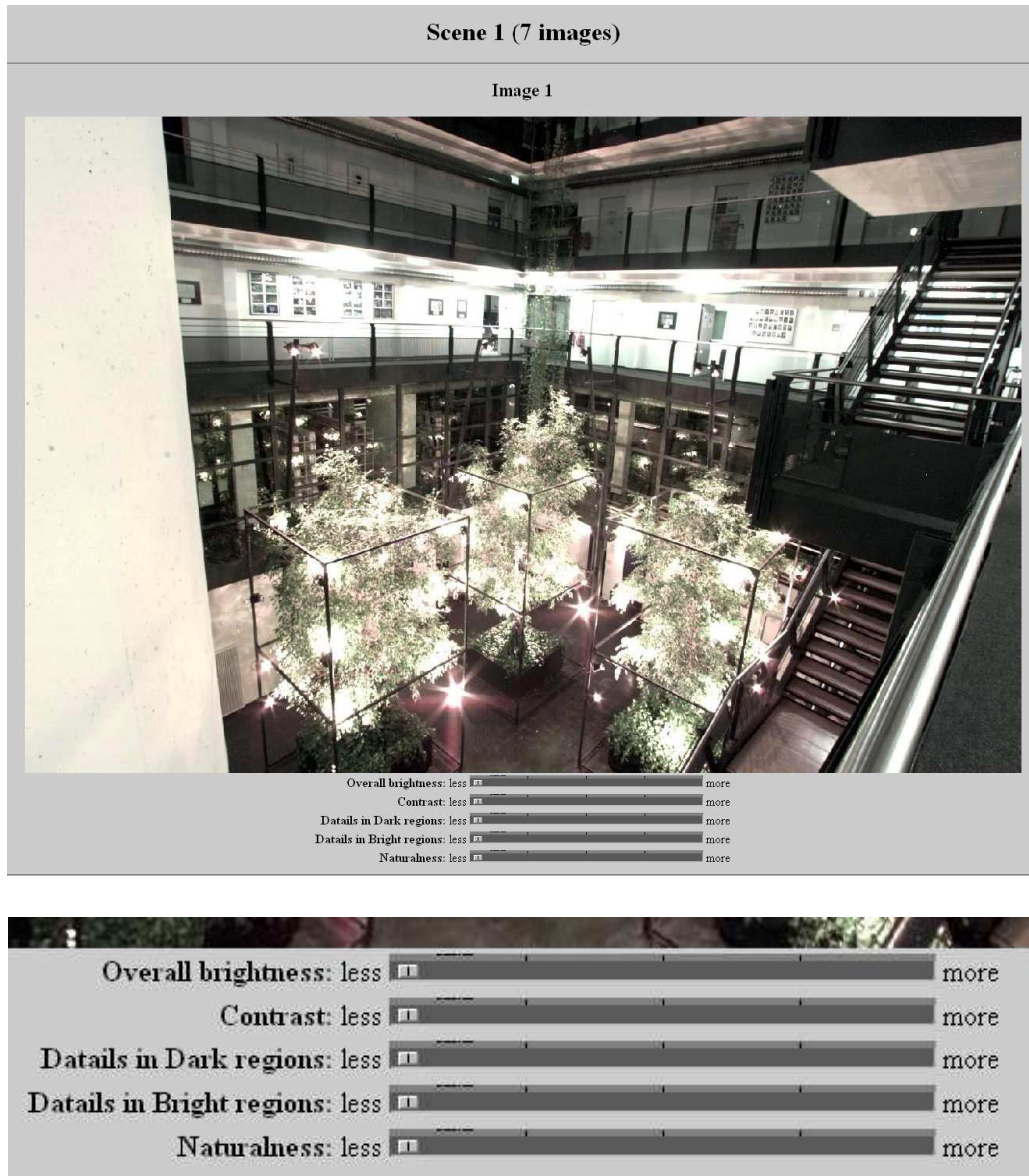


Figure 4.4: Screenshots of our perceptual experiment and its user interface part.

Overall Brightness

Figure 4.5 shows the main effect of the TMOs for overall brightness. According to the significance values, overall brightness is the most differently perceived attribute. As seen in the figure, it is manifested that images produced by the linear, Pattanaik, Ward, and Drago methods (i.e., global methods) have substantially higher overall brightness than the others. These TMOs are perceived the most differently when compared to their corresponding real-world views. Figure 4.5(b) shows the estimated differences of means. The top and bottom bars show 95% confidence interval, and the middle bar shows the mean value of difference. Note that a pair containing 0 difference (the red line in plots) in its 95% confidence interval indicates that they are not significantly different.

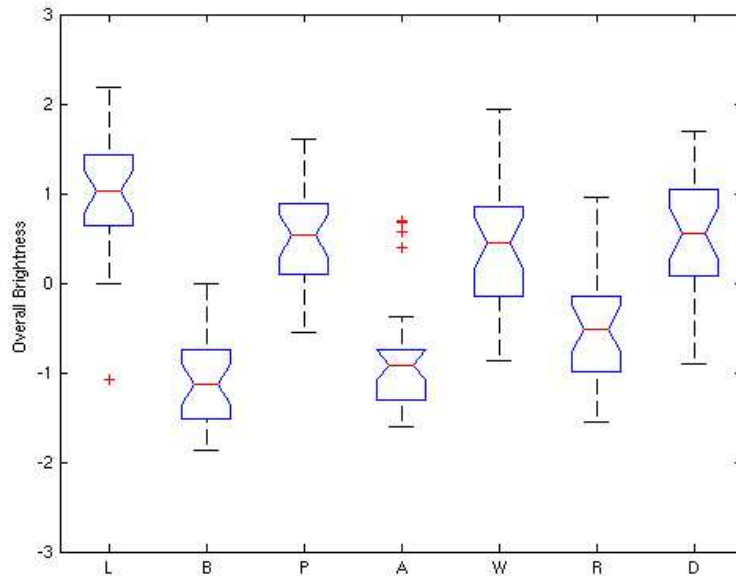
The same result is shown in Table 4.1 as significance values calculated by t -test (see Section 2.5 about t -test). As shown in Figure 4.5(b) and Table 4.1, the pair of the linear TMO and bilateral filtering has the biggest population difference. On the other hand, the pairs of bilateral filtering and Ashikhmin, of Pattanaik and Ward, of Pattanaik and Drago, of Ashikhmin and Reinhard, and of Ward and Drago are not significantly different for overall brightness.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0000	0.0012	0.0000	0.0037	0.0000	0.0140
Bilateral		0.0000	0.0546	0.0000	0.0002	0.0000
Pattanaik			0.0000	0.9507	0.0000	0.5444
Ashikhmin				0.0000	0.0873	0.0000
Ward					0.0000	0.5580
Reinhard						0.0000

Table 4.1: p -values computed by t -test for overall brightness.

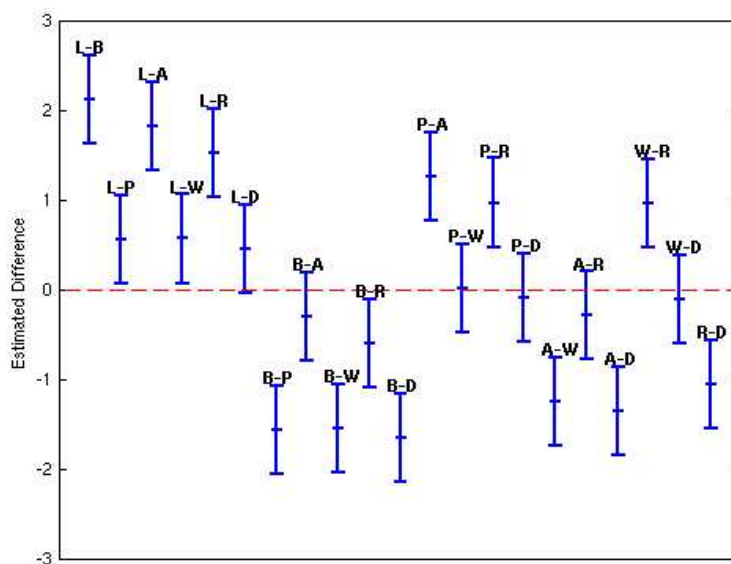
Overall Contrast

The main effect for overall contrast is shown in Figure 4.6. The linear, Pattanaik, and Ward methods have higher overall contrast than the others. Global operators have stronger contrast than local ones as shown in Figure 4.6(a). It corresponds to the expectations because global methods require high overall contrast to retain details in low-contrast regions. The estimated mean difference and t -test result are shown in Figure 4.6(b) and Table 4.2. The result of ANOVA (Figure 4.6(a)) shows significant difference between all TMOs, however, if we make pairwise comparisons for each TMO, a number of pairs are not considered significantly different (Figure 4.6(b) and Table 4.2).



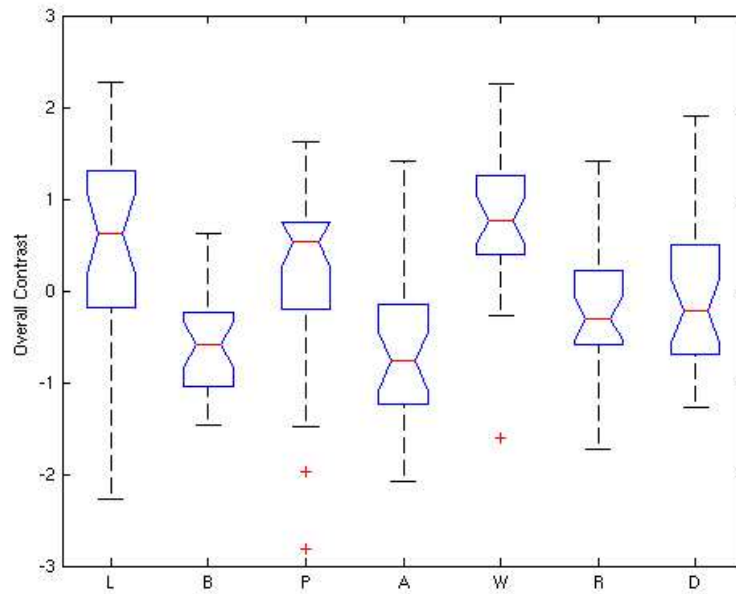
(a) Overall brightness. $F = 46.14$, $p = 0.0$.

A box shows the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data.

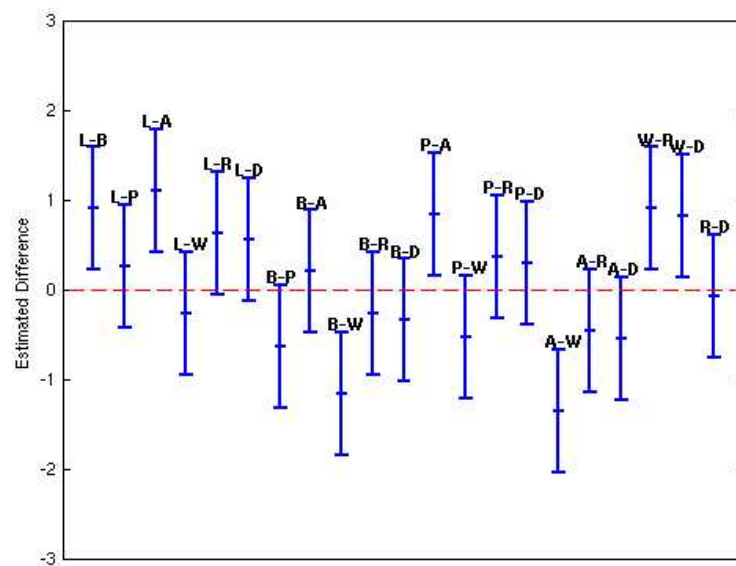


(b) Estimated mean difference. The middle bar shows the mean value of estimated difference. The top and bottom bars show 95% confidence interval.

Figure 4.5: Distributions and F and p values of overall brightness for the main effect of the TMOs. The TMOs are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods.



(a) Overall contrast. $F = 8.74$, $p = 2.1058E - 08$.



(b) Estimated mean difference.

Figure 4.6: Distributions and F and p values of overall contrast for the main effect of the TMOs. The TMOs are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0010	0.3531	0.0001	0.2713	0.0121	0.0260
Bilateral		0.0161	0.4003	0.0000	0.2197	0.1132
Pattanaik			0.0017	0.0326	0.1323	0.2305
Ashikhmin				0.0000	0.0322	0.0125
Ward					0.0000	0.0001
Reinhard						0.6942

Table 4.2: p -values computed by t -test for overall contrast.

Details in Dark Regions

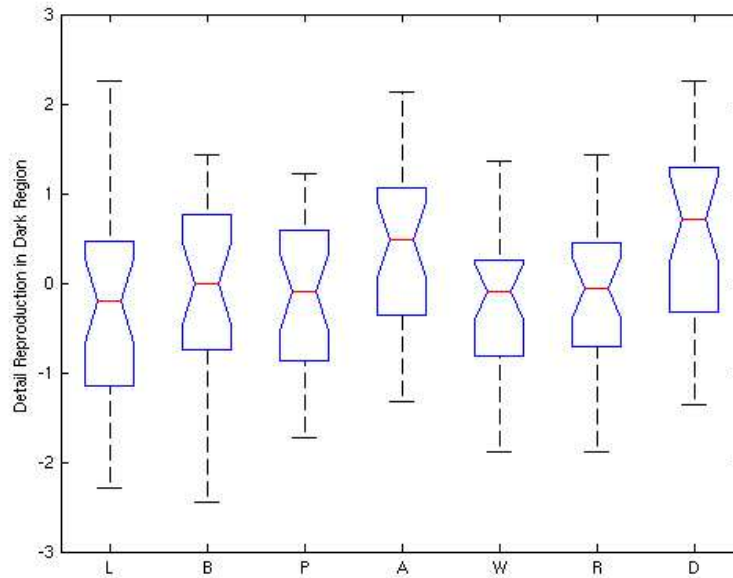
Detail reproduction in dark regions (Figure 4.7) show the least significance among the attributes, but it is still highly significant ($p = 0.0054$). The Ashikhmin and Drago methods are perceived to have the most details reproduced in dark regions. The linear, Pattanaik, Ward, and Reinhard methods have almost equal scores, and the bilateral filtering has slightly less detail reproductions than those four. Although ANOVA shows that TMOs are significantly different for detail reproduction in dark regions ($p = 0.0054$ in Figure 4.7(a)), estimated difference of means and t -test show that a number of TMOs are not significantly different if they are compared pairwise.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.4869	0.6613	0.0252	0.9355	0.6131	0.0087
Bilateral		0.7400	0.0884	0.3572	0.8153	0.0311
Pattanaik			0.0308	0.5240	0.9247	0.0096
Ashikhmin				0.0057	0.0453	0.5508
Ward					0.4793	0.0017
Reinhard						0.0147

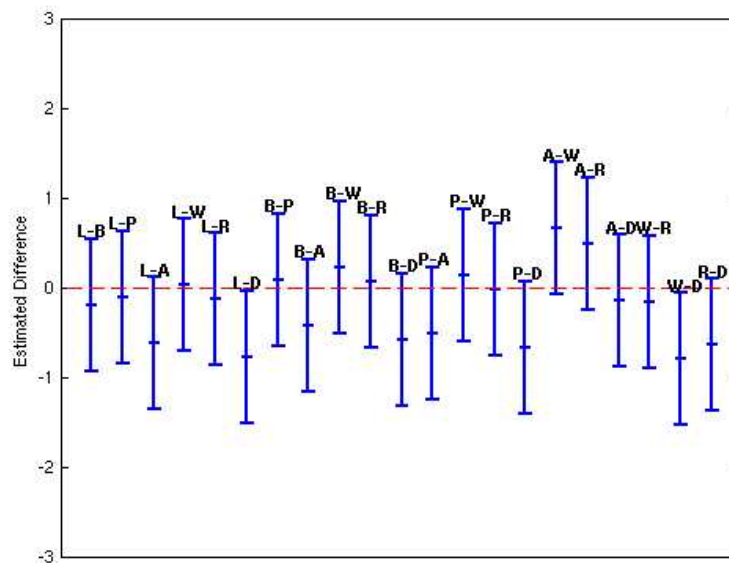
Table 4.3: p -values computed by t -test for detail reproduction in dark regions.

Details in Bright Regions

Detail reproduction in bright regions is the second most differently perceived attribute as shown in Figure 4.8. The bilateral filtering, Ashikhmin, and Reinhard methods provide significantly more detail reproductions in bright regions than the others. According to Figure 4.8(a), all of the local operators are perceived with more detail reproductions than global ones. This result comes from the fact that



(a) Detail reproductions in dark regions. $F = 3.18$, $p = 0.0054$.



(b) Estimated mean difference.

Figure 4.7: Distributions and F and p values of detail reproductions in dark regions for the main effect of the TMOs. The TMOs are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods.

local operators use different scales for small regions of an image while global operators use only one scale for the whole image and tend to saturate bright parts of an image. Figure 4.8(b) and Table 4.4 also show the distances between local and global TMOs.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0000	0.0001	0.0000	0.0003	0.0000	0.0000
Bilateral		0.0000	0.3665	0.0000	0.8794	0.4011
Pattanaik			0.0000	0.8670	0.0000	0.0001
Ashikhmin				0.0000	0.2044	0.0653
Ward					0.0000	0.0001
Reinhard						0.4053

Table 4.4: p -values computed by t -test for detail reproduction in bright regions.

Naturalness

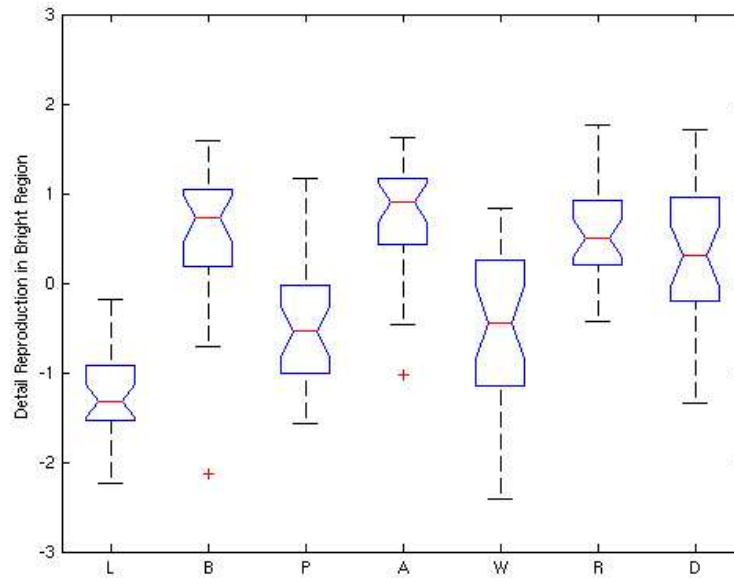
Figure 4.9 shows the main effect for naturalness. As can be seen in the figure, the Ward, Reinhard, and Drago methods are perceived to have the most natural appearance. As same as detail reproduction in dark regions, ANOVA shows that all TMOs are significantly different ($p = 8.3877E - 08$), however, each pairwise comparison does not show significant difference for almost half of pairs for naturalness.

	Bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
Linear	0.0305	0.1845	0.0864	0.0000	0.0000	0.0000
Bilateral		0.3282	0.6094	0.0329	0.0644	0.0049
Pattanaik			0.6420	0.0012	0.0017	0.0000
Ashikhmin				0.0069	0.0124	0.0004
Ward					0.5603	0.7800
Reinhard						0.2608

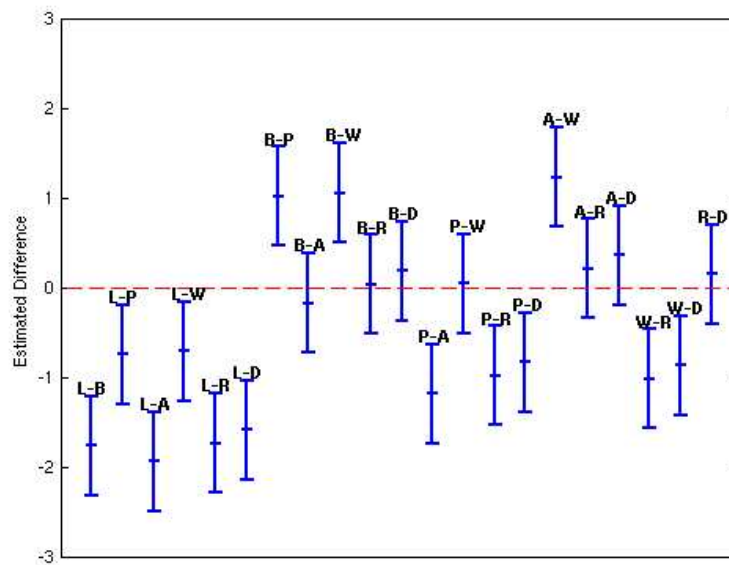
Table 4.5: p -values computed by t -test for naturalness.

4.4.2 Mahalanobis Distances

Multivariate analysis of variance (MANOVA) was run to estimate *Mahalanobis distances* between TMOs (see Section 2.5 about MANOVA). Mathematically, Ma-

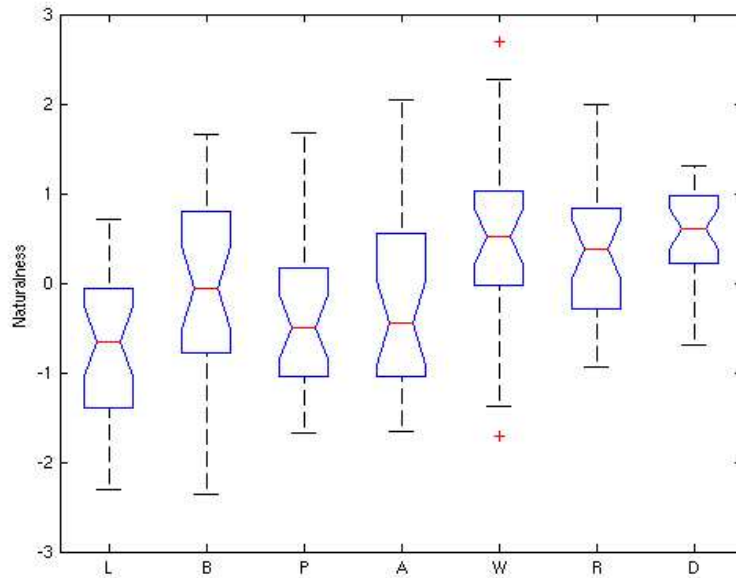


(a) Detail reproductions in bright regions. $F = 30.355$, $p = 0.0$.

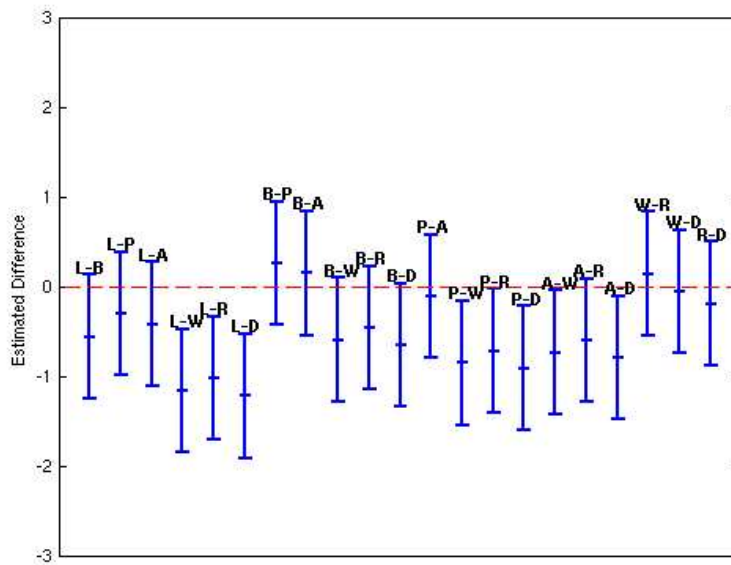


(b) Estimated mean difference.

Figure 4.8: Distributions and F and p values of detail reproductions in bright regions for the main effect of the TMOs. The TMOs are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods.



(a) Naturalness. $F = 8.11$, $p = 8.3877E - 08$.



(b) Estimated mean difference.

Figure 4.9: Distributions and F and p values of naturalness for the main effect of the TMOs. The TMOs are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods.

halanobis distance $d(x_1, x_2)$ between x_1 and x_2 is formulated as

$$d(x_1, x_2) = \langle x_1 - x_2, \mathbf{S}^{-1}(x_1, x_2) \rangle, \quad (4.1)$$

$$\mathbf{S}(x_1, x_2) = \frac{1}{n-1} \sum_{i=1}^n (x_1 - x_2)^t (x_1 - x_2) \quad (4.2)$$

where \mathbf{X} is a data matrix, \mathbf{x}_i is the i -th row of \mathbf{X} , $\bar{\mathbf{X}}$ is a row vector of means, and n is the number of rows. Mahalanobis distance is a measurement of similarity of data points relative to the probability distribution of data, which has different variances along different dimensions. Refer to [Tabachnick 1989] for more details of Mahalanobis distances.

Table 4.6 shows the Mahalanobis distances among the TMOs given after running MANOVA. According to Table 4.6, the linear tone mapping and bilateral filtering are perceived the most differently when compared with their corresponding real-world views. The second and the third most different combinations come from the combination of the linear tone mapping and Ashikhmin method and of the linear tone mapping and Reinhard method. All of the three biggest differences are found with respect to the linear tone mapping. On the other hand, the least difference is provided between bilateral filtering and the Ashikhmin method. This result is visualized in Figure 4.10 as a dendrogram plot of a hierarchical binary tree. An interesting result shown in Figure 4.10 is that those seven TMOs are divided into global and local methods by Mahalanobis distances. Three local operators (bilateral, Ashikhmin, and Reinhard) are similar to each other and four global operators (linear, Pattanaik, Ward, and Drago) are similar to each other, but both categories of global and local operators are separated by a large distance.

	bilateral	Pattanaik	Ashikhmin	Ward	Reinhard	Drago
linear	15.4530	1.6541	14.2361	2.7122	10.6089	6.6940
bilateral		7.4749	0.6674	9.2726	1.3353	8.8120
Pattanaik			6.4395	1.1613	3.9887	2.8066
Ashikhmin				8.9709	1.2405	6.2989
Ward					4.5301	2.9536
Reinhard						3.7406

Table 4.6: Mahalanobis distances among the TMOs calculated after running MANOVA. The three biggest distances are colored in red and the three smallest distances are colored in blue. All of the biggest differences are from the linear tone mapping. Those Mahalanobis distances are visualized by a hierarchical binary tree in Figure 4.10.

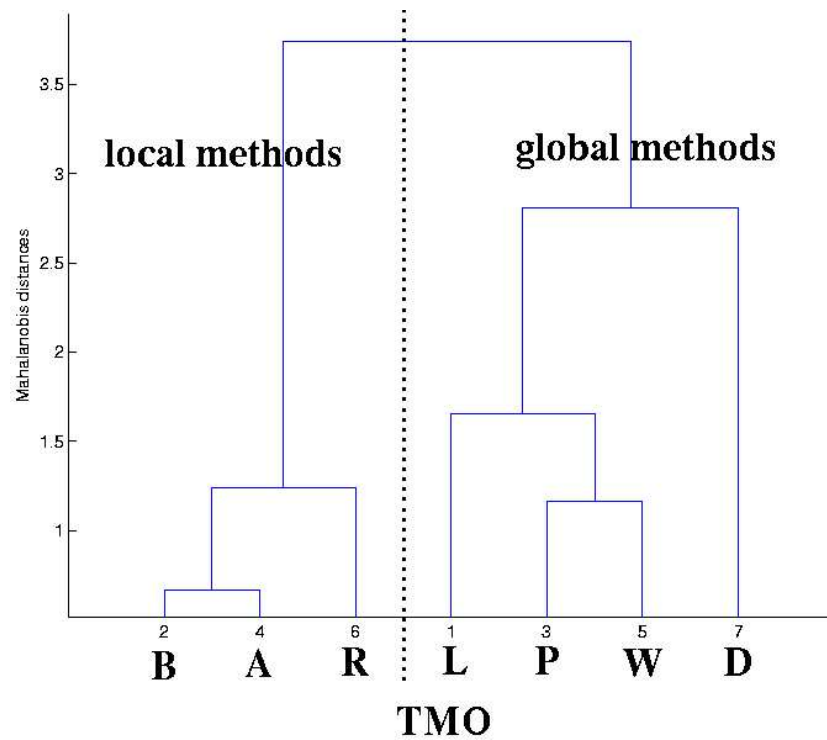


Figure 4.10: Tree-structured Mahalanobis distances to determine similarity among the TMOs given by MANOVA. As in Figure 4.5, the TMOs are labeled as linear (L), bilateral filtering (B), Pattanaik (P), Ashikhmin (A), Ward (W), Reinhard (R), and Drago (D) methods.

Note that those TMOs are divided into global and local methods for human perception.

4.5 Discussion

In this Section, we discuss the results obtained in our experiments and whenever possible we compare them with the results obtained in other independent studies [Kuang et al. 2005, Ledda et al. 2005]. It is worth nothing that Kuang et al. [Kuang et al. 2005, Kuang et al. 2004] have not used any reference scenes in their experiment. Ledda et al. [Ledda et al. 2005] showed HDR images to their subjects as the reference, but the subjects have actually never seen the real-world scenes captured in the HDR images.

In parallel to our experiment, two other sets of experiments were presented. Ashikhmin and Goral [Ashikhmin and Goral 2006] conducted two experiments consisting of three parts: image preference, image naturalness with absence of a given real-scene reference, and comparison with the real-scenes. In their experiments, they employed five TMOs

- Histogram adjustment TMO [Ward Larson et al. 1997]
- Gradient domain TMO [Fattal et al. 2002]
- Photographic TMO [Reinhard et al. 2002]
- Trilateral filtering [Choudhury and Tumblin 2003]
- Adaptive logarithmic TMO [Drago et al. 2003]

and four scenes for the one involving the real-world views. They presented all tone mapped images to subjects and asked them a question “which image is the closest to the real scene in front of you?”. The results in the experiments about preference and naturalness without real-scene views are randomly scattered. However, once the real-scenes are given as references, the results become more consistent such that *Gradient domain* TMO performed well while *trilateral filtering* did not obtain good scores when compared to the real-views.

Another set of experiments was conducted by Čadík et al. [Čadík et al. 2006]. 14 TMOs are employed in their experiments:

- linear TMO
- Ashikhmin [Ashikhmin 2002]
- Chiu et al. [Chiu et al. 1993]
- Trilateral filtering [Choudhury and Tumblin 2003]
- Adaptive logarithmic TMO [Drago et al. 2003]

- Bilateral filtering [Durand and Dorsey 2002]
- Gradient domain TMO [Fattal et al. 2002]
- Low curvature image simplifiers (LCIS) [Tumblin and Turk 1999]
- Pattanaik and Yee [Pattanaik and Yee 2002]
- Photographic TMO [Reinhard et al. 2002]
- Schlick [Schlick 1994]
- Tumblin and Rushmeier [Tumblin et al. 1999]
- Ward [Ward 1994]
- Histogram adjustment TMO [Ward Larson et al. 1997].

Their experiments contain two parts: 1) rating each tone mapped image for overall image quality, brightness, contrast, reproduction of details, and reproduction of colors with given real-world views (rating experiment) and 2) ordering printouts of all tone mapped images for overall image quality, overall contrast, brightness, color and detail reproduction with no given reference views. In their conclusions, the best overall quality is exhibited by [Reinhard et al. 2002] and [Ward Larson et al. 1997] in this order while the worst was [Chiu et al. 1993].

For our experiment, the result of the multivariate analysis (see Section 4.4) shows that the seven TMOs were perceived very differently in terms of all of the attributes when compared to their corresponding real-world views.

Overall Brightness Observer assessments of overall brightness shows the most significant differences among the TMOs, and global operators (the linear, Pattanaik, Ward, and Drago methods) have more brightness than local ones (the bilateral filtering, Ashikhmin, and Reinhard methods). This result has a strong agreement with Čadík et al.'s results. They also reported the methods of Ward and Drago with relatively good ratings about overall brightness compared to bilateral filtering and Ashikhmin TMO. Only a difference is Reinhard TMO which was rated low as same as the other local TMOs in our experiment but obtained quite good score in Čadík et al.'s experiment.

Overall Contrast Regarding overall contrast, global operators have more contrast than local ones, but the difference is less pronounced than for overall brightness. Since overall brightness and contrast are correlated [Yoshida et al. 2006], this result is expected. The linear, Pattanaik, and Ward methods show more overall

contrast reproduction than the others. However, it was shown that the bilateral filtering and photographic TMOs (i.e., local operators) had the highest rating scales in overall contrast when no reference is provided for subjects [Kuang et al. 2005]. Compared to Čadík et al.'s results, there is the same tendency as overall brightness. Both of their and our results agree with each other except Reinhard TMO.

Detail Reproduction in Dark Regions Detail reproduction in dark regions is the least significant of the attributes, but it is still highly significant. The Ashikhmin and Drago methods are perceived as providing the most details in dark regions. The bilateral filtering exhibits slightly better detail reproduction than the linear, Pattanaik, Ward, and Reinhard methods. In another independent study [Kuang et al. 2005] (performed without the reference scenes), the bilateral filtering and Reinhard methods have been reported to have high scores in detail reproductions in dark regions. It is similar to our findings. On the other hand, if an HDR image was provided as the reference on an HDR monitor (Ledda et al. [Ledda et al. 2005]), the bilateral filtering reproduced the poorest details in dark regions.

Detail Reproduction in Bright Regions Perceptual variation is the second highest for detail reproduction in bright regions. Counter to overall brightness, local operators are perceived with significantly more detail reproductions in bright regions than global ones. Even when no reference was provided, the local TMOs were considered to be better operators for detail reproduction in bright regions [Kuang et al. 2005]. In the study with the HDR references [Ledda et al. 2005], it can also be seen that the Reinhard TMO reproduced details in bright regions quite well. However, the fast bilateral filtering, Ward, and Drago TMOs were reported as having the opposite effect compared to our results.

Similar results to those of our experiments with respect to detail reproduction in dark and bright regions, have been also recently reported by Smith et al. [Smith et al. 2006] who proposed objective local and global contrast metrics. They considered 18 HDR images which have been compared to their tone mapped counterparts using their objective contrast distortion metrics.

Naturalness Concerning the naturalness attribute, the Ward, Reinhard, and Drago methods obtain highest scores. The Ward and Reinhard methods are also ranked as the second and the third preferred TMOs respectively in the research of Kuang et al., however, the fast bilateral filtering is not ranked in our experiments as high as in the experiments without reference [Kuang et al. 2004]. Ashikhmin

and Goral asked the subjects “which image is the closest to the real scene in front of you?”, and then their results show that the methods of Drago and Reinhard performed well as same as our results. In Čadík et al.’s experiment, they asked their subjects about “overall image quality” with respect to corresponding real-world views. Their question can be interpreted “how close the given image is to the reference view in front of you?”, i.e., naturalness. The results of Čadík et al. also have a strong agreement with ours about the methods of Ward and Reinhard.

For all attributes of TMOs, while the results of our experiments show some similarities to the results of other studies [Kuang et al. 2005, Ledda et al. 2005, Ashikhmin and Goral 2006, Čadík et al. 2006], a number of difference are observed as well. This may come from the difference between an experiment with or without reference. Even when a reference is provided to subjects, the results differ with an HDR image reference or with the corresponding real-world scene. Parameter setting for TMO matters as shown in [Yoshida et al. 2006] which is also presented in Chapter 5.

Similarity between TMOs In terms of the similarity of the tone mapping operators computed by Mahalanobis distances, the biggest differences are between the linear tone mapping and each of the fast bilateral filtering, the Ashikhmin method, and the photographic tone reproduction by Reinhard et al. (i.e., local methods). The least differences are between fast bilateral filtering and the Ashikhmin method, between the methods of Pattanaik and Ward, and between the Ashikhmin method and the photographic reproduction. The Mahalanobis distances are visualized in a dendrogram plot (Figure 4.10), which shows that all studied tone mapping operators are divided into global and local categories in terms of similarity.

4.6 Summary

We conducted a psychophysical experiment over seven tone mapping operators and two scenes with 14 human observers. We asked subjects to stay at the point where an HDR image was taken by camera and compare tone mapped images and their corresponding real-world views. Tone mapped images were shown one after another and subjects rated overall brightness, overall contrast, detail reproduction in bright and dark regions, and naturalness for each tone mapped image by using a slider. Our principal goal of this work was to study how people perceive tone mapped images when they are compared with the real-world views.

Our results demonstrate that qualitative differences in tone mapping operators have a systematic effect on the perception of scenes by human observers. They provide a solid basis for selecting the appropriate algorithm for a given application and for assessment of new approaches to tone mapping techniques.

Chapter 5

Analysis of Reproducing Real-World Appearance on HDR Displays

5.1 Introduction

A great variety of tone mapping operators have been developed in recent years (refer to [\[Reinhard et al. 2005\]](#) for a detailed survey) in response to accessible and simple high dynamic range (HDR) image acquisition technology. A majority of existing operators are designed to produce images that just “look good”. Some operators, especially those designed specifically for realistic image synthesis applications, use models of brightness or contrast perception to achieve a good match between the image’s appearance and the corresponding real-world scene. In practice, each operator boils down to an image processing algorithm that transforms HDR pixels into their low dynamic range (LDR) counterparts using either a monotonic function with respect to the HDR pixel intensity (global operators) or a more complex relation that involves local pixel neighborhood considerations (local operators). While new tone mapping operators are proposed, there is little understanding whether their improvements and additional complexity really lead to better images. It turns out that it is difficult to select one existing operator that consistently performs the best in terms of user preferences or fidelity to the original scene appearance for all HDR images [\[Reinhard et al. 2005\]](#).

In the following sections, we review previous work in Section 5.2, describe our

psychophysical experiments in Section 5.3. Results of our experiments are discussed in Section 5.4, followed by Section 5.6 to summarize this work.

5.2 Previous Work

Evaluation of tone mapping operators (TMOs) is an active research area [Drago et al. 2002, Kuang et al. 2004, Ledda et al. 2005, Yoshida et al. 2005], which at the current stage is more focused on choosing correct psychophysical techniques than on providing any clear guidance as to how existing operators should be improved to produce consistently high quality images. All existing evaluation methods treat each tested TMO as a “black box” and its performance is compared with respect to other operators, without explaining the reasons underlying human judgments. While some evaluation methods go one step further and attempt to analyze the reproduction quality of overall brightness, global contrast, and details (in dark and bright image regions) [Ledda et al. 2005, Yoshida et al. 2005], but again they are focused on comparing which operator is better for each of these tasks. Those studies do not provide any deeper analysis as to how pixels of an HDR image have been transformed and what the impact of such a transformation is on desired tone mapped image characteristics [Delahunt et al. 2005]. Another important question is how the outcome of the transformation depends on the particular HDR image content.

In all discussed evaluation experiments only one set of parameters per TMO and per HDR image is considered in order to reduce the number of images that must be compared by subjects. The choice of the parameters may dramatically affect the appearance of tone mapped images and thus the performance of a TMO. Experimenters commonly set such parameters based on their choice or a small pilot study, which may lead to the results that are biased by the choices of a limited number of subjects. Sometimes the original authors of TMOs are asked to prepare images according to their preference, since “they should be the best qualified to get the best results”. However, different people may have different ideas concerning the preferred image appearance and the meaning of “the best results” can be fuzzy. Even the calibration of the display used for image tuning and the actual experiment can affect the results. A limited number of TMOs offer a method of an automatic parameter estimation (e.g. [Reinhard 2003, Reinhard et al. 2005]). However, these estimation methods rely mostly on the authors observation for a small set of images or the practices borrowed from photography, rather than a study with a large number of images and subjects.

A standard pool of HDR images for tone mapping evaluation has been recently

proposed [Johnson 2005], but a common practice for every experimenter is to rely on his own version of tone mapped images, which makes cross-comparison of the results for independent evaluations difficult.

Another common problem is averaging the experimental results across subjects based on low-cross subject variability. This lack of variability can often be caused by the choices imposed on the subjects by the experiment design, which does not offer any possibility of adjusting the image appearance to the subject's real preferences within available range of parameters of the tested TMOs. The net result of published studies is that they often present contradictory results even if the same HDR images are used. Some operators shown as performing the worst in one experiment obtain the top scores in another experiment. This suggests that the TMO evaluation methodology should be improved.

In this work, instead of proposing a new TMO and then running the subjective evaluation to show that it performs better than the other operators, we take the opposite approach. We want to first identify the output tone characteristics that lead to perceptually attractive images. Therefore, we start from measuring the subjective preference and the perception of fidelity for images produced by a generic TMO, whose characteristic and parameters are well understood. Our goal is to find some universal rules that facilitate a design of the TMO that consistently produces preferred image appearance.

In this respect, there are some similarities of our approach goals with more fundamental research in psychophysics, which raises the issue of image appearance preferences as a function of various image characteristics. For example, Fedorovskaya et al. [Fedorovskaya et al. 1997] report that the relation between preference and colorfulness has a shape of inverted “U” with the maximum preference achieved for color saturation increased by 10%–20% in respect to the original image. Similar results are obtained for contrast and brightness manipulation [Jobson et al. 2002]. Higher color saturation is needed to compensate for reduced brightness of a display in order to achieve more natural image appearance [deRidder 1996]. The preferred mean luminance levels are found for images that contain human faces [Delahunt et al. 2005]. Image preferences with respect to colorfulness, contrast and brightness are studied in digital photography [Savakis et al. 2000].

What makes our study different from this fundamental research, which is motivated by the applications of color reproduction in television and photography, is that we focus on the particular problem of tone mapping HDR images for a broad range of display devices. For this purpose we use in our experiments HDR images, which are displayed on an HDR display with fully controllable minimum and maximum display luminance values. Therefore, we can investigate much wider dynamic range than is possible using traditional LDR displays and neutral

density filters.

5.3 Experiments

5.3.1 Introduction

We conducted two experiments on an HDR display to assess how people adjust the settings of a TMO. In Experiment I, the subjects were given the task of adjusting an HDR image shown on the HDR display so that it looked the best in their opinion. In Experiment II, the subjects sat in front of an HDR display showing an image and the corresponding real-world scene. Their task was to adjust the image to achieve the closest reproduction of the real-world scene on the display. Additionally, in Experiment II, we simulated several potential display devices by limiting the lowest and highest luminance outputs of the HDR display.

5.3.2 Subjects

There were in total 24 individual participants in two experiments. Four of them were female and the rest were male. The range of their age was 24 – 46 years and the average was 28. All of them reported normal or corrected to normal vision. All but two subjects were not aware of the purpose of the experiments. Eight subjects took part in both experiments, the others in only one of them. Experiment I was completed by 15 subjects. Experiment II involved a separate setup for each of three real-world scenes, therefore 13, 7, and 6 subjects completed Experiment II for each scene respectively. A single session took approximately 20 – 30 minutes for both experiments.

5.3.3 Stimuli and Apparatus

Experiment I employed 25 HDR images commonly used for testing TMOs (see Figures 5.1 – 5.3). There were 14 outdoor scenes, 9 indoor scenes, and 2 CG-rendered scenes. We did not include images of people or animals in the test set, since these are rare for HDR images. Some of the images depicted daytime scenes, the others night or evening scenes. The images were displayed in their original resolution or scaled to the resolution of the HDR display if they were too large.

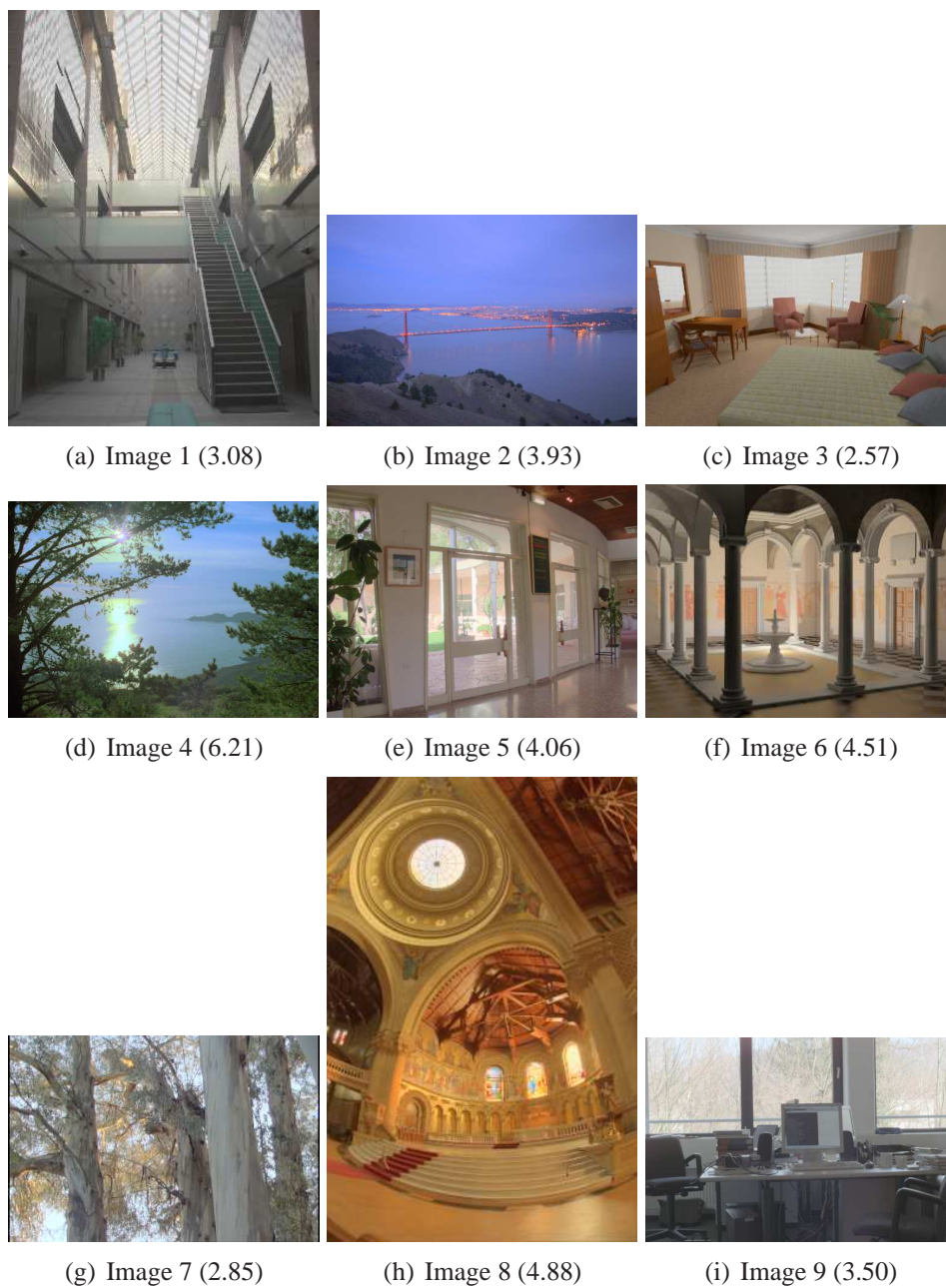


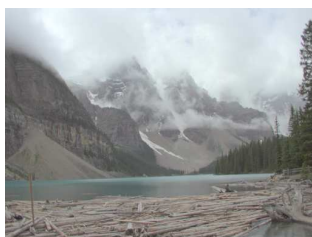
Figure 5.1: Images 1 – 9 for Experiment I. Their dynamic ranges in decimal-logarithmic units are shown in parentheses. All images are tone mapped by using [Drago et al. 2003].



Figure 5.2: Images 10 – 18 for Experiment I. Their dynamic ranges in decimal-logarithmic units are shown in parentheses. All images are tone mapped by using [Drago et al. 2003].



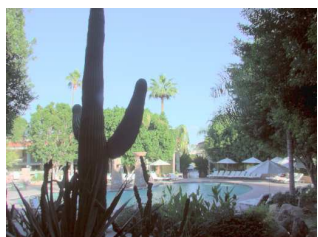
(a) Image 19 (3.54)



(b) Image 20 (2.60)



(c) Image 21 (2.98)



(d) Image 22 (4.25)



(e) Image 23 (3.84)



(f) Image 24 (3.16)



(g) Image 25 (3.48)

Figure 5.3: Images 19 – 25 for Experiment I. Their dynamic ranges in decimal-logarithmic units are shown in parentheses. All images are tone mapped by using [Drago et al. 2003].

Figure 5.4 shows the HDR images used for Experiment II. These are the HDR photographs of our experimental scenes that we set up next to the HDR display. Each of the three HDR images was created using the multiple exposure technique from 15 low dynamic range images taken with a Kodak Professional DCS560 mounted on a tripod. We used Robertson's method [Robertson et al. 1999] implemented in the *PFScalibration*¹ to calibrate a camera and create the HDR images. We selected the lens (Canon EF 50mm) and the position of the camera, so that the image displayed on the monitor closely matched the real scene. The subject's viewpoint was not restricted and the setup allowed them to have comfortable viewing of both the real scene and the display from the distance of about 1.5 times screen height from the HDR display. Images 26 and 27 (the left and the middle images in Figure 5.4) contain the same object layout but differ in the lighting condition. Both scenes were lit with the 800 Watt HMI lamp (JOKER-BUG 800), which gave approximately daylight illumination. For Image 27, the lamp was covered by the diffuser (Lightbank) to decrease the intensity of the light source. As shown in Figure 5.4, the absolute luminance values were very different with or without a diffuser for the HMI lamp. The table setup in Images 26 and 27 included a MacBeth Color CheckerTM, an 18% reflective gray card and several objects ranging in their reflectance from black to white. The experimental sessions for all images except Image 28 were conducted in the room whose lighting condition could be fully controlled and was set to a typical dark office illumination (64 lux). In the pilot study we verified that the level of ambient light does not have a significant influence on the results. For Image 28, the experiment was conducted with natural light and completed within two hours in the afternoon under stable weather conditions.

The images were shown on the Brightside 18" LED-based HDR display [Seetzen et al. 2004] which consists of an LCD panel ($1,280 \times 1,024$ pixels) and a matrix of 760 separately controlled white LEDs, acting as a back light. The minimum and maximum luminance levels of the display we used for the experiments were 0.2 and 3,000 cd/m^2 , which gave the maximum dynamic range of 4.18 log-10 units. The HDR monitor was calibrated by measuring its luminance response for a range of input values using the MINOLTA LS-100 luminance meter. Then, the measured values were used to create an inverse lookup table, which was used by the display driver. The display driver was implemented in graphics hardware as a fragment program to allow for real-time interaction with images.

Although tone mapping of images for the HDR display may seem like a futile exercise, we found several reasons for this approach. Firstly, we wanted to check if tone mapping is necessary for HDR displays, and if it is so, what kind of tone

¹<http://www.mpi-inf.mpg.de/resources/hdr/calibration/pfs.html>

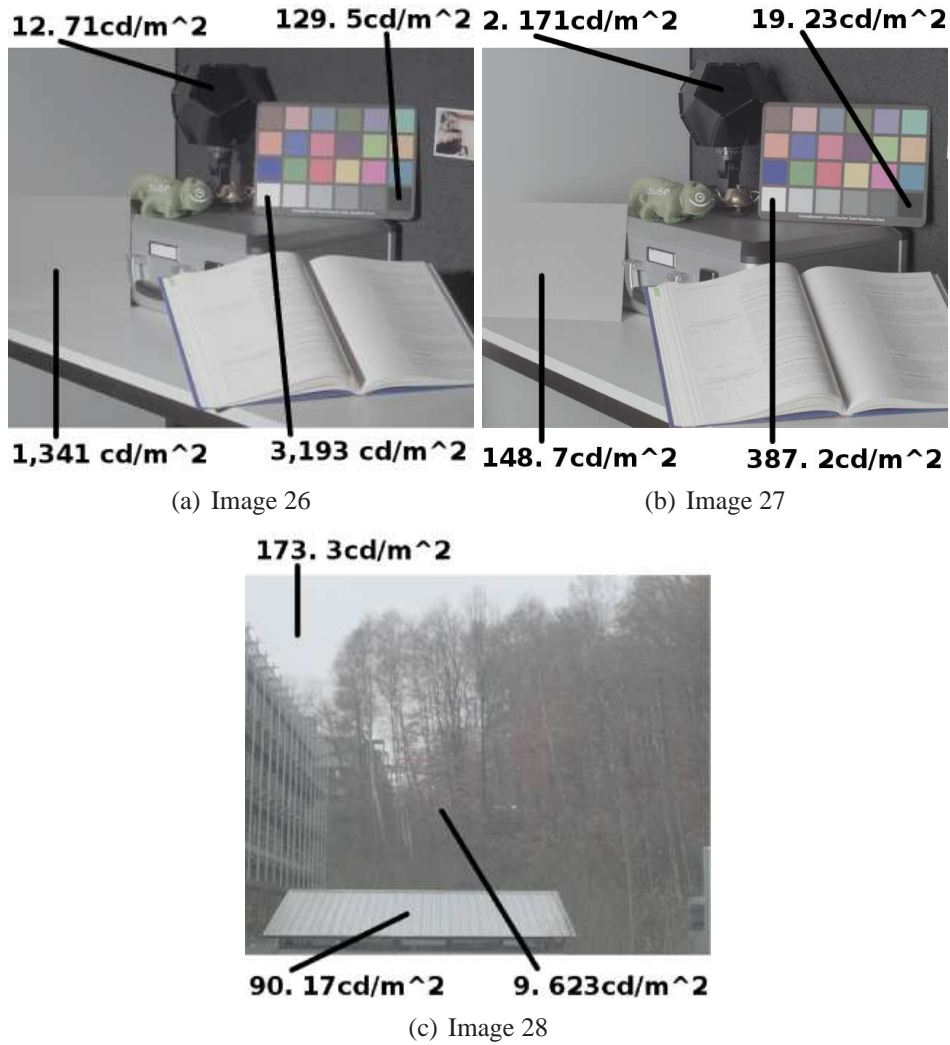


Figure 5.4: Three HDR images and several measurements of luminance of the real scenes. Their dynamic ranges in \log_{10} unit are also shown after each number of the images. These images were shown to the subjects with their corresponding real-world views in Experiment II and without reference as done in Experiment I. The Drago TMO is applied for the convenience of view.

mapping. For this reason, we included several images whose dynamic range exceeded the dynamic range of the display (refer to Figures 5.1 – 5.3). Secondly, we in fact artificially limited minimum and maximum luminance of the display in the experiments as described later (although most of the data has been collected for the full dynamic range). We also did not want to use different displays of different color characteristic in the experiments.

5.3.4 Generic Tone Mapping

The purpose of our psychophysical experiments is to collect data from human observers to determine what are the desired or important properties of a TMO. This knowledge should help in the design of new TMOs or automatic parameter estimation for the existing TMOs. Since examining all possible TMOs is not feasible in an experimental setup, we consider only a global TMO that involves linear scaling and shifting of color values in the logarithmic domain. Even though this is probably the simplest TMO that is practically used, it can mimic the behavior of many global TMOs, such as [Tumblin 1999] or [Ferwerda et al. 1996], and is in fact a part of any TMO that requires “gamma correction”, such as [Reinhard et al. 2002] (since a power function that is used in gamma correction corresponds to linear scaling in the logarithmic domain). Many TMOs produce output pixel values in an arbitrary range, which must be linearly scaled or shifted to fit the dynamic range of a display (e.g. [Fattal et al. 2002, Durand and Dorsey 2002, Reinhard et al. 2005]). For these and other TMOs such scaling (contrast adjustment) and shifting (brightness adjustment) operations are essential for the final appearance of a tone mapped image and are therefore analyzed in this work in more detail.

The generic TMO we use in the experiments is described with three parameters: brightness, contrast, and saturation of color. Brightness and contrast parameters are considered as an offset of luminance and as a difference between the maximum and minimum luminance values, respectively. To adjust color saturation, color coordinates are interpolated or extrapolated between the original pixel color and its corresponding luminance value for the D65 white point. All adjustments are performed in the logarithmic domain to approximate non-linear response of the human visual system to light. Formally, the TMO can be modeled as

$$\log_{10} R' = c \cdot \log_{10} R + b, \quad (5.1)$$

$$\begin{aligned} \log_{10} Y' &= 0.2126 \log_{10} R' + 0.7152 \log_{10} G' \\ &\quad + 0.0722 \log_{10} B', \end{aligned} \quad (5.2)$$

$$\log_{10} R'' = \log_{10} Y' + s(\log_{10} R' - \log_{10} Y') \quad (5.3)$$

where b , c , and s are brightness, contrast, and color saturation parameters respectively, Y' is the new luminance value, and R'' is the output red channel value. Equations (5.1) and (5.3) are applied for green and blue channels in the same way as for the red channel. Note that Y' is an approximation of luminance, which is used for our convenience (luminance should be a weighted sum of linear instead of logarithmic R , G and B coordinates). To assure that the adjustment of contrast has a minimum impact on the perceived brightness of a scene, the pixels of each HDR image were multiplied by a constant factor, so that the median luminance value of each image was $\bar{Y} = \log_{10}(1) = 0$. This way the multiplication by the contrast parameter in Equation (5.1) “stretched”, but did not shift image histogram. This is illustrated in Figure 5.5.

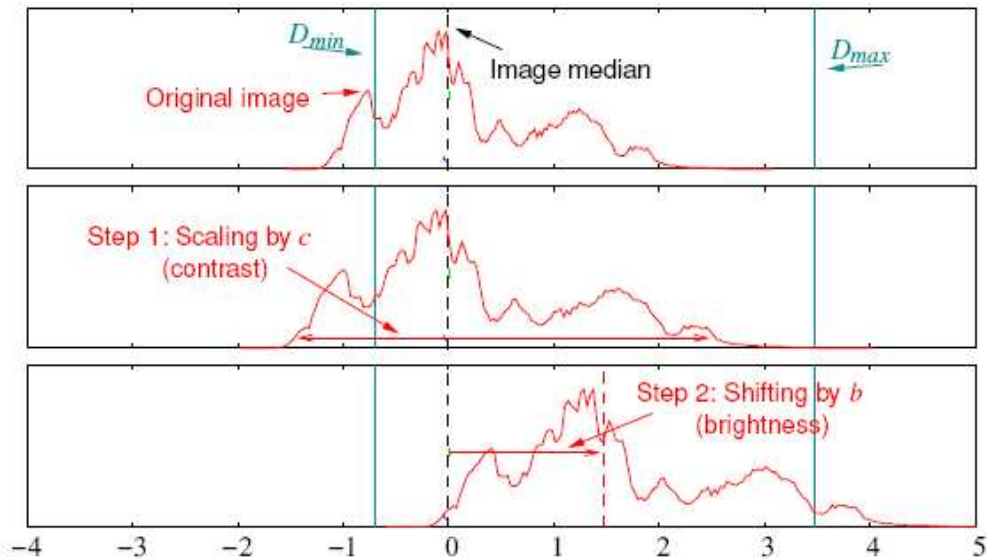


Figure 5.5: Illustration how the generic tone mapping modifies the image histogram (see Equation (5.1)).

5.3.5 Experimental Procedure

The two psychophysical experiments were conducted with and without reference scenes, respectively. For both experiments, each HDR image was shown on the HDR monitor one after another with a user interface that allowed the subjects to interactively adjust parameters of brightness, contrast, and color saturation using a mouse. The first two parameters were adjusted using a 2D slider interface and color saturation was adjusted using a 1D slider as shown in Figure 5.6. Since

we found in the pilot study that brightness and contrast are difficult to control separately, we decided to use a 2D slider that would allow adjustment of both parameters at the same time. In Experiment I, the subjects were asked to adjust these parameters until the most preferred reproduction of each HDR image was achieved in their own opinion without reference images (preference). In Experiment II, their task was to achieve the closest reproduction of the real-world view (fidelity). They were asked to reproduce the details of all objects in an HDR image as seen in the real scene and, if possible, to adjust the HDR image brightness to match the real scene (see Figure 5.7 for the settings of Experiment II).

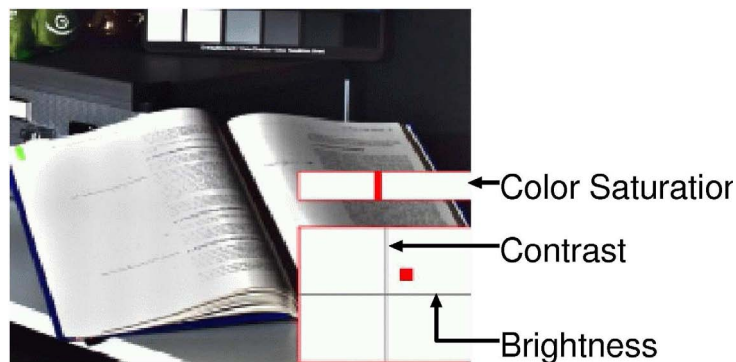
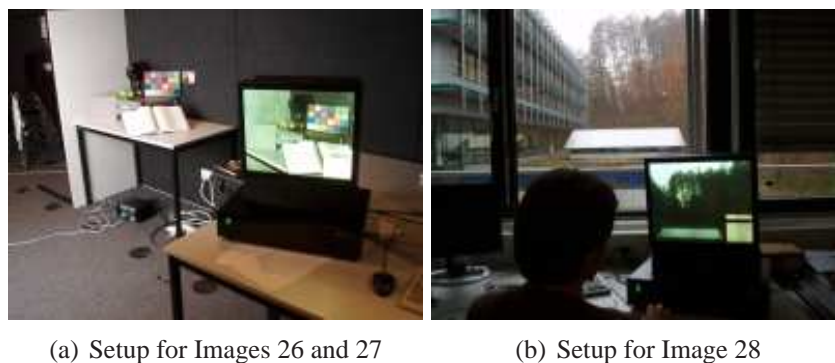


Figure 5.6: A screenshot of the user interface used in our experiments. The horizontal and vertical axes of the 2D slider (bottom right) adjust brightness and contrast, respectively. The 1D slider above changes color saturation.

The parameters of brightness, contrast, and color saturation of the generic TMO (refer to Section 5.3.4) were allowed to be adjusted within the range of $-3.0 - 5.0$, $0.1 - 4.0$, and $0.1 - 4.0$ respectively. Before starting the experiments, the ranges were checked to be large enough to reproduce everything from very dark/low con-



(a) Setup for Images 26 and 27

(b) Setup for Image 28

Figure 5.7: Setup used for Experiment II.

trast images to extremely bright/high contrast images with color settings ranging from grayscale to color-saturated image.

While Experiment I was conducted using the full dynamic range of the HDR monitor ($0.2 - 3,000 \text{ cd/m}^2$), in Experiment II, we restricted the minimum and the maximum luminance of the HDR display to simulate a range of potential display devices as listed in Table 5.1. Additionally, for each simulated display, the subjects were given a questionnaire sheet to mark the score of its reproduction, which could be “good” (3), “average” (2), or “poor” (1). The subjects were naïve as to what technically differs in each of the 14 dynamic range and brightness settings.

#	D_{min}	-	D_{max}	#	D_{min}	-	D_{max}
1	0.2	-	3,000	8	1.0	-	3,000
2	0.2	-	80	9	80.0	-	1,000
3	0.2	-	200	10	80.0	-	3,000
4	0.2	-	1,000	11	200.0	-	1,000
5	1.0	-	80	12	200.0	-	3,000
6	1.0	-	200	13	1,000.0	-	3,000
7	1.0	-	1,000	14	0.2	-	3,000

Table 5.1: The range of the minimum and maximum luminance values of the HDR display that simulates potential displays (given in cd/m^2). Note that the dynamic range between 0.2 and 3,000 cd/m^2 was used twice in the test to validate consistency of the results.

Finally, another experiment was conducted with Images 26 – 28 and four subjects using the same procedure as in Experiment II but the task was to adjust parameters to their preference (as in Experiment I) and no reference scene was given.

5.4 Results

The results for both experiments are summarized in Figure 5.8. The plot shows a large variance in the preferred TMO parameters, which indicates that the subjects used a broad range of possible parameters. There is also a strong correlation between brightness and contrast parameters. The contrast parameter is biased toward an enhanced contrast ($c > 1$ in Equation (5.1) indicates that the contrast was higher in a tone mapped image than in an original image).

Preliminary screening indicates that the results for Subject 22 are significantly different than for the other participants (probably due to improper use of the user

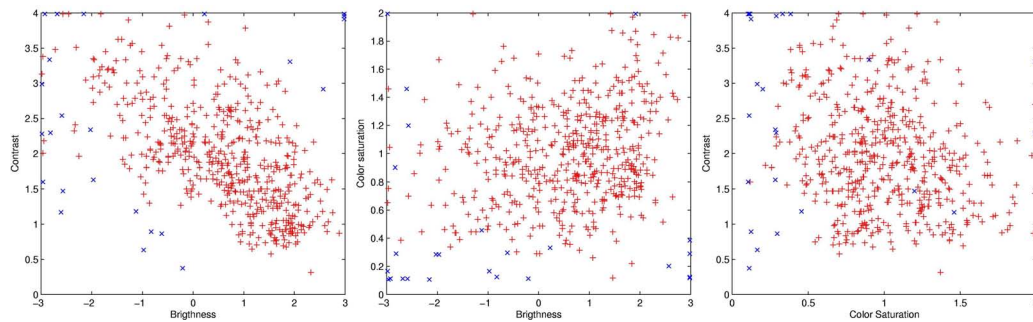


Figure 5.8: The results for all images, all subjects and for the full dynamic range of the display (red '+'). One subject regarded as an outlier is marked with blue '×'es.

interface) and therefore this data is removed from the further analysis (marked as blue '×'es in Figure 5.8).

We ran the multivariate analysis of variance (MANOVA) to test *main effects* of subjects' gender and expertise on measured parameters. The F distribution and a probability value p , which is derived from F , are used to determine whether there is a statistically significant difference between populations of samples. The higher p value, the more we can believe that the populations of samples are not statistically different. In our experiment, the gender difference is not significant ($F(3, 496) = 1.187, p \gg 0.05$ and $F(3, 360) = 1.970, p \gg 0.05$ for Experiments 1 and 2, respectively) as $p > 0.05$ shows that the difference between populations of samples (male and female in this case) is not statistically significant. Two people were aware of the experiment purpose, and therefore they were considered as experts. The population means of experts and non-experts are however not significantly different ($F(3, 496) = 0.3237, p \gg 0.05$ and $F(3, 360) = 2.2304, p \gg 0.05$ for Experiments 1 and 2, respectively). Therefore, we analyze all collected data together in the following sections.

To better understand the source of large parameter variations, we plot brightness and contrast parameter settings separately for several selected subjects and images in Figure 5.9. Similar plots for all subjects and images can be found in Figures 5.10 and 5.11. The left pane of Figure 5.9 shows that the settings can significantly differ between subjects, ranging from the preference for high contrast and low brightness (Subject 6) to the opposite preference for low contrast but bright images (Subject 2). The significant difference of subjects' settings is statistically established by MANOVA ($F(52, 1872.8) = 10.7864, p < 0.05$ and $F(64, 1349.0) = 7.6678, p < 0.05$ for Experiments 1 and 2). We can expect that two different individuals have different notions of a perfect image, therefore

the TMO settings must be affected by the subject's tastes. This is an important observation with several consequences. Firstly, a TMO designed to render the best looking images should account for the user's tastes, for example by offering user adjustable parameters. Secondly, when ranking or assessing performance of TMOs in subjective experiments (e.g., [Ledda et al. 2005]), the subjective influence should be taken as a factor in the analysis since two different subjects are likely to propose two different TMO rankings if they differ in their tastes. Finally, we cannot average parameter settings across all subjects for further analysis, since those parameters significantly differ across individuals.

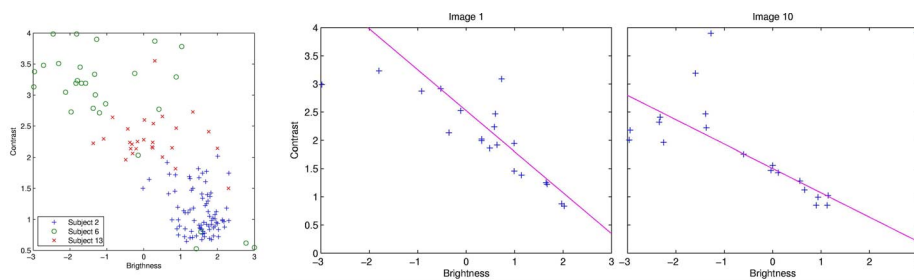


Figure 5.9: The contrast–brightness relation for three selected subjects (left) and images (center – Image 1; right – Image 13). Both contrast and brightness settings differ significantly from subject to subject and from image to image.

More consistency in the parameter settings can be observed across images. The center and right panes of Figure 5.9 show that both parameters follow a similar line of decreasing contrast and increasing brightness. While the images follow the similar pattern of parameter settings, the population means of the parameters are significantly different ($F(72, 1414.4) = 7.6420$, $p < 0.05$ and $F(6, 718) = 17.1307$, $p < 0.05$ for Experiments 1 and 2). This indicates that the TMO settings are affected by image characteristics.

5.5 Discussion

5.5.1 Contrast and Brightness Preference

To understand what the subjects' motivation for the choice of contrast and brightness parameters was, we plot the histograms of the resulting images in Figure 5.12. Although each subject adjusted the same image in different ways, he or she also followed a similar scheme when choosing TMO parameters for all the images.

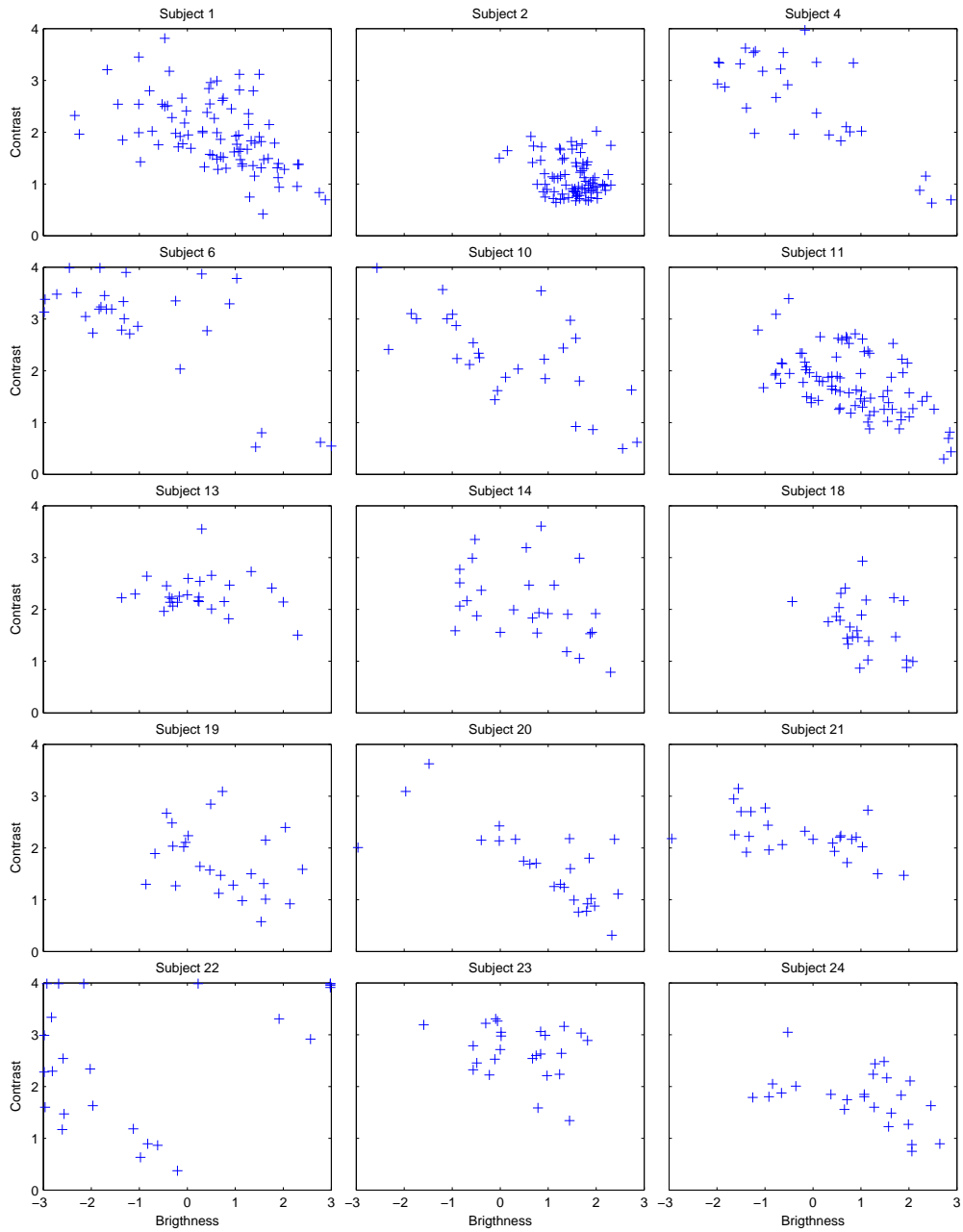


Figure 5.10: Brightness and contrast relations isolated for each subject. Each plot contains data for all images, both experiments and the full dynamic range of a display (1–3,000 cd/m^2).

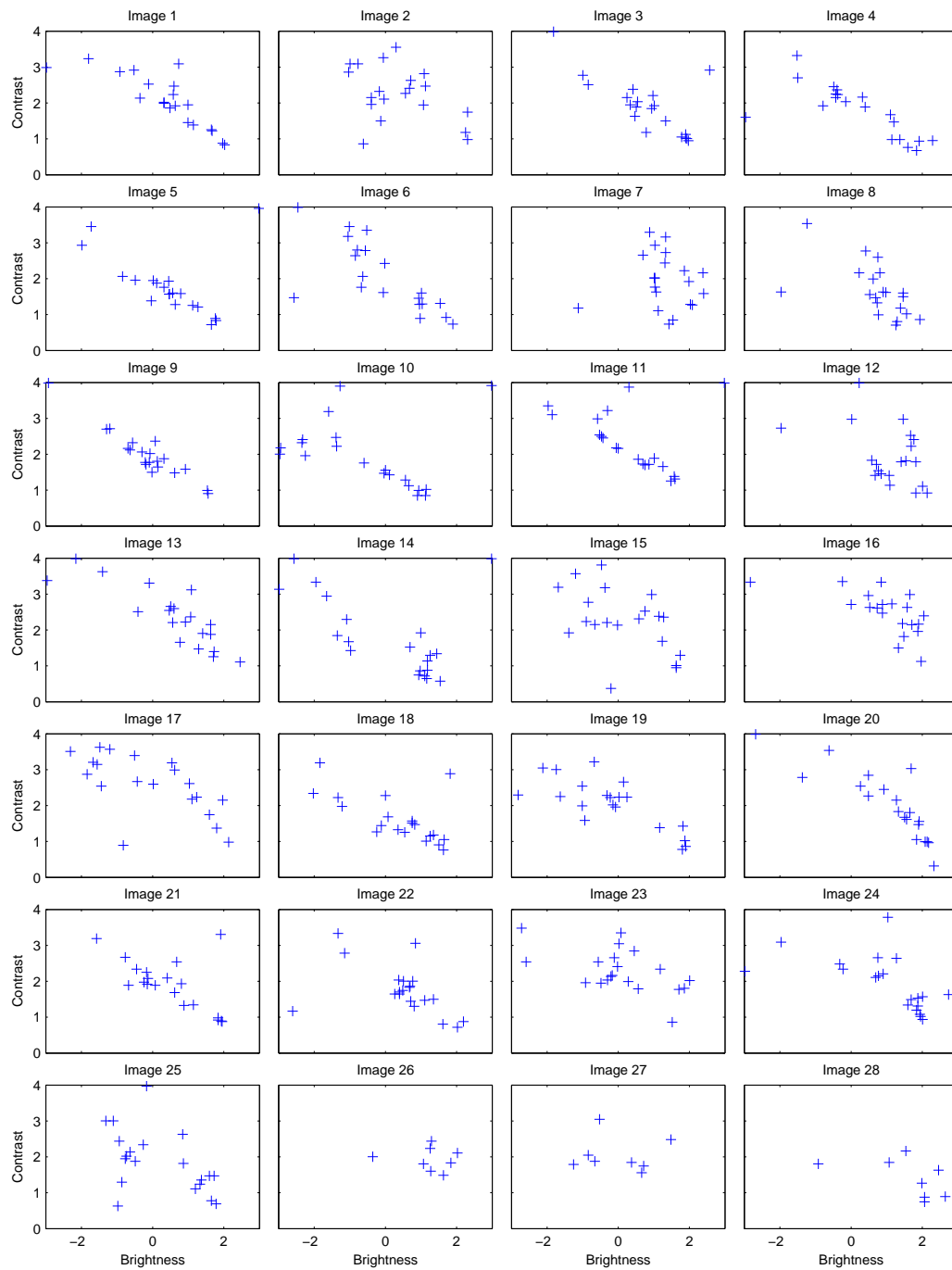


Figure 5.11: Brightness and contrast relations isolated for each image. Each plot contains data for all subjects, both experiments and the full dynamic range of a display ($1\text{--}3,000\text{ cd/m}^2$).

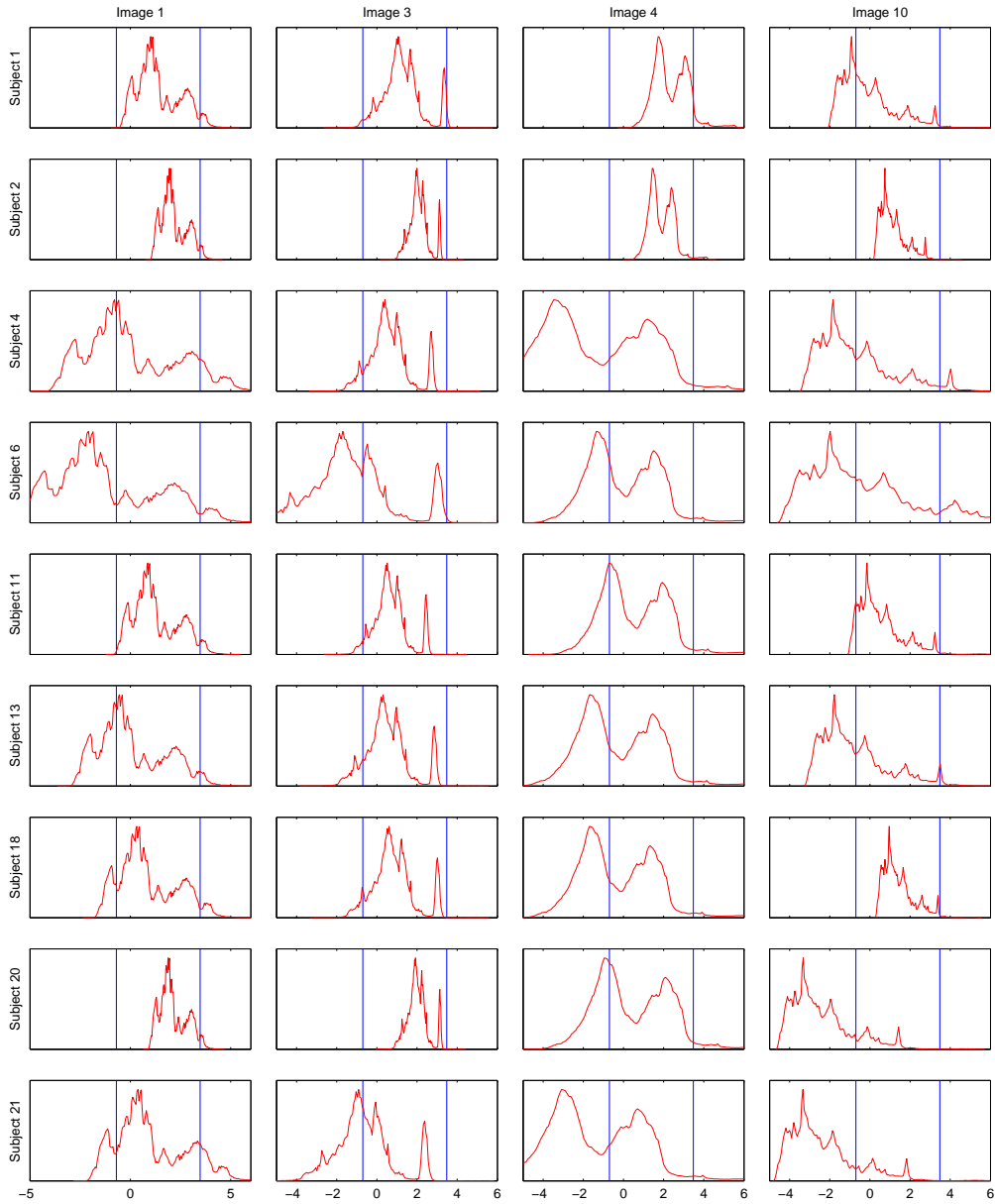


Figure 5.12: The histograms of the images after subject’s brightness and contrast adjustments. The blue vertical lines denote display minimum and maximum luminance. The horizontal axis is scaled in log luminance units. The subjects differ in their adjustments between each other, but a single subject follows a similar style for a range of images.

For example, the histograms resulting from the adjustments of Subject 6 are always more spread out than for the other subjects. This would indicate that the magnitude of contrast enhancement is correlated with the tastes of an individual. Additionally, there is another interesting observation which seems to be consistent across both all images and all subjects: the display maximum luminance, above which pixels are clipped, falls into approximately the same part of the histogram (see the blue vertical lines on the right of each plot in Figure 5.12). This indicates that people tend to “anchor” the brighter part of an image to the display maximum luminance, and then they extend or compress contrast in the direction of lower luminance to get the best looking image.

It is interesting to see whether the same observation can be generalized to a broad range of displays or if it is applicable only to an HDR display. We plot histograms in Figure 5.13 for a single subject and single tone mapped image but for several simulated displays of different brightness and dynamic ranges. The figure clearly indicates that subjects adjust images for the capabilities of a display, but they also follow the same scheme as for the HDR display ($0.2\text{--}3,000\text{ cd/m}^2$) — they map approximately the same part of the histogram to the maximum luminance of the display and then adjust contrast.

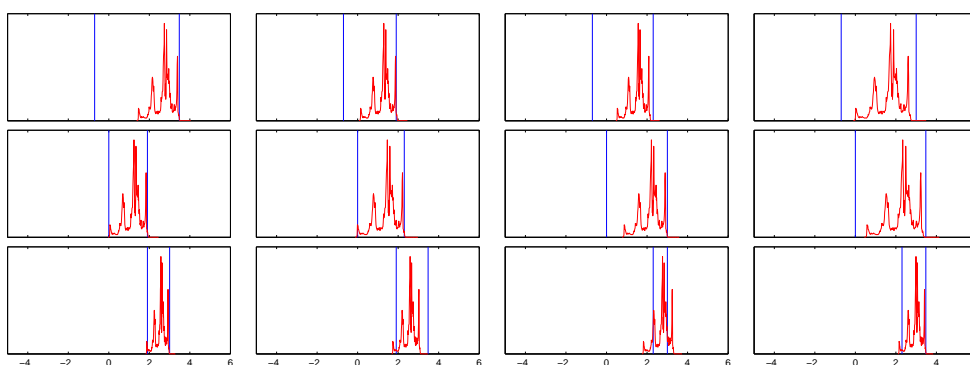


Figure 5.13: The histograms of Image 1 after Subject 1 brightness and contrast adjustments – data from Experiment 2. The blue vertical lines denote display minimum and maximum luminance. The horizontal axis is scaled in log luminance units.

5.5.2 Improved Tone Mapping Algorithm

The motivation for remodelling a TMO is to provide new parameters that would be more intuitive to use. As mentioned earlier, the settings for contrast and bright-

ness are strongly correlated. An average correlation coefficient for all images and both experiments is $\bar{R} = -0.7217 \pm 0.1622$. This suggests that both contrast and brightness could be replaced with parameters that do not exhibit such strong correlation and are therefore easier to control by the users. In the case of contrast and brightness, the 2D slider is usually adjusted along a slanted line (refer to Figure 5.9 center and right), which is neither intuitive nor convenient. A better user interface would use decorrelated parameters, so that the subjects could either use a simpler 1D sliders or move the 2D slider along the axes instead of a slanted line.

In Section 5.5.1, we analyzed and identified the strategy that the subjects use for adjusting TMO settings. Now, we show that this strategy can be modelled. We rewrite Equation (5.1) as

$$\log_{10} R' = c \cdot \log_{10}(R/Y_{max}) + \log_{10}(D_{max}) \quad (5.4)$$

where Y_{max} is the maximum luminance value in an image that we want to reproduce on a display, which we call “anchor white”. The same formula is used for the blue and green channels. The above equation mimics the operation performed by the subjects in our experiments. Firstly, the formula extends or compresses the image histogram by the scale factor c to the left side of the anchor white Y_{max} . Then, the anchor white is shifted to the display’s maximum luminance D_{max} . Note that we use the same contrast parameter c as in Equation (5.1), but we replace the brightness parameter b with the anchor white Y_{max} .

To better understand how Equation (5.4) relates to the original contrast and brightness parameters, we plot a function of c assuming constant Y_{max} . Firstly, we find the relation between b and D_{max} from Equations (5.1) and (5.4) as

$$b = \log_{10}(D_{max}) - c \cdot \log_{10}(Y_{max}). \quad (5.5)$$

Secondly, we choose two images (Images 1 and 10 in this example) and find the median percentage of the clipped pixels \bar{C} (see the third row of Figure 5.14) in order to compute Y_{max} :

$$Y_{max} = \text{percentile}(Y, 100 - \bar{C}) \quad (5.6)$$

where Y is a set of luminance (or luminance factor) values of the pixels in an image. Note that the above formula gives the location on the histogram for a given percentage of clipped pixels \bar{C} . We use the computed Y_{max} , the maximum luminance of the display $D_{max} = 3,000 \text{ cd/m}^2$ and Equation (5.5) to plot the function of c as a continuous magenta line in Figure 5.9 (center and right). The important observation is that the plotted functions for both images approximate well the correlation between contrast and brightness parameters. This indicates that the

largest variations between subjects in the resulting images are due to different selections of contrast parameter c while the anchor white Y_{max} does not vary much between subjects.

We intentionally named the parameter Y_{max} as “anchor white” to refer to the lightness perception theory [Gilchrist et al. 1999]. According to this theory, the human visual system assesses the lightness of an object based on the anchor luminance value, which acts as a reference for a white reflective surface. Such anchor luminance does not need to be the highest luminance in an image. This is especially true for the scenes that contain self-luminous surfaces, such as lights or the sun. The theory postulates that a “common denominator” for lightness estimation is a white reflectance, instead of gray, often used in photography. Our experiment confirms this since “anchoring” reflectance white to the maximum luminance of a display was a dominant strategy for adjusting the TMO settings.

The linear TMO we obtain in Equation (5.4) is easier to control than our original one, since both parameters of the contrast c and the clipping level Y_{max} modify independent aspects of image appearance. Moreover, if we transform Equation (5.4) from the logarithmic to the linear domain, we have the following formula:

$$R' = D_{max} \cdot (R/Y_{max})^c. \quad (5.7)$$

This re-parameterized form of the original TMO formula from Equation (5.1) is similar to a global contrast adjustment operation ², employed as a final-cut in many TMOs and as enhancement operation in image editing software. The importance of Equation (5.7) comes from the fact that we derived this formula based only on the analysis of the data we collected in our experiment without any prior assumptions on the parameters of the tone reproduction function. We have shown that the users try to adjust the TMO parameters along the parameter c , even if they have a non-standard user interface as used in our experiments. Moreover, we have shown that the same formula is valid for a broad range of display devices, ranging from dark CRT monitors to HDR displays.

5.5.3 Image and Subject Influence on TMO Parameters

We analyze how contrast, color saturation and the percentage of clipped pixels in dark and bright regions (dependent variables – DV) differ between subjects and

²The operation from Equation (5.7) is sometimes confusingly called *gamma-correction*. However, since the original meaning of *gamma-correction* denotes compensating the non-linearity of CRT monitors, using this term in the context of image enhancement may not be appropriate.

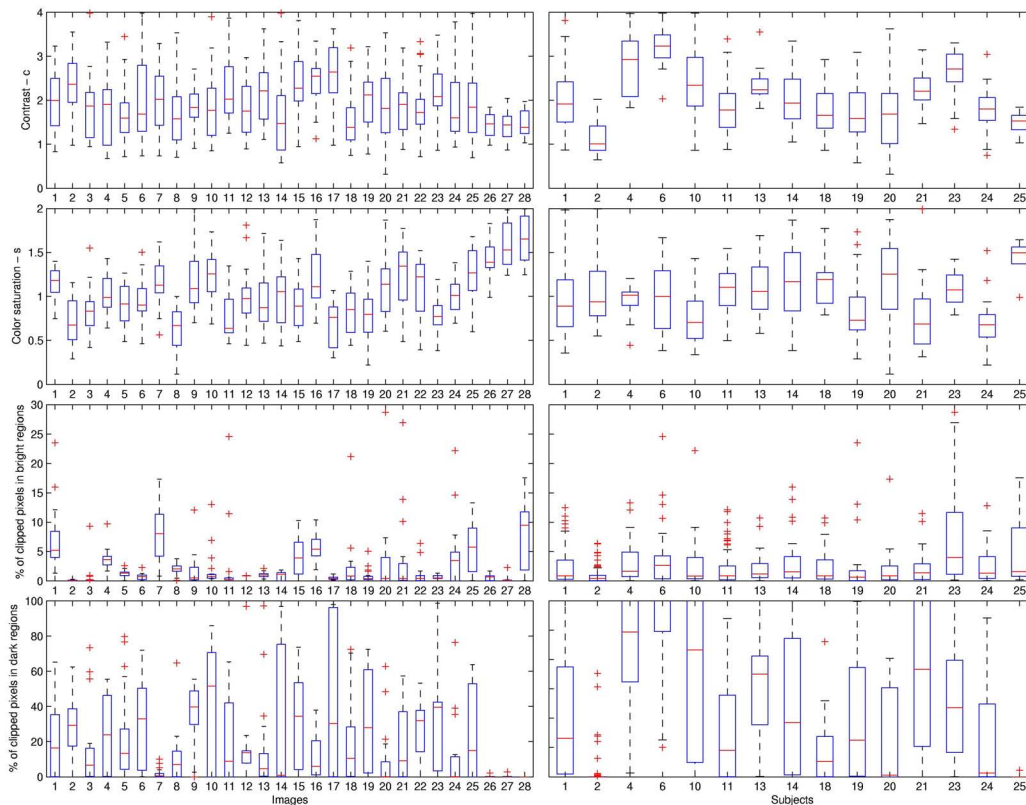


Figure 5.14: The variation of the DVs (contrast, color saturation and the percentage of clipped pixels in bright and dark regions) with respect to the IVs (images and subjects) – data from the experiments in which full dynamic range of the HDR display was used. Only 15 subjects participated in these experiments. Images are numbered as in Figure 5.1 – 5.4. The Notation: red lines – median; blue boxes – spanning from 25th to 75th percentile; whiskers – minimum and maximum values without outliers; red crosses – outliers.

images (independent variables – IV). We want to find out which of the two DVs is responsible for the large variance in the IV.

The variations of the DVs with respect to the IVs are summarized in Figure 5.14. Note that we do not include the brightness parameter in this figure. This is because brightness is strongly correlated with contrast and it is also fully determined by contrast and anchor white as described in Section 5.5.2. Anchor white, on the other hand, is related to the percentage of pixels clipped in bright regions (refer to Equation (5.6)).

From a first look at Figure 5.14 we can conclude that all four DVs are significantly different between images and subjects. This is confirmed by the two-way analysis of variance (ANOVA) for the main effects of the subjects and the images, which are run separately for each DV ($7.07 < F < 74.21$, $p < 0.001$). Contrast varies more between subjects than images (see the first row of Figure 5.14) and is probably determined mostly by subjects' personal tastes as discussed in Section 5.5.1. Color saturation and the percentage of clipped pixels in the dark regions (rows 2 and 4 in Figure 5.14) do not show any consistency between the subjects and the images and therefore it is not possible to draw any conclusion for these parameters. The third and the fourth rows of Figure 5.14 (note the difference in the scale used for these plots) show that there are significantly more pixels saturated in dark regions than in bright regions. This suggests that the subjects prefer sacrificing a significant portion of the dark part of an HDR image, probably in order to improve contrast. The same tendency can be observed in Figure 5.12, which shows that Subjects 6 and 13 decided to push a large part of the histogram below the minimum luminance of the display, while preserving the brightest pixels. This observation suggests that TMOs should follow a similar pattern and saturate more pixels in the dark regions. This is contrary to the most common approach employed in many TMOs where the same number of the darkest and brightest pixels are clipped. Such TMOs do not produce the best results if they do not provide an adjustment for the number of pixels clipped in dark regions.

Conclusions on the measured values of clipped pixels in bright regions can be drawn directly from the actual images. We observed that the most pixels are clipped for the images that contain large bright objects which should appear self-luminous in the reproduction, like the sky in Images 1, 7, 16, 25 and 28, or the sun in Image 24 (refer to Figures 5.1 – 5.3 for images and Figure 5.14 for the magnitude of clipping). Then, follow the images that contain small self-luminous objects, such as Christmas lights in Image 15 and the images that depict dark scenes without self-luminous objects (Images 12 and 27). There is also less clipping for the images of low dynamic range (Images 3, 11 and 20).

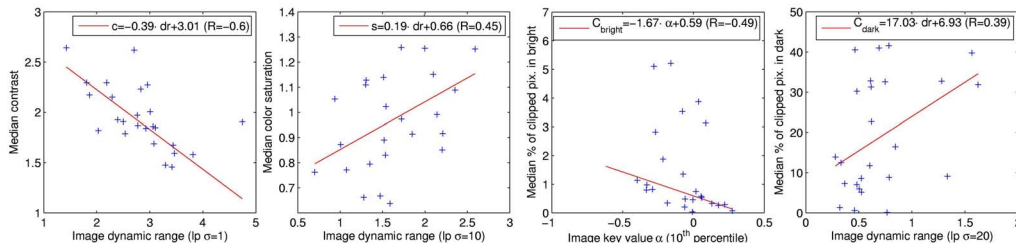


Figure 5.15: The prediction of tone mapping parameters (contrast, color saturation, the percent of clipped pixels in bright and dark regions) based on image statistics. Plot legends: robust fits of the linear model and correlation coefficients.

5.5.4 Choosing Default TMO Parameters

User adjusted TMO parameters are not desirable in many applications and it would be helpful if their values could be automatically found at least to render a “best guess” image. We want to check if there is any correlation between the DVs (the TMO parameters) and the IVs, so that, for example, an image characteristic can predict the values of contrast and the percentage of clipped pixels in the bright regions (needed to compute anchor white Y_{max}). If we find such a correlation, we can propose a method to automatically choose TMO parameters.

Although an algorithm cannot predict a user’s tastes, it may be possible to guess some TMO parameters based on the characteristics of an image. To verify this hypothesis, we compute a set of variables characterizing each HDR image; the dynamic range of an image, which is a difference between the logarithm of the highest and the lowest luminance in an image; the key value, α ; and L_{white} . α and L_{white} are used for the automatic parameter estimation in the photographic TMO [Reinhard 2003]. All these variables require the value of the maximum and the minimum luminance in an image, which can be calculated in a variety of ways. We compute the minimum and the maximum as percentiles: 0.01, 0.1, 1, 10, 20, 30 (of brightest and darkest pixels), and as the minimum or maximum value of a low-pass filtered image, where the filter is the Gaussian with different values of σ (1, 2, 5, 10, and 20). This gives in total 30 different variables that could characterize an image (3 variables times 10 estimates of the minimum and the maximum luminance).

We compute the correlation matrix to check if there is a correlation between any of the computed 30 variables and the median values of the TMO parameters for the subjects: contrast, color saturation and the percentage of clipped pixels. We use this matrix to find the variable that is the most correlating with each TMO

parameter. The relations between the most correlated variables and the TMO parameters, together with the results of the robust linear regression, are plotted in Figure 5.15. The highest correlation is found for the dynamic range of an image computed using low-pass filtered images ($\sigma = 1$) and the contrast parameter (the first plot in Figure 5.15). The negative slope of this relationship is intuitive — the images of higher dynamic range must be stronger compressed to fit into the dynamic range of a display. A weaker correlation and less intuitive relation is found for color saturation and the percentage of clipped pixels in the dark regions. These TMO parameters probably cannot be predicted using the given set of image characteristic variables. The prediction of the percentage of pixels clipped in bright regions is more reliable. It correlates with the image key coefficient α (computed using the 10-th percentile for the minimum and the maximum values in an HDR image). We observed that this prediction is less accurate for the images that contain large self-luminous objects.

The plots in Figure 5.15 show that both the contrast parameter and the number of clipped pixels in the bright regions are correlated with image content, and therefore they can be predicted. Such predictions can be used for parameter estimation in TMOs. Although the predicted values will not be optimal for many images and subjects, they could be used as the “best guess” for the TMO parameter setting. Our experiments did not include a sufficient number of images and subjects to build a reliable model for such a parameter estimation, but they proved that such estimation is possible and can be an interesting direction for further research.

5.5.5 Influence of a Display

It is interesting to know how the dynamic range and brightness of a display influences the parameters of a TMO. Figure 5.16 illustrates how the contrast setting increases as the dynamic range of a display increases. If the dynamic range of a display is too low, the subjects compress contrast. On the other hand, they expand contrast even above the contrast of an original image ($c > 1$) when a display offers higher dynamic range. However, this behavior differs slightly between both experiments: if the subjects adjust the HDR images to their preference, they enhance contrast proportionally to the dynamic range of a display (Figure 5.16 top), but if their goal is to achieve the fidelity to the real-world scene, they adjust contrast slightly above 1.0 and keep it approximately on the same level even for the HDR displays (Figure 5.16 bottom). This suggests that the TMO profiled for fidelity should not enhance contrast above the contrast of an original scene, and the TMO profiled for preference should take full advantage of the display dynamic range.

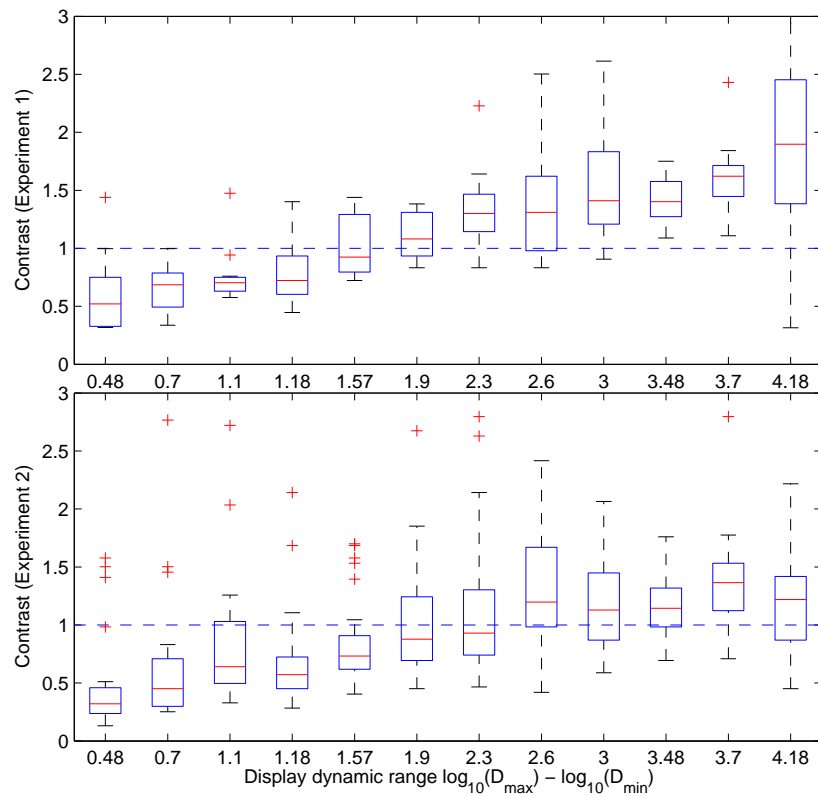


Figure 5.16: The relation between the dynamic range of a simulated display and the contrast parameters. The relation is plotted for the data from Experiment 1 (preference) at the top and Experiment 2 (fidelity) at the bottom. The subjects tend to enhance contrast more if their goal is the most preferred image. Notation is the same as in Figure 5.14.

5.5.6 Preferred Display Device

We examine how the minimum and maximum luminance values of a display can affect subjective preferences for displayed images. We use the data from the questionnaire used in Experiment 2. The preference scores for each simulated display are averaged over three scenes and over all subjects. Figure 5.17 illustrates the ranking of potential display devices simulated on the HDR display. The figure clearly shows that the subjects prefer brighter displays of higher dynamic range. A typical LCD display ($1 - 200 \text{ cd/m}^2$) is in the middle of the preference scale. Interestingly, the brighter display but of lower dynamic range ($80 - 3,000 \text{ cd/m}^2$) has higher preference score than the typical LCD. The displays of the broadest dynamic range top the ranking, but the broadest dynamic range display ($0.1 - 3,000 \text{ cd/m}^2$) comes unexpectedly lower than the $1 - 3,000 \text{ cd/m}^2$ model. However, the rankings in the top group ($1 - 3,000$, $0.2 - 1,000$, $1 - 1,000$, and $0.2 - 3,000 \text{ cd/m}^2$ models) are not significantly different from each other ($F(3, 126) = 0.82$, $p > 0.05$). The high scores for the brightest displays of the highest dynamic range indicate that both high luminance and high contrast are important for reproducing digital images.

To better understand the relation between the minimum and the maximum luminance of a display and the preference score, we fit the data to the linear model using multiple linear regression. The averaged preference score S is given by

$$S = \alpha \cdot \log_{10}(D_{min}) + \beta \cdot \log_{10}(D_{max}) - \gamma \quad (5.8)$$

where D_{min} and D_{max} are the display minimum and maximum luminance values, $\alpha = -0.47 (\pm 0.05)$, $\beta = 0.87 (\pm 0.11)$ and $\gamma = 0.25 (\pm 0.31)$. The model accounts for nearly 60% of the data ($R^2 = 0.57$). The negative α indicated that “darker” displays are more preferred (i.e., lower minimum luminance) and positive β indicates that also “brighter” displays (i.e., higher maximum luminance) are preferred. However, since the trend is stronger for β , we can assume that the maximum luminance is more important than the minimum luminance of a display. The percentages of oversaturated pixels are far smaller than that of undersaturated pixels (see the third and fourth rows in Figure 5.14). This indicates that people are more sensitive for oversaturation than undersaturation since they carefully avoided oversaturated pixels but did not pay much attention to undersaturated pixels compared to oversaturation.

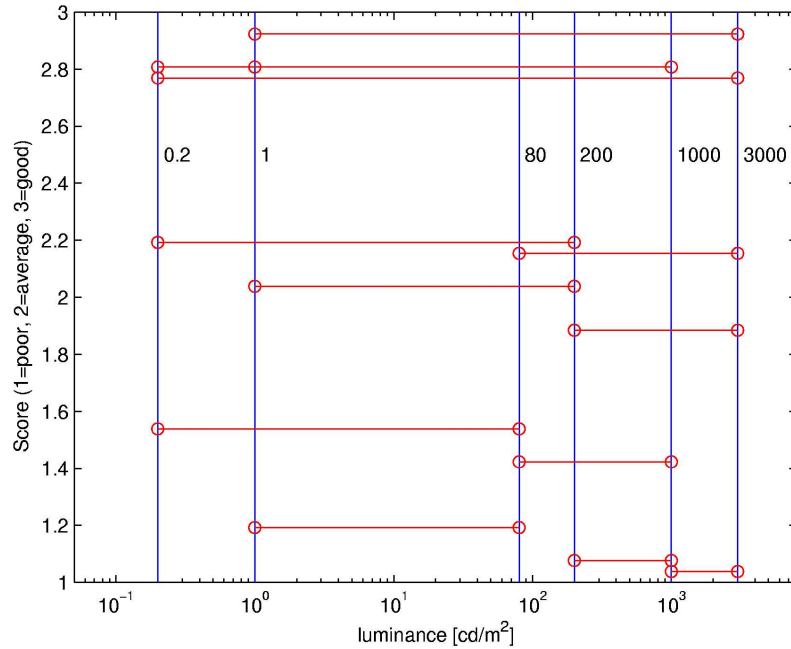


Figure 5.17: The preference of several simulated monitors, which differ in the minimum and the maximum luminance they can display. The preference score is indicated on the y-axis, and the left and right end points of the horizontal lines represent the minimum and maximum luminance of the display. The scores of the displays are shifted by a small random offset to avoid overlapping of the lines.

5.6 Summary

The major outcome of this work is a better understanding of how users adjust tone mapping operator (TMO) parameters to achieve either the best looking images (preference task) or the images that are the closest to real-world scenes (fidelity task). Based on this knowledge, we propose a better parameterization of a linear TMO in logarithmic domain, in which parameters are more intuitive and can be partly estimated from image characteristics. The TMO is controlled by two parameters: *anchor white* and *contrast*. The *anchor white* parameter is approximately consistent across subjects and depends on images — it is set to a lower value if an image contains large self-luminous objects. The *contrast* parameter is more subjective, and therefore users should be allowed to adjust it. We have shown that the parameters can be automatically estimated for a TMO based on an image characteristic to obtain a “best guess” result. The *contrast* parameter

can be predicted from the dynamic range of an image (images of higher dynamic range must be reproduced with lower contrast), and the *anchor white* parameter is related to the image key value (although the prediction can be unreliable if an image contains large self-luminous objects). We believe that the results of our analysis are also applicable to complex TMO, which can benefit from both a better selection of user adjusted parameters and an automatic parameter estimation.

The second main subject of this work is an investigation how the dynamic range and brightness of a display affects the preference for tone reproduction. For 14 simulated monitors of varying brightness and dynamic range we do not find any major difference in the strategy the subjects use to adjust images for LDR and HDR displays. We notice however that the resulting images depend on a given task. If the goal is to find the best looking image (preference), subjects tend to strongly enhance contrast (up to four times that of the original image contrast), even at the cost of clipping a large portion of the darkest pixels. On the other hand, when the task is to achieve the best fidelity with a real-world scene, the subjects avoid clipping both in the dark and bright parts of an image and they do not extend contrast much above the contrast of an original image. In both tasks, there is a tendency towards brighter images, which are achieved by over-saturating the brightest pixels belonging to self-luminous objects. The final investigation compares user's preference for displays of varying capabilities. The subjects prefer in the first order the displays that are bright, and in the second order, the displays that have low minimum luminance.

Chapter 6

Perception-Based Contrast Enhancement Model for Complex Images in HDR

6.1 Introduction

In the recent years, we witness significant progress in the display technology in terms of expanded color gamut, luminance dynamic range, and physical contrast. For example, specialized HDR displays can reproduce luminance levels ranging from 0.015 to 3,000 cd/m^2 , but even modern LCD TV sets feature remarkable luminance ranges of 0.1 – 800 cd/m^2 [Seetzen et al. 2004]. This results in much better visibility of details in deep shadows and bright highlights; it makes the reproduced images more plausible with respect to the real-world observation conditions. In particular, the black level in such displays guarantees that the darkest image regions appear black in contrast to the grey appearance of such regions on older displays with the minimum luminance higher than 2 – 5 cd/m^2 .

The dynamic range and contrast expansion of display devices require revisiting well-established image processing techniques which are often tailored for 8-bit color depths and luminance ranges typical for the once prevailing CRT displays. For example, image contrast manipulation is often based on the assumption of contrast constancy, i.e. invariance of perceived contrast over variations of display dynamic range. However, as increasing possible dynamic range of displays, the need of studying HVS for the luminance levels which used to be non-covered by

display devices is highly required. Additionally, the characteristics of HVS have been researched on simple stimuli such as sinewave gratings and Gabor patches but not on complex images.

In this work, we consider this problem in the context of complex images and for luminance ranges typical for HDR displays. Our goal is to derive a model relating physical and apparent contrast, which can be applied to improve visual uniformity of contrast changes resulting from image contrast manipulation. The main contribution of this research is to parameterize c in such a way, that a specified contrast change is perceived as a uniform modification of the image independently of luminance levels and contrasts existing in the given local area. Furthermore, we apply the parameterized model for arbitrary images in order to generate a contrast-enhanced version of them [Yoshida et al. 2007b, Yoshida et al. 2008b].

In the following sections, we first conduct perceptual experiments to establish the relation between physical and apparent contrast changes in a complex image in Chapter 6.4. Then, we derive a model encapsulating this relation, discuss the observed relations, and propose a method for perceptually uniform contrast scaling in images displayed over high dynamic range in Chapter 6.6. We conclude the paper and outline future work in Section 6.9.

6.2 Previous Work

A number of research have been proposed on contrast detection and discrimination for simple patterns (see Section 2.3.2 for contrast detection and discrimination thresholds). Legge proposed power laws for increment contrast discrimination threshold with exponents 0.6 – 0.7 by conducting psychophysical experiments on sine wave gratings stimuli at 2 and 8 cycles per degree [Legge 1980]. Whittle's law does not hold for contrast discrimination under any of their cases. Based on this work, they also presented a contrast *transducer function* which models the behavior of HVS for a given physical contrast [Legge and Foley 1980].

Two sets of psychophysical experiments on contrast were conducted in parallel. Foley and Legge conducted another forced-choice experiment to determine contrast detection and near-threshold discrimination thresholds for sine-wave gratings at 0.5, 2, and 8 cycles per degree [Foley and Legge 1981]. S-shaped models were presented for detection threshold while discrimination thresholds were formed linearly. Gottesman et al. employed magnitude estimation method for measuring perceived contrast on sine-wave gratings [Gottesman et al. 1981]. Their results proposed a power function with its exponent 0.7 which does not change accord-

ing to both luminance levels and spatial frequencies of stimuli.

Whittle measured contrast discrimination thresholds on luminance for both increment and decrement with the stimulus of 1 visual degree square and then discovered that threshold is proportional to pedestal contrast [Whittle 1986]. Based on Whittle's measurement, Mantiuk et al. proposed a transducer function [Mantiuk et al. 2006]. Whittle's law only covers the cases that contrast is smaller than 50%. Kingdom and Whittle extended this work by conducting experiments for the cases of contrasts greater than 50% and proposed U-shaped model of contrast discrimination [Kingdom and Whittle 1996].

Peli et al. investigated the contrast constancy problem for various luminance adaptation values and simple stimuli such as the Gabor patches imposed on background with different mean luminance [Peli et al. 1991]. In two independent contrast matching and contrast magnitude estimation studies, they confirmed that contrast sensitivity is significantly reduced for low luminance adaptation values below 3 – 8 cd/m^2 . The lower the physical contrast of the Gabor patches, the stronger the sensitivity reduction observed, with a typical contrast versus intensity (cvi) characteristic observed for near threshold contrast values. Effectively, this means that, on modern displays, simple contrast rescaling may lead to image distortions manifesting in changing apparent contrast relations with respect to the original image through weakening perceived contrast in dark image regions. This work was extended by including other factors such as stimulus size, dichoptic presentation, and length of adaptation [Peli 1995] and spatial frequency [Peli et al. 1996].

6.3 Contrast in Complex Images

As we reviewed in Chapter 2.3.1, contrast is the relationship between the luminance values at the peaks (L_{max}) and that at troughs (L_{min}), and there are a number of definitions to represent contrast in literature such as simple contrast, SNR, Weber's fraction and Michelson contrast. However, these definitions can be applied for simple patterns such as Gabor patch or sinewave gratings. For complex images, Peli proposed a definition of local band-limited contrast (refer to Chapter 2.3.1) [Peli 1990]. This method employs the *contrast sensitivity function* (CSF) and represents a contrast value at every pixel of an image as a function of the spatial frequency band. Since CSF works at threshold and near threshold, Peli's contrast definition cannot be directly applied for the cases of suprathreshold contrast.

No method has been developed to represent contrast in a complex image by one

numerical number not only for threshold and near-threshold contrast but also for suprathreshold contrast. We investigate the standard equation for contrast scaling in image processing [Pratt 1991]:

$$L(c) = \bar{L} \left(\frac{L}{\bar{L}} \right)^c \tag{6.1}$$

where L denotes the luminance of a pixel, \bar{L} is a luminance reference, and c denotes the *contrast factor*. The luminance reference \bar{L} defines the brightness level which remains unchanged during contrast scaling and usually equals the minimum or maximum luminance in an image, what gives normalized base in Equation (6.1). To test the perception of contrast scaling in areas of different luminance, we set the \bar{L} value to the mean luminance in the analyzed area. The *contrast factor* defines physical change to contrast in such a sense that a value of $c = 2$ increases while $c = 1/2$ decreases the physical contrast twice in logarithmic domain (see Figure 6.1 for examples). Furthermore, the *contrast factor* is a relative measure of contrast which is convenient to use and interpret within the scope of presented applications. It also allows to analyze the contrast change in terms of one number without measuring actual contrasts, which is particularly important since a single number physical contrast measure for complex images is difficult to be quantized.

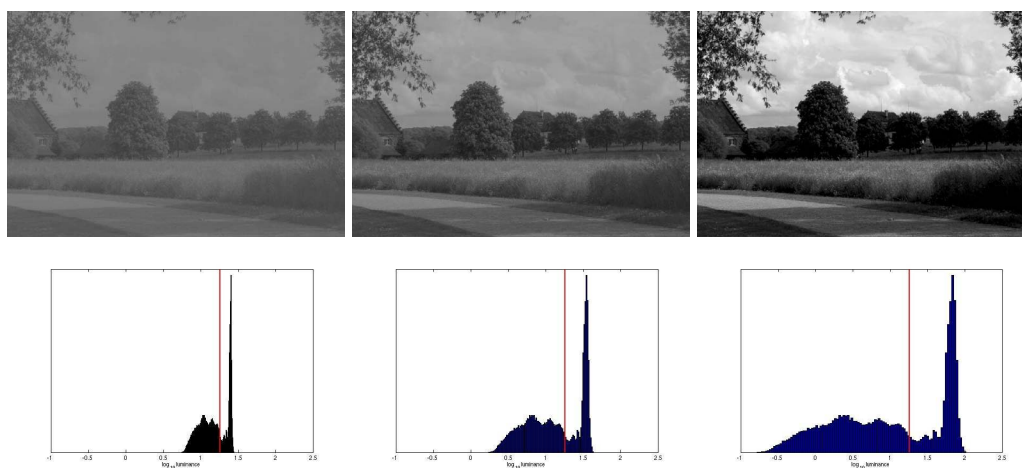


Figure 6.1: Examples of changing contrast factors $c = 0.5, 1.0,$ and 2.0 from left to right respectively. The histograms are shown in logarithmic domain with the red bars indicating the mean luminance of an image \bar{L} .

6.4 Experiments on Measuring Contrast Scaling of a Complex Image in JND

6.4.1 Introduction

We conducted two psychophysical experiments, a *contrast scaling* and a *contrast discrimination threshold* tasks, to assess how the *Human Visual Systems (HVS)* perceive physical contrast changes. The goal of the *contrast scaling* experiment is to obtain uniform scalings of perceived contrast for the human observers with respect to given physical contrast for various luminance adaptation conditions. In this experiment, we employed a *Two-Alternative Forced Choice (2AFC)* procedure for image pairs with different *contrast factors* and the same luminance levels and analyzed the obtained data using Thurstone’s Law of Comparative Judgment for *contrast scaling* experiment (refer to Section 2.5.1 for 2AFC and Thurstone’s law) [[Thurstone 1927](#)].

Thurstone’s Law of Comparative Judgment gives arbitrary uniform scaling for each set of stimuli at different luminance levels. We can compare distances between stimuli, i.e., perceived contrast magnitude, within the same set but cannot compare different sets of stimuli to each other. For rescaling the results of Thurstone’s scaling to a contrast space compatible for all stimuli sets, a *contrast discrimination threshold* experiment was conducted using the *Parameter Estimation by Sequential Testing (PEST)* (refer to Section 2.5.2 for PEST) [[Taylor and Creelman 1967](#)].

In this experiment, each subject was shown pairs of stimuli. One pair of stimuli contains reference and target images shown one after another randomly, and we asked a subject to report if they saw any difference between given two images. The details for both experiments are described in the following sections.

6.4.2 Stimuli and Apparatus

We selected a black-and-white image of the resolution 900×600 (see Figure 6.2). This is a typical landscape image with luminance and contrast patterns which we can observe in natural images. This image was segmented based on luminance levels into three different regions: “dark”, “medium”, and “bright” and our experiments were conducted on two displays: the Westinghouse high resolution digital television (HDTV) and the BrightSide DR37-P HDR display [[Seetzen et al. 2004](#)]. We used the Westinghouse display, one of the commercial LCD TVs, because it



(a) “Very dark” and “dark”.

(b) “Medium”.

(c) “Bright”.

Figure 6.2: Our test image (top) and its masks (Bottom). The average luminance levels are 0.3, 4.5, 28.8, and 158.5 cd/m^2 for “very dark”, “dark”, “medium”, and “bright” regions respectively.

has better uniformity of its back-light but obviously can not reproduce very low luminance levels. Therefore, we also employed the BrightSide HDR display which makes it possible to reproduce very low luminance levels by spatially varying LED-based dimming technology. Both displays use the same LCD sandwich type and were carefully calibrated by measuring its luminance response for a range of input values using the MINOLTA LS-100 light meter¹. Except their reproducible dynamic ranges, both displays have similar characteristic.

In order to reproduce very low luminance level, we uniformly reduced the power of LED back-lights of the BrightSide HDR display, and the former “dark” became “very dark” region. The mean luminance levels are 0.3, 4.5, 28.8, and 158.5 cd/m^2 for “very dark”, “dark”, “medium”, and “bright” areas, respectively.

Each display was placed approximately 1.5 times of its diagonal size away from a participant and viewed binocularly for both experiments. All experimental sessions were conducted in a room whose lighting condition is fully controllable and under dim illumination (65 lux).

6.4.3 Experiment I: Contrast Scaling

Contrast scaling experiment was conducted for estimating perceived contrast at physical contrast change at different luminance levels. We employed a 2AFC analyzed by Thurstone’s Law of Comparative Judgment [Thurstone 1927] which are commonly used for measuring distances between stimuli in uniform continuous scaling.

In each trial of *contrast scaling* experiment, a pair of stimuli was displayed next to each other randomly and the region of interest was specified through colored contours (see Figure 6.3). In each stimulus, a different *contrast factor* has been applied only to the selected image region. The other regions in an image are present but slightly blurred (Gaussian blur, $\sigma = 10$) not only to maintain similar local luminance adaptation in an image but also to reduce subjects distraction to non-selected areas. Subjects were asked to switch the contour off and judge in which image they were able to see more contrast in the specified areas. Every participant took approximately 20 – 30 minutes to complete this experiment.

Before the main part of the experiment, we conducted a pilot study to prepare a set of stimuli so that contrast differences are right below the visibility threshold. We prepared several different sets of stimuli in the form of $c = 1.11^n$, $c = 1.13^n$, and $c = 1.15^n$ where $n = -5, -4, \dots, 5$ and selected $c = 1.13^n$. Since 2AFC

¹ www.konicaminolta.com/instruments/products/light/luminance-meter/ls100-ls110/index.html

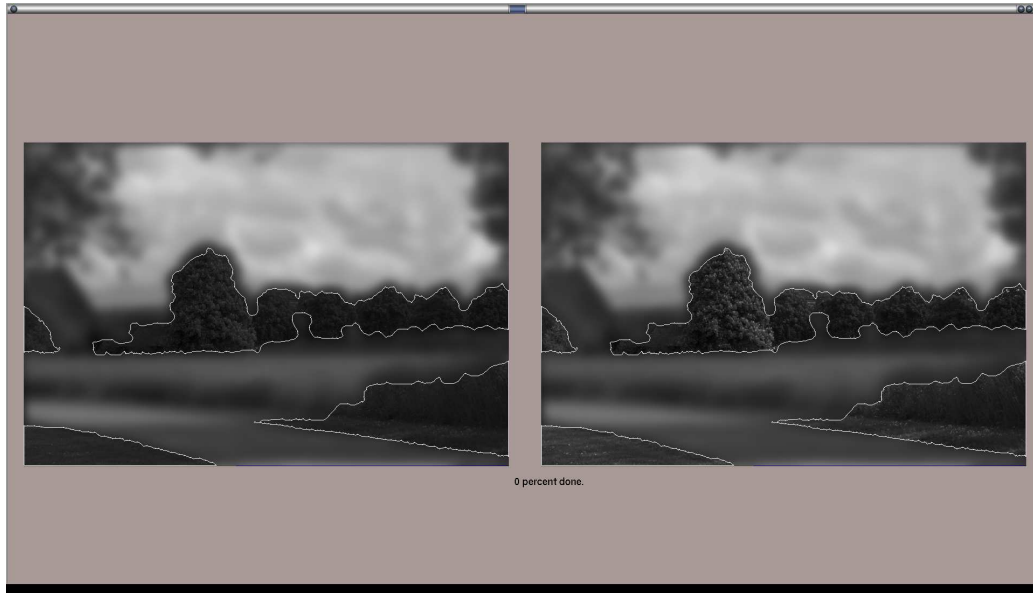


Figure 6.3: A screenshot of contrast scaling experiment. The selected areas are surrounded by colored contours to let a subject know to which areas they have to pay attention. The rest of an image is blurred to reduce a subject’s distraction and to maintain luminance local adaptation.

increases the number of trials extremely, we used only one image for our experiments. Although we used only one image for our experiments, we still had 220 pairs to compare, which is too many to judge for subjects. Therefore, we removed 68 obvious pairs and conducted the experiments comparing 152 pairs (see [Torgerson 1958] for details how to reduce experimental labor).

11 subjects between 28 – 47 years old (31 in average) participated in this experiment. Four of them were female and the rest were male. Every participant reported normal or corrected to normal vision, and everybody was naïve for the goal of the experiment.

6.4.4 Experiment II: Contrast Discrimination Threshold

Another subjective experiment was conducted for measuring *contrast discrimination thresholds* so that we can rescale the results of *contrast scaling* experiment from arbitrary units to *Just Noticeable Difference (JND)* unit. We employed the PEST [Taylor and Creelman 1967] at three reference points of *contrast factors* ($c = 0.69, 1.00, 1.44$) for all four regions. At each reference contrast, its target

contrast was started at significantly different point. One of the reference and target images was shown with colored contour surrounding the selected areas, the contour disappeared, and then another image was shown. A subject was allowed to repeat displaying each trial as many times as they wanted. In this experiment, the task of a subject was to report if there was visible difference between two images in a specified region. One trial was ended when the recent five thresholds were constant enough, i.e, a trial is finished if the standard deviation of the recent five thresholds became below 0.05.

Six people participated in the *discrimination threshold* experiment, which took 20–30 minutes for each subject. Everybody had participated in the *contrast scaling* experiment first, because we were interested in measuring *contrast discrimination threshold* for the same series of images as for the *contrast scaling* experiment.

6.5 Results

6.5.1 Experiment I: Contrast Scaling

Labels	cf01	cf02	cf03	cf04	cf05	cf06	cf07	cf08	cf09	cf10	cf11
c	0.54	0.61	0.69	0.78	0.89	1.00	1.13	1.28	1.44	1.63	1.84
Very dark	0	0.02	0.33	0.97	1.15	1.34	1.92	2.39	2.94	3.31	4.09
Dark	0	0.05	0.79	1.01	1.47	1.96	2.25	2.99	2.80	3.64	4.10
Medium	0	0.04	0.72	1.48	1.67	2.46	3.04	3.58	4.12	4.36	4.38
Bright	0	0.66	0.67	1.55	1.86	2.25	2.65	3.38	4.30	4.33	5.08

Table 6.1: Results of contrast scaling experiment analyzed by Thurstone’s Law of Comparative Judgment. The labels for contrast factors correspond to those in Figure 6.4.

Two-alternative forced choice (2AFC) was employed in this experiment. A set of 11 stimuli were compared in a pair and then analyzed by using Thurstone’s law of comparative judgment [Thurstone 1927]. The results of Thurstone’s scaling are shown in Figure 6.4 and Table 6.1. Thurstone’s scaling of 2AFC experiment is very simple and intuitive analysis, however, it returns a uni-dimension scaling in arbitrary unit. Therefore, if we analyze data for different luminance levels separately, we can not compare the results of Thurstone’s scaling to each other. To convert the results into a meaningful scale, we conducted the second experiment, *contrast discrimination threshold*.

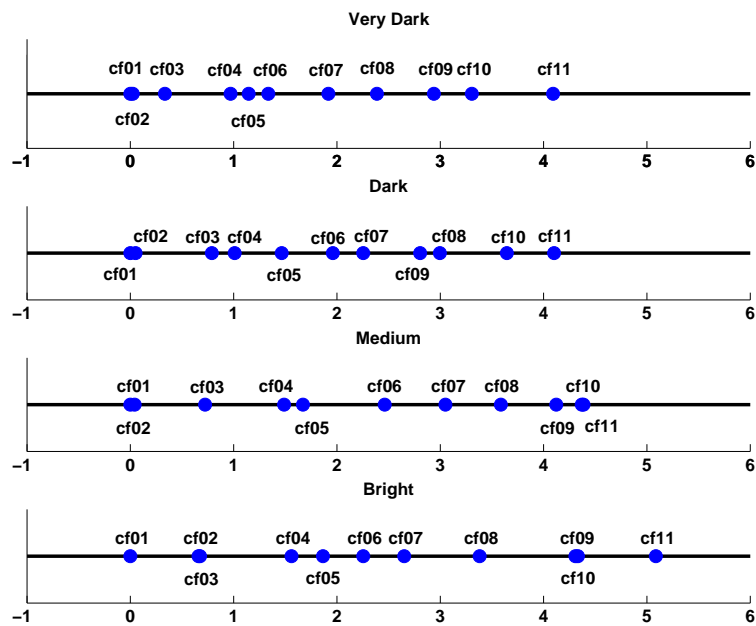


Figure 6.4: Results of contrast scaling experiment analyzed by Thurstone's Law of Comparative Judgment. The labels $cf01, \dots, cf11$ denote the contrast factors ordered from the smallest to the biggest values (see Table 6.1 for the details). Note that we can not compare them directly to each other because they are given in arbitrary units. We have to rescale them to JNDs by using the results of *contrast discrimination threshold* experiment (see Section 6.6).

6.5.2 Experiment II: Contrast Discrimination Threshold

After the *contrast scaling* experiment, there was a need to measure The results of the *discrimination threshold* experiment for contrast increments are shown in Table 6.2. Inter-observer variability was tested by one-way analysis of variance (ANOVA) before calculating *contrast discrimination threshold* in order to remove outliers. There were a few cases with outliers, but after removing them, all p -values are much higher than the significant level (0.05), i.e., they statistically behaved in the same way.

Reference <i>contrast factors</i>	$c = 0.69$	$c = 1.00$	$c = 1.44$
Very dark	0.14	0.14	0.16
Dark	0.09	0.09	0.07
Medium	0.07	0.07	0.07
Bright	0.09	0.08	0.10

Table 6.2: Contrast discrimination thresholds Δc at three reference contrast factors as measured for contrast increments.

6.6 A Model of Uniform Contrast Enhancement for Complex Images

In this section, we derive a model which adjusts the *contrast factor* for a desired perceptual contrast change as a function of luminance level. The results of *contrast scaling* experiment (Figure 6.4) are rescaled to just noticeable difference (JND) units by using the results of *contrast discrimination threshold* experiment (Table 6.2) using the following procedure:

1. Setting the origins to the *contrast detection thresholds* computed by *contrast sensitivity function* for each luminance level.
2. Rescaling the outcome of the *contrast scaling* experiment to match the result of the *contrast discrimination threshold* experiment. The distance between reference contrast and the threshold obtained by the *contrast discrimination threshold* experiment is considered as 1 JND.
3. Fitting the points obtained in Step 2 to power functions. Note that every point is rescaled in *absolute* JND units now. For practical use, we simply change the *absolute* JNDs to *relative* JNDs by setting the point of “medium” curve at $c = 1.0$ to 0 JND for *relative perceived contrast* (see Figure 6.5).

The coefficients of the power function $\alpha c^\beta + \gamma$, where c is the *contrast factor* for each luminance level, are given in Table 6.3. All R-square values of power fittings are above 0.93 for our data.

- Interpolating the curves in Figure 6.5 to construct a surface model with parameters of mean luminance level, *contrast factor*, and relative perceived contrast in JND units (see Figure 6.6). Cubic interpolation is employed.

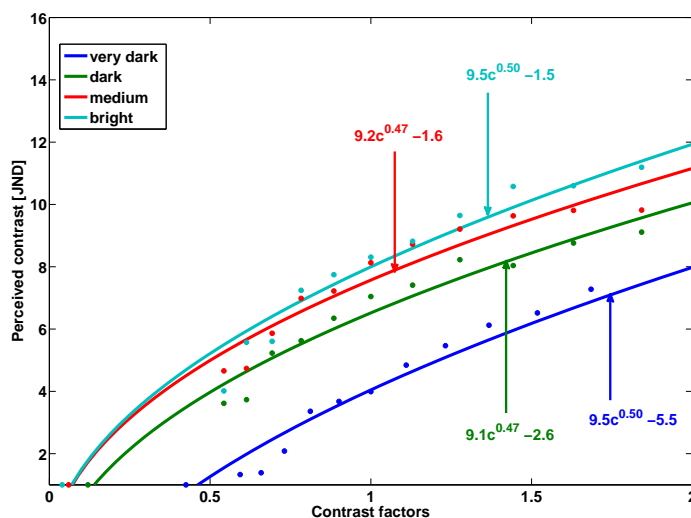


Figure 6.5: Relative perceived contrast in JNDs at different luminance levels for given contrast factors. Dots represent the rescaled data for each corresponding image region. Coefficients of each curve are shown in Table 6.3.

	α	β	γ
Very dark	9.50	0.45	-5.5
Dark	9.09	0.47	-2.6
Medium	9.21	0.47	-1.6
Highlight	9.50	0.50	-1.5

Table 6.3: Coefficients for the power function $\alpha c^\beta + \gamma$ in Section 6.6 for measured luminance levels. See also the plots in Figure 6.5.

After all steps shown above, we derive the following formula for relative perceived contrast C_p :

$$C_p(c, L) = 9.3c^{0.47} + \gamma(L) \tag{6.2}$$

where c is given *contrast factor* and L is the logarithm of mean luminance of a segmented region. The values of α and β are computed as average values in

Table 6.3. The $\gamma(L)$ coefficient part is derived as

$$\gamma(L) = \frac{0.31L - 6.1}{L + 1.7} \quad (6.3)$$

by fitting to a rational function with R-square 0.99. Figure 6.6 visualizes this model.

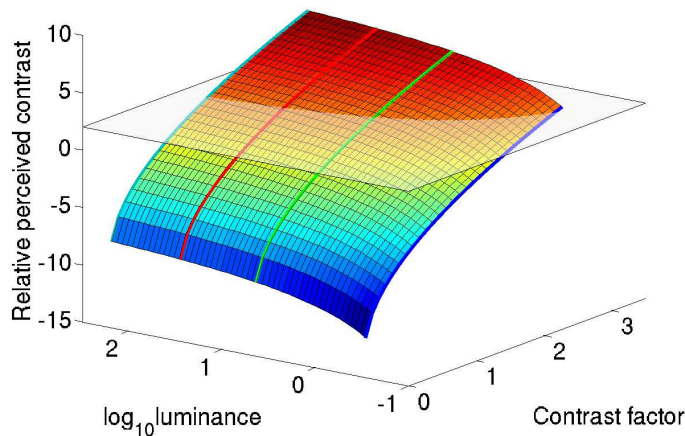


Figure 6.6: A surface model of perceived contrast in JNDs with respect to different mean luminance levels and contrast factors (see Equation (6.2)). A transparent surface shows that we need different contrast factors for different luminance levels to achieve the same perceived contrast (8 JNDs as an example). The curves derived in Figure 6.5 are also displayed on the surface.

6.7 Discussion

The studies of physical versus perceived contrast change in the context of simple patch stimuli or sinusoidal patterns have led to the derivation of *power law* for contrast discrimination [Legge 1980, Whittle 1986, Kingdom and Whittle 1996] and *contrast transducer* functions [Legge and Foley 1980, Mantiuk et al. 2006]. Although it is yet unclear how to objectively compare these findings to our studies on a complex image, we analyze and discuss apparent similarities in the following sections. Throughout the analysis we refer to the *contrast factor* c as a relative contrast measure, therefore both thresholds and scaling are expressed in it.

According to the data in Table 6.2, contrast discrimination threshold remains approximately constant for different *contrast factors* and has consistent characteristic across the luminance changes. While *contrast factor* is a measure relative to the existing contrast in the area, it means that we observe a contrast masking effect [Daly 1993] with exponent close to 1. The discrimination threshold for *contrast factor* is independent of the existing contrast in the image. The threshold remains approximately constant for middle and dark luminance values, but strongly increases for very dark luminance.

The range of local contrasts in our test image, measured at frequency of highest contrast sensitivity, spans up to 0.3 in Michelson measure. For such contrasts, Peli et al. [Peli et al. 1991] observed a similar behavior in a corresponding experiment for a simple stimuli. We also observe a slight increase in threshold for bright areas which is unusual.

The *contrast scaling* experiment derived the relation between the relative contrast measure c and JND of contrast. Such a relation is usually described by the contrast transducer function [Mantiuk et al. 2006], which is a power function. The contrast transducer converts contrast $G = \log(L_{max}/L_{min})$, to the JND of contrast. Parameterizing the contrast G with contrast scaling from Equation (6.1) we can derive the relation $G(c) = c \cdot G(1)$, where $G(1)$ is the contrast in the unmodified image. Since $G(1)$ is constant for a given image, we conclude that the contrast transducer for c should also follow the power law.

The fit of the data from the experiment to a power function results in a fair consistency of perceptual response to contrast across measured luminance levels. The exponent value $\beta = 0.47$ and scale value $\alpha = 9.32$ are approximately the same for all luminance levels and the curve is only shifted along the JND axis depending on the luminance (see Equation (6.3)). The exponent of the contrast transducer derived by Mantiuk et al. [Mantiuk et al. 2006] is approximately equal to 0.52 and is similar to our results obtained for the complex image.

We wrap the aspect of contrast in complex images in the *contrast factor* from Equation (6.1) which permits obtaining a relation between two contrasts without actually measuring them. Our handling of contrast generalizes the fact that overall image contrasts is composed from several sub-band components which have varied influence on the perceived contrast. Although we made effort that our image is representative for natural scenes, we probably make a generalization which is yet to be estimated. Currently, however, the comparison to related measurements for simple stimuli does not indicate any incorrectness.

6.8 Application

6.8.1 Uniform Contrast Scaling

We aim at maintaining perceptual uniformity in contrast scaling for complex images across wide luminance range. We employ our experimental data to parameterize *contrast factor* in Equation (6.1) so that we adjust contrast scaling by specifying the amount of perceived contrast change C_p in relative JND units of contrast:

$$L(C_p) = \bar{L} \cdot \left(\frac{L}{\bar{L}} \right)^{c(C_p, L)}. \quad (6.4)$$

The parameterized *contrast factor* $c(C_p, L)$ is obtained as a cross section of surface in Figure 6.6 at a fixed relative JND contrast change C and as a function of pixel luminance L . As a formula, $c(C_p, L)$ can be obtained as an inverse function of Equation (6.2):

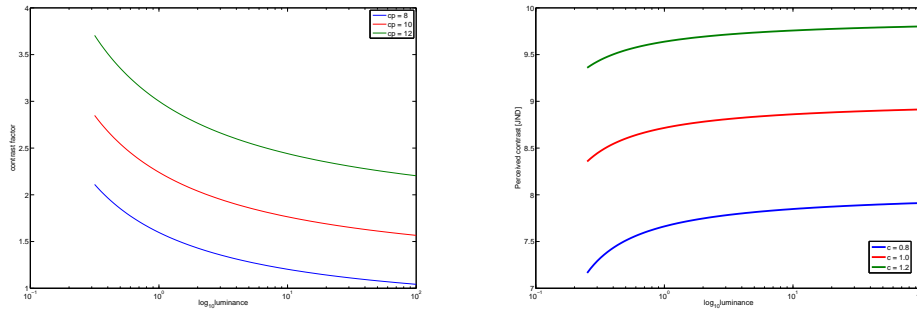
$$c(C_p, L) = \left(\frac{C_p - \gamma(L)}{\alpha} \right)^{\frac{1}{\beta}} \quad (6.5)$$

where C_p is a desired perceived contrast and $\gamma(L)$ is same as Equation (6.3).

The analysis of the parameterization $c(C_p, L)$ in Figure 6.7(a) reveals that the value *contrast factor* varies significantly for a given perceptual change of contrast. By taking the reverse, a fixed *contrast factor* leads to perceptual non-uniformity in contrast change of about 4 JND units across luminance range available on current displays (Figure 6.7(b)). Figure 6.7(a) also demonstrates an interesting observation that a desired decrease in contrast equal to -2 JND with respect to middle luminance, results in no contrast change in very dark areas. In the next sections we use Equation (6.4) to maintain perceptual uniformity in global and local contrast scaling.

6.8.2 Global Contrast Scaling

The global contrast scaling is obtained when the reference luminance \bar{L} in Equation (6.4) is constant for all pixels in the image. To maintain perceptual uniformity, the exponent of a power function is dependent on the pixel's luminance value and results in an adjusted luminance mapping function. The plot in Figure 6.8 illustrates that high luminance requires smaller contrast change than lower luminance. Such a difference in mapping is mandated by our experiment and derived based on its model (see Figure 6.6), and it stays in accordance with experiments by Peli et al. [Peli et al. 1991].



(a) Inversely computing *contrast factors* to obtain the same perceived contrast. (b) Applying the same *contrast factors* globally. It causes different perceived contrast.

Figure 6.7: Influence of luminance level on perceived contrast change and on adjustment of contrast factor to maintain perceptually uniform contrast change. The values of α and β in Equations (6.2) and (6.5) are set as $\alpha = 9.2$ and $\beta = 0.47$.

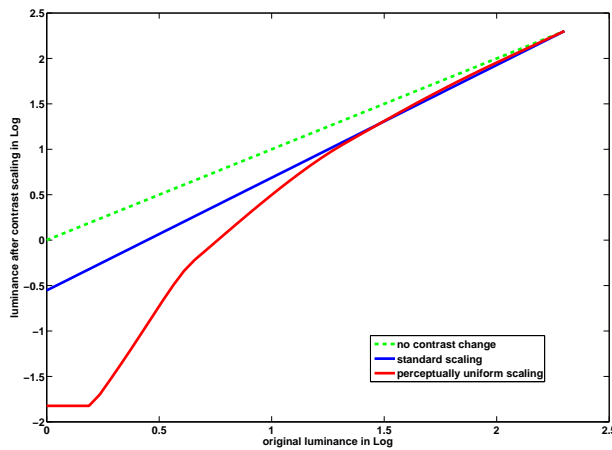


Figure 6.8: Luminance mapping in perceptually uniform global contrast scaling. Mapping clipped to minimum display luminance.

6.8.3 Local Contrast Scaling

Adjusting the reference luminance \bar{L} in Equation (6.4) to an average of certain small area around each pixel in the image, the contrast scaling equation becomes an unsharp masking filter for enhancement of local contrasts. Analogically to previous Section 6.8.2, illustrate that high luminance areas require smaller contrast enhancement than lower luminance areas. Fixed *contrast factor* leads to much weaker perceived enhancement of local contrast in dark areas (see Figure 6.9).



Figure 6.9: Standard local contrast enhancement (Top) and perceptually uniform local contrast enhancement (Bottom) by $C_p = +4$ JND with respect to the original image shown in Figure 6.2. Differences are most visible in marked areas, but are very subtle unless observed on an HDR display.

6.9 Summary and Future Work

Through psychophysical experiments, we derived a model for a perceptually uniform contrast change in complex images and demonstrated its application to global and local contrast scaling. We expect that such a new method is particularly important for displays with wide luminance range, which reduces the non-uniformity in contrast scaling of several JND units. We observed certain resemblance of our results for complex images with experiments of others performed for simple stimuli.

In the next step, we plan to extend our experiments to a more representative group of test images and to extensively compare our results with current findings in psychophysics.

Chapter 7

Brightness of the Glare Illusion

7.1 Introduction



Figure 7.1: Glare illusion in painting (left), photography (middle), and computer-generated games (right). The right image courtesy of Remedy Entertainment.

The glare illusion has been efficiently used for boosting the brightness of light sources in paintings, exploited in photography, and commonly employed in computer games (see Figure 7.1 from left to right, respectively). The illusion can evoke a very realistic sensation of self-luminous objects and can produce an impression of higher brightness than the maximum of a computer display or reflectance of white paint. While painters have to rely on their skill to produce the glare illusion, glare in photography arises naturally as the result of light scattering in lenses (referred to as *lens flare*) and can be further enhanced by cross screen or diffusion filters.

A large number of papers in computer graphics have proposed advanced visual models (diffraction and diffusion in the eye optics) to generate a realistic glare

illusion [Nakamae et al. 1990, Rokita 1993, Spencer et al. 1995, Ward Larson et al. 1997, Kakimoto et al. 2005]. However, both painters and photographers have been able to produce stunning glare illusions without any knowledge about the optical effects in the eye. Accurate visual models are also rarely used in practice, for example in game engines, as they are computationally too expensive. Instead, game artists hand-tune their digital filters to produce the best effect, even though the resulting convolution kernel is very different to the actual visual models. In this paper, we compare both approaches: an ad-hoc approach that involves the convolution with a Gaussian filter, and a physically-based approach that employs a Point Spread Function (PSF) of the eye. We measure the brightness boost that can be achieved with both methods and discuss the problems that may arise, such as deformation of the “glaring” objects due to clipping of high pixel values and undesirable Mach-band illusion that forms a bright outline around the modified objects [Yoshida et al. 2008a].

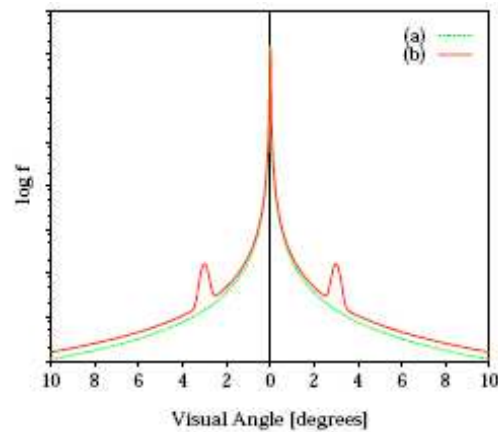
It should be noted, that the glare illusion we investigate in this research is different from *disability glare* and the *illusionary glare* effect. The glare effect consists of an illusionary glow (blooming), concentric rings of different colors (corona), and radial streaks (flare) that we can observe around bright objects and light sources. The glare effect causes the so-called *disability glare*, which is the loss of contrast visibility in the presence of strong light sources. The glare illusion, on the other hand, evokes an illusion in the center of an object rather than in its surround as the presence of a smooth gradient around an object can cause the object to appear brighter and self-luminous.

In the following sections, we introduce the related work about the glare illusion in Section 7.2. Section 7.4 contains stimuli and apparatus for our psychophysical experiment. Preliminary experiments and the main experimental setup and procedure are described in Sections 7.3 – 7.5. Results and discussion of our experiments are summarized in Sections 7.6 and 7.7. In the end, Section 7.8 concludes this work.

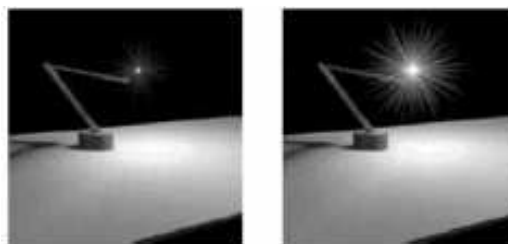
7.2 Previous Work

Rendering Glare A number of methods have been proposed for the glare rendering in computer graphics. Nakamae et al. introduced a rendering technique by considering diffraction effects at the pupil and eyelashes in images with high intensity lights [Nakamae et al. 1990]. Rokita proposed a technique to render high intensity lights, blooming and glare [Rokita 1993]. His method dealt with the spectrum of the incoming light and diffraction at the lens and on particles in

the eye. Spencer et al. presented a quantitative model to render glare [Spencer et al. 1995] (see Figure 7.2). They reviewed the physical mechanism of glare and modelled it by a PSF for each of photopic, mesopic, and scotopic cases. They also reported that the glare effects enhance the brightness of light sources. Ward Larson et al. employed Moon and Spencer's adaptation model [Moon and Spencer 1945] in their tone reproduction operator to enhance bright objects [Ward Larson et al. 1997]. Kakimoto et al. attributed the main source of glare to the diffraction on the eyelashes and pupil and simulated it using wave optics [Kakimoto et al. 2005]. Van den Berg et al. proposed a physical model to simulate the ciliary corona found by Simpson [Simpson 1953] that often accompanies the perception of real glare sources [van den Berg et al. 2005]. They assumed that the incoming light is scattered on small particles situated in the lens and the vitreous in the eye.



(a) PSFs for (a) photopic and (b) scotopic scenes.



(b) Examples of Spencer et al.'s model.

Figure 7.2: Perceptual model of rendering glare proposed by Spencer et al. [Spencer et al. 1995].

Glare in Games All of the above methods render glare effects by simulating optics of the eye. Kawase proposed a method to render glare by combining several Gaussian convolutions with different kernel sizes [Kawase 2005] (see Figure 7.3). This method has no perceptual background, still, since this approach is simple and computationally inexpensive, it is often used in computer games.

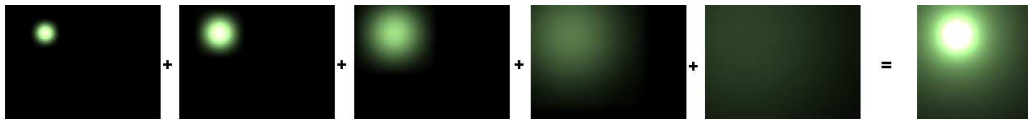


Figure 7.3: Generating glare in computer games by using multiple Gaussian filters and magnifying them with a weighting function used in [Kawase 2005].

Appearance of Glare Although much attention was put to physical and optical aspects of the glare effects and modelling disability glare [Vos 2003], the brightness boosting glare illusion has not been well studied. Zavagno and Caputo conducted psychophysical experiments to measure the impression of self-luminosity of glare [Zavagno 1999, Zavagno and Caputo 2001]. They asked subjects to increase the gradient of ramps between a bright patch and four surrounding dark squares until the center patch started being perceived as self-luminous. They found that there was a linear relation between the background luminance and the ramp gradient.

Visual Illusions The glare illusion often coexists with other illusions, which can either raise or lower perceived luminance. *Simultaneous contrast* causes a perceptual shift in color appearance when the color of the stimulus background is changed [Gerrits and Vendrik 1970, Adelson 1993] [Fairchild 1998, Chapter 6]. A stimulus is perceived as darker on a light background while the same stimulus is perceived brighter on a dark background (see Figure 2.11).

Furthermore, the steep gradient of the glare profile and its abrupt termination by clipping can elicit the *Mach-band* illusion [Ratliff 1965, Lotto et al. 1999], which is visible as a bright outline around the glaring object (see Figure 2.12). Finally, the convolution kernel used to produce glare can cause an object to grow or to change shape (see the first column of Figures 7.6 – 7.8), which results in an increase of brightness, since larger objects often appear to be brighter [Li and Gilchrist 1999]. Refer to Section 2.4 for simultaneous contrast and Mach bands.

7.3 Preliminary Experiments

Although perceived luminance is usually measured by using magnitude estimation, magnitude production, or brightness matching methods, we found that these methods resulted in too noisy data to interpret with high inter- and intra-observer variance. One possible reason for the high variance is that the illusion is often subtle and not much larger than the discrimination threshold. To reduce the discrimination threshold, we increased the background luminance from almost black (1 cd/m^2) to a much higher level (150 cd/m^2) with the constant disk luminance at 220 cd/m^2 in our pilot studies. This should have helped to reduce the discrimination threshold, as it is known that it increases with the luminance difference between a background and a target disk [Whittle 1986].

We also increased the disk size and reduced the distance between reference and target images, as larger stimuli that are closer to each other are easier to compare. In addition, we experimented with temporal comparison, however, we dismissed this idea of experimental procedure because the measurements could be affected by the Gelb effect (see [Gilchrist et al. 1999] for details on the Gelb effect). To improve accuracy, we also tried to employ stricter procedures, such as two-alternative forced choice (2AFC) combined with Parameter Estimation by Sequential Testing (PEST) [Taylor and Creelman 1967], but we did not observe a reduction in variance.

7.4 Experimental Setup for Measuring Brightness of the Glare Illusion

7.4.1 Apparatus for All Experiments

We conducted psychophysical experiments to measure the boost in brightness caused by glare illusion. The input images used in our experiments consisted of a disk ($0.3, 0.6, \dots, 1.5 \text{ vis deg}$) displayed on a background image (3.2 vis deg) containing a cloudy sky. The complex background introduced both contrast and context, which was more natural setting than a flat background. The average luminance of the background was set at $L_{\text{bg}} = 50, 100, \dots, 200 \text{ cd/m}^2$. A reference image with the glare illusion rendered around the disk was shown in the center and two target images without any glare were presented on both sides, as shown in Figure 7.4. The maximum luminance of the reference image was kept constant at $L_{\text{dmax}} = 220 \text{ cd/m}^2$ to simulate the maximum luminance of a typical display,

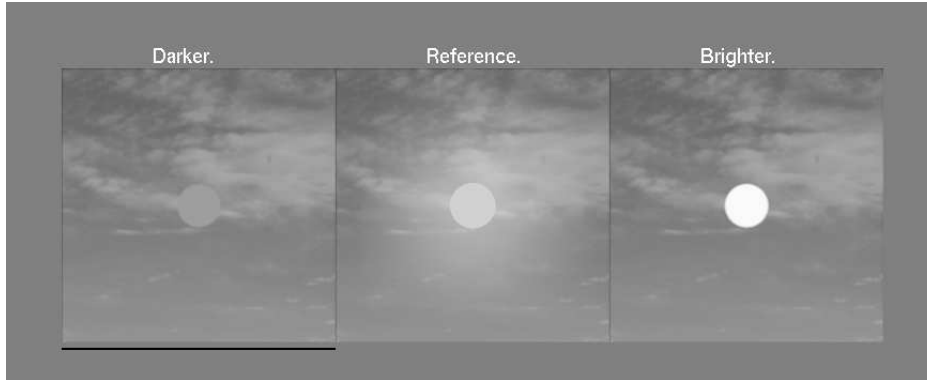


Figure 7.4: A screenshot of a single trial of the experiment. A reference image (middle) and two target disks (left and right) are shown at each trial. The black line under the target images indicates which image is activated to modify. A subject was asked to adjust the perceived luminance of the target disks which looked slightly but visibly darker (left) or brighter (right) than that of given reference image.

but the disk luminance of the target images could be increased up to the actual maximum luminance of the display used for our experiments (433 cd/m^2).

The images were displayed on a 20.8" Barco 10-bit LCD display (Coronis Color 3MP Diagnostic Luminance¹). The 10-bit precision eliminated potential contouring artifacts on smooth gradients, which could have been observed on an 8-bit display. The Barco display was carefully calibrated by measuring its luminance response for a range of input values using the MINOLTA LS-100 light meter².

7.4.2 Methods for Generating the Glare Illusion

To generate the glare illusion for the reference image, we used two strategies: a method that employed Gaussian convolution, commonly used in game engines (Method I); and the method proposed by Spencer et al. [Spencer et al. 1995] that employs a PSF of the human eye (Method II). Our input is a linear luminance image L (not gamma corrected). In both methods, we first compute for each pixel the luminance that exceeds the maximum luminance of a typical display L_{dmax} as

$$\Delta L = \begin{cases} L - L_{\text{dmax}} & \text{if } L > L_{\text{dmax}} \\ 0 & \text{otherwise.} \end{cases} \quad (7.1)$$

¹ www.barco.com/corporate/en/products/product_specs.asp?specs.asp?element=2882#

² www.konicaminolta.com/instruments/products/light/luminance-meter/ls100-ls110/index.html

Next, ΔL is convolved with an appropriate 2D digital filter. There are two reasons for applying convolution only to the values greater than L_{dmax} . Firstly, we do not want to blur the entire image; and secondly, only the pixels whose luminance can not be displayed should be boosted in brightness.

For Method I, the convolution kernel is given by

$$F(x, y) = \frac{1}{k} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (7.2)$$

where x, y are pixel indices (from $-s/2$ to $s/2$, where s is the stimulus size), $\sigma = 0.34$ vis deg, and k is a normalization factor computed as the sum of all kernel elements. For Method II, similarly as in [Spencer et al. 1995], we employ the PSF proposed by Vos [Vos 1984]:

$$PSF(\theta) = 0.384 f_0(\theta) + 0.478 f_1(\theta) + 0.138 f_2(\theta) \quad (7.3)$$

where θ is the angle between the primary object and the glare source in degrees and $f_0 \dots f_2$ are given as

$$f_0(\theta) = 2.61 \times 10^6 e^{-(\frac{\theta}{0.02})^2}, \quad (7.4)$$

$$f_1(\theta) = \frac{20.91}{(\theta + 0.02)^3}, \quad (7.5)$$

$$f_2(\theta) = \frac{72.37}{(\theta + 0.02)^2}. \quad (7.6)$$

We compute the digital filter by integrating the proposed PSF of the eye using trapezoidal numerical integration over ten samples for each pixel. The result of the convolution is added back to the original luminance map L and all values are clamped to the maximum value L_{dmax} .

A lookup table generated by the MINOLTA LS-100 light meter for display calibration is used to map the resulting luminance values to the display pixel values. We generate stimuli of twice the resolution as required and then filter and subsample them to avoid aliasing artifacts.

7.4.3 Stimuli for Experiment I

To vary the strength of the glare illusion, the input disk luminance levels for the glare rendering are set as $L_{\text{disk}} = 220, 1165, 2110, 3055, 4000, 7000$ cd/m^2 (labeled as ‘‘A’’ – ‘‘F’’ respectively) for Method I and, for Method II, $L_{\text{disk}} = 220, 1480, 2740, 4000, 7000, 10000$ cd/m^2 (‘‘a’’ – ‘‘f’’ respectively). The size of

the center disk was set at 0.3 vis deg, and the average background luminance was set at $L_{bg} = 150 \text{ cd/m}^2$. Note that the L_{disk} parameter is abstract and the same value of this parameter can result in different strengths of the glare illusion for Methods I and II. However, we selected the values L_{disk} so that the stimuli “C” and “c” do not differ visibly in size from the original disk, and the entire usable range of this L_{disk} is examined. All reference images used in the experiments are shown in Figure 7.5.

7.4.4 Stimuli for Experiments II, III, and IV

After conducting Experiment I, we extended a variety of background luminances. The average background luminances were set at $L_{bg} = 50 \text{ cd/m}^2$ for Experiment II, $L_{bg} = 100 \text{ cd/m}^2$ for Experiment III, and $L_{bg} = 200 \text{ cd/m}^2$ for Experiment IV. The other setup of an experiment was kept as same as that of Experiment I. The input disk luminance for the glare rendering were $L_{disk} = 220, 1165, 2110, 3055, 4000, \text{ and } 7000 \text{ cd/m}^2$ (labeled as “A” – “F” respectively) for Method I (Gaussian) and $L_{disk} = 220, 1480, 2740, 4000, 7000, \text{ and } 10000 \text{ cd/m}^2$ (“a” – “f”) for Method II (Spencer et al.). The center disk was remained same as Experiment I: 220 cd/m^2 and 0.3 visual degree. All stimuli and their profiles are shown in Figure 7.6 for Experiment II, Figure 7.7 for Experiment III and Figure 7.8 for Experiment IV.

7.5 Experimental Procedure for Measuring Brightness of the Glare Illusion

To further reduce the randomness in the subjective responses, we employed an arranged increment/decrement method. Subjects were asked to adjust the target images such that the perceived luminance of the left disk was as close as possible to that of the reference disk but slightly and visibly *darker*. Likewise, the right disk should be adjusted to be perceived as slightly but visibly *brighter*. Then, the matching perceived luminance is assumed to be the mean of both left and right target disk luminance, thereby producing a measure that is more robust against outliers (see Figure 7.4).

When the Mach-band illusion was seen on the reference disk, the subject was asked to ignore the Mach-band and adjust the target disks to the brightness inside the illusionary ring. As a hint of how to adjust brightness of target disks, a subject was instructed as follows: “You could try to adjust the brightness of the

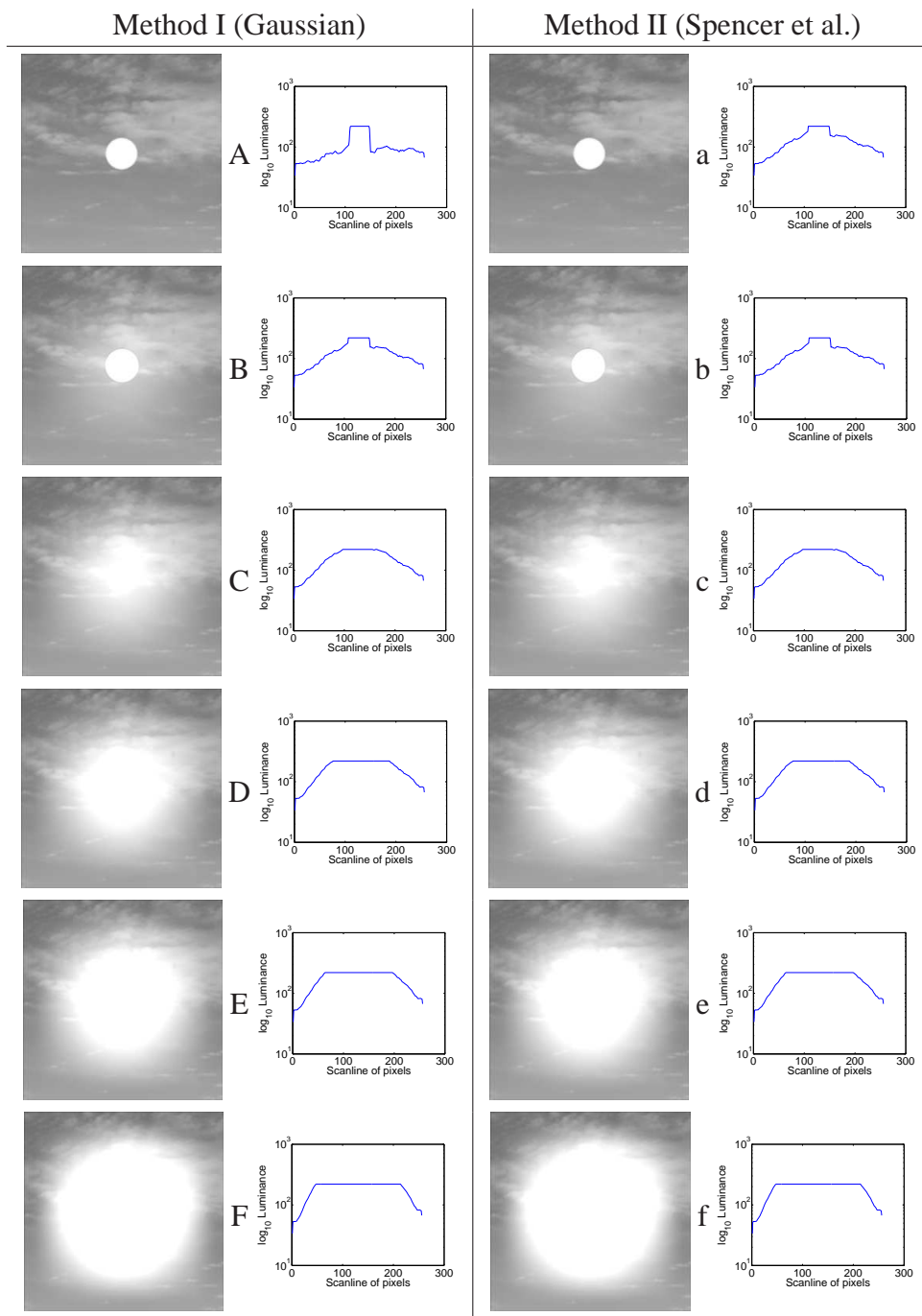


Figure 7.5: Experimental stimuli and their profiles at $L_{bg} = 150cd/m^2$. The images in the left column show the stimuli and profiles for Method I (Gaussian), while those for Method II (Spencer et al.) are arranged at the right side. The characters between stimuli and profiles (“A” – “F” and “a” – “f”) indicate the luminance of the reference disks L_{disk} (refer to Section 7.4.3 for details).

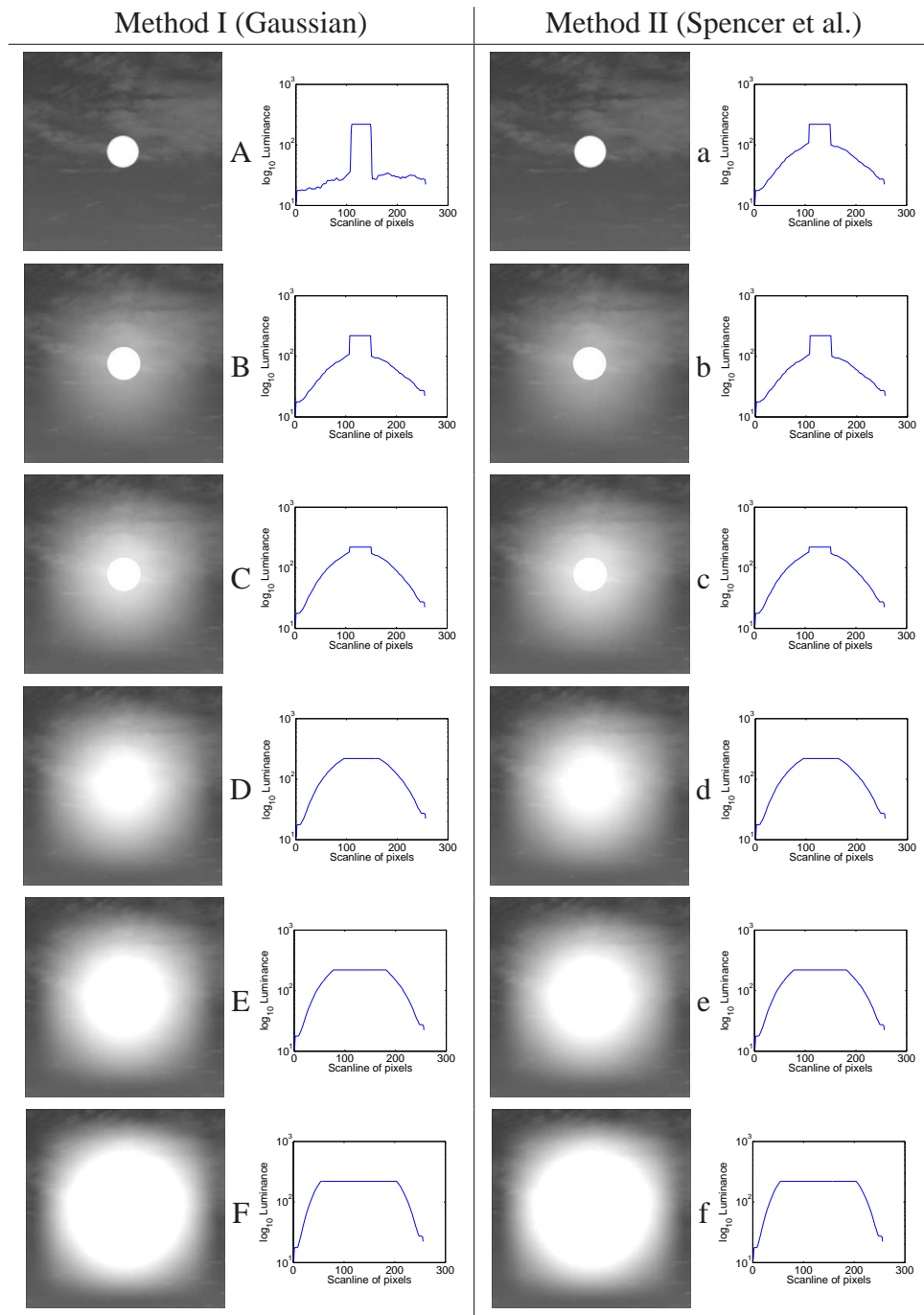


Figure 7.6: Experimental stimuli and their profiles at $L_{bg} = 50 \text{ cd/m}^2$. The images in the left column show the stimuli and profiles for Method I (Gaussian), while those for Method II (Spencer et al.) are arranged at the right side. The characters between stimuli and profiles (“A” – “F” and “a” – “f”) indicate the luminance of the reference disks L_{disk} (refer to Section 7.4.4 for details).

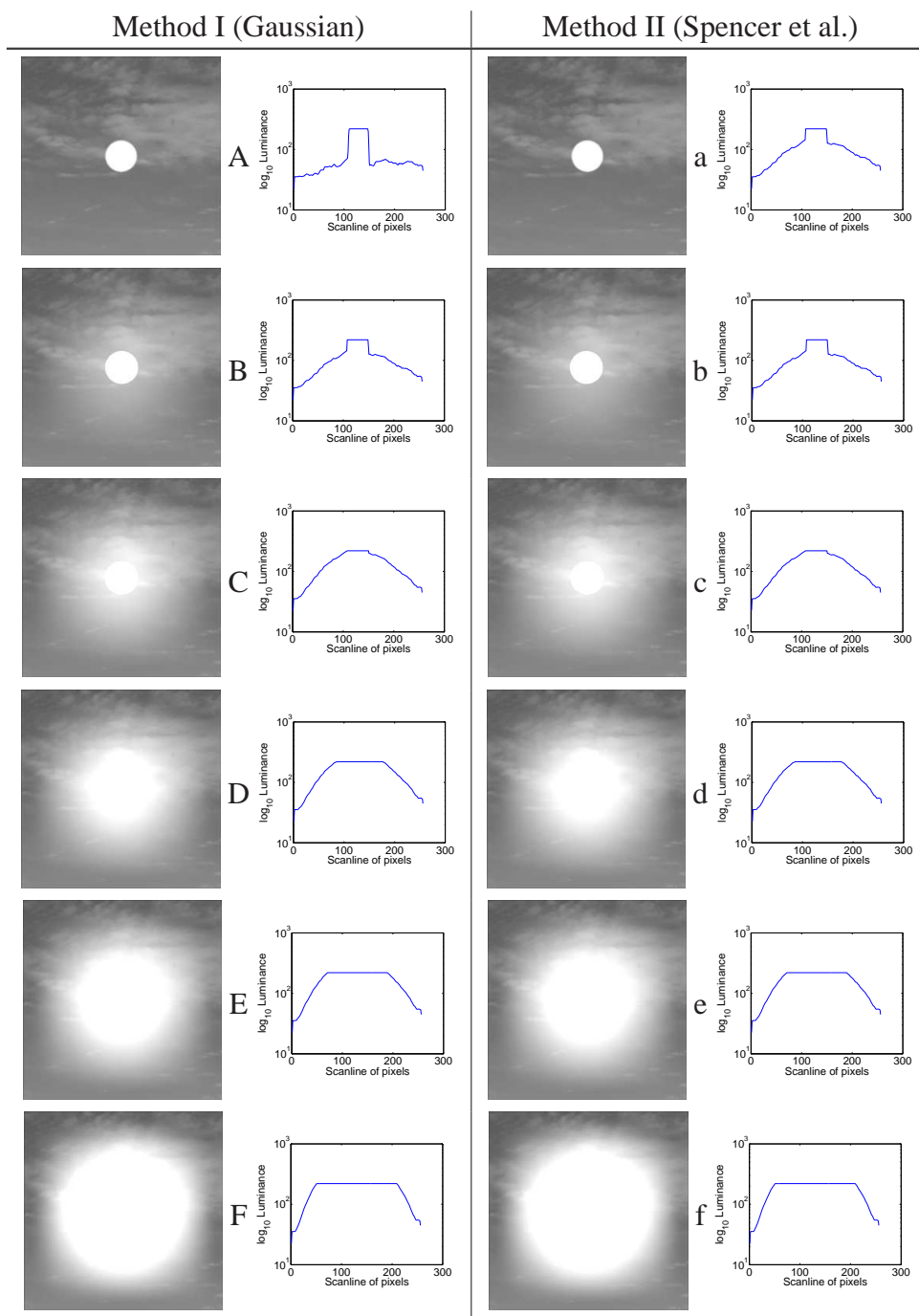


Figure 7.7: Experimental stimuli and their profiles at $L_{bg} = 100 \text{ cd/m}^2$. The images in the left column show the stimuli and profiles for Method I (Gaussian), while those for Method II (Spencer et al.) are arranged at the right side. The characters between stimuli and profiles (“A” – “F” and “a” – “f”) indicate the luminance of the reference disks L_{disk} (refer to Section 7.4.4 for details).

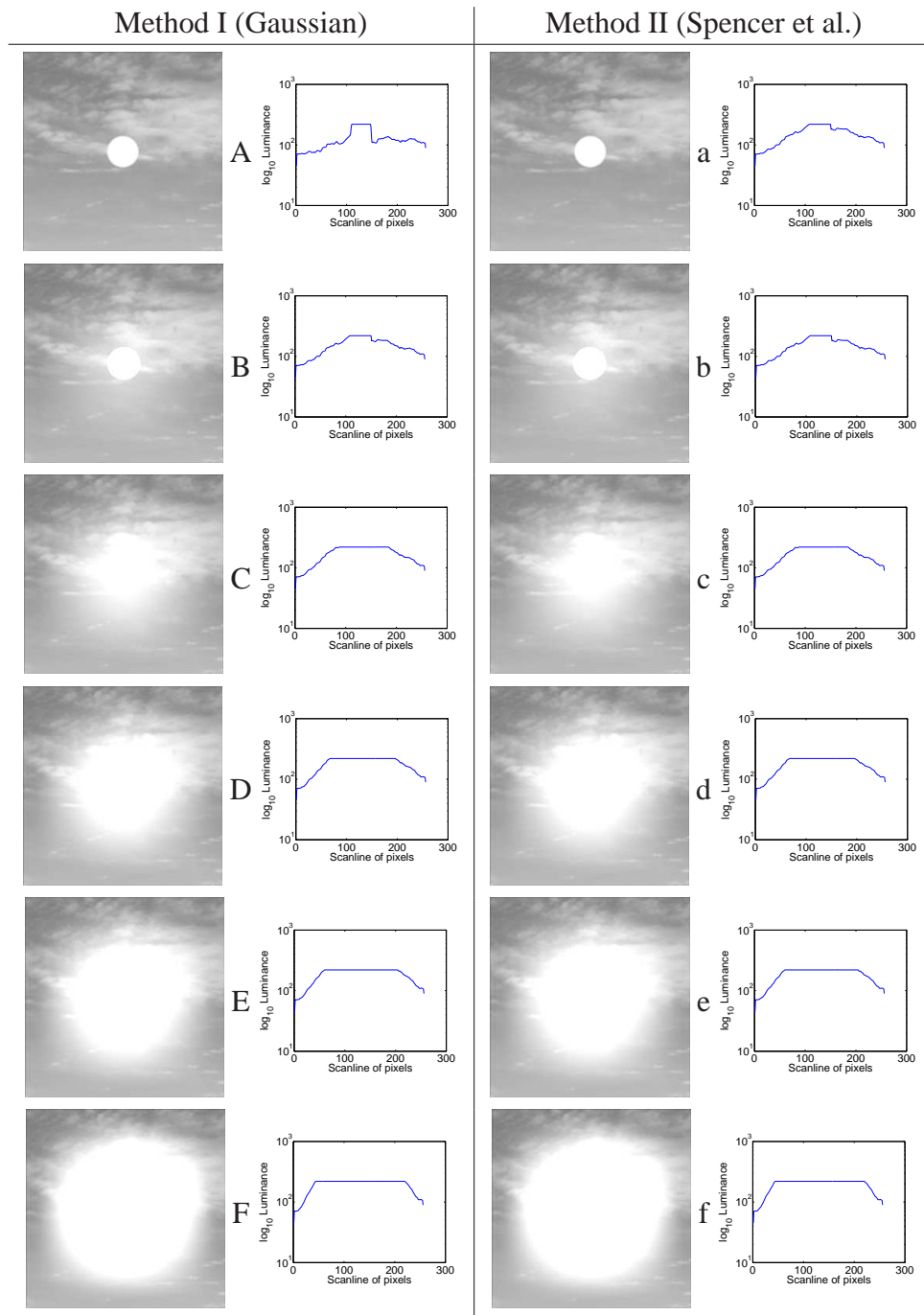


Figure 7.8: Experimental stimuli and their profiles at $L_{bg} = 200 \text{ cd/m}^2$. The images in the left column show the stimuli and profiles for Method I (Gaussian), while those for Method II (Spencer et al.) are arranged at the right side. The characters between stimuli and profiles (“A” – “F” and “a” – “f”) indicate the luminance of the reference disks L_{disk} (refer to Section 7.4.4 for details).

target disks as same as that of the reference and then go down/up until you start seeing the difference.” Both target disks were initially set to significantly different luminance levels (100 and 400 cd/m^2 for “darker” and “brighter” target images, respectively).

After adjusting the luminance of the target disks, a questionnaire followed each trial. All questions were asked for each reference image and could be answered by “yes” or “no”.

Q1: Does the reference image glow? [y/n]

Q2: Do you see a bright ring (a.k.a. Mach-band) on the reference disk? [y/n]

Q3: Are the sizes of the reference and target disks the same? [y/n]

10 subjects (7 males and 3 females) at the average age of 30 (between 26 – 40 years old) participated in our experiment. All subjects were naïve about the purpose of the experiment and had either normal or corrected-to-normal vision. The subjects were seated at a distance of 1 m from the display under dim lighting condition (60 lux). Each subject read a written instruction of the experiment, passed a training session, and then took the main part of the experiment. The whole procedure took approximately 20 minutes for a single subject.

7.6 Results

7.6.1 Experiment I: Measuring Brightness of Glare Illusion at $L_{bg} = 150 \text{ cd/m}^2$

The results of Experiment I are plotted in Figures 7.9 and 7.10. As shown in Figure 7.10, the glare effect can raise the perceived luminance by 20 – 35% compared to the actual luminance 220 cd/m^2 , and Method I boosts the perceived luminance more than Method II. Apparently, perceived luminance levels increase with increasing luminance of the disk L_{disk} that enters both methods as the main parameter. The growing trend of perceived luminance as a function of L_{disk} does not appear to be linear: while small and medium values of L_{disk} have strong effects on perceived luminance, this effect saturates for large values of L_{disk} .

It is apparent from Figure 7.9, that the upper and the lower bounds of the perceived luminance do not differ qualitatively since the general shape of the curves are close to parallel in all cases. Therefore, the measuring accuracy can be increased by using the mean of these two thresholds instead of the two separate values. In

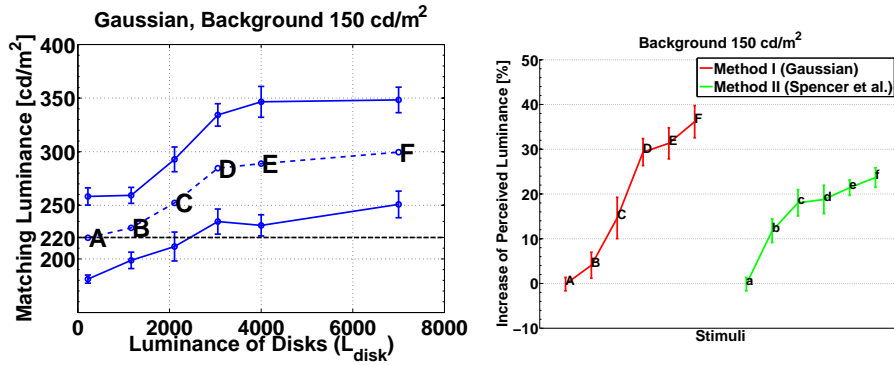


Figure 7.9: Results of the experiment on measuring brightness of glare illusion for $L_{bg} = 150 \text{ cd/m}^2$. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.3 for details).

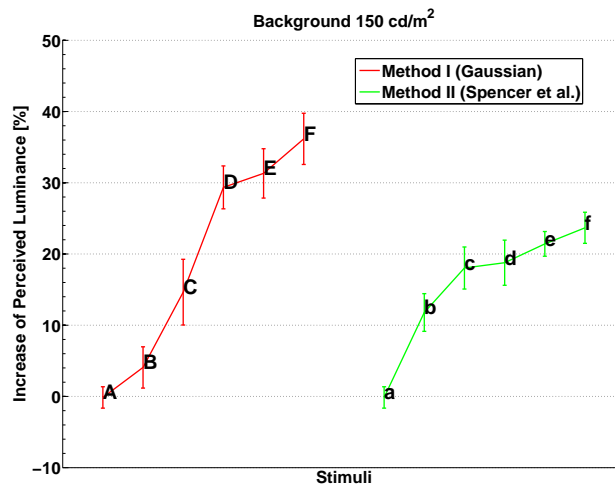


Figure 7.10: Percentage of increase for the mean perceived luminance levels at the background level $L_{bg} = 150 \text{ cd/m}^2$ for Methods I (Gaussian) and II (Spencer et al.) with errorbars of SEM. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.3 for details).

the analyzes reported in the following, only the mean of the two thresholds is used as a dependant variable.

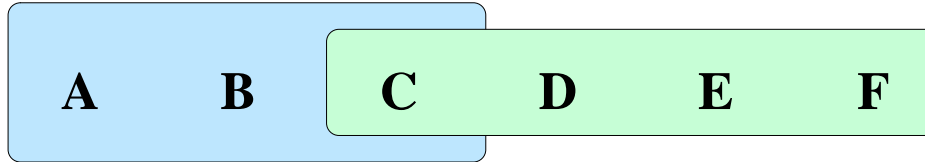
To analyze the data, we conduct a 6 (disk luminance L_{disk}) \times 2 (method) Analysis of Variance (ANOVA), treating disk luminance and method as repeated measurement factors. Because the scale of the parameter L_{disk} is not comparable for the two Methods, the analysis treats different levels (“A” – “F” and “a” – “f” for Methods I and II, respectively) rather than numerical values of L_{disk} as equivalent. These values are assumed to be of ordinal scale. There is a significant main effect of L_{disk} , $F(5) = 23.68, p < .001$, indicating that the luminance of the center disk (which entered the algorithms as the main parameter) has an influence on how bright it was perceived to be. The main effect of factor “Method” is not significant, $F(1, 9) = 1.64$, showing that the two Methods do not differ in brightness boost over all levels. However, the $L_{\text{disk}} \times$ Method interaction reaches significance, $F(5) = 7.77, p < .001$, indicating that the Methods differ at some, or at least one, of the levels of L_{disk} .

To narrow down this effect, further analyzes are carried out. From the visual inspection of the data (Figure 7.10), it is suspected that the two Methods differ only for large values of the parameter L_{disk} . Therefore, two ANOVAs which are similar to the one above are conducted for levels {A, B, C} and {D, E, F} separately. As expected, the main effect for “Method” in the ANOVA for the first group of levels {A, B, C} remains not significant ($F(1, 9) = 1.4$), while the Method factor reaches significance for the second group of levels {D, E, F} ($F(1) = 11.96, p < .01$). There are no other significant effects, in particular the Method \times L_{disk} interaction does not reach significance in both analyzes ($F(2) = 1.34$ and $F(2) = 0.28$), indicating that the interaction effect from the global analysis is sufficiently explained by this separation.

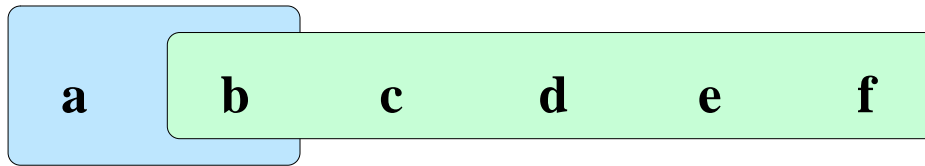
Pairwise t -tests of the two Methods for the levels D, E and F of L_{disk} are performed and reveal that, for all cases, the Gaussian method produces stronger perceived luminance ($t(14) = 1.71, p = .05$, $t(14) = 1.73, p = .05$, $t(14) = 1.96, p < .05$, for D, E and F).

To further investigate the relationship between perceived luminance and L_{disk} , pairwise contrasts between levels of L_{disk} are computed for both methods. To control the family-wise error rate, the p -values are adjusted, using the method proposed by Holm [Holm 1979]. The results of this analysis are illustrated in Figure 7.11. The indicated sets depict levels for which the perceived luminance values are statistically indistinguishable on a 95% significance level. For the Gaussian method, a “jump” in perceived luminance between the third and the fourth level arises, after which an increase in luminance does not further elevate perceived luminance. For Spencer et al.’s method, the increase occurs earlier between level B

and C.



(a) Method I (Gaussian)



(b) Method II (Spencer et al.)

Figure 7.11: Similarity groups of the L_{disk} levels at $L_{\text{bg}} = 150 \text{ cd}$ as revealed by post-hoc contrasts at a 95% significance level for the Gaussian method (a) and Spencer’s method (b). Items in the same set were statistically indistinguishable.

Method I (Gaussian), $L_{\text{bg}} = 150 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	A	B	C	D	E	F
Q1: Does the reference image glow?	0	60	90	100	100	100
Q2: Do you see a bright ring (a.k.a. Mach-band)?	0	10	20	50	80	90
Q3: Are the sizes of reference and target disks the same?	90	100	30	10	0	0
Method II (Spencer et al.), $L_{\text{bg}} = 150 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	a	b	c	d	e	f
Q1: Does the reference image glow?	0	100	90	100	100	100
Q2: Do you see a bright ring (a.k.a. Mach-band)?	0	70	80	80	80	90
Q3: Are the sizes of reference and target disks the same?	90	80	30	0	0	0

Table 7.1: Results of the questionnaire in percentages of the answer ‘yes’ for $L_{\text{bg}} = 150 \text{ cd/m}^2$. Colors indicate either above (red) or below (blue) 50%.

The answers of the subjects to the questionnaire presented after each trial are summarized in Table 7.1. For both Methods, the application of the glare models produces a “glowing” impression of the disk and is independent on how strong the glare is rendered. The results from Question 2 indicate that Method II is more likely to induce a Mach-band effect, which might be one aspect of an explanation of why Method II does not produce as strong an effect as Method I. However, another factor that probably helps to induce the difference between the Methods

is the size of the disk that increases with growing L_{disk} for Method I much stronger than for Method II (see Figure 7.5). This fact is also highlighted by the results for Question 3 of the questionnaire.

7.6.2 Experiments II – IV: Measuring Brightness of Glare Illusion at $L_{\text{bg}} = 50, 100, 200 \text{ cd/m}^2$

The results of Experiments II – IV are shown in Figures 7.12 and 7.13 (Experiment II: $L_{\text{bg}} = 50 \text{ cd/m}^2$), Figures 7.14 and 7.15 (Experiment III: $L_{\text{bg}} = 100 \text{ cd/m}^2$), and Figures 7.16 7.17 (Experiment IV: $L_{\text{bg}} = 200 \text{ cd/m}^2$). They are also analyzed by conducting $6 (L_{\text{disk}}) \times 2$ (methods) ANOVA with different levels (“A” - “F” and “a” - “f” for Methods I and II respectively) of L_{disk} rather than numerical values of L_{disk} .

In Experiment II at $L_{\text{bg}} = 50 \text{ cd/m}^2$, ANOVA indicates that the main effect of Methods is not significant, $F(1) = 3.64, p = 0.0608$. Unlike Experiment I, the main effect of L_{disk} ($F(5, 60) = 2.88, p = 0.0213$) is also not significant, while the interaction of Methods and L_{disk} ($F(4, 55) = 4.8, p = 0.0022$) is highly significant.

In Experiment III at $L_{\text{bg}} = 100 \text{ cd/m}^2$, the main effect of Methods is not significant ($F(1, 75) = 1.18, p = 0.28$) as well as Experiments I and II. On the other hand, both the main effect of L_{disk} and the interaction of Methods and L_{disk} are very significant: $F(5, 71) = 12.87, p \ll 0.01$ (the main effect of Methods) and $F(4, 66) = 4.52, p = 0.0027$ (interaction).

The results of Experiment IV at $L_{\text{bg}} = 200 \text{ cd/m}^2$ also show the same tendency that the main effect of Methods is not significant ($F(1, 75) = 0.54, p = 0.46$), but the main effects of L_{disk} ($F(5, 71) = 13.18, p \ll 0.01$) and the interaction ($F(4, 66) = 3.97, p = 0.006$) are both significant.

Tables 7.2, 7.3, and 7.4 summarize the answers of our subjects to the questionnaire for each trial. The answers for Experiments II – IV follow the same manner as that of Experiment I. the application of the glare models produces a glowing impression and is independent on the strength of the glare rendering (Question 1). According to the answers of Question 2, Method II (Spencer) produce more effects of Mach-band as well as Experiment I. Again, this could be an explanation why Method II does not produce as strong glare illusion as Method I because Mach-bands reduce brightness at the center relatively. This fact of the strength of glare illusion can be also seen in the answers of Question 3.

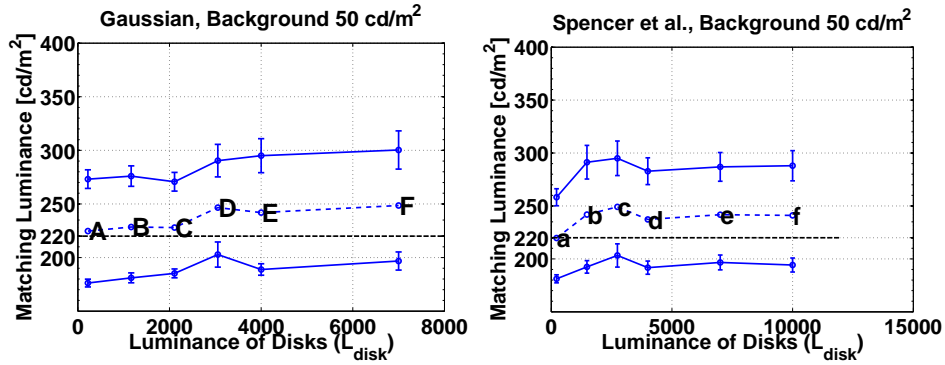


Figure 7.12: Results of the experiment on measuring brightness of glare illusion for $L_{bg} = 50 \text{ cd/m}^2$. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.4 for details).

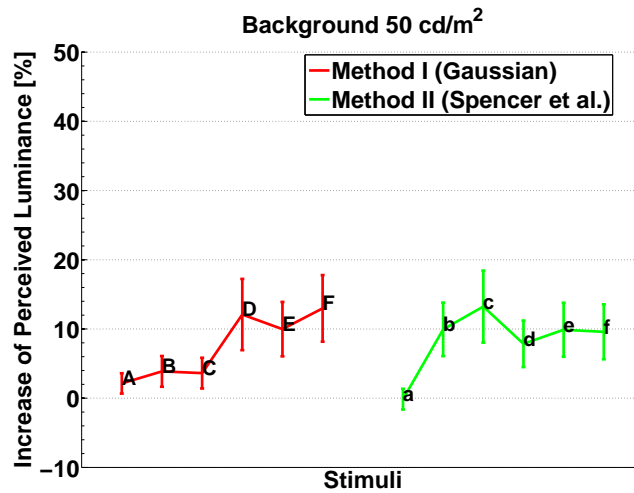


Figure 7.13: Percentage of increase for the mean perceived luminance levels at the background level $L_{bg} = 50 \text{ cd/m}^2$ for Methods I (Gaussian) and II (Spencer et al.) with errorbars of SEM. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.4 for details).

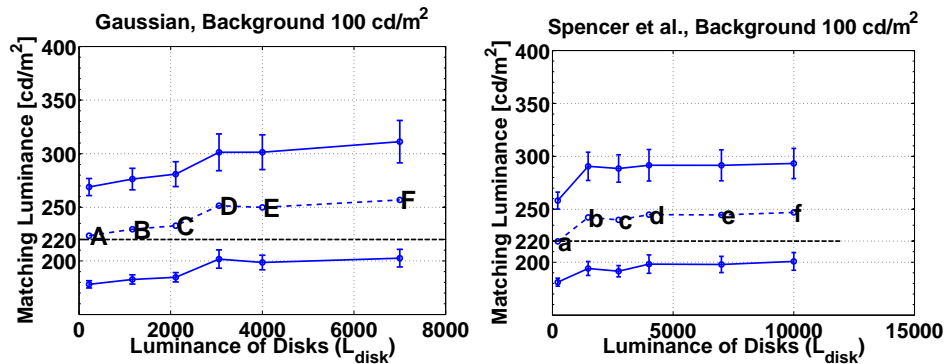


Figure 7.14: Results of the experiment on measuring brightness of glare illusion for $L_{\text{bg}} = 100 \text{ cd/m}^2$. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.4 for details).

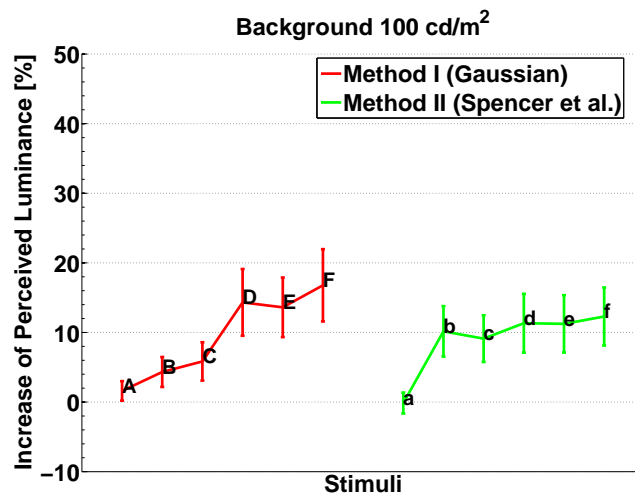


Figure 7.15: Percentage of increase for the mean perceived luminance levels at the background level $L_{\text{bg}} = 100 \text{ cd/m}^2$ for Method I (Gaussian) and Method II (Spencer et al.) with errorbars of SEM. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.4 for details).

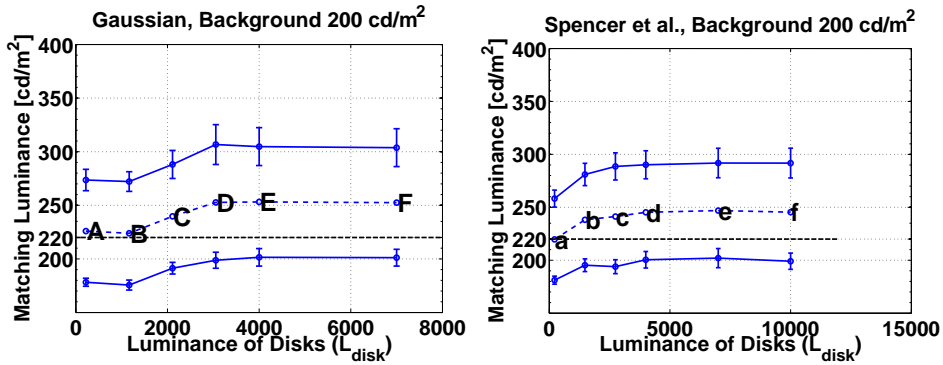


Figure 7.16: Results of the experiment on measuring brightness of glare illusion for $L_{bg} = 200 \text{ cd/m}^2$. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.4 for details).

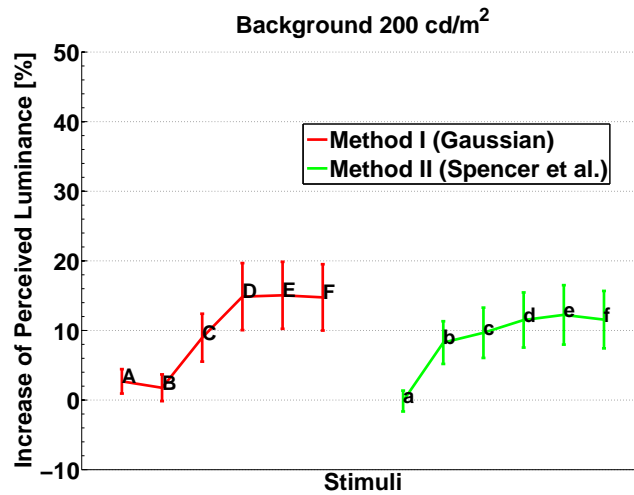


Figure 7.17: Percentage of increase for the mean perceived luminance levels at the background level $L_{bg} = 200 \text{ cd/m}^2$ for Methods I (Gaussian) and II (Spencer et al.) with errorbars of SEM. The characters (“A” – “F” and “a” – “f”) indicate the setting of the L_{disk} parameter (refer to Section 7.4.4 for details).

Method I (Gaussian), $L_{bg} = 50 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	A	B	C	D	E	F
Q1: Does the reference image glow?	0	100	100	100	100	100
Q2: Do you see a bright ring (a.k.a. Mach-band)?	0	0	33	83	83	83
Q3: Are the sizes of reference and target disks the same?	83	100	83	0	0	0
Method II (Spencer et al.), $L_{bg} = 50 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	a	b	c	d	e	f
Q1: Does the reference image glow?	100	100	100	100	100	0
Q2: Do you see a bright ring (a.k.a. Mach-band)?	67	100	100	83	83	0
Q3: Are the sizes of reference and target disks the same?	100	67	17	0	0	0

Table 7.2: Results of the questionnaire in percentages of the answer 'yes' for $L_{bg} = 50 \text{ cd/m}^2$. Colors indicate either above (red) or below (blue) 50%.

Method I (Gaussian), $L_{bg} = 100 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	A	B	C	D	E	F
Q1: Does the reference image glow?	0	86	100	100	100	100
Q2: Do you see a bright ring (a.k.a. Mach-band)?	14	14	43	86	100	100
Q3: Are the sizes of reference and target disks the same?	86	86	100	0	0	0
Method II (Spencer et al.), $L_{bg} = 100 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	a	b	c	d	e	f
Q1: Does the reference image glow?	86	100	100	100	100	0
Q2: Do you see a bright ring (a.k.a. Mach-band)?	71	100	100	100	86	0
Q3: Are the sizes of reference and target disks the same?	71	29	0	0	0	0

Table 7.3: Results of the questionnaire in percentages of the answer 'yes' for $L_{bg} = 100 \text{ cd/m}^2$. Colors indicate either above (red) or below (blue) 50%.

Method II (Gaussian), $L_{bg} = 200 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	A	B	C	D	E	F
Q1: Does the reference image glow?	0	86	100	100	100	100
Q2: Do you see a bright ring (a.k.a. Mach-band)?	0	0	14	71	71	51
Q3: Are the sizes of reference and target disks the same?	100	100	0	0	0	0
Method II (Spencer et al.), $L_{bg} = 200 \text{ cd/m}^2$.						
Luminance of disks L_{disk}	a	b	c	d	e	f
Q1: Does the reference image glow?	100	100	100	100	100	0
Q2: Do you see a bright ring (a.k.a. Mach-band)?	71	71	71	71	71	0
Q3: Are the sizes of reference and target disks the same?	86	14	0	0	0	0

Table 7.4: Results of the questionnaire in percentages of the answer 'yes' for $L_{bg} = 200 \text{ cd/m}^2$. Colors indicate either above (red) or below (blue) 50%.

7.7 Discussion

In our psychophysical experiment, we employ Spencer et al.'s model [Spencer et al. 1995] and Gaussian convolution model to produce glare illusion at different background luminances $L_{bg} = 50, 100, 150, 200 \text{ cd/m}^2$. Both Methods succeed in producing a strong glare illusion. It is shown that an increase in the chosen parameter results in a larger amount of perceived luminance. For high values of the luminance of the disks, the Gaussian Method produces a stronger boost in perceived luminance than Spencer et al.'s method.

However, the Gaussian method results in a stronger increase of disk size when large parameter values are chosen. Therefore, the finding that Gaussian kernels produce a stronger illusion should be taken with a pinch of salt since larger areas are often perceived as brighter [Li and Gilchrist 1999]. It is therefore possible, that the apparent advantage of the Gaussian method is due to the increase of the size of the glare source rather than an advantage of the Gaussian method per se. We therefore conclude, that both Methods produce a comparable increase in perceived luminance when reasonable parameters are chosen. This is interesting also from a practical point of view, since a convolution with a separable Gaussian kernel is much faster than in case of non-separable kernels required for the eye's PSF.

Yet, there are differences in how the two Methods behave in terms of potential side-effects. While the Gaussian method is relatively susceptible to distort the shape and size of the convolved object, Spencer et al.'s method is more robust regarding the choice of the parameter and therefore less likely to produce this effect (even though it does change the disk shape with growing L_{disk} , see Figure 7.5). On the other hand, Spencer et al.'s method is more likely to excite a Mach-band effect, which is often perceived as objectionable. This might be caused by the steeper gradient in the glare image rendered with Spencer et al.'s method (see scanlines in Figure 7.5) as shown in [Ratliff 1965, pp. 85].

It is interesting to note that models of the optics in the human eye [Stiehl et al. 1983, Spencer et al. 1995] do not outperform the simple Gaussian convolution approach in terms of pure effectiveness of boosting the perceived luminance. These results allow the speculation that the scattering light in human's eye is not the only factor in the Human Visual Systems (HVS) that contributes to elicit the glare illusion. Possibly, neural centers later in the visual pathway that are not captured in this type of model contribute to the perception of glare sources. Under this assumption, a model for rendering glare sources based purely on empirical evidence could be more appropriate.

7.8 Summary

If rendered properly, the glare illusion can increase the perceived luminance and therefore also the dynamic range of a display by 20–35%. Although the glare illusion is believed to be related to optical distortions in the eye, our experiment indicates that faithful simulation of the eye's optics is not necessary to achieve a strong brightness boost. The glare illusion produced by a Gaussian convolution can give the same increase of perceived luminance as a complex PSF of the eye, is less likely to cause undesirable Mach-band effects and is faster to render. On the other hand, the spiky profile of the eye's PSF does not change the object's shape and size as much as the Gaussian kernel.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

The work presented in this dissertation is mainly focused on evaluation of tone mapping operators (TMOs) and enhancement of contrast and brightness for high dynamic range (HDR) images. Since the human visual system (HVS) exhibits strongly non-linear behavior across such wide range of luminance, we often have to run psychophysical experiments ourselves to understand better the HVS characteristics in the context of HDR image perception.

As a solution for displaying HDR images on low dynamic range (LDR) display devices, a number of TMOs have been developed, however, there was no systematic evaluation to show how tone mapped images are perceived differently by human observers, and which attributes of image appearance take into account the difference between TMOs yet. We conducted a psychophysical experiment with seven TMOs which were not directly compared to each other as in other studies. The tone mapped images were compared and rated against their corresponding real-world views at the position where the HDR image was shot. Our result shows that qualitative differences between TMOs have a systematic effect on the human perception. However, it also turns out that it is hard to select one TMO which performs consistently the best in terms of image fidelity because those TMOs are too complicated to clarify a relationship among TMO parameter settings, an input HDR image, and the final tone mapped images (Chapter 4).

After this result, we went back to a simple setting of a generic TMO which employs the three most important parameters: brightness, contrast, and color saturation.

tion of an image. A series of image preference and fidelity experiments have been conducted with several types of limited dynamic range of displays, in which the human subjects adjusted those parameters, to better understand the performance of users on TMO parameter settings. The results show that this generic TMO is strongly affected by two factors: anchor white and contrast. Subjects set reference white (anchor white) to a bit lower value than the maximum luminance of an image, if the image contains large self-luminous objects. Unlike anchor white, contrast factor is more subjective, therefore users should have a control to adjust it. It is also shown that the parameters can be automatically estimated based on the characteristics of an image for best-guessing results. For example, contrast can be predicted based on the dynamic range of an image, and the anchor white has a relation to the image key value. In addition, the outcome from emulating several types of limited dynamic range of displays depicts that the resulting images depend on the purpose of the TMO: the best-looking (preference task) or the best fidelity task. These results of a simple and fundamental TMO will be applicable for more complex TMOs (Chapter 5).

The above two studies focused on evaluation of image appearance in HDR. We also studied the enhancement of contrast and brightness of an image in HDR. A usual way to scale contrast in image processing such as gamma correction gives a constant change of contrast in the whole image, however, such simple scaling leads to non-uniform perceived change in contrast because of the lower contrast sensitivity of the human eyes for the low luminances. Based on this fact and HDR display technology which can reproduce much lower luminance than that of conventional LDR displays, we conducted two perceptual experiments of perceived contrast over a complex image. Perceived contrast scaling was presented with respect to given physical contrast and different adaptation luminances. The results of contrast scaling experiment were converted to just noticeable difference (JND) units to construct a model to provide uniformly changing perceived contrast in complex images (Chapter 6).

Finally, we measured the brightness enhancement caused by the glare illusion. We observe that an object in an image looks much brighter, if it is surrounded by smooth profiles, and we call it the glare illusion. To evoke the glare illusion, we employed two convolution methods: a point spread function (PSF) of the human eye and a Gaussian kernel. The outcome of this work shows that the glare illusion increases the perceived luminance (brightness) by 20 – 35 %. It indicates that the glare illusion can visually expand the maximum luminance of a display device by 20 – 35 %. This result was obtained for both convolution methods of Gaussian kernel and PSF. This means that, although it is believed the glare illusion is related to some optical systems, faithful simulation of the human eyes is not necessary to achieve a strong brightness boost caused by the glare illusion because the Gaus-

sian kernel which has no theoretical background of human perception evokes the brightness boosts of similar strength as the PSF (Chapter 7).

8.2 Future Work

For human visual perception in high dynamic range imaging (HDRI), there are a number of unsolved problems. As we have investigated TMOs in Chapters 4 and 5, there is still no TMO which perform the best consistently in terms of image fidelity. Additionally, an optional parameter setting may depend on the contents of an image for all image classification. Automatic selection of the best TMO and its parameter settings for a given image is an open question.

There are also a number of improvements and extensions for enhancement of contrast and brightness in an image. In contrast enhancement study, we plan to extend our experiments with more representative group of test images. Then, the revised contrast enhancement model for HDR images as proposed in this dissertation should be then even more robust. Additionally, how to define contrast in complex images by a single number in an absolute unit is still a big open question.

In the study of brightness boost by the glare illusion, we have proposed a fundamental approach to investigate the glare illusion. Therefore, there are still several factors which were not measured in this dissertation, for example, the extent of the glare profile and the size of the object. A model which includes all of these factors would allow to render the glare illusion with desired strength and appearance.

Bibliography

- [Adelson and Burt 1981] ADELSON, E. H., AND BURT, P. J. 1981. Image data copression with the laplacian pyramid. In *Proc. of the Pattern Recognition and Information Processing Conference*, IEEE, 218–223.
- [Adelson 1982] ADELSON, E. H. 1982. Saturation and adaptation in the rod system. *Vision Research* 22, 1299–1312.
- [Adelson 1993] ADELSON, E. H. 1993. Perceptual organization and the judgement of brightness. *Science* 262, pp. 2042 – 2044.
- [Ashikhmin and Goral 2006] ASHIKHMIN, M., AND GORAL, J. 2006. A reality check for tone mapping operators. *ACM Transactions on Applied Perception* 3, 4, pp. 399 – 411.
- [Ashikhmin 2002] ASHIKHMIN, M. 2002. A tone mapping algorithm for high contrast images. In *13th Eurographics Workshop on Rendering*, P. Debevec and S. Gibson, Eds., The Eurographics Association, 145–155.
- [Atkinson 1988] ATKINSON, R., Ed. 1988. *Steven’s Handbook of Experimental Psychology*, 2nd ed. John Wiley & Sons, New York.
- [Baker 1949] BAKER, H. 1949. The course of foveal adaptation measured by the threshold intensity increment. *Journal of the Optical Society of America* 39, 172–179.
- [Barten 1999] BARTEN, P. G. 1999. *Contrast sensitivity of the human eye and its effects on image quality*. SPIE – The International Society for Optical Engineering. ISBN 0-8194-3496-5.
- [Bartleson and Breneman 1967] BARTLESON, C., AND BRENEMAN, E. 1967. Brightness reproduction in the photographic process. In *Photographic Science and Engineering*, vol. 11, 254–262.

- [Bartleson and Grum 1984] BARTLESON, C., AND GRUM, F. 1984. *Optical radiation measurements*, vol. 5 of *Visual Measurements*. Academic Press. pp. 467–471.
- [Bloch 2007] BLOCH, C. 2007. *The HDRI Handbook: high dynamic range imaging for photographers and CG artists*. Rocky Nook Inc.
- [Borg and Groenen 1997] BORG, I., AND GROENEN, P. 1997. *Modern Multidimensional Scaling: Theory and Applications*. Springer-Verlag, New York.
- [Čadík et al. 2006] ČADÍK, M., WIMMER, M., NEUMANN, L., AND ARTUSI, A. 2006. Image attributes and quality for evaluation of tone mapping operators. In *Proceedings of the 14th Pacific Conference on Computer Graphics and Applications*, pp. 35 – 44.
- [Campbell and Robson 1968] CAMPBELL, F., AND ROBSON, J. 1968. Application of fourier analysis to the visibility of gratings. *Journal of Physiology* 197, pp. 551 – 566.
- [Chiu et al. 1993] CHIU, K., HERF, M., SHIRLEY, P., SWAMY, S., WANG, C., AND ZIMMERMAN, K. 1993. Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface '93*, Canadian Information Processing Society, Toronto, Ontario, Canada, 245–253.
- [Choudhury and Tumblin 2003] CHOUDHURY, P., AND TUMBLIN, J. 2003. The trilateral filter for high contrast images and meshes. In *Proc. of the 13th Eurographics workshop on Rendering*, Eurographics Association, ACM International Conference Proceeding Series, 186–196.
- [Coolican 2004] COOLICAN, H. 2004. *Research methods and statistics in psychology*, 4th ed. Hodder & Stoughton.
- [Craik 1966] CRAIK, K. 1966. *The Nature of Psychology*. Cambridge University Press.
- [Daly 1993] DALY, S. 1993. The visible differences predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, MIT Press, A. B. Watson, Ed., 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.
- [Debevec 1998] DEBEVEC, P. E. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM SIGGRAPH 98*.
- [Delahunt et al. 2005] DELAHUNT, P., ZHANG, X., AND BRAINARD, D. 2005. Perceptual image quality: Effects of tone characteristics. *Journal of Electronic Imaging* 14, 2, 1–12.
- [deRidder 1996] DERIDDER, H. 1996. Naturalness and image quality: saturation and lightness variation in color images of natural scenes. *The Journal of Imaging Science and Technology* 40, 6, 487–493.
- [Devlin et al. 2002] DEVLIN, K., CHALMERS, A., WILKIE, A., AND PURGATHOFER, W. 2002. Tone reproduction and physically based spectral rendering. In *State of the Art Reports, Eurographics*, D. Fellner and R. Scopigno, Eds., The Eurographics Association, 101–123.
- [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 Eurographics*, P. Brunet and D. Fellner, Eds., 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.

- [Engeldrum 2000] ENGELDRUM, P. G. 2000. *Psychometric Scaling: A toolkit for imaging systems development*. Imcotek Press, Winchester, MA, USA.
- [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*, 249–256.
- [Fedorovskaya et al. 1997] FEDOROVSKAYA, E., DERIDDER, H., AND BLOMMAERT, F. 1997. Chroma variations and perceived quality of color images of natural scenes. *Color Research and Application* 22, 2, 96–110.
- [Ferwerda et al. 1996] FERWERDA, J. A., PATTANAİK, 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.
- [Ferwerda 2001] FERWERDA, J. A. 2001. Elements of early vision for computer graphics. *IEEE Computer Graphics and Applications* 21, 5, 22–33.
- [Foley and Legge 1981] FOLEY, J. M., AND LEGGE, G. E. 1981. Contrast detection and near-threshold discrimination in human vision. *Vision Research* 21, pp. 1041 – 1053.
- [Gerrits and Vendrik 1970] GERRITS, H., AND VENDRIK, A. 1970. Simultaneous contrast, filling-in process and information processing in man’s visual system. *Experimental Brain Research* 11, 4, pp. 411–430.
- [Gescheider 1997] GESCHEIDER, G. A. 1997. *Psychophysics: The Fundamentals*, 3rd ed. Lawrence Erlbaum Associates, Inc. ISBN 0-8058-2281-X.
- [Gilchrist et al. 1999] GILCHRIST, A. L., 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.

- [Gottesman et al. 1981] GOTTESMAN, J., RUBIN, G. S., AND LEGGE, G. E. 1981. A power law for perceived contrast in human vision. *Vision Research* 21, 791 – 799.
- [Graham 1965] GRAHAM, C. H., Ed. 1965. *Vision and Visual Perception*. John Wiley & Sons, Inc.
- [Holm 1979] HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- [Jobson et al. 2002] JOBSON, D., RAHMAN, Z., AND WOODDELL, G. 2002. The statistics of visual representation. In *Visual Information Processing XI*, SPIE, volume 4736, 25–35.
- [Johnson 2005] JOHNSON, G. M. 2005. Cares and concerns of CIE TC8-08: spatial appearance modeling and HDR rendering. In *Image Quality and System Performance II*, SPIE, volume 5668, 148–156.
- [Jones and Condit 1941] JONES, L. A., AND CONDIT, H. 1941. The brightness scale of exterior scenes and the computation of correct photographic exposure. *Journal of the Optical Society of America* 31.
- [Kakimoto et al. 2005] KAKIMOTO, M., MATSUOKA, K., NISHITA, T., NAE-MURA, T., AND HARASHIMA, H. 2005. Glare simulation and its application to evaluation of bright lights with spectral power distribution. In *Proceedings of ACM SIGGRAPH*.
- [Kawase 2005] KAWASE, M. 2005. Practical implementation of high dynamic range rendering. In *Game Developers Conference*.
- [Kingdom and Whittle 1996] KINGDOM, F. A. A., AND WHITTLE, P. 1996. Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing. *Vision Research* 36, 6, 817 – 829.
- [Kling and Riggs 1971] KLING, J. W., AND RIGGS, L. A., Eds. 1971. *Woodworth and Schlosberg's Experimental Psychology*, 3rd ed. Holt, Rinehart, and Winston, New York.
- [Krawczyk et al. 2005] KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Lightness perception in tone reproduction for high dynamic range images. In *Proceedings of EUROGRAPHICS 2005 (Computer Graphics Forum vol. 25)*, M. Alexa and J. Marks, Eds., vol. 24.

- [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*, 315–320.
- [Kuang et al. 2005] KUANG, J., JOHNSON, G. M., AND FAIRCHILD, M. D. 2005. Image preference scaling for hdr image rendering. In *Proc. of the 13th Color Imaging Conference*, 8–13.
- [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*.
- [Legge and Foley 1980] LEGGE, G. E., AND FOLEY, J. M. 1980. Contrast masking in human vision. *Journal of Optical Society of America* 70, 12 (December), pp. 158 – 1470.
- [Legge 1980] LEGGE, G. E. 1980. A power law for contrast discrimination. *Vision Research* 21, pp. 457 – 467.
- [Li and Gilchrist 1999] LI, X., AND GILCHRIST, A. 1999. Relative area and relative luminance combine to anchor surface lightness values. *Perception & Psychophysics* 61, 5, pp. 771 – 785.
- [Lotto et al. 1999] LOTTO, R. B., WILLIAMS, S. M., AND PURVES, D. 1999. An empirical basis for mach bands. In *Proceedings of National Academy of Sciences of the United States of America*, pp. 5239–5244.
- [Mann and Picard 1995] MANN, S., AND PICARD, R. W. 1995. On being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures. In *IS&T's 48th Annual Conference*, Society for Imaging Science and Technology, 422–428.
- [Mannos and Sakrison 1974] MANNOS, J., AND SAKRISON, D. 1974. The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory* 20, 3, pp. 525 – 536.
- [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, 286 – 308.

- [MathWorks, Inc.] MATHWORKS, INC. MATLAB.
http://www.mathworks.com/.
- [Miller and Hoffman 1984] MILLER, G., AND HOFFMAN, C. 1984. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Course Notes for Advanced Computer Graphics Animation*.
- [Moon and Spencer 1945] MOON, P., AND SPENCER, D. 1945. The visual effect of non-uniform surrounds. *Journal of Optic Society of America* 35, 3, pp. 233 – 248.
- [Murchison 1934] MURCHISON, C., Ed. 1934. *A Handbook of General Experimental Psychology*. Clark University Press.
- [Nakamae et al. 1990] NAKAMAE, E., KANEDA, K., OKAMOTO, T., AND NISHITA, T. 1990. A lighting model aiming at drive simulators. *Computer Graphics* 24, 4, pp. 395 – 404.
- [Nayar and Mitsunaga 2000] NAYAR, S. K., AND MITSUNAGA, T. 2000. High dynamic range imaging: spatially varying pixel exposures. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*.
- [Newsham et al. 2002] NEWSHAM, G. R., SEETZEN, H., AND VEITCH, J. A. 2002. Lighting quality evaluations using images on a high dynamic range display. In *Proc. of the ARCC/EAAE Conference on Architectural Research*, 1–9.
- [Nijenhuis et al. 1997] NIJENHUIS, M., HAMBERG, R., TEUNISSEN, C., BECH, S., DE JONG, H. L., HOUBEN, P., AND PRAMANIK, S. K. 1997. Sharpness, sharpness related attributes, and their physical correlates. In *Proc. of SPIE*, vol. 3025, 173–184.
- [O’Brien 1959] O’BRIEN, V. 1959. Contrast by contour-enhancement. *American Journal of Psychology* 72, pp. 299 – 300.
- [Pattanaik and Yee 2002] PATTANAİK, S., AND YEE, H. 2002. Adaptive gain control for high dynamic range image display. In *Proc. of ACM SIGGRAPH 2000*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, Computer Graphics Proceedings, Annual Conference Series, 55–64.
- [Pattanaik et al. 2000] PATTANAİK, 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 SIG-*

- GRAPH 2000*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, Annual Conference Series, 47–54. ISBN 1-58113-208-5.
- [Peli et al. 1991] PELI, E., YANG, J., GOLDSTEIN, R., AND REEVES, A. 1991. Effect of luminance on suprathreshold contrast perception. *Journal of the Optical Society of America A* 8, 8 (August), 1352 – 1359.
- [Peli et al. 1996] PELI, E., AREND, L., AND LABIANCE, A. T. 1996. Contrast perception across changes in luminance and spatial frequency. *Journal of the Optical Society of America A* 13, 10 (October), pp. 1953 – 1959.
- [Peli 1990] PELI, E. 1990. Contrast in complex images. *Journal of the Optical Society of America A* 7, 10 (October), pp. 2032 – 2040.
- [Peli 1995] PELI, E. 1995. Suprathreshold contrast perception across differences in mean luminance: effects of stimulus size, dichoptic presentation, and length of adaptation. *Journal of the Optical Society of America A* 12, 5 (May), pp. 817 – 823.
- [Perlin and Hoffert 1989] PERLIN, K., AND HOFFERT, E. 1989. Hypertexture. In *Proc. of ACM SIGGRAPH 89*, vol. 23, 253–262.
- [Pratt 1991] PRATT, W. K. 1991. *Digital image processing*, 2nd ed. John Wiley & Sons, Inc., New York, NY, USA. ISBN 0-471-85766-1.
- [Purves et al. 1999] PURVES, D., SHIMPI, A., AND LOTTO, R. B. 1999. An empirical explanation of the cornsweet effect. *Journal of Neuroscience* 19, 19, pp. 8542 – 8551.
- [Ratliff 1965] RATLIFF, F. 1965. *MACH BANDS: Quantitative studies on neural networks in the retina*. Holden-day, Inc.
- [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, pp. 13 – 24.
- [Reinhard et al. 2002] REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. In *SIGGRAPH 2002 Conference Proceedings*, Addison Wesley, ACM SIGGRAPH.

- [Reinhard et al. 2005] REINHARD, E., WARD, G., PATTANAİK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann.
- [Reinhard et al. 2007] REINHARD, E., KHAN, E. A., AKYÜZ, A. O., AND JOHNSON, G. 2007. *Color Imaging: fundamentals and applications*. A K Peters, Ltd.
- [Reinhard 2003] REINHARD, E. 2003. Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools* 7, 1 (January), 45–51.
- [Ritschel et al. 2008] RITSCHHEL, T., SMITH, K., IHRKE, M., GROSCH, T., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2008. 3d unsharp masking for scene coherent enhancement. In *Proceedings of ACM SIGGRAPH (Transaction on Graphics)*, vol. 27.
- [Robertson et al. 1999] ROBERTSON, M. A., BORMAN, S., AND STEVENSON, R. L. 1999. Dynamic range improvement through multiple exposures. *IEEE International Conference on Image Processing* (October).
- [Rokita 1993] ROKITA, P. 1993. A model for rendering high intensity lights. *Computers and Graphics* 17, 4, pp. 85 – 108.
- [Savakis et al. 2000] SAVAKIS, A., ETZ, S., AND LOUI, A. 2000. Evaluation of image appeal in consumer photography. In *Human Vision and Electronic Imaging V*, SPIE, volume 3959, 111–120.
- [Schlick 1994] SCHLICK, C. 1994. Quantization techniques for the visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*, Springer-Verlag Berlin Heidelberg New York, P. S. G. Sakas and S. Müller, Eds., Eurographics, 7–20.
- [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*.
- [Simpson 1953] SIMPSON, G. C. 1953. Ocular haloes and coronas. *British Journal of Ophthalmology* 37, pp. 450 – 486.
- [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 *Proc. of Eurographics 2006*.

- [Spencer et al. 1995] SPENCER, G., SHIRLEY, P., ZIMMERMAN, K., AND GREENBERG, D. P. 1995. Physically-based glare effects for digital images. In *Proceedings of ACM SIGGRAPH*, ACM.
- [Spillman et al. 1990] SPILLMAN, L., WERNER, J. S., AND SPILLMAN, L. 1990. *Visual Perception: The Neurophysiological Foundations*. Academic Press.
- [Stiehl et al. 1983] STIEHL, W. A., MCCANN, J. J., AND SAVOY, R. L. 1983. Influence of intraocular scattered light on lightness-scaling experiments. *Journal of Optical Society of America* 73, 9, pp. 1143–1148.
- [Tabachnick 1989] TABACHNICK, B. G. 1989. *Using multivariate statistics*, 2nd ed. Harper Collins Publishers, Inc. ISBN 0-06-046571-9.
- [Taylor and Creelman 1967] TAYLOR, M. M., AND CREELMAN, C. D. 1967. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America* 41, 4, 782 – 787.
- [Thurstone 1927] THURSTONE, L. L. 1927. A law of comparative judgment. *Psychological Review* 34, 273–286.
- [Thurstone 1967] THURSTONE, L. 1967. *The measurement of values*. University of Chicago Press.
- [Torgerson 1958] TORGERSON, W. S. 1958. *Theory and methods of scaling*. John Wiley & Sons, Inc., New York.
- [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, 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.

- [Tumblin 1999] TUMBLIN, J. E. 1999. *Three Methods For Detail-Preserving Contrast Reduction For Displayed Images*. PhD thesis, Georgia Institute of Technology.
- [van den Berg et al. 2005] VAN DEN BERG, T. J. T. P., HAGENOUW, M. P. J., AND COPPENS, J. E. 2005. The ciliary corona: Physical model and simulation of the fine needles radiating from point light sources. *Investigative Ophthalmology & Visual Science* 46, 7.
- [Vos 1984] VOS, J. 1984. Disability glare—a state of the art report. *CIE Journal* 3, 2, 39–53.
- [Vos 2003] VOS, J. 2003. On the cause of disability glare and its dependence on glare angle, age and ocular pigmentation. *Clin. Exp. Optom* 86, 363–370.
- [Wandell 1995] WANDELL, B. A. 1995. *Foundations of Vision*. Sinauer Associates, Inc.
- [Ward Larson et al. 1997] WARD LARSON, 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, pp. 291–306.
- [Ward 1991] WARD, G. 1991. *Graphics Gem II*. Academic Press, ch. Real Pixels, pp. 80 – 83.
- [Ward 1994] WARD, G. 1994. A contrast-based scalefactor for luminance display. *Graphics Gems IV*, 415–421.
- [Ward 2003] WARD, G. 2003. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools* 8, 2, 17–30.
- [Watson and Pelli 1983] WATSON, A., AND PELLI, D. G. 1983. QUEST: A bayesian adaptive psychometric method. *Perception & Psychophysics* 33, 2, pp. 113 – 120.
- [Watson et al. 2001] WATSON, B., FRIEDMAN, A., AND MCGAFFEY, A. 2001. Measuring and predicting visual fidelity. In *Proc. of ACM SIGGRAPH 2001*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, Computer Graphics Proceedings, Annual Conference Series, 213–220.

- [Whittle 1986] WHITTLE, P. 1986. Increments and decrements: luminance discrimination. *Vision Research* 26, 10, 1677 – 1691.
- [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 *Proceedings of Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium Electronic Imaging*, IS&T/SPIE, pp. 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 *Proc. of Eurographics 2006*.
- [Yoshida et al. 2007a] YOSHIDA, A., BLANZ, V., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2007. Testing tone mapping operators with human-perceived reality. *Journal of Electronic Imaging* 16, 1.
- [Yoshida et al. 2007b] YOSHIDA, A., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2007. Perceptual uniformity of contrast scaling in complex images. In *Proc. of ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*, pp. 137.
- [Yoshida et al. 2008a] YOSHIDA, A., IHRKE, M., MANTIUK, R., AND SEIDEL, H.-P. 2008. Brightness of the glare illusion. In *Proc. of ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*.
- [Yoshida et al. 2008b] YOSHIDA, A., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2008. Perception-based contrast enhancement model for complex images in high dynamic range. In *Proc. of Human Vision and Electronic Imaging XIII, IS&T/SPIE's 20th Annual Symposium on Electronic Imaging*, IS&T/SPIE.
- [Yoshida 2004] YOSHIDA, A. 2004. *Perceptual Evaluation of Tone Mapping Operators*. Master's thesis, Universität des Saarlandes.
- [Zavagno and Caputo 2001] ZAVAGNO, D., AND CAPUTO, G. 2001. The glare effect and the perception of luminosity. *Perception* 30, pp. 209–222.

- [Zavagno 1999] ZAVAGNO, D. 1999. Some new luminance-gradient effects.
Perception 28, pp. 835 – 838.

Chapter 9

Curriculum Vitae

Curriculum Vitae

1978	Born in Karatsu, Saga, Japan
1993 – 1996	Karatsu Higashi High School, Karatsu, Saga, Japan
1996 – 2000	B.S. in Computer Science and Engineering at the University of Aizu, Aizuwakamatsu, Fukushima, Japan
2000 Apr. - Sep.	Research student at the University of Aizu Aizuwakamatsu, Fukushima, Japan
2000 – 2002	Graduate School of Computer Science, Oregon State University, Corvallis, OR, USA
2002 – 2004	M.Sc. in Computer Science at Universität des Saarlandes, Saarbrücken, Germany
2004 – 2008	Ph.D. candidate at Max-Planck-Institut für Informatik, Saarbrücken, Germany