

Tone-Mapping High Dynamic Range Images and Videos for Bit-Depth Scalable Coding and 3D Displaying

by

Zicong Mai

M.A.Sc., The University of British Columbia, Canada, 2007

B.Eng., North China University of Technology, China, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Electrical & Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

July 2012

© Zicong Mai, 2012

Abstract

High dynamic range (HDR) images and videos provide superior picture quality by allowing a larger range of brightness levels to be captured and reproduced than their traditional 8-bit low dynamic range (LDR) counterparts. Even with existing 8-bit displays, picture quality can be significantly improved if the content is first captured in HDR format and then converted to LDR format. This converting process is called tone-mapping. In this thesis, we address different aspects of tone-mapping.

HDR video formats are unlikely to be broadly accepted without the backward-compatibility with LDR devices. We first consider the case where only the tone-mapped LDR content is transmitted and the HDR video is reconstructed at the receiver by inversely tone-mapping the encoded-decoded LDR video. We show that the appropriate choice of a tone-mapping operator can result in a reconstructed HDR video with good quality. We develop a statistical model of the distortion resulting from tone-mapping, compressing, de-compressing and inverse tone-mapping the HDR video. This model is used to formulate an optimization problem that finds the tone-curve that minimizes the distortion in the reconstructed HDR video. We also derive a simplified version of the model that leads to a closed-form solution for the optimization problem.

Next, we consider the case where the HDR content is transmitted using an LDR and an enhancement layers. We formulate an optimization problem that minimizes the transmitted bit-rate of a video sequence and also results in the tone-mapped video that satisfies some desired perceptual appearance. The problem formulation also contains a constraint that suppresses temporal flickering artifacts.

We also propose a technique that tone-maps an HDR video directly in a compression friendly color space (e.g., YCbCr) without the need to convert it to the RGB domain.

We study the design of 3D HDR-LDR tone-mapping operators. To find the appropriate tone-mapping characteristics that contribute to good 3D representation, subjective psychophysical experiments are performed for i) evaluating existing tone-mapping operators on 3D HDR images and ii) investigating how the preferred level of brightness and details differ between 3D and 2D images. The results are analyzed to find out the desired attributes.

Preface

This thesis presents research conducted by Zicong Mai, in collaboration with Dr. Panos Nasiopoulos, Dr. Rabab K. Ward, Dr. Wolfgang Heidrich, Dr. Rafał Mantiuk, Dr. Hassan Mansour, and Colin Doutre.

[P1], [P2]: Zicong Mai was the primary author of these papers and the main contributor to the identification, design, development, and testing of the presented methods under the supervision of Dr. Panos Nasiopoulos and Dr. Rabab K. Ward. Dr. Rafał Mantiuk and Dr. Hassan Mansour provided suggestions and feedback. Dr. Panos Nasiopoulos, Dr. Rabab K. Ward and Dr. Wolfgang Heidrich provided guidance and editorial input into the creation of the papers.

[P3], [P4]: Zicong Mai was the primary author of these papers and the main contributor to the identification, design, development, and testing of the presented methods under the supervision of Dr. Panos Nasiopoulos and Dr. Rabab K. Ward. Dr. Hassan Mansour provided suggestions and feedback. Dr. Panos Nasiopoulos and Dr. Rabab K. Ward also provided guidance and editorial input into the creation of the papers.

[P5], [P6]: Zicong Mai was the primary author of the paper and the main contributor to the identification, design, development, and testing of the presented method under the supervision of Panos Nasiopoulos and Dr. Rabab K. Ward. They also provided suggestions, feedback and editorial input towards the creation of the paper.

[P7], [P8]: Zicong Mai was the primary author of these papers and the main contributor to the identification, design, development, data acquisition and data analysis of the research project under the supervision of Panos Nasiopoulos and Dr. Rabab K. Ward. Colin Doutre provided inputs of the 3D aspect in the data acquisition,

experimental design and manuscript editing. Dr. Panos Nasiopoulos and Dr. Rabab K. Ward also provided suggestions, feedback and editorial guidance towards the creation of the paper.

Chapter 2

- [P1] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, W. Heidrich, "Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image/Video Compression," *IEEE Transactions on Image Processing (TIP)*, vol. 20, no. 6, pp. 1558-1571, June 2011.
- [P2] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward and W. Heidrich, "On-the-Fly Tone-Mapping for Backward-Compatible High Dynamic Rang Image/Video Compression," in *Proc. IEEE International Symposium on Circuit and Systems 2010 (ISCAS2010)*, pp. 1831-1834, May 2010.

Chapter 3

- [P3] Z. Mai, H. Mansour, P. Nasiopoulos and R. Ward, "Visually-Favorable Tone-Mapping with High Compression Performance in Bit-Depth Scalable Video Coding," *IEEE Transactions on Multimedia (TMM)*, submitted, January 2012.
- [P4] Z. Mai, H. Mansour, P. Nasiopoulos and R. Ward, "Visually-Favorable Tone-Mapping with High Compression Performance," in *Proc. IEEE International Conference on Image Processing 2010 (ICIP2010)*, pp. 1285-1288, Sep. 2010.

Chapter 4

- [P5] Z. Mai, P. Nasiopoulos and R. Ward, "Computationally Efficient Tone-Mapping of High-Bit-Depth Video in the YCbCr Domain," *The 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2012)*, to appear, March 2012.
- [P6] *US Patent 61/610734*, "Efficient tone-mapping of high-bit-depth video to low-bit-depth displays", Zicong Mai, Panos Nasiopoulos and Rabab K. Ward, *US Provisional Patent Application*, assignee Dolby Laboratories Inc., filed March 14 2012.

Chapter 5

- [P7] Z. Mai, C. Doutre, P. Nasiopoulos, and R. K. Ward, "Rendering 3D High Dynamic Range Images: Subjective Evaluation of Tone-Mapping Methods and Preferred 3D Image Attributes," *IEEE Journal of Selected Topics in Signal Processing (JSTSP)*, March 2012, accepted for publication.
- [P8] Z. Mai; C. Doutre, P. Nasiopoulos, and R. K. Ward, "Subjective Evaluation of Tone-Mapping Methods on 3D Images," in *Proc. 17th IEEE International Conference on Digital Signal Processing (DSP2011)*, pp.1-6, July 2011.

Table of Contents

Abstract.....	ii
Preface.....	iv
Table of Contents	vi
List of Tables	ix
List of Figures.....	x
List of Abbreviations	xii
Acknowledgements	xiii
Dedication	xv
1 Introduction and Overview	1
1.1 Introduction.....	1
1.2 Generating and Displaying High Dynamic Range Content.....	4
1.2.1 Generation of HDR Content.....	4
1.2.2 Display of HDR Content.....	7
1.3 Encoding High Dynamic Range Content.....	9
1.3.1 Single-Layer HDR Image/Video Compression.....	9
1.3.2 Backward-Compatible HDR Image/Video Compression.....	10
1.4 Tone-Mapping High Dynamic Range Content.....	13
1.4.1 Tone-Mapping Operators.....	13
1.4.2 Temporal Inconsistency Correction for Tone-Mapping	15
1.4.3 Color Reproduction for Tone-Mapping	16
1.4.4 Evaluation of Tone-Mapping Operators	18
1.5 3D Visual Experience	20
1.5.1 3D Display Technologies.....	20
1.5.2 Generation of 3D HDR Content	20
1.5.3 3D Quality of Experience	21
1.6 Thesis Motivation and Objectives	23
1.6.1 Thesis Motivation	23
1.6.2 Thesis Objectives.....	25
1.7 Thesis Contributions.....	27
1.8 Thesis Organization	29
2 Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image/Video Compression	32
2.1 Problem Statement.....	33
2.2 Proposed Solution.....	36
2.2.1 Tone Mapping Curve	36
2.2.2 Statistical Distortion Model	38
2.2.3 Optimization Problem.....	39
2.2.4 Closed-Form Solution.....	40

2.3 Experimental Results and Discussion	43
2.3.1 Model Validation.....	43
2.3.2 Dependence of the Tone Curves on QP.....	45
2.3.3 Further Analysis of the Closed-Form Solution	47
2.3.4 Comparison with Existing TMOs.....	49
2.3.5 Adapting Tone Curves for JVT Bit-Depth Scalable Encoding	56
2.4 Discussion.....	58
2.5 Conclusion	59
3 Visually Favorable Tone-Mapping with High Compression Performance in Bit-Depth Scalable Video Coding	60
3.1 Problem Statement.....	60
3.2 Proposed Solution.....	63
3.2.1 Tone-Mapping Parameters.....	63
3.2.2 LDR Quality Mismatch QL	64
3.2.3 HDR Bit-Rate Re	65
3.2.4 LDR Bit-Rate Rb	67
3.2.5 Temporal Consistency between Consecutive Frames	70
3.2.6 Optimization Formulation.....	71
3.3 Experimental Results and Discussion.....	72
3.3.1 Validation of the Models QL and Rb	73
3.3.2 Validation of the Model Re	75
3.3.3 Compression Performance	76
3.4 Conclusions.....	83
4 Color Correction of Tone-Mapping Directly in the YCbCr Domain.....	84
4.1 Problem Statement.....	85
4.2 Proposed Solution.....	88
4.2.1 Derivation of the Proposed Closed Form.....	88
4.2.2 Derivation of Parameters for Different Standards.....	91
4.3 Results and Discussion	92
4.4 Conclusion	94
5 Rendering High Dynamic Range 3D Images: Subjective Evaluation of Tone-Mapping Methods and Preferred 3D Image Attributes	96
5.1 Tone-Mapping Operators Evaluated.....	97
5.2 Experiment One – Evaluating Tone-Mapping Operators.....	101
5.2.1 Image Preparation.....	101
5.2.2 Experimental Framework.....	102
5.2.3 Results and Analysis.....	105
5.3 Experiment Two – 3D Effects on Brightness and Details.....	111
5.3.1 Image Preparation.....	111
5.3.2 Testing Environment and Procedures.....	116
5.3.3 Results and Analysis.....	118
5.4 Discussion.....	122
5.5 Conclusion	126

6	Thesis Summary	127
6.1	Significance of the Research	127
6.2	Potential Applications	129
6.3	Contributions.....	130
6.4	Suggestions for Future Research.....	133
	Bibliography	135
	Appendix A – H.264/AVC Intra Coding Error Model (p_C)	144

List of Tables

Table 4.1 Average performance of the proposed approach	93
--	----

List of Figures

Fig. 1.1: Demonstration of the visual difference of LDR images.....	3
Fig. 1.2: A scene taken at different exposures.....	5
Fig. 1.3: The LED layer (a) and the Dolby prototype (b) of the high dynamic range display.....	8
Fig. 1.4: System structure of backward-compatible HDR video compression proposed by H.264/AVC.....	12
Fig. 1.5: Demonstration of the difference in visual effects using different TMOs.....	13
Fig. 1.6: Conventional approaches to tone-map color images.....	17
Fig. 2.1: System overview of the proposed tone-mapping method.....	35
Fig. 2.2: Parameterization of a tone-mapping curve and the notation. The bar-plot in the background represents an image histogram used to compute $p(l)$	37
Fig. 2.3: Validation of the proposed models by comparison with the groundtruth solution.....	44
Fig. 2.4: Tone curves generated using the statistical model with different QP values for the images “AtriumNight” and “Desk”.....	46
Fig. 2.5: Distortion measures for the reconstructed HDR images using the generalized solution (see (2.16)) with different values of t , averaged over 40 images.....	48
Fig. 2.6: Comparison with other tone-mapping methods in terms of MSE and SSIM (for the reconstructed HDR image) vs. bit rate, averaged over 40 images.....	50
Fig. 2.7: Rate-distortion curves, tone curves and tone-mapped images for the image “Coby”.....	52
Fig. 2.8: Rate-distortion curves, tone curves and tone-mapped images for the image “AtriumNight”.....	53
Fig. 2.9: Rate-distortion curves, tone curves and tone-mapped images for the image “BristolBridge”.....	54
Fig. 2.10: Distortion maps of the LDR images relative to the original HDR images.....	55
Fig. 3.1: Parameterization of a tone-curve.....	63
Fig. 3.2: Correlation between the enhancement layer bit-rate and MSE.....	66
Fig. 3.3: Correlation between the temporal bit-rate and difference in average brightness between consecutive tone-mapped frames.....	69
Fig. 3.4: Demonstration of the change of base layer bit rate, PSNR, and SSIM versus base as the weights of QL and Rb vary.....	74
Fig. 3.5: Comparison of bit rate of the enhancement layer among tone-mapping schemes.....	75

Fig. 3.6: Demonstration that Rb has no or negative contribution to the overall coding performance.	77
Fig. 3.7: Demonstration of LDR visual effects and compression gains for using the JVT approach as the reference TMO.	79
Fig. 3.8: Demonstration of LDR visual effects and compression gains for using the photographic TMO as the reference TMO.	80
Fig. 3.9: Demonstration of LDR visual effects and compression gains for using the adaptive display TMO as the reference TMO.	81
Fig. 3.10: Demonstration of LDR visual effects and compression gains for using the adaptive logarithmic TMO as the reference TMO.	82
Fig. 4.1: Pipeline of the conventional method for tone-mapping the Cb and the Cr components.	86
Fig. 4.2: Framework of the proposed solution.	87
Fig. 4.3: Tone-mapped images using the conventional approach (left column) and the proposed method (right column) for chromatic correction.	94
Fig. 5.1: LDR images generated from the same HDR scene by tone-mapping operators evaluated in this paper.	100
Fig. 5.2: Scenes used in the subjective test.	103
Fig. 5.3: Results (with 95% confidence interval) for 3D effect.	106
Fig. 5.4: Results (with 95% confidence interval) for overall 3D quality.	106
Fig. 5.5: Demonstration of unnaturalness by the gradient TMO.	107
Fig. 5.6: Pair-wise ANOVA comparison of the seven TMOs and the LDR capturing, for (a) 3D Effect and (b) overall quality.	110
Fig. 5.7: Demonstration of images at different brightness and detail levels.	113
Fig. 5.8: The effect of changing the brightness and the detail levels.	115
Fig. 5.9: Illustration of the graphic user interface for subjects to fully navigate the psychophysical experiment on their own pace and to select the images at the preferred brightness or detail levels.	116
Fig. 5.10: Comparison between 2D and 3D images in terms of the preferred brightness level.	119
Fig. 5.11: Demonstration of the 2D (left) and the 3D (right) images at their average favored brightness levels.	119
Fig. 5.12: Comparison between 2D and 3D images in terms of the preferred detail level.	121
Fig. 5.13: Demonstration of the 2D (left) and the 3D (right) images at their average favored brightness levels.	121

List of Abbreviations

3D	Three-Dimensional
ANOVA	Analysis of Variance
CRT	Cathode Ray Tube
GGD	General Gaussian Distribution
HDR	High Dynamic Range
HVS	Human Visual System
JPEG	Joint Photographic Experts Group
JVT	Joint Video Team
LCD	Liquid Crystal Display
LDR	Low Dynamic Range
MPEG	Moving Pictures Experts Group
MSE	Mean Square Error
PSNR	Peak Signal to Noise Ratio
TMO	Tone-Mapping Operator
QoE	Quality of Experience
RD	Rate-Distortion
SSIM	Structural Similarity Index
QP	Quantization Parameter
SVC	Scalable Video Coding
VDP	Visible Difference Predictor

Acknowledgements

The journey of my Ph.D. has been so important and treasurable that it will have significant impact on the rest of my life. This thesis would not have been possible without encouragement and supports from a large number of people.

First and foremost I wish to express my most sincere and deepest gratitude to my supervisors Dr. Rabab Ward and Dr. Panos Nasiopoulos. I have known them for over six years since my Master's, and their understanding, support and guidance have been of great value throughout my post-graduate life. Dr. Ward granted me an opportunity to study in Canada where the new page of my life was open. As a brilliant and prestigious researcher, Dr. Ward is also a very warm-hearted and caring lady. Her comprehensive knowledge and logical way of thinking have helped me improve my research as well as my communication skills. Her expertise in dancing also inspired me to develop my skills of all aspects while working hard on research. Dr. Nasiopoulos is talented and highly-recognized in both academia and industry. He has been amazing me with his critical thinking and providing me with excellent advice since the very first stage of this degree. The challenging questions he had asked about my work sometimes made me feel nervous but always inspired me with new ideas and deeper thoughts of the research. Dr. Nasiopoulos is not only a great academic advisor but also a friend and a life mentor. This work would not have been completed without his encouragement and direction.

I am also deeply grateful to my colleagues in the group. We become more of a family, filled with wholehearted advice, close collaboration and endless fun. It was my great pleasure to work with these nice and smart people. I would like to thank Dr. Qiang

Tang, my big brother, for sharing his valuable experience on graduate study and overseas life from day one, Colin Doutre for his wide knowledge and persistent companionship as well as quick response to my questions of all kinds, Dr. Hassan Mansour for showing me analytical skills and correcting my English, Dr. Lino Coria for his jokes and useful suggestions on how to raise kids, Mrs. Di Xu for her smiling face and easy-going character, Dr. Mahsa Pourazad for working hard with me to meet deadlines and fighting for the second authorship, Mr. Sergio Infante for being my “interactive TV twin” and always ready to help, Dr. Victor Sanchez for his trouble-free personality and discussions on lossless compression, Dr. Matthias von dem Knesebeck for his kind character and finishing the beer left in the cattle unconditionally. I also want to thank Mrs. Ashfiqua Connie, Dr. Mehrdad Fatourehchi, Dr. Shan Du and Dr. Yaser Fallah for their friendship.

I would like to deeply thank Dr. Rafal Mantiuk for sharing his broad knowledge on high dynamic range imaging and working together on my first project in this degree.

To my friends in Vancouver who added different dimensions to my life: Yangwen Liang, Xudong Lv, Bernard Ng, Arthur Mak and Eric Zhang.

I own my sincere gratefulness to my parents for their love. While living in a different country, they have provided their only child with every possible support that makes this degree come true.

Last but not least, I would like to express my special gratitude to two important people in my life. First, my beloved wife Suling Yang for her constant understanding and support and for sharing the good and bad moments with me throughout my Ph.D. Second, my adorable son Ethan Mai who has created an everlasting project for me while bringing endless happiness to my life.

Dedication

To My Parents, My Wife and My Son

1 Introduction and Overview

1.1 Introduction

Human eyes are able to adapt to light conditions with a dynamic range (contrast) of over 10,000,000,000 : 1. At a single time instant, the human vision can perceive a dynamic range of the order of 100,000 : 1 [1]. Unlike the wide range of light intensity allowed by the human vision system, the majority of existing capturing and display devices support only a range between the order of 100 : 1 to 1,000 : 1 – known for this reason as “Low Dynamic Range (LDR)”. A new-generation of imaging systems promises to overcome this restriction by supporting a high dynamic range (HDR) for images and videos, containing and displaying information that covers the full visible luminance range and the entire color gamut [2].

The fact that HDR imaging offers life-like picture quality has sparked significant efforts for the development of HDR displaying, capturing, and rendering technologies in recent years. One such milestone is the innovation of applying LED light sources for LCD displays and the dual modulation mechanism used for controlling the emitted light [3]. The first prototype of an HDR display was introduced by a University of British Columbia spin-off company Brightside, which was then acquired by Dolby and led to the establishment of Dolby Canada. At the time of the writing of this thesis, the commercial version of this HDR display remains at a prototype stage and is only available to a selected professional market.

Compared to HDR display technologies, the HDR content generation was gained much more popularity among both professional and regular users. Combining LDR images/videos with multiple exposures is a common way to produce HDR content [4], [5]. A more direct but expensive solution for creating high dynamic range content is to use cameras equipped

with special sensors that provide a considerably high signal to noise ratio. As the amount of HDR data is dramatically increasing, compressing HDR data to allow efficient storage and transmission becomes an essential process. Recently, HDR compression proposals have been introduced as an extension to the international video coding standard H.264/AVC [6], [7], [8], [9].

Why is the quantity of HDR data fast growing in spite of the lack of HDR displays? This is because all LDR displays can provide much better picture quality if the content is first captured in HDR and then converted to the LDR format. The process that converts the HDR signal to an LDR format is called tone-mapping. Fig. 1.1 demonstrates the difference between an LDR image captured in LDR and a tone-mapped LDR image originally captured in HDR. Tone-mapping the HDR captured images/videos is beneficial since it produces higher quality content with much less over-saturated and under-saturated areas compared to the traditional LDR capturing process. In addition, tone-mapping in some cases makes processing of images and videos easier, and it allows higher degrees of freedom for the artists who can (during postproduction stage) decide on the final effect/style of the resultant LDR image. For instance, the white balance problem can be solved with much less difficulty, color manipulation can be done much more creatively, and the adjustment of contrast and brightness can be handled much more easily. Such a production pipeline, i.e., shooting in HDR and then rendering to LDR, has been increasingly gaining interest in movie/television production and high-end photography [2]. Many Hollywood movie studios have adopted this preferred approach. A large number of still images tagged with “HDR” can also be found on the Internet, such as the photo sharing website “Flickr”; these are actually LDR photos that have been tone-mapped from their HDR origins.



(a) Captured directly in LDR



(b) Captured in HDR and tone-mapped to LDR

Fig. 1.1: Demonstration of the visual difference of LDR images obtained by a) LDR capturing, and b) HDR capturing and tone-mapping afterwards.

HDR imaging technologies have numerous applications in different areas, including digital photography, computer gaming, home entertainment systems, digital cinema, virtual reality, and medical imaging.

In the remainder of this chapter, we first describe the current techniques for generating and displaying HDR content, followed by a background overview of the schemes that are especially designed to encode HDR data. We then discuss previous studies on tone-mapping and how its associated temporal and color problems were tackled. Next, we describe existing studies that attempt to produce 3D HDR images/videos and discuss the viewers' quality of experience (QoE) related to 3D content. Finally, we present the thesis statement and objectives, concluding with a summary of the thesis contributions as well as the organization of the chapters that follow.

1.2 Generating and Displaying High Dynamic Range Content

1.2.1 Generation of HDR Content

Different approaches have been developed for producing images and videos with a larger dynamic range. HDR recording was first studied for still image photography where multiple exposures of a static scene are captured and then combined to construct an HDR radiance map [5]. Fig. 1.2 shows an example of a scene taken at different exposures, where each exposure level is able to capture only certain details of the scene. However, this technique suffers from motion artifacts whenever moving objects exist in the scene or the camera is shaken during the recording of the multiple exposures. Using a similar idea, Kang et al. [10] employed an LDR camera with a programmable control unit and studied how to generate an HDR video from a sequence of frames with alternating bright and dark

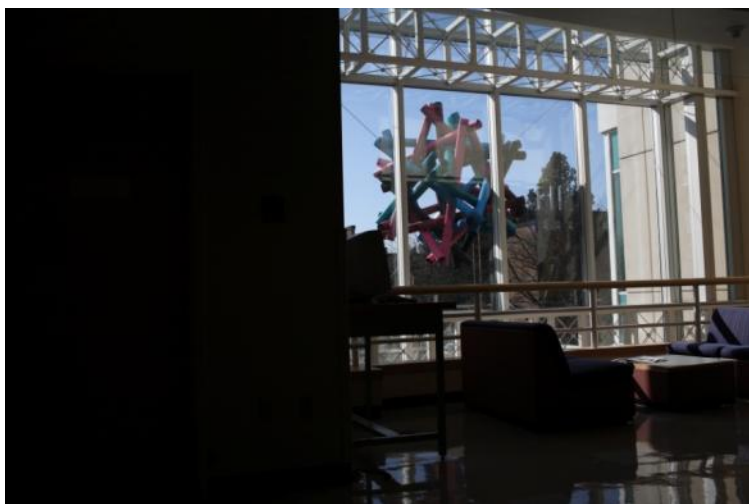
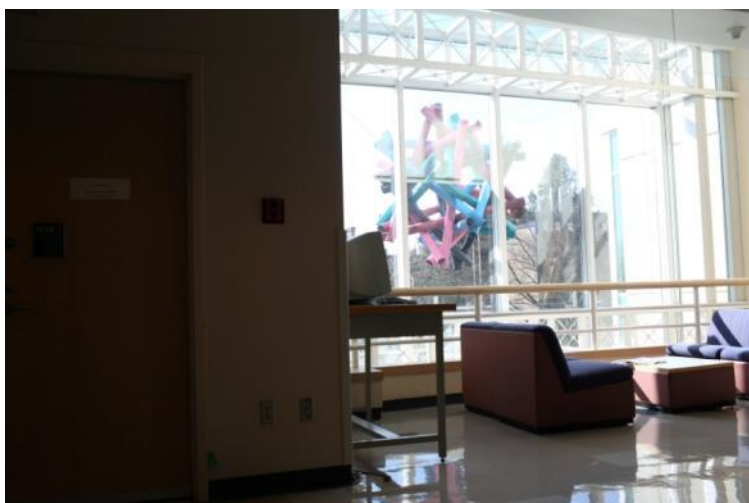


Fig. 1.2: A scene taken at different exposures.

exposures. This study overcame the temporal artifacts by applying a warping technique as well as a global and a local registration in order to compensate for pixel motion. An alternative approach is to separate the light emitted from the scene into multiple copies by a beam splitter. Each of the beams is then captured by a camera set to different exposures [11], [12], [13]. Besides generating HDR content using LDR-sensor cameras, recent developments in sensor technologies are dedicated to increasing the signal to noise ratio for recording a higher dynamic range in a single capture. In that respect, Pixim Inc. designed a new image sensor, where pixels are sampled multiple times during a single exposure (using independent analog-to-digital converters) [14]. In 2008, Fujifilm announced its “super CCD EXR” sensors where one pixel element consists of two parts: one part is responsible for the high sensitivity and the other for the low sensitivity. These two parts are then combined to produce an HDR signal. Gu et al. [15] investigated a new readout architecture for image sensors in which the readout timing and exposure length can be controlled flexibly for each row of the sensor. At the time of the writing of this thesis, the camera company RED, whose products have been used to shoot a number of Hollywood movies, released a professional camera model called “EPIC” that is able to capture a dynamic range up to 18 stops (yielding a contrast ratio of about 260,000 : 1) [16].

Another way of obtaining a high dynamic range content is through inverse Tone-mapping, a method that attempts to expand the dynamic range of existing 8-bit images and videos. The objective in this case is to non-linearly map the 256 levels to a larger range of levels. Meylan et al. [17] developed a scaling function based on the observation that light sources, highlights and specular surfaces appearing in images should have their pixels stretched more than the other parts of an image. Recovering the camera response curve is a

more accurate way to achieving this mapping. Wang et al. [18] argued that a decent approximation of the response curve may be obtained by simply inverting the gamma correction function. The technique developed by Farid [19] is able to accurately model the gamma correction function, without prior information about the scene type, camera model, or capturing parameters. Rempel et al. [20] worked closely with an HDR display prototype and based on the model for this display designed a completely automatic contrast enhancement method that results in a good visual appearance of the legacy 8-bit video on the HDR output device in real time.

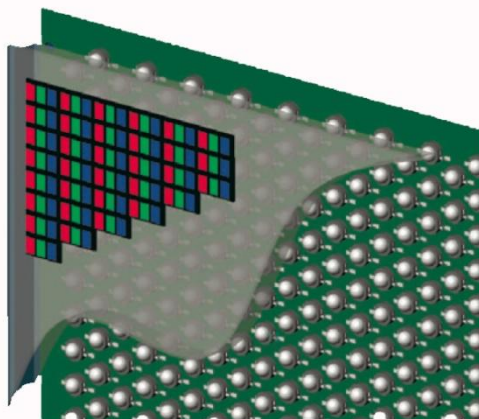
The second difficulty in achieving dynamic range expansion is to recover the missing information from the under-saturated and over-saturated regions. In [18], texture regions are first isolated from the regions containing light sources, and then texture synthesis methods are used to reconstruct the image details in the clipped areas. Di et al. [21] developed a scheme that restores the clipped pixels in color images, based on the spatial correlation in the chroma components between a clipped region and its surrounding areas.

1.2.2 Display of HDR Content

Due to their physical limitations, existing display devices provide a dynamic range that is inadequate for presenting HDR material. For cathode ray tubes (CRTs), the brightness cannot be very high since it is not safe to deposit too much energy on the screen. The limitation with the liquid crystal displays (LCDs) lies in the darkest illumination. This is because liquid crystals cannot completely block the backlight to make the screen look sufficiently dark. Plasma displays also have a limit on the dark level of their dynamic range. This level is not fully darkened due to the “pre-charged” mechanism used in this type of

displays. With this mechanism, each cell (pixel location) is pre-charged before it is lightened in order to provide fast transition between two brightness states.

The new display and projection technology known as “dual-modulation” promises to deliver a true high dynamic range visual experience by first employing narrow-wavelength LED light sources that expand the boundaries of the displayable color gamut (Fig. 1.3(a)). This expansion is again vastly enlarged with the technology that employs dual modulation or backlight dimming which greatly enhances the intra- and inter-frame contrast [3]. The prototype (Fig. 1.3 (b)) built upon this technology was shown to be capable of reaching a contrast over 150,000 : 1 . Ward [22] applied the “dual-modulation” theory to hard-copy media and developed an HDR still image viewer using a backlight system and layers of transparencies as well as stereo optics. Another emerging HDR solution to hard-copy material is the reflection print where the device does not contain any light sources and only ambient light is used. The fluorescent ink is used in the reflection prints since it makes the print look brighter as it is to transfer the ultraviolet illumination to various visible colors. The color space of fluorescent inks has also been investigated recently [23], [24].



(a)



(b)

Fig. 1.3: The LED layer (a) and the Dolby prototype (b) of the high dynamic range display.

1.3 Encoding High Dynamic Range Content

HDR images and videos need to be compressed or encoded so they can be efficiently stored and delivered. The advances in HDR display technologies have motivated the use of extended gamut color spaces. These include xvYCC (x.v.Color) for the home theater [25] and the Digital Cinema Initiative color space for digital theater applications. Yet, even these extended color spaces are too limited for the amount of contrast that can be perceived by the human eye. High dynamic range (HDR) video encoding goes beyond the typical color space restrictions and attempts to encode all colors that are visible and distinguishable to the human eye [26]. It is not restricted by the color gamut of the display technology used. Depending on whether or not the LDR data format is supported, HDR compression schemes can be classified into two categories: i) single-layer HDR video/image compression and ii) backward-compatible HDR video/image compression.

1.3.1 Single-Layer HDR Image/Video Compression

The single-layer compression approach encodes HDR signals into a single stream and does not contain a layer that is backward-compatible with the 8-bit (LDR) playback/display devices.

HDR signals preserve the colorimetric or photometric pixel values (such as CIE XYZ) within the visible color gamut and allow for intra-frame contrast to reach the magnitude of $10^6 : 1$, without introducing contouring, banding or posterization artifacts caused by excessive quantization. The photometric and colorimetric values, such as luminance ($\text{cd}\cdot\text{m}^{-2}$) or spectral radiance ($\text{W}\cdot\text{sr}^{-1}\cdot\text{m}^{-3}$), span a much larger range of values than the luma and chroma values (gamma corrected) used in typical video encoding (JPEG, MPEG, etc.). The obvious representation of these colorimetric values is via floating point numbers. These are,

however, impractical for image and video coding applications. For that reason, several color encoding and file formats have been proposed for storing HDR data, including the Radiance RGBE (.hdr) [27], OpenEXR (.exr) [28] and LogLuv TIFF (.tiff) [29] file formats. The RGBE format assigns four bytes to represent each pixel: one byte used for the mantissa of each of the RGB channels and the remaining one byte is used as a shared exponent. The exponent byte together with the mantissa part is able to represent a value of a very large range. On the other hand, OpenEXR spends 16 bits for each of the RGB channels: a sign bit, five bits for the exponent, and ten bits for the mantissa. The LogLuv TIFF format encodes the data in the logarithmic domain and supports 32 bits per pixel using one sign bit, 15 bits to encode the log scale of the luminance, and 8 bits for each of the two chrominance channels. These three formats are considered nearly lossless and require a high data rate, so they are not favorable for video applications.

In order to achieve higher coding gains, studies were conducted on HDR image/video lossy compression. Xu et al. proposed an HDR image compression scheme by adapting HDR signals to the JPEG-2000 coding requirements [30]. Mantiuk et al. [26] developed an MPEG-based HDR video compression method. They derived a color space suitable for encoding HDR signals based on threshold versus intensity functions of the human visual system. They drew a conclusion that 10-11 bits are sufficient to encode the full perceivable luminance range.

1.3.2 Backward-Compatible HDR Image/Video Compression

High dynamic range video formats are unlikely to be broadly accepted if they are not backward-compatible with existing LDR digital display devices. Such backward-compatibility can be achieved if the HDR video stream contains 1) a backward-compatible 8-

bit video layer, which could be directly displayed on existing devices, and 2) additional information which along with this 8-bit layer can yield a good quality HDR reconstructed version of the original HDR content. Such a stream may also contain a residual layer to further improve the quality of the HDR reconstruction. For still image compression, backward compatibility can be achieved by encoding a tone-mapped copy of the HDR image together with a residual image [31] or a ratio image [32] that allows the reconstruction of the original HDR image. Mantiuk et al. [33], Segall [34], and Winken et al. [7] extended this approach to video sequences. A tone-mapping curve was encoded together with the tone-mapped and residual video sequences. In [33] the residual video sequence was additionally filtered to remove the information that is not visible to the human eye.

Recently, several proposals for bit-depth scalability have been introduced to provide backward-compatible HDR video bitstreams. These proposals incorporate backward-compatible encoding of high fidelity video as an extension to the H.264/AVC video encoding standard [6], [7], [8], [9]. The extension includes tone-mapping supplemental enhancement information (SEI) messages, which encode the shape of the tone-mapping curve [6]. The scalable video coding (SVC) extension [35] is used to encode an additional residual stream needed to reconstruct the information lost due to tone-mapping, or to provide bit-depth scalability [7], [8], [34], [36].

In this thesis, we follow the main structure that is compatible with the H.264/AVC video coding standard. Fig. 1.4 illustrates the general coding structure under such framework. Depending on the source, the original HDR video can be represented by 12-bit integers or floating point numbers [27], [28]. Each frame of the HDR video sequence is first tone-mapped to an 8-bit LDR image. This LDR frame is then encoded using the H.264/AVC

standard to generate a compressed stream for the base layer. At the receiver end, the base layer is decoded to an LDR frame, which is then inversely tone-mapped to an HDR frame. Compared to the input HDR signal, such a reconstructed HDR frame has information loss introduced by both tone-mapping and video compression. On the enhancement layer, the distorted HDR frame is subtracted from the original HDR source, and the difference is encoded using H.264/AVC again for producing the residual stream. The data streams generated for the base and the enhancement layers jointly compose the compressed signal of the backward-compatible HDR video.

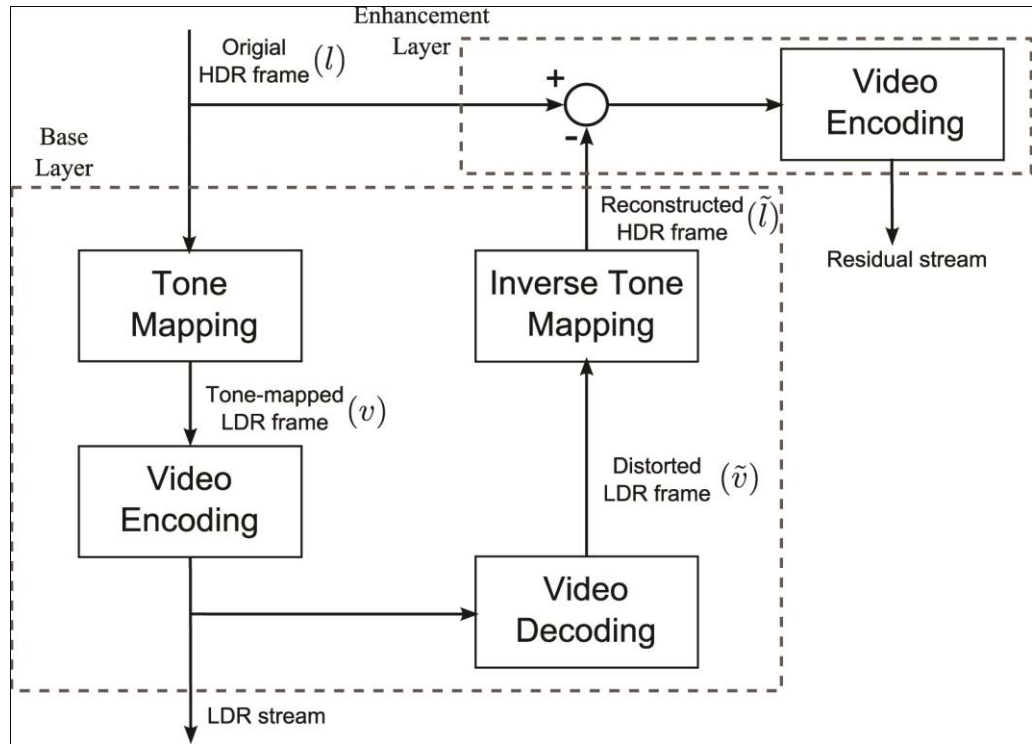


Fig. 1.4: System structure of backward-compatible HDR video compression proposed by H.264/AVC.

1.4 Tone-Mapping High Dynamic Range Content

In order to visualize HDR content on existing imaging systems, tone-mapping that converts HDR to 8-bit LDR signals is essential. In this section, we provide an in-depth overview of current tone-mapping methods and the issues related to tone-mapping including color correction and temporal inconsistency problems. We also review the studies on evaluating existing tone-mapping methods.

1.4.1 Tone-Mapping Operators



Fig. 1.5: Demonstration of the difference in visual effects using different TMOs. (a) TMO by Ashikhmin [43], (b) TMO by Mantiuk et al. [44], (c) photographic TMO, (d) Manual mapping using HDR software Photomatix [45].

The primary goal of tone-mapping is to reduce the dynamic range of an HDR scene in a way that preserves specific HDR perceptual effects in the resulting LDR version of the image/video. A large variety of tone-mapping operators (TMOs) have been developed for different purposes, such as photographic tone reproduction [38], simulating artistic drawing [39], adaption to different displays [37], user-interactive tone adjustment [40] and photoreceptor physiology based modeling [41]. A review of most of these operators can be found in [2]. Fig. 1.5 illustrates the different visual effects delivered by different TMOs. The images in (a) (b) and (c) were obtained using an HDR tool called *Luminance HDR* [42], and the TMOs are (a) TMO by Ashikhmin [43], (b) TMO by Mantiuk et al. [44], and (c) photographic TMO. As for the tone-mapped image in (d), we manually mapped the HDR image using another HDR software Photomatix [45].

TMOs can be generally categorized into two broad classes: global and local operators. Global TMOs use a single, usually nonlinear, mapping curve for every pixel in the image. For local operators, the mapping of each pixel is a spatially variant function of its neighboring pixels. In general, global TMOs are more computationally efficient and better retain the sensation of the brightness of the original HDR signal. On the other hand, local operators are able to preserve more details and provide higher local contrast.

Tone-mapping is an essential component in the backward-compatible HDR image/video compression. Li et al. [46] first considered tone-mapping explicitly so as to optimize image compression. They used forward and inverse wavelet-based tone-mapping (compressing and companding) in an iterative optimization loop to minimize HDR quality loss due to the quantization of the 8-bit tone-mapped image. As this method requires encoding the tone-mapped image using high bit-rates, it is not suitable for video. Lee et al.

[47] extended the gradient domain tone-mapping method [48] to video applications using the temporal information obtained from the video decoding process. In [49] the performance of several tone-mapping operators in terms of quality loss due to forward and inverse tone-mapping was compared. Local tone-mapping operators (spatially variant) were found to be more prone to quality loss than global operators (spatially invariant). In Chapters 2 and 3 of this thesis we compare the results of our methods with the two tone-mapping operators that performed the best in this study: the photographic TMO [38] and the adaptive logarithmic TMO [50].

1.4.2 Temporal Inconsistency Correction for Tone-Mapping

The fact that tone-mapping is usually applied independently to each frame leads to incoherence between consecutive LDR frames, known as flickering artifacts. Current research on tone-mapping has been largely limited to still images, and only a few studies try to address this temporal problem. Additionally, most studies are restricted to specific (self-contained) TMOs.

Based on Reinhard et al.'s TMO [38], Kang et al. [10] has proposed a method that prevents the flicker by smoothing the log-average HDR luminance over a number of frames. For the same TMO, Grzegorz et al. have applied a human vision model and achieve temporal adaptation using an exponential decay function [51].

Wang et al. extend the gradient tone-mapping scheme [48] to video applications by representing an HDR video sequence as a 3D volume of pixels [52]. Another time-dependent adaptation for [48] is proposed in [47], where the pixel-wise motion information is incorporated into the objective function.

Benoit et al. have recently proposed a spatio-temporal tone-mapping method using a human retina model in [53] and reported that flickering artifacts are more severe for global TMOs than local ones.

While the above papers focus on temporal extensions for certain TMOs, a tone curve filtering method which could be adapted to any global tone-mapping operators is briefly described in [37] as an add-on to their tone-mapping method for still images. Based on the human vision's peak sensitivity, the temporal variations of the tone curve is forced to be under 0.5 Hz, by applying a linear-phase low-pass FIR filter with a cutoff frequency of 0.5 Hz. A major disadvantage of this approach lies in its long buffering (30 frames) requirement, which makes the filter not suitable for real-time video applications.

1.4.3 Color Reproduction for Tone-Mapping

Other than the methods presented in [54] and [55] (where each of the chrominance channels is tone-mapped differently using the color appearance model), most TMOs have been originally designed for mapping only the luminance to LDR pixel values. To map color components, the common approach is to tone-map the luminance channel first and then to map each of the RGB channels based on (but not necessarily the same as) the tone-mapping of the luminance (Fig. 1.6). This process is called color correction.

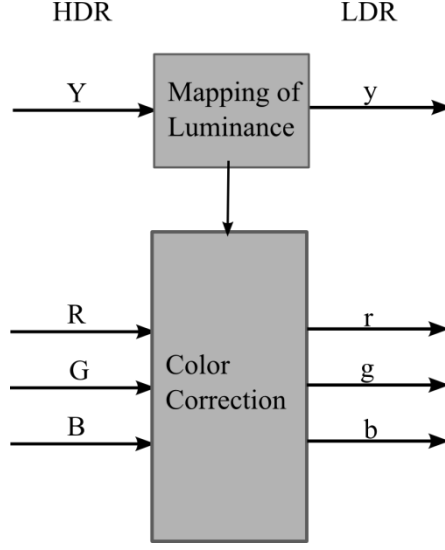


Fig. 1.6: Conventional approaches to tone-map color images.

The straightforward color correction solution for obtaining colored tone-mapped images is to map each of the color channels using the exactly same process as in luminance.

This color treatment maintains the color ratio, which can be written as follows [56]:

$$r = \frac{R}{Y} \cdot y \quad (1.1)$$

where R and Y denote the HDR version of the red and the luminance channels, respectively, while r and y are the 8-bit LDR counterparts. The above equation shows only the mapping of the red channel. The blue and the green channels can be obtained in the same way. The saturation of the resulting image/video can be further controlled using a heuristic approach [57], [58]:

$$r = \left(\frac{R}{Y}\right)^s \cdot y \quad (1.2)$$

where s is the parameter that adjusts the color saturation.

Recently, Mantiuk et al. have reviewed the color correction methods for tone-mapping and presented another color correction, in the RGB space, that preserves both color ratios and

luminance but may suffer from strong hue shift, especially for the red and the blue components [59].

The above methods, were mainly designed for RGB signals and cannot be directly applied on videos formatted in the YCbCr color space, which is the most commonly used format in practice. Little work has been done to directly convert high bit-depth videos to 8-bit ones in the YCbCr domain. In the context of high dynamic range imaging, Toda et al. studied the chromatic treatment of the Cb and Cr channels [60], but their work has mainly addressed the creation of YCbCr HDR content, not involving tone-mapping.

1.4.4 Evaluation of Tone-Mapping Operators

Tone mapping of 2D images and videos has been extensively studied and several subjective evaluations have been conducted. Drago et al. [61] performed subjective tests on six TMOs with 11 observers and found that the most salient attributes of tone-mapping are naturalness and detail. Yoshida et al. [62] evaluated seven TMOs using two real-world scenes for comparison and concluded that global (spatially invariant) TMOs are preferred for the reproduction of overall contrast and brightness while local (spatially variant) TMOs perform better for reproducing details in highlights. In 2007, Kuang et al. [63] evaluated eight tone-mapping methods by having the subjects rate four specific attributes of the images, namely global contrast, colorfulness, shadow details and sharpness. Ledda et al. [64] conducted a subjective experiment by comparing tone-mapped LDR images with the corresponding HDR images viewed on an HDR display. Ashikhmin and Goyal [65] performed a reality check for five tone-mapping methods. Their tests were divided into two parts: with and without real-world scenes as reference for evaluation. No statistically

significant difference was found between the tested TMOs in terms of subjective preference if no reference was provided.

On the other hand, the results are very different when physical scenes are used for comparison. Akyüz et al. [66] investigated the difference between the tone-mapping pipeline and the conventional LDR pipeline. In [67] Čadík et al. evaluated six different attributes of images tone-mapped using 14 tone-mapping methods and also employing with and without real-scene references. Contrary to the Ashikhmin and Goyal's work, Čadík et al. did not find statistically significant difference between the results from the tests with and without physical scenes as references. They then used the scores from the psychophysical experiments to quantify the contribution of each of the image attributes to the overall quality of the tone-mapped image.

Objective assessment methods for TMOs have also been studied. Aydin et al. [69] developed a dynamic range independent image quality measurement method taking into account the human visual system. Three types of distortion are defined, loss of visible contrast, amplification of invisible contrast and reversal of visible contrast. With their method one can assess the quality of a tone-mapped image against the original HDR counterpart. Recently, Yeganeh and Wang [70] have proposed an objective scheme to evaluate the quality of tone-mapped images versus HDR images based on the principle of the multi-scale structural similarity (SSIM) index. This principle compares the luminance, contrast and structure of local image patches. They also validated their measurement scheme with subjective experiments.

1.5 3D Visual Experience

As in the case of HDR imaging, the 3D display technology also aims at providing viewers with a more realistic visual experience. The 3D displays, however, offer a sense of scene depth, and 3D content is being produced increasingly to satisfy the demand of the rapidly growing 3D consumer market. Ideally, the 3D content should be captured in a 3D HDR representation and viewed on a 3D HDR display, to achieve a more lifelike picture quality.

1.5.1 3D Display Technologies

An increasing number of theatres and households are being equipped with 3D display systems. Currently cinemas use either polarized glasses [71] or RGB colors whose wavelengths slightly differ from one eye to the other eye [72]. Polarized glasses are starting to appear in household consumer displays, but presently most 3DTVs and computer monitors employ active shutter glasses. In active shutter based 3D, the display presents the left and the right views alternatively at a high frame rate. A 3D glasses pair switches between clear and opaque states, so that only one eye sees the image shown on the display at any instant, allowing separate images to be viewed by the left and right eyes. Progress is being made towards producing auto-stereoscopic (glasses-free 3D) displays for the consumer market. Lenticular lenses, parallax barriers, volumetric and light field displays are examples of auto-stereoscopic technologies.

1.5.2 Generation of 3D HDR Content

The generation of 3D (stereo) HDR content is quite a recent problem. One of the critical steps in achieving this goal is to recover the camera response curve from the stereo

views and, based on this curve, a radiance map of the image or the video frame is produced. In 2004, Kim and Pollefeys [73] computed the camera response curve from consecutive frames with different brightness levels, by feature matching as well as stereo matching. Inspired by this work, Troccoli et al. [74] found correspondences between frames by using all potential pairs of the input and then formulated a new weighted least square problem in the logarithmic space in order to recover the response curve. They achieved better results compared to the work by Kim and Pollefeys [73]. Ning et al. [75] improved the disparity refinement stage [74] and achieved fewer artifacts by encoding a wider dynamic range. Ramachandra et al. [76] tackled the motion blur problem resulting from capturing with long exposure. They proposed a de-blurring scheme that employed a multiscale directional mechanism using inter-view and inter-frame information. Hrabar et al. [77] proposed a new technique that constructed the 3D occupancy map by projecting the 3D points into a common plane and realized high dynamic range stereoscopic vision for outdoor robotics.

Different types of stereo HDR capturing systems have already been built. Huhle et al. [78] have demonstrated the need for conducting 3D acquisition using the HDR-capable system. They built two systems that can achieve 3D HDR capturing by setting different exposures for the adjacent cameras and developed an automatic scheme that selects the best combination of exposures for creating HDR content for their systems. Unlike other 3D HDR acquisition schemes, Zhu et al. [79] developed a system that is able to produce a 360-degree panoramic viewing angle.

1.5.3 3D Quality of Experience

Many factors and parameters may affect the perceptual quality of 3D content. The psychophysical study conducted by Cormack et al. [80] found that a change in 3D depth

perception occurs when the luminance contrast approaches a certain detection threshold. Sharp Laboratories reported that viewers are able to fuse 3D images with larger disparities if the images contain strong texture [81]. This indicates that the amount of detail present in an image affects 3D perception. Recently, the concept of 3D quality of experience (QoE) has gained increasing attention. ITU-R WP6C is making efforts to identify the requirements for broadcasting 3DTV as well as its subjective testing [82], while the ITU-T Study Group 9 has added 3D video quality to its tasks [83]. Visual fatigue has been noted as one of the main factors that affect 3D QoE; Lambooi et al. have provided a summary of its causes [84]. Goldmann et al. investigated the effect of the distance between stereo cameras on the quality of the stereoscopic videos [85]. They pointed out that only a few levels of 3D quality can be differentiated as the distance between the cameras varies. Brightness has been noted by artists to affect perceived quality of 3D content [86], but without quantifying or thoroughly studying the effect. Most recently, Pourazad et al. [87] have started to study the effect of brightness in a systematic way by capturing a series of scenes with different exposures from very dark to very bright and then conducted a subjective evaluation on the 3D perception quality. According to these results, the brightness levels preferred by viewers for the 2D and 3D cases are similar.

A few objective quality metrics have been proposed for 3D images/videos. Campisi et al. [88] developed a 3D quality metric by applying existing 2D metrics independently and then performing a fusion technique to assign an overall score. Unlike [88], Benoit et al. [89] used the depth information obtained through stereo matching based on 2D metrics and simultaneously applied the 2D metrics in order to create a larger spectrum of potential distortions. A similar approach, which considers 2D metrics together with disparity

information, has been explored by You et al. [90]. In [91], Shao et al. proposed an objective metric that combines the Structural Similarity Index (SSIM) and the distortion in color and edge signals. The above metrics require a reference signal for the assessment of 3D quality. Their main goal is the evaluation of the quality of content that has been degraded through compression or transmission. Mittal et al. [93] have proposed a no-reference 3D quality assessment algorithm. However, this method evaluates only the viewing comfort instead of the overall 3D perceived quality.

1.6 Thesis Motivation and Objectives

1.6.1 Thesis Motivation

Tone-mapping converts HDR images and videos to their LDR counterparts. It plays an essential role in the current HDR imaging technologies for two reasons: i) the majority of existing displays are LDR devices, and ii) the combination of HDR capturing and tone-mapping provides much better picture quality than regular LDR capturing.

The HDR display is the natural evolution of displays, as they offer lifelike picture quality and the opportunity for a new market for the entertainment and consumer electronics sectors. To this end, capturing, transmission and displaying of the new content will be of very high importance. The HDR video format is unlikely to be broadly accepted immediately due to the large installed base of existing LDR devices, making the need for backward-compatibility extremely vital. It is important to efficiently generate backward-compatible video streams that ensure superior video quality for both the LDR and HDR streams. In the backward-compatible HDR video coding framework, the compression process follows the

HDR-LDR tone-mapping process. In its simplest form, only the tone-mapped LDR content is transmitted and the HDR video is reconstructed by inversely tone-mapping the encoded-decoded LDR version of the video. The challenge stems from the fact that tone-mapping is designed independently from video compression. Both compression and tone-mapping introduce quantization and quality loss. To overcome this problem, we propose the minimization of the joint information losses of the two processes: tone-mapping and video compression, so as to achieve better resulting HDR reconstructed picture quality. However, finding a tone-mapper that jointly considers the two processes is not easy since the video coding process is fairly complicated and difficult to model.

To further improve the quality of the HDR reconstruction, an enhancement layer that encodes the difference between the original and the inversely-tone-mapped HDR signals may be included in a backward-compatible compression system. The task becomes even more interesting but requires more effort, if one has to add certain constraints on the image quality of the LDR version of the HDR video while improving the compression efficiency for the system containing both LDR (base) and enhancement layers. There are a number of existing tone-mapping operators that are designed to preserve specific HDR perceptual effects in their resulting LDR versions. Maintaining a desired LDR effect and at the same time offering high compression gains would be the preferable tone-mapping solution for video applications. The favorable preset tone-mapping operator can also be considered as the baseline of the LDR visual quality which would ensure the resulting tone-mapped content to deliver visually pleasing quality.

Most tone-mapping methods are derived only for the luminance component. This mapping function is then used for each of the R, G and B components to generate the LDR

color image. This color mapping approach cannot be directly applied to most videos since videos are usually encoded in the YCbCr color space. The YCbCr format is supported by the major image/video encoding standards, including JPEG 2000 and H.264/AVC. In that respect, it is desirable to have a tone-mapping color correction approach that is directly applied in the YCbCr color space.

Three-dimensional (3D) imaging technology provides realistic picture quality and has thus gained great interest in both the research and industrial communities recently. However, the introduction of 3D video technology can only be a long-lasting success if its perceived picture-quality reaches a satisfactory level that significantly surpasses existing 2D standards. The fact that 3D HDR-LDR tone-mapped images produce fewer under- and over-exposed image areas than 3D LDR captured images will help the viewers' perception of the 3D depth. The superior picture quality of the resulting HDR-LDR tone-mapped content will also add value to 3D LDR representation. While there are several subjective evaluations of tone mapping methods for 2D images have been reported, no studies have yet been carried on tone-mapping for the 3D case. The impact of tone mapping on 3D content has not been studied. Hence, there is a need to find out what are the attributes of 3D tone-mapping methods that will result in good 3D representation. The combination of 3D and HDR imaging technologies will deliver life-like visual experience to viewers.

1.6.2 Thesis Objectives

Based on the above discussion, it is evident that HDR-LDR tone-mapping plays an important role in the performance of HDR video compression and 3D visualization. These however are challenging topics. They require significant research that connects high dynamic

range imaging with the fields of video coding, 3D image/video processing and quality of experience. Therefore, the objectives of the thesis are:

1. For the backward-compatible HDR video compression case where only the tone-mapped LDR content (base layer) is compressed and transmitted and the HDR video is reconstructed by inversely tone-mapping the decoded LDR video: Develop a model of the distortion resulting from the combined processes of HDR-LDR tone-mapping, compression and de-compression of the LDR video, and inverse tone-mapping of the LDR to obtain an HDR video. Based on this model, we find a tone-mapping curve that minimizes the resulting joint information loss in the reconstructed HDR video stream.
2. In addition to transmitting the LDR tone-mapped video, an enhancement layer that encodes the residual signal between the original and the inversely-tone-mapped HDR signals can be also transmitted. This will result in a better decoded HDR video. We study the trade-offs among the LDR video visual quality, the base layer bit-rate and the enhancement layer bit-rate. Then we develop a new tone-mapping approach that provides high compression efficiency while maximizing HDR video quality as well as preserving a desired perceptual quality of the tone-mapped LDR video.
3. Derive a color correction (also called color mapping) scheme that allows tone-mapping to be conducted directly in the YCbCr color space, given the mapping of the luminance channel.
4. Address the problem of displaying tone-mapped 3D HDR-LDR content on 3D LDR displays and how it is different from the 2D display scenario.

1.7 Thesis Contributions

The main contributions of this thesis are summarized as follows:

- We develop a statistical model for the distortions resulting from the joint processes of tone-mapping the HDR to LDR, and compressing, de-compressing and inverse tone-mapping the LDR content. Using this model, we formulate a constrained optimization problem that finds the tone-mapping curve that minimizes the expected value of the distortion in the reconstructed (inversely-tone-mapped) HDR sequence.
- We modify the above model so that it reduces the computational complexity of the optimization problem and leads to a closed-form solution. The closed-form solution is computationally efficient and has a performance comparable to our developed statistical model. Moreover, the closed-form solution does not require the knowledge of QP (quantization parameter or quantization level), which makes it suitable for cases where the compression strength is unknown.
- We show that the appropriate choice of a tone-mapping curve using the distortion models we developed can significantly improve the quality of the reconstructed HDR stream that is inversely tone-mapped from the encoded-decoded LDR video.
- In the case where an enhancement layer (that encodes the residual signal between the original and the inversely-tone-mapped HDR signals) is transmitted in addition to the LDR version of the video: We develop a global (spatially invariant) tone-mapping operator that minimizes the overall bit-rate of both the base layer (i.e., the LDR video) and the enhancement layer while preserving a desired perceptual quality of the tone-mapped LDR video content. We derive two statistical models:

one formulates the bit-rate of the base layer and the other formulates the bit-rate of the enhancement layer. In order to achieve good perceptual quality, we also develop a model of the mismatch between our resulting LDR signals and a preset tone-mapping method. Using a favorable preset tone-mapping operator, TMO, as the baseline of the LDR quality guarantees that the resulting tone-mapped video looks pleasant. We incorporate these three models in our optimization problem. Our results show that the proposed solution provides superior compression efficiency with a desired perceptual quality for the tone-mapped LDR video.

- To minimize flickering artifacts, we derive an analytical expression that models the temporal consistency between consecutive resulting LDR frames as a function of the tone-mapping parameters. Incorporating this model in the constraint of our optimization problem successfully avoids flickering artifacts which may be problematic for some existing tone-mapping methods as they are designed originally for still images. This constraint also improves the coding efficiency because the temporal difference between two consecutive frames in the tone-mapped video is reduced.
- We derive a chromatic mapping method for tone-mapping that works directly in the YCbCr color space. This color space is most commonly used for video applications, but unfortunately all conventional color correction (color mapping) methods for tone-mapping require converting the signal to the RGB domain. Our method relies on a closed-form mapping solution that has significantly lower computational complexity as it avoids processes such as color space transformation and up-sampling which cannot be avoided by the conventional methods.

- We investigate the quality of displaying tone-mapped 3D content using stereoscopic LDR displays. We first conduct a subjective psychophysical experiment that evaluates existing HDR-LDR tone-mapping operators (TMOs) on the quality of the resulting 3D LDR images. To the best of our knowledge, this is the first attempt to assess TMOs from the 3D perspective. Our results show that capturing in HDR and then applying tone mapping produce 3D LDR images with better quality than those obtained by direct LDR capturing. We then undertake another set of subjective experiments to find out the viewers’ preferred levels of brightness and scene details in each of the 3D and 2D cases.

1.8 Thesis Organization

Our contributions will be presented in detail in the rest of this thesis. In what follows, we provide an outline of each of the remaining five chapters.

In Chapter 2, we consider a backward-compatible HDR video compression case where only the tone-mapped LDR content (base layer) is compressed and transmitted and the HDR video is reconstructed by inversely tone-mapping the decoded LDR video. We present our solution for finding an optimized tone mapping curve that significantly improves the reconstructed (inversely-tone-mapped) HDR quality in backward-compatible high dynamic range image/video compression. We develop a statistical model that approximates the distortion resulting from the combined processes of tone-mapping and compression and formulate a numerical optimization problem using this model to find the tone-curve that minimizes the distortion in the reconstructed HDR sequence. We also develop a simplified version of the model that leads to a closed-form solution for the optimization problem.

In Chapter 3, we consider the case where an enhancement layer that encodes the difference between the original and the inversely-tone-mapped HDR signals is also compressed and transmitted, in addition to the LDR base layer. We present a new tone-mapping approach that provides superior compression efficiency while preserving a desired perceptual quality of the tone-mapped LDR video. We develop statistical models that formulate the bit rates of each of the base and the enhancement layers, as well as the mismatch between the resulting LDR base-layer signal and the predefined base layer representation. The models are then incorporated in our optimization problem. The temporal effect of tone-mapping is also considered in order to avoid flickering artifacts.

In Chapter 4, we present our color correction (also called color mapping) method that is applied directly in the YCbCr color space. We derive a closed-form tone-mapping solution for the Cb and Cr chrominance components, given the HDR-LDR mapping of the luminance channel. This solution has significantly lower computational complexity because it avoids processes such as color space transformation and up-sampling, which are required by the conventional method.

Chapter 5 addresses the problem of displaying stereoscopic tone-mapped (HDR to LDR) images on 3D LDR displays and how it is different from the 2D LDR display case. We first present a subjective psychophysical experiment that evaluates the performance of existing tone-mapping operators on 3D HDR images. The second part of this study focuses on how the preferred level of brightness and the preferred amount of details differ between 3D and 2D images by conducting another set of subjective experiments. We analyze the results from the two tests to find out which attributes of tone-mapping methods will contribute to good 3D representation.

Finally, Chapter 6 summarizes the contributions of this thesis and proposes suggestions for future research.

Please note that we have kept the original symbols used in the respective publications to alleviate cross-referencing. There might hence be some overlaps in notations between chapters. To remove the ambiguity from these overlaps, we have explicitly redefined the notations in each of the chapters.

2 Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image/Video Compression

For video compression, the advances of offering high dynamic range visual experience in display technologies have motivated the use of extended gamut color spaces. These include xvYCC (x.v.Color) for home theater [94] and the Digital Cinema Initiative color space for digital theater applications. Yet, even these extended color spaces are too limited for the amount of contrast that can be perceived by the human eye. HDR video encoding goes beyond the typical color space restrictions and attempts to encode all colors that are visible and distinguishable to the human eye [26], and is not restricted by the color gamut of the display technology used. The main motivation is to create a video format that would be future-proof and be limited only by the performance of the human visual system (HVS).

Although HDR video offers truly life-like representation, the majority of existing digital display devices can only support 8-bit video content. Therefore, high dynamic range video formats are unlikely to be broadly accepted without supporting the backward-compatibility of these devices. Such backward-compatibility can be achieved if the HDR video stream contains: 1) a backward-compatible 8-bit video layer which could be directly displayed on existing devices, and 2) additional information which, along with this 8-bit layer, can yield a good quality reconstructed version of the original HDR content. Such a stream can also contain a residual layer to further improve the quality of the HDR reconstruction.

In this chapter, we address the problem of finding an optimal tone-curve for such a backward-compatible encoding scheme. To compute the tone-curve, we propose a method

that minimizes the difference in the video quality between the original and the reconstructed HDR video. This difference results in quality loss and is due to tone-mapping, encoding, decoding and inverse tone-mapping the original video. Minimizing this difference would reduce the size of the HDR residual signal in the enhancement layer. We also achieve the primary goal of tone-mapping which is to produce an LDR image with a visual response as similar as possible to the original HDR image. Although the initial assumptions used in our approach pose a difficult optimization problem, we demonstrate that for typical compression distortions there exists a closed-form solution that approaches the optimum.

2.1 Problem Statement

In this section, we present the challenges of obtaining a good quality reconstructed HDR representation in a backward-compatible HDR video encoding system and describe in detail the approach we propose towards overcoming these challenges.

The performance of a backward-compatible HDR video and image encoding system depends on the coding efficiency of the LDR base layer and the HDR enhancement layer. Performance gains can be achieved by finding a TMO that preserves the necessary information in the LDR base layer so that after it passes through the inverse TMO process, the resultant HDR reconstructed signal is of high quality. The coding efficiency of the base layer does not depend much on the TMO used as most TMOs attain a similar level of contrast in the LDR representation. Therefore, the performance gain lies in the effectiveness of the inverse TMO in producing a high quality inverse tone-mapped HDR representation. This in turn determines the resulting HDR quality (when no enhancement layer exists).

It can be deduced from above that the performance of the whole system strongly depends on the TMO used to produce the LDR representation which, in turn undergoes

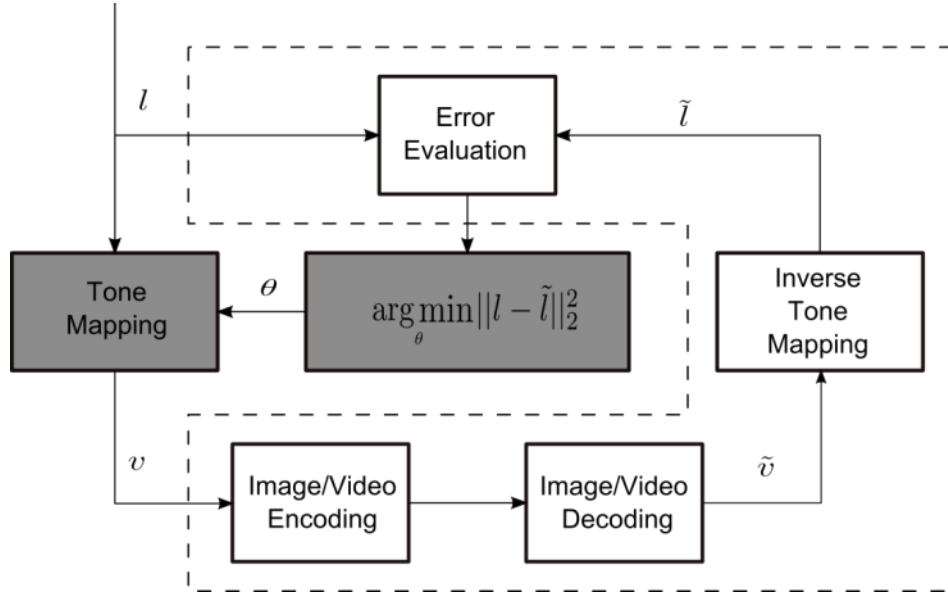
compression. Our proposed approach attempts to find the best global (spatially invariant) tone-mapping curve that minimizes the mean square error (MSE)¹ between the original HDR content and the reconstructed version obtained after tone-mapping, compression, decompression, and inverse tone-mapping. This process is illustrated in Fig. 2.1.

Let l denote the input HDR image/frame, and v the tone-mapped LDR version as shown in Fig. 2.1(a). Let \tilde{v} be the decoded LDR frame, and \tilde{l} the reconstructed HDR frame produced after inverse-tone-mapping. Also let μ be the set of parameters that control the tone-mapping operator. Our goal is to find the tone-mapping parameters that minimize MSE, which we denote $\|l - \tilde{l}\|_2^2$ using the norm notation.

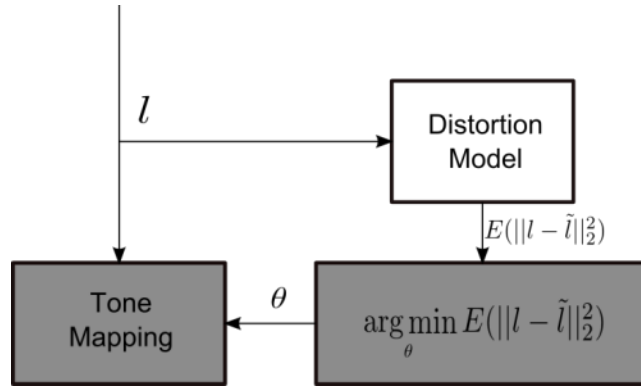
The above optimization problem can be solved by exhaustive search, repeatedly tone-mapping, encoding, decoding, and then inverse tone-mapping, until the best set of TMO parameters θ^* is found. Even though this approach guarantees an optimal solution, this framework requires an unacceptable computational cost. To overcome this problem, we estimate the distortion due to tone-mapping, encoding, decoding, and inverse tone-mapping with a statistical distortion model, as illustrated in Fig. 2.1(b). Then, we show that under certain assumptions that are valid for natural images, an immediate closed-form solution for this problem can be found.

In the following sections we consider only luminance/luma channels. To tone-map color images, we use the same tone curve for the red, green and blue color channels. Such approach was shown to well preserve color appearance for moderate contrast compression

¹ choose the mean square error as the HDR quality metric for its simplicity, despite its shortcomings in reflecting the perceptual quality of images. Moreover, the results shown in Section IV demonstrate that although we minimize the MSE we also achieve image quality gains in terms of SSIM 141[92].



(a)



(b)

Fig. 2.1: System overview of the proposed tone-mapping method. (a) demonstrates the ideal scenario where the actual H.264/AVC encoding is employed. (b) shows the practical scenario which is addressed by this chapter.

[59]. Encoding of the enhancement layer (for the residual data) is not considered in this chapter. The rationale comes from our effort to achieve the best possible HDR reconstruction and thus the smallest possible residual. As a result, the cost of encoding any additional refinement layer would be minimized.

2.2 Proposed Solution

In this section, we describe how we parameterize the tone-mapping function, approximate encoding distortions with a statistical model and then find a closed-form solution for an optimal tone curve.

2.2.1 Tone Mapping Curve

The global tone-mapping curve is a function that maps HDR luminance values to either the display’s luminance range [37], or directly to LDR pixel values. In this paper, we consider the latter case. The tone-mapping curve is usually continuous and non-decreasing. The two most common shapes for the tone curves are the sigmoidal (“S-shaped”) or a compressive power function with an exponent < 1 (gamma correction).

According to the Weber-Fechner law [95], the sensitivity of the human visual system to light is proportional to the logarithm of luminance. Thus, our tone-mapping method will operate on the logarithmic values of the luminance, which we refer to as HDR values ($l = \log_{10}(L)$ where L is the luminance of the HDR image).

To keep the problem analytically tractable, we parameterize the tone-mapping curve as a piece-wise linear function with the nodes (l_k, v_k) , as shown in Fig. 2.2. Each segment k between two nodes (l_k, v_k) and (l_{k+1}, v_{k+1}) has a constant width in HDR values equal to δ (0.1

in our implementation). The tone mapping curve can then be uniquely specified by a set of slopes:

$$s_k = \frac{v_{k+1} - v_k}{\delta} \quad (2.1)$$

which forms a vector of tone-mapping parameters μ . Using this parameterization, the forward tone-mapping function is defined as:

$$v(l) = (l - l_k) \cdot s_k + v_k \quad (2.2)$$

where v is the LDR pixel value, k is the segment corresponding to HDR value l , that is $l_k \leq l < l_{k+1}$. The inverse mapping function is then:

$$\tilde{l}(v; s_k) = \begin{cases} \frac{v - v_k}{s_k} + l_k & \text{for } s_k > 0 \\ \sum_{l \in S_0} l \cdot p_L(l) & \text{for } s_k = 0 \end{cases} \quad (2.3)$$

where $s_k \in \{s_1, s_2, \dots, s_N\}$.

When the slope is zero ($s_k = 0$), $\tilde{l}(v; s_k)$ is assigned an expected HDR pixel value for the entire range S_k in which the slope is equal zero. $p_L(l)$ is the probability of HDR pixel value l .

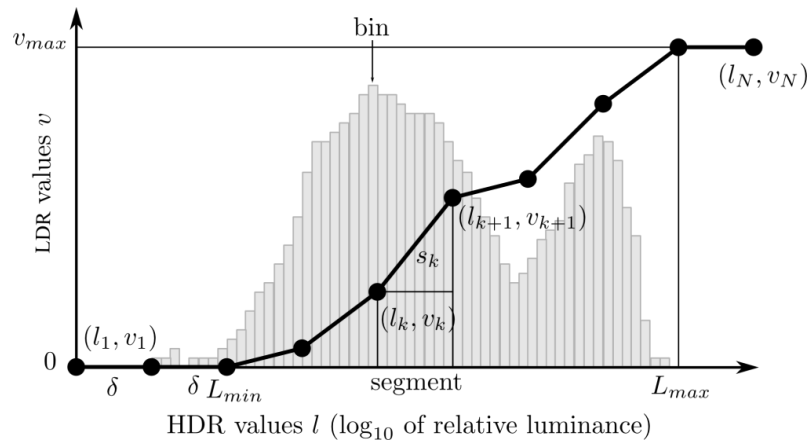


Fig. 2.2: Parameterization of a tone-mapping curve and the notation. The bar-plot in the background represents an image histogram used to compute $p(l)$.

2.2.2 Statistical Distortion Model

As mentioned earlier in section 2.1, accurately computing the distorted HDR values \tilde{l} would be too computationally demanding. Instead, we estimate the error $\|l - \tilde{l}\|_2^2$ assuming that the compression distortions follow a known probability distribution p_C . Under this assumption, the expected value of the error $\|l - \tilde{l}\|_2^2$ is:

$$E[\|\tilde{l} - l\|_2^2] = \sum_{l=l_{\min}}^{l_{\max}} \sum_{\tilde{v}=0}^{v_{\max}} (\tilde{l}(\tilde{v}; s_k) - l)^2 \cdot p_C(v(l) - \tilde{v} | v(l)) \cdot p_L(l) \quad (2.4)$$

where $p_C(v - \tilde{v} | v)$ is the probability that the encoding error equals $v - \tilde{v}$. Note that Eqs. (2.2) and (2.3) show that both v and \tilde{l} are uniquely determined by the values of l and \tilde{v} , respectively. Therefore, the conditional probabilities for these two variables and their corresponding summations have been removed from the calculation of the expected value of the error above.

The probability of the HDR pixel value $p_L(l)$ is in practice found from a histogram of HDR values and the summation over l is performed for each bin of that histogram. The number of bins is greater than or equal to the number of tone curve segments.

Since a tone curve is uniquely defined by a sequence of its slopes s_1, s_2, \dots, s_N , the expected error value in (2.4) can be expressed as a function $\varepsilon(s_k)$. For a specific tone curve defined by the sequence of slopes, the pixel value v is calculated as in (2.2). In practice the pixel values v and \tilde{v} are integer valued, such that $v, \tilde{v} \in \{0, 1, \dots, v_{\max}\}$, whereas, l and \tilde{l} are continuous real variables. The rounding operation makes the encoding error estimate in (2.4) a non-convex function. Therefore, we impose a convex relaxation on the encoding error function by removing the rounding operator from the calculation of v . Moreover, assuming that the compression error probability is independent of the LDR pixel value v , we can

simplify the expression above by removing the dependency of p_C on v . Consequently, the continuously relaxed objective function is written as:

$$\varepsilon(s_k) = \sum_{l=l_{\min}}^{l_{\max}} \sum_{\tilde{v}=0}^{v_{\max}} (\tilde{l}(\tilde{v}; s_k) - l)^2 \cdot p_C(v - \tilde{v}) \cdot p_L(l) \quad (2.5)$$

The only unknown variable is the probability distribution of the compression error $p_C(v - \tilde{v})$, which can be estimated for any lossy compression scheme. In Appendix A we model such distribution for the H.264/AVC I-frame coding. However, we will show in Section 2.24 that the distribution of the compression scheme error is not necessary to calculate a good approximation of the encoding error.

2.2.3 Optimization Problem

The optimum tone curve can be found by minimizing the function $\varepsilon(s_k)$ with respect to the segment slopes s_k :

$$\arg \min \varepsilon(s_k) \quad (2.6)$$

subject to:

$$\begin{aligned} s_{\min} &\leq s_k \leq s_{\max} \quad \text{for } k = 1, \dots, N \\ \sum_{k=1}^N s_k \cdot \delta &= v_{\max} \end{aligned} \quad (2.7)$$

The first constraint restricts slopes to the allowable range, while the second ensures that the tone curve spans exactly the range of pixel values from 0 to v_{\max} . The minimum slope s_{\min} ensures that the tone-mapping function is strictly increasing and thus invertible and $\tilde{l}(\tilde{v}; s_k)$ can be computed. The lack of this assumption introduces discontinuity and local minima, impeding the use of efficient solvers. Since s_{\min} is set to a very low value (below $0.5/\delta$) this assumption has no significant effect on the resulting tone-curves, which are

rounded to the nearest pixel values. With s_{max} we ensure that we do not try to preserve more information than what is visible to the human eye. Assuming that the luminance detection threshold equals 1% ($\frac{\Delta L}{L} = 0.01$), we can write:

$$\tilde{l}(v+1; s_k) - \tilde{l}(v; s_k) > \log_{10}(1.01) \quad (2.8)$$

so that:

$$s_{max} = ((\log_{10}(1.01))^{-1} \quad (2.9)$$

2.2.4 Closed-Form Solution

The distortion model in (2.5) gives a good estimate of compression errors, but poses two problems for practical implementation in an HDR compression scheme: 1) it requires the knowledge of the encoding distortion distribution p_C , and 2) the optimization problem can only be solved numerically using slow iterative solvers. In order to reduce the complexity of the optimization problem given in (2.7), we propose the following assumptions that allow us to cast a simpler optimization problem to which we can find a closed-form solution with almost no noticeable impact on the compression performance.

If we assume local linearity of the tone curve, so that the slope at the non-distorted pixel value v and that at the distorted pixel value \tilde{v} is the same, we can then substitute $\tilde{l}(\tilde{v}; s_k)$ and l in the distortion model (2.5) using the inverse mapping function in (2.3), thus yields:

$$\varepsilon(s_k) \approx \sum_{l=l_{min}}^{l_{max}} \sum_{\tilde{v}=0}^{v_{max}} p_C(v - \tilde{v}) \cdot p_L(l) \cdot \left(\frac{v - \tilde{v}}{s_k} \right)^2 \quad (2.10)$$

After reorganizing we get:

$$\begin{aligned}
\varepsilon(s_k) &\approx \sum_{l=l_{\min}}^{l_{\max}} \frac{p_L(l)}{s_k^2} \sum_{\tilde{v}=0}^{v_{\max}} p_C(v - \tilde{v}) \cdot (v - \tilde{v})^2 \\
&= \left(\sum_{l=l_{\min}}^{l_{\max}} \frac{p_L(l)}{s_k^2} \right) \cdot \text{Var}(v - \tilde{v})
\end{aligned} \tag{2.11}$$

Since the variance of $(v - \tilde{v})$ does not depend on the slopes s_k , it does not affect the location of the global minimum of $\varepsilon(s_k)$ and thus can be omitted when searching for the minimum.

Our local linearity assumption holds in most cases for two reasons. Firstly, the distortion distribution p_C has high kurtosis (see Appendix A) so that most of the distorted pixels are likely to lie in the same segment as the non-distorted pixel v . Secondly, even if a distorted pixel $\sim v$ moves to another segment, the slopes of two neighboring segments are usually very close to each other. This assumption has also been confirmed by our results, in which the tone-curves found using the accurate model from (2.5) and a simplified model from (2.11) were almost the same (see Section 2.5).

The most important consequence of using the simplified model from (2.11) is that the optimal tone-curve does not depend on the image compression error, as long as the compression distortions are not severe enough to invalidate the local linearity assumption. This means that the optimal tone-curve can be found independently of the compression algorithm and its quality settings.

The constrained optimization problem defined in (2.6) can now be re-written as follows:

$$\begin{aligned}
&\arg \min_{s_1, \dots, s_N} \sum_{k=1}^N \frac{p_k}{s_k^2} \\
&\text{subject to} \quad \sum_{k=1}^N s_k = \frac{v_{\max}}{\delta}
\end{aligned} \tag{2.12}$$

where $p_k = \sum_{l_k}^{l_{k+1}} p_L(l)$, and l_k and l_{k+1} define the lower and the upper bounds of a segment, respectively.

This problem can be solved analytically by calculating the first order Karush-Kuhn-Tucker (KKT) optimality conditions of the corresponding Lagrangian, which results in the following system of equations:

$$\left\{ \begin{array}{l} \frac{-2p_1}{s_1^3} + \lambda = 0 \\ \frac{-2p_2}{s_2^3} + \lambda = 0 \\ \vdots \\ \frac{-2p_N}{s_N^3} + \lambda = 0 \\ \sum_{k=1}^N s_k - \frac{v_{\max}}{\delta} = 0 \end{array} \right. \quad (2.13)$$

where λ is the Lagrange multiplier. The solution to the above system of equations results in the slopes s_k given by:

$$s_k = \frac{v_{\max} \cdot p_k^{1/3}}{\delta \cdot \sum_{k=1}^N p_k^{1/3}} \quad (2.14)$$

Note that the expression derived in (2.14) does not consider the upper bound constraint imposed on s_k in (2.7). Let I be the set of the index of a segment with a slope that exceeds the upper bound. We overcome the upper bound violation using the following adjustment:

$$s_k = \left\{ \begin{array}{ll} s_{\max} & \text{for } s \in I \\ \frac{(v_{\max} - \sum_{i \in I} s_{\max} \cdot \delta) \cdot p_k^{1/3}}{\delta \cdot \sum_{j \notin I} p_j^{1/3}} & \text{for } s \notin I \end{array} \right. \quad (2.15)$$

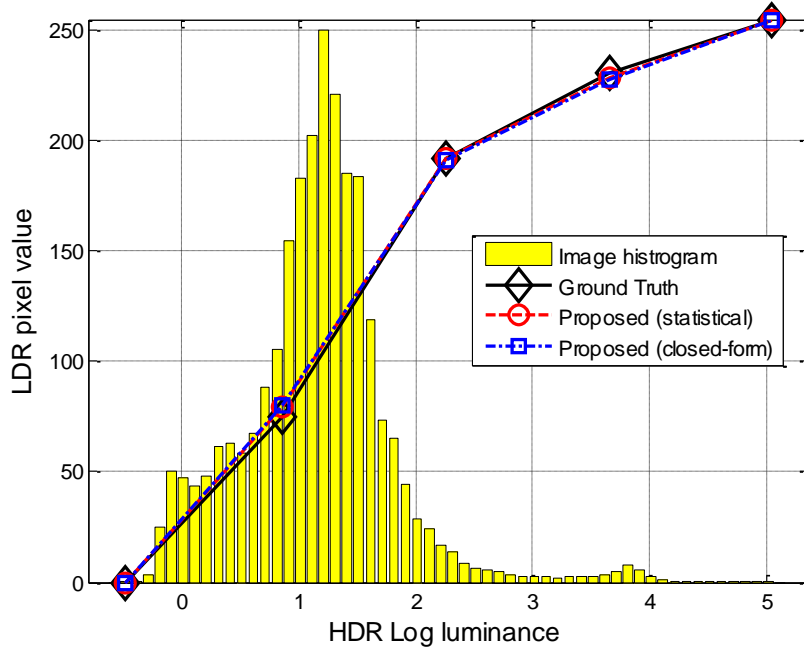
2.3 Experimental Results and Discussion

In this section we first validate the proposed methods: optimization using the statistical model proposed in Section 2.2.2 and the closed-form solution based on a simplified model derived in Section 2.2.4. Then, our models are further analyzed based on the generated tone curve and the distortion of the reconstructed HDR content. The performance of our models is also evaluated by comparing it with existing tone-mapping methods. We use H.264/AVC encoding as an example to demonstrate the results. In the experiments below, all tone-mapped images are compressed/decompressed using the intra mode of the H.264/AVC reference software [96] except for Section 2.3.5 where inter-frame mode is also used. To reconstruct an HDR image from a decoded LDR image, an inverse tone-mapping function is stored as a lookup table with each encoded image.

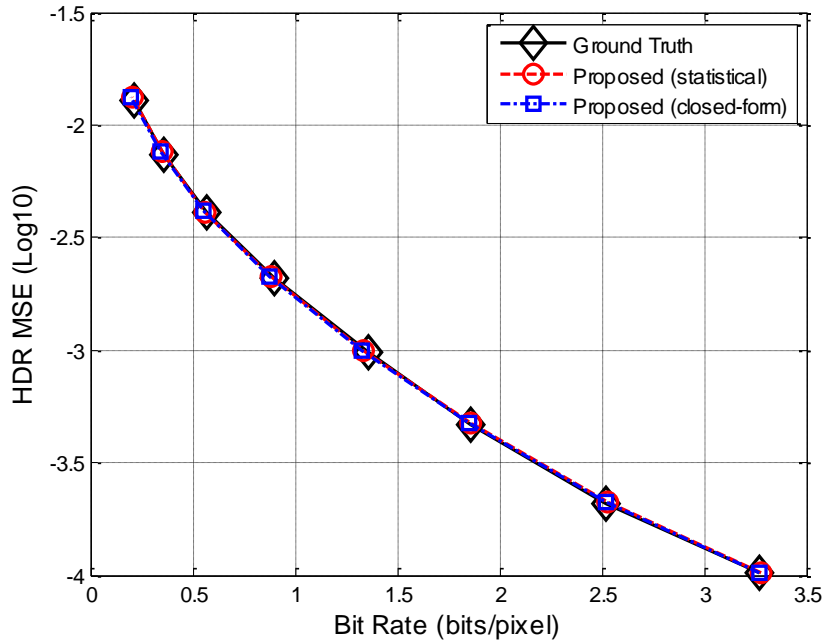
2.3.1 Model Validation

In this section, we validate that the statistical model of Section 2.2.2 results in a tone-curve that truly reflects the ground-truth results. Ground-truth results are achieved using the ideal scheme illustrated in Fig. 2.3(a), where the actual H.264/AVC encoder and decoder are employed to find the truly optimal piecewise linear tone curve. This ideal scheme is extremely computationally expensive, and its complexity increases exponentially with the number of segments. To make the experiment computationally feasible, we divided the tone curve into four segments of equal width. That is, the dynamic range of each of the segments is identical. Then the ideal scheme is used to find our ground-truth four-segment tone curve.

Fig. 2.3(a) demonstrates the tone curves generated by the statistical (Section 2.2.2), the closed-form (Section 2.2.4) and the ground-truth approaches for the H.264/AVC quantization parameter $QP = 22$. It can be seen that the tone curves produced by the proposed models are



(a) Tone curve for image “Memorial”, QP = 22



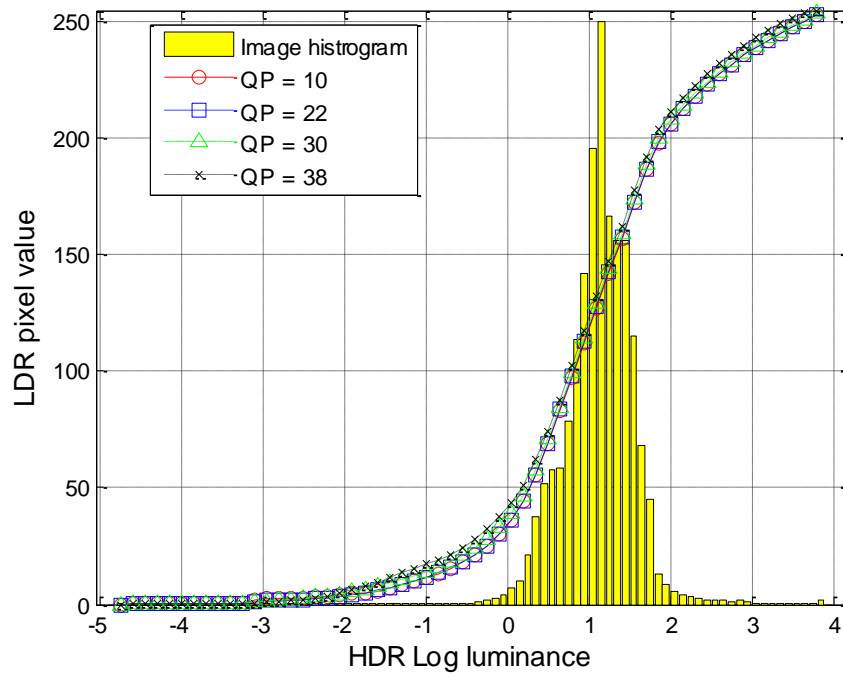
(b) HDR MSE vs. Bit Rate, image: “Memorial”

Fig. 2.3: Validation of the proposed models by comparison with the groundtruth solution. The top figure, (a), shows the tone curves computed using the statistical model, the closed-form solution and the ground-truth optimization for the image “Memorial”. The x axis denotes the HDR luminance in the log-10 scale, and the y axis is the LDR pixel value. (b) demonstrates the result of HDR MSE (in log10 scale) vs. bit rate (bits/pixel). The lower the MSE value, the better the image quality.

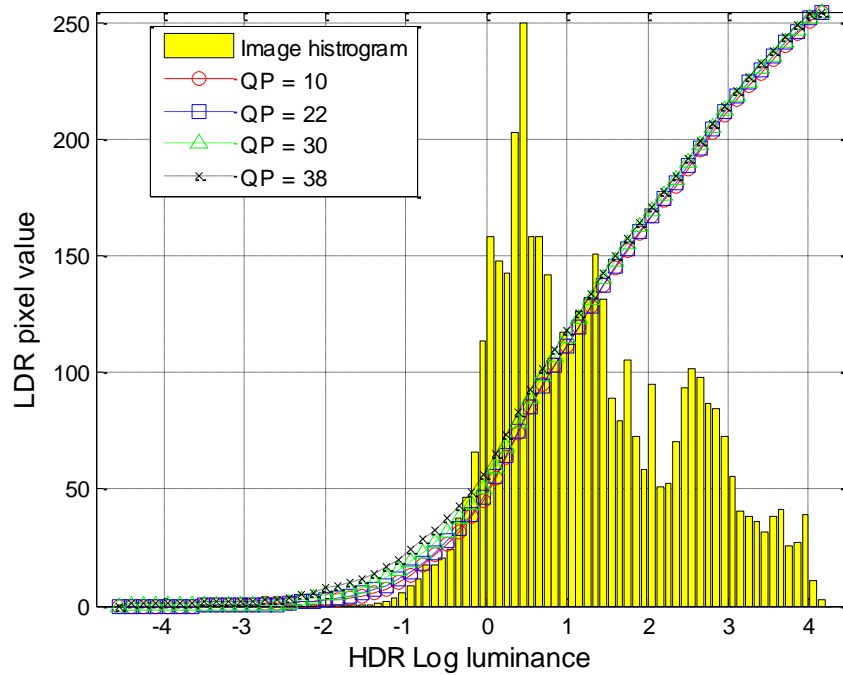
very close to the groundtruth curve. Fig. 2.3(b) shows the rate-distortion result in terms of bit rate vs HDR MSE. The results show that for different encoding bit rates, the reconstructed HDR images resulting from the two proposed models have very similar MSE relative to the ground-truth case. This further validates that the performance of the ideal scenario can be closely estimated by our statistical model and that the local linearity assumption we used to derive the closed-form solution is justified.

2.3.2 Dependence of the Tone Curves on QP

Next, we verify that the proposed statistical model can be well approximated by the closed-form solution which produces a tone curve that is independent of QP. The probability distribution of the H.264/AVC compression errors, which is a function of QP, is included in the statistical model proposed in Section 2.2.2. This suggests that the generated tone curve should vary with the value of QP. However, we observed from experiments on a large pool of HDR images encoded at different QPs that the variation in QP has no significant effect on the choice of an optimal tone curve. Fig. 2.4 illustrates this observation with an example of two images and their corresponding tone curves derived from the statistical model for different QP values (the larger the value of QP, the larger the compression error). The figures show that the tone curves are not significantly affected by the variation of QP.



(a) Tone curves for the analytical model, image: “AtriumNight”



(b) Tone curves for the analytical model, image: “Desk”

Fig. 2.4: Tone curves generated using the statistical model with different QP values for the images “AtriumNight” and “Desk”. The notation of the axis is the same as Fig. 2.3. The smaller the value of QP, the better the compression quality. 87 and 88 segments are used for “AtriumNight” and “Desk” respectively.

2.3.3 Further Analysis of the Closed-Form Solution

The tone curve resulting from the closed-form solution given by (2.14) can be generalized as follows:

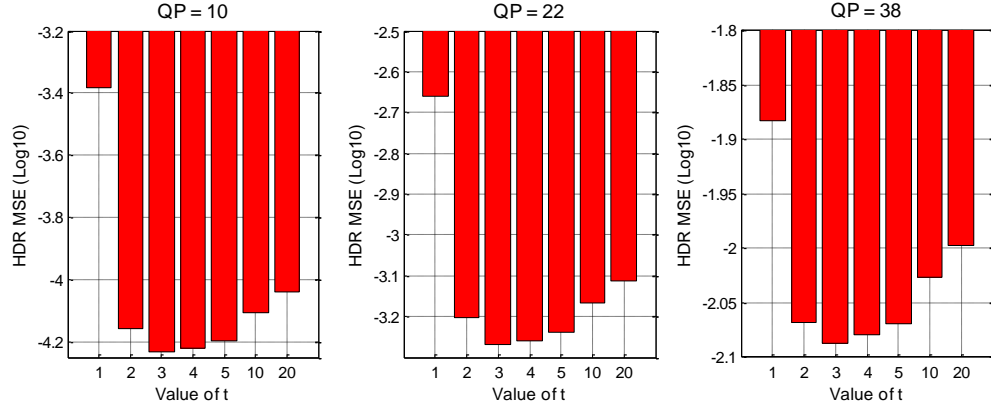
$$s_k = \frac{v_{\max} \cdot p_k^{1/t}}{\delta \cdot \sum_{k=1}^N p_k^{1/t}} \quad (2.16)$$

In our closed-form solution, t is set to equal to 3. Note that when $t = 1$, (2.16) is identical to the histogram equalization operation. Therefore, we will investigate the performance of the tone curves obtained from changing the exponent t of (2.16).

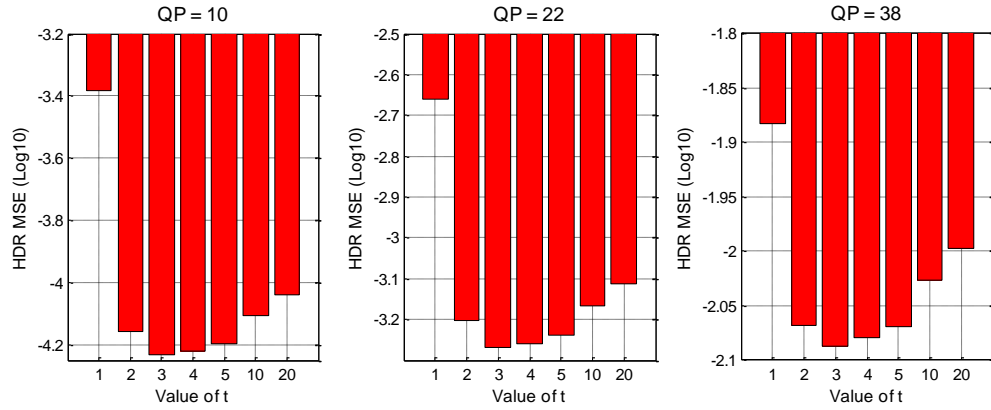
In the experiment, we set $t = 1, 2, 3, 4, 5, 10$ and 20 , and compressed the tone-mapped image using H.264/AVC at different QPs, and evaluated the distortion of the reconstructed HDR image. In addition to HDR MSE, we also used the popular quality metric known as the Structural Similarity Index (SSIM) [92] in order to find which t value gives the highest quality for a particular quality metric. Fig. 2.5 shows the resulting average performance of over 40 HDR images. The left row in the figure indicates that our closed-form solution ($t = 3$) is largely better than the histogram equalization ($t = 1$) and outperforms all other cases for HDR MSE. This can be expected, since our approach explicitly minimizes MSE. However, the same behavior cannot be expected from SSIM: the difference among all cases is minimal for light and medium compression, while the case of ($t = 2$) performs slightly better for strong compression.

From a practical point of view, HDR content is usually prepared for high-quality visual experience where only light or medium compression quantization is allowed. In this sense,

the results demonstrated in Fig. 2.5 indicate that our closed-form solution ($t = 3$) guarantees good performance.



(a) HDR MSE (Average over 40 Images)



(b) SSIM (Average over 40 Images)

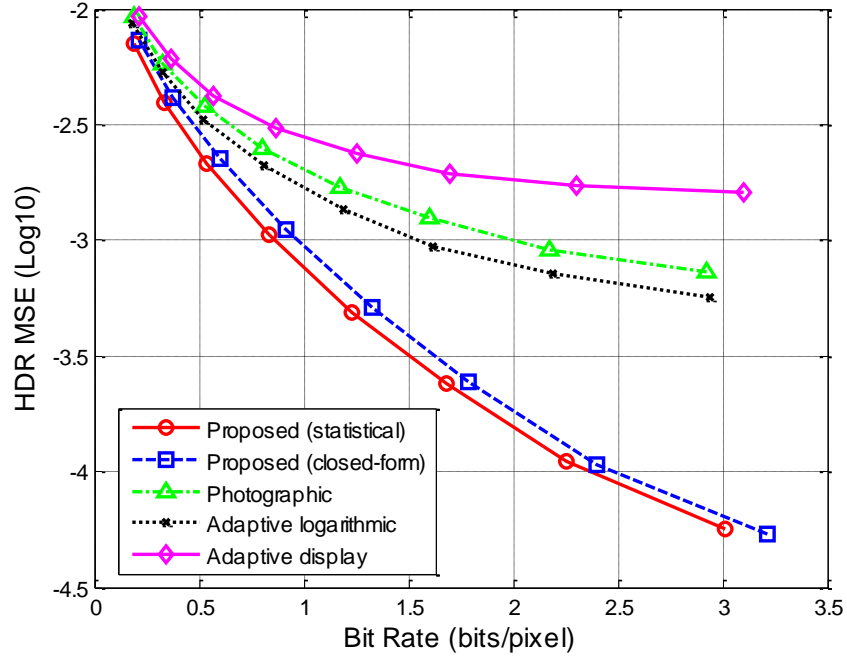
Fig. 2.5: Distortion measures for the reconstructed HDR images using the generalized solution (see (2.16)) with different values of t , averaged over 40 images. The tone-mapped images are compressed with different quality (QP = 10, 22 and 38), decoded and used to reconstruct HDR images. The left row, (a), shows the measurement of HDR MSE, where the smaller the value, the better the image quality. (b) compares the SSIM quality. Higher SSIM values mean better quality. For each of the 40 images, the segment width is set to be 0.1.

2.3.4 Comparison with Existing TMOs

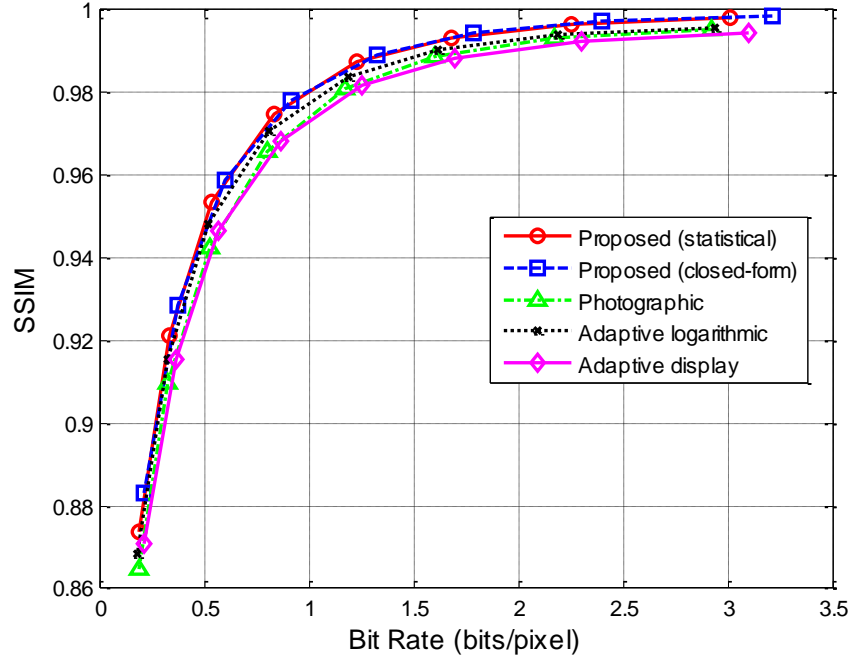
In this subsection, we compare the performance of the proposed models to existing tone-mapping methods. The chosen TMOs are the photographic TMO [38], the adaptive logarithmic TMO [50] and the display adaptive TMO [37]. In [49], a study was conducted to find how different TMOs perform when the LDR is inversely tone-mapped. Of the TMOs that were compared, the photographic TMO and the adaptive logarithmic TMO were found to outperform other popular tone-mapping methods in backward-compatible HDR image compression. However, the display adaptive TMO is a recent tone-mapping algorithm at the time of writing and it employs a similar optimization loop as our technique.

Fig. 2.6 compares the distortion of the reconstructed HDR image versus the compressed LDR bit rate for different TMOs, averaged over 40 images. This test demonstrates how successful each TMO is at delivering a good quality HDR by inverse tone mapping the corresponding LDR representation. The results show that our proposed methods clearly outperform the other methods in terms of MSE. The difference is not very large for low bit rates (heavy compression), but the performance of our model dramatically improves when the compression reaches the point of the medium and light compression. For HDR $MSE = -3$ (in \log_{10} scale), which corresponds to $QP = 25$, we save about 50% of the bit-rate compared to the best performing competitive TMO for the same quality. Fig. 2.6 also shows that the proposed statistical model results in a better MSE performance compared to the closed-form solution.

Although our models are designed for minimizing MSE, the results also indicate that the proposed TMOs show superior performance for the advanced quality metric, SSIM. In



(a) HDR MSE vs. Bit Rate (Average over 40 Images)



(b) SSIM vs. Bit Rate (Average over 40 Images)

Fig. 2.6: Comparison with other tone-mapping methods in terms of MSE and SSIM (for the reconstructed HDR image) vs. bit rate, averaged over 40 images. The correlation between image quality and the distortion measures (MSE and SSIM) can be referred to the caption of Fig. 2.5. MSE and SSIM for all methods represent the reconstruction error without the correction of the residual/enhancement layer. In the experiment, the segment width of each image histogram is set to be 0.1.

terms of SSIM, Fig. 2.6(b) shows that our proposed TMOs result in a better performance and the improvement is sustained for higher bit rates.

Fig. 2.7, Fig. 2.8 and Fig. 2.9 display the tone curve, rate-distortion curves and tone-mapped LDR images for three images. Additional results for more images are included in the supplementary material. The LDR images shown in these figures demonstrate that the images tone-mapped using our method also provide good quality. To further demonstrate the quality of the LDR images generated by the proposed models, Fig. 2.10 shows the distortion maps of the LDR images compared with their original HDR counterparts. The distortion maps were generated using the dynamic range independent image quality metric [97], which is the only available computational metric capable of comparing HDR and tone-mapped images. The metric visualizes the areas where the visible contrast is lost (green color), or distorted (red color). The distortions maps indicate that the proposed method causes less contrast loss than the other tested tone-mapping operators.

The computer graphics community often uses a perceptually-based image difference predictor, HDR visible difference predictor (VDP) [98], to compare a distorted HDR image to a reference HDR image. We compared the different TMOs in terms of HDR-VDP and found that there is no consistent improvement or degradation in performance of the different TMOs when compression is applied at the LDR layer. Therefore, we do not include performance evaluations relative to HDR VDP here.

The computational complexity of most global tone mapping operators, such as the photographic and the logarithmic operators compared in our study, is linear to the number of pixels, i.e., $O(N)$. The same holds for our closed form solution, in which the most expensive part is computing an image histogram. The solution based on the statistical model is more

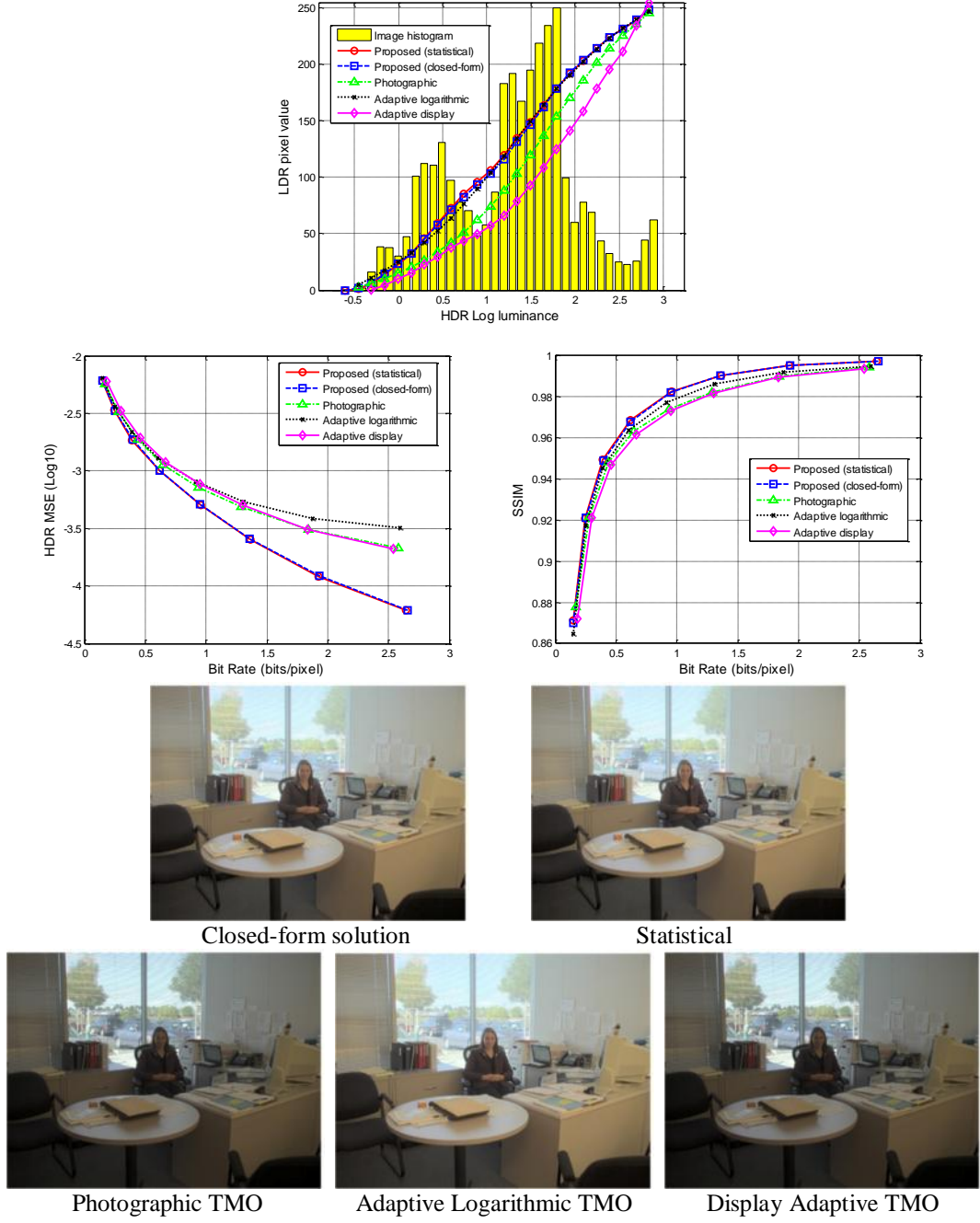


Fig. 2.7: Rate-distortion curves, tone curves and tone-mapped images for the image “Coby”. The first row demonstrates the resulting tone-curves with different TMOs, followed by the results for MSE and SSIM vs. bit rates; the second row shows tone-mapped LDR images using the proposed statistical model and the closed-form solution. The third row shows the tone-mapped images using the existing tone-mapping methods. All the tone-mapped images shown are compressed. The compression quantization parameters used for “Coby” is 10. The number of segments used for the histogram is 36.

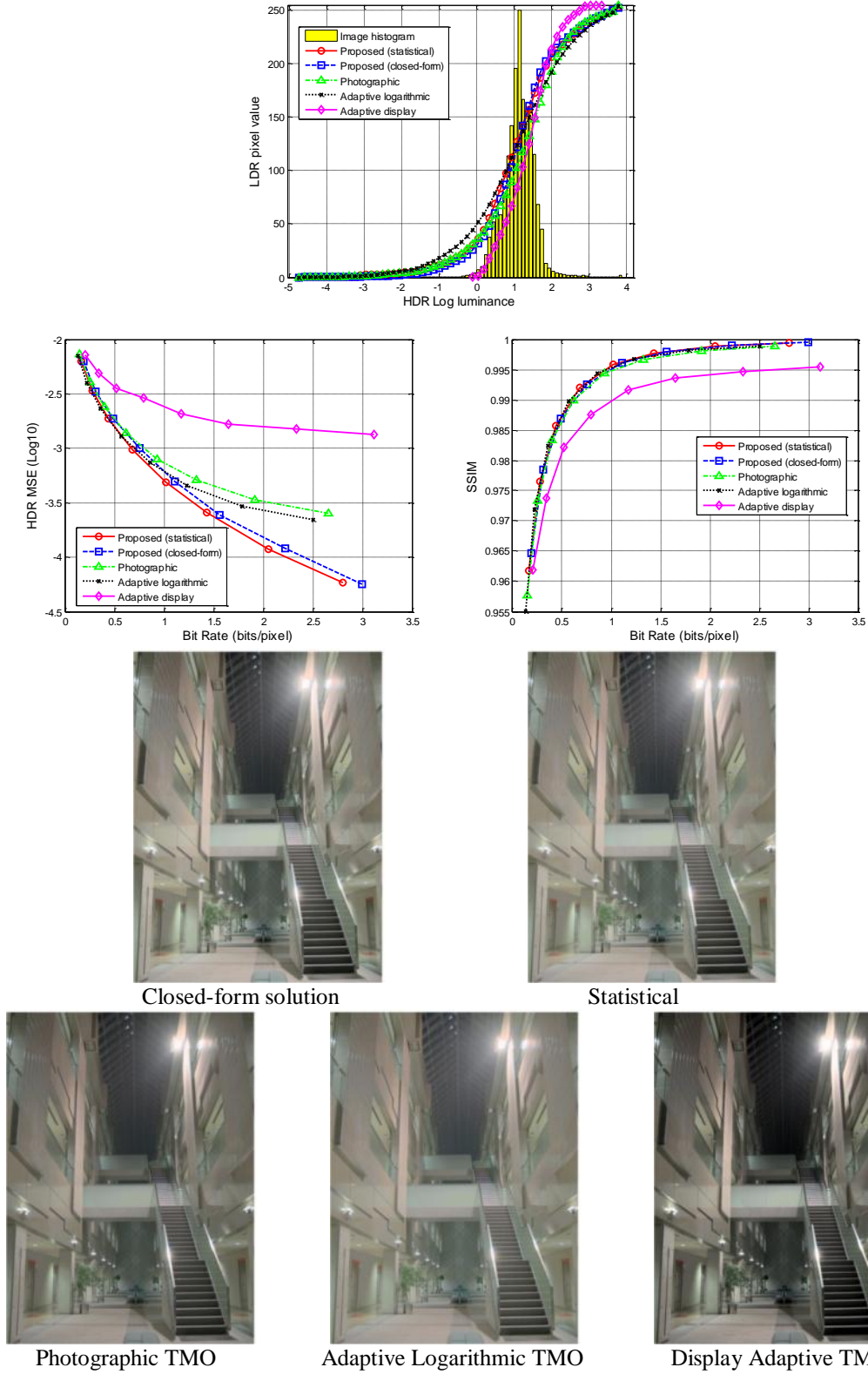


Fig. 2.8: Rate-distortion curves, tone curves and tone-mapped images for the image "AtriumNight". The notation is the same as Fig. 2.7. The compression quantization parameters used is 22. The number of segments used for the histogram is 87.

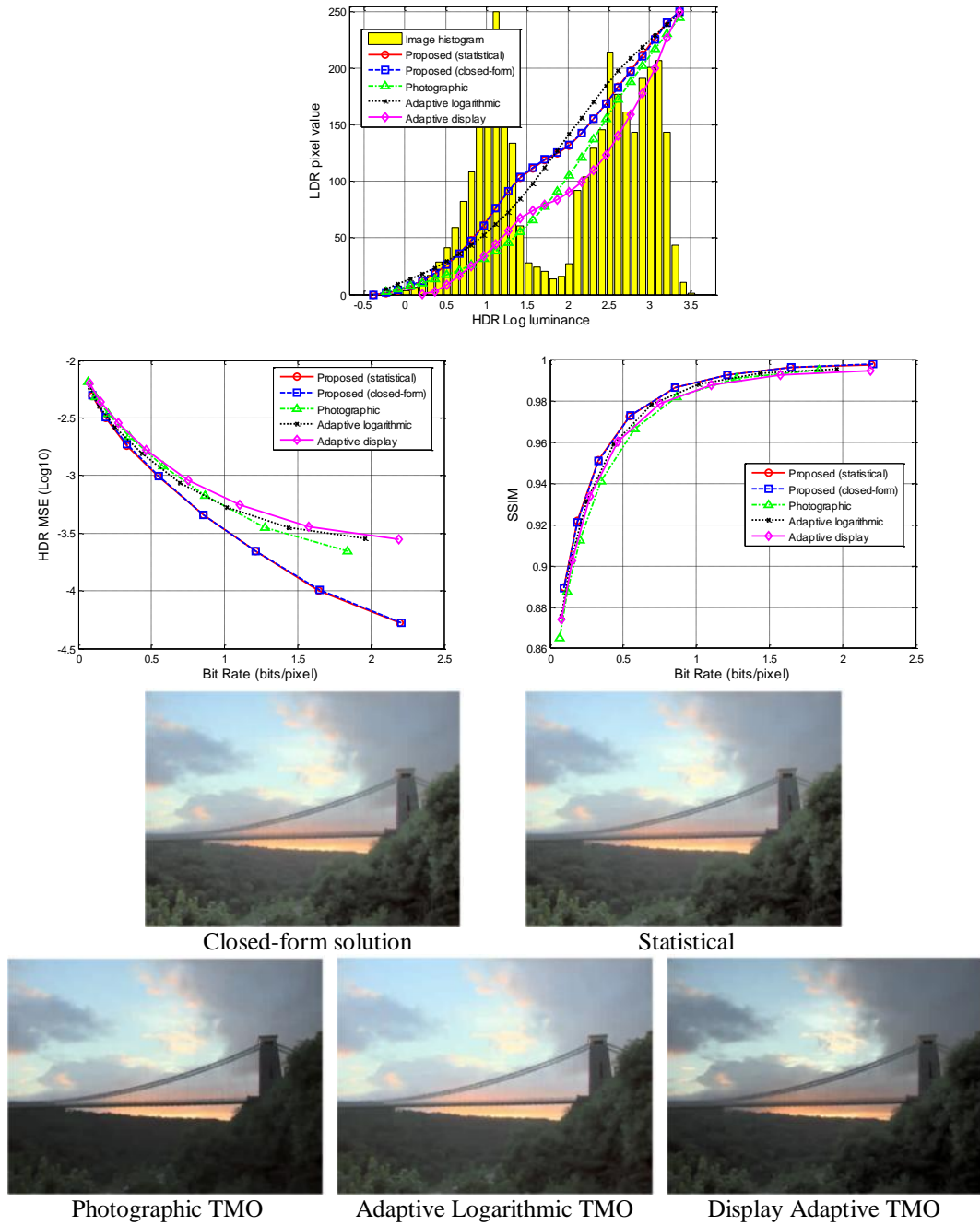
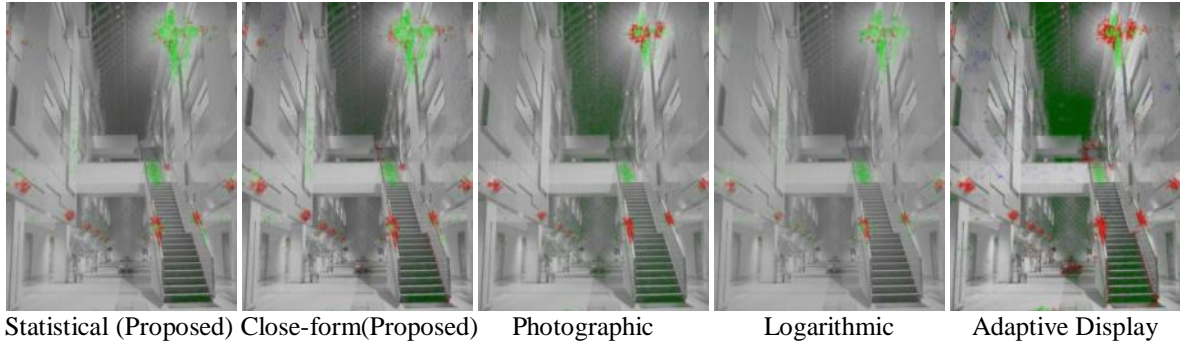


Fig. 2.9: Rate-distortion curves, tone curves and tone-mapped images for the image "BristolBridge". The notation is the same as Fig. 2.7. The compression quantization parameters used is 38. The number of segments used for the histogram is 39.

computationally expensive because it requires several iterations for the optimization procedure to converge. For comparison, the closed-form solution requires about 0.9 seconds and the statistical approach 20 seconds to complete using non-optimized MATLAB code on a 3 GHz CPU computer.



(a) AtriumNight, distortion maps of tone-mapped images (no compression)



(b) Coby, distortion maps of tone-mapped images (no compression)



(c) BristolBridge, distortion maps of tone-mapped images (no compression)

Fig. 2.10: Distortion maps of the LDR images relative to the original HDR images. The LDR images evaluated have not been compressed. In each of the distortion maps, three colors denote three different types of distortions: green for loss of visible contrast; blue for amplification of invisible contrast; red for reversal of visible contrast. The higher intensity of a color correlates with higher distortion of that type. In general, the less colored regions and lighter color intensity denote better LDR image quality.

2.3.5 Adapting Tone Curves for JVT Bit-Depth Scalable Encoding

In this section, we demonstrate the efficiency gains that can be expected when the proposed tone-mapping technique is used in combination with the Joint Video Team (JVT) bit-depth scalable extensions [35]. For that purpose, we tone-map the JVT standard sequences [99] using our closed-form method, and then compare the compression performance between the sequences generated by our method and the generic tone-mapping used in the test sequences.

The JVT test sequences are provided as 10-bit extended dynamic range frames and the corresponding 8-bit tone-mapped frames. The 10-bit frames contain gamma corrected footage from a high-end camera that can capture an extended dynamic range. We assume that the gamma correction has a similar effect as the logarithmic function that we apply to linear luminance values to account for the Weber law. Therefore, we use the 10-bit frames directly as an input to our algorithm. One important issue to consider is flickering, which our method can cause when used on video sequences. This is because the computed tone-curves solely depend on scene content, which can abruptly change from frame to frame. To prevent such flickering, we apply a low-pass filter to the generated tone-curves, identical to that in [37]. For comparison, we also generate tone-mapped sequences without the temporal filter.

Fig. 2.11 shows the comparison of the compression performance for our method (closed-form solution) with and without the temporal filtering, and the generic tone-mapping used for the JVT test sequences. The temporal filtering did not change significantly compression performance for these test sequences (compare black and red curves in Fig. 2.10). This is because the frames did not contain any abrupt scene changes that could cause flickering. For two sequences (Freeway and Waves), our method gave a significant

improvement over a generic tone-mapping that was not optimized for video compression (compare red and dashed-blue curves). The improvement is especially large for higher bit-rates. For the third sequence (Plane), the compression performance was very similar for both methods. Slightly worse results for our method for medium bit-rates can be explained by the approximations used in the closed-form solution. The tone-mapped sequence provided by the JVT was coincidentally well conditioned for video compression. However, large improvements for the two remaining sequences illustrate the gains that can be expected from the proposed method when used in combination with the JVT bit-depth scalable coding.

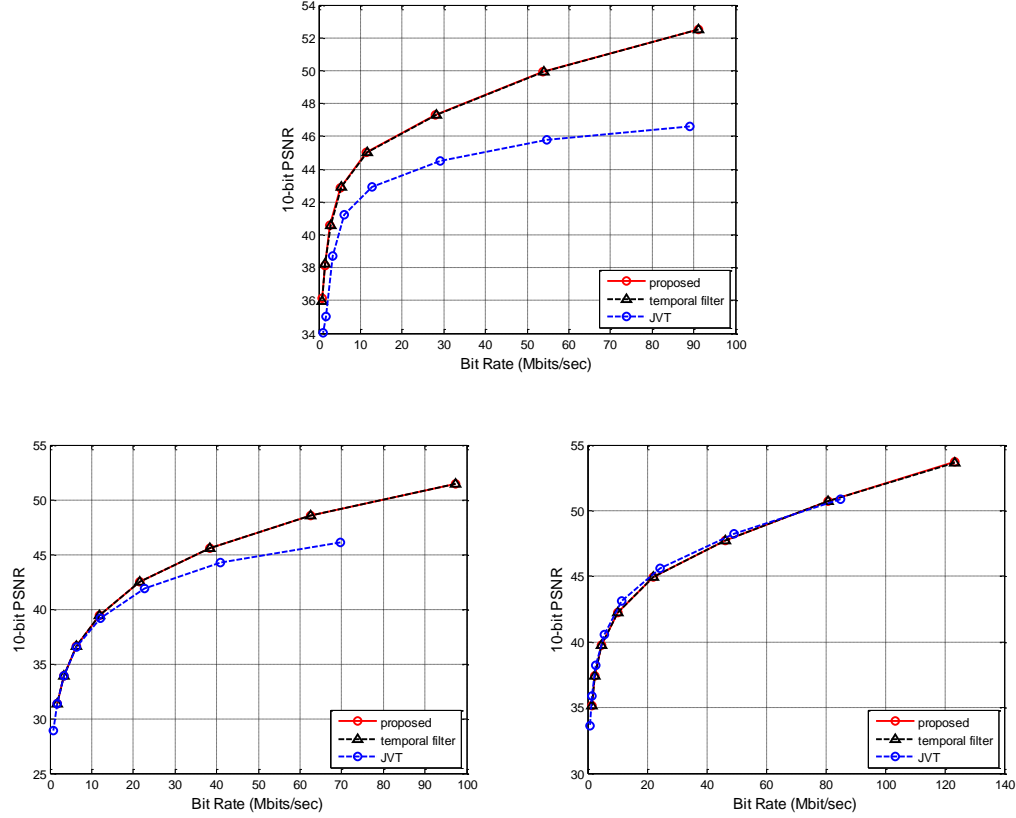


Fig. 2.11: Compression performance for the proposed method used with the JVT bit-depth scalable encoding. The comparison is made between the closed-form solution of our method (solid-circle), the same method but with the temporal filter that prevents flickering (dashed-triangle), and the generic tone-mapping used in the JVT test sequences (dashed-circle). The x-axis denotes the bit rate in Mbit/s at 30Hz of frame rate, and the y-axis is the PSNR between the original 10-bit video and its inversely-tone-mapped version. 100 segments are used in the histogram for each of three sequences.

2.4 Discussion

Our proposed tone-mapping methods can directly improve the compression efficiency of bit-depth scalable video coding. The proposed methods are designed to produce a better reconstructed HDR representation. A direct consequence of this design objective is a reduction in the size of the higher bit-depth enhancement layer which contains the difference between the reconstructed HDR image and the original one. Thus, our method can lower the total bit rate for the bit-depth scalable coding.

Although the primary goal of tone-mapping algorithms is to optimize the visual quality of the displayed LDR image, we instead designed a tone-mapping operator that optimizes the compression performance in the backward-compatible encoding scheme. Numerous algorithms in the literature have considered the primary objective that explicitly focuses on producing good quality tone-mapped images. This group of algorithms include the photographic [38], logarithmic [50] and display adaptive [37] TMOs considered in this paper. The tone-curve that meets both objectives can be approximated by the linear combination of the tone-curves produced by our method and these algorithms. Moreover, our study shows that the overall quality of images tone-mapped with our method is comparable to other tone-mapping algorithms and none of the tone-mapped images we generated was considered as unacceptable. This means that for the applications that do not require a finely adjusted backward-compatible layer, our method can be used directly.

Our considerations are limited to the global tone-curves used for the entire frame while many modern tone-mapping methods use local (spatially varying) processing to retain more details and produce better looking images. However, the study in [49] showed that local tone-mapping operators result in worse compression performance than global operators. This

suggests that there is a trade-off between choosing a tone-mapping operator that is optimal for preserving HDR information in compressed images and an operator that produces the best looking images. Such a trade-off will be studied in the next chapter.

Finally, tone-mapping each frame of a video sequence independently can produce flickering since the tone-curve can change rapidly from frame to frame. This, however, can be avoided when a low-pass filter is applied on the sequence of computed tone-curves, as done in [37].

2.5 Conclusion

In this chapter, we showed that the appropriate choice of a tone-mapping operator (TMO) can significantly improve the reconstructed HDR quality. We developed a statistical model that approximates the distortion resulting from the combined processes of tone-mapping and compression. Using this model, we formulated a constrained optimization problem that finds the tone-curve which minimizes the expected HDR MSE. The resulting optimization problem, however, suffers from high computational complexity. Therefore, we presented a few simplifying assumptions that allowed us to reduce the optimization problem to an analytically tractable form with a closed-form solution. The closed-form solution is computationally efficient and has a performance compatible to our developed statistical model. Moreover, the closed-form solution does not require the knowledge of QP, which makes it suitable for cases where the compression strength is unknown. Although our models are designed to minimize HDR MSE, extensive performance evaluations show that the proposed methods provides excellent performance in terms of SSIM and the LDR image quality, in addition to an outstanding performance in MSE.

3 Visually Favorable Tone-Mapping with High Compression Performance in Bit-Depth Scalable Video Coding

The compression efficiency of a backward compatible HDR image/video coding scheme depends on i) the LDR representation, and ii) the inversely-tone-mapped HDR signal that affects the coding performance of the enhancement layer. Both of these signals are closely determined by the tone-mapping operator. In the previous chapter we showed that the appropriate choice of a tone-mapping operator can significantly improve the reconstructed (inversely tone-mapped) HDR quality. However, the bit-rates of the base layer and the enhancement layer were not considered explicitly. In addition to compression performance, tone-mapping should deliver visually pleasing content and this is the main intention of most existing tone-mapping operators. Based on these observations, this chapter describes the design of a tone-mapping operator that preserves a preset LDR perceptual quality while minimizing the base layer bit-rate and the enhancement layer bit-rate. Our solution also considers the temporal consistency between consecutive frames in order to avoid flickering artifacts, which may be problematic for most existing tone-mapping methods as they are specifically designed for still images.

3.1 Problem Statement

In bit-depth scalable video coding, the high bit-depth input video signal is split into a tone-mapped 8-bit LDR base layer and a high bit-depth enhancement layer. The LDR base layer encodes a tone-mapped representation of the input video while the enhancement layer encodes the residual between the original video and inverse tone-mapped base layer representation. These two layers are strongly affected by the tone-mapping operator.

Moreover, the encoding of these two layers determines the coding efficiency of the resulting scalable video bitstream. Therefore, it is imperative to characterize the bit-rates of each of the base and enhancement layer as a function of the tone-mapping parameters and then find the parameters that minimize these bitrates while adhering to a prescribed LDR perceptual quality and maximizing the high-bit depth reconstruction quality. The temporal aspect of tone-mapping also affects the coding efficiency. Tone-mapping every individual frame separately may result in temporal inconsistencies (or flickering) in the LDR video sequence, which in turn negatively affects the inter-frame coding.

Our goal is to derive a global (spatially invariant) tone-mapping operator that minimizes the overall bit rate (both base layer and enhancement layer) while preserving a desired perceptual quality for the tone-mapped LDR video content. We develop numerical models R_b and R_e that estimate the bit rates of the base and the enhancement layers respectively, and incorporate them in our optimization problem. In order to achieve good perceptual quality, we introduce a quality mismatch term Q_L as the mean square error (MSE) between the LDR signal produced by our proposed scheme and the pre-set tone-mapping method. Despite the fact that MSE itself has shortcomings in reflecting the human visual response, we will show that our solution can benefit from this simple metric in a perceptual sense. Our tone-mapped video is shifted away from the pre-set TMO; however, it maintains the desired perceptual quality in the resulting LDR video. Moreover, we introduce a sensitivity term, Ω , that restricts the temporal flickering artifacts by ensuring that the average brightness of consecutive frames changes smoothly based on the HDR input.

In practice, the HDR video should conform to video coding standards that support the unsigned integer format as opposed to the floating-point formats, such as EXR and RGBE,

which are mostly used for still images. Our tone-mapping operator is designed in a way that it can be applied directly to integer-valued high-bit-depth video.

We denote by Y the original HDR signal; and let z be the reference LDR signal obtained using the predefined tone-mapping scheme. Let y be the achieved uncompressed LDR representation using our proposed scheme. Denote by \tilde{y} the compressed LDR and \tilde{Y} the reconstructed HDR achieved by inversely tone-mapping \tilde{y} . As mentioned before, we use Q_L as the estimate for the LDR quality given by the MSE between z and y ; while R_b and R_e are the estimates for the base layer (LDR) and enhancement layer (HDR residual) bit-rates, respectively.

Since our goal is to improve the compression efficiency of the bit-depth scalable bit-stream, we propose a tone-mapping scheme that solves the following constrained optimization problem:

$$\min_{\theta} w_1 R_e(\theta, t) + w_2 R_b(\theta, t) + w_3 Q_L(\theta, t, z) \quad (3.1)$$

subject to:

$$\Omega(\theta, Y_{t-1}, Y_t, y_{t-1}) \leq 0. \quad (3.2)$$

where θ denotes the tone-mapping parameters, and w_1 , w_2 , and w_3 are regularizing constants. The function Ω is the mismatch in average brightness between two consecutive LDR frames. The constraint (3.2) is used to keep temporal consistency between consecutive video frames using the information from the previous and the current HDR frames as well as the previous tone-mapped frame.

In what follows, we develop explicit formulations of each of R_b , R_e , Q_L and Ω as convex functions of the tone-mapping parameters θ . Consequently, problem (3.1) is a convex constrained optimization problem which can be solved in polynomial time. The solution to

problem (3.1) is the parameters that define the tone-mapping curve the luminance channel. In general, capitalized symbols in this chapter denote high-bit-depth signals while the lower case ones correspond to their 8-bit counterparts. A symbol with a tilde (\sim) denotes the distorted version of the variable.

3.2 Proposed Solution

In this section, we define the tone-mapping parameters θ and formulate each of the terms in our optimization problem (3.1) as functions of θ .

3.2.1 Tone-Mapping Parameters

For combined flexibility and tractability, we use a piecewise linear tone-curve as shown in Fig. 3.1.

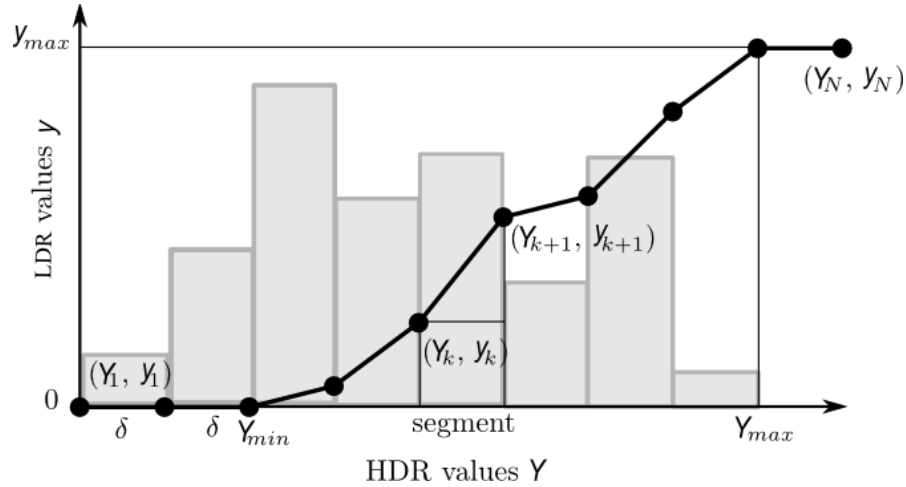


Fig. 3.1: Parameterization of a tone-curve. The grey bar-plot represents the image histogram used to compute $p(k)$. The actual number of segments is usually much greater than that in this illustration.

The tone-curve is divided into a number of segments (N) of equal and small enough width δ , with nodes (Y_k, y_k) defining a segment edge. The segment height is given by:

$x_k = y_{k+1} - y_k$, and the forward tone-mapping function is defined as:

$$y(Y) = (Y - Y_k) \cdot \frac{x_k}{\delta} + y_k, \quad (3.3)$$

where y is the LDR value that corresponds to the HDR value Y . k denotes the index of segment that Y belongs to.

The tone-curve can then be fully defined using a set of segment heights x_k , that is $\theta = [x_1, x_2, \dots, x_N]^T$. Since the aim is to find a TMO that minimizes (3.1), we model the LDR quality (Q_L), the LDR bit-rate (R_b), the enhancement layer bit-rate (R_e) and the difference of overall brightness between consecutive LDR frames (Ω) as functions of the parameter vector θ . This allows us to avoid the expensive computation of each of these costs. Without the numerical models of these cost, the calculation of each of these costs would have required us to actually tone-map and encode every frame for each change in the parameters.

3.2.2 LDR Quality Mismatch Q_L

Our objective is to find a tone-curve that achieves better compression efficiency while maintaining a predefined LDR picture quality. Therefore, we introduce the LDR quality mismatch term Q_L to measure the mean squared error (MSE) between the LDR picture z produced using a perceptual TMO and the LDR picture y resulting from the minimization of (3.1).

We define the cost $Q_L(\theta)$ as the expected value of the norm $\|z - y(\theta)\|_2^2$. We will show that $\|z - y(\theta)\|_2^2$ can be well approximated by a quadratic function of x_k . First, we find the piecewise linear approximation of the TMO that produces u . Let the nodes of the reference curve be (Y_k, z_k) and its heights be h_k ; also let $z(Y) = (Y - Y_k) \cdot \frac{h_k}{\delta} + z_k$ be the corresponding forward tone-mapping function using the reference tone curve. Then the expected value of $\|z - y(\theta)\|_2^2$ can be expressed as

$$E[\|z - y\|_2^2] = \sum_{k=1}^N \sum_{Y=Y_k}^{Y_{k+1}} ((Y - Y_k) \frac{h_k}{\delta} + z_k - ((Y - Y_k) \frac{x_k}{\delta} + y_k))^2 Pr(Y) \quad (3.4)$$

where N is the number of segments and $Pr(Y)$ is the probability of the HDR values Y in the image. Assume that the expected value of the HDR values in each segment is located in its center. That is, $E(Y) = \frac{Y_k + Y_{k+1}}{2} + z_k$ for $Y_k \leq Y \leq Y_{k+1}$.

By rearranging (3.4), we have

$$E[\|z - y\|_2^2] = \sum_{k=1}^N (\frac{h_k}{2} + \sum_{j=1}^{k-1} h_j - (\frac{x_k}{2} + \sum_{j=1}^{k-1} x_j))^2 \cdot p(k) \quad (3.5)$$

Notice that the term $\frac{h_k}{2} + \sum_{j=1}^{k-1} h_j$ can be pre-computed and is independent of the optimization.

$\frac{x_k}{2} + \sum_{j=1}^{k-1} x_j$ is linear in x . If we denote $\eta(x_k) = \sum_{k=1}^N (\frac{h_k}{2} + \sum_{j=1}^{k-1} h_j - (\frac{x_k}{2} + \sum_{j=1}^{k-1} x_j))^2$, we can now

express the expected value of the LDR mismatch as:

$$E(Q_L) = \sum_{j=1}^{k-1} \eta(x_k) p(k) \quad (3.6)$$

3.2.3 HDR Bit-Rate R_e

We model the enhancement layer bit-rate R_e as a function $\|Y - \tilde{Y}\|_2^2$, where \tilde{Y} is the high bit-depth signal produced by inverse tone-mapping (iTM) the compressed LDR base layer signal \tilde{y} . Our experiments have shown that there exists a high correlation between $\|Y - \tilde{Y}\|_2^2$ and R_e , which is illustrated in Fig. 3.2. We incorporate the model developed in our previous work [105] to approximate $\|Y - \tilde{Y}\|_2^2$ which leads to the estimation of R_e . We summarize the key points below.

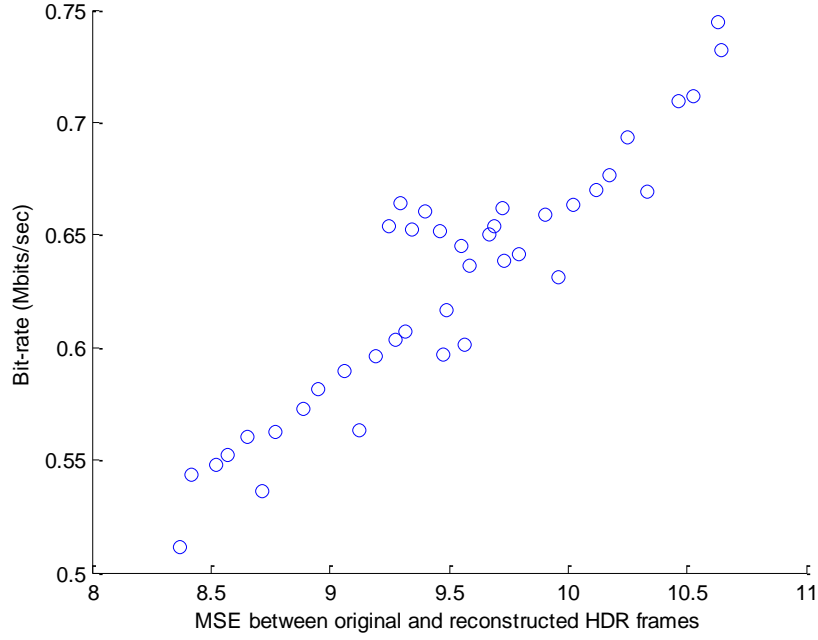


Fig. 3.2: Correlation between the enhancement layer bit-rate and $\|Y - \tilde{Y}\|_2^2$.

In [105] we estimated the expected value of $\|Y - \tilde{Y}\|_2^2$, by considering the distribution of the enhancement layer residual $Y - \tilde{Y}$. The resulting estimate is a function of the heights x_k and the HDR image histogram $p(k)$:

$$E(R_e) \approx c_1 \cdot \delta^2 \cdot \sum_{k=1}^N \frac{p(k)}{x_k^2} \quad (3.7)$$

where N denotes the number of segments of the video frame and c_1 is a constant related to the variance of the distribution function for the compression. The simplicity of this formulation contributes to the efficient performance of the optimization. The above model is incorporated into an optimization problem, and a closed solution is found:

$$s_k = \frac{y_{max} \cdot p(k)^{1/3}}{\delta \cdot \sum_{k=1}^N p(k)^{1/3}}. \quad (3.8)$$

3.2.4 LDR Bit-Rate R_b

In this section, we estimate the LDR bit-rate R_b as a linear function of the TMO parameters x_k . To that end, we follow the findings in [100] where it is shown that the intra-frame bit-rate R_b and the norm of the image gradient ζ are highly correlated, such that

$$R_b = c_2 \cdot \zeta \quad (3.9)$$

where c_2 is a constant given by the quantization parameter (QP) of the compression scheme.

The image gradient ζ can be expressed as follows:

$$\begin{aligned} \zeta &= \zeta_v + \zeta_h \\ &= \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |I(i, j) - I(i+1, j)| \\ &\quad + \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |I(i, j) - I(i, j+1)| \end{aligned} \quad (3.10)$$

where m and n are the dimensions of an image I , ζ_v and ζ_h are the vertical and horizontal gradient components, respectively.

Let $Y(i, j)$ be the current HDR pixel intensity indexed by row i and column j of the image, let Y' refer to the intensity of its immediate right/bottom pixel, and N the number of TMO segments. For every HDR pixel with intensity Y we find the conditional probability $Pr(Y' | Y)$ that the adjacent (right or bottom) pixel has an intensity equal to Y' . These conditional probabilities are computed prior to the optimization process and stored as lookup tables.

Letting $Pr(Y)$ be the probability of the current HDR value Y evaluated from the image histogram and also assuming that the mean of the HDR pixel intensities in each segment falls at its center, we use equation (3.3) to express the expected value of ζ_h in terms of x_k as shown below:

$$\begin{aligned}
\zeta_h &= \sum_{c=1}^N \sum_{Y=Y_c}^{Y_{c+1}} \sum_{a=1}^N \sum_{Y'=Y_a}^{Y_{a+1}} |y(Y) - y(Y')| Pr(Y' | Y) Pr(Y) \\
&= \sum_{c=1}^N \sum_{a=1}^N \left| \frac{\delta}{2} \frac{x_c}{\delta} + y_c - \left(\frac{\delta}{2} \frac{x_a}{\delta} + y_a \right) \right| d_h(c, a) p(c) \\
&= \sum_{c=1}^N \sum_{a=1}^N \left| \frac{x_c}{2} + \sum_{i=1}^{c-1} x_i - \left(\frac{x_a}{2} + \sum_{j=1}^{a-1} x_j \right) \right| d_h(c, a) p(c)
\end{aligned} \tag{3.11}$$

where c is the segment index of the current pixel, a is the segment index of the adjacent pixel, and $d_h(c, a) = Pr(Y_a \leq Y(i+1, j) < Y_{a+1} | Y_c \leq Y(i, j) < Y_{c+1})$. $p(c)$ is obtained from the HDR image histogram.

We can obtain a similar expression for the expected value of ζ_v . Letting $q(c, a) = (d_h(c, a) + d_v(c, a))p(c)$, the expected value of ζ is given by:

$$\begin{aligned}
E(\zeta) &= \sum_{c=1}^N \sum_{a=1}^N \left| \frac{x_c}{2} + \sum_{i=1}^{c-1} x_i - \left(\frac{x_a}{2} + \sum_{j=1}^{a-1} x_j \right) \right| q(c, a) \\
&= \sum_{k=1}^N \beta_k x_k
\end{aligned} \tag{3.12}$$

where

$$\begin{aligned}
\beta_k &= \sum_{i=1}^N \frac{q(i, k)}{2} + \sum_{j=1}^{k-1} \left(\frac{q(k, j)}{2} + \sum_{i=k+1}^N q(i, j) \right) \\
&\quad + \sum_{j=k+1}^N \left(\frac{q(k, j)}{2} + \sum_{i=1}^{k-1} q(i, j) \right)
\end{aligned} \tag{3.13}$$

Note that β_k depends only on the HDR image and can be computed prior to finding the TMO.

Though the model was derived for H.264/AVC coding, it can be applied to other coding methods. This is because most compression approaches favor low frequency components and the bit-rate of the compressed stream will be low if the gradient (amount of high frequency components) is low.

Please note that this model does not apply to inter-frames, which is reflected in the constraint of our optimization formulation by imposing the maximum difference in average brightness between consecutive LDR frames. We found that the difference in average brightness between consecutive tone-mapping frames (from the same HDR video) is correlated to the bit rate used for encoding the inter frames. Fig. 3.3 demonstrates this relationship.

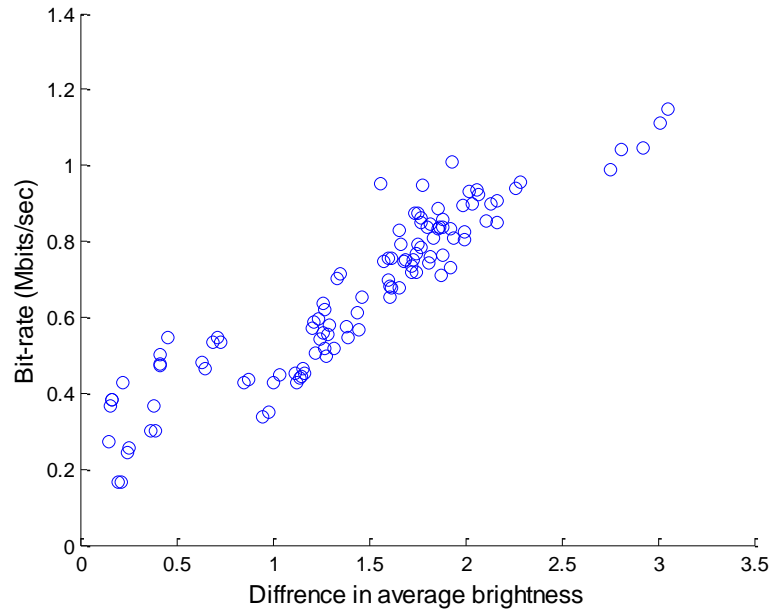


Fig. 3.3: Correlation between the temporal bit-rate and difference in average brightness between consecutive tone-mapped frames. The videos used in the test are all tone-mapped from the same HDR video sequence.

3.2.5 Temporal Consistency between Consecutive Frames

In order to avoid temporal inconsistency between consecutive tone-mapped video frames, we impose a constraint on the difference in overall brightness between the current (which is not known) and the previous LDR frames. This difference must meet two properties: i) If the current high-bit-depth frame is brighter than the previous one, then the current (being derived) LDR frame should be brighter than the previous tone-mapped frame, and vice versa. ii) The amount of difference for the LDR frames should be closely related to that for the high-bit-depth frames. Based on these observations, we formulate the difference constraint in LDR brightness (Ω) in terms of the tone-mapping parameter θ as well as the brightness of the current and the previous high-bit-depth frames and that of the previous tone-mapped frame. Denote the average brightness of a frame as $\Psi(\cdot)$ and D to be the maximum intensity value of the high bit-depth signal, e.g., 1023 for a 10-bit signal. The constraint can be written as:

$$\begin{aligned}
& (1 - \text{sgn}(\Psi(Y_t) - \Psi(Y_{t-1}))) \cdot \varepsilon \cdot \frac{255}{D} \cdot (\Psi(Y_t) - \Psi(Y_{t-1})) \\
& \leq \Psi(y_t) - \Psi(y_{t-1}) \\
& \leq (1 + \text{sgn}(\Psi(Y_t) - \Psi(Y_{t-1}))) \cdot \varepsilon \cdot \frac{255}{D} \cdot (\Psi(Y_t) - \Psi(Y_{t-1}))
\end{aligned} \tag{3.14}$$

where $\text{sgn}(\cdot)$ is the sign operator. The tolerance parameter ε is used to control how far can the brightness of the current frame differ from that of the previous one, and 20% is used in our experiment.

In practice, the values of average brightness $\Psi(Y_t)$, $\Psi(Y_{t-1})$ and $\Psi(y_{t-1})$ are known and can be computed easily. $\Psi(y_t)$ can be estimated by its expected value and be written as a function of the tone-mapping parameters:

$$\begin{aligned}
E(\Psi(y_t)) &= \sum_{k=1}^N \left(\frac{x_k}{2} + \sum_{j=1}^{k-1} x_j \right) \cdot p(k) \\
&= \sum_{k=1}^N \tau(k) \cdot x_k
\end{aligned} \tag{3.15}$$

where $\tau(k)$ is a simple linear combination of $p(k)$:

$$\tau(k) = \begin{cases} \frac{p(k)}{2} + \sum_{j=k+1}^N p(j) & \text{for } k = 1, \dots, N-1, \\ \frac{p(k)}{2} & \text{for } k = N. \end{cases} \tag{3.16}$$

The constraint in (14) can be re-organized as:

$$Q_l(t) \leq \sum_{k=1}^N \tau(k) \cdot x_k \leq Q_u(t) \tag{3.17}$$

where

$$\begin{aligned}
Q_l(t) &= (1 - \text{sgn}(\Psi(Y_t) - \Psi(Y_{t-1}))) \cdot \varepsilon \cdot \frac{255}{D} \\
&\cdot (\Psi(Y_t) - \Psi(Y_{t-1})) + \Psi(y_{t-1})
\end{aligned} \tag{3.18}$$

and

$$\begin{aligned}
Q_u(t) &= (1 + \text{sgn}(\Psi(Y_t) - \Psi(Y_{t-1}))) \cdot \varepsilon \cdot \frac{255}{D} \\
&\cdot (\Psi(Y_t) - \Psi(Y_{t-1})) + \Psi(y_{t-1})
\end{aligned} \tag{3.19}$$

3.2.6 Optimization Formulation

Using the models described above, the optimal tone-mapping curve is obtained by finding the TMO parameters $\theta^* = \{x_1, x_2, \dots, x_N\}$ that minimizes the joint cost of the expected values of R_e , R_b and Q_L :

$$\begin{aligned}
& \min_{\theta} \quad \sum_{k=1}^N (w_1 \cdot E(R_e) + w_2 \cdot E(R_b) + w_3 \cdot E(Q_L)) \\
& \text{s.t.} \quad \sum_{k=1}^N x_k = y_{max} \\
& \quad \quad \frac{x_k}{\delta} \leq 1 \\
& \quad \quad Q_l(t) \leq \sum_{k=1}^N \tau(k) \cdot x_k \leq Q_u(t)
\end{aligned} \tag{3.20}$$

The first constraint ensures that the tone curve will cover exactly the entire range of LDR pixel values (255). The second constraint indicates the slope of each segment is not greater than one. This avoids allocating more intensity levels for a segment in the LDR version than the number of levels encoded in the HDR segment. The last constraint eliminates the temporal inconsistency between consecutive tone-mapped frames. The weighting factors w_1 , w_2 and w_3 can be adjusted according to users' preference or the picture content. From our extensive experiments, we found a set of moderate values for w_1 , w_2 and w_3 that will ensure desired LDR quality and high compression efficiency, which will be discussed in detail in Section 3.3.

3.3 Experimental Results and Discussion

In this section, we validate the proposed models which estimate the base layer and enhancement layer bit-rates, ensure the LDR visual quality, avoid the temporal consistency between consecutive frames, and conduct fast color correction. Then the compression performance of our solution is demonstrated and analyzed. All the compression is conducted using the bit-depth scalable video encoder provided by the Joint Video Team (JVT) [101].

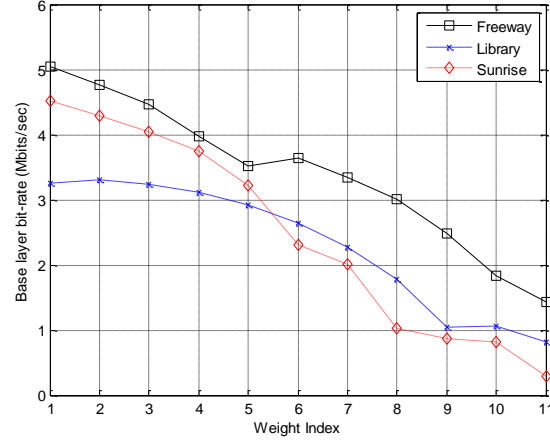
The illustration of the system structure of this encoder can be found in Fig. 1.4. The core of the encoding closely follows the H.264/AVC video compression standard. To reconstruct an HDR frame from a decoded LDR video, an inverse tone-mapping function is stored as a lookup table with each encoded frame. Input HDR videos are either 10 bits or 12 bits per color component for each pixel.

3.3.1 Validation of the Models Q_L and R_b

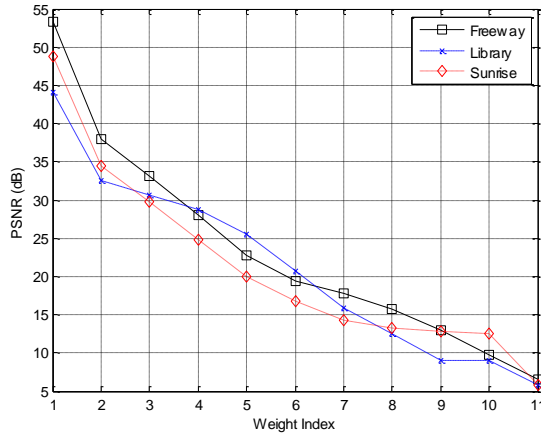
Two models are introduced in the base layer: R_b that reflects the base layer bit-rate and Q_L that controls the visual quality of the tone-mapped content. In order to validate these two models, we set the weighting factor of the enhancement layer bit rate model R_e to zero. In addition, we fix the weight for Q_L to 1 and vary the weight for R_b . In summary, $w_1 = 0$, $w_2 = 1$ and w_3 is set to different values in (3.20). Tone-mapped videos are produced by running the optimization with these settings and each video is then compressed with a set of quantization parameters. The display adaptive TMO is used in this test. Two measurements are obtained from this test: i) the difference, measured in peak signal to noise ratio (PSNR) or structural similarity (SSIM) index, between the tone-mapped videos using the reference TMO and our models with the above settings, and ii) bit rates of compressing the tone-mapped videos in the base layer.

Fig. 3.4 demonstrates the behavior of the three variables (LDR bit rate, PSNR, and SSIM) as the ratio between the weights of Q_L and R_b varies for three video. The quantization parameter (QP) in the compression is set to be 37. The weight ratio is increasing between Q_L and R_b as the "weight index" in the figures changes from 1 to 11. It can be seen from the figure that when the contribution of the base layer bit-rate model R_b increases, the resultant bit-rate from the compression becomes smaller, and, at the same time, the PSNR and the

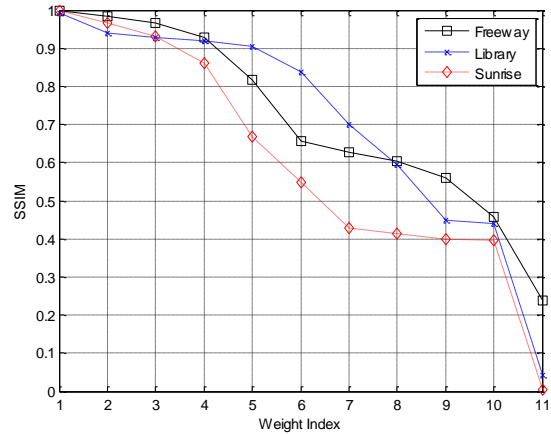
SSIM drops since the influence of the visual quality model Q_L is less. The results justify that visual effect (the difference from content generated using the reference TMO) can be modeled by Q_L and the base layer bit-rate can be well approximated by R_b .



(a) LDR bit rate vs. weight



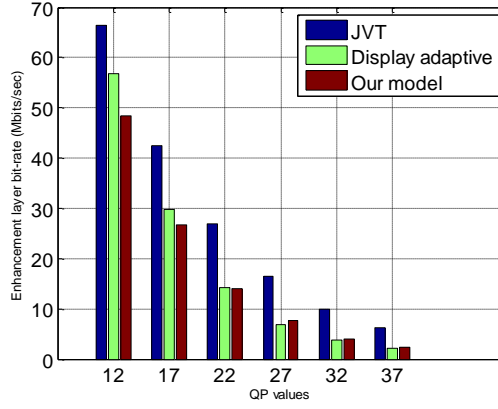
(b) PSNR vs. weight



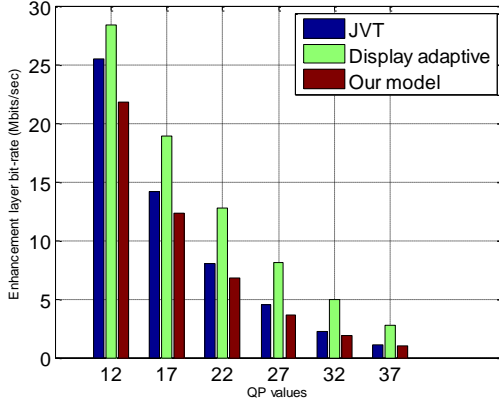
(c) SSIM vs. weight

Fig. 3.4: Demonstration of the change of base layer bit rate, PSNR, and SSIM versus base as the weights of Q_L and R_b vary. The values of weight index 1 - 11 correspond to 100, 50, 25, 12, 6, 3, 1, 0.5, 0.25, 0.125, and 0.0625 for w_3 (w_1 and w_2 are fixed to 0 and 1, respectively), respectively. Three video sequences are used: (i) Freeway (1920 x 1080, 10 bits), (ii) Library (1920 x 1080, 12 bits), and (iii) Sunrise (1920 x 1080, 12 bits).

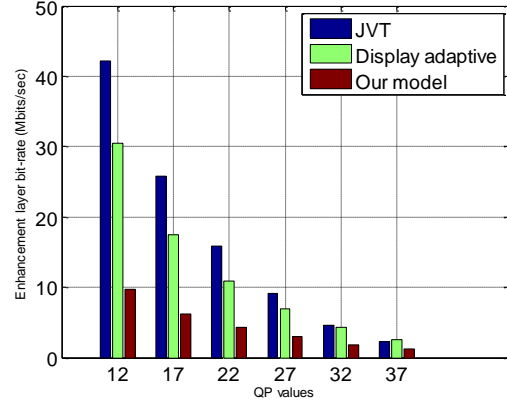
3.3.2 Validation of the Model R_e



(a) Freeway



(b) Library



(c) Sunrise

Fig. 3.5: Comparison of bit rate of the enhancement layer among tone-mapping schemes: tone-mapping used by JVT, display adaptive TMO and our proposed model. The x-axis denotes different QP values used to encode the sequences and y-axis means the bit rate of enhancement layer. The specifications of the three video sequences can be found in the caption of Fig. 3.4.

The enhancement layer bit rate model R_e accounts for the difference between the original and the reconstructed high-bit-depth videos. This difference will be encoded and, in turn, affects the bit-rate of the enhancement layer. To validate R_e , we assign the weighting factor of R_e to be 1 and the other two weighting factors (w_2 and w_3) to be zero in (3.20).

Such setting demonstrates the sole effect of the enhancement layer bit rate. HD quality videos (1920 x 1080) are used in this experiment.

Fig. 3.5 shows the enhancement layer bit rates of different video sequences using three tone-mapping schemes: i) methods that JVT uses, ii) display adaptive TMO, and iii) our model for enhancement layer bit rate R_e . For each of the videos, five QP values are used: 12, 17, 22, 27, 32 and 37. It can be observed that while the other two tone-mapping methods vary in performance for different videos, our model always guarantees the lowest bit-rate of the enhancement layer. For some sequences (e.g., "Sunrise"), the gain is improved significantly using our model. The only exception comes from the video "Freeway", where the display adaptive TMO has comparable performance to the proposed model for QPs greater than 22. The results validate that the bit rate of enhancement layer can be closely estimated out of the proposed model R_e .

3.3.3 Compression Performance

This sub-section demonstrates the compression gains that can be achieved using the proposed solution. We evaluated the picture quality using the peak signal to noise ratio (PSNR) and the structure similarity index (SSIM) against the joint bit-rate of the base and the enhancement layers.

We obtain an interesting observation that the model of the base layer bit-rate (R_b) does not contribute to improving the overall coding performance, which is illustrated in Fig. 3.6. In this test, the weights of R_e and Q_L are fixed to 1 while w_2 (the weight of R_b) varies from 2 to 400. It is seen that for a fixing quality level, the total bit rate remains the same or increases as w_2 increases. This may be because if the bit-rate of the base layer is reduced too much by encoding a tone-mapped frame with less information, then the bit-rate of the enhancement

layer will increase for encoding more residual data. Based on this observation, we decided to set the weight of R_b to zero ($w_2 = 0$) and use only the enhancement layer bit-rate R_e and LDR quality metric Q_L in the optimization process. Assigning the weighting factors of R_e and Q_L to be 5 and 1, respectively, will provide decent compression gain. The demonstration of our results is based on this setting.

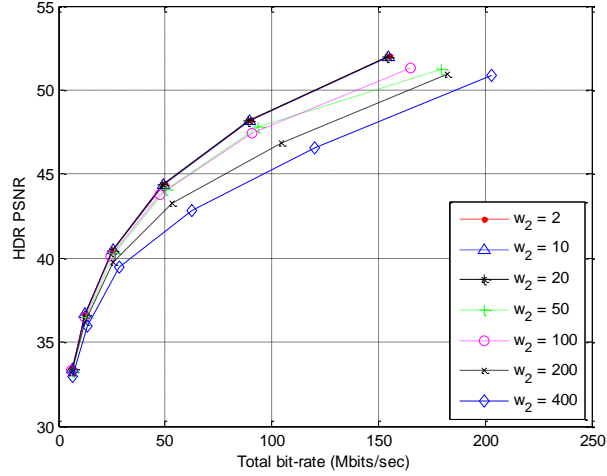


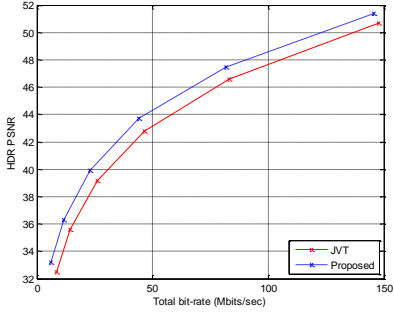
Fig. 3.6: Demonstration that R_b has no or negative contribution to the overall coding performance.

Without loss of generality, we use four tone-mapping methods as our reference TMOs: i) approach used by JVT, ii) display adaptive TMO [37], iii) photographic TMO [38], and iv) adaptive logarithmic TMO [50]. The first tone-mapping approach is used by the video coding standard community, the next two TMOs are well-known and used widely in practice, and the last one was reported to produce the best compression gain [49].

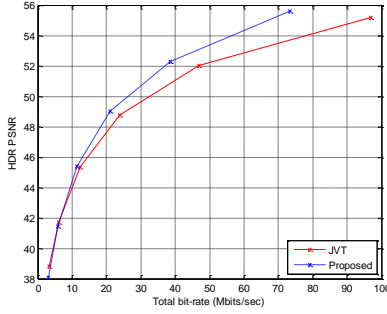
Fig. 3.7, Fig. 3.8, Fig. 3.9 and Fig. 3.10 demonstrate the visual effects and the compression gains achieved by the proposed method using the JVT, photographic, adaptive display, and adaptive logarithmic TMOs as the reference TMO, respectively. In each of the four figures, the first two rows compare the visual effects generated by the original reference

TMO and the proposed approach based on the reference. The last two rows demonstrate the rate-distortion results on different video sequences. The distortion is measured in PSNR as well as SSIM. The largest bit-rate savings can be seen from JVT tone-mapping approach on the "Sunrise" sequence, which is about 45% reduction, or equivalently an increment of PSNR by 3.5 *dB*. There is also over 20% bit-rate reduction on the other two sequences at medium quantization level ($QP = 22$ and 27) using the JVT TMO as reference, and the PSNR enhancement is about 1.2-1.4 *dB*. When the display adaptive TMO is used as the reference, the PSNR increments are about 0.7, 0.8 and 2.8 *dB* for "Freeway", "Library" and "Sunrise", respectively. With medium quantization parameters, 30-40% bit-rate savings (about 2.2 *dB* PSNR enhancement) are observed when using the photographic TMO as the reference while 20-25% bit-rate reduction (approximately 1.2 *dB* PSNR improvement) for the adaptive logarithmic TMO. The only exception falls in the "Freeway" sequence using photography TMO and the "Library" sequence using the adaptive logarithmic TMO. No compression gains are seen for these two cases. This may be because these two tone-mapped versions are already compression-friendly and little coding gain can be achieved by modifying the tone-curve. Although our model is not designed using SSIM, improvement on SSIM can be found in the results. It is seen that our method produces LDR content of good perceptual quality, in addition to providing high compression efficiency.

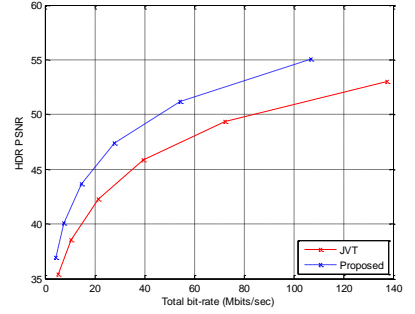
Varying the weighting factors of R_e and Q_L will allow us to adjust the trade-offs between compression efficiency and the similarity to the reference TMO. In general, increasing the weight of R_e will result in better coding gain while a lower weight of R_e will produce a tone-mapped version that is more similar to that generated using the reference TMO.



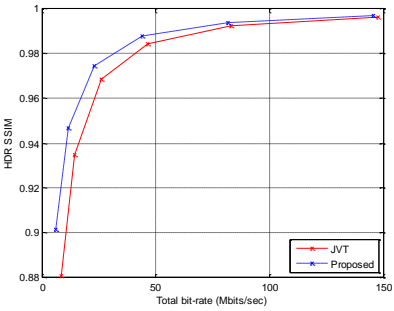
Freeway - PSNR



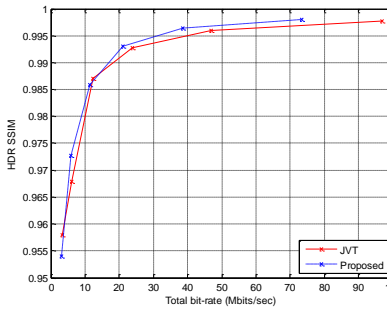
Library - PSNR



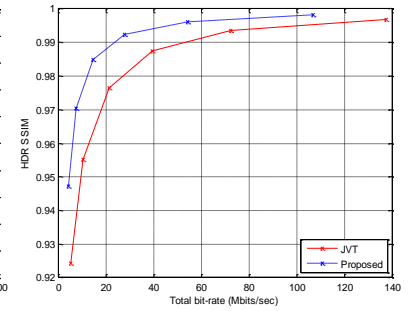
Sunrise - PSNR



Freeway - SSIM

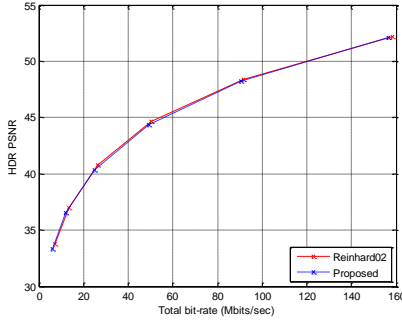


Library - SSIM

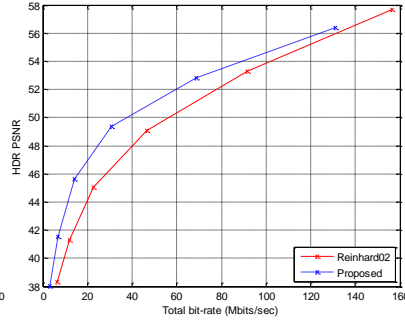


Sunrise - SSIM

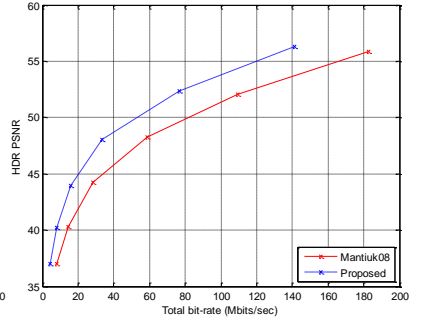
Fig. 3.7: Demonstration of LDR visual effects and compression gains for using the JVT approach as the reference TMO. Row One: The LDR version tone-mapped by the reference TMO. Row Two: The LDR version tone-mapped by the proposed method. Row Three: rate-distortion results in terms of PSNR. Row Four: rate-distortion results in terms of SSIM. For both PSNR and SSIM, higher values indicate better performance.



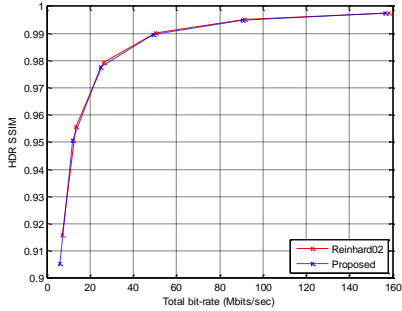
Freeway - PSNR



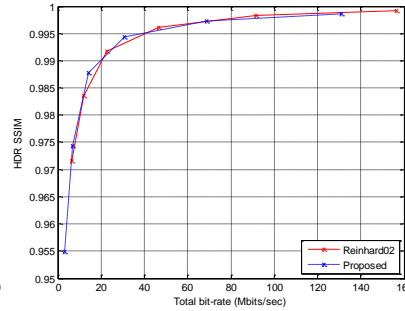
Library - PSNR



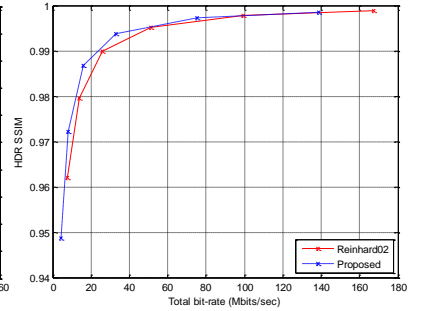
Sunrise - PSNR



Freeway - SSIM



Library - SSIM



Sunrise - SSIM

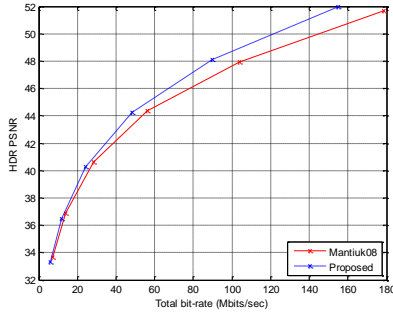
Fig. 3.8: Demonstration of LDR visual effects and compression gains for using the photographic TMO as the reference TMO. Row One: The LDR version tone-mapped by the reference TMO. Row Two: The LDR version tone-mapped by the proposed method. Row Three: rate-distortion results in terms of PSNR. Row Four: rate-distortion results in terms of SSIM. For both PSNR and SSIM, higher values indicate better performance.



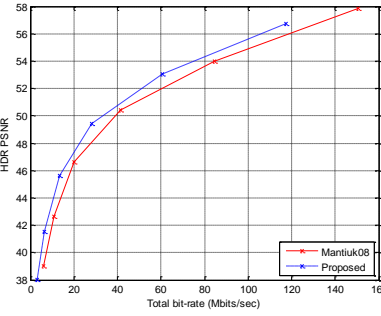
Reference



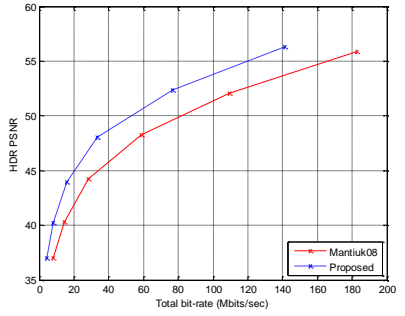
Proposed



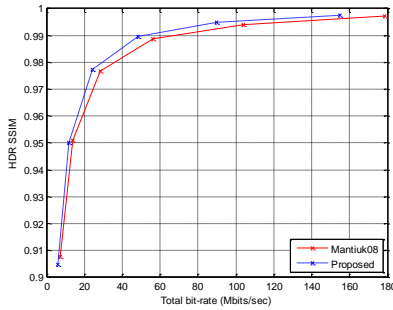
Freeway - PSNR



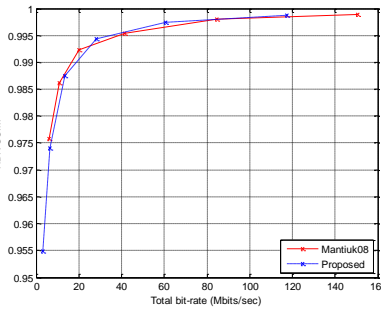
Library - PSNR



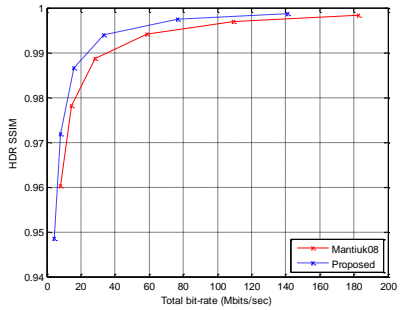
Sunrise - PSNR



Freeway - SSIM



Library - SSIM



Sunrise - SSIM

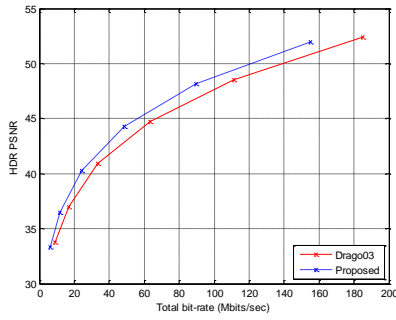
Fig. 3.9: Demonstration of LDR visual effects and compression gains for using the adaptive display TMO as the reference TMO. Row One: The LDR version tone-mapped by the reference TMO. Row Two: The LDR version tone-mapped by the proposed method. Row Three: rate-distortion results in terms of PSNR. Row Four: rate-distortion results in terms of SSIM. For both PSNR and SSIM, higher values indicate better performance.



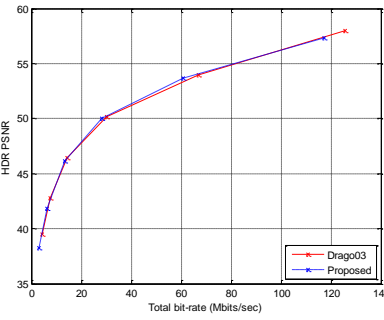
Reference



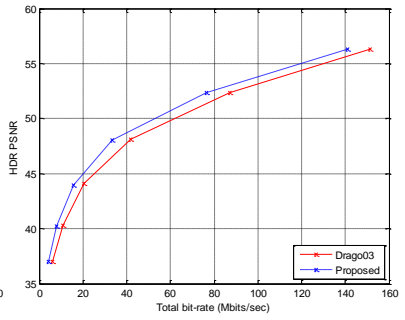
Proposed



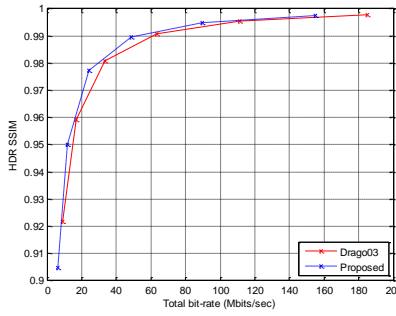
Freeway - PSNR



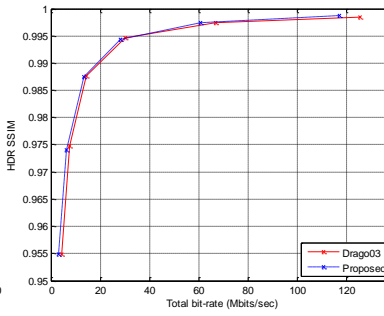
Library - PSNR



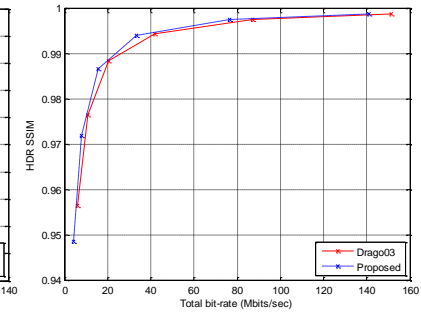
Sunrise - PSNR



Freeway - SSIM



Library - SSIM



Sunrise - SSIM

Fig. 3.10: Demonstration of LDR visual effects and compression gains for using the adaptive logarithmic TMO as the reference TMO. Row One: The LDR version tone-mapped by the reference TMO. Row Two: The LDR version tone-mapped by the proposed method. Row Three: rate-distortion results in terms of PSNR. Row Four: rate-distortion results in terms of SSIM. For both PSNR and SSIM, higher values indicate better performance.

3.4 Conclusions

In this chapter we presented a new tone-mapping approach that provides superior compression efficiency under the framework of bit-depth scalable video coding while preserving a desired perceptual quality for the tone-mapped LDR video content. We developed statistical models that formulate the bit rates of the base and the enhancement layers, as well as the quality of the tone-mapped content, and then incorporated them in our optimization problem. We also considered the temporal effect of our tone-mapper and succeeded in avoiding the flickering artifacts. Another contribution of this work lies in our efforts to tone-map high-bit-depth video directly in a compression-friendly color space (i.e., one luma and two chroma channels) without converting to the RGB domain. The experimental results showed that, in the best case, we can save up to 40% of the total bit-rate (3.5 *dB* PSNR improvement), and, in general, about 20% bit-rate can be achieved. In addition to high coding gains, our method generates tone-mapped videos with good perceptual quality.

4 Color Correction of Tone-Mapping Directly in the YCbCr Domain

The majority of TMOs are derived for only the luminance channel. As depicted in Fig. 1.6, to produce color representation the present common approach is to first derive a tone-mapping method using the luminance channel and then map each of the RGB channels using this tone-mapping scheme [59]. This process is known in this field as color correction.

The above color correction approach can be used in a straightforward manner when applied to HDR content represented in the RGB format. However, it cannot be directly applied on videos formatted in the YCbCr color space, which is the most commonly used format in practice. YCbCr is supported by the major image/video encoding standards, including JPEG 2000 and H.264/AVC. After the luminance component (Y) is tone-mapped, the conventional way to obtain the LDR color components (Cb and Cr) of a YCbCr video is to convert the signal from the YCbCr to the RGB space. Then, the conventional color correction is performed in the RGB domain, and finally the tone-mapped color signal is transformed back to the YCbCr space. Note that the main reason for generating the chrominance components Cb and Cr is that the human visual system is not very sensitive to color, and thus, most implementations end up down-sampling the two components, reducing the overall bandwidth. In that sense, tone-mapping the YCbCr signals requires that they get up-sampled prior to the RGB conversion. This also adds to the computational load.

Not many studies have been done on color correction in the YCbCr color space for tone-mapping. Toda et al. studied the chromatic treatment of the Cb and Cr channels [60], but their work mainly addresses the creation of YCbCr HDR from LDR content. In this chapter, we address the problem of mapping the high-bit-depth Cb and Cr components

directly to their 8-bit counterparts, that is, without having to go through the intermediate processes of color space transformation, up-sampling and then down-sampling.

4.1 Problem Statement

For clarity of presentation, in general, capitalized symbols denote high-bit-depth signals while the lower case ones correspond to their 8-bit counterparts. A symbol with an apostrophe (') denotes the gamma corrected version.

Fig. 4.1 demonstrates the whole process in the conventional approach for tone-mapping a high-bit-depth YCbCr video frame. The color correction should be conducted on the linearized R , G and B signals. The linearized R , G and B signals are the originally acquired signals and are proportional to the light intensity of the scene captured by a camera. Video signals, however, are usually stored using the non-linear (gamma-corrected) signals for compression purpose. Therefore, prior to color correction, a gamma decoding process needs to be applied prior to color correction so that the gamma-corrected R' , G' and B' signals are converted to their linearized counterparts R , G and B . To obtain the R' , G' and B' signals, a transform from the YCbCr to the $R'G'B'$ is required. As mentioned above, the available Cb and Cr are usually sub-sampled from the original chromatic components Cb and Cr. These are denoted by C_{B_sub} and C_{R_sub} while the luminance is not sub-sampled. The input C_{B_sub} and C_{R_sub} signals need to be up-sampled before the color transform can be applied. The color correction process tone-maps the RGB and luminance Y signals and generates 8-bit r , g , and b signals in the linearized form. The Gamma correction process and the color transform are then performed to produce the full-resolution low dynamic range signal in the YCbCr domain. Finally, the 8-bit Cb and Cr components are sub-sampled.

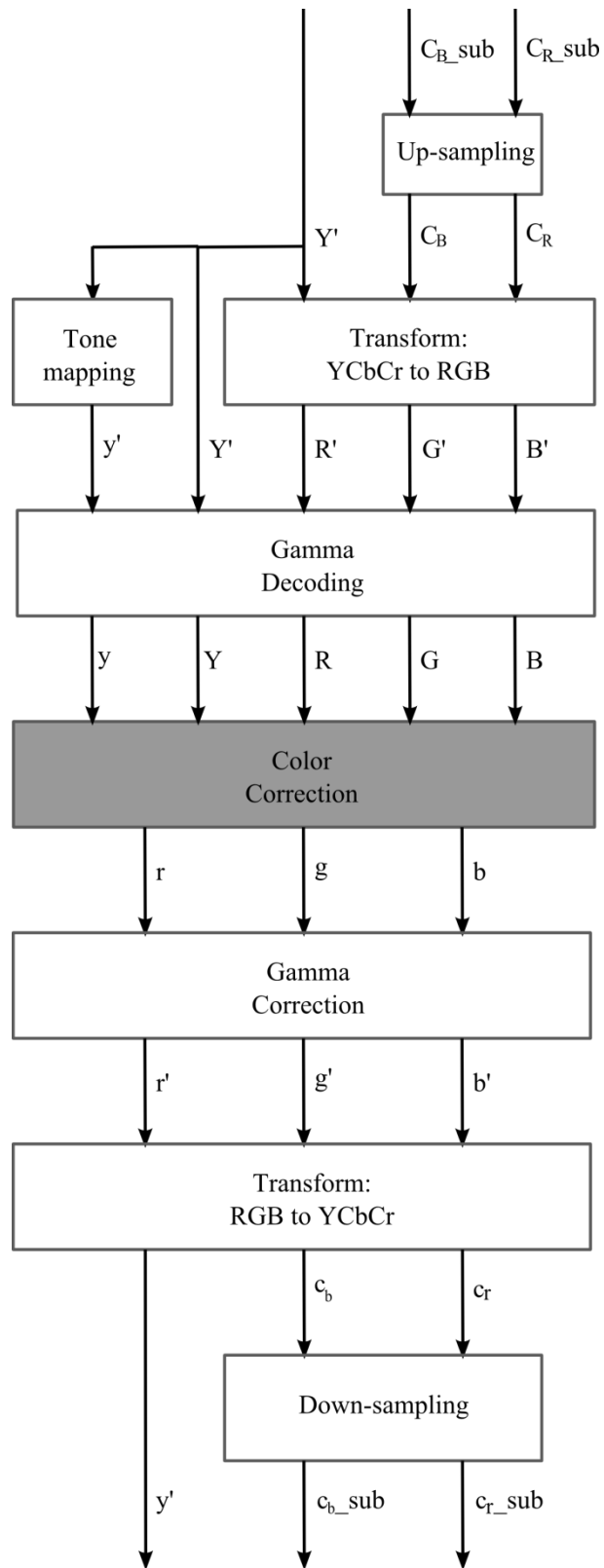


Fig. 4.1: Pipeline of the conventional method for tone-mapping the Cb and the Cr components.

As can be observed, the above process involves many intermediate stages in order to achieve the tone-mapping of the Cb and Cr components. Our proposed approach aims at generating the tone-mapped YCbCr signal with the least possible computational complexity. In our solution, a high-bit-depth YCbCr video sequence is directly tone-mapped to an 8-bit YCbCr stream (see Fig. 4.2), thus all the intermediate processes, including up-sampling, YCbCr to RGB transforming gamma decoding, gamma correction, RGB to YCbCr transforming and down-sampling are bypassed. This simplified process can be generally written as:

$$\begin{aligned} c_b &= f_1(Y', y', C_B) \\ c_r &= f_2(Y', y', C_R) \end{aligned} \quad (4.1)$$

Our approach also benefits from the fact that it avoids all the round-off errors introduced by the intermediate steps of the conventional method, and, in turn, produces a more accurate LDR representation.

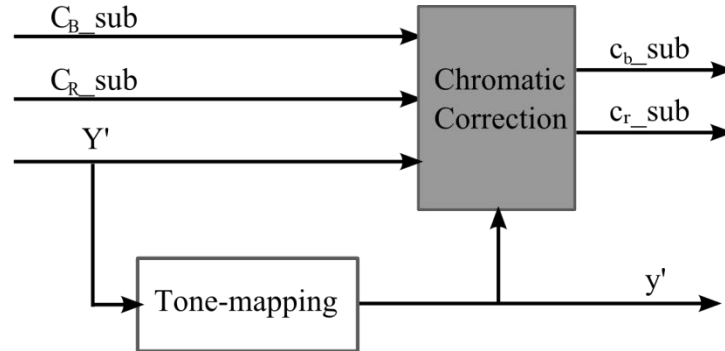


Fig. 4.2: Framework of the proposed solution.

4.2 Proposed Solution

The proposed solution maps the Cb and the Cr components of the high-bit-depth HDR signal directly to its 8-bit LDR counterpart, given the mapping of the luminance channel. Our approach adopts one of the most common color correction methods [57], [58]:

$$b = \left(\frac{B}{Y}\right)^s \cdot y \quad (4.2)$$

where B and Y denote the HDR version of the blue and the luminance channels, respectively, while b and y are the 8-bit LDR counterparts. s is used to adjust the color saturation of the resulting LDR content. Similar relationships can be applied to the red and the green channels.

It is challenging to obtain a nice closed-form solution due to the color transforms and the non-linear processes in the conventional method. If one simply cascades all the functions in the blocks of the full pipeline of Fig. 4.1, a complicated function will result. The computation of this function will be comparable to that of the conventional approach and no advantage is obtained. For example, using (4.3), the standard transform functions between YCbCr and RGB does not provide a nice solution.

$$\begin{bmatrix} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_1 & \varphi_1 \\ \alpha_2 & \beta_2 & \varphi_2 \\ \alpha_3 & \beta_3 & \varphi_3 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (4.3)$$

4.2.1 Derivation of the Proposed Closed Form

Below, we derive an efficient closed-form mapping formula for the Cb component. The derivation can easily be adapted for the Cr component.

Instead of using (4.3), we start with the definition of the Cb component, which reflects the difference between the luminance and the blue channels and can be written as:

$$\begin{aligned} C_B &= M_B(B'-Y') + T \\ c_b &= m_b(b'-y') + t \end{aligned} \quad (4.4)$$

where C_B and c_b denote the Cb component of the high-bit-depth and the 8-bit signals. T and t are the offsets for HDR and LDR signals respectively. Let D denote the maximum intensity value of the high bit-depth signal, e.g., 1023 for a 10-bit signal. The values of T and t are assigned to be $(D+1)/2$ and 128. The constants M_B and m_b are defined as:

$$\begin{aligned} M_B &= \frac{1}{2 \cdot (1 - K_B)} \cdot \frac{Q_B}{D} \\ m_b &= \frac{1}{2 \cdot (1 - k_b)} \cdot \frac{q_b}{255} \end{aligned} \quad (4.5)$$

Note that the values of B , b , Y and y in (4.2) refer to the values in the linearized (not gamma-corrected) color space. However, the input and the output in our tone-mapping pipeline refer to variables that have been or should be gamma-corrected (Fig. 4.1). The following formulae can be used to convert the linearized signals to gamma-corrected signals and vice versa. Let Γ and γ be the values used in the gamma correction for high-bit-depth and 8-bit signals. Equations (4.6) are used for the luminance channel and equations (4.7) can be applied to the color channels:

$$\frac{Y'}{D} = \left(\frac{Y}{D}\right)^{\frac{1}{\Gamma}} \quad , \quad \frac{y'}{255} = \left(\frac{y}{255}\right)^{\frac{1}{\gamma}} \quad (4.6)$$

$$\frac{B'}{D} = \left(\frac{B}{D}\right)^{\frac{1}{\Gamma}} \quad , \quad \frac{b'}{255} = \left(\frac{b}{255}\right)^{\frac{1}{\gamma}} \quad (4.7)$$

The values of D and 255 are used to normalize the signals to 0-1, which is necessary to keep the output range the same as the input range (0-1). The gamma values for LDR (γ) and HDR (Γ) are allowed to be different.

Equations (4.4) can be re-organized as:

$$\begin{aligned} B' &= \frac{C_B - T}{M_B} + Y' \\ b' &= \frac{c_b - t}{m_b} + y' \end{aligned} \quad (4.8)$$

We rewrite equations (4.6) and (4.7) in terms of B and b , and substitute them in (4.2).

After simplification we obtain:

$$b' = \left(\frac{B'}{Y'} \right)^{\frac{\Gamma \cdot s}{\gamma}} \cdot y' \quad (4.9)$$

We substitute the expressions of B' and b' obtained by equations (4.8) into equation (4.9). After re-organizing, the closed-form formula that maps the C_b component from high bit depth to 8 bits can be written as:

$$c_b = m_b \left[\left(\frac{C_B - T}{M_B \cdot Y'} + 1 \right)^{\frac{\Gamma \cdot s}{\gamma}} - 1 \right] \cdot y' + t \quad (4.10)$$

Note that the above equation is not defined when $Y' = 0$. For this reason, we assign 128 to c_b for this special case since $Y' = 0$ means neutral color information can be found in the c_b channel. Thus, the complete mapping solution for the c_b channel is given by:

$$c_b = \begin{cases} m_b \left[\left(\frac{C_B - T}{M_B \cdot Y'} + 1 \right)^{\frac{\Gamma \cdot s}{\gamma}} - 1 \right] \cdot y' + t & \text{for } Y' \neq 0 \\ 128 & \text{for } Y' = 0 \end{cases} \quad (4.11)$$

Similarly, for the c_r component we have:

$$c_r = \begin{cases} m_r \left[\left(\frac{C_R - T}{M_R \cdot Y'} + 1 \right)^{\frac{\Gamma \cdot s}{\gamma}} - 1 \right] \cdot y' + t & \text{for } Y' \neq 0 \\ 128 & \text{for } Y' = 0 \end{cases} \quad (4.12)$$

where C_R and c_r denote the Cr component of the high-bit-depth and the 8-bit signals. The constants M_R and m_r have similar definition as M_B and m_b , and their values may vary for different standards or specifications.

If the original high-bit-depth signal is in the 4:4:4 format (all components in full resolution), the resolution of Y' and y' do not need to be changed when applying equations (9) and (10). If the input is in the 4:2:0 format (the Cb and the Cr components are down-sampled by a factor of 2), we should sub-sample Y' and y' before using the proposed solution to map the chromatic components.

4.2.2 Derivation of Parameters for Different Standards

Below, we derive the parameters M_b , m_b , M_R , m_r , T and t for equations (9) and (10) for the ITU-R BT.601 and ITU-R BT.709 standards. Assuming the input is a 10-bit signal, the value of D is equal to $2^{10}-1 = 1023$.

According to ITU-R BT.601, $K_B = k_b = 0.114$ and $q_b = 224$. Let us use the maximum value for Q_B , which is 1022. Using equations (2) we get $M_B = 0.5638$, and $m_b = 0.4957$. Likewise, we can compute M_R and m_r for the Cr component: $M_R = 0.7126$ and $m_r = 0.6266$. The values of T and t can be easily determined to be $T = (1023 + 1)/2 = 512$ and t is 128.

Regarding ITU-R BT.709, we can perform similar computations and obtain: $M_B = 0.5384$, $m_b = 0.4734$, $M_R = 0.6344$, $m_r = 0.5578$, $T = 512$, and $t = 128$.

For both cases, the parameter s controls the saturation of the output signal. The larger the value of s , the more saturated is the resulting image/video. Its moderate value is 0.8. The gamma value Γ depends on how the high-bit-depth video is encoded; 2.2 is one of the most commonly used values. The gamma value γ is assigned based on how the 8-bit output should be gamma corrected. The two widely-used values for γ are 1.8 and 2.2

4.3 Results and Discussion

In this section, we demonstrate the performance of the proposed approach and compare it with that of the conventional framework shown in Fig. 4.1 using the color correction formula in (4.2). We choose to compare our method against the conventional approach since in the case of tone-mapping there is no original or ground-truth tone-mapped version for other approaches to be evaluated against. In fact, our proposed method may yield better visual results since it avoids round-off errors that occur in the conventional pipeline. The similarity between the resulting video frames generated with the proposed and the conventional approaches is a good indicator of how successfully our method is in yielding results that resemble those obtained by the conventional approach.

Our chromatic correction approach is independent of any existing tone-mapping operator that is used to generate the 8-bit luminance component. Without loss of generality, in the experiment we use a tone-mapping method designed for backward compatible HDR video compression [105]. We set $s = 0.8$ and $\Gamma = \gamma = 2.2$. The constant assigned in (4.11) and (4.12) are based on the ITU-R BT.709 standard. Five video sequences, each 60 frames long, were used in the test.

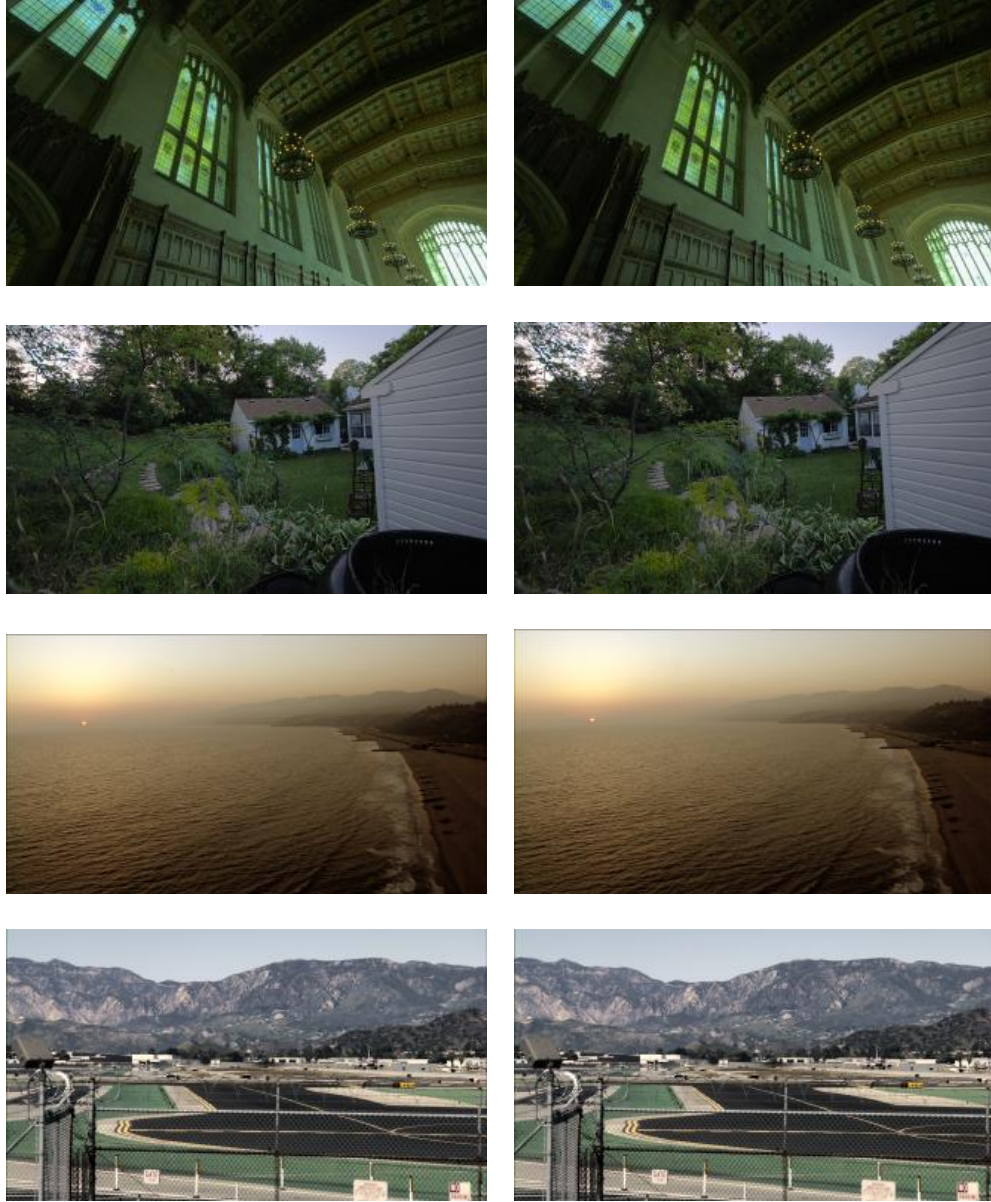
Table 4.1 shows the peak signal to noise ratio (PSNR) and structural similarity (SSIM) index between the proposed method and the full correction pipeline for 300 video frames. For both quality measures, larger values suggest higher correlation between the two mapping approaches. We observe that the resulting PSNR values range from 52.8 dB to 58.4 dB for the Cb component and range from 52.5 dB to 56.7 dB for the Cr component. The average PSNR is 55.1 dB and 54.6 dB for the Cb and Cr, respectively. These PSNR values are considered to be very high. The SSIM quality metric is agreed to be more faithful to the

human visual system and reaches the maximum value of 1 only if the two compared frames are exactly the same. It can be observed from the table that the minimum SSIM value is as high as 0.984 and the average SSIM value is 0.991 for both Cb and Cr, which indicates that the two signals are believed to be perceptually equivalent. The high PSNR and SSIM values indicate that the Cb/Cr components generated using the proposed method and the full conventional pipeline are virtually identical.

Table 4.1 Average performance of the proposed approach

	Min	Max	Average
PSNR for Cb (dB)	52.8	58.4	55.1
PSNR for Cr (dB)	52.5	56.7	54.6
SSIM for Cb	0.985	0.995	0.991
SSIM for Cr	0.984	0.995	0.991

Fig. 4.3 provides a demonstration of the visual similarity between the mapping schemes. The video frames that have been tone-mapped by our efficient method (right) appear to be visually the same to those produced by the conventional method (left). The high similarity stems from the fact that the two schemes are theoretically the same. The difference is caused only by round-off errors introduced in the intermediate steps through the conventional pipeline. Again, the difference does not suggest that our method generates worse results. The fact that our method bypasses round-off errors in its implementation may promise a better tone-mapped signal.



Conventional

Proposed

Fig. 4.3: Tone-mapped images using the conventional approach (left column) and the proposed method (right column) for chromatic correction.

4.4 Conclusion

This chapter addresses the problem of mapping high-bit-depth Cb and Cr components directly to their 8-bit counterparts. We derive a closed-form mapping solution for the two chrominance components. This solution has significantly lower complexity since it avoids

processes such as color space transformation and up-sampling which are required by the conventional method. We choose to compare our method against the conventional approach since in the case of tone-mapping there is no original or ground-truth tone-mapped version for other approaches to be evaluated against. The results show that the Cb and the Cr signals generated by our method the conventional approach are very similar, with average PSNR of about 55 dB and average SSIM at 0.991. The resulting color content also demonstrates the perceptual equivalence between our fast mapping method and the conventional approach. The high correlation between the two methods stems from the fact that our method is derived closely based on the conventional pipeline and manages to by-pass all of its intermediate processes. By avoiding all the round-off errors introduced by the intermediate steps of the conventional method, our approach will actually provide higher fidelity.

5 Rendering High Dynamic Range 3D Images: Subjective Evaluation of Tone-Mapping Methods and Preferred 3D Image Attributes

High dynamic range images provide superior picture quality by allowing a larger range of brightness levels to be captured and reproduced than traditional 8-bit low dynamic range images. Even with existing 8-bit displays, picture quality can be significantly improved if content is first captured in HDR format, and then is tone-mapped to convert it from HDR to the LDR format. 3D display technology also aims at providing viewers with a more realistic visual experience by offering a sense of depth. An increasing number of theatres and households have been equipped with 3D display systems. More and more 3D content is being produced to satisfy the demand of the rapidly growing 3D consumer market as well. Ideally, content could be captured in a 3D HDR representation and viewed on a 3D HDR display, to achieve a more lifelike picture quality. However, existing 3D displays can support only 8-bit LDR content. In order for HDR content to be displayed on existing imaging systems, LDR signals need to be generated for each view of the 3D HDR pair. That is, tone-mapping needs to be applied to each view. The fact that tone-mapped images produce less under- and over-exposed areas will help the fuse of the 3D depth, and the superior picture quality of HDR tone-mapped content will also add value to the 3D representation.

Several subjective evaluations of tone mapping methods have been reported [61], [62], and [63]. However, they are restricted to 2D images and videos. The impact of tone mapping on 3D content has not been studied.

In this work we first conduct subjective tests that evaluate tone-mapping methods on 3D displays. The objective is not to rank existing TMOs, but to understand which attributes of tone-mapping methods will contribute to good 3D representation.

By varying the tone mapping method and the related parameters, LDR images with different visual effect can be achieved. In particular, the tone mapping process can adjust the brightness and the levels of ‘details’ in the output images. Many tone mapping methods have been developed that try to preserve local contrast [38], [48], [54], [102], [103], [104] giving images with higher amounts of detail, i.e., images that look sharper and have stronger texture. The work of Cormack et al. [80] has found that luminance contrast has influence on the depth perception when the stimuli are near the stereo detection threshold. Brightness and sharpness/texture have also been noted by artists to affect the visual comfort and quality of 3D content [86]. Consequently, the optimal tone mapping parameters may be different for 3D images than those for 2D images.

In the second part of this chapter, we present a study on whether the preferred levels of brightness or details are different for tone-mapped 2D and 3D content.

This chapter is organized as follows: An overview of tone-mapping and a short description of the seven TMOs used in our experiments are presented in Section 5.1. The experimental setup, results and analysis of our first test are provided in Section 5.2. The details of the second experiment can be found in Section 5.3. Section 5.4 presents further discussion, conclusions and directions for future work.

5.1 Tone-Mapping Operators Evaluated

Tone-mapping operators are used for reducing the dynamic range in such a way that the resulting image or video may be displayed on LDR devices. TMOs can be categorized

into two broad classes: global and local operators. Global TMOs use a single, usually nonlinear, mapping curve for every pixel in the image. In local operators, the mapping of each pixel is a spatially variant function of its neighboring pixels. In general, global TMOs are more computationally efficient and better retain the sensation of the brightness of the original HDR signal. On the other hand, local operators are able to preserve more details and provide higher local contrast.

Since dozens of tone-mapping operators are available today, it is impractical to evaluate all of them. Therefore, in our experiments, seven representative TMOs, four global and three local, are selected that cover most design principles of tone-mapping. Below are short descriptions of these seven tone-mapping operators. The italic text appearing in parenthesis immediately after each name is used to represent the corresponding TMO in the rest of this paper.

Photographic tone-mapping (*Photo*) [38]: It simulates the dodging and burning techniques which are employed in the darkroom printing process to adjust exposures of selected areas. This scheme supports both global and local versions. The local version computes a local brightness adaptation level for each pixel by constructing a Gaussian pyramid and identifying a largest possible neighborhood that contains relatively low contrast. Then such adaption level is used together with a sigmoidal function to compress the dynamic range. In this paper we tested only the local operator since the behaviors of the two versions are similar. Most previous studies also considered only the local version [61], [62], [63].

Adaptive logarithmic tone-mapping (*Log*) [50]: This global TMO takes into account the human vision system and is formulated as a logarithmic compressive function. The base of the logarithm is varied according to the input intensity.

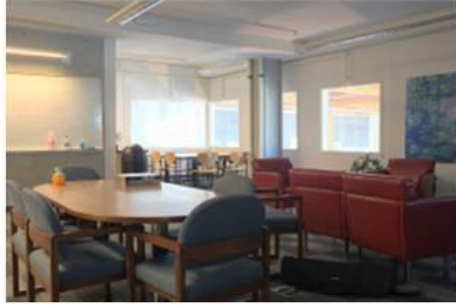
Fast bilateral filtering (*Bi*) [102]: This is a local TMO. It first generates a base layer using a bilateral filter, which preserves the edges and smoothes the rest of the HDR image. Then, a detail layer is produced by dividing the HDR image by the base layer. The LDR image is derived by combining the detail layer and a compressed dynamic range version of the base layer.

Tone-mapping for backward-compatible HDR compression (*Comp*) [105]: It is a global tone-mapping method that was inspired by the observation that most tone-mapped content, especially videos, will undergo lossy image/video encoding such as H.264/AVC compression. This TMO thus aims at minimizing the join information loss due to the tone-mapping and the image/video encoding processes.

Gradient domain tone-mapping (*Grad*) [48]: It achieves dynamic range reduction in the gradient domain. This tone-mapping method attenuates large gradients by manipulating gradients of the image in different spatial scales. The LDR image is generated by solving a Poisson equation in an adjusted gradient field.

Display adaptive tone-mapping (*Disp*) [37]: This tone-mapping operator is formulated using a global piece-wise function which is uniquely determined by minimizing the perceptual contrast distortion between the original HDR signal and the resulting tone-mapped content on various displays, varying from e-paper to high contrast output devices.

Tone-mapping by photoreceptor physiology (*Phy*) [41]: Based on the findings in physiology that vision adaption occurs in the photoreceptor, this global TMO achieves tone reproduction by modeling the behavior of photoreceptors. An advantage of this operator is that it can be easily controlled by three intuitive user parameters: brightness, contrast and chromatic adaptation.



Photo



Log



Bi



Comp



Grad



Disp



Phy



Single Exposure

Fig. 5.1: LDR images generated from the same HDR scene by tone-mapping operators evaluated in this paper. The image captured using one single LDR exposure is also included in this figure.

In order to give the reader an idea of the visual properties of each TMO, in Fig. 5.1 we show the LDR images generated by each of the operators tested in this paper. The same scene, captured with a single LDR exposure, is also shown. Note that the tone-mapped versions of the image avoid the clipping that is present in the LDR capture. It can also be seen that the local TMOs (particularly *Bi* and *Grad*) produce images with more details/texture than the global TMOs (*Log*, *Comp*, *Disp*, and *Phy*).

5.2 Experiment One – Evaluating Tone-Mapping Operators

In the first subjective test, our goal is to evaluate the preference of the 3D LDR images tone-mapped by different TMOs in terms of two aspects: i) the 3D effect (depth impression) and ii) the 3D overall quality.

5.2.1 Image Preparation

Eight 3D scenes, four indoors and four outdoors, were captured with multiple exposures, and then stereoscopic HDR images were generated by blending these exposures [5]. In particular, we used two identical cameras and placed them side-by-side, in parallel. To create a pleasing 3D effect, the baseline (distance between the two cameras) was depending on the distance between the cameras and the scene. For each scene, the settings on both cameras are the same: both were set in “Manual” mode with ISO equal to 100; to alter the exposure, the shutter speed stayed the same while the aperture value changed.

Fig. 5.2 shows an LDR version of the eight scenes. Eight different LDR images were derived from each scene: seven of them were tone-mapped using the selected TMOs described in the previous session while one of them is a well-exposed image captured in LDR. Please note that several of the selected tone-mapping operators require their user

parameters to be specified. Therefore, prior to the main subjective test, we conducted a pilot study, where several subjects (they did not participate in the main test) were asked to choose the best possible parameters for each of the TMOs for each scene. For the same scene and the same TMO, average values were used to generate tone-mapped images.

The complete database of the original HDR image pairs and their LDR versions can be found in [106]. This supplementary material also contains statistics (min, max, mean of luminance/luma) and the histograms of all the images.

5.2.2 Experimental Framework

Sixteen subjects participated in our main experiment. They had diverse cultural backgrounds and differed in age (24 – 38 years old). All of them had normal or corrected vision and little experience with 3D and HDR content. The display device used in our test was a *Samsung UN55C7000* 3D TV, 55 inch, 1920 x 1080p at 240Hz, which uses active shutter glasses. The viewing condition is designed according to ITU-R Recommendation BT.500-11 [107].

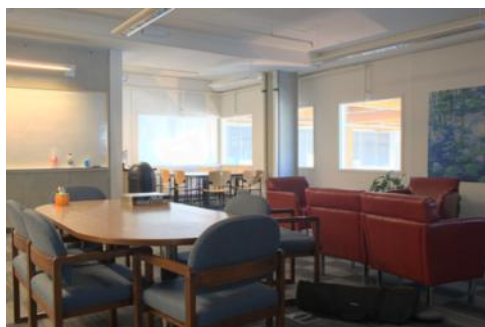
In this psychophysical experiment, subjects needed to evaluate stereoscopic LDR images on a 3D display. Eight scenes, four indoors and four outdoors, were captured in 3D with multiple exposures, and then stereoscopic HDR images were generated by blending these exposures [5]. Fig. 5.2 shows an LDR version of the eight scenes. Eight different LDR images were derived from each scene: seven of them are tone-mapped using the selected TMOs described in the previous session while one of them is a well-exposed image captured in LDR. The display order of the different tone mapped versions of each scene was randomized. In total, 64 images were tested. In order to help the subjects better adapt their



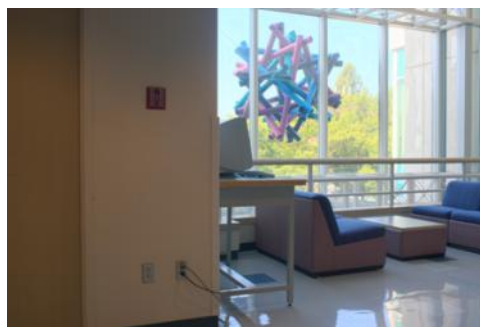
Bulletin



LibraryTree



MeetingTable



ICICS



ChemEng



Sauder



LabWindow



Stairs

Fig. 5.2: Scenes used in the subjective test. Since HDR content cannot be shown on the paper or most monitors, all the images displayed above are tone-mapped using the photographic TMO.

visions to a new image, we inserted a plain grey slide between tested images to cushion the transition.

For each of the 64 3D images in the test, subjects were asked to evaluate two aspects: i) the 3D effect and ii) the overall quality. The 3D effect score rates the impression of depth and how comfortable the 3D scene is perceived. On the other hand, the overall quality is a measure of how much the user ‘likes’ each image. It reflects a combination of image attributes including contrast, naturalness, sharpness, detail reproduction as well as 3D effect.

At the beginning of our test sessions, the purpose and the evaluation procedure was explained to the subjects. Subjects were verbally instructed to “rate the 3D effect based on depth impression”. They were also asked to rate the overall 3D quality based on how much they liked the entire image, including 3D depth perception, naturalness, contrast, sharpness and detail reproduction. This introduction was followed by a training session where material similar to the images in the actual test was presented. The training material was the eight tone-mapped versions of one extra scene that was not used in our main tests. The objective of this training is to allow the subjects to adapt their eyes to the test environment and become familiar with their task. The actual test started right after the training session.

A scoring bar ranging from 0 – 10 is associated with each aspect to be evaluated; with a higher score meaning better quality. It is a continuous grading scheme, which means rating a non-integer number such as 4.5, is permitted. For each scene, subjects are asked to view the eight LDR images consecutively at first. Subjects would press a key to switch to the next image, and therefore could control the amount of time they spent viewing each image. After viewing the eight LDR versions of a scene, the subjects were allowed to go back and perform paired comparisons on images that they had rated with low confidence. The combination of

the two evaluation techniques, rating scale and paired comparison, will keep the experiment time-efficient and, at the same time, guarantee high accuracy of the collected data. The pace of the entire test could be fully controlled by the subject; it took about 30 minutes on average to finish one user study.

5.2.3 Results and Analysis

As described previously, our experiment consists of 64 images – 8 (TMOs) x 8 (scenes). Two aspects are evaluated for each image: 3D effect and overall quality. Below we conduct a statistical study on the data collected. To account for the fact that different people would have different standards regarding the quality scaling, prior to the main process and analysis and for each aspect evaluated, we normalize the scores as $(S_{i,j} - \mu_i) / \sigma_i$, where $S_{i,j}$ is a score from the i th subject and the j th image. The variables μ_i and σ_i denote the average value and the standard deviation over all the images for the i th subject, respectively.

First of all, we performed outlier detection based on the ITU-R Recommendation BT.500-11 [107] and found one outlier in our case, so only the data of the remaining subjects were used.

The results for the 3D effect and the overall 3D quality (with 95% confidence intervals) are shown in Fig. 5.3 and Fig. 5.4, respectively. In both figures, the results for each of the eight scenes are provided on the left while the average performance is given on the right. The x-axis denotes the tone-mapping schemes, and the y-axis shows the normalized quality standings.

All TMOs had significantly higher scores for 3D effect compared to the single exposure LDR (Fig. 5.3). The LDR images often have clipped regions (over- or under-exposed regions), which interfere with the 3D effect. The lack of information in these

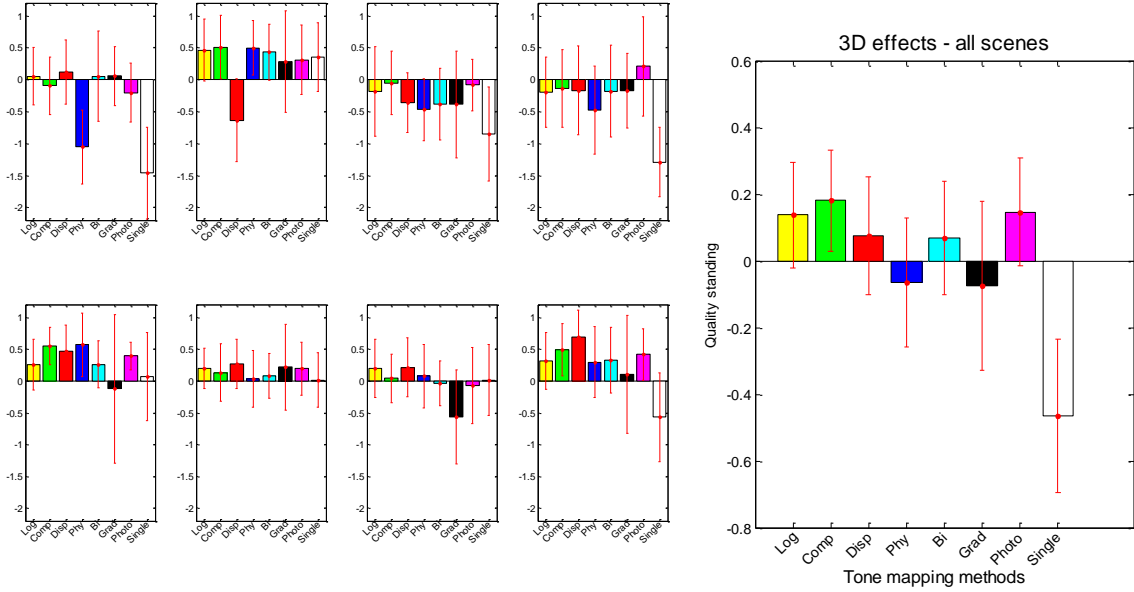


Fig. 5.3: Results (with 95% confidence interval) for 3D effect. The eight smaller figures on the left corresponds to ratings for individual scenes. The top row (from left to right): “Bulletin”, “ChemEngEntrance”, “ICICSBuilding” and “LabWindow”; The bottom row (from left to right): “Library”, “SauderSchool”, “MeetingTable” and “Stairs”. The larger figure on the right shows the average results over the eight scenes. In each of the plot, the x axis denotes the quality standing: the higher the better. The y axis represents the tone-mapping schemes: (1) *Log*, (2) *Comp*, (3) *Disp*, (4) *Phy*, (5) *Bi*, (6) *Grad*, (7) *Photo*, and (8) *Captured with normal LDR cameras with a single exposure*.

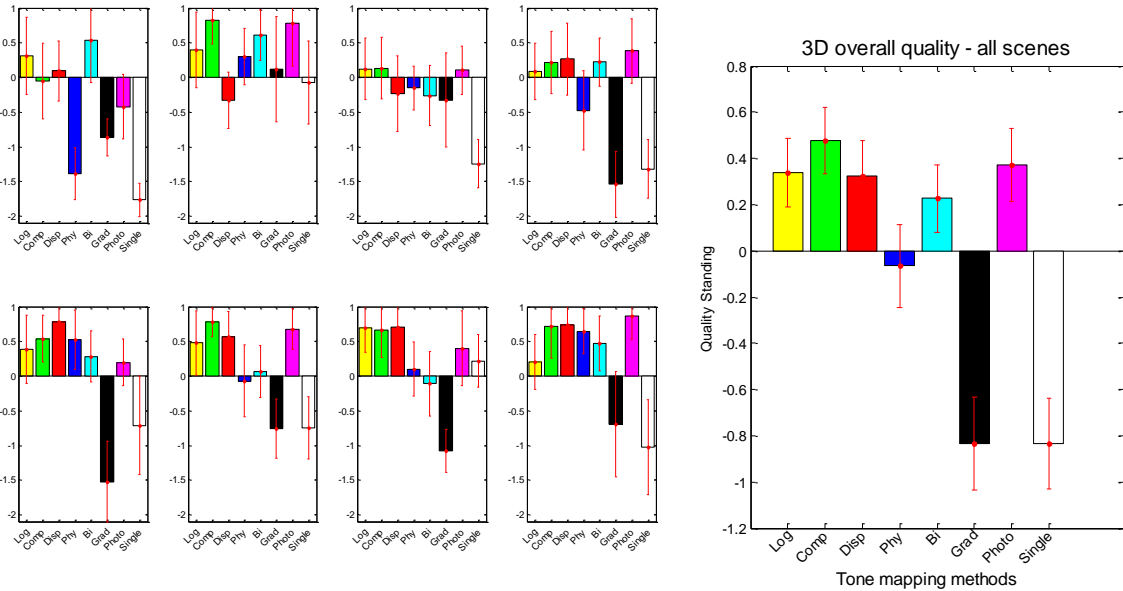


Fig. 5.4: Results (with 95% confidence interval) for overall 3D quality. Notations are the same as Fig. 5.3.

regions will make the 3D perception uncomfortable, since the viewer’s brain has little evidence on how to fuse the images. This impairs the perceived 3D effect, and also lowers the overall 3D quality. This drawback of LDR capturing shows that there is significant benefit to capturing in HDR format and then performing tone-mapping, even though an LDR display is used to view the images in the end.

Relatively poor results can also be observed for the global operator *Phy* (blue) and the local operator *Grad* (black). The poor performance of *Phy* is probably because it produces noticeable under- or over-exposed areas for scenes containing ultra high dynamic range, such as “*Bulletin*” and “*LabWindow*”.



Fig. 5.5: Demonstration of unnaturalness by the gradient TMO.

The average results of the 3D effect and the overall 3D quality have a similar trend except for the *Grad* TMO. This TMO produces images with such high amount of detail that the images tone-mapped by it may be perceptually unreal (see Fig. 5.5). This explains why it has a much lower rating in terms of overall quality, which takes into account the naturalness of an image. The other seven schemes generate more natural looking LDR content, and their performance seems to indicate that the overall perceived 3D quality has high correlation with the 3D effect. However, as the gradient tone-mapping results show, poor 2D image attributes can lead to poor overall quality even when there is a good 3D effect. This suggests that a

good 3D effect is a necessary but not sufficient condition for a good overall 3D quality. The reason for this may be that subjects cannot isolate well the 3D effect from other image attributes unless these attributes exceed a certain tolerance level.

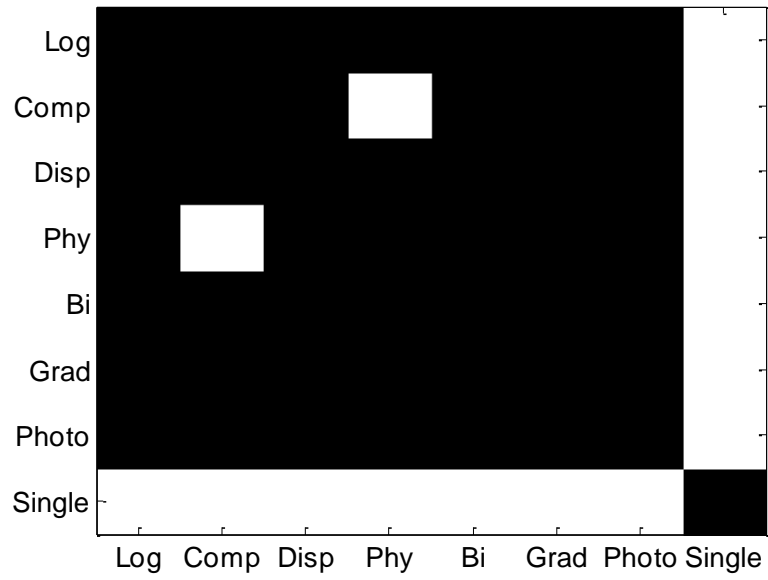
Another observation from Fig. 5.3 and Fig. 5.4 is that several of the global TMOs (*Log*, *Comp* and *Disp*) and local TMOs (*Bi* and *Photo*) have similar (good) performance in terms of both 3D effect and overall 3D quality. These TMOs were rated highly compared to *Grad*, *Phy* and single exposure capturing, but have similar performance to each other. These results are somewhat similar to the conclusions of [65]; that the top tone-mapping operators have similar subjective quality.

It is worth noting that the performance of some TMO's varies greatly from image to image. For example, the TMO *Phy* has the lowest scores for 3D effect on the “*Bulletin*”, yet it has the highest 3D effect scores for the “*Library*”. Looking at the image histograms (available online [106]), we can see that the “*Bulletin*” scene has pixels with low HDR values that are clipped in the *Phy* tone-mapped image. The “*Library*” scene has less dynamic range, so few pixels get clipped in the *Phy* tone-mapped image and for that reason it achieves a good 3D effect. Although the single exposure yields very poor results overall, for a few scenes (“*ChemEng*”, “*MeetingTable*”) it has comparable performance to the best operators. We also notice that the display adaptive TMO (*Disp*) has generally good 3D effect scores except for the “*ChemEng*” scene. Its tone-mapped image has higher contrast than the images produced by other methods, and its image histogram shows that most of its pixels have lower values. On the other hand, *Comp* and *Log* are most consistent in giving good results for overall 3D quality across all images whereas the other top performers (*Disp*, *Bi*, *Photo*) all have relatively poor scores for at least one image. Therefore, *Comp* and *Log* might be

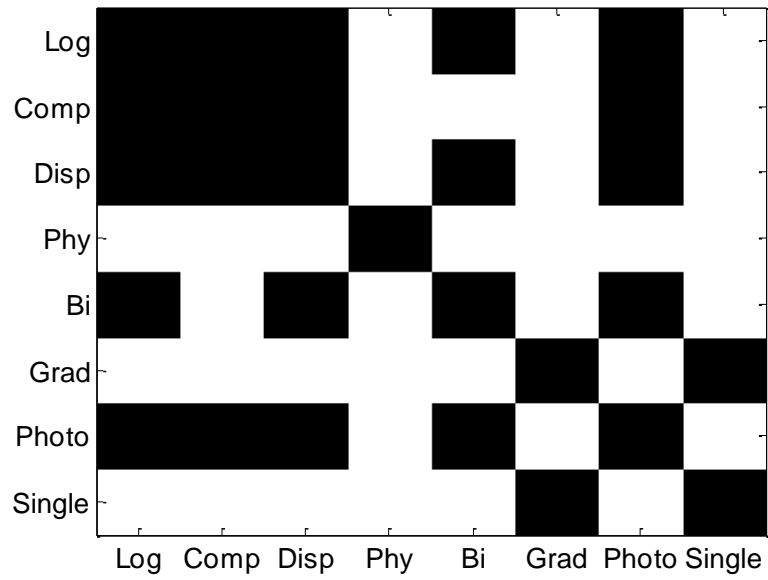
preferred in applications that require reliably providing good LDR quality for a wide range of different inputs.

The variance of the collected data is relatively high. This may be due to the nature of the study, where there is no reference provided for subjects to compare the tested images against. The scores, particularly for “overall 3D quality”, are a matter of personal taste much more so than other studies which try to measure the distortion of an altered version of an image compared to an original. Another factor is that subjects have less prior knowledge (reference from experience) about “good 3D effect”, because people generally have much less experience viewing 3D images compared to 2D ones. Therefore, they may have more difficulty quantifying the 3D effect.

To further analyze the difference between the different TMOs, we perform the statistical significance test. Its objective is to determine if there are significant differences between any two of the tone-mapping methods for both 3D effects and overall 3D quality. Hogg’s one-way analysis of variance (ANOVA) [108] is applied for every pair of the eight methods for each aspect. We state the null hypothesis as: there is no significant distinction between the collected values for the two tested tone-mapping methods. The smaller the returned p value, the more significant is the difference. A typical threshold of the p value that rejects the null hypothesis is 0.05. Fig. 5.6 visualizes results of this pair-wise comparison. A black box corresponds to a p value that is greater than 0.05. That is, black means there is no statistically significant difference between the TMO on the x-axis and that on the y-axis. A white box indicates there is a statistically significant difference between the two methods. Fig. 5.6 (a) confirms that for 3D effects there is no statistically significant difference between global and local TMOs except for *Phy* which performs clipping for images with very high



(a)



(b)

Fig. 5.6: Pair-wise ANOVA comparison of the seven TMOs and the LDR capturing, for (a) 3D Effect and (b) overall quality. Black means there is no statistically significant difference between the TMO on the x axis and that on the y axis. White stands for the opposite.

dynamic range. The LDR capturing approach (*single exposure*) has significant difference from all the TMO for 3D effects.

Previous work has shown that contrast can influence perceived depth [80], so one might expect that different TMO's yield different 3D effect scores. However, this was not what we witnessed from our tests. One possible explanation for the similar performance of several TMO's is that the luminance contrast produced by them is close to the supra-threshold level, where any further contrast increase does not have a strong influence on depth perception [80].

Fig. 5.6(b) shows the significance results for the overall 3D quality where attributes such as naturalness and color saturation are taken into account. *Phy*, *Grad* and *Single exposure* are generally found to have statistically significant difference from the other TMOs, for the reasons discussed previously in this section. The results for the top performer, *Comp*, are not significantly different from those of *Log*, *Disp* and *Photo*, suggesting that several of the top TMOs give very similar performance for 3D images.

5.3 Experiment Two – 3D Effects on Brightness and Details

The second part of our study focuses on how the preferred level of brightness and the preferred amount of details differ between tone-mapped 3D and 2D images.

5.3.1 Image Preparation

The same eight scenes from the first experiment were used (Fig. 5.2). In order to investigate the effect of brightness and the amount of details, tone-mapped 3D images at different levels of brightness and with different amount of details were generated from the eight HDR images respectively.

To vary the brightness levels in the LDR version with consistent effect, we chose to apply the popular *Photo* TMO since it provides a user parameter (key value α) for changing brightness. This tone-mapping operator simulates the dodging and burning techniques in darkroom printing. For each scene (i.e., each HDR image), we altered the value of α such that i) 41 LDR images with different brightness levels were produced, from very dark to very bright, and ii) the difference in the overall image brightness between two consecutive levels is as constant as possible. Fig. 5.7(a) illustrates tone-mapped images at different brightness levels generated using the above approach. In order to quantify the brightness levels for the different tone mapped images, we use mean image brightness B_{img} , which is defined as:

$$B_{img} = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n I(i, j) \quad (5.1)$$

where $I(i, j)$ denotes the pixel value at the location of (i, j) in an image I , and m and n are the dimensions of the image I . Technically speaking, this quantity should be referred as the “mean image luma” because it is calculated based on gamma-corrected pixel values. However, here we chose the term “brightness” since it reflects the brightness of perceived image through the display and in that respect it is a more intuitive term than “luma”. Fig. 5.8(a) shows how the mean image brightness changes as the brightness level increases for the image “*LibraryTree*”. It can be seen that the curve looks like a gamma-correction function ($x^{1/\gamma}$, where γ usually ranges from 1.8 to 2.6). When the images are shown on a display, the response curve of the display (x^γ) will make the final display luminance close to a straight line. That is, the difference in the perceived luminance is relatively constant between two consecutive brightness levels.

To produce different amounts of details in the LDR version, we choose the popular *Bi* TMO [102]. As a reminder, this TMO filters an HDR image into a base (low-pass) layer as

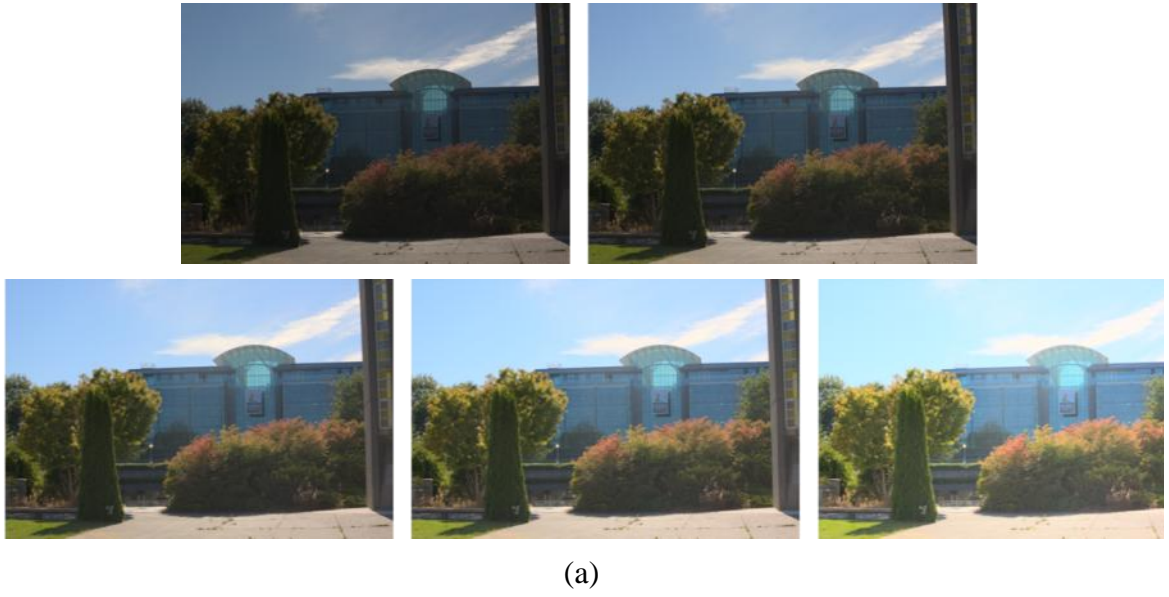


Fig. 5.7: Demonstration of images at different brightness and detail levels: (a) - the scene “Library” at different brightness levels; (b) the scene “ICICS” at different detail levels.

well as a detail layer and then combines the modified versions of them to yield its tone-mapped image. The original *Bi TMO* uses a fixed ratio between the base and the detail layers when combining them. In our modification, this ratio λ may be changed for adjusting the contribution of the detail layer against the base layer as can be seen in (5.2).

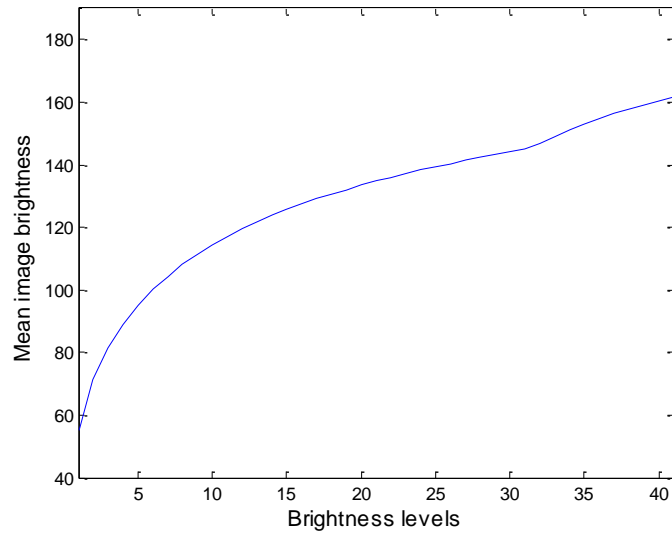
$$I_{LDR} = L_b + \lambda \cdot L_d \quad (5.2)$$

where I_{LDR} denotes the resulting tone-mapped image. L_b and L_d represent the base layer and the detail layer, respectively. The effect of changing this ratio λ on the amount of details is demonstrated in Fig. 5.7(b). Similar to brightness, to quantify the detail levels, we use mean image gradient D_{img} :

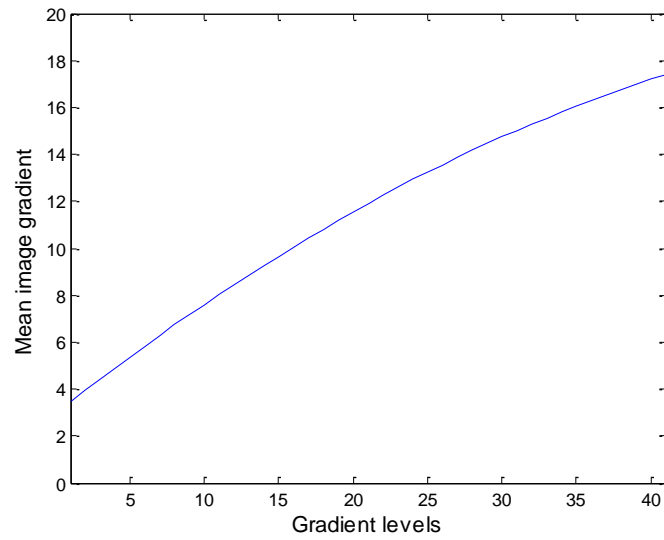
$$\begin{aligned} D_{img} = & \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n |I(i, j) - I(i+1, j)| \\ & + \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n |I(i, j) - I(i, j+1)| \end{aligned} \quad (5.3)$$

where the notations are defined the same as (5.1). It is seen from Fig. 5.8(b) that the mean image gradient varies approximately linearly as the gradient level rises.

For each of the scenes, four sets of tone mapped versions of the screen were created: i) 3D representation with various brightness levels, ii) 3D representation at various detail levels, iii) 2D representation at various brightness levels, and iv) 2D representation at various detail levels. For each version, we prepared 41 different levels. In total, there are 8 (scenes) x 4 (versions) x 41 (levels) = 1312 images in our system that have to be used in the test.



(a)



(b)

Fig. 5.8: The effect of changing the brightness and the detail levels. (a): Changes of the image average brightness as the brightness level varies; (b): of the image average gradient as the detail level varies.

5.3.2 Testing Environment and Procedures

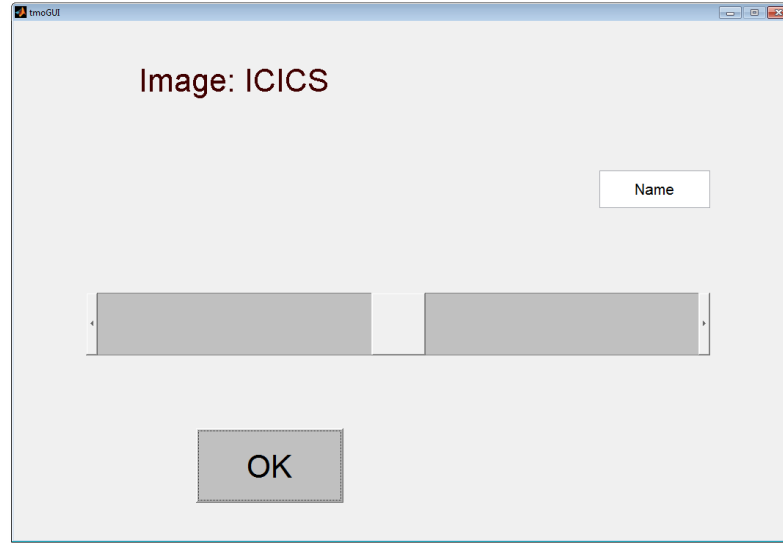


Fig. 5.9: Illustration of the graphic user interface for subjects to fully navigate the psychophysical experiment on their own pace and to select the images at the preferred brightness or detail levels.

Eighteen subjects participated in our subjective experiment. All of them had normal or corrected vision with no/marginal experience in 3D technology. The viewing conditions of our tests were set based on the ITU-R Recommendation BT.500-11 [107]. Viewers keep having their 3D active shutter glasses on when watching 2D videos. This guarantees that the brightness reduction is the same when watching 3D and 2D videos.

A secondary display, a 19" 2D screen, was placed closed to the viewer, where a graphic user-interface was created for subjects to select their preferred image (level) for each of the versions. Fig. 5.9 demonstrates the user interface. A slider bar can be found near the center of the interface. The slider consists of 41 values and each value corresponds to each brightness/detail level. The position of the slider is reset to the middle of the bar when a new scene is loaded. Once a particular position/value is chosen, an image at the selected level (brightness or amount of detail) is updated and shown immediately. As the slider is moved

from left to right, an image will become brighter in the case of evaluating brightness versions and will have more details in the case of evaluating detail versions. An “OK” button sits below the slider and is used by subjects to confirm their selection. The selected value on the slider will then be recorded.

Before the test started, participants were provided with a training section which ensures they were comfortable in using the scoring interface and had their eyes adapted to the viewing conditions. In the test, they were asked to move the slider for selecting the preferred images for each version of each scene. In the case of 3D, they were instructed to “select the preferred brightness or detail level that provided the best 3D depth impression”, and in the case of 2D they were instructed to “select the preferred brightness or detail level that provided the best overall image quality”.

The procedure for each subject would proceed as follows. First, a 3D version of one scene would be shown to the subject, and he/she would adjust the slider on the GUI to select the level of brightness they thought gave the best 3D quality. Then the subject would press the OK button on the GUI, after which 5 seconds of grey is shown to allow the subject to rest his/her eyes. Then, the next scene would be displayed and the subject would repeat the process. After the user has done this for all 8 scenes, the slider is changed to control the amount of detail in the images, while still displaying 3D images. Again the user would adjust the slider to select what is thought to be the best amount of details for each scene. After the subject has selected the preferred amount of details for all eight scenes, the entire procedure is repeated in 2D viewing scenario. That is, the user first selects what he/she considers the optimal amount of brightness for each scene, and then selects the optimal amount of details for each scene viewed in 2D.

5.3.3 Results and Analysis

After recording the results from the psychophysical tests, we performed statistical analysis on the collected data. First, we tested for outliers based on ITU-R Recommendation BT.500-11 [107] and in our case no outlier was identified, so the data for all subjects were used.

The plot in Fig. 5.10(a) shows the mean image brightness (B_{img} , defined in (5.1)) of the preferred 3D and the preferred 2D representations for each of the eight scenes (x-axis). The height of each bar is obtained by first calculating the mean image brightness of the preferred level from each of the 18 subjects and then averaging these 18 values. A general observation is that the mean brightness of the preferred 3D images is slightly higher than that of the preferred 2D counterparts for all the scenes except scene six (“*Sauder*”). All the mean brightness values fall in the interval between 115 and 135.

In order to gain a better understanding of the differences between 3D and 2D viewing, for each subject we calculated the difference between the average brightness of their preferred images in 3D and 2D modes as:

$$\Delta_B = B_{pref,3D} - B_{pref,2D} \quad (5.4)$$

where $B_{pref,3D}$ is the average brightness of the image the user selected with the slider in 3D mode (their ‘preferred’ image), and $B_{pref,2D}$ is the corresponding average brightness of the preferred image they selected in 2D mode. A positive value for Δ_B indicates the subject prefers a brighter image when viewing in 3D compared to viewing in 2D. Fig. 5.10(b) shows the difference averaged over the eighteen participants, and also the 95% confidence interval for the differences. It is seen that although the average values of seven out of the eight scenes are above zero, all of the 95% confidence bars cross the zero axis. Therefore, there is

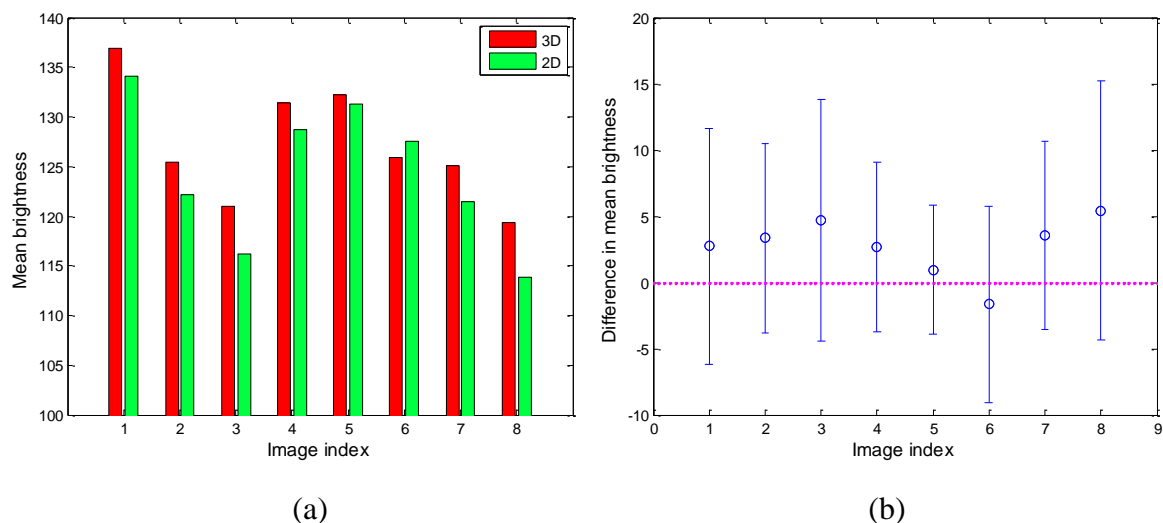


Fig. 5.10: Comparison between 2D and 3D images in terms of the preferred brightness level: (a) – mean image average, (b) subtraction of the mean image average of 2D from that of 3D images. The x-axis denotes the image index, and 1 – 8 correspond to 'ICICS', 'MeetingTable', 'LabWindow', 'Bulletin', 'Stairs', 'SauderBuilding', 'LibraryTree', 'ChemEngEntrance', respectively. The y-axis denotes pixel value in (a) and the difference of pixel values in (b).

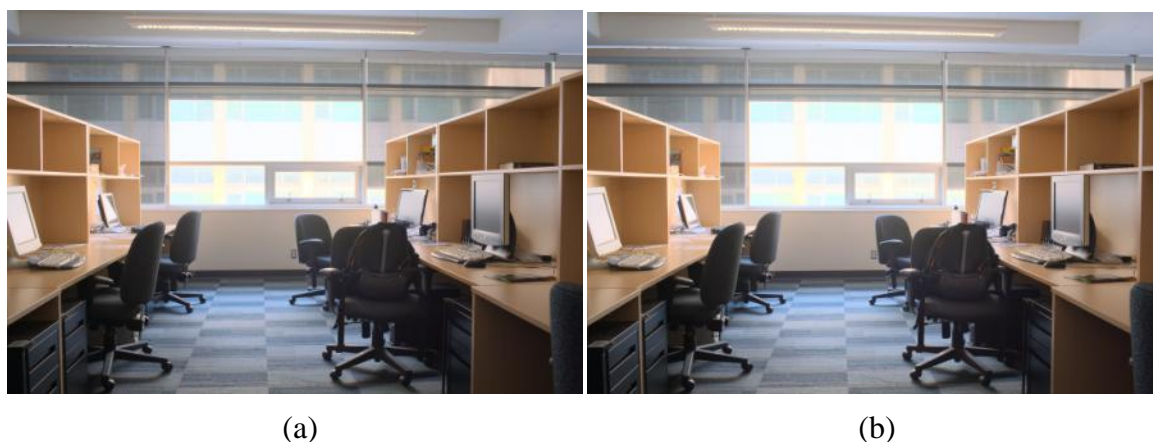


Fig. 5.11: Demonstration of the 2D (left) and the 3D (right) images at their average favored brightness levels. Marginal brightness difference can be seen between the 2D and the 3D versions.

no statistically significant difference in the preferred brightness level between a 3D image and its 2D counterpart. Fig. 5.11 illustrates the appearance of the 3D and the 2D images at their average favored brightness levels. Marginal brightness difference can be seen between the 3D and the 2D versions.

The results of the preferred mean image gradient of the 3D and the 2D representations are shown in Fig. 5.12(a). Each point is the average of the preferred gradient selected by each of the subjects for each of the scenes. For all scenes people selected a higher level of details in 3D viewing mode compared to 2D viewing. Similar to the brightness case, we compute the difference in mean gradient between the preferred 3D and the preferred 2D images for every single subject and every scene. Fig. 5.12(b) shows such difference averaged over 18 subjects for each of the scenes, and the 95% confidence intervals are also provided. Positive values on the vertical axis mean that people favor more details in 3D images than their 2D counterparts. It is observed from the plot that all the average difference is above zero. Moreover, for many of the scenes, the 95% confidence interval is either completely (scenes 3, 4, 5 and 6) or majorly (scenes 1, 2, 7 and 8) above zero. A reliable conclusion can thus be drawn that people prefer a higher level of details (i.e., a sharper image with more texture) when viewing in 3D than when viewing in 2D. This could be explained by the conclusion in Cormack et al. [80], that the 3D effect can be improved when the contrast is near the visibility threshold. By adjusting the tone-mapping process to produce more detail, the 3D effect in image regions where the contrast is near the threshold may be improved. We demonstrate the visual difference between the detail levels preferred for 3D and 2D images in Fig. 5.13.

Comparing Fig. 5.10(b) and Fig. 5.12(b), we notice that the confidence interval of the former (around seven brightness levels) is larger than that of the latter (around one detail

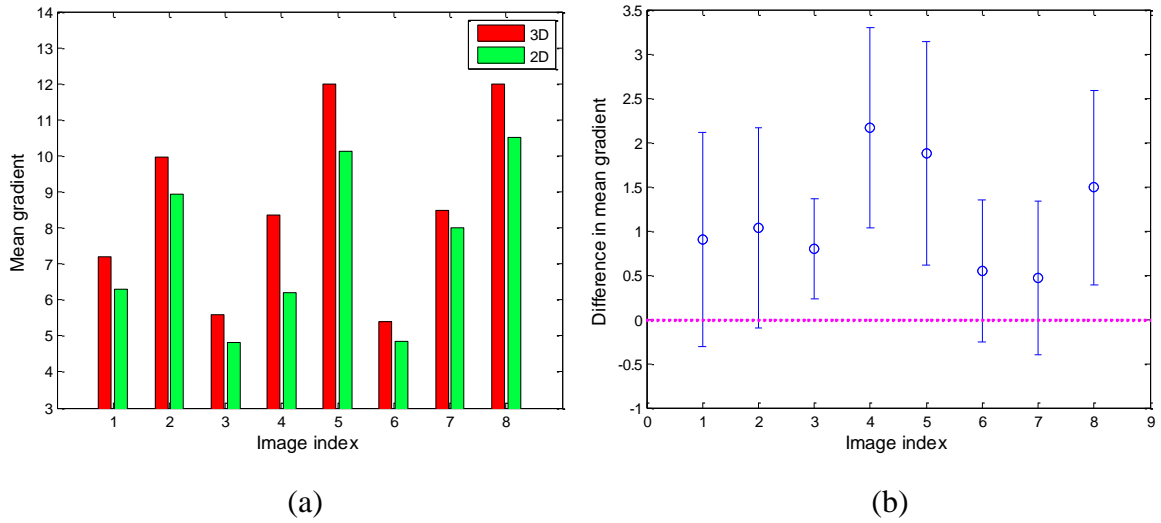


Fig. 5.12: Comparison between 2D and 3D images in terms of the preferred detail level: (a) – mean image detail, (b) subtraction of the mean image detail of 2D from that of 3D images. The horizontal axis denotes the image index, and the image order is the same as Fig. 4. The vertical axis denotes gradient value in (a) and the difference of gradient values in (b).

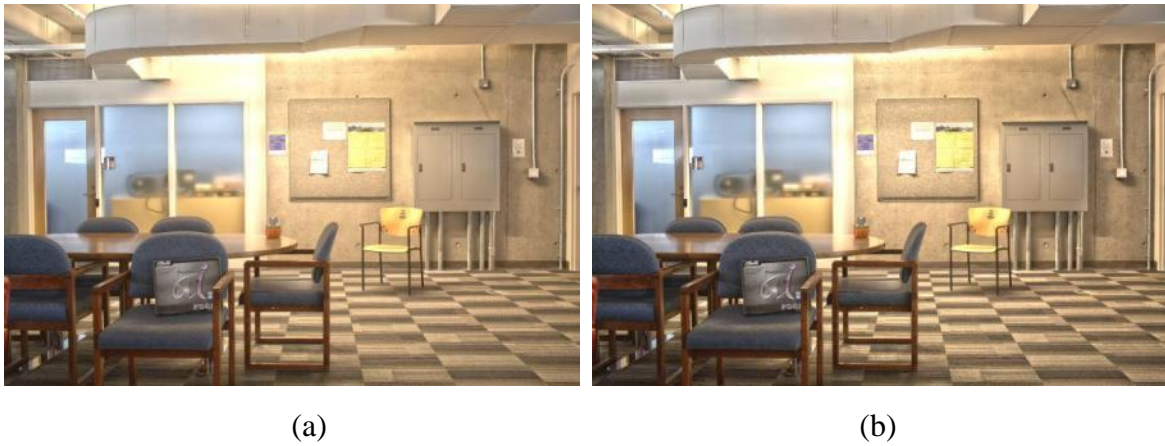


Fig. 5.13: Demonstration of the 2D (left) and the 3D (right) images at their average favored brightness levels. Marginal brightness difference can be seen between the 2D and the 3D versions.

level). In the preferred range for brightness, detecting the difference between close levels is not a simple task, a fact that can also be verified in Fig. 5.11 where minimal brightness difference can be identified between two similar levels. This is the reason for the higher variance in Fig. 5.10(b), which also leads to our conclusion that statistically there is no significant difference between 3D and 2D in terms of preferred brightness levels although the average preferred brightness level for 3D is higher than that for 2D. On the other hand when investigating details, there is a perceivable distinction between consecutive levels. This explains why the confidence interval for preferred detail levels is smaller about 1 – 1.5 levels.

5.4 Discussion

Although previous TMO evaluation studies are limited to 2D images, relating them to our results of the “3D overall quality” test would give an overall idea whether 3D viewing has any influence on the choice of tone-mapping methods. Our observation that the gradient TMO (*grad*) is among the lowest rated methods is consistent with the results from Čadík et al. [67] and Kuang et al. [109]. This is because *grad* produces images with unnatural appearance. However, our finding that LDR capturing with a good exposure performs much worse than the tested TMO’s (except for *grad*) is contrary to the conclusions of Akyüz et al. [57] and Čadík et al. [67]. Both of these studies concluded that a carefully tone-mapped HDR image is not statistically better than a well-exposed LDR image. Three causes may lead to such contradiction:

- i) In the case of 2D, the overall brightness and contrast are found to be the most significant factors that contribute to good image quality while detail reproduction falls behind [57] [67]. Hence, even if there is information loss in some regions, the image would be rated

good as long as the overall image displays sufficient brightness and contrast. This can be achieved by a well-exposed LDR image that favors one particular region of interest. It has been reported that viewers' eye movements are more widely distributed when watching 3D content [110] [111]. That is, when watching 3D, viewers do not just focus on the main region of interest but are more likely to explore the entire image area. Consequently, viewers are probably more likely to notice regions with loss of detail when viewing 3D, and this may give them a poor impression of the image. A well-exposed LDR image often has lots of clipped areas while a well-tuned tone-mapping method delivers a good overall visibility.

ii) As discussed before, lack of detail in the clipped regions can affect the 3D effect by making the two images harder to match in those regions. Having a poor 3D effect will obviously also reduce the overall 3D image quality.

iii) One explanation for the good performance of the well-exposed 2D image is that people were used to viewing under- and over-exposed images [57]. While viewers are used to clipping artifacts in 2D images, they are less accustomed to viewing under- and over-exposed images in 3D, so they may find the associated artifacts more objectionable.

In our results, no clear distinction is found between the performance of local and global tone-mapping methods. This is in agreement with the psychophysical evaluations presented by Ledda et al [64] as well as Ashikmin and Goyal [65]. On the contrary, Yoshida et al. [62] and Kuang et al. [63] observed an obvious dissimilarity between global and local TMOs, with local ones performing better. Other studies have come to the opposite conclusion; Akyüz et al. [66] and Čadík et al. [67] concluded that global operators are in favor relative to local TMOs. We believe that naturalness, good overall brightness/contrast, and sufficient

details in most regions are properties of a tone-mapping operator that is able to produce 3D images with good perceived quality.

It is interesting to find that in our first experiment several global and local TMOs have similar performance, despite some of these TMOs producing images with higher levels of details than others, while in our second experiment, we found more details are preferred for 3D presentations compared to their 2D counterparts. A possible reason is that the different TMOs in the first experiment result in images with different visual effect/styles. Global TMOs are known to produce more natural looking images [62], and images with more natural appearance may strengthen monoscopic depth cues and therefore give good 3D effect. The fact that local TMOs produce more details/texture could also strengthen the 3D effect. Since global and local TMOs both have positive qualities, in the end they could yield comparable 3D quality. In the second experiment, the 3D and the 2D images with different detail levels were generated by varying one parameter of a single tone-mapping operator, so the style of those images would be alike and the difference in detail levels will be the main factor for differentiating the viewers' preference. Another reason for the seemingly contradictory results between the two experiments is that in preparing the images used in the first experiment, we first tuned the parameters of each tone-mapping method to get the best possible 3D quality. It could be that tuning the parameters of each method brought most of the contrast to the supra-threshold level, at which point increasing the contrast further does not improve the 3D effect [80]. This may explain the comparable performance of several TMO's in the first experiment. In the second experiment, the users could adjust the level of detail over a very wide range, so more of the image content may have been close to the contrast detection threshold, where contrast does have an impact on 3D effect.

It is also worth noting that the limitations of the 3D display used could influence the preferred levels of brightness and contrast. Virtually all 3D displays suffer from crosstalk, an effect where the content of the image intended for one eye will “leak” through to the other eye, resulting in the viewer seeing “double” or “ghost” images [112]. LCD based 3D displays have a similar problem, where the time it takes for a pixel to switch from one level to another can darken the appearance of some pixels [113]. Crosstalk is more visible in images with high contrast [112], so it is possible that viewers would prefer images with more contrast and different brightness when viewing a 3D display with lower crosstalk.

Tone-mapping the left and the right views independently might result in different brightness and detail levels in the LDR versions, even if the tone-mapping parameters are set exactly the same. This may happen to both global and local TMOs when a light source is blocked in one view while it appears on the other view. Another possibility for global TMOs is that the brightness in the non-overlapped areas of the two views (i.e., the left edge of the left image and the right edge of the right image) differs significantly. For local TMOs, this undesired phenomenon might also stem from the fact that the same object captured by two cameras can have different structures. A topic for future work could be designing a local tone-mapping method that jointly considers the information in the two views in order to generate images high in detail which is consistent between the views.

Although visual fatigue/discomfort has been found to be one of the important factors that affect the overall 3D quality of experience [84], we did not explicitly ask subjects to rate this aspect in our test. Accurately assessing fatigue requires prolonged viewing of a stimulus [114], which was not possible in these tests in which subjects had to rate many different images while being allowed to control the amount of time they viewed each image. In future

studies we would like to identify all the factors that may affect the 3D QoE and evaluate each of these 3D image attributes separately.

5.5 Conclusion

This chapter addressed the problem of presenting stereoscopic tone-mapped HDR images on 3D LDR displays and how it is different from the 2D scenario. Two experiments have been conducted towards the objectives of this study.

In the first part of our study, we presented a subjective evaluation of tone mapping operators on 3D images. We evaluated seven tone-mapping schemes, as well as LDR capturing, in two aspects: 3D effect and overall 3D quality. Several TMOs were found to have similar good performance. In general, there is no clear distinction between global and local TMOs. All of the tested TMOs had significantly higher scores for 3D effect than LDR capturing, which shows that capturing in HDR and then applying tone mapping produces better 3D images than direct LDR capturing.

The second part of our study focuses on how the preferred level of brightness and the preferred amount of details differ between tone-mapped 3D and 2D images. “Preferred” means the best depth impression for 3D images and the best overall picture quality for 2D images. Images at a large range of different brightness and detail levels were generated from high dynamic range capturing followed by a tone-mapping process. We conducted an intensive subjective experiment that allows participants to select 3D and 2D images with their preference on brightness and details, respectively. Our results show that while people selected slightly brighter images in 3D viewing compared to 2D, the difference is not statistically significant. However, compared to 2D images, the subjects consistently preferred having a greater amount of details when viewing in 3D.

6 Thesis Summary

6.1 Significance of the Research

The future of high dynamic range (HDR) imaging technologies is bright. This is because they have the potential of significantly increasing the viewer's quality of experience. Since almost all current displays support only a low dynamic range (LDR), it is desirable to design tone-mapping schemes that convert HDR to the LDR format so as to enable the viewing of images and videos using existing LDR devices. More importantly, existing displays are also able to deliver better picture quality if the content is first captured in HDR and then tone-mapped to LDR.

In this thesis we address the different aspects of tone-mapping, that bridge high dynamic range technology and tone-mapping with video coding and 3D quality of experience.

In Chapter 2, we consider a backward-compatible HDR video compression case where only the tone-mapped LDR content (base layer) is compressed and transmitted and the HDR video is reconstructed by inversely tone-mapping the decoded LDR video. We present our solution for finding an optimized tone mapping curve that significantly improves the reconstructed (inversely-tone-mapped) HDR quality in backward-compatible high dynamic range image/video compression. We develop a statistical model that represents an approximation of the distortions resulting from the combined processes of tone-mapping, compression, de-compression and inverse tone-mapping. We formulate a numerical optimization problem using this model to find the tone-curve that minimizes the distortion in the reconstructed HDR sequence. We also develop a simplified version of the model that

leads to a closed-form solution for the optimization problem, allowing implementation in real-time.

In Chapter 3, we consider the case where an enhancement layer that encodes the difference between the original and the inversely-tone-mapped HDR signals is also compressed and transmitted, in addition to the LDR base layer. We present a new tone-mapping approach that provides superior compression efficiency while preserving a desired perceptual quality of the tone-mapped LDR video. We develop statistical models that formulate the bit rates of each of the base and the enhancement layers, as well as the mismatch between the resulting LDR base-layer signal and the predefined base layer representation. The models are then incorporated in our optimization problem. The temporal effect of tone-mapping is also considered in order to avoid flickering artifacts.

In Chapter 4, we present our color correction (also called color mapping) method that is applied directly in the YCbCr color space. We derive a closed-form tone-mapping solution for the Cb and Cr chrominance components, given the HDR-LDR mapping of the luminance channel. This solution has significantly lower computational complexity because it avoids processes such as color space transformation and up-sampling, which are required by the conventional method.

Chapter 5 addresses the problem of displaying stereoscopic tone-mapped (HDR to LDR) images on 3D LDR displays and how it is different from the 2D scenario. We first present a subjective psychophysical experiment that evaluates existing tone-mapping operators in the case where 3D HDR images are tone-mapped and then displayed on 3D LDR devices. The second part of this study focuses on how the preferred level of brightness and the preferred amount of details differ between 3D and 2D images, by conducting another set

of subjective experiments. We analyze the results from the two sets of subjective tests to find out which attributes of tone-mapping methods contribute to good 3D representation.

6.2 Potential Applications

The applications of tone-mapping include digital photography, video recording, digital video broadcasting and medical imaging, for a wide range of modern capturing, transmission, storage and display systems.

The two proposed tone-mapping operators in Chapter 2 and Chapter 3 can be applied within the framework of the bit-depth scalable video coding, an extension supported by the major video compression standard H.264/AVC. The former one is more computationally efficient, so it can be used for on-the-fly broadcasting; the latter one allows selecting a desired visual style of the LDR video, so it can be integrated into any post-production software which offers interaction between the user and the raw/unedited video. Regardless of the applications, the two tone-mapping operators both deliver high compression gains as well as satisfactory picture quality. The color correction scheme proposed in Chapter 4 can be incorporated into any encoding system that supports backward-compatible HDR video compression.

In addition, the low computational complexity of the proposed schemes in Chapters 2 and 4, allows these schemes to be embedded in video capturing systems which require low power consumption and have limited storage space. A large consumer market for this includes video cameras and video enabled mobile phones. Existing capturing devices only allow real-time generation of tone-mapped still images (although they are marketed as “HDR”). Our real-time tone-mapping approach (the closed-form solution) and colour tone-mapping in the YCbCr domain, will be a good foundation in real-time generation of video. In

a different scenario, where only HDR is transmitted, the proposed techniques can also be added to the set-top-box in households and perform instantaneous HDR video tone-mapping in order to render 8-bit signals for viewing on an LDR display.

The findings in Chapter 5 can be applied when developing a good tone-mapping operator for 3D applications. A tone-mapping designed for 3D signals is essential since i) combining 3D with HDR or with tone-mapping will significantly increase the perceived picture quality, and ii) all displays for 3D images/videos will probably be of the LDR type for some time. Moreover, regular 3D image/video recording can also benefit from our results on the difference between 2D and 3D in viewers' preferred brightness and detail levels.

6.3 Contributions

The main contributions of this thesis are summarized as follows:

- We first considered a backward-compatible HDR video compression case where only the tone-mapped LDR content (base layer) is compressed and transmitted and the HDR video is reconstructed by inversely tone-mapping the decoded LDR video. We developed a statistical model that represents how the distortions from the joint processes of tone-mapping, compression, de-compression and inverse tone-mapping arise. Using this model, we formulated a constrained optimization problem that finds the HDR-LDR tone-curve which minimizes the expected value of the distortions in the reconstructed (inversely-tone-mapped) HDR sequence.

We modified the above model so that it reduces the computational complexity of the optimization problem and leads to a closed-form solution. The closed-form solution is computationally efficient and has a performance comparable to our developed statistical model. Moreover, the closed-form solution does not require

the knowledge of QP (quantization parameter or quantization level), which makes it suitable for cases where the compression strength is unknown.

We showed that the appropriate choice of a tone-mapping curve using the distortion models we developed can significantly improve the quality of the reconstructed HDR stream that is inversely tone-mapped from the encoded-decoded LDR video

- We then considered the case where an enhancement layer (that encodes the residual signal between the original and the inversely-tone-mapped HDR signals) is also transmitted in addition to the LDR base layer. We developed a global (spatially invariant) tone-mapping operator that minimizes the overall bit rate of both the base layer and the enhancement layer while preserving a desired perceptual quality of the tone-mapped LDR video content. We derived two statistical models: one formulates the bit rate of the base layer and the other formulates that of the enhancement layer. In order to achieve good perceptual quality, we also developed a model of the mismatch between our resulting LDR signals and a preset tone-mapping method. Using a favorable preset TMO as the baseline of the LDR quality guarantees that the resulting tone-mapped video looks pleasant. We incorporated these three models in our optimization problem. Our results showed that the proposed solution provides superior compression efficiency with a desired perceptual quality for the tone-mapped LDR video.

To minimize flickering artifacts, we derived an analytical expression that models the temporal consistency between consecutive resulting LDR frames as a function of tone-mapping parameters. Incorporating this model in the constraint of our

optimization problem successfully avoids flickering artifacts, which is problematic for some existing tone-mapping methods as they are designed originally for still images. This constraint also improves the coding efficiency because the temporal difference between two consecutive frames in the tone-mapped video is reduced.

- We derived a chromatic mapping method for tone-mapping that works directly in the YCbCr color space, which is most commonly used for video applications, but unfortunately all conventional color correction (color mapping) methods for tone-mapping require converting the signal to the RGB domain. Our method relies on a closed-form mapping solution that has significantly lower computational complexity as it avoids processes such as color space transformation and up-sampling which cannot be avoided by the conventional methods.
- We investigated the problem of displaying tone-mapped 3D content using stereoscopic LDR displays. We first conducted a subjective psychophysical experiment that evaluates the quality of 3D LDR images that result from using existing tone-mapping operators (TMOs). To the best of our knowledge, it is the first attempt to assess TMOs for 3D visual signals. Our results showed that capturing in HDR and then applying tone mapping produce 3D LDR images with better quality than those obtained by direct LDR capturing. We then undertook another set of subjective experiments to find out the viewers' preferred levels of brightness and scene details in each of the 3D and 2D cases.

6.4 Suggestions for Future Research

The introduction of high dynamic range video technologies can only be a long-lasting success if a large amount of HDR video content is available. This requires HDR video capturing to be cost-effective and easy to achieve technically. Unfortunately, current technologies suffer from either complicated hardware setup, such as using beam splitters with multiple LDR cameras [11], [12], [13], or from extremely expensive HDR camera sensors [14], [15], [16]. These two drawbacks will deter regular consumers from using HDR video recording, which slows the evolution of the HDR market. These problems could be addressed by allowing a simple stereo setup consisting of regular cameras with different exposures, placed side by side. Possible future work would be that of studying how to combine all the views (that have spatial disparity and exposure disparity) into an HDR signal. Another issue worth exploring is the exposure difference among the cameras so that HDR reconstruction can be achieved while the number of cameras used is minimized. This camera setup might have the added advantage in that if an HDR representation is produced for one of the views, then we may be able to produce HDR videos for any other views. Consequently, 3D HDR videos can be obtained.

The compression gain of HDR videos can be improved by changing the components in the encoding engine to be more HDR oriented. Rate-distortion (RD) optimization is used inside the compression system for making coding decisions. Current solutions to the RD optimization problem are all limited to LDR distortion assessment models. Applying HDR quality metrics [69], [115], [116], [117] in the formulation of RD optimization can improve the quality of the resulting coded HDR video. In the future, we would like to investigate modifying the RD optimization process, using distortion evaluation metrics that are designed

for HDR signals. A challenge arises from the fact that video encoding is a block-by-block process while the existing HDR-specific metrics require information of the entire frame. A solution may be achieved by using temporal information and knowledge of adjacent frames, together with the blocks that are already reconstructed in the current frame.

Tone-mapping the left and the right views independently might result in different brightness and detail levels in the LDR versions, even if the tone-mapping parameters are set exactly the same. This may happen to both global and local TMOs when a light source is blocked in one view while it appears in the other view. Another possibility for global TMOs is that the brightness in the non-overlapped areas of the two views (i.e., the left edge of the left image and the right edge of the right image) differs significantly. For local TMOs, this undesired phenomenon might also stem from the fact that the same object captured by two cameras can have different structures. A topic for future work could be designing a tone-mapping method that jointly considers the information in the two views in order to generate images high in detail which is consistent between the views.

Bibliography

- [1] J. A. Ferwerda, "Elements of Early Vision for Computer Graphics," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 22–33, 2001.
- [2] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, "High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting," Morgan Kaufmann, 2005.
- [3] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High Dynamic Range Display Systems," *ACM Transactions on Graphics (Proc. of SIGGRAPH 04)*, vol. 23, no. 3, pp. 757–765, 2004.
- [4] M. Robertson, S. Borman, and R. Stevenson, "Dynamic Range Improvement through Multiple Exposures," in *Proc. of the 1999 International Conference on Image Processing (ICIP1999)*, pp. 159-163, 1999.
- [5] P. E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in *Proc. the 24th annual conference on Computer graphics and interactive techniques (Proc. of SIGGRAPH 97)*, pp. 369-378, 1997.
- [6] G.J. Sullivan, Haoping Yu, S.-i. Sekiguchi, Huifang Sun, T. Wedi, S. Wittmann, Yung-Lyul Lee, A. Segall, and T. Suzuki, "New Standardized Extensions of Mpeg4-avc/H.264 for Professional-Quality Video Applications," in *Proc. of the 2007 International Conference on Image Processing (ICIP2007)*, vol. 1, pp. I–13–16, 2007.
- [7] M. Winken, D. Marpe, H. Schwarz, and T. Wiegand, "Bit-Depth Scalable Video Coding," in *Proc. of the 2007 International Conference on Image Processing (ICIP2007)*, vol. 1, pp. I–5–8, 2007.
- [8] A. Segall and J. Zhao, "Bit-Stream Rewriting for SVC-to-AVC Conversion," in *Proc. of the 2008 International Conference on Image Processing (ICIP2008)*, pp. 2776–2779.
- [9] Y. Wu, Y. Gao, and Y. Chen, "Bit-Depth Scalability Compatible to H.264/AVC-Scalable Extension," *Journal of Visual Communication and Image Representation*, vol. 19, no. 6, pp. 372–381, 2008.
- [10] S. B. Kang , M. Uyttendaele , S. Winder , and R. Szeliski, "High Dynamic Range Video," *ACM Transactions on Graphics (Proc. of SIGGRAPH2003)*, vol.22, no.3, pp. 319- 325, 2003.
- [11] M. Aggarwal and N. Ahuja, "Split Aperture Imaging for High Dynamic Range," in *Proc. IEEE International Conference on Computer Vision*, pp. 10-16, 2001.
- [12] H. Wang, R. Raskar, and N. Ahuja, "High Dynamic Range Video using Split Aperture Camera," In *Proceedings of the 6th Workshop on Omnidirectional Vision (OMNIVIS)*, 2005.
- [13] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A Versatile HDR Video Production System," *ACM Transactions on Graphics (Proc. of SIGGRAPH2011)*, vol. 30, no. 4, article 41, 2011.
- [14] B. Fowler, A. E. Gamal, and D. Yang, "Techniques for Pixel Level Analog to Digital Conversion," in *Proc. SPIE*, vol. 3360, pp. 2–12, 1998.

- [15] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded Rolling Shutter Photography: Flexible Space-Time Sampling," in *Proc. IEEE International Conference on Computational Photography (ICCP2010)*, pp. 1-8, 2010.
- [16] "RED Camera," Available: <http://www.red.com/products/epic>. Last access: October 15, 2011.
- [17] L. Meylan, S. Daly, and S. Susstrunk, "The Reproduction of Specular Highlights on High Dynamic Range Displays," in *Proc. the 14th Color Imaging Conference*, pp. 333-338, 2006.
- [18] L. Wang, L. Wei, K. Zhou, B. Guo, and H.-Y. Shum, "High Dynamic Range Image Hallucination," in *Proc. 18th Eurographics Symposium on Rendering*, pp. 321-326, 2007.
- [19] H. Farid, "Blind Inverse Gamma Correction," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1428-1433, 2001.
- [20] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heihrich, L. Whitehead, and G. Ward, "Ldr2Hdr: On-the-Fly Reverse Tone Mapping of Legacy Video and Photographs," *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 26, no. 3, article 39, 2007.
- [21] D. Xu, C. Doutre, and P. Nasiopoulos, "Correction of Clipped Pixels in Color Images," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 3, pp. 333-344, 2011.
- [22] G. Ward, "A Wide Field, High Dynamic Range, Stererographic Viewer," in *Proc. PICS*, 2002.
- [23] R. D. Hersch, "Spectral Prediction Model for Color Prints on Paper with Fluorescent Additives," *Applied Optics*, vol. 47, no. 36, pp. 6710-6722, 2008.
- [24] R. D. Hersch, P. Doneze, and S. Chosson, "Color Images Visible Under UV Light," *ACM Transactions on Graphics (Proc. of SIGGRAPH 07)*, vol. 26, no. 3, pp. 75, 2007.
- [25] IEC61966-2-4, "Colour measurement and management - part 2-4: Colour management - extended-gamut YCC colour space for video applications - xvYCC," 2006.
- [26] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-Motivated High Dynamic Range Video Encoding," *ACM Transactions on Graphics (Proc. SIGGRAPH04)*, vol. 23, no. 3, pp. 730-738, 2004.
- [27] G. Ward, "Real Pixels," *Graphics Gems II*, pp. 80-83, 1991.
- [28] R. Bogart, F. Kainz, and D. Hess, "Openexr Image File Format," *ACM SIGGRAPH 2003, Sketches and Applications*, 2003.
- [29] G. Ward Larson, "Logluv Encoding for Full-Gamut, High-Dynamic Range Images," *Journal of Graphics Tools*, vol. 3, no. 1, pp. 15-31, 1998.
- [30] R. Xu, S.N. Pattanaik and C.E. Hughes, "High-Dynamic-Range Still-Image Encoding in JPEG 2000," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 57-64, 2005.

- [31] K. E. Spaulding, G. J. Woolfe, and R. L. Joshi, "Using a Residual Image to Extend the Color Gamut and Dynamic Range of an sRGB Image," in *Proc. of IS&T PICS Conference*, pp. 307–314, 2003.
- [32] G. Ward and M. Simmons, "JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG," in *Proceedings of the 13th Color Imaging Conference*, pp. 283–290, 2005.
- [33] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward Compatible High Dynamic Range MPEG Video Compression," *ACM Transactions on Graphics (Proc. SIGGRAPH06)*, vol. 25, no. 3, 2006.
- [34] A. Segall, "Scalable Coding of High Dynamic Range Video," in *Proc. of the 2007 International Conference on Image Processing (ICIP2007)*, vol. 1, pp. I–1–I–4.
- [35] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.
- [36] S. Liu, W.-S. Kim, and A. Vetro, "Bit-Depth Scalable Coding for High Dynamic Range Video," in *Proc. of SPIE Visual Communications and Image Processing 2008*, vol. 6822, 2008.
- [37] R. Mantiuk, S. Daly, and L. Kerofsky, "Display Adaptive Tone Mapping," *ACM Transactions on Graphics (Proc. SIGGRAPH08)*, vol. 27, no. 3, paper 28, 2008.
- [38] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," *ACM Transactions on Graphics (Proc. SIGGRAPH02)*, vol. 21, no. 3, pp. 267–276, 2002.
- [39] J. Tumblin and G. Turk, "LCIS: A Boundary Hierarchy for Detail-Preserving Contrast Reduction," In *Computer Graphics Proceedings (SIGGRAPH1999)*, , pp. 83–90, 1999.
- [40] D. Linschinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive Local Adjustment of Tonal Values," *ACM Transactions on Graphics (Proc. SIGGRAPH06)*, vol. 25, no. 3, pp. 646-653, 2006.
- [41] E. Reinhard, and K. Devlin, "Dynamic Range Reduction Inspired by Photoreceptor Physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 13-24, 2005.
- [42] "Luminance HDR," Available: <http://qtpfsgui.sourceforge.net>. Last access: November 1, 2011.
- [43] M. Ashikhmin, "A Tone Mapping Algorithm for High Contrast Images," in *Proc. the 13th Eurographics workshop on Rendering*, pp. 1-11, 2002.
- [44] R. Mantiuk ,K. Myszkowski , H.-P. Seidel, "A Perceptual Framework for Contrast Processing of High Dynamic Range Images," *ACM Transactions on Applied Perception (TAP)*, vol.3 no.3, pp. 286-308, 2006.
- [45] "Photomatix HDR software," Available: <http://www.hdrsoft.com>. Last access: October 28, 2011.

- [46] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and Componding High Dynamic Range Images with Subband Architectures," *ACM Transactions on Graphics (Proc. SIGGRAPH2005)*, vol. 24, no. 3, pp. 836–844, 2005.
- [47] C. Lee and C.-S. Kim, "Gradient Domain Tone Mapping of High Dynamic Range Videos," in *Proc. of the 2007 International Conference on Image Processing (ICIP2007)*, vol. 3, pp. III–461–III–464, 2007.
- [48] R. Fattal, D. Lischinski, and M. Werman, "Gradient Domain Gigh Dynamic Range Compression," *Computer Graphics Forum (Proc. Of Eurographics)*, vol. 21, no. 3, pp. 249–256, 2002.
- [49] R. Mantiuk and H.P. Seidel, "Modeling a Generic Tone-mapping Operator," *Computer Graphics Forum (Proc. of Eurographics'08)*, vol. 27, no. 2, pp. 699–708, 2008.
- [50] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive Logarithmic Mapping for Displaying High Contrast Scenes," *Computer Graphics Forum (Proc. of Eurographics)*, vol. 22, no. 3, pp. 419–426, 2003.
- [51] G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perceptual Effects in Real-Time Tone Mapping," in *Proc. the 21st spring conference on Computer graphics*, pp. 195–202, 2005.
- [52] H. Wang, R. Raskar, and N. Ahuja, "High Dynamic Range Video using Split Aperture Camera," *IEEE 6th Workshop on Omnidirectional Vision (OMNIVIS) in conjunction with ICCV'05*, 2005.
- [53] A. Benoit, D. Alleysson, J. H  rault, and P.L. Callet, "Spatio-Temporal Tone Mapping Operator Based on a Retina Model", in *Proc. Computational Color Imaging Workshop (CCIW2009)*, pp.12-22, 2009.
- [54] M. Fairchild and G. Johnson, "The iCAM Framework for Image Appearance, Image Differences, and Image Quality," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 126–138, 2004.
- [55] J. Kuang, G. Johnson, and M. Fairchild, "iCAM06: A Refined Image Appearance Model for HDR Image Rendering," *Journal of Visual Communication and Image Representation*, vol. 18, no. 5, pp. 406–414, 2006.
- [56] C. Schlick, "Quantization Techniques for the Visualization of High Dynamic Range Pictures," in *Photorealistic Rendering Techniques*, Springer-Verlag, New York, pp. 7–20, 1994.
- [57] A. Akyuz and E. Reinhard, "Color Appearance in High-Dynamic-Range Imaging," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 126–138, 2004.
- [58] S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, "Time Dependent Visual Adaptation for Fast Realistic Image Display," *27th Ann. Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH2000)*, pp. 47–54, 2000.
- [59] R. Mantiuk, R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color Correction for Tone Mapping," *Computer Graphics Forum (Proc. Of EUROGRAPHICS)*, vol. 28, no. 2, pp. 193–202, 2009.

- [60] M. Toda, M. Tsukada, A. Inoue, and T. Suzuki, "High Dynamic Range Rendering for YUV Images with a Constraint on Perceptual Chroma Preservation," in *Proc. 16th IEEE International Conference on Image Processing (ICIP2009)*, pp. 1817 – 1820, 2009.
- [61] F. Drago, W. L. Martens, K. Myszkowski, and H.-P. Seidel, "Perceptual Evaluation of Tone Mapping Operators," In *ACM Proc. SIGGRAPH 2003 Conference on Sketches and Applications*, pp. 1, 2003.
- [62] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel, "Perceptual Evaluation of Tone Mapping Operators with Real-World Scenes," *SPIE: Human Vision & Electronic Image X*, San Jose, CA, USA, pp. 192-203, 2005.
- [63] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild, "Evaluating HDR Rendering Algorithms," *ACM Transactions on Applied Perception*, vol. 4, no. 2, article 9, 2007.
- [64] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of Tone Mapping Operators using a High Dynamic Range Display," in *ACM Transactions on Graphics (Proc. SIGGRAPH2005)*, pp. 640-648, 2005.
- [65] M. Ashikhmin, and J. Goyal, "A Reality Check for Tone-Mapping Operators," *ACM Transactions on Applied Perception (TAP)*, vol.3 no.4, pp. 399-411, 2006.
- [66] A. O. Akyüz , R. Fleming , B. E. Riecke , E. Reinhard , and H. H. Bühlhoff, "Do HDR Displays Support LDR Content: a Psychophysical Evaluation," *ACM Transactions on Graphics (TOG)*, vol.26 no.3, 2007.
- [67] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR Tone Mapping Methods using Essential Perceptual Attributes," *Computers & Graphics*, vol. 32, no. 3, pp. 330-349, 2008.
- [68] Z. Mai, C. Doutre, P. Nasiopoulos, and R. K. Ward, "Subjective Evaluation of Tone-Mapping Methods on 3D Images," in *Proc. 17th International Conference on Digital Signal Processing (DSP2011)*, 2011.
- [69] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. P. Seidel, "Dynamic range independent image quality assessment," *International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, 2008.
- [70] H. Yeganeh , and Z. Wang, "Objective Assessment of Tone Mapping Algorithms," in *Proc. 17th IEEE International Conference on Image Processing (ICIP2010)*, pp. 2477-2480, 2010.
- [71] M. Cowan, J. Greer, L. Lipton and J. Chiu, "Enhanced ZScreen Modulator Techniques," US Patent No. 7,760,157, 2010.
- [72] M. Richards and W. Allen, "Method and System for Shaped Glasses and Viewing 3D Images," US Patent Application No. 2010/0066976, 2010.
- [73] S. J. Kim and M. Pollefeys, "Radiometric self-alignment of image sequences," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2004)*, pp. 1645-651, 2004.

- [74] A. Troccoli, S.B. Kang and S. Seitz, "Multi-View MultiExposure Stereo," in *Proc. Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pp. 861-868, 2006.
- [75] N. Sun, H. Mansour, R. Ward, "HDR Image Construction from Multi-Exposed Stereo LDR Images," in *Proc. of the IEEE International Conference on Image Processing (ICIP2010)*, Hong Kong, 2010.
- [76] Ramachandra, V., Zwicker, M., Truong Nguyen, "HDR Imaging From Differently Exposed Multiview Videos", in *Proc. 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, vol. 28, no. 30, pp. 85-88, 2008.
- [77] S. Hrabar, P. Corke, and M. Bosse, "High Dynamic Range Stereo Vision for Outdoor Mobile Robotics," in *Proc. IEEE International Conference on Robotics and Automation 2009 (ICRA 2009)*, pp. 430 – 435, 2009.
- [78] B. Huhle, O. Pirinen, S. Fleck, A. Gotchev, and W. Strasser, "Why HDR is Important for 3DTV Model Acquisition," in *Proc. 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp. 45 - 48 , 2008.
- [79] J. Zhu, G. Humphreys, D. Koller, S. Steuart, W. Rui, "Fast Omnidirectional 3D Scene Acquisition with an Array of Stereo Cameras," in *Proc. Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, pp. 217 – 224, 2007.
- [80] L.K. Cormack, S.B. Stevenson, and C.M. Schor, "Interocular Correlation, Luminance Contrast and Cyclopean Processing," *Vision Research*, vol.31, no.12, pp. 2195–2207, 1991.
- [81] G. Jones, D. Lee, N. Holliman, and D. Ezra, "Controlling Perceived Depth in Stereoscopic Images," in *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems*, vol. 4297, pp. 42–53, 2001.
- [82] ITU-R, "Digital Three-Dimensional (3D) TV Broadcasting," Question ITU-R 128/6, 2008.
- [83] ITU-T, "Objective and Subjective Methods for Evaluating Perceptual Audiovisual Quality in Multimedia Services within the Terms of Study Group 9," Question 12/9, 2009.
- [84] M.T.M Lambooi, W.A. IJsselsteijn, and I. Heynderickx, "Visual Discomfort in Stereoscopic Displays: a Review," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490, Jan. 2007.
- [85] L. Goldmann, F. D. Simone, and T. Ebrahimi, "Impact of Acquisition Distortion on the Quality of Stereoscopic Images," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2010.
- [86] B. Mendiburu, "3D Movie Making – Stereoscopic Digital Cinema from Script to Screen." Elsevier, 2008.
- [87] M. T. Pourazad, Z. Mai, P. Nasiopoulos, R. K. Ward, and K. Plataniotis, "Effect of Brightness on the Quality of Visual 3D Perception," in *Proc. 18th IEEE International Conference on Image Processing (ICIP2011)*, 2011.

- [88] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic Images Quality Assessment," in *Proc. of 15th European Signal Processing Conference (EUSIPCO '07)*, Poznan, Poland, 2007.
- [89] A. Benoit, P. Le. Callet, P. Campisi, and R. Cousseau, "Quality Assessment of Stereoscopic Images," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp.1-13, 2008.
- [90] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual Quality Assessment for Stereoscopic Images Based on 2D Image Quality and Disparity Analysis," in *Proc. Int. Workshop on Video Process. and Quality Metrics for Consumer Electronics -VPQM*, 2010.
- [91] H. Shao, X. Cao, and G. Er, "Objective Quality of Depth Image Based Rendering in 3DTV System," in *Proc. 3DTV Conference*, pp. 1–4, 2009.
- [92] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [93] A. Mittal, A.K. Moorthy, J. Ghosh and A.C. Bovik, "Algorithmic Assessment of 3D Quality of Experience for Images and Videos," in *Proc. IEEE DSP Wkshp*, Sedona AZ, pp. 338 – 343, 2011.
- [94] IEC61966-2-4, "Colour Measurement and Management - Part 2-4: Colour Management - Extended-Gamut YCC Colour Space for Video Applications - xvYCC," 2006.
- [95] Selig Hecht, "The Visual Discrimination of Intensity and the Weber-Fechner law," *The Journal of General Physiology*, vol. 7, no. 2, pp. 235C267, 1924.
- [96] "H.264/AVC JM14.2 Reference Software," Available: <http://iphome.hhi.de/suehring/tml/>. Last access: November 15, 2011.
- [97] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic Range Independent Image Quality Assessment," *ACM Transactions on Graphics (Proc. SIGGRAPH08)*, vol. 27, no. 3, 2008.
- [98] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting Visible Differences in High Dynamic Range Images Model and Its Calibration," in *Proc. of Human Vision and Electronic Imaging X. (of Proceedings of SPIE)*, vol. 5666, pp. 204–214, 2005.
- [99] P. Topiwala and H. Yu, "New Test Sequences in the VIPER 10-bit HD Data," Tech. Rep., ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-Q090, 2005.
- [100] X. Jing and L. P. Chau, "A Novel Intra-Rate Estimation Method for H.264 Rate Control," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS2006)*, pp. 5019-5022, 2006.
- [101] M. Winken, H. Schwarz, D. Marpe and T. Wiegand, "CE2: SVC bit-depth scalable coding," Technical report, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-X057, 2007

- [102] F. Durand, and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics (Proc. SIGGRAPH02)*, vol. 21, no. 3, pp.257–266, 2002.
- [103] K. Chiu, M. Herf, P. Shirley, S. Swamy, C.Wang, and K. Zimmerman, "Spatially nonuniform scaling functions for high contrast images," in *Proc. Interface '93*, pp. 245–253, 1993.
- [104] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display," in *ACM Transactions on Graphics (Proc. SIGGRAPH88)*, pp. 287–298, 1998.
- [105] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, W. Heidrich, " Optimizing a tone curve for backward-compatible high dynamic range image/video compression ", *IEEE Transactions on Image Processing (TIP)*, vol. 20, no. 6, pp. 1558-1571, 2011.
- [106] "Supplementary material," Available:
http://www.ece.ubc.ca/~zicongm/subjective_test_3dtmo/3DTMO_HTMLReport.html .
 Last access: December 2011.
- [107] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Tech. Rep. BT.500-11, 2002.
- [108] R. V. Hogg, and J. Ledolter, "*Engineering statistics*," New York: MacMillan, 1987.
- [109] J. Kuang, H. Yamaguchi, G. M. Johnson, and M. D. Fairchild, "Testing HDR Image Rendering Algorithms," in *Proc. Color Image Conference*, pp. 315-320, 2004.
- [110] J. Häkkinen, T. Kawai, J. Takatalo, R. Mitsuya, and G. Nyman, "What Do People Look at When They Watch Stereoscopic Movies?" in *Proc. SPIE Conf. Stereoscopic Displays and Applications XXI*, vol. 7524, San Jose, 2010.
- [111] L. Jansen, S. Onat, and P. König, "Influence of Disparity on Fixation and Saccades in Free Viewing of Natural Scenes," *Journal of Vision*, vol. 9, no. 1, pp. 1–19, 2009.
- [112] A. J. Woods, "Understanding Crosstalk in Stereoscopic Displays" (Keynote Presentation) at *Three-Dimensional Systems and Applications (3DSA) Conference*, Tokyo, Japan, 19-21 2010.
- [113] S. Shestak, D. Kim, and S. Hwang "Measuring of Gray-to-Gray Crosstalk in a LCD Based Time-Sequential Stereoscopic Display" proc. of the Society for Information Display (SID 2010), Seattle, pp. 132-135, 2010.
- [114] M. Pölönen, T. Jarvenpaa, J. Hakksinen, "Comparison of Near-to-Rye Displays: Subjective Experience and Comfort," *J. Display Technol.*, vol. 6, no. 1, pp. 27–35, 2010.
- [115] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, "Extending Quality Metrics to Full Luminance Range Images," in *Proc. of Human Vision and Electronic Imaging XIII*, vol. 6806, p. 68060C, 2008.
- [116] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting Visible Differences in High Dynamic Range Images Model and its Calibration," in *Proc. of Human Vision and Electronic Imaging X*, vol. 5666 of *Proceedings of SPIE*, pp. 204-214, SPIE, 2005.

- [117] R. Mantiuk, K. J. Kim, A. G. Rempel and W. Heidrich, “HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions,” *ACM Transactions on Graphics (Proc. SIGGRAPH11)*, vol. 30, no. 4, 2011.

Appendix A – H.264/AVC Intra Coding Error Model (p_C)

Let $v, \tilde{v} \in \{0, 1, \dots, v_{max}\}$, be the original and decoded values of a LDR pixel, respectively. We denote by $p_C(v - \tilde{v})$ the probability that decoded pixel luma level has shifted by a factor $(v - \tilde{v})$ from its original luma level.

For a specific image, $p_C(v - \tilde{v})$ can be estimated by subtracting each pixel value of the de-compressed image from that of the original image and then fitting a distribution for these differences by sampling such distribution on a large set of compression-distorted images. Let ω be equal to $(v - \tilde{v})$. We found that the probability can be well approximated with the General Gaussian Distribution (GGD):

$$P_C(\omega; \mu, \sigma, \alpha) = \frac{\lambda(\alpha, \sigma) \cdot \alpha}{2 \cdot \Gamma(1/\alpha)} \cdot e^{-[\lambda(\alpha, \sigma) \cdot |\omega - \mu|]^\alpha} \quad (\text{A.1})$$

Where μ denotes the mean, σ is the standard deviation and α denotes the shape parameter. The functions λ and Γ are expressed as follows:

$$\lambda = \frac{1}{\sigma} \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)} \right]^{1/2} \quad (\text{A.2})$$

$$\Gamma(m) = \int_0^\infty t^{m-1} \cdot e^{-t} dt, z > 0 \quad (\text{A.3})$$

and (A.3) is called the gamma function. In order to find a GGD fitting, the values of μ , σ and α need to be assigned. The distribution mean is set equal to 0 since all compression schemes make every effort to keep decoded pixel values unchanged. To find the standard deviation σ and the shape parameter α that best fit the image histograms, we use the least square regression. Note that σ and α vary for different images and different values of quantization parameters (QPs). Fig. A.1 shows example error distributions and the resulting fitting curves.

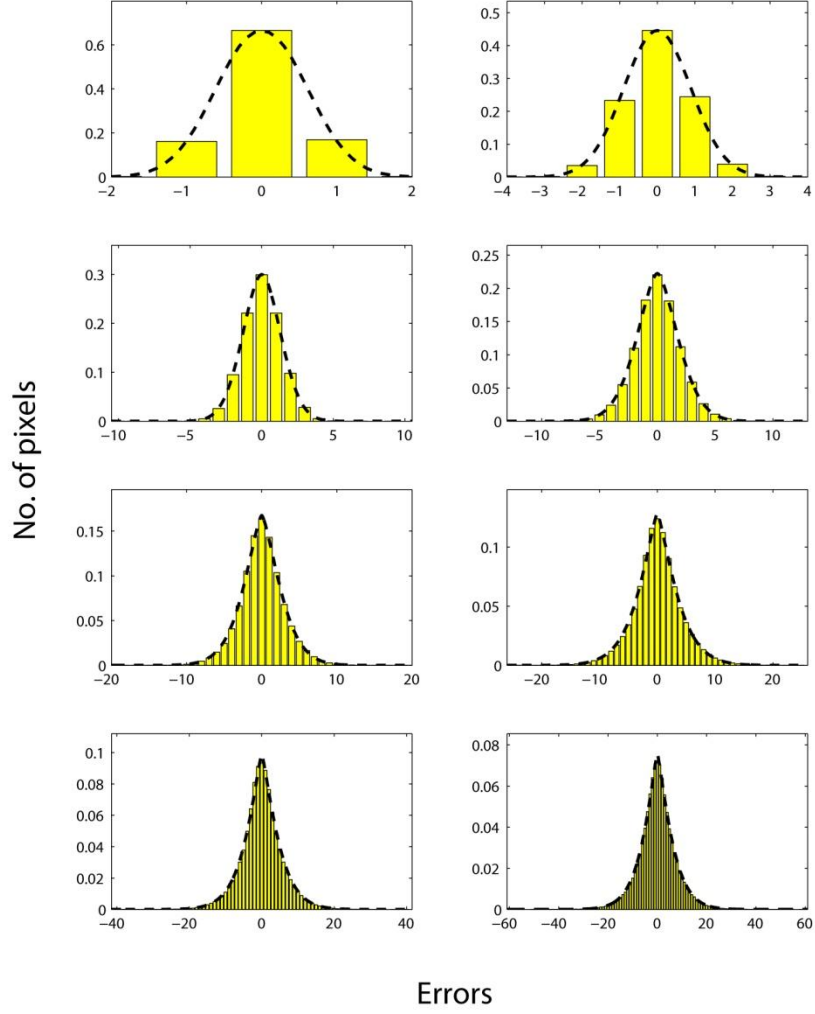


Fig. A.1.: Error distributions fitted with GGD distribution curves. Yellow bars denote the histogram of the compression error and dash black line is the fitted GGD curve.

We collected the estimated σ and α for a large number of images and for different QPs using H.264/AVC intra-frame mode. Fig. A.2 demonstrates the results of σ and α vs. the value of QP for different images. We found that σ and α can be well described by the functions of QP:

$$\sigma = a \cdot QP^2 - b \cdot QP + c \quad (\text{A.4})$$

$$\alpha = 1 + e^{(d \cdot QP + g)} \quad (\text{A.5})$$

where a , b , c , d and g are constants equal to 0.00625, 0.12457, 1.2859, -0.1 and 1.32, respectively.

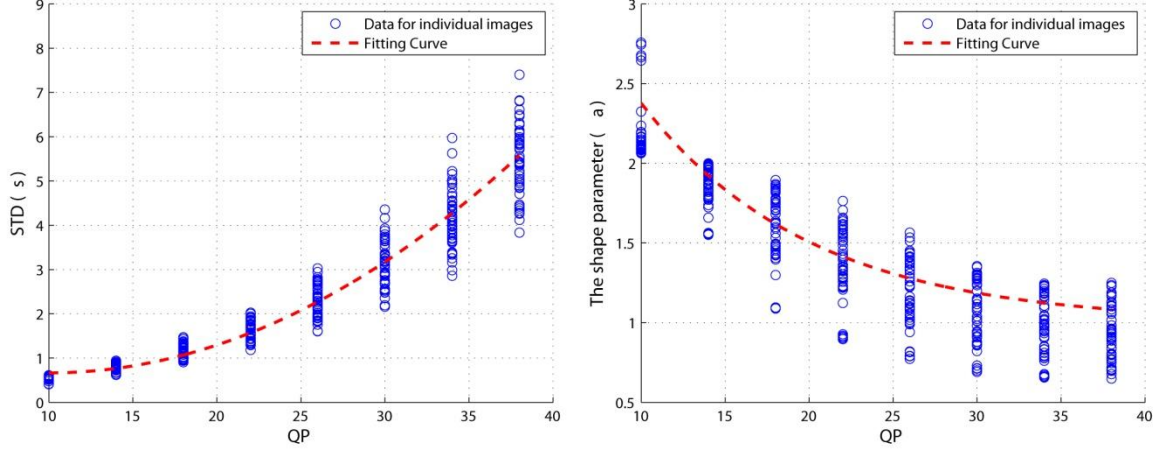


Fig. A.2.: The standard deviation and the shape parameter vs. the value of QP for H.264/AVC encoding over a large pool of images. The blue circle corresponds to a single image at a certain QP. The red line is the averaging curve used to remove the image dependency.

We also modeled the error distributions for H.264/AVC predicted frames (P frames), bi-directional predicted frames (B frames) and JPEG compression. We found that their compression errors can be well estimated by GGD too. Please refer to the supplementary documents (http://ece.ubc.ca/~zicongm/tmo_hdr_enc/HTMLReport.html) for details.