Backward-compatible high dynamic range imaging : acquisition, display and compression

Wei, Zhe

2018

Wei, Z. (2018). Backward-compatible high dynamic range imaging : acquisition, display and compression. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/75404>

<https://doi.org/10.32657/10356/75404>

BACKWARD-COMPATIBLE HIGH DYNAMIC RANGE IMAGING: ACQUISITION, DISPLAY AND COMPRESSION

WEI ZHE

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctorate of Philosophy

2018

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....

Date

.....

author

Acknowledgements

First and foremost, I would like to express my greatest gratitude to my supervisor, Prof. Wen Changyun, for his continuous support of my Ph.D study, for his patience, advise and immense knowledge. He taught me a lot on research problems formulating as well as academic writing and he always encouraged me to work independently. He is so kind a supervisor that he treats all his students like friends or even family members. I feel lucky to be one of Prof. Wen's group.

Beside my supervisor, I would like to thank Dr. Li Zhengguo for his valuable knowledge sharing, for his insightful comments and guidance on problem solving and for his tremendous help in polishing papers. He gave me a lot of suggestions on research directions as well as career planing. I will never forget his kindness to me.

I would also like to thank Huang Jiangshuai, Guo Fanghong, Guan Zheming, Ding Jie, Guan Mingyang, Zou Ying and Kou Fei, who are also Prof. Wen's students. We shared precious memories during group meetings and get-togethers. Especially, I express my great thanks to Kou Fei who helped me a lot on experiments. I thank my previous labmates Zhong Jingshan, Yu Hang, Qi Yongbin and Wang Mengmeng. They created a harmonious laboratory atmosphere and I had a wonderful time with them during the first two years of my Ph.D study. I also thank my current labmates

Zhang Shuai and Zhai Deqing. We had great times playing basketballs together. Especially, I would like to thank Zhai Deqing who always endures my complaints about research and cheered me up. I also thank my roommates Jiang Chaoyang and Zhu Qingchang for their accomplishment in the last three years. I appreciate that they are always ready to give me a hand.

Last but not the least, I would like to thank my parents for their unconditional love and their spiritual support throughout my whole Ph.D study. They always backed me up and encouraged me to pursue my dreams.

Abstract

High dynamic range imaging (HDR) techniques can provide excellent visual experience which is comparable to the human visual system, thus it is supposed to be the future of imaging techniques. However, because of relative slow pavement in development and deployment of HDR display technology, most consumers, who are only equipped with conventional low dynamic range (LDR) displays cannot benefit from the new imaging technology because the HDR images are not directly viewable on LDR displays. Backward-compatible solutions are needed so that LDR consumers can gradually accept and upgrade legacy LDR imaging techniques to HDR ones.

In this thesis, we provide three backward-compatible solutions in the *acquisition*, *display* and *compression* stages of the high dynamic range imaging pipeline. In the *acquisition* stage, we propose a detail-enhanced exposure fusion algorithm to directly acquire an HDR-like LDR image by merging a bracket of differently exposed LDR images. The acquired LDR image, which preserves information in brightest and darkest regions as well as the fine-details of the scene, can provide superior HDR-like visual experience without generating intermediate HDR images. In the *display* stage, we propose an efficient saliency-aware tone mapping operator to reproduce the HDR image on conventional LDR displays. Our method takes the overall brightness, contrast and details preservation and visual attention into consideration. We manage

to view HDR images on LDR displays without degrading its visual quality too much. In the *compression* stage, we propose a two-layer local inverse tone mapping algorithm to estimate an HDR image from its locally tone-mapped version. And we successfully adopt it in HDR image scalable coding to improve the coding efficiency for efficient backward-compatible distribution of HDR images.

In conclusion, our backward-compatible solutions enable LDR consumers to enjoy superior visual experience provided by HDR imaging techniques. Our work should be helpful for transition from LDR imaging techniques to HDR ones.

Table of contents

Acknowledgements	i
Abstract	iii
List of Abbreviations	xi
List of Figures	xiii
List of Tables	xix
1 Introduction	1
1.1 High Dynamic Range Imaging	1
1.2 Motivation and Objective	3
1.3 Our Contributions	4
1.4 Outline of the thesis	7
2 Background Knowledge and Literature Review	9

2.1	The High Dynamic Range Ecosystem	9
2.2	High Dynamic Range Image Acquisition	11
2.2.1	Radiance Map Recovery	12
2.2.2	Exposure Fusion	13
2.2.3	Assessment of Exposure Fusion	18
2.3	Tone Mapping	20
2.3.1	Global Tone Mapping Curves	20
2.3.2	Local Tone Mapping Operators	21
2.3.3	Assessment of Tone Mapping	24
2.4	High Dynamic Range Image Compression	25
2.4.1	Single-layer HDR Image Compression	25
2.4.2	Scalable HDR Image Compression	26
2.4.3	Assessment of HDR Image Compression	28
3	Detail-Enhanced Multi-Scale Exposure Fusion	31
3.1	Introduction	32
3.2	WGIF Based Multi-Scale Exposure Fusion	33
3.2.1	Pyramid Based Exposure Fusion	34
3.2.2	Weighted Guided Filtering	38
3.2.3	Discussion on WGIF-based Fusion	42

3.3	Detail-Enhanced Exposure Fusion	44
3.3.1	Problem Formulation	45
3.3.2	Vector Field Generation	46
3.3.3	Efficient Solution	52
3.4	Comparison among Exposure Fusion Algorithms	56
3.4.1	Experiment 1	56
3.4.2	Experiment 2	61
3.4.3	Experiment 3	68
3.4.4	Experiment 4	68
3.5	Conclusions	70
4	An Efficient Saliency-aware Local Tone Mapping Operator for High Dynamic Range Images	73
4.1	Introduction	74
4.2	Global Tone Mapping Curve	76
4.3	Saliency-aware Local Tone Mapping Operator	83
4.3.1	Saliency-aware Global Tone Mapping Curve	83
4.3.2	Saliency-aware Detail extraction	86
4.4	Experimental Results	92
4.4.1	Qualitative Comparisons	92

4.4.2	Quantitative Comparisons	97
4.4.3	User Study	101
4.4.4	Limitation	103
4.5	Conclusion	103
5	Local Inverse Tone Mapping for Scalable High Dynamic Range Im-	
	age Coding	107
5.1	Introduction	108
5.2	Two-layer Local Inverse Tone Mapping	111
5.2.1	Local Tone Mapping Recap	112
5.2.2	Generation of a Pair of Base Layers	113
5.2.3	Tone Curve Estimation	114
5.2.4	HDR Image Estimator	117
5.3	Experimental Results	119
5.3.1	Dataset and Evaluation Protocol	119
5.3.2	Parameter Selection	122
5.3.3	Validation of Inverse Tone Mapping	123
5.3.4	Lossy Scalable Coding	125
5.4	Conclusion	132
6	Conclusion and Future Work	133

6.1	Conclusion	133
6.2	Recommendation of Future Work	134
A	1D Solver	137
B	Subjective Evaluation Record	140
Author's Publications		147
Bibliography		149

List of Abbreviations

Abbreviation	Meaning
HDR	high dynamic range
LDR	low dynamic range
HVS	human visual system
CRF	camera response function
GF	guided image filter
BF	bilateral filter
WGIF	weighted guided image filter
WLS	weighted least square
FWLS	fast weighted least square
TMO	tone mapping operator
iTMO	inverse tone mapping operator
GTMC	global tone mapping curve
LTMO	local tone mapping operator

List of Figures

1.1	A complete high dynamic range imaging pipeline.	2
2.1	Illustration of the HDR ecosystem.	9
2.2	Illustration of the HDR radiance map recovery and exposure fusion. .	11
2.3	Illustration of different tone mapping algorithms. (a) simple linear scaling. (b) tone mapping using the algorithm in [1].	19
2.4	Flow of scalable HDR image compression.	26
3.1	The workflow of the proposed detail-enhanced multi-scale exposure fusion.	32
3.2	Fusion results using different Gaussian pyramid levels. (a) input images. (b) κ layers. (c) $\kappa - 1$ layers. (d) $\kappa - 2$ layers. (e) $\kappa - 3$ layers.	37
3.3	Comparison of Gaussian pyramid of the weight maps and their smoothed versions using the WGIF.	40

3.4	Fusion results using different WGIF refined pyramid levels. (a) κ layers. (b) $\kappa - 1$ layers. (c) $\kappa - 2$ layers. (d) $\kappa - 3$ layers.	41
3.5	Illustration of noise reduction by the WGIF.	43
3.6	Fusion results using different structure tensor. (a) and (c) Enhanced fusion results using the structure tensor in [2]. (b) and (d) Enhanced fusion results using the proposed tensor. It can be observed that the proposed tensor results in less noise.	49
3.7	Effect of the direction correction in Eq. (3.24). (a) Intermediate fusion result. (b) Enhanced exposure result without direction correction. (c) Enhanced exposure result with direction correction. (d)-(f) Zooms-ins of (a)-(c).	51
3.8	Effect of the number of iterations T . (a) $T = 1$. (b) $T = 3$. (c) $T = 15$. The brightness and contrast of detail layers are manually adjusted for better illustration. (d)-(f) zoom-ins of (a)-(c)	55
3.9	Comparison with the fusion algorithm in [3]. (a)-(c) input images. (d) our WGIF-based fusion result. (f) result by [3]. (e) and (g) zoom-ins of (d) and (f).	58
3.10	Comparison with the fusion algorithm in [4]. (a)-(c) input images. (d) our WGIF-based fusion result. (f) result by [4]. (e) and (g) zoom-ins of (d) and (f).	59
3.11	Comparison with the fusion algorithm in [5]. (a)-(c) input images. (d) our WGIF-based fusion result. (f) result by [5]. (e) and (g) zoom-ins of (d) and (f).	60

3.12	Comparison with the fusion algorithm in [6]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [6]. (c) and (e) zoom-ins of (b) and (d).	62
3.13	Comparison with the fusion algorithm in [7]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [7]. (c) and (e) zoom-ins of (b) and (d).	64
3.14	Comparison with the fusion algorithm in [8]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [8]. (c) and (e) zoom-ins of (b) and (d).	65
3.15	Comparison with the fusion algorithm in [9]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [9]. (c) and (e) zoom-ins of (b) and (d).	66
3.16	The BT scores of the seven fusion algorithms. The scores indicate the relative ranking where a higher score stands for a higher rank.	71
4.1	The tone mapping function f is modelled as a piecewise linear function from log HDR luminance l^h to log LDR luminance l^l	77
4.2	Comparison among the GTMCs in [1], [10] and ours. (a) GTMC in [1]. (b) GTMC in [10]. (c) Our GTMC in Eq. (4.5). (d) Histogram of the HDR image.	80
4.3	(a) without the third smoothness term in Eq. (4.5). (b) with the smoothness term.	81
4.4	Tone mapping curves for different γ in Eq. (4.5).	82

4.5	Comparison among different saliency-aware cues in our GTMC refinement. (a) Our proposed GTMC in Eq. (4.5). (b) Using the saliency histogram only. (c) Using the salient-aware brightness estimation only. (d) Using both saliency histogram and the salient-aware brightness estimation.	84
4.6	Illustration of better detail reproduction by the proposed saliency-aware filter. (a) Our saliency-aware GTMC only. (b) Saliency map. (c) Our saliency-aware GTMC plus the edge-preserving filter in [11]. (d) Our saliency-aware GTMC and saliency-aware edge-preserving filter.	89
4.7	One more example for illustration of better detail reproduction by the proposed saliency-aware filter. The captions are same as figure 4.6. .	90
4.8	Comparisons among our LTMO and tone mapping algorithms in [1] and [10]. From first row to last row are the tone mapping results using the GTMC in [1], our LTMO and the LTMO in [10].	94
4.9	Comparisons among our LTMO and tone mapping algorithms in [1] and [10]. From first row to last row are the tone mapping results using the GTMC in [1], our LTMO and the LTMO in [10].	95
4.10	Comparison of other TMOs and the proposed method. (a) bilTMO [12]. (b) disTMO [13]. (c) perTMO [14]. (d) the proposed LTMO. . .	96
4.11	Comparison of other TMOs and the proposed method. (a) bilTMO [12]. (b) disTMO [13]. (c) perTMO [14]. (d) the proposed LTMO. . .	98
4.12	Saliency maps of Figs. 4.8 - 4.11.	99

4.13	The BT scores of the six TMOs. The scores indicate the relative ranking where a higher score stands for a higher rank.	102
4.14	One failure case in which the saliency detection model [15] fail to produce reliable salience map. (a) Tone mapping with the saliency cue. (b) Tone mapping without the saliency cue. (c) Saliency map. (d) and (e) are zoom-ins of (a) and (b).	104
5.1	Workflow of scalable coding for HDR images.	109
5.2	Workflow of iTMO estimation in encoder side.	114
5.3	Workflow of HDR image estimation via the proposed two-layer iTMO.	118
5.4	The 20 HDR images used in our experiments. The dataset is downloaded from [16].	120
5.5	Average HDR-VDP-Q scores over 10 HDR images in the parameter study with varying μ in Eq. 5.6.	123
5.6	Average HDR-VDP-Q scores over 10 HDR images in the parameter study with varying λ in Eq. 5.3.	124
5.7	Inverse tone mapping results for images tone-mapped using the global TMO in [1]. It is obvious that perceptually better HDR estimations are obtained with the proposed global iTMO estimation method.	125
5.8	Inverse tone mapping results for images tone-mapped using the local TMO in [17]. The average HDR-VDP-Q scores of the linear expansion, local iTMO estimation without decomposition and with decomposition are 61.05, 60.01 and 62.86, respectively.	126

5.9	Mean compression performance of using the proposed global iTMO. All images are tone-mapped with the global TMO in [1]. The quality parameter for the base layer is fixed as 90.	128
5.10	Mean compression performance of using the proposed global iTMO. All images are tone-mapped with the global TMO in [1]. The quality parameter for the base layer is fixed as 65.	128
5.11	Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [17]. The quality parameter for the base layer is fixed as 90.	129
5.12	Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [17]. The quality parameter for the base layer is fixed as 65.	129
5.13	Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [14]. The quality parameter for the base layer is fixed as 90.	130
5.14	Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [14]. The quality parameter for the base layer is fixed as 65.	130
5.15	Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [14]. The quality parameter for the base layer is fixed as 30.	131

List of Tables

3.1	Average MEF-SSIM scores of 10 test image sets of using different pyra- mid levels	42
3.2	MEF-SSIM scores of different fusion algorithms.	69
3.3	Comparison of computation time (in second).	69
3.4	Preference matrix.	70
4.1	Average TMQI scores [18] over 30 images used in our experiment . .	100
4.2	Average FSITM scores [19] over 30 images used in our experiment . .	101
4.3	Preference matrix of the six TMOs.	101
5.1	Statistics of the 20 HDR images. The minimal, maximum luminance and the dynamic range are in $\log_{10}$ units.	121
5.2	Choices of μ and λ in the parameter study.	122
B.1	Preference matrix of subject 1.	140
B.2	Preference matrix of subject 2.	140

B.3	Preference matrix of subject 3.	141
B.4	Preference matrix of subject 4.	141
B.5	Preference matrix of subject 5.	141
B.6	Preference matrix of subject 6.	142
B.7	Preference matrix of subject 7.	142
B.8	Preference matrix of subject 8.	142
B.9	Preference matrix of subject 9.	143
B.10	Preference matrix of subject 10.	143
B.11	Preference matrix of subject 11.	143
B.12	Preference matrix of subject 12.	144
B.13	Preference matrix of subject 13.	144
B.14	Preference matrix of subject 14.	144
B.15	Preference matrix of subject 15.	145

Chapter 1

Introduction

1.1 High Dynamic Range Imaging

The dynamic range of physical scenes, which is defined as the ratio of the largest and smallest luminance, can be quite high. For instance, in a typical night scene, the ambient luminance of starlight is about 10^{-3} cd/m^2 while that of fluorescent lamps is about 10^3 cd/m^2 . The dynamic range of a night scene can be as high as six orders of magnitude. Conventional digital images which encode the display-referred luma using 8-bits can only represent a dynamic range of about two orders of magnitude, which is why they are called low dynamic range (LDR) images. For scenes of high dynamic range (HDR), conventional LDR cameras cannot capture the brightest and darkest regions of the scenes simultaneously. This results in over/under-saturation in these extremely bright/dark regions. However, our human visual system (HVS) has a dynamic range of about five orders of magnitude and has no problem in perceiving the details in the brightest and darkest regions of the HDR scenes. The limited visual experience provided by conventional LDR images is much worse than the HVS perceived. To address this problem, high dynamic range imaging techniques have been proposed. Unlike the LDR image which stores luma that are display-referred, the HDR image representation corresponds to the radiance of the physical scene and

can encode the HDR scene in a single image.

Figure 1.1: A complete high dynamic range imaging pipeline.

HDR imaging is an evolutionary technique and it changes the conventional LDR imaging pipeline in all stages. A complete HDR imaging pipeline is shown in Fig. 1.1. In the image acquisition stage, the scene-referred HDR image can be captured by hardware solutions [20, 21], which utilize CCD sensors that can record 10/12-bit value in logarithmic domain, or software solutions [22, 23], which reconstruct the HDR image from multiple differently exposed images. In the display stage, the LDR images are naively displayed on conventional LCD/LED monitors. However, it is not applicable to HDR images because of the limited dynamic range of those LDR displays. New types of monitors [24, 25], which are designed for HDR images, are prototyped to support the dynamic range up to five orders of magnitude. With these new HDR displays, HDR images can be directly viewed. Because of the scene-referred representation, the HDR image data is naively recorded using float numbers which results in much higher bit-depth than the LDR luma. To efficiently store and transmit HDR images, new image formats [26–28] and compression standards [29, 30] have been proposed. These new formats and compression standards enable

efficient distributions of HDR contents. With all the HDR imaging techniques in the acquisition, display and compression stages, HDR images provide excellent viewing experience and are more preferred than the LDR ones.

1.2 Motivation and Objective

Despite HDR techniques provide superior viewing experience, HDR images are not immediately accepted because of the relatively slow pavement in HDR display technologies. Though the first HDR monitor was prototyped ten years ago, HDR monitors are still not widely deployed currently because of their high prices. The majority of consumers still rely on LDR displays and a huge investment is needed to upgrade all the LDR equipments to HDR ones, which cannot be completely accomplished in recent years. Therefore we will live with both HDR and LDR images in a quite long term. In this transition phase, an important issue is how to enable LDR consumers, which are equipped with LDR displays only, to benefit from the HDR imaging techniques so that the new HDR imaging techniques will be gradually accepted and the consumers are willing to update legacy LDR imaging techniques to HDR ones. Based on this backward-compatible consideration, three research problems arise:

- *Acquisition:* In the HDR imaging pipeline, an HDR image, which represents the radiance of the scene, is reconstructed in the acquisition stage. However, this HDR image is not applicable for viewing on LDR displays. Can we directly synthesize an LDR image which preserves both details in brightest and darkest regions of the scene like the HDR image to provide HDR-like visual experience on LDR displays without generating the intermediate HDR image?

- *Display:* For consumers who are only equipped with LDR displays, is it possible to reproduce the already acquired HDR image on the LDR displays without degrading the visual experience too much?
- *Compression:* Though new file formats and compression standards can enable efficient transmission of HDR images. The decoded HDR images will also not be directly viewable if the receivers are only equipped with LDR displays. Can we provide an efficient compression algorithm, e.g. a scalable coding algorithm, so that the compression is compatible to both HDR and LDR consumers?

1.3 Our Contributions

Based on above three research problems, we propose three solutions, one for each question, in this thesis. The contributions of this thesis can be summarized as follows:

- To solve the first backward-compatible HDR image acquisition problem, we propose an exposure fusion algorithm to merge the ‘best’ parts of the differently-exposed inputs to directly form an HDR-like LDR image. Our exposure fusion algorithm mainly consists of two modules: multi-scale exposure fusion and detail extraction. In the multi-scale fusion module, we propose to refine the fusion weight pyramid using an edge-preserving filter. With the smoothed weight pyramid, the fusion result can keep contrast in both brightest and darkest regions of the scene without introducing halo artifacts. In our detail extraction module, we propose an exposure-aware weighted structure tensor to compute the vector field of the inputs. This vector field is then used to extract a detail layer to further improve the visual quality of the intermediate multi-scale fusion result.

Note that our weighted structure tensor results in less noise and less amplification of unwanted gradients in the extracted detail layer. In addition, we propose an efficient solution of the detail extraction problem. Our solver is about four times faster than the original one. We validate the effectiveness of the two modules using substantial qualitative and quantitative comparisons. The experimental results prove that the proposed method produces fusion results with higher visual quality than the state-of-the-arts.

- For the second problem on backward-compatibly displaying HDR images, we propose a saliency-aware tone mapping algorithm to reduce the dynamic range of the HDR image in order to view it on LDR displays. Our tone mapping algorithm is developed under the base-detail separation tone mapping framework. In this framework, the global tone mapping algorithm for the base layer controls the brightness and contrast of the tone mapping results while the detail layer controls the extent of detail preservation. In our tone mapping algorithm, we first formulate a novel global tone mapping problem for dynamic range compression of the base layer. Our formulation takes both the brightness and contrast loss into consideration and produces perceptual bright tone mapping results with minimal contrast loss. We also propose to introduce the saliency cue to further improve the perceptual quality of the tone mapping results. We refine our global tone mapping algorithm by exploiting the saliency histogram instead of the intensity histogram and manage to reduce the contrast loss in salient regions. In addition to the dynamic range compression for base layer, we also introduce the saliency cue into the detail layer extraction process. We incorporate a saliency-aware weight into the edge-preserving filter and manage to selectively preserve more details in salient regions. With the new tone map-

ping formulation and saliency cues, our algorithm enables LDR consumers to view HDR images on LDR displays without degrading the visual experience too much.

- To address the issue of backward-compatible HDR image compression, we propose a local inverse tone mapping algorithm to estimate an HDR image from its tone-mapped version. This estimation is applied in scalable HDR image compression to improve the coding efficiency. Unlike previous works on local inverse tone mapping, we propose a two-layer inverse tone mapping approach. In our method, the base layer of the HDR image is estimated by expanding the base layer of its tone-mapped LDR image while the detail layer of the HDR image is estimated directly using the detail layer of its tone-mapped version. To find the proper inverse tone mapping function for dynamic range expansion of the LDR base layer, we formulate an optimization problem to estimate the inverse tone mapping function using the pair of HDR and LDR base layers. This function is then used to estimate the raw HDR image when only the tone-mapped LDR image is given. We verify that our estimated inverse tone mapping function can accurately predict the HDR image from the tone-mapped images even the closed-form inverse of the tone mapping algorithm does not exist. We also apply the proposed two-layer inverse tone mapping algorithm into the JPEG-XT compression standard, which is a scalable coding standard for HDR image. We show that the proposed method can improve the coding efficiency by testing two challenging local tone mapping operators.

1.4 Outline of the thesis

This thesis focuses on addressing above mentioned three research problems. The remaining parts of this thesis are outlined as follows:

- Chapter 2 introduces the main steps of the HDR imaging pipeline, including the HDR ecosystem, reconstruction of HDR image, synthesis of HDR-like image by merging a stack of differently exposed image, visualization of HDR image on conventional LDR displays by dynamic range reduction using tone mapping, new file formats and compression standards for HDR images and the scalable compression algorithms.
- Chapter 3 introduces our exposure fusion technique to directly synthesizing an HDR-like LDR image. The details of the multi-scale fusion and detail extraction modules are explained sequentially, followed by the qualitative and quantitative comparisons with state-of-the-arts and the conclusions for this chapter.
- Chapter 4 introduces our local tone mapping algorithm for visualization of HDR images on LDR displays. Our formulation for the dynamic range compression of base layer will be first introduced. Then we explain how we incorporate the saliency cue into base layer dynamic range compression and detail layer extraction. Finally, objective and subjective evaluations are presented and conclusions are given.
- Chapter 5 introduces a two-layer inverse local tone mapping algorithm for efficient compression of HDR images in the scalable HDR coding framework. We start from the inverse tone mapping function estimation using the pair of HDR and LDR base layers. Then we explain how to estimate the HDR image using

the estimated inverse tone mapping function. Afterwards, we validate the effectiveness of the proposed method by experiments without involvement of coding algorithms and also by integrating it into the JPEG-XT compression standard.

- Chapter 6 concludes the thesis and provides suggestions for future research.

Chapter 2

Background Knowledge and Literature Review

2.1 The High Dynamic Range Ecosystem

Due to the superior ability in representing and reproducing the real HDR scene, the HDR imaging technique has attracted great attention from the consumer electronics industry. In last few years, camera and television manufacturers have been actively involved in developing HDR products and standards.

Figure 2.1: Illustration of the HDR ecosystem.

The complete HDR ecosystem is illustrated in Fig. 2.1. For easy capture of HDR contents, a number of modern consumer-level cameras, such as Cannon 650D, Nikon 7100 and Sony FDR-AX700, provide built-in HDR mode which enables recording of HDR images and videos in one single click. For vivid reproduction of HDR scenes using the HDR signals, new OLED televisions, which have higher luminance range and more gray levels than conventional CRT ones, have been developed. Thanks to the HDR-capable OLED televisions, the HDR contents, which are usually represented in 10-16 bits per channel, are displayed without much loss of contrast in both brightest and darkest regions. In addition to the enhanced contrast, the higher bit-depth representation is also capable of representing more lifelike colors than the conventional sRGB color space [31]. This extended color space is defined as the Wide Color Gamut (WCG) feature in ITU-R BT.2020 [32]. The HDR, WCG, together with a higher resolution than 4K, are the three key features of next generation UHDTV as recommended in [32]. For efficient production and delivery of HDR contents, a number of HDR formats, such as PQ, HDR-10, HDR-10+ and HLG [32–34], have been proposed. These formats define an Opto-Electronic Transfer Function (OETF), which transfers the HDR signals into digital code words, and a corresponding Electro-Optical Transfer Function (EOTF), which turns the encoded digital words into visible light. In the conventional LDR processing pipeline, the gamma function is applied to encode and decode luminance of LDR contents while the OETF and EOTF are used for HDR contents in the HDR processing pipeline. In addition, new compression standards, such as JPEG 2000 [30] and JPEG XR [29], have been proposed by professionals to reduce the bandwidth needed for transmission of HDR contents.

With all the HDR related techniques discussed above, consumers will be able to have a superior viewing experience in the near future. However, HDR-capable displays are

not widely deployed currently because of their high prices. The majority of consumers still rely on LDR displays. It is desired to enable LDR consumers to benefit from the HDR imaging techniques for promotion of the new techniques. Therefore, backward-compatible solutions are needed in this transition stage from LDR to HDR, which inspires the research topics covered in this thesis. In the remaining sections of this chapter, we will discuss the backward-compatible solutions in the acquisition, display and compression stages of the HDR imaging pipeline.

Figure 2.2: Illustration of the HDR radiance map recovery and exposure fusion.

2.2 High Dynamic Range Image Acquisition

Because of extremely high price of HDR hardware [20, 21], most HDR images are captured by taking a bracket of differently-exposed LDR instead of directly using the

HDR sensors. As shown in Fig. 2.2, if the radiance map, which roughly represents the radiance of the HDR scene, is recovered from multiple exposures, it can be displayed on HDR monitors to provide superior visual experience. However, for LDR consumers, which account for the majority of the display market, the radiance map cannot be directly viewed because of its higher dynamic range. Exposure fusion techniques are desired to synthesize a LDR image from the multi-exposure inputs to provide an HDR-like visual experience on LDR displays. In the rest of this section, the radiance map recovery and exposure fusion techniques will be introduced.

2.2.1 Radiance Map Recovery

Denote the radiance map and the camera response function (CRF) as E and f , respectively. If the exposure time for the i -th image is Δt_i , we have [22]

$$Z_i = f(E \cdot \Delta t_i) \quad (2.1)$$

where Z_i is the intensity of the LDR image. If the inverse of the CRF f^{-1} is known, the radiance map E can be easily recovered from Eq. (2.1). Note that because of the clipping in LDR images, a single exposure may only cover a small range of the radiance map. Thus the final radiance map is usually computed using multiple exposure LDR images in a weighted average manner as

$$E(p) = \frac{\sum_{i=1}^N \frac{1}{\Delta t_i} w(Z_i(p)) f^{-1}(Z_i(p))}{\sum_{i=1}^N w(Z_i(p))} \quad (2.2)$$

where p is the pixel index, N is the total number of exposures and $w(\cdot)$ is the blending weight. Note that the CRF is a combination of several non-linear functions. It is

usually not known in advance. To estimate the CRF, many methods have been proposed [22, 23].

Assuming the CRF to be a monotonically increasing function, Debevec et al. [22] formulate the CRF as a least square error minimization problem in logarithmic domain. By assuming the CRF to be a polynomial function, Mitsunaga et al. obtain a more robust CRF estimation in [23]. With these efficient solutions for CRF estimation, the HDR radiance map can be easily recovered. Nowadays, more and more camera manufacturers provide built-in radiance map recovery function which makes the number of HDR images increases exponentially. However, these HDR images cannot be directly viewed on LDR displays because of their higher dynamic ranges. Additional processing such as tone mapping techniques discussed in latter sections are required for visualization of HDR images on LDR displays.

2.2.2 Exposure Fusion

Because the radiance map cannot be directly viewed on LDR displays, exposure fusion techniques are demanded to synthesize an HDR-like LDR image for LDR consumers. The idea of exposure fusion is combining the ‘best’ parts in each exposure so that the details in both the brightest and darkest regions of the scene can be preserved in a single LDR image. Till now, there is a huge number of exposure fusion algorithms. In this section, we only introduce some representative exposure fusion algorithms by categorizing them into following four groups. For a more complete review, please refer to [35, 36].

Single-scale Exposure Fusion

An intuitive idea for exposure fusion is averaging all differently-exposed inputs using a uniform weight. However, this simple approach is unlikely to produce satisfactory result because a region may be ‘good’ in some exposures while ‘bad’ in the others. An more appropriate approach should be blending the inputs according to some quality measures which indicate the ‘good’ and ‘bad’ regions. Based on this idea, the exposure fusion algorithms in [4, 5, 37–39] have been proposed.

Authors in [4, 37] propose to blend all the inputs pixel-wisely. Based on the idea that well-exposed pixels should contribute more to the fusion result, Vonikakis et al. [4] estimate the illumination of each exposure by the Retinex theory [40] and propose to assign the blending weights according to the estimated illumination. In their work, different weighting functions are used for different exposures because the criteria of well-exposed pixels varies from shortest to longest exposures. Raman et al. [37] employ the bilateral filter to decompose each exposure into a base layer and a detail layer. The blending weight is defined as the magnitude of the detail layer. The fusion schemes [4, 37] are simple and easy to implement. However, the single-scale fusion strategy may produce unsatisfactory fusion results in regions where the blending weights changes rapidly in different exposures as discussed in [3].

In addition to the pixel-wise blending methods, authors in [5, 38, 39] propose to fuse all the inputs block-wisely. In [38], each input image is partitioned into uniform blocks. The blocks which maximize the entropy are selected for composition of the final fusion result. Aiming to maximize the detail information remained in the fusion result, authors in [39] propose to determine the block-wise blending weight by the sum of the gradients in each block. The algorithm in [5] treats pixels in each block as

a vector and represents the vector using three components which are the mean of the vector elements, magnitude of the vector and the normalized unit vector. The three components are blended and the final fused blocks are reconstructed from the three blended components. Note that the block-wise fusion may result in inconsistency where the transition between blocks is sharp, an additional smoothing step is usually required to post-process the block-wise composition.

Multi-scale Exposure Fusion

To better preserve details in the final fusion result, authors in [8] propose to decompose each input image into a base layer and a detail layer. After separately blending the two layers, the fusion result is obtained by combining the blended two layers together. For simplicity, the two-layer decomposition in [8] is achieved by Gaussian smoothing and the weight map for each exposure is computed using the saliency measure only. To avoid inconsistency along edges in the fusion result, authors in [8] propose to refine the weight map using the guided image filter (GF) [41]. Because of the efficiency of the Gaussian smoothing and GF, the algorithm in [8] is very fast. However, only two layers are used in the fusion process. Similar to the single-scale fusion algorithms, inconsistency may still exist in the final fusion result.

For smooth blending of multiple exposures, authors in [3, 42] propose a multi-scale pyramid based fusion strategy. In their work, the input images and the weight maps are represented using Laplacian and Gaussian pyramids [43], respectively. Like the two-layer fusion in [8], the blending in [3, 42] is conducted layer-wisely. And the final fusion result is reconstructed from the fused pyramid. Similarly, Shen et al. also propose a multi-scale fusion algorithm in [44] using the same pyramid representation. Different from methods in [3, 42], Shen et al. propose a new weighting function which

is based on the hybrid contrast and just noticeable saliency measure. Except for the novel weighting, Shen et al. also propose to boost the detail layers for improved sharpness which is similar to the work in [45].

Unlike above Gaussian smoothing based pyramids, Fattal et al. [46] propose a fusion scheme using a bilateral filter (BF) [47] based pyramid. By iteratively smoothing the input image using the BF with gradually varying parameters, Fattal et al. manage to decompose the input images into multiple scales. Because the BF has a better ability in preserving edges than the Gaussian filter, the fusion results of [46] are unlikely to produce halos. However, the naive implementation of the BF is much slower than Gaussian filter. Therefore, the speed of [46] is slower than [3, 42, 44].

In addition to above mentioned spatial domain fusion schemes, some algorithms [48–51] have also been proposed for multi-scale fusion in frequency domain. In these methods, a discrete wavelet transform is first applied on each input image to transform it into frequency domain. After the transform, the coefficients of each subband instead of the pixel values are averaged. The final fusion result is obtained by inverse discrete wavelet transform using the averaged coefficients. The weighting functions used in [48, 49] are a little heuristic. In [50], the authors propose to assign the weights according to the locally just noticeable difference threshold which leads to a higher perceptual quality. In [51], the authors propose a cross-scale fusion rule which takes both intrascale and interscale consistencies into consideration to generate better fusion results.

Gradient Domain Exposure Fusion

Considering that gradients usually indicate the structural information of the image, authors in [2, 52–54] propose their fusion algorithms in gradient domain. In this kind of methods, a desired vector field is first computed from the multiple inputs. The final fusion result is reconstructed from the vector field by solving a Poisson equation which is quite time-consuming when the resolution of the image is high.

In [2], a structure tensor is first computed using the gradient matrix of the multiple inputs. Then the desired vector field is obtained by singular value decomposition of the tensor. Based on the observation that directly reconstructing the fusion result from the vector field computed in [2] may enhance invisible gradients much and cause exaggerated over-enhancement, authors in [53] propose a scaling function, which adapts to the magnitude of the gradients, to adjust the vector field. By introducing a saliency-aware weight, Wang et al. [52] propose a saliency-aware structure tensor and achieve better structure preservation in salient regions with the weighted tensor. Different from the structure tensor based methods, Hara et al. [54] propose to compute the desired vector field by finding the optimal blending weights of the multiple inputs that maximize the amount of gradient magnitude. Because of the extra optimization problem, the algorithm in [54] is much slower than those in [2, 52, 53].

Fusion by Probabilistic Inferring

Authors in [6, 7, 55] propose to solve the exposure fusion problem using probabilistic models. In [6], Shen et al. measure the quality of the desired fused result by local contrast and color consistency and find the optimal blending weights that maximize the two quality measures by solving the inferring problem under a generalized random

walks framework. In their later work in [7], Shen et al. propose two novel perceptual measures which are the perceived local contrast and color saturation. With the new quality measures and a hierarchical multivariate Gaussian conditional random field model, Shen et al. manage to obtain perceptual better fusion results in [7]. Taking the Weber's law into consideration, Song et al. [55] formulate a maximum-a-posterior (MAP) problem to obtain the optimal fusion result which both maximizes the visible contrasts and embeds the gradients of the scene. It is shown in [55] that the MAP problem can be easily simplified to a variational fusion problem which is similar to the works in [56, 57]. Note that the probabilistic fusion algorithms usually result in a large-scale linear system. The speed of these methods could be an issue, especially for images with very high resolution.

2.2.3 Assessment of Exposure Fusion

Because the HVS is the ultimate receiver of exposure fusion applications, it is straightforward to evaluate different exposure fusion algorithms subjectively. Though the subjective evaluation gives insights on improving existing fusion algorithms, it is quite expensive and time consuming. Objective image quality assessment models are desired for efficient evaluation and comparison of exposure fusion algorithms.

It is very challenging to find an objective metric that perfectly matches the subjective evaluation result. In order to develop an objective assessment model on which most subjects will agree, a lot of work [58–62] have been done to design objective assessment models that correlate well with the subjective ground truth. In [58], the authors propose to measure the quality of fusion results using differences in frequency domain. The fused image and the corresponding sources are divided into local regions and the

differences in frequency domain are computed for each region. Finally, the overall quality of the fused image is pooled as a weighted summation of the quality measures for each local region. In [59], the authors compute the relationship between the contrast maps of the fused image and sources and the overall quality map is pooled by weighted summation. In [60], a new ratio of spatial frequency error is derived and applied to measure the quality of fusion results. In [61], the authors conduct a subjective evaluation and model an objective metric based on multi-scale structure similarity index (SSIM) [63]. They have compared their proposed metric with the other state-of-the-arts in [61]. The result turns out that their metric correlates well with the collected subjective ground truth which makes their metric suitable for evaluation of exposure fusion algorithms. For readers who are interested in objective fusion assessment models, please refer to the comparative study in [62].

Figure 2.3: Illustration of different tone mapping algorithms. (a) simple linear scaling. (b) tone mapping using the algorithm in [1].

2.3 Tone Mapping

Due to the discrepancy between the dynamic range of HDR images and LDR displays, the HDR images cannot be directly viewed on conventional LDR displays. Simple mapping operations such as linear scaling may result in poor visual experience as shown in Fig. 2.3(a). Tone mapping algorithms such as the one shown in Fig. 2.3(b) are desired for visualization of HDR images on LDR displaying without degrading the HDR visual experience much. In the last twenty five years, a huge number of tone mapping algorithms have been proposed. In this section, we only review and explain the most closely related TMOs by classifying them into global tone mapping curves (GTMCs) and local tone mapping operators (LTMOs) as below. For a more complete review of tone mapping algorithms, please refer to [64, 65].

2.3.1 Global Tone Mapping Curves

The GTMCs adjust the HDR input using a spatially-invariant curve. Based on how the mapping curve is derived, we category them into following two groups.

“S” Shape Tone Mapping Curves

Inspired by the sigmoidal photoreceptor response function of the human visual system (HVS), GTMCs in [1, 66] are proposed. Though the curves are defined in different forms of functions, they share the same “S”-like shape. For these “S”-shape mapping curves, the middle tone of the HDR image is usually well-preserved. However, the GTMCs gradually saturate in high and low tones. Thus they face problems in preserving contrast in the near-saturated high and low tone regions because of the heavy

dynamic range compression.

Histogram Based Tone Mapping Curves

Different from the pre-defined GTMCs in [1, 66], authors in [10, 13, 67] propose to define their GTMCs according to the histogram distribution of the HDR image. By modeling the GTMC as a piecewise linear function, Gregory et al. [67] propose to define the tone mapping curve by histogram adjustment with a ceiling constrain. Based on the consideration that the optimal GTMC should minimize the contrast loss caused by the dynamic range compression, Mantiuk et al. [13] formulate a novel contrast loss minimization problem using the HVS model in [68, 69]. Lately, they simplify the optimization problem in [13] and propose a more efficient solution in [10].

The three GTMCs [10, 13, 67] allocate the dynamic range adaptively and they may preserve contrast better than the GTMCs in [1, 70] because of the adaptive strategy. However, the GTMCs do not explicitly take the brightness into consideration. In some cases it cannot produce images as bright as those GTMCs in [1, 70] because the histogram-based GTMCs may cause brightness shift.

2.3.2 Local Tone Mapping Operators

Unlike GTMCs, LTMOs use spatially-varying mapping functions, which depend on the local features around the pixel, to adjust the HDR input. In this section, we briefly review the LTMOs in [1, 10, 12, 14, 71–74] and category them into following three groups.

Local Adaptation

Based on the observation that the adaptation level of HVS can be approximately determined by the background luminance level, Chiu et. al [71] propose to design a spatially nonuniform scaling function for the tone mapping purpose. In their work, the background luminance level is estimated using the Gaussian smoothed version of the input HDR image and then adopted to scale down the HDR luminance. As discussed in [71], the user-specific variance of the Gaussian smoothing kernel should be carefully selected. Otherwise, severe halos will be produced. Following the same local adaptation idea, authors in [1] propose an automatic dodging-and-burning approach to refine the intermediate LDR image produced by their GTMC in [1]. By exploiting the difference-of-Gaussian to find the largest local support window which does not contain sharp edges, each pixel is mapped according to the average of the local window. The LTMO in [1] can improve the local contrast of their GTMC version. However, the spatially-varying mapping may cause inconsistency among adjacent pixels. In addition, this method has a limited ability in preserving fine-scale details.

Base-detail Separation Based Tone Mapping

To achieve better details preservation, Durand et al. [12] first proposes to address the tone mapping problem using an edge-preserving base-detail separation framework. In their work, an HDR image is decomposed into a base layer, which contains the dynamic range information, and a detail layer, which contains fine-scale details, using the bilateral filter (BF). By compressing the dynamic range of the base layer using the GTMC in [75] and adding the detail layer back, Durand et al. manage to obtain tone mapping results with fine details. Note that though Durand et al. select the

GTMC in [75] for base layer dynamic range compression, it can be replaced by other GTMCs such as [10, 72, 73] to produce more visually appealing results.

One main problem of the base-detail separation based LTMO in [12] is that the BF has a limited ability in preserving edges and it may result in halos in some cases as discussed in [41]. Aiming to develop filters that better preserve edges, many works on edge-preserving filters [17, 41, 76] have been proposed. By assuming an affine relation between the guidance image, which is usually selected as the original image when applying the filter [41] for smoothing purpose, and the smoothed result, He et al. [41] propose an efficient alternative, namely the guided filter (GF), to the BF. However, as illustrated in [41], the GF may still result in halos because the smoothing is conducted within local support windows. By taking all pixels into consideration and formulating the edge-preserving smoothing as a total variation problem, the filters in [17, 76] can obtain halo-free tone mapping results. However, solving the optimization problems in these two methods can be time-consuming especially for images with high resolutions. Recent work in [11] approximates the optimization problem in [17] using a sequence of low dimensional optimization problems which makes it possible for efficient tone mapping using the global optimization based filter in [17].

Gradient Domain Tone Mapping

Unlike above methods which works in spatial domain, authors in [74] propose to reduce the dynamic range by manipulating the gradients of the HDR image. In their work, the gradients for each pixel are adjusted according to a scaling function which attenuates gradients of large magnitude while keeps gradients of small magnitude. With the modified gradients, the tone mapping result can be reconstructed by solving a Poisson equation. In [74], the gradients are computed using the logarithm of

the HDR luminance. By replacing the logarithmic operator with a visible contrast model, Mantiuk et al. propose a similar LTMO in [14] for tone mapping in contrast domain. Because small scale gradients are kept, these two methods preserve the details well. However, they do not explicitly take the brightness into consideration and may produce artistic results with reduced global contrast and exaggerated details.

2.3.3 Assessment of Tone Mapping

Similar to the assessment of exposure fusion, subjective evaluations, such as previous works in [77, 78], are reliable but time-consuming approaches for comparison of different tone mapping algorithms. Note that the tone-mapped image and the HDR image have different dynamic ranges, conventional image quality assessment approaches, such as PSNR and SSIM, which assume that the reference and test images have the same dynamic range, may not be suitable for assessment of tone mapping results.

While there is a huge number of metrics for assessment of images with the same dynamic range, the number of metrics which can deal with different dynamic ranges is very small. The metric in [79] takes the HVS model into consideration and transforms the HDR and tone-mapped LDR images into contrast domain. It produces three distortion maps which indicate the loss of visible contrast, amplification of invisible contrast and reversal of visible contrast, respectively. The metric only gives distortion maps instead of an overall quality score. Sometimes, it is hard to tell which result is better without a certain numerical score. The TMQI metric in [18] gives a single numerical score for each test image. Taking structural fidelity and naturalness into consideration, authors in [18] score each test image using a combination of these two kinds of measures. The authors conduct a subjective study using 5 well-known

TMOs and learned the combination weights from the collected subjective data. They have also cross-validated the learned model on other subjective datasets and find the model correlates well with all the subjective data. The metric in [19] also gives a numerical score for each test image. It combines the TMQI and a locally weighted mean phase angle based feature similarity index. From the experimental results in [19], the combined metric gives a more accurate quality assessment than TMQI.

2.4 High Dynamic Range Image Compression

Because the HDR images represent a much higher dynamic range than the LDR ones, a higher bit-depth is required for HDR images. Usually, the HDR values are stored using float numbers. If uncompressed, this will result in about four times of the file size of the LDR ones which are encoded using 8-bit integers. New compression schemes for HDR images are required for efficient storage and transmission of the HDR content. In addition to the new compression algorithms for the higher bit-depth data, the backward-compatibility to conventional 8-bit formats should also been taken into consideration because we are in the transition phase from LDR images to HDR images and the majority of the display market still depend on the 8-bit LDR images. Based on whether the HDR compression algorithms provide backward-compatibility to LDR images, we category them into following two groups.

2.4.1 Single-layer HDR Image Compression

To efficiently encode the HDR data, authors in [26–28] propose new file formats for HDR images. Assuming that the exponents of the three colors do not vary much,

Ward [26] proposes to encode the shared exponent and fractions of the three colors in 32 bits. In his later work in [27], he proposes another signed 32-bit representation of HDR pixels in Luv color space. In [27], 15 bits are assigned to luminance and 16 bits are assigned to the two chromaticity. By encoding each color component using half-precision floating-point format, Bogart et al. propose a 48-bit representation for the three color channels of each pixel in [28]. In general, the new HDR file formats [26–28] can losslessly compress the HDR data with 10-16 bits per pixel for each color component.

To further improve the coding efficiency, lossy compression schemes [29, 80, 81] have been proposed for HDR images. In [80], the raw HDR image is transformed into the logarithmic domain and then uniformly quantized so that it can be compressed using the JPEG 2000 standard [30]. Lately, Miller et al. find that the compression gain can be further improved by applying a perceptual uniform transform instead of the logarithmic transform in [81]. In addition to the JPEG 2000 standard, new standard such as the JPEG XR [29] is also proposed to support file formats with higher bit-depth to enable lossy HDR compression.

2.4.2 Scalable HDR Image Compression

Figure 2.4: Flow of scalable HDR image compression.

In addition to new HDR image compression standards, researchers are also interested

in scalable HDR image compression [82–87] because both HDR and LDR images exist concurrently. The flow of the scalable HDR image compression is shown in Fig. 2.4. In this scalable compression framework, an HDR image is represented using a LDR image plus the residual and auxiliary information. For consumers who are only equipped with LDR devices, only the decoded LDR image will be used to display the transmitted image. For other consumers who need the raw HDR image, both the decoded LDR image and the residual and auxiliary information will be combined together to reconstruct the transmitted HDR image.

In [82, 83], Ward et al. propose to use the tone-mapped image as the LDR image and the ratio of the raw HDR and tone-mapped LDR image as the residual image. They implement their scalable compression scheme based on the widely-used JPEG standard. Similarly, Mantiuk et al. [84] propose another scalable compression scheme based on the MPEG standard. In [82, 83], the residual image is computed by directly using the tone-mapped LDR image while a reconstruction function is introduced in [84] to estimate the HDR image from the LDR image and compute the residual image using the estimated HDR image. With the reconstruction function, there is less information in the residual image which results in lower bit-depth for encoding the residual image. Following the same idea, Banterle et al. introduce the inverse tone mapping operators (iTMOs), which act as the reconstruction function in [84], into the scalable compression. By expanding the dynamic range of the LDR image, the iTMOs can estimate an HDR image from single LDR input. And they are very helpful in improving the coding efficiency of scalable compression algorithms because of the accurate HDR image estimation.

In this scalable compression framework, the TMOs determine the visual quality of the LDR image and affect the visual experience of LDR consumers. The accuracy of

the HDR image estimation, which is accomplished by iTMOs, determines how much information left in the residual image and affects the coding efficiency. Because of the reversibility of GTMCs, lower bit-rates for residual image have been achieved for GTMCs [85, 88]. In [85, 86], the authors even take both the GMTTC and its inverse into consideration simultaneously and design a GTMC that maximizes the coding efficiency. However, as discussed in previous tone mapping section, the GMTTCs cannot produce as good looking results as LTMOs, especially in terms of detail preservation. The visual quality of the tone-mapped LDR image in [85, 86] may not be satisfactory for LDR consumers. To provide satisfactory visual experience, the scalable compression scheme should also be able to work with LTMOs. However, it is very challenging to directly estimate the iTMO of a LTMO because of the non-linearity in the local tone mapping process [84, 89]. And it is reported in [83, 84, 90] that the size of the compressed HDR images is usually larger when using the LTMOs instead of the GTMCs to generate the tone-mapped LDR images.

2.4.3 Assessment of HDR Image Compression

Although the LDR quality assessment metrics, such as PSNR and SSIM, can be computed directly on HDR pixel values, they may not be able to predict the quality of HDR images reliably [91]. The reason is that while the luma values in LDR images are approximately perceptually uniform, this is not true for luminance values in HDR images. A proper perceptual uniform transform is desired before applying PSNR and SSIM on HDR images. In [92], the authors introduce a simple luminance encoding scheme which maps physical luminance values into perceptually uniform representations and apply this transformation to extend PSNR and SSIM to handle HDR images. In addition to the extensions of PSNR or SSIM, other perceptual quality

assessment models such as the HDR-VDP-2 [93] have also been proposed to assess the quality of HDR images. Starting from the visual differences predictor (VDP) [69] and its extension HDR-VDP [94], the HDR-VDP-2 has undergone several revisions. A recent benchmark study in [90] shows that among all the metrics discussed in this section, the HDR-VDP-2 is the best metric in terms of prediction quality of HDR images. Therefore, the HDR-VDP-2 is recommended for quality assessment of HDR images.

Chapter 3

Detail-Enhanced Multi-Scale Exposure Fusion

In this chapter, we address the backward-compatible HDR acquisition problem and propose a multi-scale fusion algorithm to directly synthesize an HDR-like LDR image from a bracket of differently-exposed LDR images. Our algorithm mainly consists of two modules. In the first module, we refine the multi-scale fusion algorithm in [3] to keep as much information in the LDR inputs as possible. In our second module, we propose an exposure-aware structure tensor for generation of a vector field and efficiently extract a detail layer from the vector field to further improve the visual quality of the multi-scale fusion result.

The rest of this chapter is organized as follows. The overview of the proposed exposure fusion algorithm is given in Section 3.1. The refinement of the multi-scale exposure fusion in [3] is explained in Section 3.2. Details of the detail extraction module is introduced in Section 3.3. Substantial experimental results are presented in 3.4. Finally, concluding remarks are given in Section 3.5.

3.1 Introduction

As extensively reviewed in Chapter 2, many exposure fusion algorithms have been proposed in last decade. Recently, a subjective user study conducted in [61] shows that the mutli-scale fusion algorithm in [3] produces the best fusion quality on average. Thus it is fair to select it as our baseline. Though the algorithm in [3] produces

Figure 3.1: The workflow of the proposed detail-enhanced multi-scale exposure fusion.

halo-free fusion results with good global contrast, we find that it cannot preserve information in the brightest and/or darkest regions of the HDR scene well. A new fusion algorithm is desired to overcome this problem. In addition, fine details in the inputs are usually lost because of the weighted average operation in the fusion

process. A detail extraction module is also desired to preserve fine-scale details for better visual quality of the fusion result.

Based on above two considerations, we propose a detail-enhanced multi-scale fusion algorithm. The workflow of the proposed algorithm is shown in Fig. 3.1. Our method mainly consists of two modules: multi-scale fusion and detail extraction. Based on the observation that using fewer scales for fusion helps to preserve information in brightest and/or darkest regions [4], we propose to reduce the pyramid layers in [3]. However, simply reducing pyramid layers may introduce halos. Thus we propose to refine the blending weights using the weighted guided filter (WGIF) [95] to suppress halos from appearing in the final fused result. As first introduced in [9], the detail layer can be extracted from the vector field of the multiple inputs. Following the same idea, we also propose to extract the detail layer from the vector field. Different from the work in [9], we propose an weighted structure tensor in the computation of vector field which leads to less noise in the extracted detail layer. In addition, the detail extraction is achieved by solving a large-scale optimization problem in [9]. In our work, we propose to approximate the optimization problem in [9] using a sequence of small-scale optimization problems. With our efficient approximation, the speed of the detail extraction module in [9] can be increased by more than four times. In the next two sections, the details of the two modules in our algorithm will be explained.

3.2 WGIF Based Multi-Scale Exposure Fusion

In this section, our WGIF based multi-scale fusion for intermediate fusion result generation is introduced. We first introduce the conventional multi-scale fusion algorithm in [3] and then demonstrate how we improve our baseline algorithm in [3] by incor-

porating the WGIF in weight map refinement to better preserve information in the brightest and darkest regions of the scene without introducing halo artifacts.

3.2.1 Pyramid Based Exposure Fusion

Suppose we have a stack of N differently-exposed LDR color images. Denote the i -th LDR image and its luminance component as Z_i and Y_i , respectively. The weights to blend each exposure can be obtained from the three measures of contrast (C_i), color saturation (S_i) and well-exposedness (E_i) as in [3].

The contrast measure C_i is defined as

$$C_i(p) = |Y_i(p) - k * Y_i(p)| \quad (3.1)$$

where p is the index of the pixel. Symbol $*$ is the convolution operator and k is a 3×3 Laplacian kernel.

Color saturation S_i is computed as

$$S_i(p) = \sqrt{\frac{(R_i(p) - \mu_i(p))^2 + (G_i(p) - \mu_i(p))^2 + (B_i(p) - \mu_i(p))^2}{3}} \quad (3.2)$$

where R, G, B are the red, green, blue color channels and $\mu_i(p) = \frac{R_i(p) + G_i(p) + B_i(p)}{3}$.

The well-exposedness E_i is defined as

$$E_i(p) = \exp\left(-\frac{R_i(p) - 0.5}{2\sigma^2}\right) \cdot \exp\left(-\frac{G_i(p) - 0.5}{2\sigma^2}\right) \cdot \exp\left(-\frac{B_i(p) - 0.5}{2\sigma^2}\right) \quad (3.3)$$

where $\sigma = 0.2$ as in [3]. The pixel-wise weight is then computed as

$$\tilde{W}_i(p) = C_i(p)^{e_C} \times \{S_i(p)\}^{e_S} \times \{E_i(p)\}^{e_E} \quad (3.4)$$

where the exponents e_C , e_S and e_E are introduced to control the influence of the three measurements and the default values of the three exponents are 1 as in [3]. Finally, the fused weight $\tilde{W}_i(p)$ is normalized as

$$W_i(p) = \frac{\tilde{W}_i(p)}{\sum_{j=1}^N \tilde{W}_j(p)}. \quad (3.5)$$

As discussed in [3], simply applying the weight in Eq. (3.5) on input LDR images may result in inconsistency in regions where the blending weights change quickly. To overcome the artifacts introduced by sharp weight map transitions, authors in [3] proposed to perform the exposure fusion in a multi-scale manner. Specifically, the input images are represented as Laplacian pyramids and the weight maps generated using Eq. (3.5) are represented using Gaussian pyramids [43]. Then the blending is conducted layer-wisely to avoid artifacts. Let $L\{Z_i\}^{(l)}$ and $G\{W_i\}^{(l)}$ be the Laplacian pyramid of input Z_i and the Gaussian pyramid of the weight map W_i , respectively. All the N inputs are blended layer-wisely as

$$L\{Z^{int}\}^{(l)}(p) = \sum_{i=1}^N \left[L\{Z_i\}^{(l)}(p) \cdot G\{W_i\}^{(l)}(p) \right] \quad (3.6)$$

where Z^{int} is the intermediate fusion result. Finally, the Laplacian pyramid $L\{Z^{int}\}^{(l)}$ collapses to produce Z^{int} . Note that all the RGB color channels use the same blending weights W_i in Eq. (3.6).

In [3], the number of the pyramid layers is computed as

$$\kappa = \lfloor \log_2 \min(r, c) \rfloor \quad (3.7)$$

where $\lfloor x \rfloor$ is the largest integer which is not greater than x . r and c are the height and width of the input image. In our experiments, we found that the value of κ affects the fusion result much. A larger κ such as that used in Eq. (3.7) results in good global contrast. A smaller κ will reduce the brightness difference between the brightest and darkest regions and reveal details in these near-saturated regions. However, using too small pyramid level is more likely to produce halos in the final fusion result.

To show the effects of the parameter κ , we present the fusion results using pyramid levels κ , $\kappa - 1$, $\kappa - 2$ and $\kappa - 3$ in Fig. 3.2. From Fig. 3.2(b) to Fig. 3.2(e), the pyramid level decreases. As show in Fig. 3.2(b), the naturalness of the scene is preserved well when the pyramid layer is set as κ . The sky is very bright and the shadow of the tree is very dark. However, these regions are near saturated and some information such as the boundary of the mountain in the middle is lost in Fig. 3.2(b). As κ decreases, information in near-saturated regions appears gradually as in Figs. 3.2(c)-3.2(e). However, the weight pyramid $G\{W_i\}^{(l)}$ does not always meet the edges of the input images. This may cause inconsistence along edges and are likely to produce halos as shown in Fig. 3.2(e). The improved fusion quality is not always guaranteed by simply decreasing the level of pyramid. An edge-protecting scheme should be used to refine the weight pyramid to suppress halos from appearing along edges.

(a) input images

(b)

(c)

(d)

(e)

Figure 3.2: Fusion results using different Gaussian pyramid levels. (a) input images. (b) κ layers. (c) $\kappa - 1$ layers. (d) $\kappa - 2$ layers. (e) $\kappa - 3$ layers.

3.2.2 Weighted Guided Filtering

In this section, we elaborate how each layer in the weight pyramid is refined using an edge-preserving filter. As discussed in [41, 95], the two filters have the ability to transform the structure of the guidance image to the smoothed image as such they are good candidates for our weight pyramid refinement.

Denote the refined weight map as $\hat{W}$. Assuming that the refined weight map $\hat{W}$ has a similar structure as the luminance Y in local supporting windows, for each pixel p in the window we have

$$\hat{W}(p) = a_p Y(p) + b_p \quad (3.8)$$

where a_p and b_p are the two constants to be determined. With the affine relation in Eq. (3.8), it can be easily verified that $\nabla \hat{W}(p) = a_p \nabla Y(p)$ which means that the smoothed weight map $\hat{W}$ has the similar structure of the luminance Y .

As shown in [41, 95], the constant pair (a_p, b_p) for each pixel p can be obtained by solving an optimization problem formulated on the local support window as

$$\arg \min_{a_p, b_p} \sum_{q \in \Omega_r(p)} \left[\left(a_p Y(q) + b_p - \hat{W}(q) \right)^2 + \frac{\epsilon}{\Gamma(p)} a_p^2 \right] \quad (3.9)$$

where $\Omega_r(p)$ is a square window centered at pixel p with a radius r . Here, we use the formulation in [95] instead of that in [41] because the filter in [95] refines previous work in [41] by introducing an edge-aware weight $\Gamma(p)$ which results in a better ability in preserving edges with little gain in computational cost. And the edge-aware weight $\Gamma(p)$ is defined by the local variance of the 3×3 window centered at p as in [95]. r and ϵ are free parameters which control the extent of smoothing and are selected as 4 and $1/1024$ in our implementation.

The closed form solution of Eq. (3.9) can be computed as

$$a_p = \frac{\frac{1}{|\Omega_r(p)|} \sum_{i \in \Omega_r(p)} Y(i)W(i) - \bar{Y}_p \bar{W}_p}{\sigma_{Y_p}^2 + \frac{\epsilon}{\Gamma(p)}}, \quad (3.10)$$

$$b_p = \bar{W}_p - a_p \bar{Y}_p \quad (3.11)$$

where $|\Omega_r(p)|$ is the number of pixels in the local window. Because a fixed window radius is used for all pixels, $|\Omega_r(p)|$ is a constant which is determined by r only. $\bar{Y}_p$ and $\bar{W}_p$ are the means of the luminance and weight in the window. And σ_{Y_p} is the variance of Y_p in $|\Omega_r(p)|$. Note that the computation of Eq. 3.10 actually relies on the mean, variance and co-variance in the local support window. It can be efficiently computed using the integral image technique [96].

Once (a_p, b_p) is obtained, the smoothed weight map $\hat{W}(p)$ can be computed according to Eq. 3.8. Note that each pixel p will be involved in multiple windows. These overlapped windows which cover p will result in different pairs of parameters (a_p, b_p) according to Eq. 3.10. Thus different window gives different smoothed output $\hat{W}(p)$ at p . A simple strategy is to average $\hat{W}(p)$ over all the windows covering p to get the final smoothed output. Note that the number of windows which cover p equals to the number of pixels in local support window $\Omega_r(p)$. So the average of $\hat{W}(p)$ can be computed using the average of pairs (a_p, b_p) which can be easily derived from Eq. 3.8. Denote the means of a_p and b_p in window $\Omega_r(p)$ as $\bar{a}_p$ and $\bar{b}_p$, the final smoothed weight $\hat{W}(p)$ by [95] is obtained as

$$\hat{W}(p) = \bar{a}_p Y(p) + \bar{b}_p. \quad (3.12)$$

Because the edge-preserving filter is employed to refine the weight map pyramid. We

decompose the luminance channel into a gaussian pyramids similarly to the weight pyramid and apply the filtering layer-wisely. Specifically, if the luminance pyramid is denoted as $G\{Y_i\}^{(l)}$, we apply the filtering using the corresponding layers of $G\{Y_i\}^{(l)}$ and $G\{W_i\}^{(l)}$.

Figure 3.3: Comparison of Gaussian pyramid of the weight maps and their smoothed versions using the WGIF.

Some examples of the smoothed weight maps are shown in Fig. 3.3. In this figure, we show the 1st, 3rd and 6th layers of the weight pyramids of the 2nd, 4th and 6th input images in the first column. The corresponding smoothed weight maps are denoted as $E\{W_i\}^l$ and shown in the second column.

The fusion results using the refined weight maps are shown in Fig. 3.4. Similar to the gaussian pyramid, information in brightest and darkest regions gradually appears as pyramid level κ decreases. However, halos are less likely to occur in the final fusion results when using the refined weight map pyramid.

(a) κ (b) $\kappa - 1$ (c) $\kappa - 2$ (d) $\kappa - 3$

Figure 3.4: Fusion results using different WGIF refined pyramid levels. (a) κ layers. (b) $\kappa - 1$ layers. (c) $\kappa - 2$ layers. (d) $\kappa - 3$ layers.

To find the best pyramid level, we employ the MEF-SSIM metric in [61] to evaluate the fusion results quantitatively. The average MEF-SSIM scores of 10 test image

sets for different pyramid levels are shown in Table 3.1. It can be observed that as κ decreases, the quality of the fusion image improves. However, when the pyramid level is too small, the MEF-SSIM score decreases. The performance of the MEF-SSIM score coincides with our observation that decreasing pyramid level results in improved visual quality while too smaller levels introduce halos and lower the quality. It can be observed the highest MEF-SSIM score is achieved when using $\kappa - 2$ layers in both two types of pyramids. And the highest MEF-SSIM score for our WGIF-based pyramid is higher than that of the Gaussian pyramid. Thus the pyramid level in our WGIF-based fusion is set as $\kappa - 2$ in all our following experiments.

	κ	$\kappa - 1$	$\kappa - 2$	$\kappa - 3$
Gaussian	0.9772	0.9778	0.9783	0.9768
WGIF	0.9809	0.9813	0.9824	0.9807

Table 3.1: Average MEF-SSIM scores of 10 test image sets of using different pyramid levels

3.2.3 Discussion on WGIF-based Fusion

An attractive feature of the WGIF is that it is friendly to multi-scale fusion strategy from the computational cost point of view. Note that the most complex part of the WGIF is the computation of coefficients maps $\bar{a}_i$ and $\bar{b}_i$. In the multi-scale representation, the coefficient maps of a layer can be efficiently interpolated from the coefficient maps of its down-sampled version, which is exactly the next layer in the pyramid. Thus the WGIF can be applied on a limited number of layers and the coefficient maps of remaining layers can be obtained by interpolation to reduce the computation cost. In our implementation, we only compute the coefficients of 4-th and the $(\kappa - 2)$ -th layers which are denoted as $\{\bar{a}_i\}^{(4)}$, $\{\bar{b}_i\}^{(4)}$ and $\{\bar{a}_i\}^{(\kappa-2)}$, $\{\bar{b}_i\}^{(\kappa-2)}$.

The coefficients maps $\{\bar{a}_i\}^{(l)}$ for the other layers are computed as

$$\{\bar{a}_i\}^{(l)} = \begin{cases} UP(\{\bar{a}_i\}^{(4)}, 2^{4-l}) & \text{if } 1 \leq l < 4 \\ UP(\{\bar{a}_i\}^{(\kappa-2)}, 2^{\kappa-2-l}) & \text{if } 5 \leq l < \kappa - 2 \end{cases} \quad (3.13)$$

where $UP(\{\bar{a}_i\}^{(4)}, 2^{4-l})$ stands for upsampling $\{\bar{a}_i\}^{(4)}$ by scale 2^{4-l} . The coefficient maps $\{\bar{b}_i\}^{(l)}$ can be computed in the same way.

(a)

(b)

(c)

(d)

Figure 3.5: Illustration of noise reduction by the WGIF.

There are two benefits of filtering the weight map using the WGIF: 1) noises in the

weight maps are smoothed out as such a more consistent weight is assigned to adjacent pixels. 2) the structure of the luminance channel is transferred to the weight map as such the weight map corresponds to the main structures, i.e. the edges in the input images, which is useful to reduce halos. In our experiments, we found that besides suppressing halos the WGIF is also helpful in reducing noises. The reduction of noise is illustrated in Fig. 3.5. The first column is the weight map and fusion result without WGIF refinement and the second column is the weight map and fusion result after WGIF refinement. It can be observed that the weight map contains less noise after the WGIF refinement and it results in less noise in the fusion result as shown in the zoom-ins in Figs. 3.5(c) and 3.5(d).

3.3 Detail-Enhanced Exposure Fusion

The multi-scale exposure fusion provides a good fusion result. However, because of the weighted average, details in the original image bracket are usually lost. To finally obtain high quality fusion result with good detail synthesis, we propose a detail extraction method to obtain a detail layer from the multi-exposure inputs. Our work is inspired by the development of local tone mapping operators. As discussed in Chapter 2, some local tone mapping operators are developed under an edge-preserving decomposition framework. Under this framework, the contrast and brightness are controlled by the design of the global tone mapping curve and the details are added back to improve the visual quality. Similarly, it is also desired to extract details from the bracket of differently-exposed images for better visual quality.

In the rest of this section, our detail extraction module will be explained. We first estimate a vector field that represents the structural information of the image stack

and then reconstruct the detail layer from the vector field by solving an optimization problem. Note that we only extract the detail layer from the luminance channels, other color channels can be adjusted according to the luminance change.

3.3.1 Problem Formulation

In image processing, the gradients of an image are very important information which usually indicate the structure of the image. As discussed in [9], if the vector field of the detail layer is known, the detail layer can be reconstructed by solving a quadratic optimization problem. Note that the detail layer is usually multiplied with the intermediate image to improve the visual quality. The magnitude of the detail layer in spatial domain is desired to be 1 so that it will enhance local contrast without changing the brightness of the whole image. For simplicity in formulating the detail extraction problem, we follow the work in [9] and formulate the optimization problem in logarithmic domain.

Denote the logarithmic detail layer and the vector field at each pixel p as $l_d(p)$ and $(v_1(p), v_2(p))$, respectively. The optimization problem for detail extraction is formulated as

$$\arg \min_{l_d} \sum_p l_d^2(p) + \alpha \sum_p \left[\left(\frac{v_1(p) - \frac{\partial l_d(p)}{\partial x}}{\psi(v_1(p))} \right)^2 + \left(\frac{v_2(p) - \frac{\partial l_d(p)}{\partial y}}{\psi(v_2(p))} \right)^2 \right] \quad (3.14)$$

The function $\psi(z)$ is defined as [17]

$$\psi(z) = \sqrt{|z|^\gamma + \epsilon} \quad (3.15)$$

where γ and ϵ are set as 1.2 and 10^{-4} as in [17]. Parameter α is a balance factor of

the two terms and is usually selected as 0.25.

The first term of Eq. (3.14) ensures that the magnitude of the l_d is around 0. And the second term ensures that ∇l_d is close to the vector field. Note that l_d only contains small scale gradients. The weights $\psi(v_1(p))$ and $\psi(v_2(p))$ are introduced to prevent sharp edges from appearing in l_d .

With the formulation in Eq. (3.14), l_d can be obtained with given vector field. In the HDR tone mapping scenario, the vector field can be easily computed using the gradients of the HDR input. However, in the exposure fusion case, the inputs are multi-exposure LDR images. The vector field cannot be computed directly. In next section, we will show how the vector field is derived from multiple inputs.

3.3.2 Vector Field Generation

Denote the luminance channel of each input image and the intermediate image generated by the multi-scale fusion scheme discussed in Section 3.2 as Y_i and Y^{int} , respectively. Note that the detail extraction problem is formulated in logarithmic domain, the vector field is also computed in the logarithmic domain. Denote the logarithmic luminance of the input LDR images as $\hat{Y}_i(p) = \log(Y_i(p) + 1)$ where 1 is introduced to avoid computing logarithmic of 0.

The gradients of the multiple LDR inputs at pixel p can be computed as

$$\nabla \hat{Y}(p) = \begin{bmatrix} \frac{\partial \hat{Y}_1(p)}{\partial x} & \frac{\partial \hat{Y}_1(p)}{\partial y} \\ \vdots & \vdots \\ \frac{\partial \hat{Y}_N(p)}{\partial x} & \frac{\partial \hat{Y}_N(p)}{\partial y} \end{bmatrix}. \quad (3.16)$$

We use forward difference to compute the gradients. Denote p_r and p_b as pixels on the right and bottom of current pixel p . The derivatives in Eq. (3.16) becomes

$$\begin{aligned}\frac{\partial \hat{Y}_i(p)}{\partial x} &= \hat{Y}(p_r) - \hat{Y}(p) \\ \frac{\partial \hat{Y}_i(p)}{\partial y} &= \hat{Y}(p_b) - \hat{Y}(p)\end{aligned}\quad (3.17)$$

With the gradient matrix $\nabla \hat{Y}(p)$, a structure tensor $\Phi_{\hat{Y}}(p)$ can be computed as

$$\Phi_{\hat{Y}}(p) = (\nabla \hat{Y}(p))^T \nabla \hat{Y}(p). \quad (3.18)$$

The concept of the tensor $\Phi_{\hat{Y}}(p)$ was introduced in the image processing literature by Di Zenzo [97]. Note that though the gradient matrix $\nabla \hat{Y}(p)$ is an $N \times 2$ matrix where N is determined by the number of input images, the structure tensor $\Phi_{\hat{Y}}(p)$ is always a 2×2 matrix.

The eigenvalues of the structure tensor provide a rich and discriminative description of the local geometry of the image. Denote the two eigenvalues of $\Phi_{\hat{Y}}(p)$ as ρ_1 and ρ_2 . Without loss of generality, we assume ρ_1 is no less than ρ_2 . When both of the two eigenvalues are small, the pixel p is located in a homogeneous region. When ρ_1 is large while ρ_2 is small, the pixel p is located close to an image edge. When both of them are large, the pixel p is close to an image corner.

With the singular value decomposition, the matrix $\nabla \hat{Y}(p)$ is represented as

$$\nabla \hat{Y}(p) = U_h(p) \Lambda_h(p) V_h^T(p) \quad (3.19)$$

where $U_h(p) \in \mathbb{R}^{N \times 2}$, $V_h(p) = [\vec{v}_{h,1}(p), \vec{v}_{h,2}(p)] \in \mathbb{R}^{2 \times 2}$. $U_h^T(p) U_h(p) = I_N$ and

$V_h^T(p)V_h(p) = I_2$ where I_N and I_2 are the $N \times N$ and 2×2 identity matrix. $\vec{v}_{h,1}(p)$ and $\vec{v}_{h,2}(p)$ are the two columns of $V_h(p)$. $\Lambda_h(p)$ is a diagonal matrix $\text{diag}\{\sigma_1(p), \sigma_2(p)\}$.

It can be verified that

$$\begin{aligned}\Phi_{\hat{Y}}(p) &= (\nabla U_h(p)\Lambda_h(p)V_h^T(p))^T \nabla U_h(p)\Lambda_h(p)V_h^T(p) \\ &= V_h(p)\Lambda_h^2(p)V_h^T(p)\end{aligned}\tag{3.20}$$

Equation (3.20) shows that the eigenvectors of the gradient matrix is also the eigenvectors of the structure tensor. A vector field can be constructed according to the eigenvalues and eigenvectors, i.e. the vector field is computed as $\sigma_1(p)\vec{v}_{h,1}^T(p)$ in [2].

However, it is shown in Figs. 3.6(a) and 3.6(c) that noises are amplified when using such method to generate the vector field. The reason for noise amplification is that not all gradient information from the input images are reliable. Pixels which are not well-exposed may contain noisy gradient information and result in unwanted gradient amplifications. To address this problem, we propose a weighted gradient matrix which takes the exposedness of the pixels into consideration.

Note that for each pixel p , the gradients are computed using p , p_r and p_b as in Eq. (3.17). With the consideration that the computation of the gradient is reliable only when the pair of pixels, i.e. (p, p_r) or (p, p_b) , is well-exposed, we proposed to define the reliability of the gradients as a multiplication of the well-exposedness of the pair of pixels. The exposedness weight of the gradients is computed as

$$\begin{aligned}\tilde{W}_{x,i}(p) &= \frac{h(Y_i(p))h(Y_i(p_r))}{\max_j\{h(Y_j(p))h(Y_j(p_r))\}} \\ \tilde{W}_{y,i}(p) &= \frac{h(Y_i(p))h(Y_i(p_b))}{\max_j\{h(Y_j(p))h(Y_j(p_b))\}}\end{aligned}\tag{3.21}$$

where $h(\cdot)$ is a hat function which assigns large weights to well-exposed pixels and

(a)

(b)

(c)

(d)

Figure 3.6: Fusion results using different structure tensor. (a) and (c) Enhanced fusion results using the structure tensor in [2]. (b) and (d) Enhanced fusion results using the proposed tensor. It can be observed that the proposed tensor results in less noise.

small weights to near-saturated pixels. And $h(\cdot)$ is defined as

$$h(z) = \begin{cases} z + 1, & \text{if } z \leq 127 \\ 256 - z, & \text{otherwise.} \end{cases} \quad (3.22)$$

With the weights in Eq. (3.21), the gradients of best-exposed pixels stay unchanged

while the others are suppressed to avoid amplification of noise. By introducing the weights into Eq. (3.16), a new gradient matrix can be computed

$$\nabla \tilde{Y}(p) = \begin{bmatrix} \tilde{W}_{x,1}(p) \frac{\partial \hat{Y}_1(p)}{\partial x} & \tilde{W}_{y,1}(p) \frac{\partial \hat{Y}_1(p)}{\partial y} \\ \vdots & \vdots \\ \tilde{W}_{x,N}(p) \frac{\partial \hat{Y}_N(p)}{\partial x} & \tilde{W}_{y,N}(p) \frac{\partial \hat{Y}_N(p)}{\partial y} \end{bmatrix}. \quad (3.23)$$

And the weighted structural tensor can be computed accordingly. The enhanced fusion results using the proposed weighted structural tensor are shown in Figs. 3.6(b) and 3.6(d). It is obvious that the proposed weighted tensor produces much less noise than [2].

Note that in the singular value decomposition, the eigenvectors are lines without a preassigned orientation. When the direction of the desired vector field is opposite to the gradient of the intermediate fusion result Y^{int} , the extracted detail layer will decrease the contrast of Y^{int} or even introduce unexpected distortions when multiplied with Y^{int} . To ensure that the detail layer always enhance the intermediate result, we propose to correct the directions of the desired vector field (v_1, v_2) as

$$\begin{aligned} \tilde{v}_1(p) &= v_1(p) \operatorname{sign} \left(v_1(p) \frac{\partial Y^{int}}{\partial x} \right) \\ \tilde{v}_2(p) &= v_2(p) \operatorname{sign} \left(v_2(p) \frac{\partial Y^{int}}{\partial y} \right) \end{aligned} \quad (3.24)$$

where $\operatorname{sign}(x)$ is the sign of x .

After this direction correction, the direction of the desired vector field is consistent with the gradient of the intermediate fusion result Y^{int} which ensures that the later extracted detail layer enhanced Y^{int} . The effect of the direction correction is shown

in Fig. 3.7. It can be observed from Fig. 3.7(b) that inconsistency between directions of the vector field and gradients of the intermediate fusion result causes distortions, i.e. the cloud regions in the zoom-ins in Fig. 3.7(e), in the final enhanced result while with the correction in Eq. (3.24), the enhanced result in Fig. 3.7(c) looks more shaper than the intermediate fusion result in Fig. 3.7(a) without the distortion in Fig. 3.7(b).

Figure 3.7: Effect of the direction correction in Eq. (3.24). (a) Intermediate fusion result. (b) Enhanced exposure result without direction correction. (c) Enhanced exposure result with direction correction. (d)-(f) Zooms-ins of (a)-(c).

3.3.3 Efficient Solution

In this section, we will present how we efficiently solve the optimization problem in Eq. (3.14). Because the problem in Eq. (3.14) is in quadratic form, the closed form solution can be derived using convex optimization techniques [98]. Denote the objective function in Eq. (3.14) as J . It can be represented in matrix form as

$$\begin{aligned} J(L_d) = & L_d^T L_d + \alpha (V_1 - D_x L_d)^T A(V_1) (V_1 - D_x L_d) \\ & + \alpha (V_2 - D_y L_d)^T A(V_2) (V_2 - D_y L_d) \end{aligned} \quad (3.25)$$

where L_d is a long vector which is constructed by stacking the columns of the detail layer together. D_x and D_y are forward difference operators. $A(V_1)$ and $A(V_2)$ are diagonal matrices which are $\text{diag}\{\frac{1}{\psi^2(v_1(p))}\}$ and $\text{diag}\{\frac{1}{\psi^2(v_2(p))}\}$, respectively.

Let $\frac{\partial J(L_d)}{\partial L_d} = 0$, it can be derived that

$$(I + \alpha L_v) L_d = \alpha B \quad (3.26)$$

where

$$L_v = D_x^T A(V_1) D_x + D_y^T A(V_2) D_y \quad (3.27)$$

$$B = D_x^T A(V_1) V_1 + D_y^T A(V_2) V_2 \quad (3.28)$$

Note that the linear system in Eq. (3.26) is very similar to that in [17]. The coefficient matrix $(I + \alpha L_v)$ is a five-point spatially inhomogeneous Laplacian matrix and the system can be solved using traditional numerical solutions [99]. Solving Eq. (3.26) requires computation of the inverse of the huge Laplacian matrix. This is quite time-

consuming especially when computed on low-end devices such as mobile phones. To accelerate the detail extraction process, we propose to approximate the optimization problem in Eq (3.14) in a separable manner.

Our approximation is inspired by the separable smoothing filters in [11, 100]. These filters are proposed to smooth an image while preserving its edges. Inspired by the fact that the two-dimensional gaussian smoothing can be implemented by sequential one-dimensional smoothing. Authors in [11, 100] proposed to approximate their spatially-varying two-dimensional smoothing with simpler one-dimensional solvers. Following the same idea, we formulate our one-dimensional optimization problem for each row and column. In the following part, we present our one-dimensional optimization problem for each row. The optimization problem for each column can be similarly formulated.

Our one-dimensional optimization problem for each row is formulated as

$$\arg \min_{L_d^k} \sum_{x=0}^{c-1} \left[(L_d^k(x) - L_d^{k-1}(x))^2 + \alpha_{\lceil k/2 \rceil} \left(\frac{\frac{\partial L_d^k(x)}{\partial x} - v_1(x)}{\psi(v_1(x))} \right)^2 \right]. \quad (3.29)$$

Here we only consider one row and L_d^k stands for a row vector of the image instead of the vector of the whole image as in Eq. (3.25). L_d^{k-1} is the result of last smoothing process and it is initialized as all zeros, i.e. $L_d^0 = 0$.

The first term in Eq. 3.29 ensures that the one-dimensional smoothing result is close to last round smoothing result and the second term penalizes L_d^k when it is far away from the desired vector field. Note that the formulation in Eq. 3.29 is actually an low-pass filter [17]. Parameter $\alpha_{\lceil k/2 \rceil}$ is reduced after a complete round of horizontal and vertical smoothing so that the the total amount of smoothing is similar to that

Algorithm 1 Separable detail extraction

Input:Vector field of the stack of the input images (V_1, V_2) Iteration number T Smoothing balance factor α **Output:**Detail layer of the input images L_d Initialize $L_d^0 \leftarrow 0$ Compute weight matrix $\psi(V_1)$ and $\psi(V_2)$ using V_1, V_2 **for** $k = 1; k \leq 2T; k++$ **do** Compute $\alpha_{\lceil k/2 \rceil}$ using Eq. (3.30) **for** $y = 0; y < r - 1; y++$ **do** Compute y -th row of L_d^k using Eq. (3.29) **end for** $k \leftarrow k + 1$ **for** $x = 0; x < c - 1; x++$ **do** Compute x -th column of L_d^k using Eq. (3.29) **end for****end for**

without the approximation. Parameter $\alpha_{\lceil k/2 \rceil}$ is computed as

$$\alpha_{\lceil k/2 \rceil} = \frac{3}{2} \frac{4^{T-\lceil k/2 \rceil}}{4^T - 1} \alpha \quad (3.30)$$

where T is a user-specific parameter which is the total number of the separable iteration. And the maximum k in Eq. (3.29) is $2T$ because one round of smoothing consists of two one-dimensional smoothing. The complexity of solving all the 1D optimization problems for the whole image is $\mathcal{O}(S)$ where S is the total number of pixels in each input image. Thus the complexity for the separable approximation with T iterations is $\mathcal{O}(2TS)$. The steps of the separable detail extraction is summarized in Algorithm 1. And the detailed solution of the one-dimensional optimization problem (3.29) can be found in Appendix A.

Parameter T affects the result of detail extraction. Small T results in shorter execu-

Figure 3.8: Effect of the number of iterations T . (a) $T = 1$. (b) $T = 3$. (c) $T = 15$. The brightness and contrast of detail layers are manually adjusted for better illustration. (d)-(f) zoom-ins of (a)-(c)

tion time under the risk of poor approximation which results in “streaking artifacts” as shown in Fig. 3.8(a) and its zoom-in Fig. 3.8(d). Much longer time is required with little gain in accuracy of approximation when T is very large, i.e. as shown in Fig. 3.8(c). Typically, $T = 3$ gives a good balance between time consumption and accuracy of approximation as shown in Fig. 3.8(b). Note that the brightness and contrast of detail layers in Fig. 3.8 are manually adjusted for better illustration. Once the detail layer is computed, it is transformed back to spatial domain and multiplied with the intermediate image to produce the final detail-enhanced fusion result.

3.4 Comparison among Exposure Fusion Algorithms

In this section, we present comparison results between the proposed exposure fusion algorithm and six competitive state-of-the-arts. We present two qualitative experiments and one quantitative experiment in this section. To show the effectiveness of the proposed WGIF-based multi-scale fusion, the detail enhancement component is excluded when our method is compared with the algorithms in [3–5] in experiment 1. For more comparisons with algorithms in [6, 8, 101] in experiment 2, the detail enhancement module is added. Finally, we evaluate all the algorithms using the objective metric in [61].

3.4.1 Experiment 1

In this experiment, we compare our method with the three fusion algorithms in [3–5]. Note that this experiment is designed to show the effectiveness of the proposed WGIF-based fusion. The detail enhancement module is excluded in this experiment.

We first compare our WGIF-based fusion with our baseline in [3] using the inputs in Figs. 3.9(a)-3.9(c). This set of image is a typical daytime high dynamic range scene. Because that the lighting condition in daytime is usually quite good, photographers have no problem in capturing the indoor scene, i.e. the railing and the sinage in the shopping mall. However, the sky outside is so bright that it becomes over-exposed and nothing can be seen in these regions. By gradually decreasing the exposure time, the cloud in the sky begins to appear. However, the indoor scene becomes too dark when using lower exposure. Thus the whole scene cannot be perfectly captured using single exposure. By applying our method and the algorithm in [3], images like Figs.

3.9(d) and 3.9(f) are obtained. It can be observed that both the indoor scene and the outdoor sky are kept in the two fused results which result in better visual quality than using single exposure. The global contrast of both fused images are good. However, our method has a better ability in preserving information in the brightest sky region. The clouds can be found in Fig. 3.9(d) while the corresponding region is totally white in Fig. 3.9(f). This is caused by using fewer layer in our method. Note that because of the WGIF, no halos are found in our fusion result even in the challenging areas, where the brightness changes rapidly in different exposures, as shown in the zoom-in in Figs. 3.9(e).

Then we compare our WGIF-based fusion with the single-scale fusion algorithm in [4] using a nighttime high dynamic range scene as shown in Figs. 3.10(a)-3.10(c). Unlike previous daytime scene, the ambient light is very weak at night. As such when a normal exposure is used, only the lights emitted from lamps can be well-captured while other areas such as the street and cars can be hardly seen. With gradual increase in exposure time, the street and cars become visible while the information inside the bright supermarket gradually loses because of saturation. As shown in Figs. 3.10(d) and 3.10(f), both our method and the algorithm in [4] preserves the brightest (the supermarket) and darkest (the street and cars) regions of the scene. However, it can be observed that average brightness of Fig. 3.10(f) is darker than our result in Fig. 3.10(d). It seems that the fusion weight in [4] is not suitable for such example and results in the brightness shift. Also, it can be found that there is a ghost black box bounding the lamp as shown in Fig. 3.10(g). This is caused by the single-scale strategy used for fusing all exposures in [4]. Note that our method uses a multi-scale fusion strategy, the difference between varying exposures are blended more smoothly. As such no such ghosting artifact is found in our result as shown in Fig. 3.10(e).

Figure 3.9: Comparison with the fusion algorithm in [3]. (a)-(c) input images. (d) our WGIF-based fusion result. (f) result by [3]. (e) and (g) zoom-ins of (d) and (f).

(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 3.10: Comparison with the fusion algorithm in [4]. (a)-(c) input images. (d) our WGIF-based fusion result. (f) result by [4]. (e) and (g) zoom-ins of (d) and (f).

Figure 3.11: Comparison with the fusion algorithm in [5]. (a)-(c) input images. (d) our WGIF-based fusion result. (f) result by [5]. (e) and (g) zoom-ins of (d) and (f).

In addition to above two methods, we also compare our method with the algorithm in [5]. Different from [3, 4], authors in [5] propose to fuse the differently-exposed inputs patch-wisely. They partition each image into overlapped blocks and decompose each block into three components. Then each component is averaged according to a local structure aware weight and the final fused result is obtained by reconstructing blocks from the three modified components. The inputs of this comparison, which is another daytime high dynamic range scene, are shown in Figs. 3.11(a)-3.11(c). The two fused results are shown in Figs. 3.11(d) and 3.11(f). It can be observed that both the global contrast of these two results are quite good. However, similar to our baseline in [3], the algorithm in [5] cannot preserve the clouds in the sky. Comparing Figs. 3.11(d) and 3.11(f), it seems that the contrast of the coffee machine fused by [5] is slightly better than ours. However, as shown in the zoom-in in Fig. 3.11(g), halos can be found along the edges of the coffee machine in Fig. 3.11(f). Our method suppresses halos by the weighted guided filter and there is no such artifact in the zoom-in shown in Fig. 3.11(e).

3.4.2 Experiment 2

The above first experiment shows that our WGIF-based fusion can produce visually appealing fusion results in a variety of high dynamic range scenes. In this second experiment, we show how our detail-enhanced component can further improve the visual quality by better preservation of the details. In this experiment, we add the detail-enhanced module and compare our full detail-enhanced method with four more exposure fusion algorithms [6–9].

We first compare our method with the algorithm in [6]. Unlike our method which

(a)

(b)

(c)

(d)

(e)

Figure 3.12: Comparison with the fusion algorithm in [6]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [6]. (c) and (e) zoom-ins of (b) and (d).

fused the inputs by weighted averaging, authors in [6] formulate exposure fusion as a probabilistic inferring problem. Their work desires to find a fusion result which preserves the local contrast and color consistency of the multiple inputs. The fused results of these two methods are shown in Figs. 3.12(b) and 3.12(d). It can be observed that both methods preserve local contrast well. However, the global contrast of Fig. 3.12(d) is not as good as Fig. 3.12(b). The reason may be that the algorithm in [6] tries to keep the consistency in adjacent pixels. Note that our method includes a detail enhancement module and the texture and painting on the walls are much sharper than that produced by [6] as shown in the zoom-ins in Figs. 3.12(c) and 3.12(e).

Considering that perceptual measures may help producing more visually appealing fusion results, authors in [6] proposed to improve their work by using the perceived local contrast and color saturation measures in [7]. In addition to the improved perceptual measures, the authors also proposed a more flexible probabilistic fusion model in [7]. With the new perceptual measures and fusion model, the algorithm in [7] outperforms that in [6]. An example of the fusion result by [7] is shown in Fig. 3.13(b). This image is provided by the authors. The fusion result by our method using the same input bracket is shown in Fig. 3.13(d). Comparing these two images, it can be found that the proposed algorithm preserves the global contrast better than the algorithm in [7]. In addition, the algorithm in [7] does not take details into consideration. As shown in zoom-ins in Figs. 3.13(c) and 3.13(e), the details, such as the leaves and textures of the brick wall and basket, are preserved better by our method.

We then compare our method with the algorithm in [8]. Authors in [8] try to preserve the details of the input sequence and propose to apply the weighted average fusion

(a)

(b)

(c)

(d)

(e)

Figure 3.13: Comparison with the fusion algorithm in [7]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [7]. (c) and (e) zoom-ins of (b) and (d).

(a)

(b)

(c)

(d)

(e)

Figure 3.14: Comparison with the fusion algorithm in [8]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [8]. (c) and (e) zoom-ins of (b) and (d).

(a)

(b)

(c)

(d)

(e)

Figure 3.15: Comparison with the fusion algorithm in [9]. (a) input images. (b) our detail-enhanced fusion result. (d) result by [9]. (c) and (e) zoom-ins of (b) and (d).

in two layers. They first decompose each input image into a base layer and a detail layer. Then different weights are assigned to the two layers to blend them. Finally the blended base and detail layers are combined to obtain the final fusion result. The fused results are shown in Figs. 3.14(b) and 3.14(d). It can be observed that the result by [8] is not as sharp as ours. The reason may be that the details are high frequency signals. When averaged, the details are smoothed by low-pass filter and some high frequency components are removed which results in poorer detail preservation. Our detail enhancement module extract details from the structure tensor as such the details are not smoothed and can be added back to improve the sharpness. Note that because the algorithm in [8] employs a single-scale fusion strategy for the base layer, inconsistency is found as marked by the red arrow in Fig. 3.14(d).

Finally, we compare our method with the algorithm in [9] which also includes a detail enhancement module. The fused results are shown in Figs. 3.15(b) and 3.15(d). From these two images, it can be found that both methods preserve the details well. However, as shown in zoom-ins in Figs. 3.15(c) and 3.15(e), our method outperforms [9] in the brightest regions. The reason is that the algorithm in [9] relies on the multi-scale fusion in [3] while our WGIF-based fusion outperforms [3] in brightest and darkest regions. The vector field estimated in [9] is based on averaging the gradients of multiple inputs. As pointed out in [61], this approach may result in unnatural enhancement of unwanted edges which can be observed in the roof patch in Fig. 3.15(e). Our detail layer is extracted from the weighted structure tensor and is unlikely to produce such artifacts as shown in Fig. 3.15(c). In addition, the detail extraction problem in [9] is solved directly while our separable approximation increases the speed by about four times.

3.4.3 Experiment 3

In above two experiments, we have already shown the qualitative comparison between our detail-enhanced fusion algorithm and other six competitive state-of-the-arts [3–6, 8, 9]. In this third experiment, we present the quantitative comparisons among the mentioned seven fusion algorithms using the objective MEF-SSIM metric in [61]. The ten images showed in this chapter are used in this quantitative comparison and the MEF-SSIM score for each image set is shown in Table 3.2. From the table, it can be observed that our fusion algorithm achieves the highest MEF-SSIM score on average.

The computation time of each algorithm on each image set is shown in Table 3.3. Although the proposed method is not the fastest one, its speed is competitive to some of the state-of-the-arts such as the algorithms in [5] and [8]. Note that except for the proposed method and the algorithm in [9], the other algorithms do not contain a detail enhancement component. Comparing to the detail enhanced exposure fusion algorithm in [9], the proposed method achieves around 2.5x speeds up.

3.4.4 Experiment 4

In addition to above qualitative and quantitative comparisons, we also conduct a subjective comparison to evaluate the performance of existing exposure fusion algorithms. Six males and four females are involved in the subjective evaluation. We employ the pair-wise comparison method as in [78] in this user study. Each time, a pair of exposure fusion results, which are generated by two different fusion algorithms, are displayed side by side in front of the participants. The participants are asked to choose which one is more preferred. During the experiment, the sequence and spatial

Set	[3]	[4]	[5]	[6]	[8]	[9]	ours
1	0.950	0.913	0.941	0.904	0.956	0.925	0.969
2	0.986	0.951	0.980	0.954	0.986	0.956	0.990
3	0.974	0.944	0.965	0.945	0.956	0.952	0.981
4	0.969	0.952	0.960	0.960	0.971	0.944	0.983
5	0.976	0.962	0.971	0.958	0.980	0.954	0.975
6	0.983	0.948	0.982	0.957	0.935	0.951	0.979
7	0.971	0.955	0.961	0.952	0.974	0.951	0.984
8	0.991	0.968	0.981	0.969	0.981	0.976	0.993
9	0.982	0.948	0.970	0.955	0.941	0.970	0.988
10	0.988	0.962	0.985	0.965	0.979	0.962	0.991
Average	0.977	0.950	0.970	0.952	0.963	0.954	0.983

Table 3.2: MEF-SSIM scores of different fusion algorithms.

Set	Image Size	[3]	[4]	[5]	[6]	[8]	[9]	ours
1	1025*769*9	5.6	10.0	19.9	1.7	28.9	30.9	18.1
2	1064*768*5	3.2	5.4	13.4	1.0	15.2	27.7	12.0
3	530*795*3	1.2	2.0	5.8	0.4	5.2	13.5	5.8
4	2128*1426*5	12.1	22.0	53.6	4.1	61.0	113.1	44.8
5	752*500*4	1.2	1.9	3.6	1.4	4.3	10.4	4.1
6	512*384*5	1.1	1.7	3.3	0.3	4.2	10	3.9
7	2128*1426*3	7.7	13.6	42.1	3.2	36.6	100.9	34.6
8	512*340*6	1.5	2.6	5.8	1.2	6.3	15.6	5.9
9	1600*1200*7	10.6	19.7	42.4	3.0	54.0	69.3	35.5
10	800*600*7	2.9	4.8	10.5	1.5	13.7	17	9.8
Avg/Set		4.7	8.4	20.0	1.8	22.9	40.8	17.4
Rank		2	3	5	1	6	7	4

Table 3.3: Comparison of computation time (in second).

ordering of the fusion results are randomized for a fair comparison.

	[3]	[4]	[5]	[6]	[8]	[9]	ours
[3]	0	85	62	74	64	21	14
[4]	15	0	19	21	24	13	11
[5]	38	81	0	76	59	17	14
[6]	26	79	24	0	37	15	10
[8]	36	76	41	63	0	23	13
[9]	79	87	83	85	77	0	23
ours	86	89	86	90	87	77	0

Table 3.4: Preference matrix.

The preference matrix is concluded in Table 3.4. In this matrix, each element $e_{i,j}$ means there are $e_{i,j}$ votes prefer the fusion algorithm in i -th row to that in j -th column. From Table 3.4, it can be observed that more people prefer the proposed method than the others. To show the relative ranking of the seven exposure fusion algorithms, we employ the Bradley-Terry model [102] to analyse the subjective results. With the Bradley-Terry model, a relative score of the seven exposure fusion algorithms can be computed and plotted in Fig. 3.16. It can be observed that the proposed fusion algorithm out-performs the others statistically.

3.5 Conclusions

A detail-enhanced multi-scale fusion algorithm is proposed in this chapter to directly acquire an HDR-like image from a bracket of differently exposed LDR images. With the proposed WGIF-based fusion, a more informative and perceptually appealing LDR image can be merged without introducing unwanted artifacts. In addition, an efficient detail extraction module is proposed to enhance the visual quality of the

Figure 3.16: The BT scores of the seven fusion algorithms. The scores indicate the relative ranking where a higher score stands for a higher rank.

intermediate fusion result by using an exposure-aware structure tensor. Experimental results show that the proposed detail-enhanced fusion algorithm produce images with better visual quality than other stat-of-the-arts.

We have addressed the backward-compatible HDR acquisition problem in this chapter. In next chapter, we will introduce our backward-compatible solution in HDR image display stage for visualization of already acquired HDR images on conventional LDR displays.

Chapter 4

An Efficient Saliency-aware Local Tone Mapping Operator for High Dynamic Range Images

Because the dynamic ranges of HDR images are usually much higher than the conventional LDR display, they are not directly viewable on LDR displays. In this chapter, we propose a backward-compatible solution for visualization of HDR images on LDR displays. Our algorithm is an LTMO which is developed under the base-detail separation based tone mapping framework. Taking both the overall brightness and contrast loss into consideration, we formulate a novel optimization problem to obtain a better GTMC which is used for the dynamic range compression of the base layer in our LTMO. Inspired by the fact that the saliency cue is useful to obtain perceptual better results. We propose to incorporate the saliency cue in our LTMO to further improve the visual quality of the tone mapping result.

The remaining part of this chapter is organized as follows. In Section 4.1, we discuss the current tone mapping solutions and briefly explain how we come up with the new LTMO. The steps of how we obtain an optimal GTMC and incorporate the saliency cue in both base layer dynamic range compression and selective detail layer extraction

are detailed in Sections 4.2 and 4.3. Experimental results are presented in Section 4.4 and concluding remarks are given in Section 4.5.

4.1 Introduction

As extensively reviewed in Chapter 2, there are mainly two kinds of TMOs, namely the GTMCs and LTMOs, for visualization of the HDR image on conventional LDR displays. The GTMCs employ spatially-invariant mapping functions for dynamic range compression and produce naturally looks with little fine-scale details. The LTMOs modify pixels according to their local features and manage to preserve local contrast and fine-scale details better. However, the spatially-varying operations may introduce artifacts and sometimes also change the brightness which results in a poor visual appearance.

The LTMO in [12] provide a convenient framework to extend the GTMCs into LTMOs by keeping the advantages of both methods and many LTMOs [10, 72, 73] are developed under this framework. In this framework, an HDR image is decomposed into a base layer and a detail layer. The dynamic range compression is applied on the base layer according to a selected GTMC and the detail layer is added back to preserve fine-scale details. Existing GTMCs such as [1, 66, 70] employ pre-defined mapping functions which are derived from HVS model or practices in photography. They cannot adapt to image content well and result in severe contrast loss because of the heavy compression in high and low tones. Other GTMCs such as [10, 13], which depend on the HDR histogram distribution, allocate the dynamic range more smartly but may not produce as perceptual bright tone mapping results as [1, 66, 70] because of the brightness shift caused by the biased histogram distribution. A new GTMC

is desired to keep the advantages of above two kinds of GTMCs and overcome the disadvantages.

Unlike the LTMOs discussed in Chapter 2, some works [103, 104] tried to incorporate low-level vision cues such as the saliency in developing LTMOs. The idea of introducing saliency into tone mapping is attractive because our HVS only processes part of the visual stimulus while leaving the rest relatively unprocessed [105, 106]. The saliency cue has already been adopted in a variety of image and video processing applications [107–113] to improve the perceptual quality. Thus it can also be expected to obtain perceptually better tone mapping result by incorporating such a visual attention cue. However, the saliency cue is exploited for local adjustment of the GTMC to increase the local contrast in [103, 104]. These works are only validated in a limited number of examples and local adjustment of the GTMC may introduce artifacts as in [104]. A more proper approach is needed to incorporate the valuable saliency cue into the tone mapping process without introducing artifacts.

Based on above two considerations, we propose an efficient saliency-aware LTMO under the edge-preserving decomposition framework in this chapter. We first propose a new optimization problem to obtain a better GTMC. Then we incorporate the saliency cue to further refine the GTMC and formulate a saliency-aware edge-preserving filter for better detail preservation in salient regions. Our new GTMC tries to reduce the contrast loss of the GTMCs in [1, 66, 70] and alleviate the brightness shift of the GTMCs in [10, 13]. Specifically, we obtain our GTMC by solving a contrast loss minimization problem with a brightness preserving constrain. Because the HVS assign higher visual importance to salient regions, we also propose to introduce the saliency cue into the development of the GTMC to further improve the contrast preservation in visually important regions. In addition to the GTMC, we also propose

a selective base-detail separation for better detail preservation in salient regions. Note that in the works of edge-preserving filters [11, 17, 41, 76], the main purpose is to achieve better base-detail separation. These filters are proposed without considering the relative importance of different regions in the HDR image. However, the HVS assigns visual importance according to the visual attention model [105, 106]. The visual experience is expected to be improved when the details of salient regions are preserved better. This inspires us to refine the filters for tone mapping by introducing a saliency-aware weight in the filter formulation. In next two sections, we will introduce the formulation of our new GTMC and how we incorporate the saliency cue in our LTMO in details.

4.2 Global Tone Mapping Curve

In this section, our GTMC will be introduced. We start with a histogram-based contrast loss minimization tone mapping problem. Then we reformulate it by imposing a new brightness constraint to avoid possible brightness shift of the histogram-based method. Like previous TMOs [1, 70, 75], the dynamic range compression is applied on the luminance channel only. The other color channels are computed according to the adjusted luminance.

Denote the luminance of the HDR and LDR images as L^h and L^l and their log luminance as l^h and l^l , respectively. In our implementation, the base of the logarithm is 10. The global tone mapping problem can be defined as designing a monotonically increasing function f between L^h and L^l or l^h and l^l . Note that the contrast between adjacent pixels can be easily modelled as the differences in logarithmic domain, the histogram-based global tone mapping operators [10, 13] usually modelled

f as a function between l^h and l^l .

Denote the contrast of the input HDR and the output LDR as G^h and G^l respectively. By modeling the tone mapping curve as a piece-wise linear function as shown in Fig.

Figure 4.1: The tone mapping function f is modelled as a piecewise linear function from log HDR luminance l^h to log LDR luminance l^l .

4.1, the contrast loss introduced by tone mapping can be easily computed as in [10]:

$$\begin{aligned}
 \mathbb{E}[\|G^h - G^l\|_2^2] &\approx \sum_{i=1}^N p(l_i^h) (G^h(p) - k_i G^h(p))^2 \\
 &= \sum_{i=1}^N p(l_i^h) (1 - k_i)^2 \sum_{G^h} (G^h)^2 \\
 &\propto \sum_{i=1}^N p(l_i^h) (1 - k_i)^2
 \end{aligned} \tag{4.1}$$

where l_i^h is the median log luminance of the i -th bin and $p(l_i^h)$ is the probability of the i -th bin which is calculated from the histogram of the HDR image. N is the total number of bins which is usually selected as 30 in our implementation. k_i is the slope

of the i -th bin which is defined as

$$k_i = \frac{l_{i+1}^l - l_i^l}{\delta} \quad (4.2)$$

where δ is the length of the bin.

By minimizing the contrast loss in Eq. (4.1), the slopes of each bins can be calculated. The formation in Eq. (4.1) enables adaptive adjustment of tone curve according to the HDR histogram. However, unlike the famous histogram equalization approach, the GTMC in [10] reproduces contrast more conservatively. It forces the tone curve slopes to 1 instead of infinity for bins with very high probability. This avoids over-enhancement of bins with very high probability.

The GTMC in [10] performs well in contrast preservation. However, a possible drawback of this kind of methods is that the brightness is not explicitly considered. If the histogram is highly biased in high tones, the GTMC in [10] may produce underexposed LDR images because of brightness shift which is caused by accumulation of small slopes in low tones. One example of the possible underexposed appearances is shown in Fig. 4.2(b).

To address the brightness shift problem, we borrow ideas from the global tone mapping algorithm in [1]. The GTMC in [1] is a pre-defined function derived from best practices in photography. The key idea of the GTMC in [1] is to map the mean luminance of the HDR to the middle brightness of the LDR output and the GTMC defined in [1] is

$$L^l(p) = \frac{\hat{L}^h(p)(1 + \frac{\hat{L}^h(p)}{(\hat{L}_{white}^h)^2})}{1 + \hat{L}^h(p)} \quad (4.3)$$

where $\hat{L}^h$ is the normalized luminance which is defined as $\hat{L}^h(p) = \frac{\alpha}{L^h} L^h(p)$ and

$\bar{L}^h = 10^{\bar{l}^h}$ where $\bar{l}^h$ is the mean log luminance of the HDR image. $\hat{L}_{white}^h$ is a user-specific parameter to control the burn out and usually it is set as infinity in HDR tone mapping. α is a key parameter which determines the overall brightness of the tone mapping result. The parameter α is computed as

$$\alpha = 0.18 \times 4^{\left(\frac{2\bar{l}^h - l_m^h - l_M^h}{l_M^h - l_m^h}\right)} \quad (4.4)$$

where l_m^h and l_M^h are the minimal and maximum of the log luminance as in [114].

The GTMC in [1] produces a good overall brightness. However, the tone mapping function is a pre-defined ‘‘S’’ shape function. The dynamic range is allocated empirically for each bin in Fig. 4.1 even though the probability of that bin is very low. Note that the ‘‘S’’ shape function preserves middle tone well while compresses low and high tones heavily. This empirical tone mapping may result in heavy contrast loss in high and low tones as shown in Fig. 4.2(a).

To address the problems faced by the GTMCs in [1] and [10], we propose to take both the brightness and contrast loss into consideration. Suppose that the mean log luminance of the HDR image $\bar{l}^h$ falls in the m -th bin, we formulate our GTMC as

$$\arg \min_{k_1 \dots k_N} \sum_{i=1}^N p(l_i^h) (1 - k_i)^2 + \gamma \left(\delta \cdot \sum_{i=m+1}^N k_i - dr_{high} \right)^2 + \eta \sum_{i=2}^N (k_i - k_{i-1})^2 \quad (4.5)$$

subject to:

$$\begin{aligned} k_i &\geq 0 \text{ for } i = 1 \dots N \\ \sum_{i=1}^N k_i \cdot \delta &\leq dr. \end{aligned} \quad (4.6)$$

where dr is the dynamic range which is $\log_{10}(256)$ in our case. dr_{high} is the dynamic range allocated to HDR pixels whose luminance is larger than $\bar{L}^h$ and it is calculated

as $dr_{high} = l^M - \bar{l}$, where $\bar{l}$ is the log of the LDR luminance computed using $\bar{L}^h$ in Eq. (4.3) and l^M is 0 because the largest HDR pixels is mapped to 1 in Eq. (4.3).

(a)

(b)

(c)

(d)

Figure 4.2: Comparison among the GTMCs in [1], [10] and ours. (a) GTMC in [1]. (b) GTMC in [10]. (c) Our GTMC in Eq. (4.5). (d) Histogram of the HDR image.

The first term of Eq. (4.5) accounts for contrast loss caused by the tone mapping while the second term counts the estimated brightness shift of the middle tone. The third term is introduced to ensure the smoothness of the GTMC. The constraints in Eq. (4.6) ensures that the tone curve increases monotonically and will not exceed the target dynamic range. In our experiments, we find that without the third smoothness term, quantization artifacts may occur as shown in Fig. 4.3(a). In our implementations, we use $\eta = 0.01$ for all cases.

(a)

(b)

Figure 4.3: (a) without the third smoothness term in Eq. (4.5). (b) with the smoothness term.

The parameter γ in Eq. (4.5) controls the relative importance of the brightness shift term. If $\gamma = 0$, the tone curve is determined by the histogram only. If $\gamma = \infty$, the tone curve is forced to across the estimated average brightness $\bar{l}$. The effect of choosing

different γ is shown in Fig. 4.4. We have tested $\gamma \in \{0, 0.01, 0.04\}$ for tone mapping of the sample HDR image in Fig. 4.2. It can be observed from Fig. 4.4 that for all γ , the derived tone curve adapts to the HDR histogram. As γ grows, the middle tone of the tone mapping curve derived from Eq. (4.5) approaches the curve pre-defined as in [1]. Since the GTMC in [1] produces a good overall brightness, we choose $\gamma = 0.04$ in our implementation because this value gives similar overall brightness as the GTMC in [1].

Figure 4.4: Tone mapping curves for different γ in Eq. (4.5).

Once the slope of each segment of the tone mapping curve is determined according to Eq. (4.5), the luminance can be retargetted using the curve. For color channels, we follow the same color mapping in [115] to keep the linearity between the color channels and the tone-mapped luminance. Suppose the HDR input $l^h(p)$ is in the i -th bin. The color tone mapping for each color channel is defined as

$$C^l(p) = \left(\left(\frac{C^h(p)}{L^h(p)} - 1 \right) s(k_i) + 1 \right) L^l(p) \quad (4.7)$$

where C^h and C^l are the RGB color channels of the HDR and tone-mapped LDR

images. $s(k_i)$ is determined by the slope of i -th segment of the tone mapping curve as

$$s(k_i) = \frac{(1 + a_1)k_i^{a_2}}{1 + a_1k_i^{a_2}}. \quad (4.8)$$

The constants a_1 and a_2 are defined as 2.3892 and 0.8552 as in [115].

One preliminary comparison between the GTMCs in [1, 10] and ours is shown in Fig. 4.2. Because of the brightness preserving term, the tone curve of our GTMC is forced to map middle tone of the HDR image to middle brightness of the LDR image which results in a good overall brightness in the tone-mapped image in Fig. 4.2(c). In addition, because of the contrast loss term, our tone curve adapts to the HDR histogram which allows our GTMC to better preserve contrast in bright regions in Fig. 4.2(c), i.e. the sky regions in the middle of the image.

4.3 Saliency-aware Local Tone Mapping Operator

4.3.1 Saliency-aware Global Tone Mapping Curve

As shown in the previous section, our GTMC benefits from the new optimization problem in Eq. (4.5) and produces images of high visual quality. However, both the contrast measure and brightness estimation in previous formulation are calculated without consideration of visual importance. As discussed in [103, 104], saliency is quite useful in producing perceptually good tone mapping results. As such, we propose to incorporate saliency-aware weighting into our GTMC in (4.5) to further improve its visual quality.

There are a lot of saliency detection models [15, 105, 116–118]. Because of the effi-

(a)

(b)

(c)

(d)

Figure 4.5: Comparison among different saliency-aware cues in our GTMC refinement. (a) Our proposed GTMC in Eq. (4.5). (b) Using the saliency histogram only. (c) Using the salient-aware brightness estimation only. (d) Using both saliency histogram and the salient-aware brightness estimation.

ciency, the model in [15] is adopted to produce the saliency map. Note that the original model works for LDR images. We apply logarithmic transform to pre-process the HDR image before generating the saliency map. By such transformation, the contrast can be computed using spatial differences. With the saliency map $S(p)$ generated by model [15], we define a saliency-aware weight as $\omega_s(p) = S(p)/\bar{S}$ where $\bar{S}$ is the average value of the saliency map.

We have tried two modifications of our previously proposed GTMC in Eq. (4.5). Firstly, we calculate a saliency histogram instead of the intensity histogram based on the consideration that the contrast loss in less salient regions is less important than that in salient regions. With the saliency histogram, a saliency probability $\tilde{p}(l_i^h)$ for each bin can be computed. Secondly, we exploit the saliency weight and estimate the log luminance of the HDR image as

$$\tilde{l}^h = \frac{\sum_p \omega_s(p) l^h(p)}{\sum_p \omega_s(p)}. \quad (4.9)$$

Because global tone mapping operators usually reproduce the middle tone well. With the weighted average $\tilde{l}^h$, the salient regions are more likely to fall in the middle tone as such it might improve the visual quality of the salient regions.

The tone mapping results of using the two modifications are shown in Fig. 4.5. Fig. 4.5(a) is the tone mapping result generated by our GTMC in Eq. (4.5). By introducing the saliency histogram, the contrast of salient regions such as the man and trees outdoor are preserved better by sacrificing less salient contrast such as the light and shadows in the floor and ceiling as shown in 4.5(b). With the saliency-aware brightness estimation, salient regions such as the man wearing glasses are closer to middle tone which results in slightly better visual quality in these regions as in

Fig. 4.5(c). However, the average brightness is estimated using all pixels. The improvement is not always guaranteed for all regions. In other words, the salient regions outside in 4.5(c) become too bright. With both the saliency histogram and saliency-aware brightness estimation, the result in 4.5(d) is of the best visual quality. However, comparing Figs. 4.5(b) and 4.5(d), it can be observed that the improvement of the shifted brightness estimation is marginal. As such we only use the saliency histogram in our implementations.

4.3.2 Saliency-aware Detail extraction

In this section, our saliency-aware filter, which is designed for better detail preservation in salient regions, will be introduced. Previous research on edge-preserving filters [11] show that the fast weighted least square (FWLS) based filter is a good candidate for real-time applications. So we select it as our baseline.

According to [11], an edge-preserving smoothed version of the raw image can be obtained by solving the following optimization problem:

$$\operatorname{argmin}_{b^h} \sum_p \{(b^h(p) - l^h(p))^2 + \lambda \sum_{q \in \Omega(p)} \omega_{p,q} (b^h(p) - b^h(q))^2\} \quad (4.10)$$

where b^h is the base layer of the HDR image. $\Omega(p)$ is a set consists of the two pixels on the right and bottom of the pixel p . The pairwise weight $\omega_{p,q}$ is defined as

$$\omega_{p,q} = (|l^h(p) - l^h(q)|^\beta + \epsilon)^{-1} \quad (4.11)$$

where β is selected as 1.2 and ϵ is selected as 0.0001 as suggested in [17]. The first term of the objective function in Eq. (4.10) ensures b^h to be close to the HDR data l^h

while the second term ensures the smoothness of b^h . When λ is small, the first term dominates and results in a sharp b^h . When λ is large, the second term dominates which results in a smoother b^h . Typically, $\lambda = 0.25$ gives a good base-detail separation in the HDR tone mapping application.

By introducing the saliency weight $\omega_s(p)$, the saliency-aware detail extraction can be formulated as

$$\operatorname{argmin}_{b^h} \sum_p \{ (b^h(p) - l^h(p))^2 + \lambda \cdot \omega_s(p) \sum_{q \in N(p)} \omega_{p,q} (b^h(p) - b^h(q))^2 \}. \quad (4.12)$$

In this formulation, when pixel p is in salient regions, the second term dominates and the base layer is forced to be smooth as more details are extracted. When p is in less salient regions, the smoothed base layer is closer to original signal and results in less details in these regions. However, the less salient regions are usually textureless flat regions which contains very few details, thus the decrease of details in these regions is negligible.

As discussed in [17], the closed form solution of Eq. (4.12) needs to compute an inverse of a large Laplacian matrix which is quite time-consuming. Recent work in [11] shows that the original optimization problem in [17] can be greatly accelerated by approximating the original 2D smoother with sequential 1D smoothers. Based on this idea, we can also solve the optimization problem in Eq. (4.12) in a sequential smoothing manner.

Suppose the input HDR image is of resolution $H \times W$. When smoothing along the horizontal direction, the optimization problem for each row becomes

$$\operatorname{argmin}_{b^h} \sum_{x=1}^W \{ (b^h(x) - l^h(x))^2 + \lambda_k \omega_s(x) \cdot \omega_{x,x+1} (b^h(x) - b^h(x+1))^2 \} \quad (4.13)$$

where x is the horizontal index of the pixel. Note that b^h here is a row vector and the vertical index is ignored in Eq. (4.13) for simplicity. Parameter $\lambda_k = c_k \cdot \lambda$ where

the scaling factor c_k is introduced to reduce the amount of spatial smoothing for each iteration as discussed in [11]. c_k is computed as

$$c_k = \frac{3}{2} \frac{4^{T-k}}{4^T - 1}. \quad (4.14)$$

The closed form solution of Eq. (4.13) in vector form is

$$b^h = (I + \lambda_k A)^{-1} l^h, \quad (4.15)$$

where the $W \times W$ matrix A is

$$A(m, n) = \begin{cases} -\omega_s(m-1) \times \omega_{m,n} & \text{if } n = m-1 \\ -\omega_s(m) \times \omega_{m,n} & \text{if } n = m+1 \\ A(m, m-1) + A(m, m+1) & \text{if } n = m \\ 0 & \text{otherwise} \end{cases} \quad (4.16)$$

Because only one row is considered in Eq. (4.13), the dimension of the Laplacian matrix $(I + \lambda_k A)$ is W . Comparing to the closed form solution of Eq. (4.12) which relies on the inverse of a Laplacian matrix of dimension HW , the solution of Eq. (4.15) can be efficiently computed because it only requires computing the inverse of a lower dimension tri-diagonal matrix. In addition, the optimization problem of each row is formulated independently, a parallel computing scheme can be employed to further reduce the computational time.

Once smoothing along horizontal direction is completed, a followed smoothing along vertical direction is performed in the same manner to obtain the smoothing result for

Figure 4.6: Illustration of better detail reproduction by the proposed saliency-aware filter. (a) Our saliency-aware GTMC only. (b) Saliency map. (c) Our saliency-aware GTMC plus the edge-preserving filter in [11]. (d) Our saliency-aware GTMC and saliency-aware edge-preserving filter.

(a)

(b)

(c)

(d)

Figure 4.7: One more example for illustration of better detail reproduction by the proposed saliency-aware filter. The captions are same as figure 4.6.

the k -th iteration. As discussed in [11], a single iteration of the separating smoothing usually leads to the “streaking artifact”. Multiple iterations are needed for a satisfactory smoothing result. As suggested in [11], 3 iterations are enough for a good approximation of the 2D non-linear smoother.

The base-detail decomposition is the most time-consuming part in the proposed tone mapping algorithm. With the separable smoothing strategy, the time complexity of the decomposition is $O(n)$. Therefore the speed of the decomposition is competitive to other famous $O(n)$ edge-preserving filters such as the GF [41]. A more complete comparison between the separable smoothing and other edge-preserving filters can be found in [11]. In our unoptimized Matlab implementation, the separable smoothing takes about 1.2s to process a 1M pixel image. In addition, it takes about 0.3s to compute the saliency-aware weight and about 0.1s to solve the optimization problem in Eq. (4.5). Overall, it takes about 1.8s for the proposed method to process a 1M pixel image.

Some examples of using the proposed saliency-aware for better detail reproduction are shown in Figs. 4.6 and 4.7. In Fig. 4.6, the tone mapping result using our saliency-aware GTMC only is shown in Fig. 4.6(a). The tone mapping results by adding the edge-preserving filter in [11] and the proposed saliency-aware filter are shown in Figs. 4.6(c) and 4.6(d), respectively. Fig. 4.6(b) is the saliency map of the HDR input. Because all the results employ our proposed GTMC, the overall brightness and contrast of the tone-mapped images is quite good. However, the GTMC faces a fundamental problem in preserving details. The images in Fig. 4.6(a) is not as good as those in Figs. 4.6(c) and 4.6(d) in terms of detail preservation. As shown in the zoom-ins, the proposed saliency-aware filter preserves details better in the salient regions. Thus Fig. 4.6(d) is more preferred than Fig. 4.6(c) which does not take the

saliency cue into consideration. Similarly, the example in Fig. 4.7 also shows better detail preservation in salient regions.

4.4 Experimental Results

In this section, we first present some qualitative comparison results between our proposed LTMO and other tone mapping algorithms. Then we report the quantitative results using two metrics which are specially designed for quality assessment of tone-mapped images. Finally, the effectiveness of the proposed method is validated through a user study.

4.4.1 Qualitative Comparisons

We first compare our LTMO with the most closely-related works in [1, 10]. The results of [10] are from their project website and the results of [1] are produced by our implementation. The tone mapping results are shown in Figs. 4.8 and 4.9. From first row to last row, the images are tone-mapped by the GTMC in [1], our LTMO and the LTMO in [10], respectively. It can be observed that though the contrast are well-preserved by the LTMO in [10], the overall brightness of the tone mapping results are darker than the other two. This is caused by designing the tone curve depending on the image histogram only which may result in brightness shift. The results of [1] are perceptually bright. However, the brightest regions are clipped because the GTMC in [1] allocates the dynamic range without consideration of image statistics. For bins with very few pixels, the GTMC in [1] still allocates some dynamic range according to the pre-defined tone curve. Our method finds a quite good balance

between the other two methods. Our results are perceptually bright while without too much clipping. Because our LTMO and the LTMO in [10] are developed under the base-detail separation framework, the details are preserved better than that of GTMC in [1] as shown in the zoom-ins. Also note that our LTMO takes the saliency cue into consideration, the contrast and details in salient regions like the table in Fig. 4.8, and the bulbs and the woman in the mirror in Fig. 4.9 are preserved much better. The saliency maps of these two examples can be found in Fig. 4.12.

In addition to comparison with these two methods, we also compare our TMO with other state-of-the-art approaches. Specifically, we compare our method with the bilTMO [12], the disTMO [13] and the perTMO [14]. The bilTMO first introduces the edge-preserving bilateral filter into tone mapping and produces results with good detail reproduction. The disTMO is a global TMO which tries to find a tone mapping curve that minimizes the difference of HVS response between viewing the HDR images on HDR display and viewing the tone-mapped image on the LDR display. The perTMO is a local TMO which operates in perceptually linear gradient domain. It compresses the contrast in the gradient domain and then reconstructs the tone mapping result by solving a large scale poisson equation. The three TMOs are included in the open source pfstools [119] software. In our experiment, we use the pfstools with default parameters for the three TMOs. Some tone mapping results using our TMO and the three mentioned algorithms are given in Figs. 4.10 and 4.11.

First of all, it is obvious to find that the brightness of Figs. 4.10(c) and 4.11(c) is darker than the others. The reason is that the perTMO works in gradient domain and the brightness of image cannot be explicitly controlled. The contrast of the tone mapping results by the disTMO and our TMO, which are shown in Figs. 4.10(b), 4.11(b) and Figs. 4.10(d), 4.11(d), is better than the others because both the disTMO

Figure 4.8: Comparisons among our LTMO and tone mapping algorithms in [1] and [10]. From first row to last row are the tone mapping results using the GTMC in [1], our LTMO and the LTMO in [10].

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.9: Comparisons among our LTMO and tone mapping algorithms in [1] and [10]. From first row to last row are the tone mapping results using the GTMC in [1], our LTMO and the LTMO in [10].

Figure 4.10: Comparison of other TMOs and the proposed method. (a) bilTMO [12]. (b) distTMO [13]. (c) perTMO [14]. (d) the proposed LTMO.

and ours take the histogram of the HDR image into consideration and find a good tone mapping curve which minimizes the contrast loss. As shown in the zoom-ins, the bilTMO, perTMO and our TMO reproduce details well. Because the disTMO is a global TMO, it faces the fundamental problem in preserving the details which results in a slightly blurring appearance as shown in the zooms-ins. Among the bilTMO, the perTMO and our TMO, the perTMO preserves most details. However, because of the quite dark appearance, the tone mapping results looks very unnatural which degrades the visual quality of the perTMO. The bilTMO employs a simple linear scaling function for dynamic range compression. This tone mapping curve cannot reproduce contrast as good as ours and the disTMO. It results in many clippings as shown in Figs. 4.10(a) and 4.11(a). Our method takes both the brightness and contrast into consideration and produces perceptually bright appearance with proper contrast preservation. In addition, our saliency-aware detail extraction reproduce details well especially in salient regions, i.e., the book and the napkin in 4.10(c) and the chair in Fig. 4.11(c). The overall visual quality of our method is better than the others. The saliency maps of these two examples can be found in Fig. 4.12.

4.4.2 Quantitative Comparisons

To further validate the effectiveness of the proposed TMO, we evaluate our TMO using two objective metrics. We employ the TMQI [18] and the FSITM [19] which are designed for evaluating the quality of tone mapping results in our study. The TMQI judges the quality of tone-mapped images from the structural fidelity TMQI-S and statistical naturalness TMQI-N perspectives. The TMQI-S component is calculated by modifying the SSIM [63] and the TMQI-N component is modeled statistically from natural LDR images. Their quality index TMQI-Q is defined as a combination of the

Figure 4.11: Comparison of other TMOs and the proposed method. (a) bilTMO [12]. (b) disTMO [13]. (c) perTMO [14]. (d) the proposed LTMO.

(a)

(b)

(c)

(d)

Figure 4.12: Saliency maps of Figs. 4.8 - 4.11.

two criteria. The index TMQI-Q is a score between 0 and 1 with 0 stands for the worst quality while 1 for the best. The FSITM combines a similarity measurement of phase-derived features and the TMQI score to provide better assessment of tone mapping results [19]. For each color channel, the FSITM gives a score which ranges from 0 as the worst and 1 as the best.

Method	TMQI-Q	(TMQI-S, TMQI-N)
bilTMO [12]	0.8559	(0.8230, 0.4111)
disTMO [13]	0.8851	(0.8766, 0.4848)
perTMO[14]	0.8669	(0.8871, 0.3847)
reaTMO [10]	0.8529	(0.8660, 0.3360)
phoTMO [1]	0.8849	(0.8461, 0.5331)
our LTMO	0.8902	(0.8757, 0.5249)

Table 4.1: Average TMQI scores [18] over 30 images used in our experiment

We collect 30 HDR images for this objective study. Twenty of the 30 HDR images are from the public HDR database [16]. The others are from previous works [1, 12, 74] on HDR tone mapping. The quantitative evaluation results are given in Table 4.1 and Table 4.2. Six methods, namely bilTMO [12], disTMO [13], perTMO [14], reaTMO [10], phoTMO [1] and ours are evaluated in the objective study. Because the code of reaTMO is not publicly available, we have implemented this TMO using their global TMO and the FWLS [11]. We have also implemented the phoTMO because the phoTMO in the pfstools does not conform to the color correction in [115] and may produce severe color saturation problems.

From Table 4.1, it can be observed that our LTMO achieves the highest TMQI-Q score. Though the perTMO achieves the highest TMQI-S, the unnatural looking results in a quite low TMQI-N. Our LTMO finds a good balance between the TMQI-S and the TMQI-N. And the overall quality (TMQI-Q) of our method is the best.

Method	FSITM-R	FSITM-G	FSITM-B
bilTMO [12]	0.8510	0.8530	0.8522
disTMO [13]	0.8379	0.8373	0.8332
perTMO[14]	0.8236	0.8248	0.8213
reaTMO [10]	0.8635	0.8629	0.8595
phoTMO [1]	0.8692	0.8718	0.8655
our LTMO	0.8749	0.8755	0.8720

Table 4.2: Average FSITM scores [19] over 30 images used in our experiment

The results of using each color channels in the FSITM are shown in Table 4.2. It can also be observed that our LTMO outperforms the others in terms of this metric for all color channels.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	107	162	129	179	54
dis [13]	343	-	220	192	305	162
per [14]	288	230	-	177	263	173
rea [10]	321	258	273	-	307	168
pho [1]	271	145	187	143	-	59
ours	396	288	277	282	391	-

Table 4.3: Preference matrix of the six TMOs.

4.4.3 User Study

Because it is very challenging to evaluate the visual quality of tone-mapped images, we perform a subjective study to test the effectiveness of the proposed method. The subjective evaluation is conducted in a standardized room conforming to the ITU recommendations [120]. We employ the pairwise comparison method as in [78] in our study. The LDR display used in our experiment is 23-inch (HP Z23n) with maximum

luminance of 250 cd/m^2 . Each time, a pair of images with the same content tone-mapped by different TMOs are presented side by side in front of observers. And the observers are asked to choose which one they prefer more. During the experiment, the sequence and spatial ordering of tone-mapped images are randomized for a fair comparison.

Figure 4.13: The BT scores of the six TMOs. The scores indicate the relative ranking where a higher score stands for a higher rank.

We have tested 6 TMOs in our experiments. The numbers of participants and HDR images tested are 15 and 30, respectively. Thus we have $15 \times 30 = 450$ votes for each comparison between any two TMOs. With the pairwise comparison, a total preference matrix is obtained as shown in Table 4.3. In the preference matrix, each elements $e_{i,j}$ stands for the total number of votes that prefers the TMO in the i -th row than the TMO in the j -th column. The preference matrix for each subject can be found in Appendix B. From the last row of Table 4.3, it is obvious that our LTMO is preferred

than the others in the pairwise comparison. To show the relative ranking of the six TMOs, we employ the Bradley-Terry model [102] to analyse the subjective results. With the Bradley-Terry model, a relative score of the six TMOs can be computed and plotted in Fig. 4.13. It can be observed that our LTMO out-performs the others statistically.

4.4.4 Limitation

The performance of the proposed method depends on the success of saliency detection. However, saliency detection is such a challenging task that it is difficult to find a saliency detection model which works reliably for all images. One failure case is shown in Fig. 4.14(c). In this case, the saliency detection model [15] fail to detect the girl and cloth as salient objects. Therefore, our saliency-based tone mapping will not be able to protect the contrast and details in these regions. In such case, viewers may prefer the tone mapping results without the saliency cue. In our experiments, this failure of saliency detection rarely happens and employ a more reliable saliency detection model is supposed to address this problem.

4.5 Conclusion

An edge-preserving decomposition based local tone mapping operator is proposed in this chapter for visualization of HDR images on LDR displays. Compared to other stat-of-the-art TMOs, the proposed method produces perceptually bright result with reduced contrast loss especially in salient regions. In addition, the proposed method has a better ability in reproducing details of salient regions. These two improvements

(a)

(b)

(c)

(d)

(e)

Figure 4.14: One failure case in which the saliency detection model [15] fail to produce reliable saliency map. (a) Tone mapping with the saliency cue. (b) Tone mapping without the saliency cue. (c) Saliency map. (d) and (e) are zoom-ins of (a) and (b).

result in a better visual quality of the tone mapping result. Substantial experimental results validate the improved visual appearance of the proposed method.

Till now, we have introduced our backward-compatible solutions in HDR acquisition and display stages. In next chapter, we will introduce our backward-compatible HDR compression algorithm for efficient distribution of HDR images to both HDR and LDR consumers.

Chapter 5

Local Inverse Tone Mapping for Scalable High Dynamic Range Image Coding

Though the HDR image provides superior visual appearance, it is not widely accepted currently because the majority of the display market still depend on the 8-bit LDR image format. To promote HDR images in this transition phase from LDR to HDR images, the scalable compression algorithms are desired for distribution of HDR images to both HDR and LDR consumers. In this chapter, we propose a two-layer local inverse tone mapping to improve the coding efficiency of the scalable HDR image compression algorithms.

The rest of the chapter is outlined as follows. The scalable HDR image compression is introduced in Section 5.1. Our two-layer local inverse tone mapping is explained in details in Section 5.2. Experimental results are shown in Section 5.3 and concluding remarks are given in Section 5.4.

5.1 Introduction

Scalable coding schemes [121, 122] provide a possible backward compatible solution to distribute a high dynamic range (HDR) image in the transition phase from low dynamic range (LDR) displays to HDR displays. The workflow of this scalable coding is shown in Fig. 5.1. As shown in the workflow, two bitstreams are transmitted to the decoder. The base layer bitstream contains the tone mapping result of the LDR image which is obtained by using an arbitrary tone mapping operator (TMO) according to a user's preference [10, 13, 123]. The enhancement layer bit-stream contains the meta data to recover the HDR image, such as the parameters of the inverse TMO (iTMO) and the residual of the HDR image and the inverse tone-mapped LDR image. In the decoder side, LDR consumers only decode the base layer bitstream to view the tone mapping result on the LDR displays while HDR consumers decode all bitstreams to reconstruct the HDR image transmitted and reproduce it on the HDR displays. Under such scalable coding framework, the TMO determines the appearance of the base layer which is intended for the LDR consumers while the iTMO affects the bit-rate for the enhancement layer. A more accurate iTMO requires fewer bits to encode the residual.

Currently, several scalable coding schemes have been proposed based on the global TMO and its iTMO because of the efficiency and reversibility of global TMOs [84, 85, 88]. These methods are almost real-time and achieve a low bit-rate for the enhancement layer. However, the global TMOs [1, 64] cannot produce as good looking results as local TMOs [11, 12, 14, 17, 41] in terms of detail preservation. Thus from such consideration, the global TMO based scalable coding schemes might not be friendly to LDR consumers. Local TMO based scalable coding schemes are demanded to

provide better user experience for LDR consumers. Unfortunately, it is very difficult

Figure 5.1: Workflow of scalable coding for HDR images.

to estimate the local iTMO because a local TMO is a function of both luminance and local contrast. Directly modeling a local iTMO as a mapping from LDR luminance to HDR luminance results in inaccurate estimations [84, 89]. Assuming that pixels within a small neighborhood have the same local contrast, a local iTMO can

be simplified as a mapping function block-wisely. An example is reported in [87], by modelling a local iTMO as a number of global iTMOs which are estimated using the pair of HDR and its tone-mapped LDR patches for each block. Actually, it is noted that if the local contrast for each pixel of the tone-mapped image is nearly zero, the local iTMO can be simplified as a mapping function globally. Following this idea, we believe that an edge-preserving smoothing technique can be employed to obtain a piece-wise smooth version of the tone-mapped image for developing a local iTMO globally. However, to the best of our knowledge, no work has been done on estimating a local iTMO using such technique.

In this chapter, we propose a two-layer local iTMO by using a fast edge-preserving decomposition technique. The proposed iTMO is inspired by the local TMOs [11, 12, 17, 41, 73, 124] which are designed based on an edge-preserving decomposition of the HDR image. In these local TMOs, an HDR image is decomposed into a base layer and a detail layer via an edge-preserving smoothing filter. The base layer which is piece-wise smooth and contains dynamic range information is compressed via a global TMO while the detail layer is preserved to keep local contrast. Thus it is reasonable to take the decomposition into consideration when a local iTMO is designed. Unlike the local TMO, the linearized LDR image instead of the HDR image is decomposed into a base layer and a detail layer in logarithmic domain by using the fast edge-preserving filter in [11] in the proposed method. The base layer of the HDR image is generated by subtracting the detail layer of the linearized LDR image from the HDR image in logarithmic domain. An iTMO function is estimated using the pair of base layers rather than the pair of HDR and LDR images themselves as in existing methods. The estimation of the iTMO function between the pair of base layers is similar to the estimation of camera response functions (CRFs) among multiple differently exposed

LDR images. Inspired by the method in [22] which shows that the CRF and the global tone mapping function share the same property of smooth and monotonous increasing, a novel quadratic optimization problem is formulated to estimate the iTMO function between the pair of base layers. It is worth noting that other existing CRF estimation methods such as [125, 126] are also applicable. With the iTMO function and the edge-preserving decomposition of the linearized LDR image, an HDR image can be estimated in the decoder side by using the decoded LDR image only. The proposed local iTMO is simple and is applicable to the compression of HDR images. In next section, we will introduce the proposed local iTMO. In latter experimental results section, we will demonstrate how the proposed iTMO is applied to improve the coding efficiency.

5.2 Two-layer Local Inverse Tone Mapping

As illustrated in Fig. 5.1, the two key components in the scalable coding are the iTMO function estimation and HDR image estimator. In this section, these two components will be introduced. Our work is inspired by the base-detail decomposition based local tone mapping operators. We decompose the decoded LDR image and estimate the iTMO function using the base layers of the decoded LDR image and HDR image. The iTMO function estimation process is shown in Fig. 5.2. Similarly, the HDR image estimator is also designed using the base-detail decomposition and the process is shown in Fig. 5.3.

5.2.1 Local Tone Mapping Recap

Denote the logarithmic of the luminance of the HDR image and the tone-mapped LDR image as Y and X , respectively. In the base-detail decomposition based local TMOs, Y is decomposed into a base layer Y_b and a detail layer Y_d as

$$Y = Y_b + Y_d. \quad (5.1)$$

The base layer Y_b is obtained by edge-preserving filters such as the ones in [11, 12, 17, 41]. It is piece-wise smooth and contains large-scale variations such as sharp edges and therefore dynamic range information. Thus it can be compressed to reduce the dynamic range. On the other hand, Y_d contains small-scale variations such as details and is kept to preserve local contrast. Based on such properties, we have

$$X_b = f(Y_b); \quad X_d = Y_d, \quad (5.2)$$

where X_b and X_d are the base layer and detail layer of the tone-mapped image, respectively, $f(\cdot)$ is a global TMO, for example the one in [1].

Under this edge-preserving decomposition based local tone mapping framework, the quality of decomposition could affect the tone mapping result significantly. Local edge-preserving filters such as bilateral filter (BF) [12] and guided image filter (GIF) [41, 95] cannot always preserve edges well so that halos could be produced [41]. Global edge-preserving filters such as weighted least square (WLS) [17] can produce a halo-free tone mapping result but it is very slow. The fast WLS (FWLS) [11], which is a fast approximation of WLS, produces a similar smoothing result of the WLS with a competitive speed to the BF [12] and the GIF [41]. Because of the fast speed and good

ability in preserving edges, the FWLS is selected for the base-detail layers separation in our implementation.

5.2.2 Generation of a Pair of Base Layers

The FWLS in [11] is adopted to decompose the linearized LDR image which is obtained by gamma correction with the value of gamma as 2. According to [17], the base layer X_b can be obtained by solving the following optimization problem:

$$\operatorname{argmin}_{X_b} \sum_p \{(X_b(p) - X(p))^2 + \lambda \sum_{q \in N(p)} \omega_{p,q} (X_b(p) - X_b(q))^2\} \quad (5.3)$$

where p is the location of the pixel, $N(p)$ is the set of the right and bottom pixels of pixel p and $\omega_{p,q}$ is defined as $\omega_{p,q} = (\|X(p) - X(q)\| + \epsilon)^{-\alpha}$ where ϵ is a small number to avoid division by 0. α controls the smoothness of the result and is selected as 1.2. Parameter λ determines the smoothness of X_b . A larger λ results in a smoother X_b . The FWLS [11] solves Eq. (5.3) in a separable manner. An intermediate smoothing result is first obtained by solving the optimization problem along horizontal direction. Then the intermediate result is smoothed along vertical direction. When smoothing along the horizontal direction, an optimization problem is formulated for each row. Denote the i -th row of the base layer and the LDR image as X_b^i and X^i respectively. The smoothing problem for i th row is formulated as

$$\operatorname{argmin}_{X_b^i} \sum_x \{(X_b^i(x) - X^i(x))^2 + \lambda_k \omega_{x,x+1} (X_b^i(x) - X_b^i(x+1))^2\} \quad (5.4)$$

where $\lambda_k = \frac{3}{2} \frac{4^{T-k}}{4^T - 1} \lambda$ is the parameter for k -th iteration. The WLS is accelerated because the closed-form solution of Eq. (5.4) is equivalent to solving a small scale lin-

ear system with a tri-diagonal coefficient matrix. In addition, since the optimization problem is formulated for each row, the fast WLS can be further accelerated using the parallel computing techniques. The formulation of the smoothing along vertical direction is similar to Eq. (5.4).

Figure 5.2: Workflow of iTMO estimation in encoder side.

5.2.3 Tone Curve Estimation

The iTMO function is estimated using the pair of base layers of the HDR image and its tone-mapped LDR image as shown in Fig. 5.2 in the encoder side. The LDR image is linearized by gamma correction. After decomposing the linearized LDR image into

a base layer and a detail layer using the FWLS filter, the detail layer of the LDR image is subtracted from the HDR image in logarithmic domain to generate the base layer of the HDR image:

$$Y_b = Y - X_d. \quad (5.5)$$

Because the global TMO in Eq. (5.2) adjusts HDR luminance using a monotonous increasing function, the inverse function is well-defined. It is worth noting that the CRFs among differently exposed LDR images are also monotonous increasing. It is expected that existing CRF estimation methods such as the one in [22] can be extended to estimate the iTMO function. Based on this observation, a new quadratic optimization problem is formulated as

$$E(G) = \sum_{i=1}^n \{w(X_b(i))(g(X_b(i)) - Y_b(i))^2\} + \mu \sum_{j=x_{min}+1}^{x_{max}-1} w(j)g''(j)^2 \quad (5.6)$$

where n is the number of pixels, i is the index of the pixel, $g(\cdot)$ is the global iTMO function, X_b is the base layer of the LDR image and Y_b is the base layer of the HDR image.

The first term ensures that the estimated image using the iTMO function is close to the original HDR data and the second term ensures the smoothness of the iTMO function. Weight w is introduced because the iTMO function may fit data poorly when near extremes. With $\mu = 0$, Eq. (5.6) may become ill-posed when the number of pixels in some bins is nearly zero. When μ is large, the resulting iTMO function is very smooth. Note that without w and the smoothness term, Eq. (5.6) becomes the same fitting problem as in [84].

We quantize X_b into 256 levels so that $x_{min} = 0$ and $x_{max} = 255$. Like the global

TMO, the global iTMO function can also be implemented using a look-up table. Denote the look-up table as a vector $G = [g(0) \ g(1) \ \cdots \ g(255)]^T$. The second order derivative $g''(j)$ is computed as

$$g''(j) = g(j-1) - 2g(j) + g(j+1). \quad (5.7)$$

$w(i)$, which is introduced to penalize near-saturated pixels, is defined as

$$w(i) = \begin{cases} i+1, i \in [0, 127] \\ 256-i, i \in [128, 255]. \end{cases} \quad (5.8)$$

Now we define the following three matrices: an $n * n$ diagonal matrix W where $W(i, i) = w(X_b(i))$, an $n * 256$ matrix H

$$h_{i,j} = \begin{cases} 1, if \ X_b(i) = j-1 \\ 0, otherwise \end{cases} \quad (5.9)$$

and a $254 * 256$ matrix M

$$M = \begin{bmatrix} 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \quad (5.10)$$

Then the matrix form of the energy function is represented as

$$E(G) = (HG - Y)^T W (HG - Y) + \mu (MG)^T \hat{W} MG \quad (5.11)$$

where $\hat{W} = \text{diag}\{w(1), w(2), \dots, w(254)\}$.

Let $\frac{\partial E(G)}{\partial G} = 0$. We have

$$(H^T W H + \mu M^T \hat{W} M)G = H^T W Y. \quad (5.12)$$

Denote H as $H = [h_{c_1} \ h_{c_2} \ \dots \ h_{c_{256}}]$, where h_{c_i} is the i th column of H . Because for each row of H , only one element is non-zero which is 1, thus

$$h_{c_i}^T * h_{c_j} = \begin{cases} n_{i-1}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \quad (5.13)$$

where n_{i-1} is the number of pixels with intensity $i - 1$. Therefore

$$H^T W H = \text{diag}\{w_0 * n_0, w_1 * n_1, \dots, w_{255} * n_{255}\} \quad (5.14)$$

where $w_i = w(i)$.

Then Eq. (5.12) is a linear system with a $256 * 256$ tri-diagonal coefficient matrix. It can be solved efficiently.

5.2.4 HDR Image Estimator

As shown in Fig. 5.1, the HDR image estimator is used to estimate the HDR image from the decoded LDR image and it exists in both encoder and decoder side. The workflow of the HDR image estimator is shown in Fig. 5.3. After the decomposition of linearized LDR image which is obtained by gamma correction with the value of gamma as 2, the LDR base layer is expanded with the iTMO function and combined

with the LDR detail layer of the decoded LDR image as

$$\hat{Y} = \exp(g(X_b) + X_d), \quad (5.15)$$

where $g(X_b)$ is the expanded LDR base layer using the iTMO function and $\hat{Y}$ is the luminance predictor of HDR image.

Figure 5.3: Workflow of HDR image estimation via the proposed two-layer iTMO.

Then the RGB color channels of the HDR image predictor can be computed as

$$\{\hat{R}, \hat{G}, \hat{B}\} = \{R_l, G_l, B_l\} \cdot \frac{\hat{Y}}{Y_l}, \quad (5.16)$$

where $\{R_l, G_l, B_l\}$ are the color channels of the linearized tone mapped image and Y_l is the corresponding luminance channel.

5.3 Experimental Results

In this section, the experimental results will be presented. We first introduce the HDR image set and the evaluation metric used in our experiment. Then we conduct a preliminary study to determine the best parameters in the local iTMO. Using the best parameters, we validate the effectiveness of the proposed iTMO by applying it in HDR image estimation without involvement of coding. Finally, we incorporate our local iTMO into the scalable JPEG-Xt standard to improve the coding efficiency when using LTMOs to generate tone-mapped base streams.

5.3.1 Dataset and Evaluation Protocol

The twenty HDR images for the experiments are from the publicly available dataset [16]. The dataset covers a variety of scenes and dynamic ranges. The tone-mapped LDR images of the dataset [16] are shown in Fig. 5.4 and the dynamic range of each HDR image is shown in Table 5.1.

Note that in our experiments, the quality of the inverse tone-mapped image which is of high dynamic range is predicted. In this scenario, famous metrics for LDR images

Figure 5.4: The 20 HDR images used in our experiments. The dataset is downloaded from [16].

such as PSNR, RMSE and SSIM are not applicable. In our experiment, the HDR-VDP-2 (version 2.2.1) [93] which is specially designed for comparing images of the same dynamic range is employed to evaluate the perceptual quality of the decoded HDR image objectively. By pooling the visual difference measure of each pixel, the HDR-VDP-2 gives an overall quality predictor “Q” which ranges from 0 to 100 (100 as the best quality). Recently, the performance of the HDR-VDP-2 “Q” predictor is calibrated and validated in [90] with substantial user studies. Thus it is suitable to use it to predict the quality of estimated HDR images.

index	min	max	dynamic range
1	-0.34	3.51	3.85
2	1.24	3.34	2.10
3	1.29	3.25	1.96
4	0.65	3.10	2.45
5	-0.62	3.51	4.13
6	-0.67	3.58	4.25
7	-0.46	3.37	3.83
8	-0.26	2.77	3.03
9	-0.43	3.44	3.87
10	-0.67	3.54	4.21
11	-0.79	3.55	4.34
12	-0.66	3.57	4.23
13	0.06	3.01	2.95
14	0.27	3.34	3.07
15	-0.21	3.31	3.52
16	-0.46	3.40	3.86
17	-0.77	3.54	4.31
18	0.16	3.08	2.92
19	-0.79	3.48	4.27
20	-0.14	2.97	3.11

Table 5.1: Statistics of the 20 HDR images. The minimal, maximum luminance and the dynamic range are in $\log_{10}$ units.

5.3.2 Parameter Selection

There are two parameters in our inverse tone mapping operator: μ in Eq. 5.6 and λ in Eq. 5.3. To see the effects of these two parameters, we conduct a preliminary study with 10 HDR images randomly selected from the dataset described before. In this parameter study, the global tone mapping operator in [1] is selected to generate global tone mapping results. The local tone mapping results are generated by extending the global tone mapping operator in [1] with the edge-preserving filter in [11]. The inverse tone mapping results are validated using the average HDR-VDP-Q score over the 10 images.

μ	0.01	0.1	1	10	100	1000
λ	0.1	0.5	1	2	4	8

Table 5.2: Choices of μ and λ in the parameter study.

We tested six possible choices of μ . The values of μ used are shown in the first row of Table. 5.2.

The result of varying parameter μ is shown in Fig. 5.5. Note that the x -axis is in logarithmic scale. It can be observed that the inverse tone mapping operator is not sensitive to choices of μ . The reason is that the μ is introduced to avoid ill-posed formulation of Eq. 5.6 when the number of pixels in some bins is nearly zero. Though larger μ results in a smoother iTMO function which results in better performance in the HDR image estimation as shown in Fig. 5.5, the gain is marginal (less than 0.05 average HDR-VDP-Q score). As such we simply select μ as 1 in all our other experiments.

The parameter λ controls the extent of base-detail separation. To validate its effect on our local iTMO development, we tested six possible values of λ in this parameter

Figure 5.5: Average HDR-VDP-Q scores over 10 HDR images in the parameter study with varying μ in Eq. 5.6.

study. The six values are shown in the second row of Table. 5.2. The result of using various λ is shown in Fig. 5.6. Unlike μ , the parameter λ affects the local iTMO. Too large or too small λ will not give a proper separation of the base and detail layers and will result in a lower average HDR-VDP-Q score as shown in Fig. 5.6. Our preliminary experiment showed that $\lambda = 1$ gives a good separation of the two layers and we use $\lambda = 1$ as our default parameter for all the remaining experiments.

5.3.3 Validation of Inverse Tone Mapping

In this section, we validate that the proposed two-layer iTMO can recover the HDR image accurately without involvement of coding. In the first step, we validate that the global iTMO function estimation as described in Eq. (5.6) is useful to inverse tone

Figure 5.6: Average HDR-VDP-Q scores over 10 HDR images in the parameter study with varying λ in Eq. 5.3.

map the global TMOs. In this experiment, the global TMO [1] is selected and inverse tone mapped using Eq. (5.6). The inverse tone mapping result is compared with the simple linear expansion, which multiplies linearized tone-mapped LDR image with a constant to approximate the absolute luminance of the LDR image when viewed on an HDR display, as shown in Fig. 5.7. It is obvious that for all images, the proposed global iTMO estimation outperforms the linear scaling.

The local TMO in [17] is selected to evaluate the effectiveness of the proposed two-layer iTMO on local TMOs. In addition to comparing with the linear scaling, the proposed two-layer iTMO is also compared with directly estimating the tone mapping function without the decomposition which corresponds to notation “w/o decomposition” in Fig. 5.8. The proposed two-layer iTMO is notated as “w/ decomposition” in Fig. 5.8. As shown in Fig. 5.8. The iTMO estimation with the two-layer decom-

Figure 5.7: Inverse tone mapping results for images tone-mapped using the global TMO in [1]. It is obvious that perceptually better HDR estimations are obtained with the proposed global iTMO estimation method.

position outperforms the linear expansion while directly estimating the inverse tone mapping function leads to worse perceptual quality. The reason is that the detail of the locally tone-mapped image, which is an additive noise in the global tone mapping function estimation, results in a less accurate mapping function. The average HDR-VDP-Q scores of the linear expansion, local iTMO estimation without decomposition and with decomposition are 61.05, 60.01 and 62.86, respectively.

5.3.4 Lossy Scalable Coding

We also verify the efficiency of the proposed scheme by applying it to scalable HDR image coding. In our implementation, we use the JPEG-XT reference software, which is implemented by the Dolby Laboratories Inc, as the base. In their implementation,

Figure 5.8: Inverse tone mapping results for images tone-mapped using the local TMO in [17]. The average HDR-VDP-Q scores of the linear expansion, local iTMO estimation without decomposition and with decomposition are 61.05, 60.01 and 62.86, respectively.

the residual image which is transmitted in the enhancement layer bitstream is computed as the difference between the HDR image and the decoded LDR image in logarithmic domain. We replace the decoded LDR image with the HDR image estimated from the decoded LDR image using the proposed iTMO when computing such residual image. Three experiments are conducted to verify the effectiveness of HDR image coding via the proposed iTMO.

The first experiment is designed to show that a better compression performance of global tone-mapped HDR image can be achieved with the HDR estimation proposed in Eq. (5.6). The HDR image is tone mapped with the global TMO in [1] and a corresponding inverse tone mapping function is obtained by solving Eq. (5.6). With the inverse tone mapping function, a color HDR image is estimated using Eq. (5.16).

In the JPEG-XT framework, the compression parameters q and Q control the quality of base layer and enhancement layer, separately. In our experiments, we test 8 bit rates in total, with q selected as 65 and 90, which corresponds to imperceptible and slightly annoying JPEG compression on base layer, respectively. When q is fixed, Q is selected as 10, 35, 65, 90 to imitate from heavy to slight compression on enhancement layer. The HDR-VDP-Q scores of these two cases are presented in Fig. 5.9 and Fig. 5.10. For each pair of q and Q , the average HDR-VDP-Q score is computed as the arithmetic mean of the HDR-VDP-Q scores of the 20 images. The average bit per pixel (bpp) is computed in a similar manner. Note that the bpp is computed using the total file size of the base image, the residual image and the meta data, i.e. the iTMO function. The error bar shows the standard variation of the HDR-VDP-Q score. From Fig. 5.9 and Fig. 5.10, it can be concluded that with similar bit rate, a perceptually better HDR image can be encoded and decoded via the global iTMO described in Eq. (5.6).

The second experiment is to show the effectiveness of the two layer local iTMO. In this experiment, the local TMO is designed using the edge-preserving filter in [17] to decompose the luminance of HDR image into two layers and adopting the global TMO in [1] to scale down the base layer. The encoding parameters are selected as the same as the previous experiment. The notation “w/o decomposition” in Figs. 5.11-5.14 stands for estimating iTMO function without the decomposition while “w/ decomposition” stands for estimating iTMO function with the decomposition. It can be observed from Figs. 5.11 and 5.12 that with a fixed q , the proposed method outperforms JPEG-XT as Q grows. This is because that our iTMO estimates the HDR more accurately, which results in less high frequency information in the residual image. So a smaller Q is needed to compress the enhancement layer.

Figure 5.9: Mean compression performance of using the proposed global iTMO. All images are tone-mapped with the global TMO in [1]. The quality parameter for the base layer is fixed as 90.

Figure 5.10: Mean compression performance of using the proposed global iTMO. All images are tone-mapped with the global TMO in [1]. The quality parameter for the base layer is fixed as 65.

Figure 5.11: Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [17]. The quality parameter for the base layer is fixed as 90.

Figure 5.12: Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [17]. The quality parameter for the base layer is fixed as 65.

Figure 5.13: Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [14]. The quality parameter for the base layer is fixed as 90.

Figure 5.14: Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [14]. The quality parameter for the base layer is fixed as 65.

Figure 5.15: Mean compression performance of using the proposed local iTMO. All images are tone-mapped with the local TMO in [14]. The quality parameter for the base layer is fixed as 30.

The third experiment is to show the robustness of the proposed local iTMO. The local TMO in [14] is employed to generate the tone-mapped LDR images with default parameters while the iTMO is designed by using the FWLS as in Section III. The local TMO [14] tries to preserve as much detail as possible so that it is very challenging to model the inverse of this local TMO. The encoding parameters are selected as the same as that in our previous experiments. It can be shown from Fig. 5.13 and Fig. 5.14 that though the improvement is not as large as tone mapping with the FWLS, JPEG-XT can benefit from the proposed two-layer local iTMO, especially when a higher perceptual quality of decoded HDR image is required. It should be mentioned that the coding gain could be expected to be increased if the entropy coding is re-designed based on the proposed iTMO. In addition, it can be expected that the proposed method is applicable to the piece-wise linear tone curves in [10, 13, 123]. Note that the enhancement layer is multiplied with the base layer to reconstruct the

HDR image. Parameter Q controls the quality of the enhancement layer. When Q is small (the cases when bpp is smaller than 3 and 2 in Figs. 5.13 and 5.14), there are a lot of noise in the enhancement layer. Therefore after multiplication with the base layer, the quality of the reconstructed HDR image may also contain much noise. That is why the improved compression performance is not always guaranteed when the bpp is very small.

In addition to above two settings for base layer compression, we also tested $q = 30$ in this third experiment. Note that when $q = 30$, very little details remain in the LDR base layer. Therefore, the performance of the iTMO with and without base-detail separation will be quite similar. As shown in figure 5.15, the compression efficiencies of the proposed method and JPEG-XT are quite close. Note that when $q = 30$, visible compression artifacts will appear in the LDR base layer. This setting is not friendly to LDR consumers.

5.4 Conclusion

In this chapter, we propose a two-layer local inverse tone mapping scheme to estimate the HDR image using its tone-mapped LDR version. The detail layer of the HDR image is estimated using the detail layer of the linearized LDR image while the base layer of the HDR image is estimated by expanding the dynamic range of the base layer of the linearized LDR image. We have adopted our two-layer local inverse tone mapping into scalable HDR image compression and the experimental results show that the proposed two-layer iTMO can estimate the HDR image accurately and improve the coding efficiency.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

HDR imaging techniques provide excellent visual experience and it is supposed to be next generation of imaging technology. However, the majority of consumers still rely on LDR display technology currently and this situation will last for a long time because of slow pavement in commercializing HDR displays. Considering that it is very important to provide backward-compatible solutions to LDR consumers to ensure gradual transition from LDR imaging techniques to HDR ones, in this thesis, three backward-compatible solutions in the HDR imaging pipeline developed by us have been presented. The contributions of this thesis can be summarized as follows:

- We propose a multi-scale exposure fusion algorithm to directly acquire an HDR-like LDR image from a bracket of differently exposed inputs without reconstruction of the HDR image. With our WGIF-based fusion strategy, we are able to obtain informative fusion results which are halo-free and preserve information in the brightest and darkest regions of the scene. In addition, we provide an efficient detail extraction module to preserve fine-scale details in the HDR scene which further improves the visual quality of the fusion result. We have com-

pared our algorithm with other state-of-the-art algorithms both qualitatively and quantitatively. The comparison results confirm that our algorithm out-performs the others in terms of visual quality.

- We develop a saliency-aware local tone mapping algorithm under the base-detail separation tone mapping framework. With our new formulation for the GTMC, the obtained tone mapping results are perceptually bright and preserve the contrast of the HDR image well. By additionally incorporating the saliency cue into both the GTMC formation and detail extraction, we manage to achieve better contrast and detail preservation in salient regions, which results in a better perceptual quality. We have conducted both objective and subjective evaluations to validate the effectiveness of the proposed algorithm. Both experimental results show that our tone mapping results are of high visual quality and is more preferred.
- We propose a two-layer local inverse tone mapping to estimate an HDR image from its tone-mapped version. We have adopted our iTMO into the JPEG-Xt standard, which is proposed for scalable HDR image compression. Our experimental results show that the coding efficiency is improved when using our iTMO to estimate the HDR image in the scalable compression framework.

6.2 Recommendation of Future Work

There are several interesting topics for future studies. In addition to the extended dynamic range, the HDR technologies are also associated with the wide color gamut. In this thesis, we mainly focus on processing the luminance of the HDR image. How to

deal with color channels, such as how to correct color distortions caused by exposure fusion and tone mapping [115] or how to compress color channels which are defined in WCG [127] will be interesting research topics.

In the development of the exposure fusion, tone mapping and compression algorithms, several perceptual objective quality assessment metrics [18, 61, 93] have been involved. In our work, these metrics are employed for performance evaluation only. Except for the objective evaluation, these metrics give insights on how to develop better algorithms for HDR images. And it is also possible to optimize HDR image processing algorithms in terms of these metrics. How to develop such automatically tuned algorithms is worth investigation in the future.

In our exposure fusion algorithm in Chapter 3, we assume that differently-exposed LDR images are captured using a fixed camera and the scene is still. However, in practice, both camera motions and object motions may exist. These motions will result in ghosting artifacts in the final fusion result. How to extend our work to cope with these motions and produce ghost-free fusion results will be an interesting problem for future research.

In this thesis, we mainly focus on HDR images. Recently, as the development of HDR capture devices, both HDR videos [128] and stereoscopic HDR videos [129] are becoming available. However, because of the relative slow pavement of HDR displays, it is also desired to develop tone mapping algorithm for visualization of these HDR contents on LDR displays. Different from HDR image tone mapping algorithms, the temporal consistency or spatial-temporal consistency of these new HDR contents should be taken into consideration. Also, the speed of the tone mapping algorithm should be fast enough to support real-time video playback. How to develop tone mapping algorithms for visualization of these contents on conventional LDR displays

will be a challenging problem. Also, how to extend the scalable coding algorithms such as the one in Chapter 5 to provide backward-compatibility for these new HDR contents is of great interest.

Because of the insensitivity to lighting conditions, researchers are trying to apply HDR imaging techniques to challenging vision tasks under complex lighting such as feature points detection [130], facial expression recognition [131] and SLAM [132] et al. How to extend the conventional vision algorithms to work for HDR images/videos or how to tonemap HDR contents to provide illumination-invariant representation for vision algorithms will be another interesting research topic.

Appendix A

1D Solver

The one-dimensional optimization problem formulated for each row/colume in Eq. 3.29 is

$$\arg \min_{L_d^k} \sum_{x=0}^{n-1} \left[(L_d^k(x) - L_d^{k-1}(x))^2 + \alpha_{\lceil k/2 \rceil} \left(\frac{\frac{\partial L_d^k(x)}{\partial x} - v_1(x)}{\psi(v_1(x))} \right)^2 \right]. \quad (\text{A.1})$$

Here, n can be the width (if formulated for each row) or height (if formulated for each column) of the image.

Letting $\frac{\partial L_d^k(x)}{\partial x} = L_d^k(x+1) - L_d^k(x)$, it can be derived that the optimal solution is obtained by solving the following linear equation:

$$A * L_d^k = D, \quad (\text{A.2})$$

where the matrices A , L_d^k and D are given as

$$A = \begin{bmatrix} b_0 & c_0 & 0 & \cdots & 0 \\ a_0 & b_1 & c_1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a_{n-3} & b_{n-2} & c_{n-2} \\ 0 & \cdots & 0 & a_{n-2} & b_{n-1} \end{bmatrix}, \quad (\text{A.3})$$

$$L_d^k = [L_d^k(0) \ L_d^k(1) \ L_d^k(2) \ \cdots \ L_d^k(n-1)]^T, \quad (\text{A.4})$$

$$D = [d_0 \ d_1 \ d_2 \ \cdots \ d_{n-1}]^T, \quad (\text{A.5})$$

a_i , b_i , c_i and d_i are computed as

$$\begin{aligned} a_i &= c_i = -\frac{\alpha_{[k/2]}}{\psi^2(v_1(i))}; \ 0 \leq i \leq n-2, \\ b_0 &= 1 + \frac{\alpha_{[k/2]}}{\psi^2(v_1(0))}, \\ b_{n-1} &= 1 + \frac{\alpha_{[k/2]}}{\psi^2(v_1(n-2))}, \\ b_i &= 1 + \frac{\alpha_{[k/2]}}{\psi^2(v_1(i))} + \frac{\alpha_{[k/2]}}{\psi^2(v_1(i-1))}; \ 1 \leq i \leq n-2, \\ d_0 &= L_d^{k-1}(0) - \alpha_{[k/2]} \frac{v_1(0)}{\psi^2(v_1(0))}, \\ d_{n-1} &= L_d^{k-1}(n-1) + \alpha_{[k/2]} \frac{v_1(n-2)}{\psi^2(v_1(n-2))}, \\ d_i &= L_d^{k-1}(i) + \alpha_{[k/2]} \left(\frac{v_1(i-1)}{\psi^2(v_1(i-1))} - \frac{v_1(i)}{\psi^2(v_1(i))} \right); \ 1 \leq i \leq n-2. \end{aligned} \quad (\text{A.6})$$

The coefficient matrix A can be easily decomposed into two triangular matrices as

$$A = B * C \quad (\text{A.7})$$

where B and C are

$$B = \begin{bmatrix} \tilde{b}_0 & 0 & 0 & \cdots & 0 \\ a_0 & \tilde{b}_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a_{n-3} & \tilde{b}_{n-2} & 0 \\ 0 & \cdots & 0 & a_{n-2} & \tilde{b}_{n-1} \end{bmatrix}, \quad (\text{A.8})$$

$$C = \begin{bmatrix} 1 & \frac{c_0}{\tilde{b}_0} & 0 & \cdots & 0 \\ 0 & 1 & \frac{c_1}{\tilde{b}_1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \frac{c_{n-2}}{\tilde{b}_{n-2}} \\ 0 & \cdots & 0 & a_{n-2} & 1 \end{bmatrix}, \quad (\text{A.9})$$

where $\tilde{b}_i$ is

$$\tilde{b}_i = \begin{cases} b_0, & i = 0 \\ b_i - \frac{a_{i-1}c_{i-1}}{\tilde{b}_{i-1}}, & 1 \leq i \leq n-1. \end{cases} \quad (\text{A.10})$$

The optimal solution is then computed as

$$L_d^{k,*} = C^{-1} * B^{-1} * D. \quad (\text{A.11})$$

Appendix B

Subjective Evaluation Record

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	7	17	16	15	5
dis [13]	23	-	23	17	25	14
per [14]	13	7	-	7	6	11
rea [10]	14	13	23	-	16	8
pho [1]	15	5	24	14	-	2
ours	25	16	19	22	28	-

Table B.1: Preference matrix of subject 1.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	9	4	7	9	5
dis [13]	21	-	10	10	13	9
per [14]	26	20	-	17	23	18
rea [10]	23	20	13	-	18	18
pho [1]	21	17	7	12	-	14
ours	25	21	12	12	16	-

Table B.2: Preference matrix of subject 2.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	10	18	15	14	2
dis [13]	20	-	24	18	18	7
per [14]	12	6	-	4	10	7
rea [10]	15	12	26	-	13	10
pho [1]	16	12	20	17	-	3
ours	28	23	23	20	27	-

Table B.3: Preference matrix of subject 3.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	2	9	3	8	5
dis [13]	28	-	12	15	17	15
per [14]	21	18	-	14	21	6
rea [10]	27	15	16	-	24	7
pho [1]	22	13	9	6	-	5
ours	25	15	24	23	25	-

Table B.4: Preference matrix of subject 4.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	12	23	15	13	13
dis [13]	18	-	20	16	21	20
per [14]	7	10	-	6	9	6
rea [10]	15	14	24	-	15	12
pho [1]	17	9	21	15	-	6
ours	17	10	24	18	24	-

Table B.5: Preference matrix of subject 5.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	11	12	9	10	3
dis [13]	19	-	16	17	21	11
per [14]	18	14	-	11	19	13
rea [10]	21	13	19	-	19	16
pho [1]	20	9	11	11	-	7
ours	27	19	17	14	23	-

Table B.6: Preference matrix of subject 6.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	3	16	10	10	4
dis [13]	27	-	22	15	20	13
per [14]	14	8	-	9	6	9
rea [10]	20	15	21	-	20	8
pho [1]	20	10	24	10	-	1
ours	26	17	21	22	29	-

Table B.7: Preference matrix of subject 7.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	6	11	10	16	3
dis [13]	24	-	21	13	22	12
per [14]	19	9	-	6	11	10
rea [10]	20	17	24	-	21	14
pho [1]	14	8	19	9	-	2
ours	27	18	20	16	28	-

Table B.8: Preference matrix of subject 8.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	3	10	7	13	2
dis [13]	27	-	13	14	22	10
per [14]	20	17	-	17	23	13
rea [10]	23	16	13	-	22	13
pho [1]	17	8	7	8	-	2
ours	28	20	17	17	28	-

Table B.9: Preference matrix of subject 9.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	6	9	7	11	2
dis [13]	24	-	11	11	22	11
per [14]	21	19	-	14	22	8
rea [10]	23	19	16	-	24	13
pho [1]	19	8	8	6	-	2
ours	28	19	22	17	28	-

Table B.10: Preference matrix of subject 10.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	8	8	7	16	1
dis [13]	22	-	11	10	21	11
per [14]	22	19	-	13	23	12
rea [10]	23	20	17	-	22	9
pho [1]	14	9	7	8	-	4
ours	29	19	18	21	26	-

Table B.11: Preference matrix of subject 11.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	9	5	7	11	6
dis [13]	21	-	9	8	21	9
per [14]	25	21	-	13	22	17
rea [10]	23	22	17	-	21	8
pho [1]	19	9	8	9	-	4
ours	24	21	13	22	26	-

Table B.12: Preference matrix of subject 12.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	9	8	5	10	1
dis [13]	21	-	13	11	24	8
per [14]	22	17	-	13	19	12
rea [10]	25	19	17	-	25	8
pho [1]	20	6	11	5	-	1
ours	29	22	18	22	29	-

Table B.13: Preference matrix of subject 13.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	6	7	6	14	1
dis [13]	24	-	8	10	20	6
per [14]	23	22	-	18	25	14
rea [10]	24	20	12	-	25	12
pho [1]	16	10	5	5	-	4
ours	29	24	16	18	26	-

Table B.14: Preference matrix of subject 14.

	bil [12]	dis [13]	per [14]	rea [10]	pho [1]	ours
bil [12]	-	6	5	5	9	1
dis [13]	24	-	7	7	18	6
per [14]	25	23	-	15	24	17
rea [10]	25	23	15	-	22	12
pho [1]	21	12	6	8	-	2
ours	29	24	13	18	28	-

Table B.15: Preference matrix of subject 15.

Author's publications

Journal Papers

1. Zhengguo Li, Zhe Wei, Changyun Wen, and Jinghong Zheng, "Detail-Enhanced Multi-Scale Exposure Fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243-1252, Mar. 2017.
2. Zhe Wei, Changyun Wen and Zhengguo Li, "Local Inverse Tone Mapping for Scalable High Dynamic Range Image Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 2, pp. 550-555, Feb. 2018.
3. Fei Kou, Zhe Wei, Weihai Chen, Xingming Wu, Changyun Wen and Zhengguo Li, "Intelligent Detail Enhancement for Exposure Fusion," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 484-495, Feb. 2018.

Bibliography

- [1] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, “Photographic tone reproduction for digital images,” *ACM Trans. Graph.* **21**, 267–276 (2002).
- [2] D. A. Socolinsky and L. B. Wolff, “Multispectral image visualization through first-order fusion,” *IEEE Trans. Image Process.* **11**, 923–931 (2002).
- [3] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” in “*Comput. Graph. Forum*,” , vol. 28 (2009), vol. 28, pp. 161–171.
- [4] V. Vonikakis, O. Bouzos, and I. Andreadis, “Multi-exposure image fusion based on illumination estimation,” in “*Proc. IASTED SIPA*,” (2011), pp. 135–142.
- [5] K. Ma and Z. Wang, “Multi-exposure image fusion: A patch-wise approach,” in “*IEEE Int. Conf. on Image Process.*,” (IEEE, 2015), pp. 1717–1721.
- [6] R. Shen, I. Cheng, J. Shi, and A. Basu, “Generalized random walks for fusion of multi-exposure images,” *IEEE Trans. Image Process.* **20**, 3634–3646 (2011).
- [7] R. Shen, I. Cheng, and A. Basu, “Qoe-based multi-exposure fusion in hierarchical multivariate gaussian crf,” *IEEE Trans. Image Process.* **22**, 2469–2478 (2013).
- [8] S. Li, X. Kang, and J. Hu, “Image fusion with guided filtering,” *IEEE Trans. Image Process.* **22**, 2864–2875 (2013).

-
- [9] Z. G. Li, J. H. Zheng, and S. Rahardja, “Detail-enhanced exposure fusion,” *IEEE Trans. Image Process.* **21**, 4672–4676 (2012).
 - [10] G. Eilertsen, R. K. Mantiuk, and J. Unger, “Real-time noise-aware tone mapping,” *ACM Trans. Graph.* **34** (2015).
 - [11] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. Do, “Fast global image smoothing based on weighted least squares,” *IEEE Trans. Image Process.* **23**, 5638–5653 (2014).
 - [12] F. Durand and J. Dorsey, “Fast bilateral filtering for the display of high-dynamic-range images,” *ACM Trans. Graph.* **21**, 257–266 (2002).
 - [13] R. Mantiuk, S. Daly, and L. Kerofsky, “Display adaptive tone mapping,” *ACM Trans. Graph.* **27**, 68 (2008).
 - [14] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “A perceptual framework for contrast processing of high dynamic range images,” *ACM Trans. Appl. Perception* **3**, 286–308 (2006).
 - [15] S. Frintrop, T. Werner, and G. Martin Garcia, “Traditional saliency reloaded: A good old model in new shape,” in “IEEE Int. Conf. Comput. Vision and Pattern Recognition,” (2015), pp. 82–90.
 - [16] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, “Subjective quality assessment database of hdr images compressed with jpeg xt,” in “Quality of Multimedia Experience (QoMEX), 2015 Seventh Int. Workshop on,” (2015), pp. 1–6.
 - [17] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, “Edge-preserving decompositions for multi-scale tone and detail manipulation,” *ACM Trans. Graph.* **27**, 67 (2008).
 - [18] H. Yeganeh and Z. Wang, “Objective quality assessment of tone-mapped images,” *IEEE Trans. Image Process.* **22**, 657–667 (2013).

- [19] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam, and M. Cheriet, “Fsim: A feature similarity index for tone-mapped images,” *IEEE Signal Process. Lett.* **22**, 1026–1029 (2015).
- [20] “Phanton hd,” <http://www.phantomhighspeed.com>.
- [21] “Redcompany,” <http://www.red.com>.
- [22] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in “Proc. 24th Annu. Conf. Comput. Graph. and Interactive Techniques,” (ACM, 1998), p. 31.
- [23] T. Mitsunaga and S. K. Nayar, “Radiometric self calibration,” in “IEEE Conf. on Comput. Vision and Pattern Recognition,” , vol. 1 (IEEE, 1999), vol. 1, pp. 374–380.
- [24] P. Ledda, G. Ward, and A. Chalmers, “A wide field, high dynamic range, stereographic viewer,” in “Proc. of the 1st Int. Conf. on Comput. Graph. and Interactive techniques in Australasia and South East Asia,” (ACM, 2003), pp. 237–244.
- [25] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, “High dynamic range display systems,” *ACM Trans. on Graph.* **23**, 760–768 (2004).
- [26] G. Ward, “Real pixels,” *Graph. Gems II* pp. 80–83 (1991).
- [27] G. W. Larson, “Logluv encoding for full-gamut, high-dynamic range images,” *J. Graph. Tools* **3**, 15–31 (1998).
- [28] R. Bogart, F. Kainz, and D. Hess, “Openexr image file format,” *ACM SIGGRAPH 2003, Sketches & Applicat.* (2003).
- [29] “Jpeg xr standard,” <https://jpeg.org/jpegxr/>.
- [30] “Jpeg 2000 standard,” <https://jpeg.org/jpeg2000/>.

- [31] “Parameter values for the hdtv standards for production and international programme exchange,” https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.709-6-201506-I!!PDF-E.pdf.
- [32] “Parameter values for ultra-high definition television systems for production and international programme exchange,” https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2020-2-201510-I!!PDF-E.pdf.
- [33] “St 2084:2014 - smpte standard - high dynamic range electro-optical transfer function of mastering reference displays,” SMPTE ST 2084:2014 pp. 1–14 (2014).
- [34] “High dynamic range video,” https://en.wikipedia.org/wiki/High-dynamic-range_video#cite_note-SMPTE2094DynamicMetadataJanuary2017-27.
- [35] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, “Pixel-level image fusion: A survey of the state of the art,” *Inform. Fusion* **33**, 100–112 (2017).
- [36] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” *Inform. Fusion* **14**, 28–44 (2013).
- [37] S. Raman and S. Chaudhuri, “Bilateral filter based compositing for variable exposure photography,” in “Eurographics (Short Papers),” (2009), pp. 1–4.
- [38] A. A. Goshtasby, “Fusion of multi-exposure images,” *Image and Vision Computing* **23**, 611–618 (2005).
- [39] A. R. Várkonyi-Kóczy, A. Rovid, and T. Hashimoto, “Gradient-based synthesized multiple exposure time color hdr image,” *IEEE Trans. Instrum. Meas.* **57**, 1779–1785 (2008).
- [40] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, “A variational framework for retinex,” *Int. J. of Comput. vision* **52**, 7–23 (2003).

- [41] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Trans. Pattern Anal. Mach. Learn.* **35**, 1397–1409 (2013).
- [42] P. J. Burt and R. J. Kolczynski, “Enhanced image capture through fusion,” in “*IEEE Int. Conf. Comput. Vision*,” (IEEE, 1993), pp. 173–182.
- [43] P. Burt and E. Adelson, “The laplacian pyramid as a compact image code,” *IEEE Trans. Commun.* **31**, 532–540 (1983).
- [44] J. Shen, Y. Zhao, S. Yan, X. Li *et al.*, “Exposure fusion using boosting laplacian pyramid,” *IEEE Trans. Cybern.* **44**, 1579–1590 (2014).
- [45] S. Paris, S. W. Hasinoff, and J. Kautz, “Local laplacian filters: Edge-aware image processing with a laplacian pyramid,” *ACM Trans. Graph.* **30**, 68–1 (2011).
- [46] R. Fattal, M. Agrawala, and S. Rusinkiewicz, “Multiscale shape and detail enhancement from multi-light image collections,” *ACM Trans. Graph.* **26**, 51 (2007).
- [47] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in “*IEEE Int. Conf. Comput. Vision*,” (IEEE, 1998), pp. 839–846.
- [48] H. Li, B. Manjunath, and S. K. Mitra, “Multisensor image fusion using the wavelet transform,” *Graphical Models and Image Process.* **57**, 235–245 (1995).
- [49] J. Shen, Y. Zhao, and Y. He, “Detail-preserving exposure fusion using subband architecture,” *The Visual Comput.* **28**, 463–473 (2012).
- [50] P. Hill, M. E. Al-Mualla, and D. Bull, “Perceptual image fusion using wavelets,” *IEEE Trans. Image Process.* **26**, 1076–1088 (2017).
- [51] R. Shen, I. Cheng, and A. Basu, “Cross-scale coefficient selection for volumetric medical image fusion,” *IEEE Trans. Biomed. Eng.* **60**, 1069–1079 (2013).

- [52] C. Wang, Q. Yang, X. Tang, and Z. Ye, “Saliency preserving image fusion with dynamic range compression,” in “Int. Conf. Image Process.”, (IEEE, 2006), pp. 989–992.
- [53] B. Gu, W. Li, J. Wong, M. Zhu, and M. Wang, “Gradient field multi-exposure images fusion for high dynamic range image visualization,” *J. of Visual Communication and Image Representation* **23**, 604–610 (2012).
- [54] K. Hara, K. Inoue, and K. Urahama, “A differentiable approximation approach to contrast-aware image fusion,” *Signal Process. Lett.* **21**, 742–745 (2014).
- [55] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, “Probabilistic exposure fusion,” *IEEE Trans. Image Process.* **21**, 341–357 (2012).
- [56] G. Piella, “Image fusion for enhanced visualization: A variational approach,” *Int. J. of Comput. Vision* **83**, 1–11 (2009).
- [57] D. Hafner and J. Weickert, “Variational image fusion with optimal local contrast,” in “Comput. Graph. Forum,” , vol. 35 (2016), vol. 35, pp. 100–112.
- [58] H. Chen and P. K. Varshney, “A human perception inspired quality metric for image fusion based on regional information,” *Inf. Fusion* **8**, 193 – 207 (2007).
- [59] Y. Chen and R. S. Blum, “A new automated quality assessment algorithm for image fusion,” *Image and Vision Comput.* **27**, 1421 – 1432 (2009).
- [60] Y. Zheng, E. A. Essock, B. C. Hansen, and A. M. Haun, “A new metric based on extended spatial frequency and its application to dwt based fusion algorithms,” *Inf. Fusion* **8**, 177 – 192 (2007).
- [61] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Trans. Image Process.* **24**, 3345–3356 (2015).
- [62] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganier, and W. Wu, “Objective assessment of multiresolution image fusion algorithms for context enhancement

- in night vision: A comparative study,” IEEE Trans. Pattern Analysis and Mach. Intell. **34**, 94–109 (2012).
- [63] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process. **13**, 600–612 (2004).
- [64] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: Acquisition, Display and Image-Based Lighting* (San Mateo, CA, USA: Morgan Kaufmann, 2010).
- [65] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced high dynamic range imaging: theory and practice* (CRC press, 2011).
- [66] E. Reinhard and K. Devlin, “Dynamic range reduction inspired by photoreceptor physiology,” IEEE Trans. Visualization and Comput. Graph. **11**, 13–24 (2005).
- [67] G. W. Larson, H. Rushmeier, and C. Piatko, “A visibility matching tone reproduction operator for high dynamic range scenes,” IEEE Trans. Visualization and Comput. Graph. **3**, 291–306 (1997).
- [68] H. R. Wilson, “A transducer function for threshold and suprathreshold human vision,” Biological Cybern. **38**, 171–178 (1980).
- [69] S. Daly, “The visible differences predictor: an algorithm for the assessment of image fidelity,” Digital Images and Human Vision **4**, 124–125 (1993).
- [70] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, “Adaptive logarithmic mapping for displaying high contrast scenes,” in “Comput. Graph. Forum,” , vol. 22 (2003), vol. 22, pp. 419–426.
- [71] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, K. Zimmerman *et al.*, “Spatially nonuniform scaling functions for high contrast images,” in “Graph. Interface,” (CANADIAN INFORMATION PROCESSING SOCIETY, 1993), pp. 245–245.

-
- [72] Z. Li and J. Zheng, “Visual salience based tone mapping for high dynamic range images,” *IEEE Trans. Ind. Electron.* **61**, 7076–7082 (2014).
- [73] Z. Li, S. Rahardja, S. Yao, J. Zheng, and W. Yao, “High dynamic range compression by half quadratic regularization,” in “*IEEE Int. Conf. Image Process.*,” (2009), pp. 3169–3172.
- [74] R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” *ACM Trans. Graph.* **21**, 249–256 (2002).
- [75] J. Tumblin and H. Rushmeier, “Tone reproduction for realistic images,” *IEEE Comput. Graph. and Applicat.* **13**, 42–48 (1993).
- [76] L. Xu, C. Lu, Y. Xu, and J. Jia, “Image smoothing via l0 gradient minimization,” in “*ACM Trans. Graph.*,” , vol. 30 (2011), vol. 30, p. 174.
- [77] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, “Evaluation of tone mapping operators using a high dynamic range display,” *ACM Trans. Graph.* **24**, 640–648 (2005).
- [78] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, “Evaluation of hdr tone mapping methods using essential perceptual attributes,” *Comput. and Graph.* **32**, 330–349 (2008).
- [79] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “Dynamic range independent image quality assessment,” *ACM Trans. Graph.* **27**, 69:1–69:10 (2008).
- [80] R. Xu, S. N. Pattanaik, and C. E. Hughes, “High-dynamic-range still-image encoding in jpeg 2000,” *IEEE Comput. Graph. and Applicat.* **25**, 57–64 (2005).
- [81] S. Miller, M. Nezamabadi, and S. Daly, “Perceptual signal coding for more efficient usage of bit codes,” *SMPTE Motion Imaging J.* **122**, 52–59 (2013).

- [82] G. Ward and M. Simmons, “Subband encoding of high dynamic range imagery,” in “Proc. of the 1st Symp. on Appl. Perception in Graph. and Vis.”, (ACM, 2004), pp. 83–90.
- [83] G. Ward and M. Simmons, “Jpeg-hdr: A backwards-compatible, high dynamic range extension to jpeg,” in “ACM SIGGRAPH 2006 Courses,” (2006), p. 3.
- [84] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, “Backward compatible high dynamic range mpeg video compression,” in “ACM Trans. Graph.”, , vol. 25 (ACM, 2006), vol. 25, pp. 713–723.
- [85] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, “Optimizing a tone curve for backward-compatible high dynamic range image and video compression,” *IEEE Trans. Image Process.* **20**, 1558–1571 (2011).
- [86] Z. Mai, H. Mansour, P. Nasiopoulos, and R. K. Ward, “Visually favorable tone-mapping with high compression performance in bit-depth scalable video coding,” *IEEE Trans. Multimedia* **15**, 1503–1518 (2013).
- [87] M. Le Pendu, C. Guillemot, and D. Thoreau, “Local inverse tone curve learning for high dynamic range image scalable compression,” *IEEE Trans. Image Process.* (2015).
- [88] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, “Inverse tone mapping,” in “Proc. 4th Int. Conf. Comput. Graph. and Interactive Techniques in Australasia and Southeast Asia,” (ACM, 2006), pp. 349–356.
- [89] R. Mantiuk and H.-P. Seidel, “Modeling a generic tone-mapping operator,” in “Comput. Graph. Forum,” , vol. 27 (2008), vol. 27, pp. 699–708.
- [90] A. Artusi, R. K. Mantiuk, T. Richter, P. Hanhart, P. Korshunov, M. Agostinelli, A. Ten, and T. Ebrahimi, “Overview and evaluation of the jpeg xt hdr image compression standard,” *J. of Real-Time Image Process.* pp. 1–16 (2015).
- [91] R. K. Mantiuk, “Practicalities of predicting quality of high dynamic range images and video,” in “IEEE Int. Conf. Image Process.”, (2016), pp. 904–908.

-
- [92] T. O. Aydın, R. Mantiuk, and H.-P. Seidel, “Extending quality metrics to full dynamic range images,” in “Human Vision and Electronic Imaging XIII,” (San Jose, USA, 2008), Proceedings of SPIE, pp. 6806–10.
 - [93] 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,” in “ACM Trans. Graph.”, , vol. 30 (2011), vol. 30, p. 40.
 - [94] R. Mantiuk, K. Myszkowski, and H. P. Seidel, “Visible difference predictor for high dynamic range images,” in “IEEE Int. Conf. Syst., Man and Cybern.”, , vol. 3 (2004), vol. 3, pp. 2763–2769 vol.3.
 - [95] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, “Weighted guided image filtering,” IEEE Trans. Image Process. **24**, 120–129 (2015).
 - [96] F. C. Crow, “Summed-area tables for texture mapping,” ACM SIGGRAPH Comp. Graph. **18**, 207–212 (1984).
 - [97] S. Di Zenzo, “A note on the gradient of a multi-image,” Comp. Vision, Graph., and Image Proc. **33**, 116–125 (1986).
 - [98] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, NY, USA, 2004).
 - [99] Y. Saad, *Iterative Methods for Sparse Linear Systems* (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2003), 2nd ed.
 - [100] E. S. Gastal and M. M. Oliveira, “Domain transform for edge-aware image and video processing,” in “ACM Trans. Graph.”, , vol. 30 (ACM, 2011), vol. 30, p. 69.
 - [101] J. H. Z. Z. G. Li and S. Rahardja, “Detail-enhanced exposure fusion,” IEEE Trans. Image Process. **21**, 4672–4676 (2012).
 - [102] R. A. Bradley and M. E. Terry, “Rank analysis of incomplete block designs: I. the method of paired comparisons,” Biometrika **39**, 324–345 (1952).

- [103] W.-C. Lin and Z.-C. Yan, “Attention-based high dynamic range imaging,” *The Visual Comput.* **27**, 717 (2011).
- [104] Y. Mei, G. Qiu, and K.-M. Lam, “Saliency modulated high dynamic range image tone mapping,” in “Sixth Int. Conf. Image and Graph,” (IEEE, 2011), pp. 22–27.
- [105] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. and Mach. Intell.* **20**, 1254–1259 (1998).
- [106] L. Itti and C. Koch, “Computational modelling of visual attention,” *Nature Reviews Neuroscience* **2**, 194–203 (2001).
- [107] L. Marchesotti, C. Cifarelli, and G. Csurka, “A framework for visual saliency detection with applications to image thumbnailing,” in “IEEE Int. Conf. Comput. Vision,” (IEEE, 2009), pp. 2232–2239.
- [108] R. Achanta and S. Süsstrunk, “Saliency detection for content-aware image re-sizing,” in “IEEE Int. Conf. on Image Process.”, (IEEE, 2009), pp. 1005–1008.
- [109] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Trans. Pattern Analysis and Mach. Intell.* **34**, 1915–1926 (2012).
- [110] J. Wang, L. Quan, J. Sun, X. Tang, and H.-Y. Shum, “Picture collage,” in “IEEE Conf. on Comput. Vision and Pattern Recognition,” , vol. 1 (IEEE, 2006), vol. 1, pp. 347–354.
- [111] N. Jacobson and T. Q. Nguyen, “Video processing with scale-aware saliency: Application to frame rate up-conversion,” in “IEEE Int. Conf. on Acoust. Speech and Signal Process.”, (IEEE, 2011), pp. 1313–1316.
- [112] R. Margolin, L. Zelnik-Manor, and A. Tal, “Saliency for image manipulation,” *The Visual Comput.* **29**, 381–392 (2013).

- [113] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Trans. Image Process.* **13**, 1304–1318 (2004).
- [114] E. Reinhard, “Parameter estimation for photographic tone reproduction,” *Journal of Graph. Tools* **7**, 45–51 (2002).
- [115] R. Mantiuk, A. Tomaszewska, and W. Heidrich, “Color correction for tone mapping,” in “*Comput. Graph. Forum*,” , vol. 28 (2009), vol. 28, pp. 193–202.
- [116] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in “*IEEE Conf. on Comput. Vision and Pattern Recognition*,” (IEEE, 2007), pp. 1–8.
- [117] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, “Saliency detection via graph-based manifold ranking,” in “*IEEE Conf. on Comput. Vision and Pattern Recognition*,” (2013), pp. 3166–3173.
- [118] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical saliency detection,” in “*IEEE Conf. on Comput. Vision and Pattern Recognition*,” (2013), pp. 1155–1162.
- [119] <http://pfstools.sourceforge.net/>.
- [120] *Methodology for the subjective assessment of the quality of television pictures* (Rec. ITU-R BT.500-13, 2012).
- [121] P. Korshunov and T. Ebrahimi, “A jpeg backward-compatible hdr image compression,” in “*SPIE Optical Eng.+ Applicat.*,” (Int. Soc. for Optics and Photonics, 2012), pp. 84990J–84990J.
- [122] I. R. Khan, “Two layer scheme for encoding of high dynamic range images,” in “*Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*,” (2008), pp. 1169–1172.
- [123] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic, “Temporally coherent local tone mapping of hdr video,” *ACM Trans. Graph.* **33**, 196 (2014).

- [124] Z. Li and J. Zheng, “Visual-saliency-based tone mapping for high dynamic range images,” *IEEE Trans. Ind. Electron.* **61**, 7076–7082 (2014).
- [125] M. D. Grossberg and S. K. Nayar, “Determining the camera response from images: What is knowable?” *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1455–1467 (2003).
- [126] S. Mann and R. Mann, “Quantigraphic imaging: Estimating the camera response and exposures from differently exposed images,” in “*IEEE Conf. Comput. Vis. and Pattern Recognition*,” , vol. 1 (2001), vol. 1, pp. I–842.
- [127] M. L. Pendu, C. Guillemot, and D. Thoreau, “Inter-layer prediction of color in high dynamic range image scalable compression,” *IEEE Trans. Image Process.* **25**, 3585–3596 (2016).
- [128] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, “Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays.” in “*Digital Photography*,” (2014), p. 90230X.
- [129] A. Banitalebi-Dehkordi, “Introducing a public stereoscopic 3d high dynamic range (shdr) video database,” *3D Research* **8**, 3 (2017).
- [130] B. Přibyl, A. Chalmers, P. Zemčík, L. Hooberman, and M. Čadík, “Evaluation of feature point detection in high dynamic range imagery,” *J. of Visual Communication and Image Representation* **38**, 141–160 (2016).
- [131] E. O. Ige, K. Debattista, and A. Chalmers, “Towards hdr based facial expression recognition under complex lighting,” in “*Proc. of the 33rd Comput. Graph. Int.*,” (ACM, 2016), pp. 49–52.
- [132] S. Li, A. Handa, Y. Zhang, and A. Calway, “Hdrfusion: Hdr slam using a low-cost auto-exposure rgb-d sensor,” in “*Fourth Int. Conf. on 3D Vision*,” (IEEE, 2016), pp. 314–322.