

A Simplified HDR Image Processing Pipeline for Digital Photography

Jakkarin Singnoo

A thesis submitted for the Degree of
Doctor of Philosophy

University of East Anglia
School of Computing Sciences

October 2012

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior written consent.

Abstract

High Dynamic Range (HDR) imaging has revolutionized the digital imaging. It allows capture, storage, manipulation, and display of full dynamic range of the captured scene. As a result, it has spawned whole new possibilities for digital photography, from photo-realistic to hyper-real. With all these advantages, the technique is expected to replace the conventional 8-bit Low Dynamic Range (LDR) imaging in the future. However, HDR results in an even more complex imaging pipeline including new techniques for capturing, encoding, and displaying images. The goal of this thesis is to bridge the gap between conventional imaging pipeline to the HDR's in as simple a way as possible. We make three contributions. First we show that a simple extension of gamma encoding suffices as a representation to store HDR images. Second, gamma as a control for image contrast can be 'optimally' tuned on a per image basis. Lastly, we show a general tone curve, with detail preservation, suffices to tone map an image (there is only a limited need for the expensive spatially varying tone mappers). All three of our contributions are evaluated psychophysically. Together they support our general thesis that an HDR workflow, similar to that already used in photography, might be used. This said, we believe the adoption of HDR into photography is, perhaps, less difficult than it is sometimes posed to be.

Acknowledgements

First of all, I would like to thank my Supervisor Prof. Graham Finlayson who gave me the opportunity to be his student at the school of Computing Sciences at University of East Anglia (UEA), and for his valuable help with writing this thesis. I cannot put my gratitude into words, but merely thank you for always believing in me.

Thank to my colleagues and friends at the Colour Lab at UEA, alphabetically, Dave, Javier, Mary, Mike, Michal, Perla, Roberto, Stuart, and Toby for the numerous things they have done for me in all domains (especially Dave for correcting my written English at the beginning of my study).

I would like to thank all of my undergrad lecturers at the department of Imaging Science and Printing Technology (PHOTO) at Chulalongkorn University (CU), especially to Ajarn Erng who always supported and encouraged me throughout my study.

Thanks also to my PHOTO friends who were always available to hang out with me every time I went back home, alphabetically, Kai, Luknoo, Lukpla, Mei, Tik, Toon, and Tum.

And most of all, I would like to thank my Mom and Dad for understanding and tirelessly supporting me to pursue my dream, without them I would not have had the opportunity to do what I love. Last but not least, I would like to thank you to my brother, Diew.

Contents

Abstract	i
Acknowledgements	ii
List of figures	vi
List of tables	ix
Glossary	xi
1 Introduction	1
1.1 Summary of the Thesis	6
1.2 Publications	10
2 Background	12
2.1 HDR Imaging Pipeline	12
2.2 HDR Content Generation	14
2.2.1 Multiple Exposure Technique	14
2.2.2 Direct Capturing	16
2.2.3 Digital Image Synthesis	17
2.3 HDR Image Encoding	17
2.3.1 IEEE Float Format	18
2.3.2 Radiance RGBE Encoding	18
2.3.3 OpenEXR Encoding	20
2.3.4 scRGB Encoding	21
2.3.5 HDR Encoding Comparison	22
2.4 HDR Display	23
2.5 Tone Mapping Operators	25
2.5.1 Global Operators	29
2.5.2 Spatially-Varying Operators	31
2.6 Post-Processing	37

2.6.1	Black and White Point Correction	37
2.6.2	Gamma Adjustment	38
2.6.3	Color Reconstruction	39
2.7	Image Statistics	40
2.7.1	Dynamic Range	41
2.7.2	Image Entropy	41
2.8	Image Quality Measurement	44
2.8.1	Objective Image Quality Measurement	45
2.8.2	Subjective Image Quality Measurement	47
2.9	Staircase Psychophysical Experiment	53
2.10	Summary	56
3	Modified TIFF for Encoding HDR Photographs	58
3.1	Background	59
3.2	The Modified Framework: Using Integer TIFF to Encode HDR	62
3.3	Experiment	64
3.3.1	Objective Image Quality Measurement	64
3.3.2	Subjective Image Quality Measurement	74
3.3.3	Storage Size	76
3.4	Conclusions	77
4	Automatic Gamma Adjustment	81
4.1	Background	81
4.2	Experiments	83
4.3	Results	86
4.4	Discussion	88
4.4.1	Moroney's Gamma Adjustment	88
4.4.2	CIECAM Gamma Adjustment	89
4.4.3	Results	91
4.5	From the Experiment to an Automatic Gamma Adjustment Operator	93
4.6	Conclusion	93
5	Optimal Global Approximation to SV-TMOs	96
5.1	Background	97
5.2	The Optimal Approximation Operator	101
5.2.1	Tone-Curve Optimization	102
5.2.2	Detail Recovery	109
5.3	Preference Experiment	111
5.3.1	Algorithm Parameters	112
5.3.2	Experimental Results	113
5.4	Conclusions	115

6	Gamma Adjustment for HDR Post Processing	116
6.1	Revisiting the Gamma Adjustment Equation	117
6.2	Preference Experiment	118
6.3	Results and Discussions	119
6.4	Conclusion	119
7	Conclusion and Future works	123
7.1	Conclusion	123
7.2	Future Works	126
	References	128

List of Figures

1.1	Examples of scenes that contain a wider range of luminance values than conventional imaging devices can record or display.	2
1.2	Example of scene that is impossible to be reproduced by conventional photography.	3
1.3	Photographs in which dynamic range of the scene is fully captured. Noted that there are no under- or over-exposed areas.	3
1.4	Examples of artistic photographs resulted from dynamic range clipping. The under-exposed foregrounds is created for artistic purpose.	4
1.5	HDR visualization problem	7
2.1	Schematic diagram of HDR Imaging Pipeline.	13
2.2	HDR content generation. By fusing a sequence of bracketed exposures using a camera response function, an HDR image can be generated.	16
2.3	Schematic diagrams of HDR display technologies.	25
2.4	Variation in dynamic range of LDR and HDR imaging systems.	26
2.5	HDR imaging techniques applied to the scene shown in Figure 1.2	27
2.6	Reproduction images should be visually similar to the observed scene.	28
2.7	An example of the Histogram adjustment by Larson et al.	30
2.8	A comparison between a global tone-mapped image and a local tone-mapped image.	32
2.9	A diagram illustrates the halo artifact.	33
2.10	Halo artifact introduced by spatially-varying TMOs.	34
2.11	A small filter size leads to halo artifacts, whereas a large filter size alleviate these artifacts (they are much reduced) but the detail in the output can be more muted (the power of local operation is reduced).	35
2.12	An example of histogram-based black and white point correction.	38
2.13	The effect of the saturation factor on the tone-mapped image.	40
2.14	Histogram equalization.	43
2.15	Relationship between the scale difference D and the probability P of preferring the one stimulus above the other stimulus.	52
2.16	An example of single staircase experiment.	54

2.17	An example of double staircase experiment.	56
3.1	The image dataset.	66
3.2	Effect of bit-depth on the appearance of the tone-mapped images.	68
3.3	Effect of gamma on the appearance of the tone-mapped images.	69
3.4	Average percentage of pixels at a particular ΔE S-CIELAB of different encodings for the entire test images.	70
3.5	Effect of bit-depth and γ on the iCAM06 tone-mapped image.	71
3.6	Percentage of pixels at a particular ΔE S-CIELAB of different encodings of the scene shown in Figure 3.5.	72
3.7	The perceived brightness of the quantization steps of different gamma (γ of 1, 2.2, and 3.0).	73
3.8	Overall preference scores for encodings over 10 images.	77
3.9	Preference scores for 10 test images by image.	78
3.10	Preference scores for 10 test images by encoding.	79
3.11	The average file size (size efficiency) for each test encodings. Error bars show extrema in the test.	79
4.1	Some images in the dataset.	85
4.2	Average preferred gamma from 12 observers plotted against the theoretical optimal gamma ($-1/\text{mean}(\log_2(x))$) for 42 testing images.	87
4.3	The approximation of A/A_W by the square root of the ratio of input intensities Y/Y_W	91
4.4	Different gamma predictors plotted against the average chosen gamma from the experiment, the corresponding lines to each dataset are the linear fitting to each of the predictor.	92
4.5	Examples of our automatic gamma adjustment. Left column shows original images. Right column shows result images.	94
5.1	PAVA approximation and its reference spatially-varying TMO.	99
5.2	Image processing of the optimal approximation operator.	102
5.3	Tone-curves resulted from PAVA and its robust smoothed version (PAVA _s). 106	
5.4	Compare PAVA output (left) and PAVA _s output (right) (we see a smoother tone rendering in the PAVA _s than the original PAVA).	107
5.5	An example of a poor fit of our tone-curve approximation.	108
5.6	Details recovered from the cross-bilateral of the proposed operator.	109
5.7	A comparison chart of HDR pixel encodings.	110
5.8	5 well-known images used in the experiment.	112
5.9	Overall preference scores.	114
6.1	Sigmoidal regression versus linear regression of the experimental results conducted in Chapter 4.	118

6.2	(left column) images were obtained with the optimal global approximation TMO. (right column) images were obtained with the combination process proposed here (generated by applying an automatic gamma adjustment to the right column images).	120
6.3	Overall preference scores over five test scenes.	121
6.4	Preference scores for each scene.	122

List of Tables

2.1	A summary of HDR capturing devices (Banterle et al., 2011).	16
2.2	A comparison chart of HDR pixel encodings.	22
2.3	Properties of HDR display by type (Banterle et al., 2011).	24
3.1	The summary of percent of pixels above ΔE S-CIELAB of 2 and 5 for each encoding.	69
4.1	Gamma predicting equations along with the correlation coefficient (R^2) of the four models.	93
5.1	Tone mapping operators used in the experiment together with their labels and categories used in the paper.	113
5.2	The number of times each TMO is preferred over the others.	115

Glossary

2AFC	two-alternative forced choice
BF	bilateral filter
CDF	cumulative distribution function
CG	computer graphics
CRF	camera response function
CRT	cathode ray tube
DCI	digital cinema initiatives standard
DR	detail recovery
G-TMO	global TMO
GND	graduated neutral density
HDR	high dynamic range
HE	histogram equalization
HVS	human visual system
JND	just noticeable difference
JPEG	(joint photographic experts group) image file format
LED	light-emitting diode
LCD	liquid crystal display
LDR	low dynamic range
MSE	mean square error
PAVA	pool-adjacent-violators-algorithm
PAVA _s	smooth-PAVA
QP	quadratic programming
RGB	red, green, blue
RGBE	HDR radiance image encoding (.hdr)
RMSE	root mean square error
SSR	single-scale retinex
SV-TMO	spatially-varying TMO
TC	tone-curve
TIFF	tagged image file format
TMO	tone mapping operator
VDP	visual difference predictor

Chapter 1

Introduction

Typical imaging represents color with 8-bit per pixel for each of the red, green, and blue color components. With 24-bit per pixel, more than 1.6 million different colors can be reproduced (often referred as “millions of colors” in many software applications). Although this seems to be enormous, it should be noted that there are only 256 different shades per color component. Having only 256 different shades per color component means conventional images cover only about two orders of magnitude (to calculate this see Section 2.7.1) which is not enough to represent many real scenes. Indeed, the Dynamic range (ratio of the brightest to darkest pixel values) of real scenes can span 4, 5 or more orders of magnitude. Examples of HDR - high dynamic range scene are shown in Figure 1.1.

A scene that contains both indoor and outdoor lit objects is likely to have large range of intensities. As a result, using a conventional 8 bit encoding, either the outdoor background is overexposed or the indoor foreground is underexposed depending on the exposure value setting used in the capturing process. In fact, some scenes might be impossible to capture and reproduce using a single exposure (i.e. using one image). Figure 1.2 further illustrates the problem.

The inability to store the full dynamic range of the captured scene, in ‘one shot’,

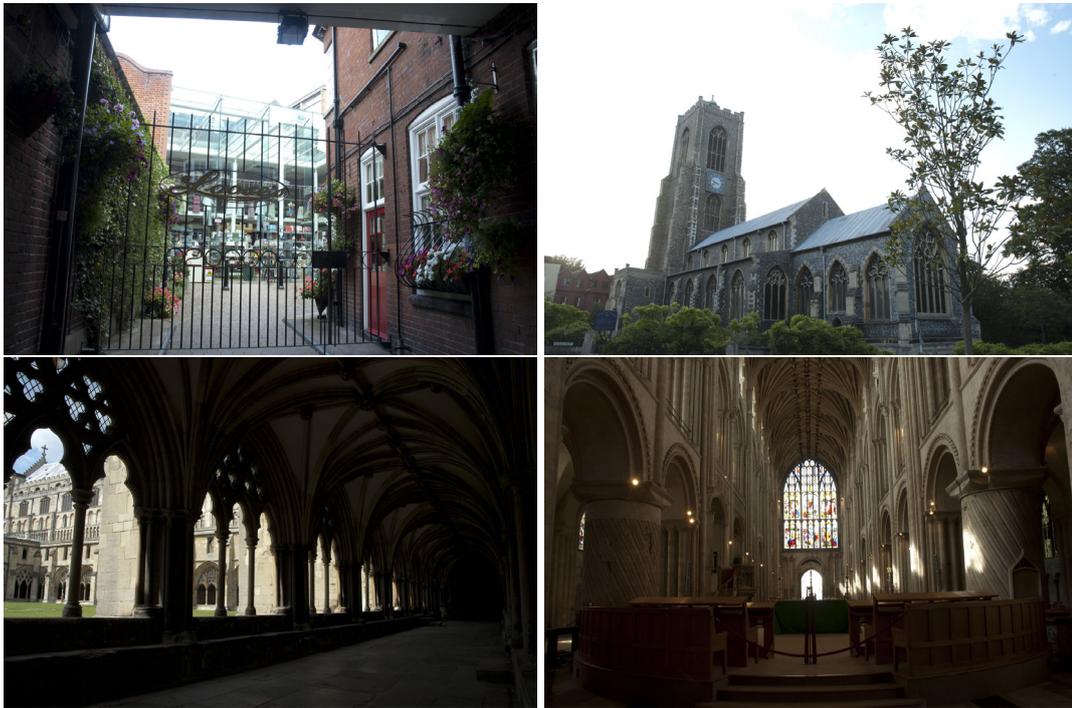


Figure 1.1: Examples of scenes that contain a wider range of luminance values than conventional imaging devices can record or display.

means that there is no way to capture the full details of these scenes using conventional image capturing devices. As a result, if a photographer wants to fully represent the real-scene, he/she is forced to capture the scenes with the dynamic range that the hardware is capable of. Such scenes often refer as a “good-light” scene and may be achieved by keeping the light source behind the camera. Examples of such photographs are shown in Figure 1.3. We also encounter “good-light” scenes when we ask a photographer to take family pictures. Often they take great care to control the lighting.

In practice, professional photographers sometimes intend to discard some of the dynamic range of the captured scene for artistic purpose. Figure 1.4 illustrates such photographs. Although the limitation on affordable dynamic range of capturing devices sometimes creates a desired atmosphere for a number of applications, however photographers have to be very careful with the exposure setting in order to make sure that

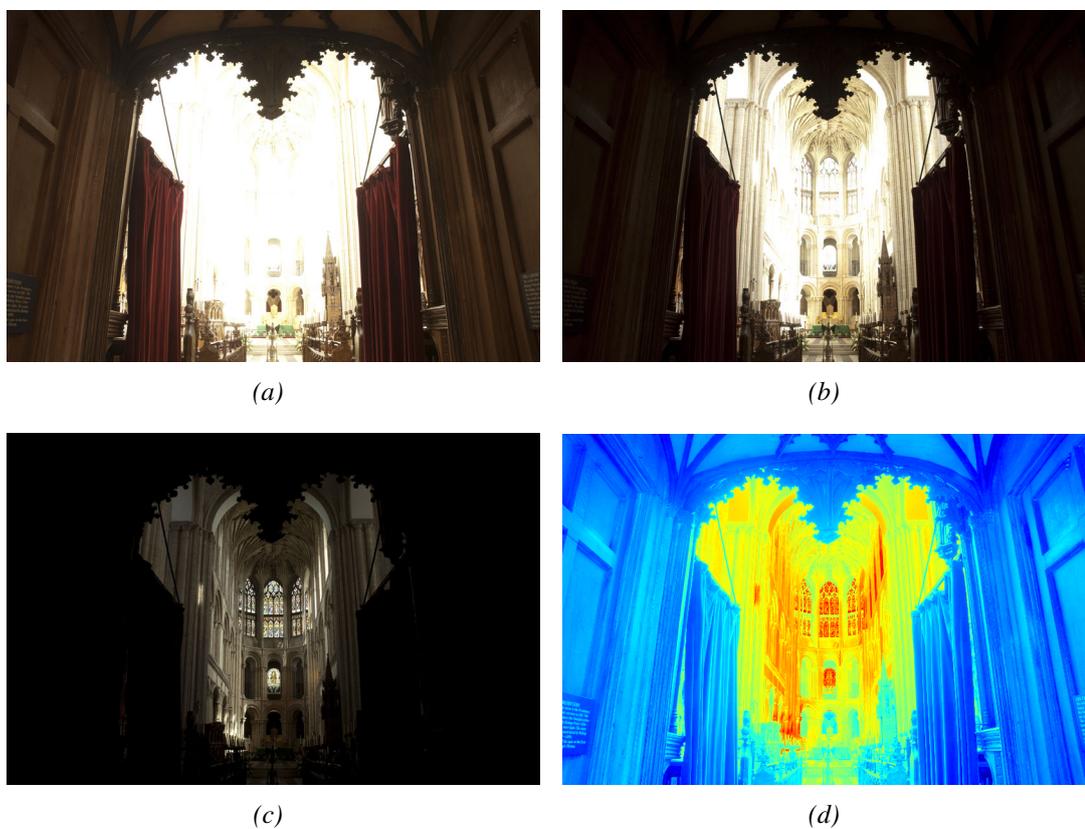


Figure 1.2: Example of scene that is impossible to be reproduced by conventional photography. (a) an over-exposed image (contains shadow details). (b) a normal-exposed image. (details in both shadow and highlight are lost) (c) an under-exposed image (contains highlight details). (d) the corresponding HDR (shown in false color).



Figure 1.3: Photographs in which dynamic range of the scene is fully captured. Noted that their are no under- or over-exposed areas.

whether the shadow or the highlight part of the scene are properly exposed. Thus, it is almost always more desirable to capture the full dynamic range of the scene; later, if desired, the image can then be processed to imitate this effect.

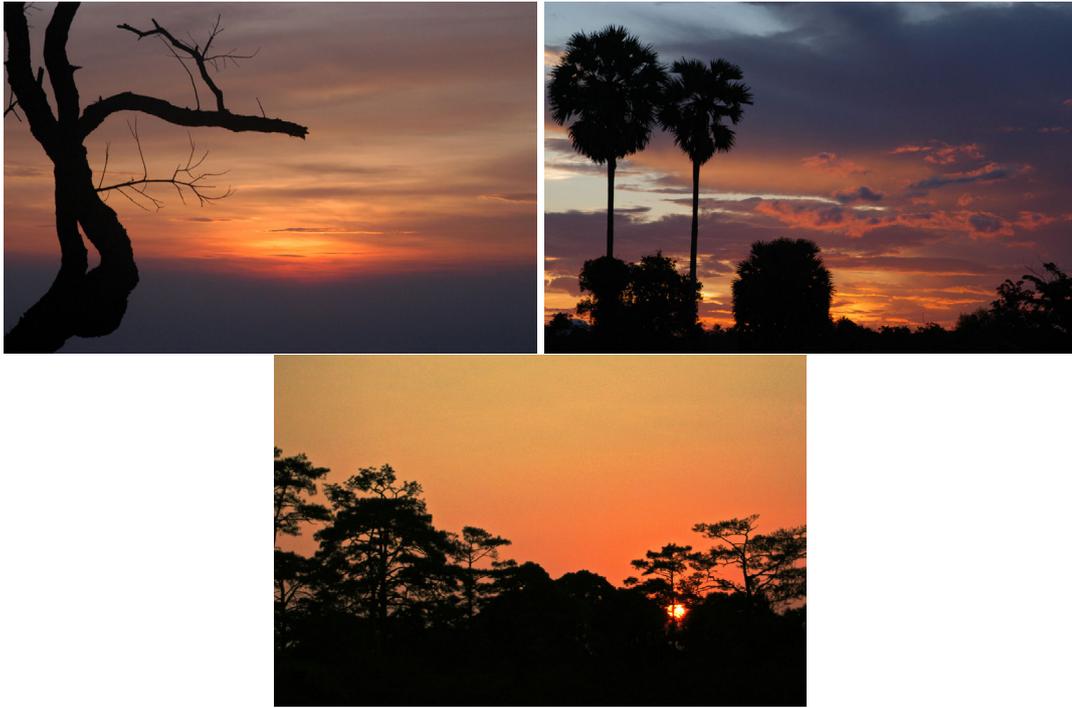


Figure 1.4: Examples of artistic photographs resulted from dynamic range clipping. The under-exposed foregrounds is created for artistic purpose.

Given single ‘one shot’ cameras, it is useful to consider the range of options conventionally available to capture a desired dynamic range. One classical way to handle this problem is to reduce the dynamic range of the scene to fit the affordable dynamic range of the camera. There are many photographic techniques to achieve this. For example, one could apply the Graduated Neutral Density (GND) filter, to restrict the amount of light across an image in a smooth geometric pattern, to a specific part of the scene. However, GND filter works for scenes with simple lighting geometries, such as the linear blend from dark to light encountered commonly in landscape photography which tend to have a bright sky as a background and a much darker terrain as a foreground.

Additionally and problematically, some GND filters can add a very slight color cast to the image. Another frequently used technique is to add artificial light to foreground objects in the scene using flash. However, the use of flash can also have a negative effect on the lighting characteristics of the scene, since, foreground objects could be disproportionately brightened. In addition, the flash may introduce unpleasant artifacts such as specularities, and harsh shadows which are not exist in the real scene. As a consequent, these photographs often fail to represent the true ambient illumination of the captured scenes.

A modern way to solve the problem of dynamic range - and the primary focus of this thesis - is to fully capture the dynamic range of the scene by using a set of techniques called High Dynamic Range (HDR) imaging. HDR enables us to capture, store, transmit, and utilize real-world lighting of the scene beyond that which conventional imaging devices offer (which is also exceed the conventional image encoding that rely on these devices). The highlight and shadow detail of a scene can be fully recorded either by direct capture using special HDR capabled capturing devices (Panoscan, 2012; Spheron, 2012; Weiss-AG, 2012) or by capturing multiple exposures of the same scene then merging them to reconstruct the image that contains full dynamic range of the captured scene (Debevec and Malik, 1997; Mann and Picard, 1994) (see Figure 1.2d).

Capturing HDR is only one part of the problem that needs to be solved. While we might be able to capture all the detail, conventional displays do not output the range of brightnesses to reproduce the captured scene. For example, most Cathode-Ray-Tube (CRT) displays output about only two orders of magnitudes due to the fact that phosphors cannot be excited beyond a certain level. Although, Liquid Crystal Displays (LCD) have larger dynamic range, they produce an order of magnitude more output brightness range, it still too little compared with the scene dynamic range. Although, display devices that can natively visualize HDR contents have been available for some times (Seetzen et al., 2004), such displays are likely limited to only researchers and

impractical to most users since they are extremely expensive. The displays are also extraordinarily expensive in terms of power usage. Thus, in this thesis we limit our discussion of HDR visualization only on the technique used for conventional displaying devices. This means that only a fraction of the captured dynamic range of the HDR can be displayed.

Figure 1.5 illustrates the HDR visualization problem. Figure 1.5a shows that the LDR image has a very limited dynamic range (both highlight and shadow are clipped). In contrast, the HDR representation shown in Figure 1.5b contains all pixel information, however only a small fraction is visible. Neither of these two reproductions is desirable.

In order to properly display HDR images, an additional process that performs *dynamic range compression* is needed. This process is known as tone mapping. Tone mapping is carried out by a Tone-Mapping Operator (TMO) (see Figure 1.5c). The purpose of a TMO is to reduce the dynamic range of the HDR image to fit the displayable dynamic range afforded by a given displaying device in the way that preserves or enhances the appearance of the compressed image to the original.

1.1 Summary of the Thesis

While it is straightforward to articulate what we would like to do: make HDR work! There is, in fact, a very large literature that attempts to do this. The problem is not easy and, hitherto, adopting HDR in a photographic work flow is complex (not for the faint hearted). Here we wish to address this complexity. We ask ‘is this complexity necessary?’ The goal of this thesis is thus to simplify the HDR imaging pipeline in the context of digital photography. To achieve this goal three aspects of the pipeline have been investigated.

First, the HDR encoding was investigated. Our aim is to bridge the the gap between the conventional and HDR imaging workflows. In HDR encoding, RGBE radiance



Figure 1.5: HDR visualization problem. (a) Single LDR image of a clearly HDR scene. (b) HDR image directly displayed. (c) Tone-mapped image showing a superior visual information compared to the previous two images (this image is tone-mapped using the proposed algorithm in Chapter 5). It should be noted that these images are printed on paper so that the range of intensities is actually no larger than a conventional image. Since typically it is not possible to directly print images with a much higher dynamic range than one that is afforded by the print medium. This indicates that even in the absence of an HDR capable displaying device, images are still benefited from HDR technique.

format (Larson and Shakespeare, 1998) is one of the most commonly used. RGBE encodes image brightnesses in scientific notation (mantissa plus exponent). Cleverly, a single exponent (coded as an 8-bit number) is used along with three 8-bit mantissas for R, G and B. The RGBE encoding suffices to accurately represent all (even extreme) HDR images, however, conventional image handling applications do not always support the RGBE encoding and those that do take a considerable amount of time to encode or decode compared with conventional image formats.

The expense of using RGBE is directly related to the efficient way (scientific notation) that images brightnesses are encoded. However, in the context of an actual image encoding/decoding where the image can have more than 10 Mega pixels, converting scientific notation to a linear brightness signal is very expensive compared with, for example, inverting the gamma in an 8-bit coding such as sRGB since inverting the gamma typically can be done using an 1-D lookup table. So, in Chapter 3, we consider the possibility of using conventional 12 and 16-bit/component integer TIFF gamma-encoded image encoding for storing these dynamic ranges. We will show that the proposed encoding competes with the state-of-the-art RGBE in both objective and subjective ways. Further, with an additional compression step, our method is also efficient in terms of bits/pixel.

In the second part of the the thesis, we rigorously investigate the simplest and commonly used TMO: applying a gamma to an image. Although, gamma adjustment requires only one gamma parameter, traditionally this parameter is manually specified by user. Recently, Finlayson and Xu (2012) proposed a technique to optimally derive the gamma in the information theoretic sense (from the entropy of image). Over all choices of gamma there is a single closed form solution that delivers an image that has maximum ‘differential’ entropy. The premise of that work, though it was not tested, was that more information would correlate with image preference. In Chapter 4, a psychophysics experiment called “Double staircase” method has been conducted on a set of luminance images. which allows a user to choose the gamma they find brings out most detail and is preferred. Given real gamma adjustments we can relate the theoretical optimal gamma to the adjustments actually made by observers. Broadly, we find that optimum gamma does predict psychophysical data well. Significantly, the prediction is better than other (modified) formula reported in the literature (Moroney’s non linear masking (Moroney, 2000) and CIECAMs (Alessi et al., 1998; Moroney et al., 2002))

Applying a tone mapping operator probably is one of the most significant stages of

the HDR imaging pipeline. In recent years the problem of tone mapping has attracted much attention, many TMOs have been proposed e.g. (Ashikhmin, 2002; Chiu et al., 1993; Durand and Dorsey, 2002; Fairchild et al., 2004; Jobson et al., 1996; Larson et al., 1997; Meylan and Susstrunk, 2006; Pattanaik et al., 1998; Rahman et al., 1996, 1997b; Reinhard et al., 2002; Tumblin et al., 1999; Tumblin and Rushmeier, 1993). In conventional photography a so-called ‘global’ operator is used. Input brightnesses are mapped to output counterparts using a single transfer curve (the same input brightness is always mapped to the same output). Unfortunately, in mapping HDR images a global curve approach often produce outputs which are lackluster (dull and flat) and lack detail. As a consequence the majority of modern approaches propose that ‘local’ tone mapping should be used (the same brightness maps to different outputs in different parts of the image). The results are, in some sense, dramatically improved. HDR images can be mapped for display where all the detail is preserved and where the detail is made tremendously exaggerated. The cost of using spatially varying TMOs is twofold. First the algorithms are often very costly: it can take minutes to process an image. Second, these algorithms tend to add artifacts in the reproduction e.g. ‘halos’ at high contrast edges. In this work, we propose a global tone-mapping operator that optimally, in a sum of least-squares sense, approximates spatially-varying tone-mapping operators. Specifically, in Chapter 5 we proposed a hybrid processing that has advantages of both global and spatially-varying TMOs. Our idea is, effectively, we apply a good spatially varying tone mapper to a small thumbnail image. Then, in an optimal least-squares sense, approximate this algorithm using a global transfer curve. This curve is applied to the full resolution input. We deal with detail loss using a commonly used detail recovery step reported in the literature. Remarkably, this workflow though global can produce excellent reproductions (at least as good as spatially varying counterparts).

We find the proposed global tone-map is based on a modification of the simple but elegant constrained optimization technique called Pool-Adjacent-Violators-Algorithm

(PAVA). Together with a detail recovery step that brings back any lost details that are missing by PAVA, in an edge sensitive manner using bilateral filter.

Typically, the final step of the HDR visualization is to apply a set of post-processing process. Several post-processing have been proposed to improve the visual appearance of the tone-mapped images. In Chapter 6, we demonstrate the use of the gamma predictor proposed in Chapter 4 in the context of HDR visualization (TMO follows by post-processing process). Specifically, we combined the optimal global approximation TMO proposed in Chapter 5 and the gamma predictor proposed in Chapter 4 and validated this by conducting a paired comparison experiment. Not only do we find that our PAVA TMO approximates spatially varying tone mapping that by tweaking the gamma (using our developed equation) results in a reproduction that is further preferred.

Chapter 7 concludes the whole thesis together with a list of future directions of the works proposed in this thesis.

1.2 Publications

In this thesis, the following publications have been made:

- J. Singnoo and G. D. Finlayson. RGBE vs Modified TIFF for Encoding High Dynamic Range. In *Proceedings of IS&T's Fifth European Conference on Color in Graphics, Imaging and Vision*, pages 431–436, Joensuu, Finland, June 2010.
- J. Singnoo and G. D. Finlayson. Understanding the Gamma Adjustment of Images. In *Proceedings of IS&T and Eighteenth Color Imaging Conference*, pages 134–139, San Antonio, Texas (USA), November 2010.
- J. Singnoo and G. D. Finlayson. Optimal global approximation to spatially varying tone mapping operators. In *Proceedings of IS&T's Sixth European Conference on Color in Graphics, Imaging and Vision*, pages 182–188, Amsterdam,

Netherlands, May 2012.

Chapter 2

Background

This chapter provides general information concerning HDR imaging pipeline, other related background, and notations used in the rest of the thesis. A brief overview of HDR imaging pipeline is described in Section 2.1. Section 2.2 gives details on the HDR content generation. Section 2.3 focuses on the HDR image encoding. Section 2.4 gives details on the HDR display. In Section 2.5, tone mapping operators are discussed. The post processing stage of the HDR imaging pipeline is introduced in Section 2.6. Section 2.7 describes related image statistics of the thesis. The image quality measurement used with respect to HDR is described in Section 2.8. Finally, section 2.9 describes the method of staircase in psychophysics.

2.1 HDR Imaging Pipeline

An imaging pipeline is a term used to describe the processes that are typically used between an image acquisition (image capturing) to image reproduction. In the broadest sense, HDR imaging pipeline may be divided into four main stages: capturing, storing, processing, and visualising. Figure 2.1 illustrates the pipeline.

The first stage of the pipeline is the generation of the HDR image. HDR photographs

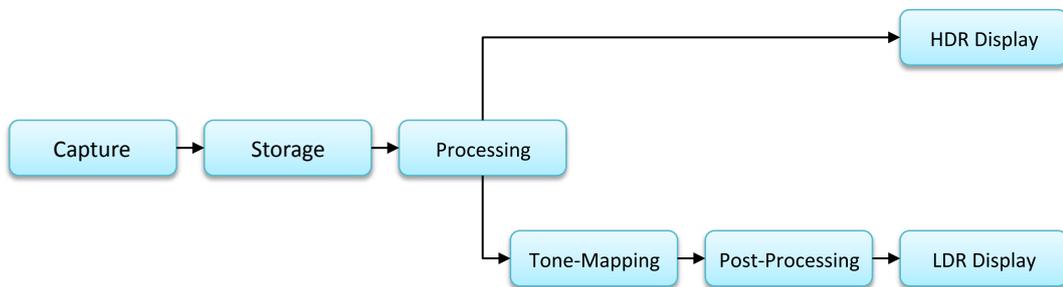


Figure 2.1: Schematic diagram of HDR Imaging Pipeline.

can be generated in a number of ways. These include ones that generated using a multiple-exposure technique (Debevec and Malik, 1997; Mann and Picard, 1994), direct capture (Panoscan, 2012; Spheron, 2012; Weiss-AG, 2012), or even inferring HDR from a single LDR image (Banterle et al., 2006). In the multiple exposure technique - the most common method - three or more different exposures can be taken (e.g. normal, short, and long exposures). These exposures are registered and blended. This results in an image with the dynamic range excess of 16 bits/component compared with the 12 bits captured from typical imaging devices.

Once HDR data has been generated, it is time to store the content. Unlike LDR, HDR needs a more efficient encoding technique to store its contents. An ideal HDR representation takes single precision floating point values to encode one single color component (32-bit/component) (Hough, 1981). This means that 96-bit of memory is required to encode just one single pixel (assuming there are three RGB color components). This means that an HDR image can be four times larger than its LDR (8-bit/component, 24-bit in total). To address the high memory storage problem, many encodings either store floating point or integer data in a special arrangement that allows recovery of the original data have been developed that compete in terms of file size, encoding and decoding times, precision, etc.

The stored HDR images can be processed in different ways, much like conventional

LDR images. However, due to the fact that HDR contains much larger data, they can additionally be used for several other applications such as special effects, or image based lighting (Debevec, 2005), for which LDR images are not suitable.

In the following step, the HDR image is visualized. As discussed earlier, there are two methods to visualize HDR content. One is to develop special display hardware that can accommodate high dynamic range. The other is to compress the dynamic range of the HDR image to fit the dynamic range afforded by typical display devices through the software process called tone mapping.

In addition, it is also typical to apply a number of post-processing processes as a final step of the pipeline. Examples of such processes are: black and white point correction, gamma adjustment, and color reconstruction (Schlick, 1994).

The following sections give an overview of all of these steps of the HDR imaging pipeline. The intention is to provide the necessary background of the whole thesis.

2.2 HDR Content Generation

Among several techniques to generate an HDR image, three of them are discussed here. The first technique is probably the most commonly used is the technique of combining a sequence of different exposures. The second technique is the direct capture using a specialized hardware. The third technique is the technique of digital image synthesis which frequently can be augmented by photographic methods.

2.2.1 Multiple Exposure Technique

As mentioned earlier, photographers are forced to set the exposure setting (exposure time and aperture number) to a specific value in order to capture the desired contents of a scene. There are many situations in which the dynamic range of the scene is extreme that is impossible to capture the desired content without either overexposing or underexposing

the rest of the image (see Figure 1.2). To fully capture the dynamic range of such a scene, a sequence of bracketed exposures can be taken. Each exposure lets in more or less light (different EV values). However, these images cannot be directly merged to construct a HDR image. This is because cameras (like film) have a non-linear response that determines how radiance of the captured scene is mapped to pixel intensities. This non-linear response is often known as the Camera Response Function (CRF) f . The CRF is often hard to know beforehand because it is the formation of several non-linear functions that take place in the photographic process.

Mann and Picard (1994) was the first to propose a method to calculate f . By fitting the values of corresponding pixels at different exposures to find a fixed $f(x) = ax^\gamma + b$. Note that, to solve for the camera response function Mann and Picard places a strong constraint on the form of the function. However, practical CRFs found in typical image capturing devices do not fit with this formula. Later, Debevec and Malik (1997) proposed an alternative least-squares method to recover the CRF. Once the CRF function is derived, it can be used to convert pixel intensities to radiance values (assuming the exposure is known). Since f is monotonically increasing, this means that its inverse function can be derived. The exposure X at each pixel is calculated as:

$$X = f^{-1}(I(x)) \quad (2.1)$$

The irradiance E can be then recovered as

$$E = \frac{X}{\Delta t} \quad (2.2)$$

where Δt is the exposure time.

Once the CRF has been generated, it may be stored and later used for subsequent scenes. The process of HDR content generation is illustrated in Figure 2.2.

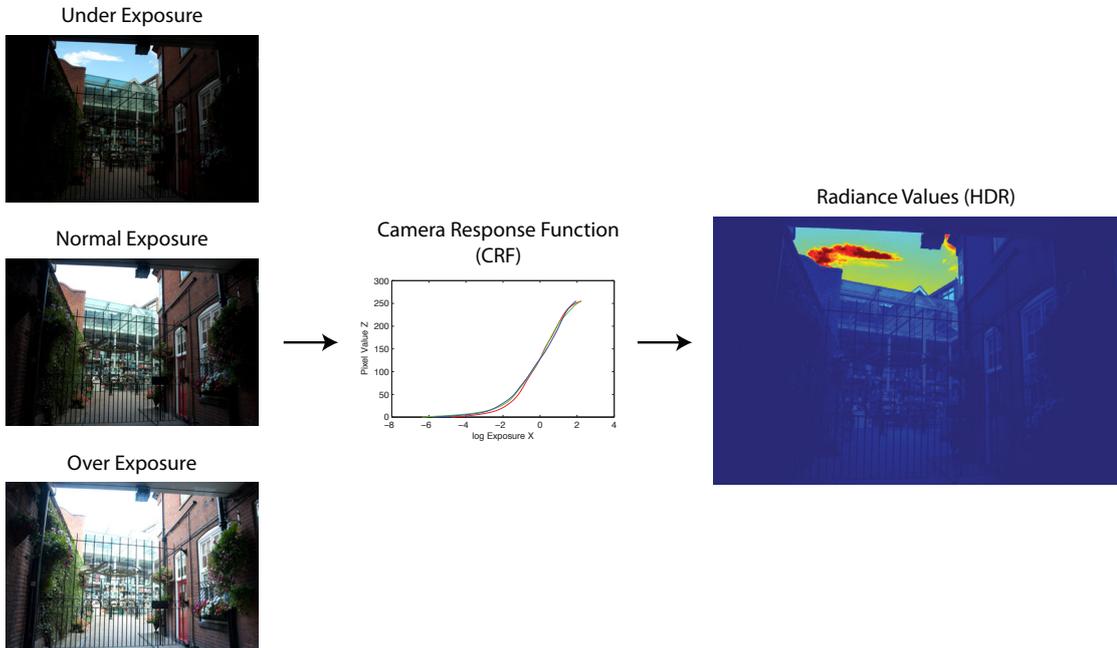


Figure 2.2: HDR content generation. By fusing a sequence of bracketed exposures using a camera response function, an HDR image can be generated.

2.2.2 Direct Capturing

Recently, direct HDR capture devices have been introduced to the market. Although these camera are expected to replace conventional cameras in the near future, currently are rather expensive. Among all of them, three of them have been included here: Civetta360 by (Weiss-AG, 2012), SpheronCamHDR by (Spheron, 2012), and Panoscan MK-3 by (Panoscan, 2012). Table 2.1 summarises the three cameras.

Camera	Dynamic Range (f-stops)	Maximum Resolution (Pixels)
Civetta360	30	14144 × 7072
SpheronCamHDR	26	10624 × 5312
Panoscan MK-3	11	12000 × 6000

Table 2.1: A summary of HDR capturing devices (Banterle et al., 2011).

Recently, several camera manufacturers such as Canon, Fuji, Nikon, Sony, etc. have

introduced a HDR capturing features in their consumer products. Such features include multi-exposure capturing, automatic exposure bracketing, and automatic exposure merging.

2.2.3 Digital Image Synthesis

Computer Graphics (CG) rendering methods are another method of creating HDRI. Digital image synthesis is the process of capturing perspective of virtual scenes, which are composed of objects, materials, and lighting, using a virtual camera. There are two algorithms usually employed for rendering which are ray tracing (Whitted, 1979, 1980) and rasterization (Akenine-Möller et al., 2008; Akenine-Möller et al., 2002). Since in this thesis, we scope our study only on digital photography, thus the method of digital image synthesis is not interested and will not be further investigated. For more detail about the technique, reader is referred to (Banterle et al., 2011).

2.3 HDR Image Encoding

An image representation is normally known as an image format or an image encoding. The main purpose of the encoding is to maintain the consistency between the original and the reproduction. In other words, to reproduce the same appearance of an output to the input.

Conventional image encodings store image intensities between 0 and 1 in different step-size depending on the bit-depth, typically 8-, 12-, or 16-bit/color component required by each encoding. A value of zero represents total black, while a value of one represents total white. An HDR image format can store data outside this range i.e. either negative or above the value of one.

In order to fully represent the whole dynamic range of the captured scene HDR encoding requires a more accurate representation (more bit depth) compared to the LDR.

Parallel to LDR image encodings, a small number of HDR image encodings have been developed (Bogart et al., 2003; IEEE, 1985; Larson, 1998; Larson and Shakespeare, 1998; scRGB, 2003; Ward and Simmons, 2004, 2005). Among all of them, four of the encodings related to this work are described here.

2.3.1 IEEE Float Format

In thinking about encoding large numbers of brightnesses, we can start with the Standard IEEE 32-bit/component TIFF float format (single precision floating point numbers) (Hough, 1981; IEEE, 1985). This format is an ideal HDR image representation: we have the same representation as that used for all real number computations on our computer. This format, as illustrated in Table 2.2, covers 79 orders of magnitude but takes 96-bit per pixel. However, as can be guessed, the main drawback of this format is that it takes up more space than any other encodings. Moreover, the floating-point values do not compress well using advanced entropy compression (ZIP) the most one can get is less than 10% (Reinhard et al., 2005). Because of the large file size, the format rather serves as a lossless intermediate representation than an actual practical HDR encoding. In addition, and more importantly to this thesis, since the format is the ideal representation, it also serves as a reference for evaluating different HDR representations.

2.3.2 Radiance RGBE Encoding

RGBE radiance format (.hdr) was introduced in 1998 (Larson and Shakespeare, 1998), and has become the most common alternative to the TIFF float format for storing HDR photographs. Efficiently, the encoding uses the common exponent - mantissa scientific notation e.g. $1245 = 1.245 \times 10^3$, here, 1.245 is the mantissa and 3 is the exponent. The encoding appends the same 8-bits exponent channel for three 8-bit RGB mantissas at each pixel. This results in a 32-bits/pixel encoding. This relatively low memory usage

(compared to the IEEE). As illustrated in Table 2.2, RGBE covers a dynamic range of about 76 orders of magnitude, making the format suitable for almost all of the scenes. The RGBE encoding is encoded as follow:

$$E = \lceil \log_2(\max(R_W, G_W, B_W)) + 128 \rceil \quad (2.3)$$

$$R_M = \left\lfloor \frac{256R_W}{2^{E-128}} \right\rfloor \quad (2.4)$$

$$G_M = \left\lfloor \frac{256G_W}{2^{E-128}} \right\rfloor \quad (2.5)$$

$$B_M = \left\lfloor \frac{256B_W}{2^{E-128}} \right\rfloor \quad (2.6)$$

and decoded as:

$$R_W = \left\lfloor \frac{R_M + 0.5}{256} 2^{E-128} \right\rfloor, \quad (2.7)$$

$$G_W = \left\lfloor \frac{G_M + 0.5}{256} 2^{E-128} \right\rfloor, \quad (2.8)$$

$$B_W = \left\lfloor \frac{B_M + 0.5}{256} 2^{E-128} \right\rfloor. \quad (2.9)$$

Mantissas E are stored in an 8-bit unsigned integer. The range of luminances that can be stored in the format is $[2^{-128}, 2^{127}]$ (approximately $[10^{-38}, 10^{37}]$). The format on average can be compressed approximately 25% (Reinhard et al., 2005).

RGBE, however, introduces complexity in encoding/decoding HDR images because it applies the exponent to the mantissas of each individual pixel. Using the format necessitates a large amount of encoding/decoding time and making it incompatible with many existing image handling applications. In addition, although the ability to encode 76 orders of magnitude is impressive, as mentioned earlier, real scenes with the dynamic

range of more than ten orders of magnitude is hard to find. Thus, 76 orders of magnitude are far beyond the practical dynamic range and unnecessary.

RGBE has an extension to an encoding based on CIE XYZ (CIE, 1931). This format, which has particular properties, which interesting, is not directly relevant to our study. And so will not be discussed further.

2.3.3 OpenEXR Encoding

OpenEXR (Bogart et al., 2003; Magic, 2003) or simply EXR for short, proposed by Industrial Light and Magic (ILM) in 2003, is another HDR image format widely used in computer graphics applications e.g. visual effects and animation. OpenEXR encodes each color component using 16-bit/component floating point number (half precision of the IEEE 754 standard (IEEE, 1985)). The encoding is defined as (Banterle et al., 2011):

$$H = \begin{cases} 0 & \text{if } (M = 0 \wedge E = 0), \\ (-1)^S 2^{E-15} + \frac{M}{1024} & \text{if } E = 0, \\ (-1)^S 2^{E-15} \left(1 + \frac{M}{1024}\right) & \text{if } 1 \leq E \leq 30, \\ (-1)^S \infty & \text{if } (E = 31 \wedge M = 0), \\ \text{NaN} & \text{if } (E = 31 \wedge M > 0), \end{cases} \quad (2.10)$$

where S is a signed bit, E is the exponent occupies 5 bits, and M is the mantissa occupies 10 bits. Thus, the total bits required to encode a single pixel is 48. The encoding covers around 10.7 orders of magnitude (around 30 f-stops of exposure).

2.3.4 scRGB Encoding

scRGB was first proposed by Microsoft and Hewlett-Packard, and accepted as an IEC standard (61966-2-2) (scRGB, 2003). scRGB actually is an extension to the existing conventional sRGB also developed by the two companies (Stokes et al., 1996) i.e. the ITU-R BT.709-5 primaries (ITU, 2002) were chosen to be consistent with sRGB. The encoding extends 8-bit to linear 16-bits per R, G, and B, or to gamma encoded 12-bits per channel (the encoding can encode the data exceeds the $[0, 1]$ range). scRGB allows work in a larger color gamut, wider dynamic range with a price of additional bits compared to the conventional sRGB.

The scRGB has two variants: linear and non-linear (gamma encoded) variants. The linear variant (scRGB48) requires 48 bits/pixel to encode an RGB. A linear encoding is used here as it simplifies graphics hardware and image processing operations. However, this come with a price of dynamic range it can cover since a linear encoding spends many of its precision at the highlight where our eye can detect little difference in adjacent intensities. As a result, this variant covers only about 3.5 orders of magnitude. The non-linear variant (scRGB-nl) of the encoding is of particularly interest to this thesis. The variant requires 36 bits/pixel (25% fewer bits compared to the first variant) to encode 3.2 orders of magnitude. Gamma encoding scheme is employed here together with a linear subsection near zero (as found in sRGB encoding). Unlike RGBE, scRGB effectively gives an HDR encoding using a ‘fixed’ exponent.

However, in order to design this encoding to be consistent with the sRGB, a number of constraints are applied e.g. the same primaries were chosen (the value of 1 of each RGB component of the encoding represent the sRGB’s primaries), or the linear encoding found in the 16 bit variant. These constraints however limit both application compatibility (not so many applications that support signed encoding), and the affordable dynamic range that the encoding can handle which is the key of HDR image encoding.

2.3.5 HDR Encoding Comparison

How do we compare different encodings? Well, we note that a continuous interval (say $[10^{-2}, 10^2]$) is encoded using a finite number of binary strings. So there is a step from one number to the next.

Table 2.2 lists HDR encodings described above along with their properties. The conventional LDR 24-bit RGB encoding (sRGB) is also included in the table as a point of comparison. The dynamic range is given as order of magnitude (logarithm base 10 of the maximum over the minimum intensities, more details on how to calculate this can be found in Section 2.7.1). In addition, the actual intensities are also given in parentheses. Dynamic range of gamma encodings are hard to pin down because the relative step-size is not constant across the whole values. Human can detect luminance changes as small as 2%, however, in the darkest areas of the image this may go up to 5% unnoticed (adjacent steps in the encoding are easily distinguished at step sizes above 5%). Thus, we chose 5% as the minimum value in order to calculate the dynamic range of gamma encodings, and report ‘Variable’ as their relative step.

The goal of most HDR encoding is to keep this relative step size as low as possible to avoid possible quantization artifact. Reinhard et al. (2005) suggested that encodings with variable step size are ill suited as the HDR image encoding.

Encoding	Color Space	Bit/pixel	Dynamic Range (\log_{10})	Relative Step
sRGB	RGB in [0,1] range	24	1.6 orders (1.0:0.025)	Variable
IEEE	RGB	96	79 orders	0.000003%
RGBE	Positive RGB	32	76 orders ($10^{38} : 10^{-38}$)	1.0%
EXR	RGB	48	10.7 orders (65000:0.0000012)	0.1%
scRGB48	RGB	48	3.5 orders (7.5:0.0023)	Variable
scRGB-nl	RGB	36	3.2 orders (6.2:0.0039)	Variable

Table 2.2: A comparison chart of HDR pixel encodings. The dynamic range quoted for sRGB, ScRGB48, and scRGB-nl are based on the point at which their relative steps pass 5% (Reinhard et al., 2005).

As can be seen from the table, a conventional 8-bit/channel RGB gamma-encoding format or “sRGB” cannot encode HDR data since the format covers only 1.6 orders of magnitude. It is clear that a standard IEEE 32-bit float format is a full representation, however requires too much storage space (96-bit/pixel). RGBE seems to be the encoding that while covers a dynamic range of 76 orders, only 32 bit/pixel is required for storing such a wide range of data. However, as discussed earlier, 76 orders of magnitude is much far beyond the useful dynamic range of real-world scenes. scRGB both linear and gamma-encoded variants are not real high dynamic range since the dynamic ranges that they can cover are less than what human can perceive, and too limited compared to the useful dynamic range. In addition, they cover only about an order of magnitude beyond the conventional 24-bit sRGB encoding. For these reasons, the scRGB is sometimes categorized as a Mid Dynamic Range (MDR) image encoding and is not generally used to encode HDR images.

For clarification, in this thesis, Low, Mid, and High dynamic range are defined to have the magnitude of less than three, larger than three but less than nine, and larger than nine orders, respectively. Since nine orders of magnitude is equal to the dynamic range that our HVS can cope with (Seetzen et al., 2004).

2.4 HDR Display

Real scenes typically contain a range of illumination much larger than can be reproduced by conventional images. As discussed earlier, conventional image formats encode only about two orders of magnitudes. This limited dynamic range designed to only match the conventional CRT displays, and to all displays that are designed to be compatible with them. New LED monitors tend to have a wider dynamic range (some claim to have up to millions to one of contrast ratio). However, although they cover larger dynamic range, the extended range often lies in the dark end where the HVS can not

discriminates this difference under normal viewing environments (Reinhard et al., 2005). This means that the useful dynamic range of these displays is still somewhere around the magnitude of two or three.

A limitation of conventional displays is that they have a single uniform backlight. It is hard to generate hard black using an LCD panel even when the RGB is set to zero. Of course a spatially-varying backlight would make black a black, spatially-varying backlight are the key to the HDR display.

In recent years, novel display technologies that can natively display HDR contents without the need of TMO were proposed. Based on the pioneering work by Seetzen et al. (2004), HDR displays can be divided into two types. The first type is a combination of digital light projector (DLP) and an LCD panel. For the second type, the projector is replaced with a panel of high intensity light-emitting diodes (LEDs). Both DLP and LED HDR displays use the same principle. The principle is to use a DLP/LEDs to modulate the light. As shown in Figure 2.3, an HDR display substitutes the fluorescent backlight unit with a DLP projector or LED panel to modulate the light source while the front LCD panel encodes details and colors. As a result, the display can reproduce images with a higher dynamic range than typical displays as the backlight is now spatially-varying - dark areas appear darker and higher contrast compared to the uniform backlight modulation. Table 2.3 gives properties of the two types of the HDR display.

Type	Max Luminance (cd/m^2)	Min Luminance (cd/m^2)	Dynamic Range
Projector-based	2,700	0.054	50,000:1
LED-based	3,000	0.015	200,000:1

Table 2.3: Properties of HDR display by type (Banterle et al., 2011).

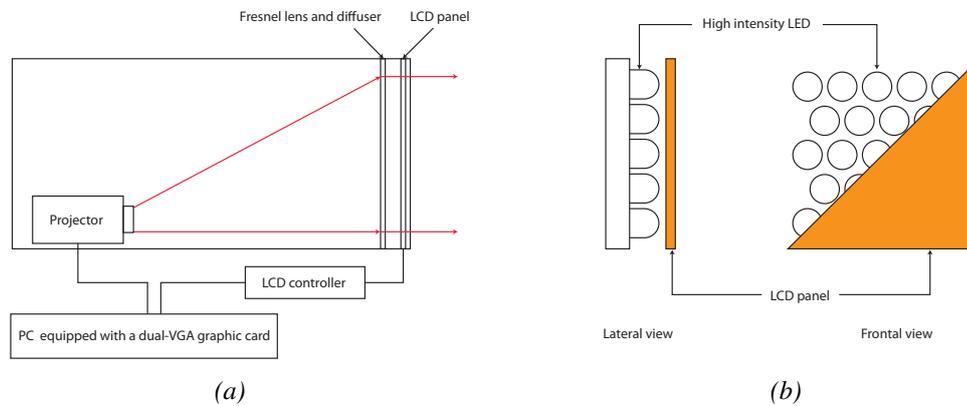


Figure 2.3: Schematic diagrams of HDR display technologies. (a) A projector-based scheme (Seetzen et al., 2004). (b) An LED-based scheme. (Banterle et al., 2011).

2.5 Tone Mapping Operators

HDR images can have a much higher dynamic range than that can be reproduced by conventional imaging devices. As we have seen, HDR radiance maps can have a dynamic range up to ten orders of magnitude. As mentioned earlier, HDR photographs cannot be properly displayed on typical displaying devices due to the huge difference in dynamic range. Figure 2.4 compares the data flows of LDR and HDR imaging systems. As can be seen, the difference is significant, the dynamic range of the captured HDR image exceeds that of the conventional display’s capability significantly. Simple tone mapping methods e.g. linear, gamma, log, as found in LDR tone reproduction are not enough to compress this large difference. As a consequence, much of the information of the HDR representation is lost (either by truncation or discretization). To visualize HDR contents on these devices, an additional process that performs dynamic range compression to properly reproduce the output images with a similar appearance to the real scene (similar to what we perceived) is needed. Such a process is known as a “Tone-Mapping Operator” (TMO).

Most TMOs process only luminance information; chromatic information is left un-

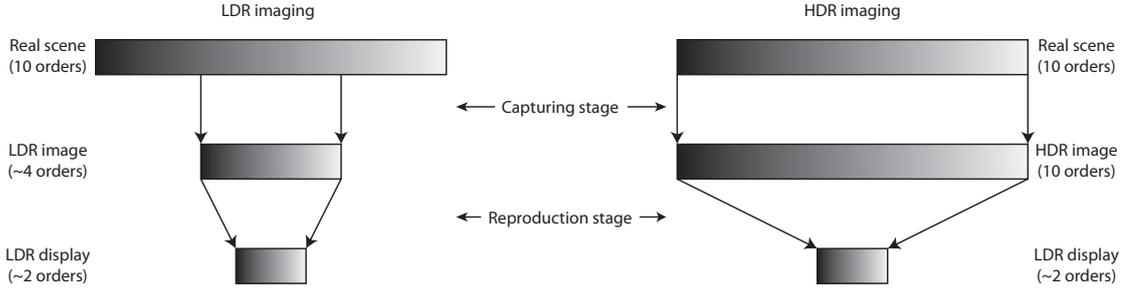


Figure 2.4: Variation in dynamic range of LDR and HDR imaging systems. The middle bars represent the dynamic range that the each system can record. As can be seen, LDR imaging can record only a partial range of the real world illuminant. HDR imaging on the other hand, can record all of the real world illuminant. The bottom bars represent the dynamic range of a conventional display (8 bit representation). While, there is a minor change in dynamic range compression of the LDR imaging, there is a significantly change in dynamic range compression of the HDR imaging.

altered (all of TMOs discussed in this work operate in this manner). Thus, for the sake of simplicity, we limit our discussion only on the luminance of the HDR. Later in Section 2.6.3, we will show how to reconstruct the color information from the tone mapped luminance. Mathematically, a tone mapping operator f that maps HDR encoding value to display pixel value is defined as:

$$L_d = f(L_w) : \mathbb{R}^{w \times h} \rightarrow \mathbb{D}^{w \times h} \quad (2.11)$$

Where L_d is the display luminance, L_w is the luminance of the HDR image, w and h are width and height of the image. In the case of conventional displays $\mathbb{D} = [0, 255]$.

Figure 2.5 compares HDR images resulted from three different rendering techniques; 2.5a is displayed without tone-mapping, 2.5b is displayed with linear-scaling, and 2.5c is tone-mapped with a proper TMO. As can be seen, most of pixels of the direct displaying is quantized into the darkest display intensity, thus the image looks dark. A simple compressive function such as linear scaling also fails to produce plausible outputs 2.5b. It is sufficiently only if the dynamic range of the image is slightly higher than the target dynamic range. For image with a higher dynamic range, small differences between intensities will be quantized to the same display intensity and res-

ults in the lost of visible details. This indicates that a simple compressive operator is insufficient and naïve to serve as a tone-mapper since the key issue in the tone-mapping is to compress an image, while retaining one or more image's attributes. In comparison, a proper tone-mapped image 2.5c compresses the dynamic range in a controlled fashion. As a result, both of the highlight and shadow of the scene are well preserved and does not have either over-exposed or under-exposed areas, and it looks more visually pleasing.



Figure 2.5: HDR imaging techniques applied to the scene shown in Figure 1.2. (a) HDR image directly displayed, (b) HDR image linearly scaled, and (c) tone-mapped HDR image showing a superior amount of information compared with the two previous images. (this image is tone-mapped using the algorithm proposed in Chapter 5). Both direct displaying and linear scaling often end up with the image that hardly resemble to the captured scene. In comparison, the tone-mapped image shows details in both the light and dark areas.

So how does tone-mapping work? Well the aim is simple to articulate. We wish to

map the input brightness range to that which is displayable so we can see all the detail in an image. Further, the reproduced image should correlate with our visual memory. Figure 2.6 illustrates the concept.

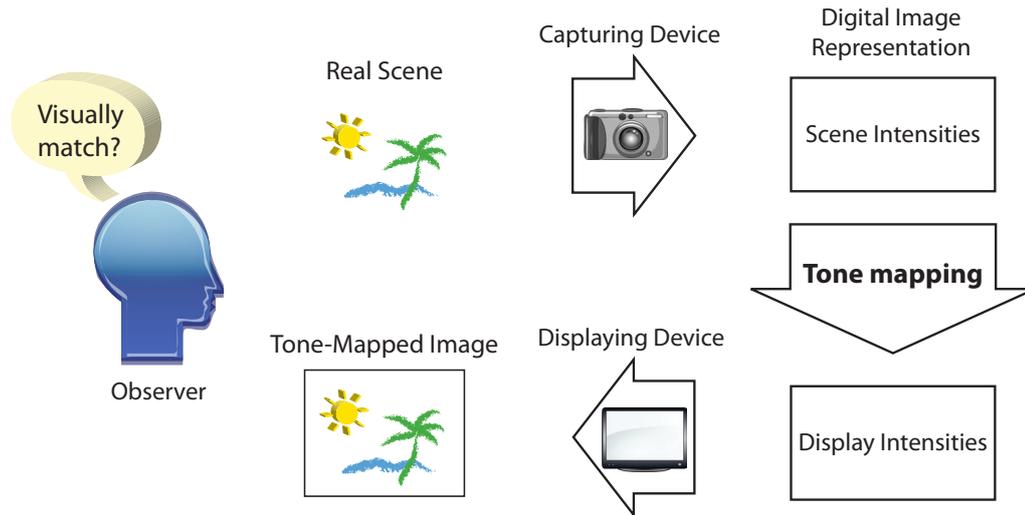


Figure 2.6: Reproduction images should be visually similar to the observed scene.

Various tone-mapping operators have been developed in recent years, ranging from a simple gamma adjustment to a more complex histogram adjustment, or even more complex techniques such as lightness perception models (Reinhard et al., 2005). In general, TMOs can be categorized in different classes based on the image processing techniques they use: global (spatially-uniform) operators, local (spatially-varying) operators, frequency/gradient operators, and segmentation operators (Banterle et al., 2011). However, in the broadest sense, they could be classified into global and local operators since frequency/gradient and segmentation operators are categorized as local operator (Dicarlo and Wandell, 2000a; Ferwerda, 1998; Tumblin and Turk, 1999).

For the rest of this section, the details of some of the most commonly used tone-mapping operators are given in the order organized by image processing techniques described above. These operators at the same time are the operators that relate to the

work presented in this thesis. For a comprehensive study of tone mapping operators, the reader is referred to (Banterle et al., 2011; Devlin, 2002; Reinhard et al., 2005).

2.5.1 Global Operators

The classical way of handling dynamic range is based on global operation. Global TMOs (henceforth G-TMOs) are non-linear functions that map an input HDR image to the output LDR image for display. The terms global and spatially-uniform stem from the fact that the same operator f is applied across the entire spatial domain (to all pixels of the image) independent from surrounding pixels in the image. Typically, f is parameterized by simple image statistics based on image attributes drawn from the full image and are later used to optimize the dynamic range compression. Such global statistics include minimum luminance, maximum luminance, and logarithmic mean of the luminance. Once the global function is found, every pixel in the image is mapped in the same way.

Global operators range from simple operators that based on basic functions (linear scaling, power law [gamma], logarithmic, sigmoid, etc.) to more complex functions that can be vary according to image properties in which typically store control parameter as a 1D look-up table or as commonly known as Tone-Curve (TC).

One of the very first TMOs is proposed by Tumblin and Rushmeier (1993) and later is revised in (Tumblin et al., 1999). The operator attempts to match the brightness of the displaying devices to the brightness of the perceived scene. Brightness is referred to the perception of luminance (Wyszecki, 2000). They build upon the results of the brightness measurement by Stevens and Stevens (1963, 1960) that indicates there is exist a power law relationship between luminance and perceived brightness.

Larson et al. (1997) proposed one of state-of-the-art global TMOs that is simple but effective and is based on the classical imaging technique of histogram equalization (Gonzalez and Woods, 2006) (see Section 2.7.2). Histogram equalization adjusts the

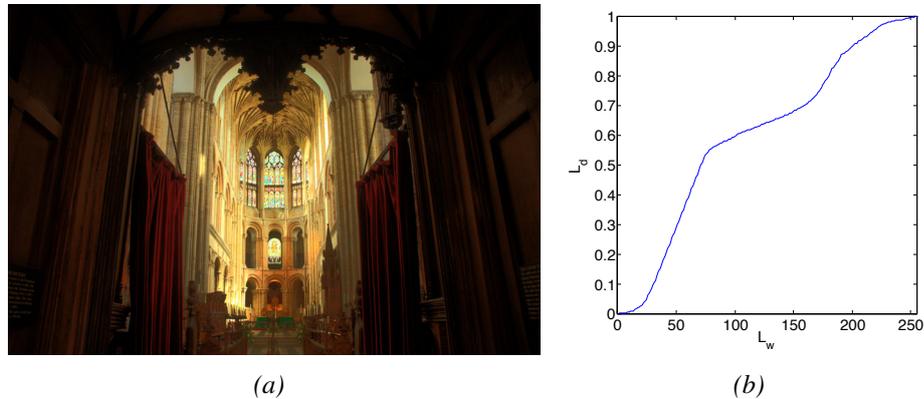


Figure 2.7: An example of the Histogram adjustment by Larson et al.. (a) The tone-mapped image. (b) The mapping function was created by reshaping the cumulative histogram of the image shown in Figure 1.2.

input values so that the probability of each display value occurs in the output image equally. Histogram equalization often makes unnatural images and introduces artifacts such as ‘contouring’ into the image. While this method is a significant improvement, histogram equalization type artifacts remain. An example of the mapping function generated by this concept is illustrated in Figure 2.7.

In general global TMOs are simple (they make use of global image statistics) and fast (since they can be implemented using tone-curves) which are suitable for applications that have limited resources such as in-camera image processing. However, by their very nature G-TMOs compress or stretch the input signal. If the slope of the G-TMO function is less than one then detail is compressed in the output images. Such compression often happens in highlight areas of an image and if it does the highlight of the output image appears flat. Often G-TMOs are unable to maintain local contrast and the subtle details of the original scene.

2.5.2 Spatially-Varying Operators

Local or Spatially-varying TMOs (SV-TMOs) take into account the spatial context when they adjust pixel intensities. In other words, the parameters of the non-linear function f change at each pixel according to the local features extracted from the neighboring pixels rather than from the full image (Ashikhmin, 2002; Chiu et al., 1993; Durand and Dorsey, 2002; Fairchild et al., 2004; Jobson et al., 1996; Meylan and Susstrunk, 2006; Pattanaik et al., 1998; Rahman et al., 1996, 1997b; Reinhard et al., 2002). The logic is that a bright pixel in the bright neighborhood should be perceived differently from the one in a dark neighborhood. Figure 2.8 shows an example of an HDR scene that is impossible for global TMOs to maintain local details, and thus requires local processing when displayed on a conventional display. As can be seen, although the overall appearance of the two images are very similar, local TMO preserves details in both highlight and shadow areas, which fail to preserve with global TMO (compare the street lamp and the spiral staircase in both images).

Unfortunately, spatially-varying technique can however cause halos around high contrast edges if the function is not applied appropriately. To understand why SV-TMOs introduce halos to the tone-mapped image, let's consider the surround-based retinex (Land, 1974) which is known to be fundamental of many spatially-varying operators. In surround-based retinex, output pixel values are calculated by the ratio (or difference in log space) between each pixel and a weighted average of its surround. The simplest form of center-surround retinex algorithm is called ‘‘Single Scale Retinex (SSR)’’ and is defined as:

$$R(x, y) = \log I(x, y) - \log [G(x, y) * I(x, y)] \quad (2.12)$$

$$= \log \frac{I(x, y)}{G(x, y) * I(x, y)} \quad (2.13)$$

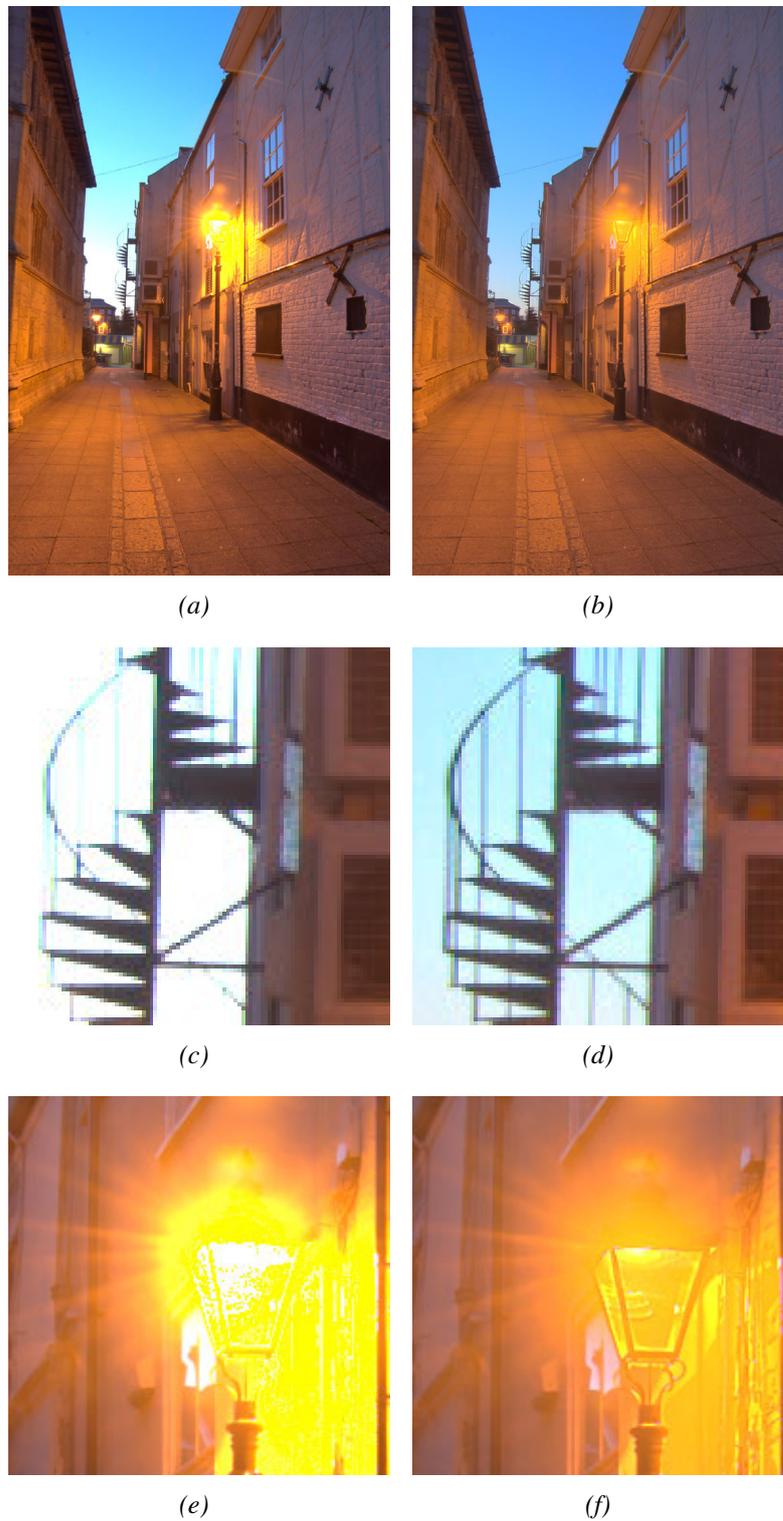


Figure 2.8: A comparison between a global tone-mapped image (a) and a local tone-mapped image (b). The local tone-mapped image preserves more detail, as for instance seen in the insets (d) and (f) compared to (c) and (e), respectively.

Where $I(x, y)$ is the input image, $*$ denotes convolution operator, and $G(x, y)$ is a Gaussian weighting filter.

The cause of halos originates from the fact that neighborhood weighting filters (e.g. Mean, Median, or Gaussian which in this case is Gaussian) used in the operator blur across two areas of very different intensity i.e. high contrast edges. If a dark area is close to a bright area (as illustrated in Figure 2.9), the bright pixels overly influence the processing of the dark pixels, which results in a black halo around the bright area, in the same sense, the dark pixels also influence the processing of the bright pixels, which results in a white halo around the dark area. In other word, the intensity inversion appear around the edge of the tone compressed image is known as halos. Figure 2.10 shows example images with halos.

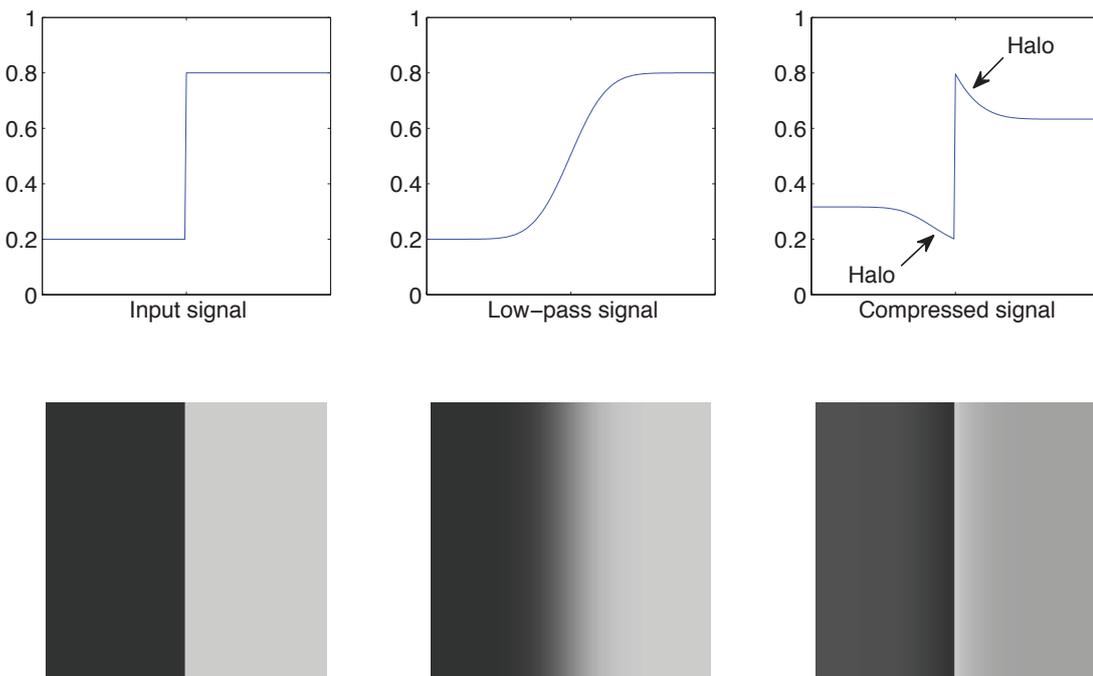


Figure 2.9: A diagram illustrates the halo artifact. (bottom left) Original image with a high contrast edge in the middle. (bottom middle) Smoothing filter such as Gaussian blurred out the edge. (bottom right) A tone-compressed output resulted from the input minus the blurred image in log unit. The intensity inversion appeared around the edge is known as halos.

Larger filter size can alleviate this problem since such large contrasts are averaged



Figure 2.10: Halo artifact introduced by spatially-varying TMOs. Both images were tone-mapped using Retinex operator (The original HDR images are courtesy of Dani Lischinski and Paul Debevec).

out. However, if the filter size approaches infinity, the local average becomes identical to the global average. In other word, a very large filter size results in the similar result to the global operator. Therefore, the compressive power of the local operator is limited to be no better than a global operator, and indicates the trade-off between the occurrence of artifacts and the dynamic range compression ability. This phenomenon is also illustrated in Figure 2.11. The middle image was obtained with center surround retinex using a small spatial filter. As can be seen, detail visibility is improved but unwanted halo artifacts appear around high contrast edges. The right image was obtained with the same operator but with a large filter. Here, there is no halo, however, the local contrast is not satisfying enhanced. This tradeoff is discussed in (Ashikhmin, 2002; Dicarlo and Wandell, 2000b; Durand and Dorsey, 2002; Fattal et al., 2002a; Reinhard et al., 2002; Tumblin and Turk, 1999) and in general reviews in (Devlin, 2002; Reinhard et al., 2005). Thus, when come to choose the neighboring pixels, care should be taken otherwise, halo artifacts around strong edges can appear.

To reduce halos without sacrificing the dynamic range compression capability, a number of local techniques (Ashikhmin, 2002; Dicarlo and Wandell, 2000b; Durand and Dorsey, 2002; Fattal et al., 2002a; Jobson et al., 1996; Meylan and Susstrunk, 2006; Rahman et al., 1996, 1997b; Reinhard et al., 2002; Tumblin and Turk, 1999) have been

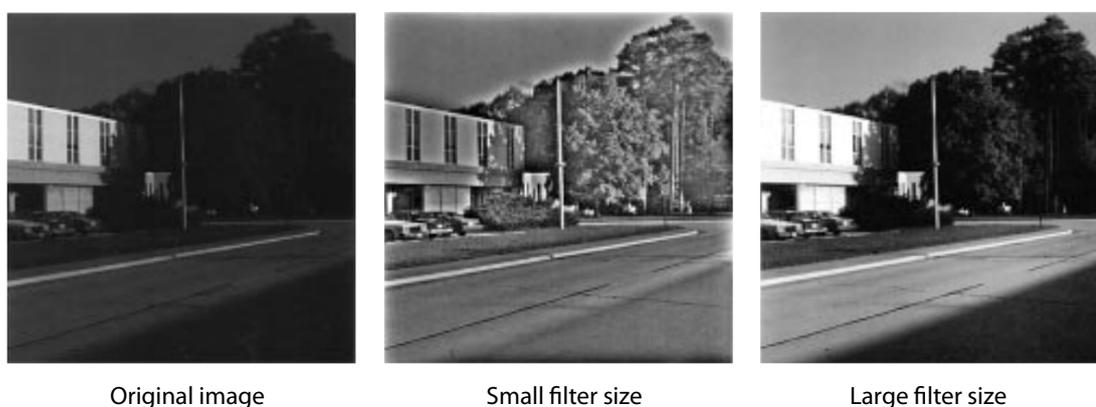


Figure 2.11: A small filter size leads to halo artifacts, whereas a large filter size alleviate these artifacts (they are much reduced) but the detail in the output can be more muted (the power of local operation is reduced). (left) Original image. (middle) Small filter is used. (right) Large filter is used (Images are courtesy of Rahman et al. (1997a)).

proposed. Among all of them, three techniques related to the work presented here are described.

Reinhard et al. (2002) presented a photographic tone reproduction technique, which is analogous to zone-system and dodging-and-burning techniques (Adams, 1971a,b, 1972) technique found in traditional photography. The concept is to find the largest circular surrounding area at each pixel that does not contain any high contrast edges in order to avoid halos. To achieve this, a set of Gaussian weighted average that vary in size is generated, if the difference between the two average is close to 0, there are no high contrast edges in this surround. The method provides an efficient way of compressing the dynamic range while reducing halo artifacts. However, the circular surround limits its performance.

Meylan and Susstrunk (2006) proposed a tone-mapping operator based on a center-surround retinex algorithm with some modifications, and named it “Retinex-Based Adaptive Filter”. Typically, retinex algorithm uses symmetrical filter. However, as discussed above, symmetrical filter results in halos. Here, to prevent halos, an adaptive filter whose shape of it depends on the high contrast edges in the images is used instead.

Another alternative approach to alleviate the problem of halo includes the use of edge-preserving filter. Such filters are designed for removing small details while keeping sharp edges untouched, therefore, halos are reduced. One of the well-known edge preserving filter is a Bilateral filter. A bilateral filter is a non-linear filter proposed by Tomasi and Manduchi (1998) that smooths image details while preserve strong edges. In contrast to traditional filtering that handles only geometric closeness, bilateral filtering adds photometric distance enforcement in which it can explicitly average only perceptually similar intensities. By using bilateral filtering an image can be decomposed into two different scales: a large-scale component where edges are preserved - called the base layer - and a small-scale component called the detail layer defined to be the original image minus the base layer (in log-domain). Durand and Dorsey (2002) adopted this concept and proposed a HDR tone-mapping operator which preserves local contrasts. Their idea is to tone map only on the base layer and later add back detail to the tone-mapped based image. This makes intuitive sense. If the tone mapper is the identity function then the output = base + (original-base) i.e. the original image (which is what we would expect). Adding the details back ameliorates the ‘detail loss’ problem of traditional tone mappers.

Generally, neighboring pixel values, and the size of the local neighborhood for each pixel are often used to derive the parameters of these operators. SV-TMOs provide an efficient way of compressing the dynamic range than global TMOs since the operators take local contrast into account. However, the introduction of halos limits their performance. Further, SV-TMOs are far more complicated in term of computational cost than global TMOs.

2.6 Post-Processing

After tone-mapping, it is possible to apply several post-processing steps to either enhance the appearance of the output image or correcting some of the issues introduced during the tone-mapping process. In this section, three post-processing techniques that have a high impact on the overall appearance of the output images, a black and white point correction technique, a color reconstruction technique, and a gamma adjustment, are described.

2.6.1 Black and White Point Correction

Black and white point correction is an operation that matches the perceived black to the darkest display luminance and the perceived white to the brightest display luminance. The operation can be embedded in the tone reproduction operator such as histogram equalization technique, or can be used as a post-processing in addition to the TMO. In this work, the latter case is considered.

The reason that one might need to do this is because the black and white points of the HDR cannot simply be determined by the darkest and brightest pixel. Indeed, they could be an outlier due to the signal noise. Instead, group of pixels of lowest and highest should be used. One possible way to do this is to use a histogram-based technique on a luminance component.

The image I is linearly scaled by matching the black and white points b and w (the pixel values that fall within the specified bin numbers) into the range $[0, 1]$. The process is defined as:

$$I_{new}(p) = \min \left(1, \frac{\max(0, I(p) - b)}{w - b} \right) \quad (2.14)$$

Image pixel with values below b or above w are clipped to 0 or 1 respectively. Figure 2.6.1 illustrates an example of black and white point correction where the black

and white points b and w are defined by 1% and 99% of the total bins. As can be seen, 2.12b the histogram of the input image indicates there are only few pixels have lowest or highest pixel values. By correcting black and white points, the pixel values of the output image are redistributed.

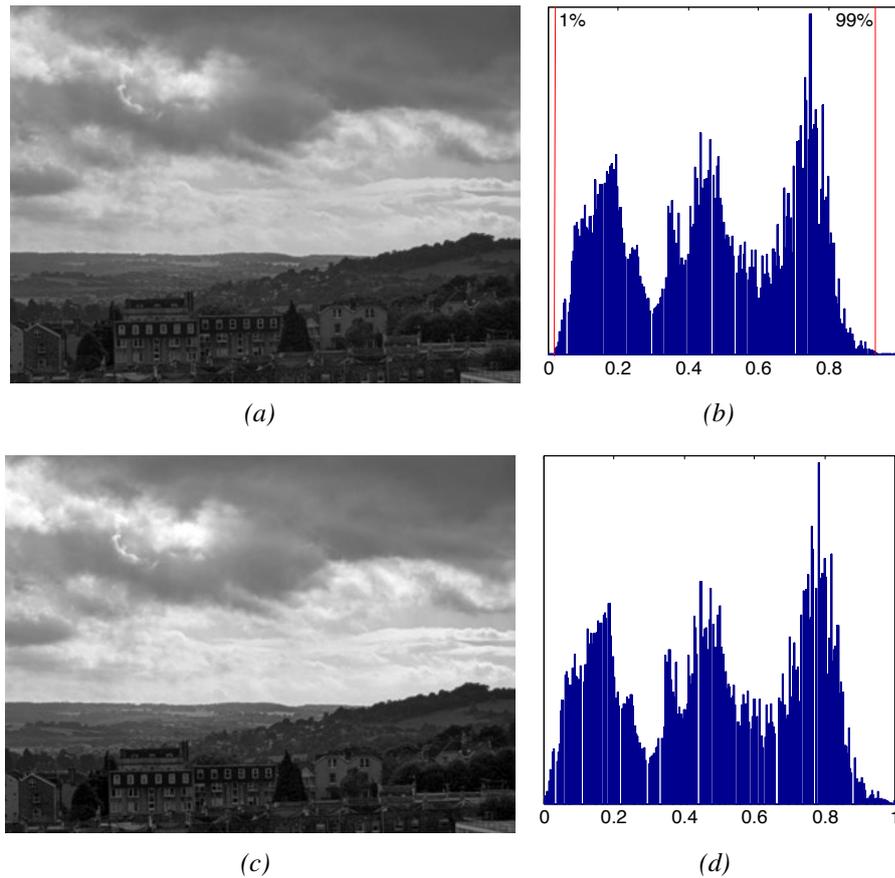


Figure 2.12: An example of histogram-based black and white point correction. (a) Input image. (b) Histogram of the input image. (c) Output image after the black and white point correction. (d) Histogram of the output image.

2.6.2 Gamma Adjustment

Gamma adjustment is one of the simplest operation to adjust the final image rendering, encode image data, and compensate the display characteristics. It is often known as

“gamma encoding” when it is used for the middle case, and “gamma correction” when it is used for the latter case. Gamma adjustment in the context of HDR post-processing provides contrast adjustment to the tone-mapped image. The simplest form of the operator is defined by the following power-law expression: ¹

$$L_{out} = L_{in}^{\gamma} \quad (2.15)$$

where the input (L_{in}) and output (L_{out}) values are non-negative real values, typically in the normalized range of $[0, 1]$. If γ is larger than 1, the output image will be darkened, in contrast, if γ is smaller than 1 the output will be brightened. Gamma adjustment can be thought as the contrast adjustment operator. Typically, γ is a constant manually controlled by the user. In Chapter 4, we propose an effective way to estimate the choice of γ based on the image statistics called image entropy.

2.6.3 Color Reconstruction

To avoid altering color of the image, most TMOs typically operate on luminance component e.g. corresponding to the Y component in CIEXYZ color space, rather than directly operate on the red, green, and blue color components. It is well known that direct RGB modifying often alters the physical color property of the input image and the HVS is extremely sensitive to this. Schlick (1994) proposed the technique of preserving the ratios between color components. Instead of scaling all three color components with a non-linear function, the luminance values are derived from the HDR $L_w(x, y)$. Next, $L_w(x, y)$ is then compressed into display values $L_d(x, y)$. Finally, to reconstruct a color output image, the ratio of $L_d(x, y)$ to $L_w(x, y)$ is used as follows:

¹Power-law function is also commonly known as “gamma”, referring to the value of the exponent.

$$I_{r,d}(x, y) = \left(\frac{I_r(x, y)}{L_w(x, y)} \right)^s L_d(x, y), \quad (2.16)$$

$$I_{g,d}(x, y) = \left(\frac{I_g(x, y)}{L_w(x, y)} \right)^s L_d(x, y), \quad (2.17)$$

$$I_{b,d}(x, y) = \left(\frac{I_b(x, y)}{L_w(x, y)} \right)^s L_d(x, y) \quad (2.18)$$

where the exponent s is a saturation factor in which $s > 1$ increases the saturation, $s < 1$ decreases the saturation.

Figure 2.6.3 illustrates the effect of this per channel gamma correction saturation factor s found in Equation 2.18.

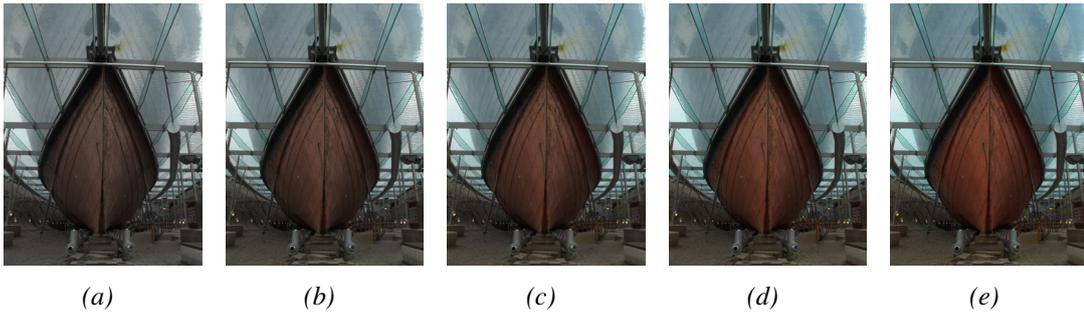


Figure 2.13: The effect of the saturation factor on the tone-mapped image. From left to right, the saturation factor is set as 0.6, 0.8, 1.0, 1.2, and 1.4, respectively.

2.7 Image Statistics

Several attributes of images can be reported as image statistics. These statistics reveal valuable information of the image. In addition, for some applications, such as machine vision, these statistics can be utilized to make automated operations. In this section, two image statistics related to the study are explored: dynamic range, and the entropy of image.

2.7.1 Dynamic Range

Dynamic range is the ratio between the largest and the lowest value of a variable quantity. In photography dynamic range refers to the ratio between the brightest and the darkest luminance of that scene (cd/m^2). In display technology, the dynamic range is also known as a contrast ratio or simply contrast, the terms denote the brightness ratio of the black and white pixels displayed on the screen. In the field of HDR, the measurement of dynamic range is becoming important, since the dynamic range directly related to be used to assess how difficult it may be to reproduce tone-mapped images.

In general there are two main approaches to measure the dynamic range of the scene. The simplest approach is to calculate the ratio between the smallest and the largest pixel value of the image. However, there is always noise during the capturing process especially in the darkest regions. These outliers may be reduced by ignoring a small percentage of the darkest and brightest pixels. A more complex way to better measure the dynamic range is to report the signal-to-noise ratio in decibels (DB) unit (Reinhard et al., 2008). In this thesis, the former approach is used. The dynamic range is calculated using the ratio between the smallest and the largest pixel value of the image. In addition, instead of reporting the ratio, one can express it in logarithmic domain. For example, if the maximum and minimum intensities of a given scene are $10^2 cd/m^2$ and $10^{-2} cd/m^2$ respectively, the dynamic range of this scene is 10^4 which can be also reported as four orders of magnitude or equivalently denoted as 10,000:1.

2.7.2 Image Entropy

Image entropy is a quantity which is used to describe the amount of information of a given image. In other word, how expensive it is to encode the image using an optimal encoding scheme. As an example, an image whose every intensity occurs with equal frequency is more expensive to encode than one where a few intensities occur more

frequently. Images that contain a large area of a clear sky, have very little contrast and large runs of pixels with the same or similar intensities values have low entropy. An image that is totally flat will have zero entropy. In addition, as can be guessed, the concept of image entropy are closely related to the field of the image compression. For example, we can encode intensities that occur more often with a small number of bits than ones that less occur.

An image entropy is calculated as

$$H = -\sum_{i=1}^n p(x_i) \log_b p(x_i) \quad (2.19)$$

where $p(x_i)$ is the probability of the occurrence of the i^{th} luminance, and \log_b is the base 2 logarithm.

Important to the work here, the idea of entropy relates to the conspicuity of detail in images. To understand why entropy relates to visible detail in images, in the following sections tone reproduction operators based on image entropy are given.

Histogram Equalization

Histogram equalization (HE) is a tone reproduction technique normally used to improve the global contrast of images. HE adjusts contrast of images based on image histograms by employing a monotonic, non-linear function which reallocates pixel intensities of the input image such that the output image has a uniform intensity distribution (flat histogram). Through the adjustment, the histogram becomes better distributed (close contrast intensities of the input image are spread across the whole range of the histogram). As a consequence, this often results in the increasing of global contrast of images allowing areas with lower contrast to gain a higher contrast. Figure 2.14 gives an example of HE on a grayscale image.

The method is suitable for images that are either bright or dark. Especially, HE can

lead to better views of scientific images such as satellite, x-ray images, and any images that typically process with false-color technique, and to improve detail in photographs that are under or over-exposed. The main advantage of the method is that the technique can be invertible. This means that if the histogram equalization function is known, then the original image can be fully recovered. Another advantage of the method is computationally inexpensive. However, the major drawback of the technique is it may boost the contrast of background noise, while decreasing the usable intensities. It also can produce undesirable effects such as visible banding when applied to images with low bit-depth.

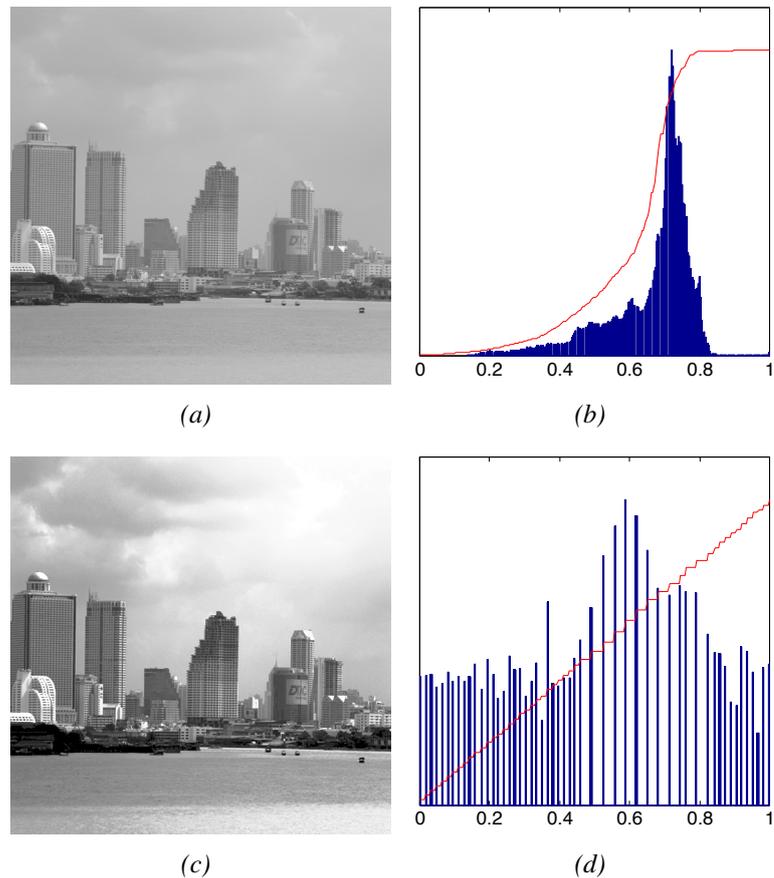


Figure 2.14: Histogram equalization. (a) An unequalized image (b) Corresponding histogram (blue) and cumulative histogram (red) (c) The same image after histogram equalization (d) Corresponding histogram (blue) and cumulative histogram (red)

The Optimal Gamma

Recently work (Finlayson and Xu, 2012) solved for the gamma that resulted in the image with the maximum entropy. Remarkably, assuming a continuous probability distribution of brightness in the interval $[0, 1]$, the optimal gamma can be calculated as:

$$\gamma_{optimal} = -\frac{1}{mean(log(x))} \quad (2.20)$$

Simple algebraic manipulation will convince the reader that once the optimal gamma is applied there is no benefit in finding the optimal gamma for the result (the second gamma will always be 1 the ‘identity’ gamma adjustment).

The gamma adjustment that maximizes entropy works similarly to HE though the amount the image changes will be less. This the authors argue is to the technique’s advantage. In addition, a gamma adjustment in log space is just a scalar multiplication (i.e. contrast adjustment). The mapping relating input and output log values (pre and post gamma adjustment) is just a straight line (whose slope equals gamma). In contrast, HE might apply a very wiggly ‘high frequency’ tone curve and this can result in the well known problems of there being too much contrast and counteracting (edges appearing that were not visible in the original).

2.8 Image Quality Measurement

Digital images are subject to being distorted by various processes during their lifetime, ranging from acquisition, storage, processing, compression, transmission, and reproduction, any of these can degrade the quality of image. The field involves a study of quality of image often referred to image quality measurement. Image quality measurement is roughly classified into two major classes as objective and subjective image quality measurements. The objective measurement is evaluated through physical meas-

measurements of image properties, where the subjective measurement is evaluated through judgment by human observers. This section provides an overview of the image quality measurement using objective and subjective methodologies.

2.8.1 Objective Image Quality Measurement

Although the human observer is the final judge for an imaging system that is designed for human consumption, it is difficult to extract objective data from a given image that rely on human observations. Since, human perception is a combination of psychology, physiology, and environment (Wolin et al., 1998). In addition, apart from these difficulties, there is the need of quantitative analysis. Objective image quality measurement attempts to quantify the perceptual differences (visibility of errors) between a test (distorted) image and a reference image. To evaluate the dis/similarities between test images, an objective image quality metric is used. The image quality measurement metric takes two inputs, one as the reference, the other as the test image. The result metric is a map (often known as an error metric) where each value represents how closely the difference between the two images. Objective image quality metrics can be classified in to three categories according to the availability of a reference (distortion-free) image to the test (distorted) image: full-reference (a reference image is available), no-reference (the reference image is unavailable), and reduced-reference (the reference image is only partially available). In this thesis we scope our discussion on the full-reference image quality metric.

The simplest full-reference objective image quality metric is the Mean-Square Error (MSE). MSE is computed by averaging the squared intensity differences of reference and test image pixels.

Given $x = x_i | i = 1, 2, \dots, N$ and $y = y_i | i = 1, 2, \dots, N$ are two images, where N is the total number of pixels of the images, the MSE between these images is calculated as:

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (2.21)$$

In fact, MSE is equivalent to a variance in statistics. In statistics, MSE is sometimes refers to the variance. Similarly to standard deviation that is the square root of variance, applying a square root to MSE yields the Root Mean Square Error (RMSE). RMSE is a frequently used measure of the differences between the pair of images. In this work, RMSE is used in Chapter 5 (Section 5.2.1).

RMSE although are simple and are mathematically defined in the context of optimization, it is often fail to match the perceived visual quality (Girod, 1993). After all, RMSE is just a scalar value represented the difference between two images, this means that two distorted images that have very different types of errors could possibly have the same RMSE, and some of which are more visible than others. To solve this problem, the development of the image quality metrics that take advantage of know characteristics of the HVS is introduced. In fact, many of the perceptual quality metrics are based on the modification of the RMSE (to measure the error according to the human perception) (Wang et al., 2004).

Several metrics have been proposed to cope with different evaluation tasks. Depending on what needs to be measured, different metrics is chosen. For example, if one wants to measure color different, conventional color different metric such as ΔE CIELAB fit the task. However, if the goal is to compare the perceptual similarity of images, then an error metric derived from HVS is needed here. For example, a spatial extension to the CIELAB color metric called “S-CIELAB” (Zhang and Wandell, 1996) is one of the widely used evaluating metrics for measuring color reproduction errors since S-CIELAB simulate the spatial blurring by the HVS.

One of the well-known error metric that takes into account HVS for HDR TMO comparison is HDR-Visual Difference Predictor (HDR-VDP) (Mantiuk et al., 2005, 2004). Only luminance component is processed which is a key stimulus in the HVS.

The main goal of the visual difference metric is to predict visible difference in HDR images. The metric is an extension of the existing VDP to HDR images. VDP is based on a model of HVS. As with other error metrics, HDR-VDP takes the reference and the tested images as its inputs. HDR-VDP generates a probability map where each value represents how difference between the tested images may be perceived by the HVS. HDR-VDP simulates the contrast reduction in the HVS by simulating light scattering in the cornea, lens, and retina. This takes into account the nonlinear response of human photoreceptors to light (Just Noticeable Difference [JND]).

Here, care should be taken on choosing which metric is the most suitable for the task. This is because up until now there is no such metric that fully simulates the HVS. Indeed, these metrics are merely a simulation of early stages of the HVS that can identify few aspects of the HVS. Since the HVS is a complex and highly non-linear, but most models of early vision are based on linear or quasilinear operators.

Objective image quality measurement can also provide the repeatability (automated) and reliability lacking in the subjective measurement, allowing the measurement of a large volume of test subjects. Reviews on the perceptual image quality metrics can be found in (Pappas and Safranek, 2000; Wang et al., 2004).

2.8.2 Subjective Image Quality Measurement

Psychophysics, psychometric scaling or simply mind-measuring is the field of study of the relationships between the physical measurement of stimuli and the perceptions that these stimuli bring to mind (Fairchild, 2005; Gescheider, 1997).

In psychophysics, the term “scaling” means the creation of a scale of the observers response to a physical stimulus. The result of a scaling is scales of “nesses” or image quality. According to Stevens (1946), there are four types of psychophysical scales: nominal, ordinal, interval, and ratio scales. A nominal scale uses names or labels to distinguish stimuli. An ordinal scale uses labels or numbers to sort stimuli either in

ascending or descending order based on particular property. An ordinal scale has the property of greater than ($>$) or less than ($<$). However, it contains no information on the distances between stimuli along the scale. In this way, an ordinal scale can be used in rank order method (which will be described shortly). In the case of interval scale which is the focus of this work, a scale adds the equality of intervals and differences to the ordinal scale (equal distances anywhere along the scale have the same significance). In this way, the scale contains information about differences between stimuli for answering the “how close” question which is the key question of the image quality measurement. However, in general, the origin and multiplier of the scale are unknown and arbitrary. Finally, a ratio scale is an interval scale with a zero origin.

There are several methods of psychophysical experiments, among all of them, three methods are widely used and briefly summarised here:

- **Rank order method** An observer has to rank a series of stimuli based on a given criterion. The method is time consuming since all stimuli are compared at once. We note that Ranking can be indirectly conducted using paired comparisons.
- **Rate method** Each observer has to rate an attribute of a given stimulus on a scale. Here a reference can be included. The method is very fast compared to the method of Ranking. However, observers can have different perceptions of the rating scale making this method the least accurate.
- **Paired comparison method** An observer determines which stimulus in a given pair is more preferred (one shown on the left side and one shown on the right side of the screen) or closer to a reference (normally is shown in the middle of the screen). This method gives the the most reliable result among the three methods. However, the method is the most time consuming among the three methods.

In the next section, the paired comparison method is described in detail.

Paired Comparison Method

The method of paired comparisons has been developed by Gustav Fechner, who described it in 1860 (David, 1988). The method is also known as Two-Alternative Forced Choice (2AFC) since the method forces the observer to choose between two alternatives. The method of paired comparisons is perhaps the most straightforward way of presenting stimuli for comparative judgement. In general, paired comparison refers to any process of comparing stimuli in pairs to judge which of each pair is preferred.

In a paired comparison experiment, the samples are presented to an observer in pairs (typically shown on the left and right sides of the screen). The observer's task is to choose one out of the two samples presented that has the characteristics that the study administrator has focused on. This pairwise presentation is repeated for all possible $n(n - 1)t/2$ pairs, where n is the number of test stimuli and t is number of test scenes in the experiment.

When we ask the observers to respond to stimuli, we are asking for a judgment. The judgment implies that the observer is acting “objectively”, whereas the choice implies some form of personal preference in the response. Thus, the stimuli are rated according to the reaction they produce on human observers.

Originally, the paired comparisons method gives ordinal scales. However, as demonstrated by Thurstone (1927), the ordinal scale resulted from paired comparisons can be transformed to an interval scale using the law of comparative judgement. The law assumes that the discriminational process (the process by which human makes judgments of stimuli) is random and has the probability density function follows a normal or gaussian distribution on the psychological continuum (“ness” scale) (Engeldrum, 2000). And since the scale is expressed in terms of the probability density function, the mean of the function is the scale value. Generally, the law states that the proportion of times that stimulus A was judged greater than B ($A > B$) was an indirect measure of the distance on the “ness” scale between A and B. And the difference in scale values between stim-

uli A and B ($S_A - S_B$) is equal to the distance between the means of their response distributions. and is defined as:

$$S_A - S_B = Z_{(A-B)} \sqrt{\sigma_A^2 + \sigma_B^2 - 2\rho\sigma_A^2\sigma_B^2} \quad (2.22)$$

where S_A and S_B are the mean response of A and B, σ_A and σ_B are the standard deviation of A and B, ρ is the correlation coefficient between A and B, and $Z_{(A-B)}$ is the standard normal deviated corresponding to the proportion of time A is chosen over B.

However, the equation which represents the complete version of Thurstone's law of comparative judgment is unsolvable. To solve this problem, some further assumptions about the sample standard deviations and the correlation coefficient are needed. The most extensively used assumption is the Case V assumption that enables the practical application of the law since the interval scales can directly derived from the proportion matrix resulted from the paired comparison experiment. The Case V assumption assumes standard deviations for the two distributions are equal ($\sigma_A^2 = \sigma_B^2$) and the correlation coefficient is zero ($\rho = 0$). Thus, the Equation 2.22 is simplified as:

$$S_A - S_B = Z_{(A-B)} \sigma \sqrt{2}, \quad (2.23)$$

And since the empirical proportion is the estimate of the probability of how one stimulus is judged over the other stimulus. This implies that

$$P(A > B) = H(S_A - S_B), \quad (2.24)$$

$$S_A - S_B = H^{-1}[P(A > B)] \quad (2.25)$$

where $P(A > B)$ is the probability that stimulus A is judged over B, $H()$ is the Cu-

mulative Distribution Function (CDF) that transforms the scale differences into probabilities. For Thurstone's case V, the CDF is defined as a Gaussian distribution function as:

$$H(S_A - S_B) = \frac{1}{\sqrt{2\pi}} \int_{-(S_A - S_B)}^{\infty} e^{-\left(\frac{x^2}{2}\right)} dx \quad (2.26)$$

By setting $\sigma\sqrt{2}$ to unity (representing the unit of one on an interval scale), the proportion of times one stimulus is chosen over another is equal to the standard deviation alone. Thus, for example, assuming that there is a set of n stimuli, say three stimuli (A, B, and C), the Thurstone's model equation in a matrix form is now expressed as:

$$\begin{bmatrix} S_A - S_A & S_A - S_B & S_A - S_C \\ S_B - S_A & S_B - S_B & S_B - S_C \\ S_C - S_A & S_C - S_B & S_C - S_C \end{bmatrix} = \begin{bmatrix} Z_{A-A} & Z_{B-A} & Z_{C-A} \\ Z_{A-B} & Z_{B-B} & Z_{C-B} \\ Z_{A-C} & Z_{B-C} & Z_{C-C} \end{bmatrix} \quad (2.27)$$

In fact, the scale values \bar{S} has an average of zero, meaning that the column average itself gives the scale value directly. Therefore, the scale value of each stimuli can be determined by averaging each column of the z-score matrix. By doing this, the corresponding scale for each stimulus is derived.

Finally the scales are accompanied by confidence intervals (CI).² Typically, a 95% confidence interval has been reported and is calculated as:

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}, \quad (2.28)$$

where \bar{X} is the sample mean or in our case the scale itself, and σ is the standard deviation of the population (in our case, population is the number of observers).

²In statistics, a CI is an interval between two numbers, where there is a certain specified level of confidence in which a population parameter lies.

From the equation, it is clear that the width of the interval is controlled by the sample standard deviation i.e. the larger the sample standard deviation, the wider the confidence interval. This simply means that turbulent data generates wider intervals than data with a smaller standard deviation. In addition, since the number of the population affects the standard deviation, as population increases, the interval gets narrower i.e. one way to obtain more accurate estimation is to increase the sample size.

To analyse the data resulted from a paired comparison experiment. The proportional matrix is calculated from the number of left and right answers for each stimulus pair. From this matrix, a psychometric difference scale (the subjective image quality) is constructed for each stimulus pair.

A scale difference D between two scale values can be calculated, which can be interpreted as the probability P that one stimulus is preferred above the other. The relationship between the scale difference D and the probability P is shown in Figure 2.15. For example, when the difference D between two stimuli equals 1 or 2, the probability P that one stimulus is preferred above the other is 76 or 92%, respectively.

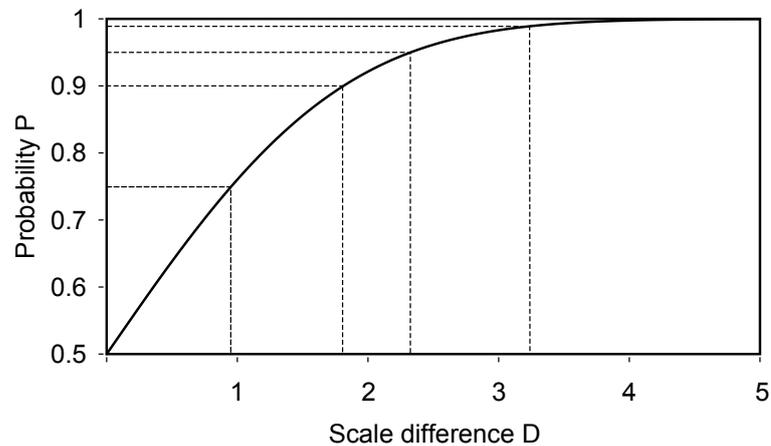


Figure 2.15: Relationship between the scale difference D and the probability P of preferring the one stimulus above the other stimulus. The dashed lines represent the 75, 90, 95 and 99% confidence intervals (Denecker et al., 2002).

Denecker et al. (2002) suggests that in order to be sure that for two stimuli, one

is preferred over the other, the scale difference between them should be larger than 1. However, in practice, it is generally sufficient to say that one stimulus is preferred over another if their confidence intervals do not overlap. For example, if the 95% confidence interval has been reported and the scale difference of the pair of stimuli is larger than this interval, we can conclude that at the 95% of confidence we reject the hypothesis that the pair of stimuli are similar and the stimulus that has higher preference scale is more preferred over the other. Besides, it is important to understand that the difference in scale values between two stimuli is only the probability that one stimulus is preferred over the other. It is not possible to conclude that how much better one stimuli is compared to the other or how difference between the two stimuli is. Since the observer's task of the experiment is not to answer how much different the perceived quality of the two stimuli is.

2.9 Staircase Psychophysical Experiment

A staircase method often referred as the “method of up and down” is one of the widely used methods in modern psychophysics. The method generally uses the previous one or more responses within an adaptive track to select the next trial placement toward the desired stimulus value. The simplest application of this method is shown in Figure 2.16. There, imagine we want to find the gamma value (refers to Section 2.6.2) that gives the best output image. As illustrated in the figure, the staircase starts with a descending set of stimuli with the initial value of gamma of 1.6. The observer judges “too dark” (D), the experimenter therefore decreases the stimulus value (gamma) one step (0.1 gamma unit) down which in this case results in a brighter image. This staircase continues with decreasing the stimulus value as the observer keeps responding (D). At the tenth trial, the stimulus is judged “too bright” (B) (a reversal is occurred), this time the experimenter reverses one step up, and so on. This stimulus series is exactly illustrated in

trials 1-10 in Figure 2.16, which shows a total of 20 trials.

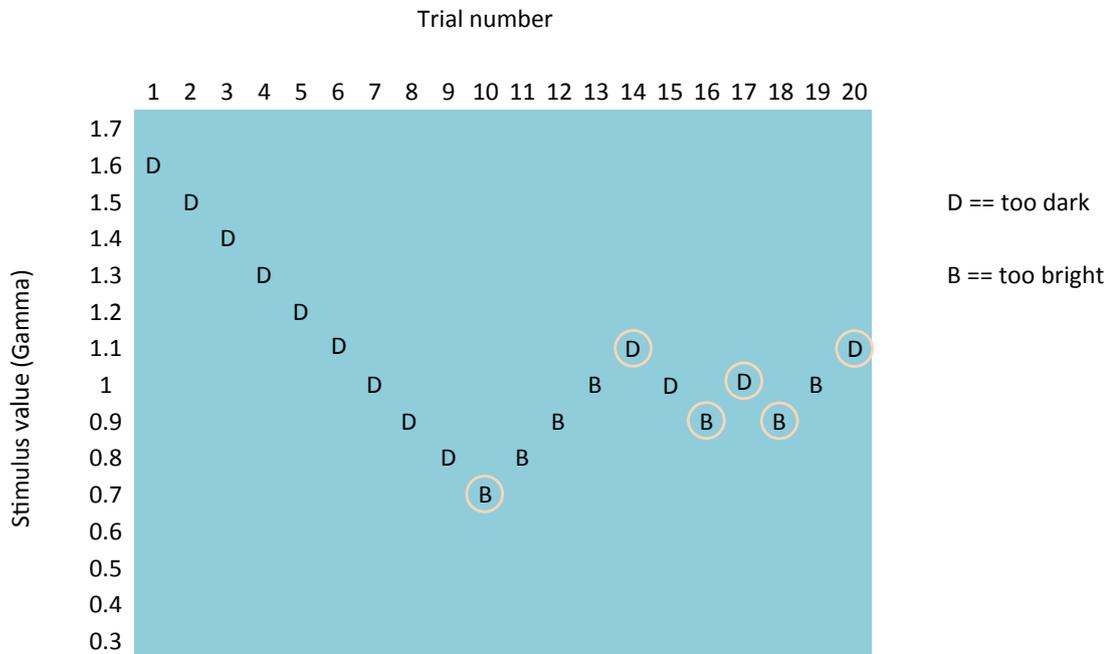


Figure 2.16: An example of single staircase experiment.

In fact, before one can utilise this method, there are three questions that needed to be answered: what is the best initial stimulus value, which value is the most suitable step-size, and how we terminate the experiment. Consider the first question, as one might guess, if we use a single staircase, the observer can easily guess that response on trial $n - 1$ determines a stimulus value of the trial n which could lead him or her to foresee the approach of the desired stimulus value and change his or her response before the value is actually reached. To eliminate the response bias that is inherently present in the single staircase, the experimenter may run two staircases simultaneously, randomly switching from one to the other, with one starting with a large stimulus value and the other with a small stimulus value. This application of the method is often known as “double (interleaved) staircase” (Cornsweet, 1962; Wetherill and Levitt, 1965). In this way, not only the initial stimulus value problem is solved, but also the ability of

the observer to keep track of the direction (either increase or decrease) alone which stimulus intensity will vary, is removed. Of course, within each staircase this tracking ability remains, but by randomly interleaving the two staircases can eliminate it from the observer.

Regarding the step-size, if it too large, then the observer's response will oscillate between the two stimuli resulting in no real preferred value. In contrast, if the step-size is too small, then the experiment becomes inefficient because the observer might worry about his or her choice of the same judgment for long consecutive trials. One way to find out the appropriate step-size is to run pilot experiments.

One of the benefits of double staircase application is when the two staircases begin at far apart starting positions, the step-size can be large at the very beginning then, after the two staircases converge, the step-size is halved (unknown to the observer). This variable step-size strategy speeds up the process, makes the experiment consume less time, and also minimise the problems related to the observer's tiredness.

The experiment is terminated when a fixed number of reversals is observed. At this stage we have a 'trajectory' of adjustments. Effectively, the observer consecutively makes the stimuli larger and smaller around the preferred point. To extract the final adjustment made by the observer it is necessary to average over the settings made for the last part of the staircase procedure. Typically, six to nine reversals are taken to estimate the preferred value.

Figure 2.17 replaces the single staircase shown in Figure 2.16 with a double staircase example. The staircases start at gamma of 1.6 and 0.4, converge with a 0.2 step-size, after the two staircases converge (trial 9), the step-size is halved to 0.1. In the figure, trials 1, 4, 5, 7, 8, ... belong to staircase A, while trials 2, 3, 6, 9, 10, ... belong to staircase B. To calculate the preferred gamma, the total of eight reversals (peaks and valleys) are collected and averaged. Thus, for staircase A we have $(1 + 0.9 + 1 + 0.9)/4 = 0.95$, for staircase B we have $(0.9 + 1 + 0.8 + 1)/4 = 0.925$.

These make the final gamma of $(0.95 + 0.925)/2 = 0.9375$.

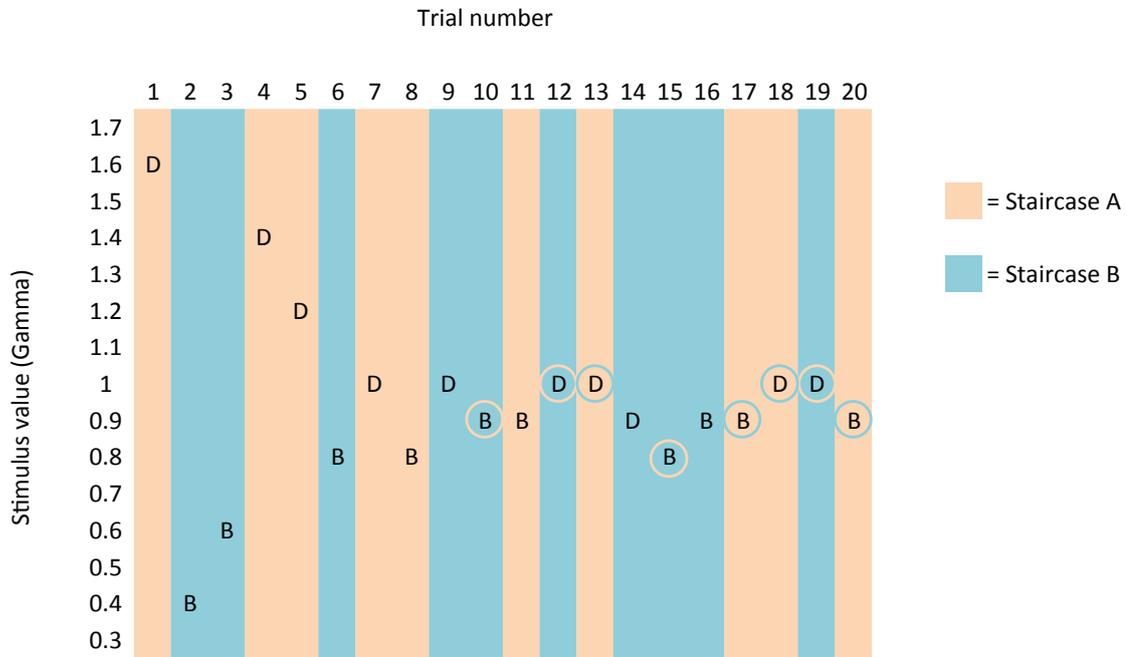


Figure 2.17: An example of double staircase experiment.

2.10 Summary

In this chapter we reviewed the HDR imaging pipeline. We started by reviewing methods of generating HDR images. We then reviewed several HDR image encoding. We then moved on to the techniques used in visualizing HDR contents, placing a greater emphasis on tone-mapping operators and surveyed several of them both global and spatially-varying. We also reviewed three related post-processing processes that typically applied to the tone-mapped images. Our discussion moved on to image statistics. Two image statistics related to the work presented here - dynamic range and image entropy - have been described. Next, we discussed the image quality measurement both objective and subjective approaches. Finally, the staircase psychophysical experimental

procedure was described. The following chapters present our research into how HDR images can be simply integrated within a standard photographic pipeline. We assume that the reader has familiarity with the contents presented in this chapter.

Chapter 3

Modified TIFF for Encoding HDR

Photographs

In this chapter, we seek to arrive at a new simple format: a format that suffices for encoding the practical dynamic ranges found in real scenes. And, is also easy to encode and decode (in fact we wish our encoding to be no more complex than a normal jpeg or tiff). We also seek an encoding which is as efficient as possible in terms of storage used. Specifically, in this paper we investigate the possibility of using a gamma-encoding scheme that stores images in an integer fashion that is found on many conventional LDR image formats for storing HDR photographs. Since this scheme is based on a conventional format, it should be fully compatible with any imaging applications.

This chapter is organized as follows. In Section 3.1, related backgrounds are provided. In Section 3.2, we described the proposed method in detail. In Section 3.3, we evaluate the proposed method using both computational image quality metrics and psychophysical experiments. We also discuss the effects of parameter variations that are used in the method. In addition, the size efficiency is also tested here. Section 3.4 concludes the chapter.

3.1 Background

Here, we would like to point out what is the gamma encoding found in conventional images. Gamma encoding is a nonlinear operation used to code intensities. Its simplest form is defined as in Equation 2.15 as:

$$L_{out} = L_{in}^{\gamma}$$

It should be noted that the gamma encoding is neither the gamma adjustment (the adjustment of image contrast), nor gamma correction (the correction of display characteristic). These three terms are often misunderstood by many people because they apply the same mathematical function. Many people assume that a gamma function is used in encoding images only because it compensates for the transfer function of the displaying devices. In regard to this chapter, this is actually wrong. In gamma encoding, we are neither adjusting the image contrast, nor correcting the display characteristic, rather we are attempting to minimize the visible quantization of the image with a limited number of encoded values (bits) using a non-linear relationship of the lightness sensitivity of our HVS, in other words, to maximize the use of the bits relative to human vision.

Applying gamma incorrectly can lead to lower quality images. For example, pixel values of images with linear encoding (gamma of 1.0, i.e. no gamma encoding) don't take into account human perception. In this case, many of the pixel values are wasted on highlights that we human cannot distinguish, while few pixel values are used for the shadows or midtones. Even worse, when this data is converted to a perceptually uniform encoding, typically we lose many important values due to quantization (The process of constraining data from a relatively large or continuous set of values to a relatively small discrete set, e.g. converting between float value into integer) leading to visible quantization artifacts.

With 8-bit/component images for example, converting from a gamma of 1.0 to a

gamma of 2.2 means that there is only 184 intensity values out of 256 values left (losing one quarter of the values). If we do not want to lose the levels we have when we want to convert the gamma of 1.0 to 2.2, we have to increase the bit-depth of image to 12-bit to avoid this degradation. Later in this work we will demonstrate that using a proper value of gamma at first can result in a good HDR representation.

Our method takes inspiration from existing RGB gamma-encoded HDR formats. We begin with scRGB (scRGB, 2003) (an open standard first developed by Microsoft and Hewlett-Packard). There are two variants of this encoding. The first variant is a 16-bit/component linear RGB encoding (scRGB). That is we can store 65536 brightness levels per channel at an overall cost of 48 bits per pixel. The second variant is a 12-bit/component RGB non-linear encoding (scRGB-nl) using a standard 2.2 gamma (with a linear subsection near zero). This format can be thought of as the extension to the existing sRGB color space, since it uses the same color primaries and white/black points as the sRGB color space (Stokes et al., 1996) but allows negative values and values above 1.0, and offers a larger gamut and larger dynamic range. The dynamic ranges of the two variants in theory cover 3.5 and 3.2 orders of magnitude for scRGB and scRGB-nl respectively.

In addition, it is well known that human vision is sensitive to relative changes. Thus, it is crucial for the format to have low relative changes as possible. In other word, the relative change has to be below the human detectable threshold. In the context of image encoding, relative step-size measures how different the relative change is between a pair of adjacent encoding values, and is generally held to double a size of the error caused from the quantization process. A relative quantization error is the difference between the correct value and the representation divided by the correct value. Given this, the relative step-size (μ) at the specified value x_i is defined as:

$$\mu(x_i) = \frac{\|x_i - x_{i-1}\|}{\|x_i\|} \quad (3.1)$$

Floating point encodings (IEEE RGB, RGBE) have an almost constant relative step-size that is below the human detectable threshold over the entire range. Unfortunately, relative step-sizes of gamma encoding continue to increase as the value gets smaller. If one encodes data below the value that has the relative step-size over this threshold, these data would exhibit visible banding if one were to view it on an HDR display. In fact, it is difficult to characterize encodings with variable quantization step in terms of their affordable dynamic range since the human detectable threshold is ambiguous. Reinhard et al. (2005) calculates the dynamic range of gamma-encoding based on the value whose relative step equal to 5%. For example, the minimum value whose relative step equal to 5% of an 8-bit sRGB is 0.025, which gives the dynamic range of 0.025:1 or 1.6 orders of magnitude. Recall the dynamic range of the two variants of scRGB, counting the issue of a 5% relative step-size making the dynamic range of these variants reduced and cover only 3.5 (7.5:0.0023) and 3.2 (6.2:0.0039) orders of magnitude for scRGB and scRGB-nl respectively (scRGB, 2003; Ward, Unknown). We can encode smaller values but when these are brought to a visible display range there will be quantization artifacts.

Since, HDR radiance data can span up to 6, 7 or even 8 orders of magnitude, none of gamma-encodings can cover the necessary dynamic range. This is why gamma-encoding is more like a Mid Dynamic Range (MDR) image format (cover about 3 to 5 orders of magnitude). Thus, directly encoding HDR images that has wider dynamic range using the gamma-encoding does not work and will end up with either losing detail outside the affordable dynamic range (both in highlight and shadow) or introducing visible banding artefacts in the regions that have value close to zero (dark areas) of the reproduction (assuming the corresponding error exceeds the human detectable threshold) when visualise on HDR display. In this Chapter, a modified framework to the standard gamma-based LDR format for storing HDR photographs is proposed.

3.2 The Modified Framework: Using Integer TIFF to Encode HDR

Our proposed framework for the HDR encoding is based on a standard TIFF image format. The benefit of TIFF is it provides metadata components (that can be extensible). There are several frameworks in TIFF standard, among all of them 12 and 16-bit/component RGB integer frameworks were chosen since they are widely supported and are used in the two variants of sRGB. However, unlike the sRGB, here instead of storing in float, we base the proposed encodings on integer since it is more widely supported by many applications.

With this in mind, the proposed frameworks consist of three R, G, and B color components, either 12 or 16-bits are required for each color component (36 and 48-bits/pixel respectively). For the encoding, the concept is the same for the 12 and 16-bit variants, but they achieve different accuracy. In both cases, each color component is first normalized by dividing by the maximum value of the three color components (I_{max}). The gamma (γ) is applied, the result is then scaled to fit the $[K, 1]$ range where K is the minimum value that depends on the number of encoding bits (2.4×10^{-4} and 1.5×10^{-5} for 12 and 16-bit, respectively), and is quantized into a specific integer range which is different for the two frameworks ($[1, 4098]$ for the 12-bit, and $[1, 65536]$ for the 16-bit). We avoid storing zero values because later when decoding, zero cannot be scaled back to the original floating point value. The whole process is computed as

$$Y = \left(\frac{I}{I_{max}} \right)^{\frac{1}{\gamma}} \quad (3.2)$$

$$K = 1/N \quad (3.3)$$

$$N = 2^{\text{no. of bits}} \quad (3.4)$$

$$I_m = \left\lfloor N \left(\frac{Y(1-K)}{1-Y_{min}} + K \right) \right\rfloor \quad (3.5)$$

I_m is not different from a standard integer TIFF. This means that naïve applications that are not supporting floating-point operation will discard the decoding operation and simply read I_m (treat the HDR as a conventional 12 or 16-bit TIFF). Although this is more likely to give us unnatural images, they still gain some benefit from the HDR capture, since they covers all the dynamic range of the scene and both permit larger exposure shifts and contrast manipulation during the image manipulation than we could with the limited range of conventional LDR images. In addition, in order to bring back the original data, the maximum (I_{max}) and minimum (I_{min}) values of the three color components are need to be recorded as additional tags called “scale factor” in the metadata.

For HDR enabled applications (where the application notices the private tags and can process in floating point fashion), the original dynamic range of the HDR can be fully retrieved by consulting these tags in the decoding process as

$$Y = \left(\frac{I_m}{N} \right)^{\gamma} \quad (3.6)$$

$$I = \frac{Y(I_{max} - I_{min})}{1 - Y_{min}} + Y_{min} \quad (3.7)$$

By performing these processes, we can eliminate an uncertainty of dynamic range of HDR, conquer the limited affordable dynamic range of the format, allowing compat-

ibility with various imaging applications.

The proposed solution is lossy, in that it does not encode all the original data, but the goal is to produce a representation that gives a visually faithful tone-mapped image. Thus we can take advantage of the filtering properties of human vision to mask the loss of data (as with JPEG and other lossy encoding schemes for LDR images). However, although the initial encoding is lossy, once the data have been encoded, subsequent reading and writing of the files does not incur any further degradation.

3.3 Experiment

To evaluate the image quality of the proposed encoding frameworks, we compare the tone-mapped images resulting from our encoding with the RGBE radiance encoding, using the original 96-bit TIFF float format as a reference. Two experiments, both objective and subjective approaches, have been conducted. In addition, the objective experiment also serves as a study of the effects of parameter variations: bit-depth and gamma. Space efficiency is tested at the end of the section.

3.3.1 Objective Image Quality Measurement

There are two goals for the objective part: one is to compare the encoding preference, one is to empirically examine the effects of gamma and bit-depth on the appearance of the tone-mapped image. Our work here is related to the HDR encoding comparison article by (Ward, Unknown). In that work, to evaluate the encoding preference, he measures the ΔE CIELAB 1994 color difference metric between each of the test HDR encodings to the 96-bit float TIFF. To visualize the results, quality curves of different encodings are generated using these color different metrics.

Since, here we are interested in how perceptually different the tone-mapped images resulting from the test encodings are compared to the ones resulting from the 96-bit

TIFF float (ideal representation). Thus, we rather investigate the perceptual color difference of the tone-mapped images. Since they are conventional 8-bit images, we can measure the color difference directly. The color difference metric that we choose is ΔE S-CIELAB (Zhang and Wandell, 1996), since it takes into account spatial characteristics of the human visual system for measuring perceptual color differences. In other words, it measures how similar the reproduction is to the original when viewed by an observer.

A total of 65 HDR images were used as reference images in the test. All of them were generated using between five and eleven 1-stop exposure bracketed photographs and saved as 96-bit TIFF float format. Figure 3.1 shows the dataset. To ensure that our results will be consistent across a range of TMOs, two widely available TMOs: iCAM06 (Kuang et al., 2007) and Retinex-based Adaptive filter (Meylan and Susstrunk, 2006), which differ significantly in their mode of operation are used to generate tone-mapped images of the three encodings: Our encoding, RGBE, and 96-bit TIFF (as a reference).

Testing Encoding Parameters

We begin the investigation of the parameter variations by first defining a range of test γ . To our knowledge, there are three values of γ that are practically used: the first one is the γ of 2.2 found in conventional sRGB encoding and resembling the gamma of scRGB-nl. The second one is the γ of 2.6 found in Digital Cinema Initiatives (DCI) standard (DCI, 2005, 2008). The third one is the γ of 3.0 found in lightness calculation of CIELAB, CIELUV perceptual uniform color spaces (Bodmann et al., 1979). This indicates that the optimal gamma that we are searching for should be somewhere around these values. With this in mind, the range of gamma tested here is between 2.0 to 3.2 with a step size of 0.2. Secondly, we test different bit-depths ranging from the bit-depth of 8 (as found in LDR image) to 16 with a step size of 2, since it is directly related to the image file size. The effects of bit-depth and gamma parameter variations on the appearance of

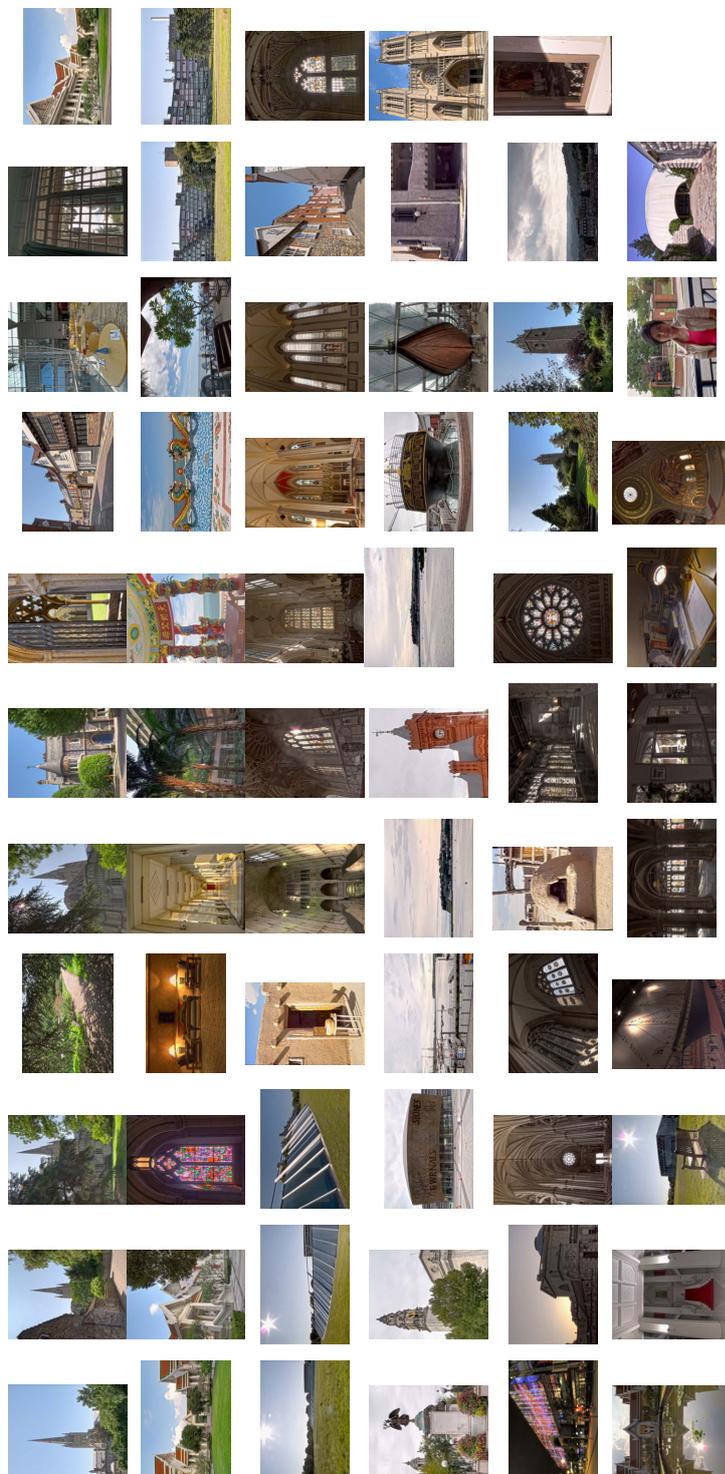


Figure 3.1: The image dataset; 65 images in total, both from our own images and the well-known images by Čadík et al. (2008); Debevec and Malik (1997); Fattal et al. (2002b). The last row is the ten images used in the subjective psychophysical experiment.

the tone-mapped image are illustrated in Figures 3.2 and 3.3, respectively. There, each curve represents the average percentage of pixels above a particular ΔE S-CIELAB obtained by weighing each of the test images equally. To interpret these plots, a steeper slope curve that reaches a small percentile at the smaller ΔE would be the more ideal encoding, since there are fewer pixels perceived differently to the reference. In addition, here there are two important values of ΔE that need to be considered, which are the value of 2, and 5, since these values correspond to the percentage levels indicating a noticeable color difference to observers in ideal viewing conditions, and noticeable in side-by-side images, respectively. Practically, we could say that one might notice a difference between a pair of images if there are more than 2% of the pixels that have a ΔE greater than 5.

The results show that for the bit-depth, as one might guess, the more bit-depth that we use, the more ideal the shape of the curve we will get. Since the percentages of 8 and 10-bit exceed the ΔE of 5, this indicates that these two value of bit-depths incapable to represent the HDR. For the gamma, the larger the gamma we use, the lower ΔE we will get. In addition these figures also include the curve of the RGBE and shows that for the bit-depth (using gamma of 2.2) that the 14-bit is better than the RGBE. For the gamma, the gamma of 2.6 of the 12-bit framework is better than the RGBE.

We note that, although a greater value of gamma results in the more ideal output works in principle, in practice we might end up losing more details in the bright areas (bright area often has small number of pixels) if we use too large γ , as the consecutive steps in these areas are likely to be quantized to have the same value (the consecutive step size of that region is smaller than the quantization step size). And since the gamma of 3.2 is insignificantly better than the gamma of 3.0, we decided to base the maximum gamma of the encoding at the value of 3.0. With this in mind, from this point forward we will constrain the choices of gamma parameter to 2.2 and 3.0 and constrain the choices of bit-depth to 12 and 16-bits.

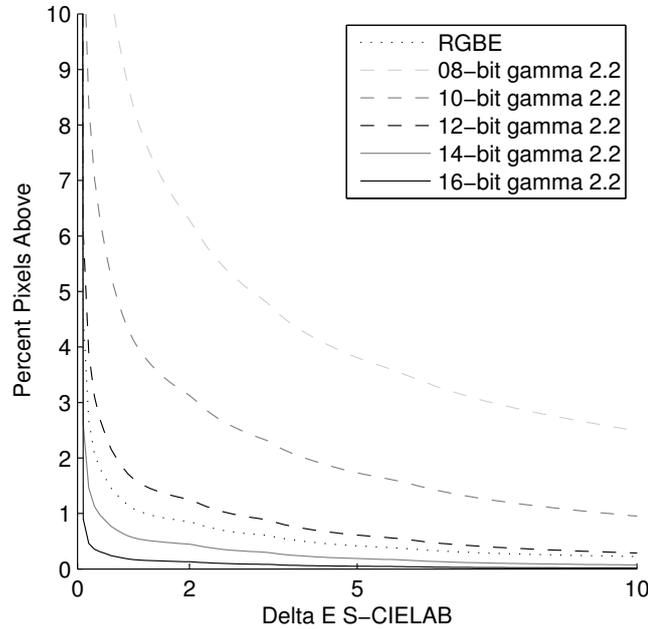


Figure 3.2: Effect of bit-depth on the appearance of the tone-mapped images.

Encoding Comparison

Here, we evaluate the preference of the four variants (12 and 16-bit, with gamma of 2.2 and 3.0) of the proposed encoding compared to the state-of-the-art RGBE radiance format. Figure 3.4 shows the result. There it is clear that, our 16-bit framework both a γ of 2.2 and 3.0 have been out performed by the RGBE. In fact, this is not surprising at all, since the format uses 48-bit/pixel. In comparison, our 12-bit framework, γ of 3 gives a slightly better result than RGBE, while γ of 2.2 gives the poorest result.

Table 3.1 summarizes the results. We report two percentile quantities taken from the plot in figure 3.4 for the percentage of pixels above ΔE values of 2, and 5. Since there are only 0.61% of pixels of the worst encoding (12-bit, $\gamma = 2.2$) that have ΔE more than 5, this implies that on average both RGBE and all of the four variants of our encoding frameworks generate perceptually identical tone-mapped images to the

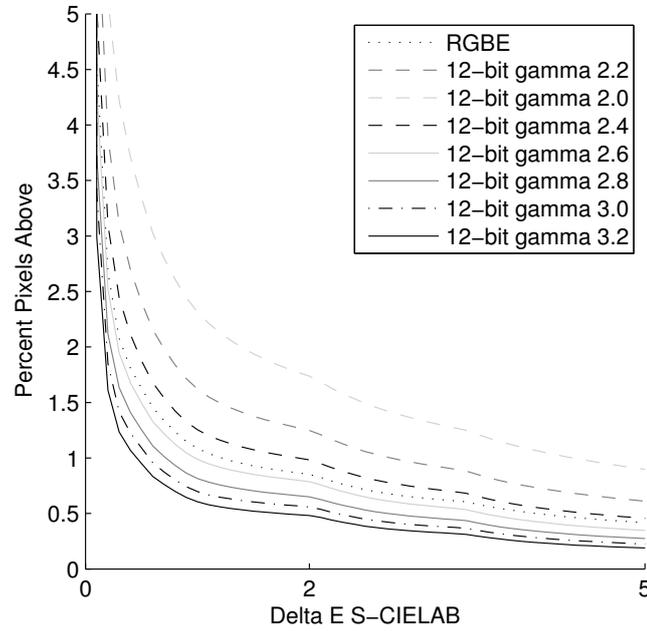


Figure 3.3: Effect of gamma on the appearance of the tone-mapped images.

	RGBE	12-bit (γ of 2.2)	12-bit (γ of 3.0)	16-bit (γ of 2.2)	16-bit (γ of 3.0)
$\Delta E > 2$	0.85%	1.25%	0.56%	0.13%	0.04%
$\Delta E > 5$	0.42%	0.61%	0.23%	0.05%	0.01%

Table 3.1: The summary of percent of pixels above ΔE S-CIELAB of 2 and 5 for each encoding.

original floating point format.

Problems of the 12-bits framework

Although the average percentages at the ΔE value of 2 and 5 of all test gamma (of the 12-bit framework) are lower than 2, in fact there is one image (shown in Figure 3.5) that the tone-mapped output of the 12-bit framework (with the $\gamma \leq 2.2$) contains a visible banding artifact in the dark flat achromatic region (black paint of the ship) which indic-

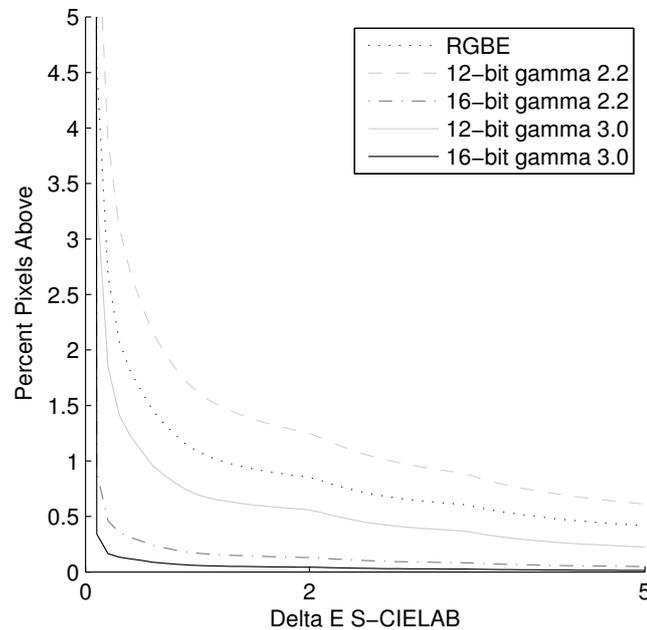


Figure 3.4: Average percentage of pixels at a particular ΔE S-CIELAB of different encodings for the entire test images.

ates a failure in the reproduction. If we want to encode this image without introducing any banding artifacts, yet with the γ of 2.2, a minimum of 14 bits is required (see top middle left image). And since 14 bits is not common for conventional applications, using 14 bits could lead to the compatibility issue. In this case 16 bits is the most practical alternative since it is supported by most of the applications, although it comes with a sacrifice of storage space. However, since we can vary the γ , the bit-depth becomes less important here. To better visualize this artifact, false-color images are also given in the bottom row of the figure, which represent the ΔE CIELAB color differences between the corresponding image (the above image) and the 96-bit TIFF reference (top left). A color bar (bottom left) indicates a magnitude of the ΔE of these images. In addition, for easy comparison, Figure 3.6 plots results of different selected gamma and bit-depth parameters of this image.

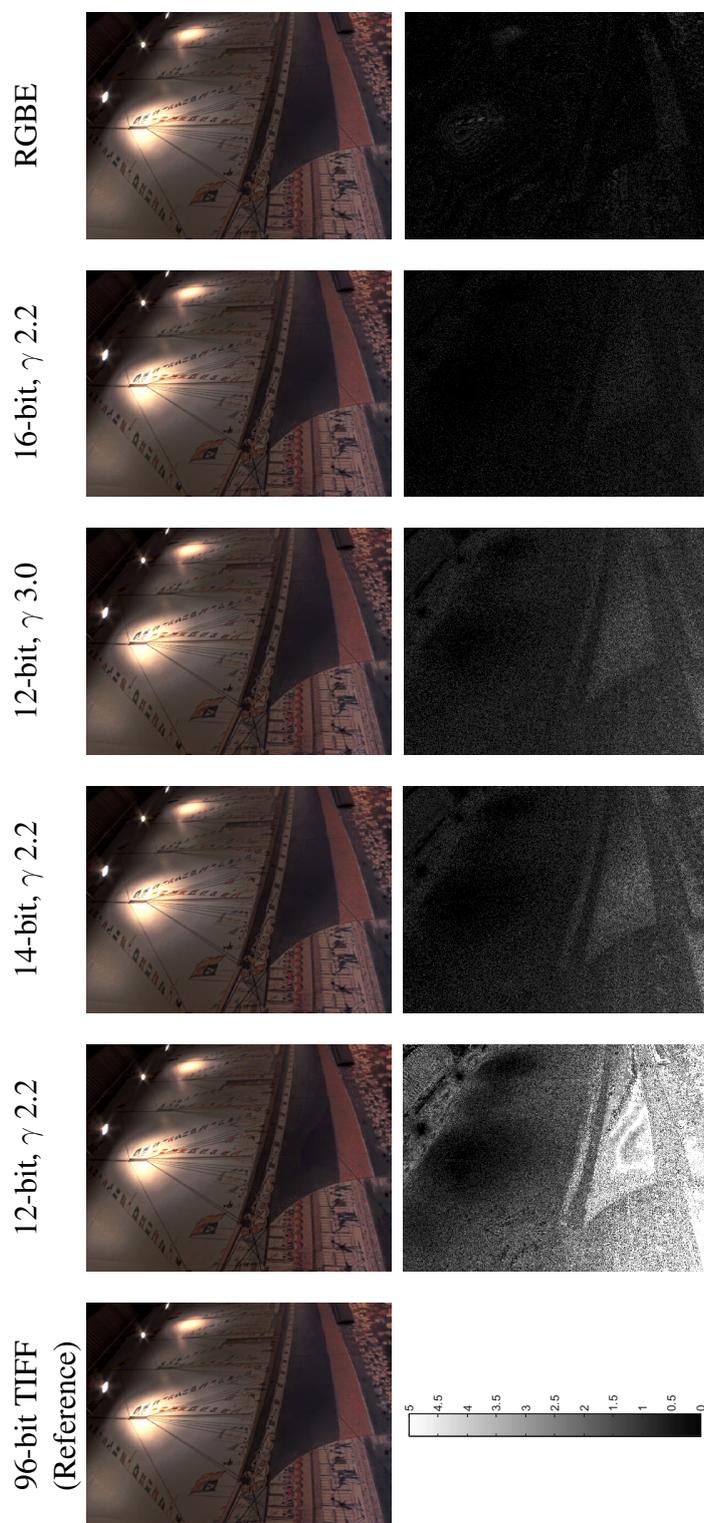


Figure 3.5: Effect of bit-depth and γ on the iCAM06 tone-mapped image. Top row from left to right: the iCAM06 tone-mapped version of the original (32-bit float point) SS Great Britain image, the same image passed through our 12-bit γ of 2.2, our 14-bit γ of 2.2, our 12-bit γ of 3.0, 16-bit γ of 2.2, and RGBE. Bottom row from left to right: A color bar indicates a magnitude of the ΔE , false color image represent ΔE between the corresponding (above) image and the reference 96-bit TIFF (Top left) image.

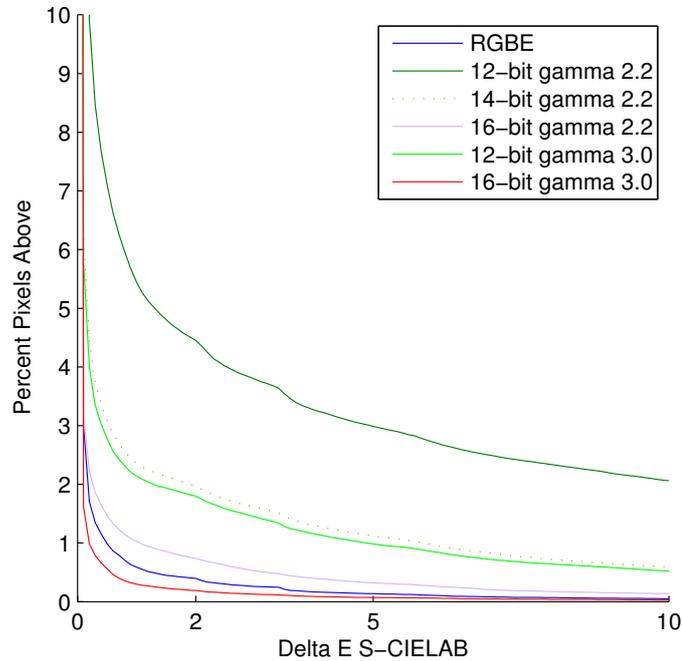


Figure 3.6: Percentage of pixels at a particular ΔE S-CIELAB of different encodings of the scene shown in Figure 3.5.

As indicated in both Figure 3.5 and 3.6, for this image, it is clear that the γ of 3.0 dramatically reduces the ΔE of the tone-mapped image resulting from the 12-bit framework compared to the gamma of 2.2. We can say that the overall color appearance of the 12-bit γ of 3.0 is visually equal to the 14-bit γ of 2.2 since the different of the two plots quite similar and is less than the ΔE of 1. Indeed the 12-bit γ of 3.0 yields lower ΔE compared to the 14-bit γ of 2.2 (the error image is darker). This indicates that gamma has strong effect on the appearance of the tone-mapped images.

To this point it is clear that the γ of 3.0 seems to be the winner in terms of minimizing color different while reserving some storage for the highlight. As one might guess from the context of lightness perception of the HVS, our HVS perceives lightness approximately equal to the γ of 3.0. This can be demonstrated by plotting the perceived brightness against the quantization steps as illustrated in Figure 3.7. There the most de-

irable curve is supposed to be linear. Although γ of 2.2 gives a more even distribution of quantization steps compared to a linear quantization, the behavior in dark regions is still not ideal (we still see larger steps in darker regions than we do in the brighter regions). As one might guess, this can lead to visible banding artifacts in the dark regions when tone mapped. When using a γ of 3.0, we see a more even distribution of quantization steps through the whole steps.

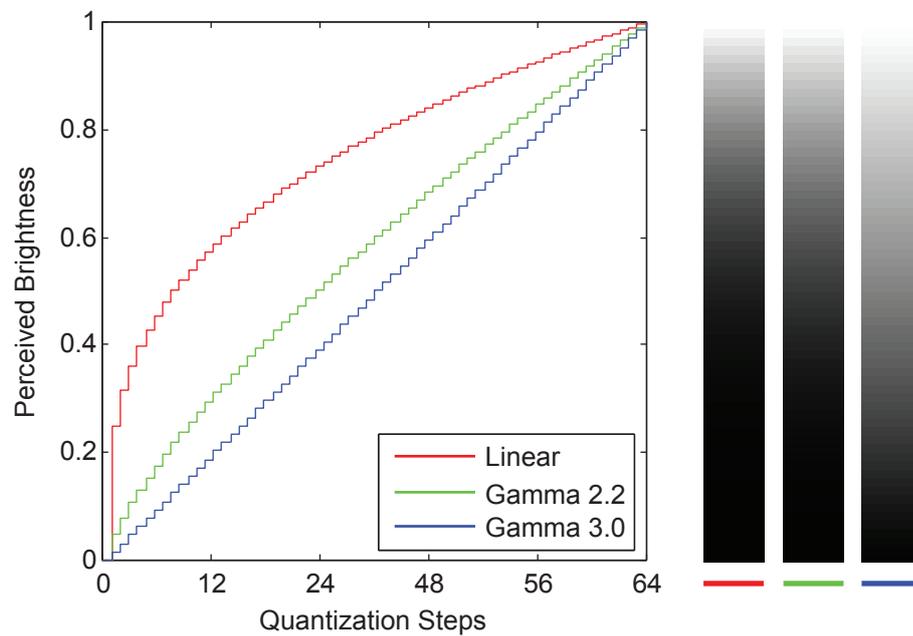


Figure 3.7: The perceived brightness of the quantization steps of different gamma (γ of 1, 2.2, and 3.0). The number of steps in this quantization is 64 (2^6) to emphasize the visible steps. Grayscale ramps on the right are corresponding representations of different value of gamma.

We note that, although a proper gamma results in a perceptually similar TMO image, it is still true that a larger bit-depth is more accurate than the lower one (the quantization step is narrower in a larger bit-depth). We realize that apart from the tone-mapping application, there are many of possible applications that one can apply to the HDR, for example in the case of exposure shifting applications, like when we want to convert a camera RAW format to an 8-bit JPEG format. Such applications manipulate an in-

tensity distribution of the image, especially when we want to boost up the dark details, this might amplify the relative error in the dark regions, making it exceed the visible threshold that results in banding artifacts. Using larger gamma will give us a more room to perform this task (especially on dark area) without exposing the artifact (with the scarifying of accuracy in the bright area).

3.3.2 Subjective Image Quality Measurement

For the subjective image quality measurement part, a psychophysical experiment has been conducted to test the preference of our encoding frameworks in the context of a tone-mapping application. The goal is to generate preference scales in order to evaluate the perceived image quality of the proposed encodings compared with HDR radiance format.

Experiment Overview

In order to evaluate the perceptual image quality of the encoding, a psychophysical experiment based on a Two-Alternative Forced Choice (2AFC) procedure for image pairs (paired comparison) (Engeldrum, 2000) was conducted. Six tone-mapped images were generated by applying two tone-mapping operators used in the previous section (iCAM06 (Kuang et al., 2007) and Retinex-based Adaptive filter (Meylan and Susstrunk, 2006)) to three different HDR encodings, our 12-bit and 16-bit gamma of 2.2, and HDR RGBE encodings. These images were evaluated by twelve observers (six male and six female) with normal color vision, naïve for the goal of the experiment, under the same experimental conditions. We chose only to test a 2.2 gamma encoding. This gamma is the most commonly used and for our suite of images produced visually similar tone mapped images to 3.0.

Experimental Design

In the experiment, the participants were asked to choose whether an image shown on the left or right was more similar to the reference (rendered 96-bit float format) image in the middle. A color calibrated monitor (HP DreamColor LP2480zx) displayed the three comparison images, each of which had a resolution 640 x 480, at 60.0 Hz. Ten real-world images with a diversity of dynamic ranges and spatial configurations (shown in the bottom row in Figure 3.1) were incorporated in the experiment. With this in mind, there were a total of 120 comparisons (3 encodings, 2 tone-mapping operators, and 10 images with 2 repetitions of each image) in the experiment. The whole procedure for one participant took approximately 15 to 20 minutes.

Experimental Results

In order to rank the encodings, the percentages of observers choice are transformed into z-scores. In this way, the z-scores now act as preference scores. The scores are computed using Thurstones Law of Comparative Judgment Case V (Thurstone, 1927).

Figure 3.8 shows average preference scores for 10 scenes. Figure 3.9 shows the scores obtained for individual test images. Figure 3.10 summarizes the results for each encoding.

We added error bars to the perceptual scale data in order to assess if two samples were significantly different. The 95% confidence interval can then be expressed as $CI = zscore \pm 1.96SD$ where SD is the standard deviation of the interval scale, for 12 observers CI is 0.36.

The higher the preference score the encoding has, the more often it was chosen by the observer. The results show that our 12-bit framework gives the lowest scores while our 16-bit framework has slightly higher scores than the radiance RGBE encoding in total. The results also show that both iCAM and Meylan algorithms give almost the same score; this indicates the consistency of our encoding across tone-mapping

algorithms. Only the 12-bit framework of SS Great Britain image contains artifacts, and has the lowest scores among all images.

From the results, we conclude that there is a difference between the 12 bit γ of 2.2 encoding to the reference. However, the γ of 3.0 completely eliminates this artifact and perceptually appears to be the same in side-by-side comparisons.

In all of the scenes we tested, the tone mapped images appear identical as the original to the observers as does the RGBE format, even in the darkest region of the scenes. Only in one scene shown in Figure 3.5, which comprises of very flat dark achromatic region, in that only the 12-bit γ of 2.2 encoding introduces visibly artifact as mentioned above.

As a conclusion, on average, the observers cannot discriminate the different between the reference (tone-mapped of TIFF float) and the three encodings.

To this point, we note that the results obtained both by objective and subjective experiments exhibit a strong correlation between themselves, for example, the 12-bit γ 2.2 tone-mapped representation of SS Great Britain has the poorest scores in both of the experiments.

3.3.3 Storage Size

In addition to the appearance of the tone-mapped images, we are also interested in the statistical behavior of the proposed encodings with regard to file size. Here, we compare our encodings with RGBE encoding with/without ZIP compression on our image dataset. The results are shown in Figure 3.11. Compared with the original 32-bit float point format, RGBE format takes 28.75% of the original file size. Our 12 and 16-bit encodings take 37.5% and 49.92% respectively. For ZIP compression, our encodings get the benefit of integer values when dealing with compression. As can be seen, both of our encodings dramatically reduce the file size when ZIP compressed, whereas the RGBE format is reduced in size by relatively small amount (RGBE is floating point).

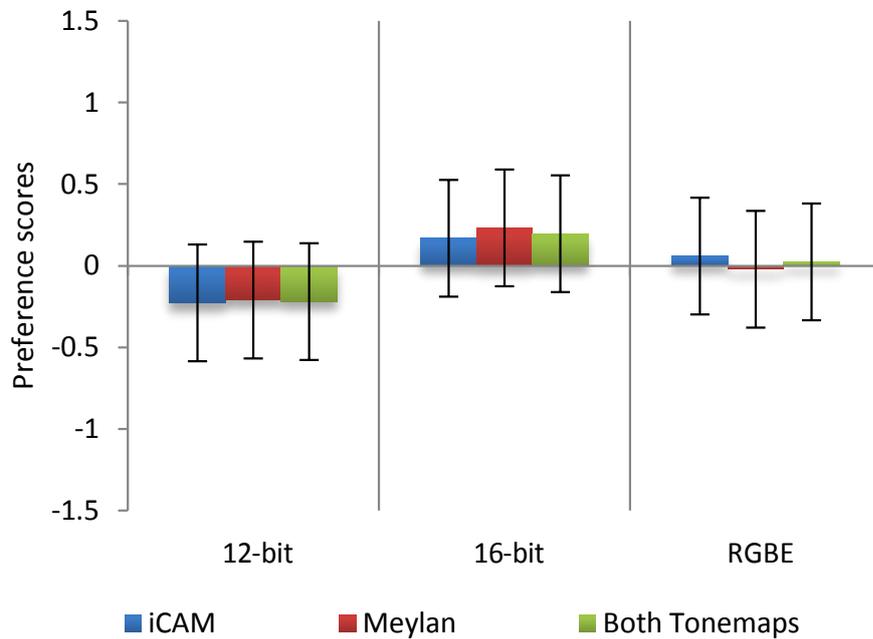


Figure 3.8: Overall preference scores for encodings over 10 images (The encodings are labeled as our 12-bit Gamma TIFF encoding (12-bit), our 16-bit Gamma TIFF framework (16-bit) and HDR Radiance format).

From the results, it is clear that our 12-bit encoding requires less storage space than RGBE format (22.77% compare to 24.11%), while our 16-bit encoding still requires the most storage space (30.32%). This indicates that our 12-bit framework in practice, consumes less storage space than RGBE on average, irrespective of the actual bit-depth required by the formats.

3.4 Conclusions

The proposed encoding can bridge the gap between conventional and HDR imaging pipelines, it is simple, fast, practical, and compatible with any imaging applications that support conventional 12 or 16-bit integer TIFF.

Thorough evaluations, both perceptually based color difference calculations and im-

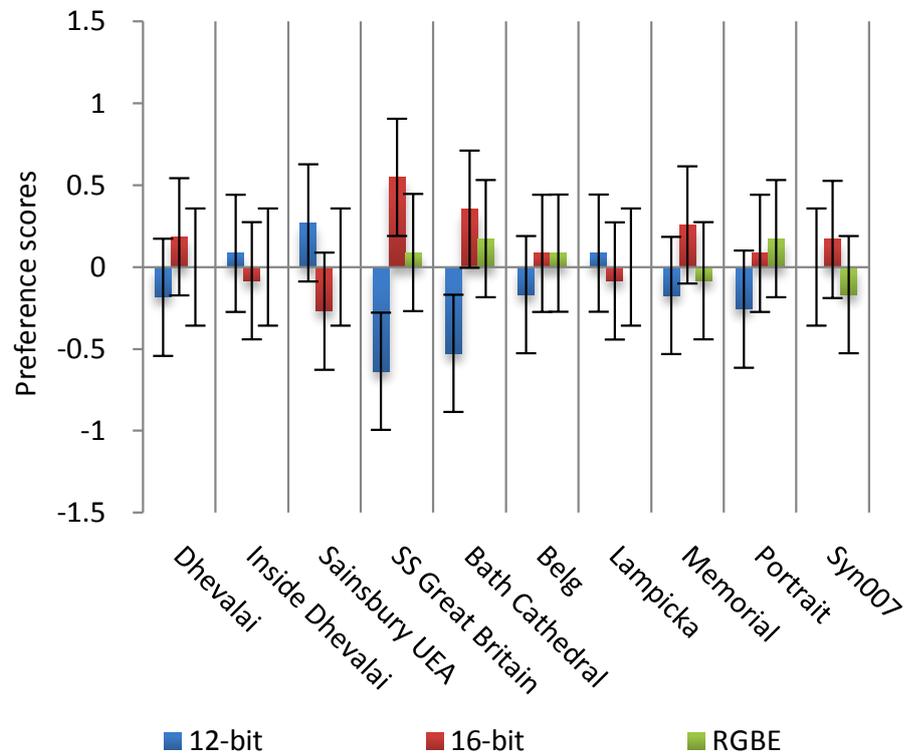


Figure 3.9: Preference scores for 10 test images by image.

age preference through psychophysical experiment for a set of HDR encodings have been conducted in this study. The results show that our encoding frameworks, both 12 and 16-bit, perform well in the context of a tone-mapping application. Compared to the RGBE format the 16-bit representation is slightly better in terms of error reduction, although the 12-bit representation gives poorer results as evidenced by the one banding artifact in the dark achromatic region of one of the test tone-mapped images. However, this artifact can be eliminated by increasing the γ parameter, and found a value of 3.0 to work best.

It is true that, for the extreme exposure shifting applications (as one might encounter when trying to make a night scene image as bright as it has been taken at day time) of HDR images (of the kind not offered in tone mappers) then the RGBE image format is

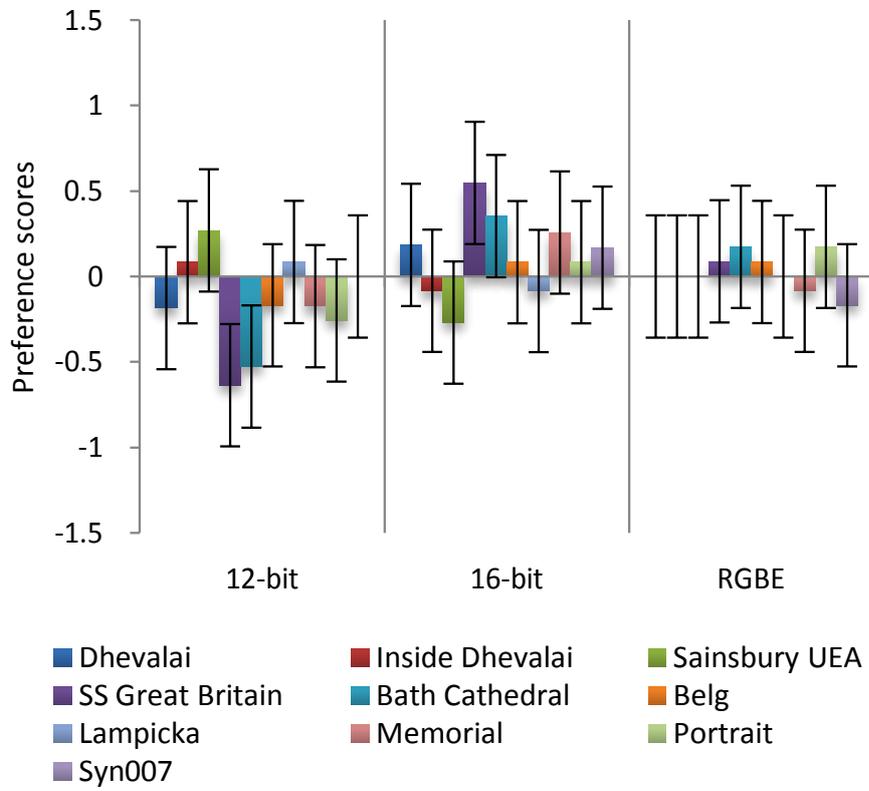


Figure 3.10: Preference scores for 10 test images by encoding.

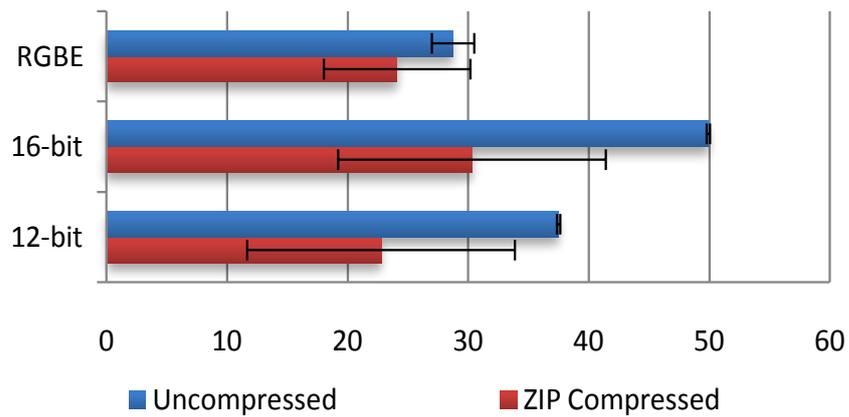


Figure 3.11: The average file size (size efficiency) for each test encodings. Error bars show extrema in the test.

likely to still be a preferred encoding standard. But, all experiments indicate that for the practical HDR workflow used by the majority of HDR enthusiasts our simple TIFF encoding would suffice (offering advantages both in simplicity, storage and computational processing).

Chapter 4

Automatic Gamma Adjustment

In this chapter, a study on how to achieve optimum results without the need of parameter adjustment for gamma adjustment operator is investigated. A psychophysical double staircase experiment is conducted. The goal of this experiment is to investigate the relationship between the theoretic derived gamma that maximizes the entropy of image and the preferred gamma obtained from the psychophysical experiment. As a result, an automatic gamma adjustment tone reproduction operator applied for conventional images is proposed.

This chapter is organized as follows. Related backgrounds is presented in section 4.1. In section 4.2, the double staircase psychophysical experiment is discussed. In section 4.3, experimental results are given. We also discuss relevant gamma predictors in Section 4.4. In section 4.5, an automatic gamma adjustment is proposed. Finally, in section 4.6, a conclusion is given.

4.1 Background

Gamma adjustment in the context of tone reproduction operator provides contrast adjustment. The simplest form of the operator (refer to Equation 2.15) is given as:

$$L_{out} = L_{in}^{\gamma}$$

where L_{in} and L_{out} are the input and output intensities, respectively, and are typically in the range $[0, 1]$.

Typically, gamma might be adjusted by a user in a package such as Photoshop or, the concern of this work, automatically using some sort of formula: relating an image statistic to (hopefully) choose the appropriate gamma. So, how might we find such a gamma adjustment formula? In the standard approach we investigate gamma adjustment in a purely empirical manner. That is, run an experiment where observers adjust gamma and then fit a formula to predict observer results. Indeed, this is a perfectly valid strategy and one that was extensively employed in the development of CIECAM.

In contradistinction to this approach we will also adopt a purely theoretical formula (derived only from mathematical argument) for choosing gamma. Then, we relate our predictions to observer choices to validate our approach. To derive our theoretical gamma, we ask the following question. Assuming that we wish to choose a gamma to best bring out image detail which gamma should we choose. We answered this question using the concept of image entropy from information theory (refer to Section 2.7.2).

Shannon Entropy (Shannon, 1951) encapsulates the idea of how expensive it is to code data. As an example a language where every character occurs with equal frequency is more expensive to encode than one where a few characters occur more frequently. In English for example we know that the letter 'e' occurs often but that 'z' is rare. We can exploit this information in coding letters with binary strings. We might code 'e' with a small number of bits and 'z' with a larger number. In the context of an image we can think of each brightness level as a distinct character. When we raise an image to the power of gamma we are mapping one set of characters with an inherent probability distribution to another where the original and gammed image will have different distributions of brightnesses. It follows then that the original image and gammed counterpart

will have different entropies. Finlayson and Xu (2012) proposed the formula that resulting in the gamma that maximize image entropy (refer to Section 2.7.2) and called it the “optimal gamma”. However, as can be guessed, having the maximum entropy state does not mean having the most appealing output. In the context of this work we are interested in whether the gamma adjustments that maximize image entropy match those that chosen by observers.

So, if we adjust an image using the information theoretic optimal choice of gamma, does it correspond with the adjustments made by observers? To test our assumption, we first need to design a psychophysical experiment to acquire observer gamma choices and then, in a second stage, relate these adjustments to the optimal gamma. In the next section, the experiment conducted in this study is described in detail.

4.2 Experiments

Here we use a double (interleaved) staircase psychophysical experiment (Cornsweet, 1962; Levitt, 1971; Wetherill and Levitt, 1965) that was previously discussed in Section 2.9. Informally, the idea is that a user will adjust a pair of reproductions of an image that are respectively too bright and too dark. They progressively and respectively darken and lighten the pair (this is the double staircase) until the image appearance converges. A single staircase is not used here because often an observer will ‘overshoot’ the preferred gamma setting and make an image too bright (and they will only realize this when a much darker counterpart is shown). This darker counterpart is an informal invocation of the second staircase.

The stimulus in the experiment is the gamma adjustment (value of γ). As described in Section 2.6.2 , if we increase the value of gamma, this will darken the image. In contrast, if we decrease the gamma, we will brighten the image. With this in mind, in the experiment, the participants were asked to choose whether an image shown on

the screen appear either “too dark” or “too bright”. If the response is “too dark”, the next displayed stimulus will be two steps brighter to the current one ($\gamma_{(n+1)} = \gamma_n - 2 * accuracy$). In contrast, if the response is “too bright”, the next stimulus will be two steps darker ($\gamma_{(n+1)} = \gamma_n + 2 * accuracy$). Once the two staircases converge, then the step-size will be halved ($\gamma_{(n+1)} = \gamma_n \pm accuracy$). The accuracy of the experiment is the γ of 0.1. The experiment finishes when the step size has a minimum accuracy and the observer adjusts the image around a preferred gamma i.e. they make the image brighter then darker (i.e. reversal) around the preferred gamma. The number of reversals to be collected in the experiment is six.

We use two image sets in our experiments: ours (27) and the (15) standard images from Kodak (Franzen, 2004) (42 images in total). Our images were chosen to have different average intensities ranges from very dark to very bright. The Kodak images are often used as references for photographic reproduction and have a much more limited brightness range. For our experiments we use only the brightness information. Grey-scale are calculated from the Kodak images. Figure 4.1 shows some of the test images. The starting points for the bright and dark staircases are gamma-applied images that are overly bright and dark. Specifically we raise each image to the gamma which makes the average log value equal to -0.5 (overly bright) and -5 (overly dark). All images are displayed to observers in random order.

All the original images in our set are ‘gamma corrected’ for display (to deal with the display non linearity). We assume the images are coded as sRGB and invert this gamma as a first step. That is, all our images are assumed to be linear. Thus when an observer chooses a gamma of, say, 1.5 the image we display is equal to the linear image raised to the power of 1.5/2.2. Here the 1.5 changes the image contrast and 1/2.2 applies the display gamma.

The total number of participants in the pilot experiment are 12 (eight males and four females) with normal color vision, naïve (the participants have not seen the original be-



Figure 4.1: Some images in the dataset. The top two rows are our dataset. The bottom two are the greyscale images obtained from the Kodak dataset. Obviously, some of our images are either too dark or too bright, thus we expect the observer to choose either a fractional gamma or a gamma of higher than one to make the output images appear brighter or darker, respectively.

fore doing the experiment) for the goal of the experiment under the control environment conditions. The whole procedure per participant took approximately 60 minutes (two 25 minute sessions with a 10 minute break in the middle).

4.3 Results

To derive the gamma predicting model based on the image's entropy, first pre-calculate the information theoretic gamma ($\gamma_{optimal} = -1/mean(log(x))$). From mathematical argument alone, this gamma will maximize the image entropy (make details theoretically most visible). We then plot them against the chosen gamma received from the experiment. Figure 4.2 shows this relation. The error bar in the plot is the standard deviation.

From the pilot results, we found the linear relation between the optimal gamma and the chosen gamma that we have got from the experiment. As can be seen, the chosen gamma increases as the image's entropy increase. We believe this is a highly significant result as it provides some evidence that the observer is behaving in as an 'optimal information processor'. There are several psychophysical papers which propose that aspects of vision can be explained by appealing to the notion of information theory and optimal information processing. For example, Buchsbaum and Gottschalk (1983) proposed that color opponency could be explained by efficient information coding. Here we are proposing gamma adjustment is proportional to the gamma that maximizes entropy.

The reader might be interested to consider why the relation has a slope and intercept (as opposed to just a slope). The equation relating observer gamma to optimal gamma can be written as $\gamma_{display} = 0.62\gamma_{optimal} + 0.38$ (we invert the line equation relating x to y). The intercept 0.38 effectively imposes a minimum gamma (a gamma below this number is not possible). We believe this bound makes sense when we think about image

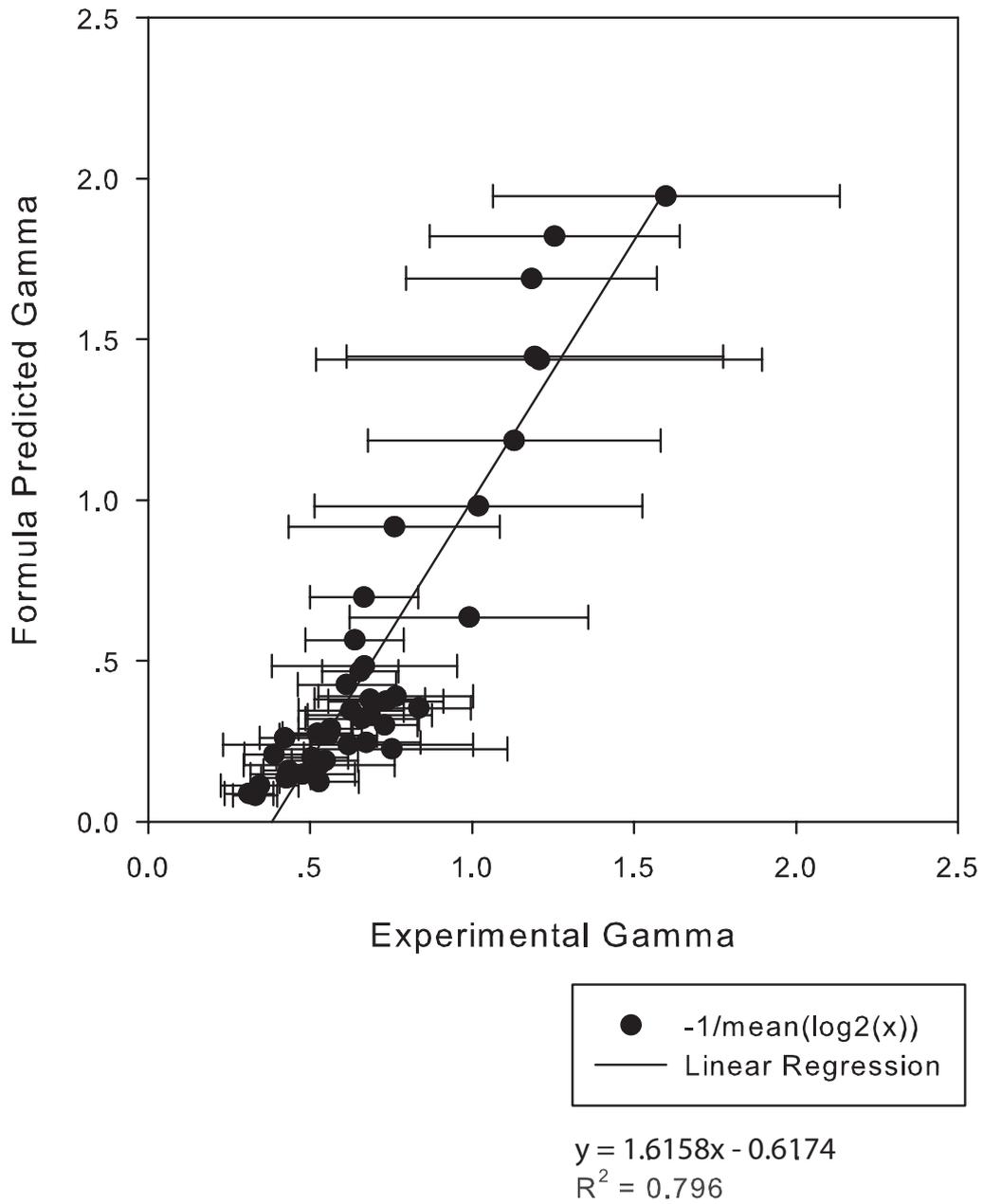


Figure 4.2: Average preferred gamma from 12 observers plotted against the theoretical optimal gamma ($-1/\text{mean}(\log_2(x))$) for 42 testing images.

content. First, our dark images clearly looked like night time scenes so an observer would not make these scenes so bright as to resemble daylight conditions. Also, there were often two illumination fields in each scene e.g. outdoor dark and indoor artificial lights. Thus the gamma has to compromise brightening the shadows and not destroying the highlight detail.

In addition, the reader might doubt why there are few images that have the entropy value above the value of 0.5, this is because actually our entropy calculation is calculated in log-2 space. This results in the cluster at the origin of the coordinate. Furthermore, results show that image with a narrow histogram tends to have wider range of chosen gamma (indicate there is a substantial versatility in observer preferences) [larger error bars] than the ones with broaden histogram and because most of the images with high value of log-mean normally have narrow histogram that condenses in the high value of intensities, this is the reason why brighter images often have a larger error bar compared to dark images. In the next section, we are going to provide more evidence that support this assumption by investigating three related works.

4.4 Discussion

4.4.1 Moroney's Gamma Adjustment

The Non-Linear Masking operator by Moroney (2000) performs local gamma correction. The operator uses power function where the exponent is computed pixel-by-pixel from a mask that derived from a low-pass filtered negative (1-image) monochrome version of the input image. This can be written as the following equation:

$$Output = Input^{2 \frac{0.5 - Mask}{0.5}} \quad (4.1)$$

If an input is bright, the negative is dark and so the mask value is less than 0.5. In

this case $2^{\frac{0.5-Mask}{0.5}}$ is bigger than one, so we have the exponent larger than one and we darken the image. A similar argument can be made for dark areas.

To make the operator acts as a global gamma predictor, we modify equation 4.1 by replacing the mask with a mean intensity of image μ , and replace the base of exponent of 2.0 with 2.5, since it gives a wider range of possible exponents. The modified equation can be written as:

$$Output = Input^{2.5\left(\frac{\mu-0.5}{0.5}\right)} \quad (4.2)$$

If μ is greater than 0.5, the exponent will be more than 1, in contrast, if μ are less than 0.5, the exponent will be less than 1. The range of possible exponent is between 0.4 and 2.5 (2.5^{-1} to 2.5^1).

4.4.2 CIECAM Gamma Adjustment

In the same sense, the lightness calculation found in the two versions of the well known Color Appearance Model; CIECAM97s Alessi et al. (1998) and CIECAM02 Moroney et al. (2002), are very similar to Moroney's operator, since the equations found in both versions are power functions. The Lightness J is calculated from the following equation:

$$J = 100 \left(\frac{A}{A_w} \right)^{cz} \quad (4.3)$$

where A/A_w is the ratio of the achromatic response of the sample to the response of the adopted white point, c is the predefined surround factor and z is the base exponential nonlinearity. There are two versions of CIECAM for the calculation of z :

$$z_{CIECAM97s} = 1 + F_{LL} n^{\frac{1}{2}} \quad (4.4)$$

$$z_{CIECAM02} = 1.48 + n^{\frac{1}{2}} \quad (4.5)$$

$$n = \left(\frac{Y_b}{Y_W} \right) \quad (4.6)$$

where F_{LL} found in CIECAM97s is a lightness contrast factor, n is the background induction factor and equal to the ratio of a luminance of the source background Y_b to a luminance of the white Y_w , and ranges from 0 for a background luminance factor of 0 to 1 for a background luminance factor equal to the luminance factor of the adopted white point (Fairchild, 2005).

In CIECAM the lightness ration A/A_W is somewhat laborious to calculate and not easy to relate to the original image (which is a key concern for us in our experiments). We solve this by relating CIECAM lightness to Colorimetric Luminance. Specifically, we generated an achromatic ramp of intensities and then calculated the ratios of Y/Y_W and A/A_W for the ramp. We then solved for the gamma which when applied to Luminance, resulted in values closest to CIECAM Lightness. We found that this relation is roughly equal to the square root of the ratio of input intensity:

$$\frac{A}{A_W} \approx \left(\frac{Y}{Y_W} \right)^{\frac{1}{2}} \quad (4.7)$$

Figure 4.3 illustrates this approximation.

This implies that if we apply a square root to the lightness equation, the exponent of this new equation is now comparable with the other predictors. The new lightness equation can be written as:

$$J \approx 100 \left(\frac{Y}{Y_W} \right)^{\left(\frac{c_z}{2} \right)} \quad (4.8)$$

For the average surround ($c = 0.69$, $F_{LL} = 1.0$), this result in the exponent of between 0.35 and 0.69 for CIECAM97s and between 0.51 and 0.86 for CIECAM02.

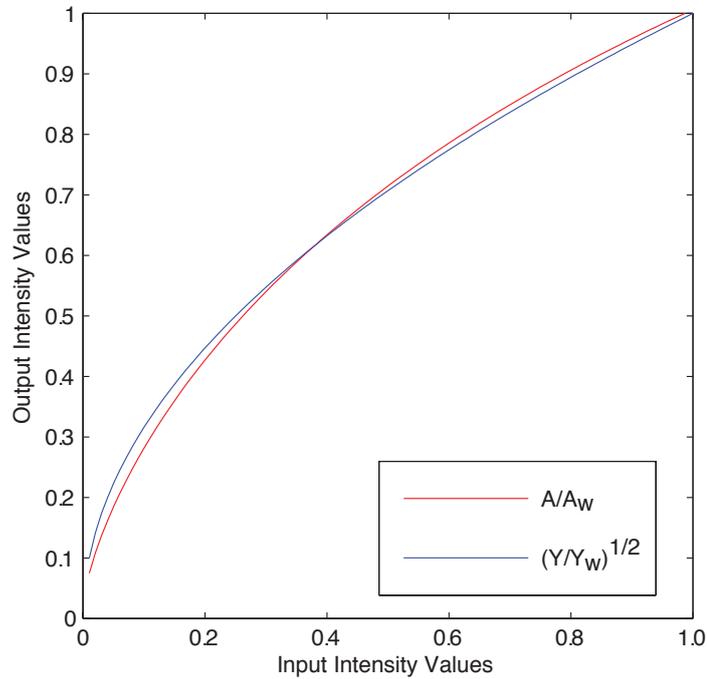


Figure 4.3: The approximation of A/A_W by the square root of the ratio of input intensities Y/Y_W .

4.4.3 Results

Figure 4.4 shows the plot between the observer's preferred gamma and the predicted gamma of the four predictors; ours, modified Moroney, CIECAM97s and CIECAM02. As can be seen, the four predictors have linear fittings to the perceived gamma from the experiment and positive slopes. Compared with the three antecedent predictors, our new formula has highest correlation to the linear fitting. This is assuming gamma adjustment is a linear function of the four formulae. It is the new 'optimal gamma' that best predicts the real observer adjustment. Furthermore, the modified version of Moroney and our predictor cover a broad range of useful gamma, indicates that both of them are appropriate to predict the chosen gamma, whereas the two CIECAM fail for predicting the gamma, since a little difference in prediction can make a large difference

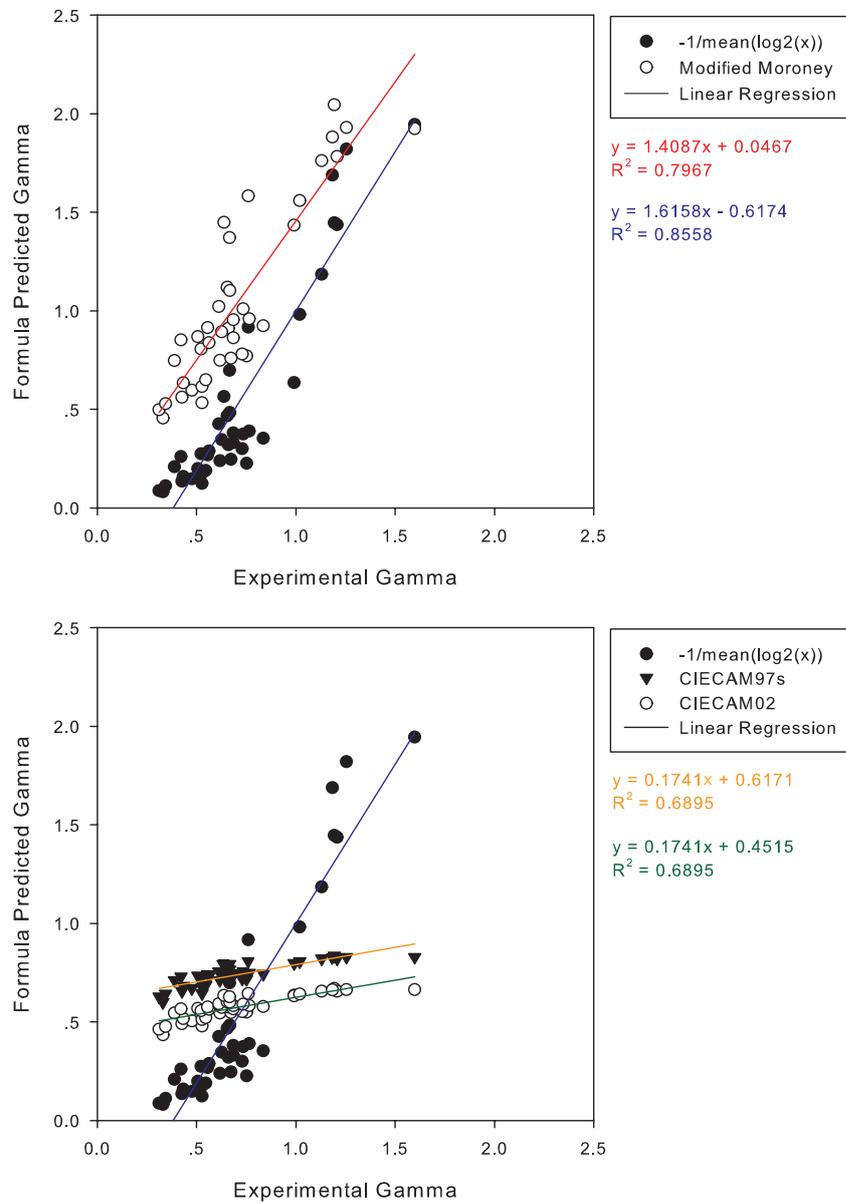


Figure 4.4: Different gamma predictors plotted against the average chosen gamma from the experiment, the corresponding lines to each dataset are the linear fitting to each of the predictor. Top, the modified version of Moroney and our predictor ($-1/\text{mean}(\log_2(x))$) are plotted. Bottom, the two versions of CIECAM (97s and 2002) are plotted along with our predictor.

to the observer's preferred gamma. Table 4.1 summarizes the four gamma predicting functions derived from the plot in Figure 4.4.

	Predicting equation	(R^2)
Ours	$y = 1.62x - 0.62$	0.86
Modified Moroney	$y = 1.41x + 0.05$	0.80
CIECAM97s	$y = 0.17x + 0.61$	0.69
CIECAM02	$y = 0.17x + 0.45$	0.69

Table 4.1: Gamma predicting equations along with the correlation coefficient (R^2) of the four models.

4.5 From the Experiment to an Automatic Gamma Adjustment Operator

There are numerous possible applications of the model. One application is to use the derived equations from the previous section as an automatic global tone reproduction operator.

We will evaluate experimentally how well this idea works as a post-processing process of the HDR imaging pipeline in Chapter 6. As a taster, Figure 4.5 shows five of our images and the automatically chosen display gamma. These images are derived from the following equation:

$$L_{out} = L_{in}^{\frac{1.62}{\gamma_{optimal} + 0.62}} \quad (4.9)$$

4.6 Conclusion

A simple and common way to adjust an image, making it brighter and darker, is to raise the image to the power of gamma: to carry out gamma adjustment. Recent theoretical



Figure 4.5: Examples of our automatic gamma adjustment. Left column shows original images. Right column shows result images.

work has provided a formula that, for a given image, calculates the gamma that maximizes entropy. That is, it returns the gamma which, theoretically, is the optimal in an information theoretic sense. In this paper, we provide evidence that there is a simple linear relationship between the information theoretically inspired optimal gamma and the gamma adjustment made by observers. Further, a similar (though, lesser fit) linear prediction is delivered by Moroney's non linear tone masking formula and via CIECAM-type formulae. Plausibly, these functions are also increasing the information content of image. Finally, we propose that the linear prediction formula, relating the theoretical optimum result to our experimental data, can be used to adjust gamma in images.

The predictor later will be further investigated whether it can serve as a post-processing process in the HDR imaging pipeline (we will discuss this in Chapter 6).

Chapter 5

Optimal Global Approximation to Spatially Varying TMOs

In this Chapter, we have developed a new approach to generate global tone-mapping operator (G-TMO) by matching the original image with the reference spatially-varying tone-mapped image using simple but elegant optimization technique called Pool Adjacent Violator Algorithm (PAVA).¹ The main advantage of the proposed technique is that we are certain not to introduce halo artifacts, due to the fact that the proposed algorithm is in global operation. However, the disadvantage is that the full local power of local processing is lost. To overcome this problem, we introduce adding detail back to the tone-mapped image using cross-bilateral filtering. The idea behind this is to take advantage of the global operation that does not introduce the spatial artifacts, combined with the advantage of bilateral filtering that bring back subtle details. We demonstrate that when these two approaches work together, they can achieve both good dynamic range compression and preserve subtle detail that often left out by global operator, without introducing halos. Remarkably, our new approach provides images similarly preferred to its spatially-varying TMOs.

¹To our knowledge, this work is the first application of PAVA to image reproduction.

This chapter is structure as follows: Section 5.1 presents relevant works and backgrounds. Then, Section 5.2 presents the proposed tone mapping operator. In Section 5.3, a psychophysical experiment is conducted in order to evaluate the image preference of the proposed operator. Finally, a conclusion is given in Section 5.4.

5.1 Background

As discussed earlier, G-TMOs compress or expand the input intensities. If the slope of the G-TMO function is less than one then detail is compressed in the output images. Such compression results in output images that unable to maintain local contrast and subtle details of the captured scene. One input intensity can map to only one output intensity. It is well known that G-TMOs produce images where contrast is lacking.

Spatially-varying tone mapping operator (SV-TMOs) on the other hand, take into account the spatial context where different functions are applied for different spatial pixel locations. One input intensity can result in different output intensities depends on its spatial context. Often this leads to improved local contrast which often are more preferred by observers. However, SV-TMOs are computational more expensive compared to G-TMOs. Moreover, SV-TMOs can introduce halo and other spatial artifacts such as ringing in the area around high contrast edges.

Our proposed operator is motivated by the idea of finding the optimal global function that best approximates the spatially-varying operator.

In a previous TMO study by Mantiuk and Seidel (2008), The tone-curves map input to output (for non optimal) approximations to the different SV-TMOs are analysed statistically using Principal Component Analysis (PCA). The first principal component resembles the sigmoidal (S-shaped) curve. Thus, Mantiuk and Seidel used a 5-parameter sigmoidal function, which can model either S-shaped sigmoidal or J-shaped curves, to approximate different SV-TMOs.

In our work we rather seek to find an optimal G-TMO instead of constraining the shape of the G-TMO. Specifically, given an input HDR image and the output of an arbitrary SV-TMO, we find the G-TMO that approximates the output in a least-squares sense. We make no prior assumption about the shape of the curve. We have found that our agnosticism is important to achieving the best image outputs. Indeed we found that our G-TMOs are not, in general well modeled by a sigmoidal or J-shaped curves.

We realized from the outset that mapping an input to an output image using a G-TMO could be cast as a Quadratic Programming (QP). Yet, QP is a rather general and computationally expensive procedure. Thus, instead, we propose using the simpler (and for this problem equivalent) Pool-Adjacent-Violators-algorithm (PAVA). The benefit of using PAVA over QP is that PAVA is lower computational burden, since in the worst case, the computational complexity of the PAVA is $O(n)$ compared to $O(n^2)$ of QP (Burdakov et al., 2004).

Figure 5.1 gives an example of the PAVA approximation to the tone-mapped image obtained using Ashikhmin's operator (Ashikhmin, 2002). As can be seen, the global tone-curve derived from the PAVA results in a very similar output to its reference. However, as one might notice, subtle details of the cloud are lost. This is because, although the approximation is optimal, it still global operator, thus, the power to boost local details is absent. The output appears flatter and details that appear on SV tone-mapped are missing.

There are several approaches to recover the lost details as discussed in Section 2.5, here, the edge-preserved smoothing bilateral filtering technique is used. Bilateral filtering adds photometric distance enforcement to traditional filtering proximate pixels are averaged together in proportion to their similarity to a central pixel. By using bilateral filtering an image can be decomposed into two components: a low frequency component where edges are preserved called base layer, and a high frequency component called detail layer defined to be the original minus the base. Base and detail layers are defined



Figure 5.1: PAVA approximation and its reference spatially-varying TMO. (a) The original HDR scene, for displaying purpose, the image was first normalized and then applied a gamma of 2.2. (b) The reference SV tone-mapped image, the image was tone-mapped using Ashikhmin's operator (Ashikhmin, 2002). (c) The image resulted from PAVA.

as:

$$L_{base} = BF(reference, target) \quad (5.1)$$

$$L_{detail} = L_{in} - L_{base} \quad (5.2)$$

where both *reference* and *target* images are in log-space.

Durand and Dorsey (2002) used bilateral filtering to compress the dynamic range of the HDR images. There, a bilateral filter was used to decompose an HDR image into two-scale layers: a large-scale HDR base layer and a small scale LDR detail-layer. The base-layer is then tone-mapped to LDR using a shape-constrained G-TMO (Tumblin et al., 1999). After processing, the detail is added back onto the output image. Mathematically, the operator is defined as:

$$L_{base} = BF(L_{in}, L_{in}) \quad (5.3)$$

$$L_{Detail} = L_{in} - L_{Base} \quad (5.4)$$

$$L_{Base'} = \text{G-TMO}(L_{Base}) \quad (5.5)$$

$$L_{Out} = L_{Base'} + L_{Detail} \quad (5.6)$$

L_{in} denotes a brightness image (e.g. average of R, G and B channels) in logarithmic domain. To obtain the base layer L_{base} , the brightness image is filtered using the bilateral filter $BF()$. The detail layer is then calculated as $L_{detail} = L_{in} - L_{base}$. Next, the base layer L_{base} is tone-mapped. The final step is to reconstruct the tone-mapped image L_{out} by adding the detail layer L_{detail} to the tone-mapped base layer $L_{base'}$.

Noted that, in Equation 5.1, if the *reference* image is different from the *target* image, then the operating is called ‘‘cross-bilateral’’. In this way, the cross-bilateral

filter uses the *reference* image to determine the photometric weighting where the actual smoothing is carried out on the *target* image. Eisemann and Durand (2004) and Petschnigg et al. (2004) used a different reference image to control the distance enforcement in order to transfer detail across images.

In the next section, the proposed global tone-mapping operator that optimally approximates spatially-varying TMO using a constrained optimization PAVA technique is proposed. We will demonstrate how to utilize the cross-bilateral filtering to bring back details that are lost due to the PAVA.

5.2 The Optimal Approximation Operator

Figure 5.2 illustrates the global framework of the proposed method. For faster processing, the HDR luminance image is first down-sampled. In this work, a VGA size image (640×480) has been used.² Next, a reference tone-mapped image is generated using an arbitrary spatially-varying TMO is also down-sampled. Then, a global tone-curve is generated by matching, in a least-squares sense, the HDR input and the reference LDR down-sampled using an optimization technique called PAVA. Next, the full-size tone-mapped luminance image is obtained by applying the tone-curve to the original HDR. Details are now added back to the tone-mapped image using a cross-bilateral filtering. An additional step called unsharp-masking is then performed. Finally, the color tone-mapped image is obtained by using a luminance ratio technique.

Essentially, the above processes can be summarized into four main parts as follows: (1) pre-processing, (2) tone-curve optimization (the main contribution of the chapter), (3) detail recovery, and (4) post processing.

The main tone-mapping processing is summarized as:

²To preserve the original dynamic range of the HDR when down-sampling as much as possible, a nearest-neighbor technique has been used.

$$L_{out} = DR(G-TMO(L_{in})) \quad (5.7)$$

$G-TMO()$ is the global tone-curve applied to the input image, which can be either derived from a down-sampled or full-size image. $DR()$ denotes the detail recovery process (see 5.2.2 below). The output color image can be then reconstructed using Equation 2.18. An additional unsharp masking step is available to tweak preferred contrast (not necessary for the experiments we report later).

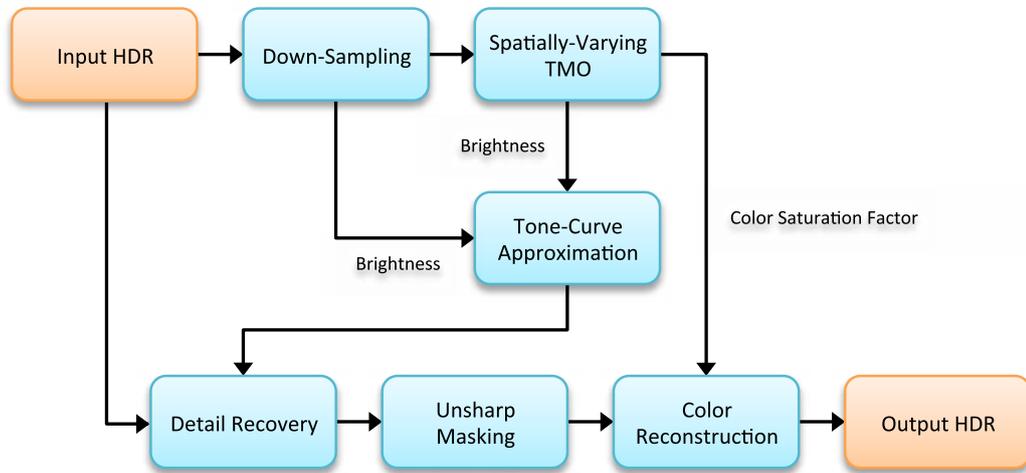


Figure 5.2: Image processing of the optimal approximation operator.

5.2.1 Tone-Curve Optimization

Let us now consider the function $G-TMO$. The optimal global tone-curve applied to an image is not an arbitrary function. Rather, it should be monotonically increasing both to avoid intensity inversions and to allow image manipulations to be undone. Almost all curve adjustments made to images (brightening, contrast changes and gamma) are monotonically increasing functions.

Given an HDR image and its spatially-varying tone-mapped LDR image, we want

to find a 1-D surjective and monotonically increasing function that best maps HDR to LDR. We point out that we have a choice of how to *encode* the HDR brightness image. Throughout this paper we will choose to represent our input data in the *log* domain. We do this because relative differences are most meaningful to human observers (we have approximately a *log* visual response) and, practically, most tone mappers take the *log* of the input image as input.

Pool-Adjacent-Violators-Algorithm (PAVA)

PAVA (Barlow and Brunk, 1972; Barlow et al., 1972; de Leeuw et al., 2009) is one of the most widely used technique for solving monotonic (isotonic) regression problem in the field of statistics (Härdle, 1992; Robertson et al., 1988). PAVA is a simple iterative algorithm for solving monotonic (either increasing or decreasing) regression problems. It pools values together until an optimal solution is found. Generally speaking, PAVA starts from the leftmost. PAVA move to the right until it encounter the violation. Then PAVA replace the violated pair by their average, and back-average to the left until the monotonicity is satisfied. Then, PAVA repeats this process to right until it finally reach the right most.

Let us assume we have a two-dimensional data $\{(X_i, Y_i)\}_{i=1}^n$ where X_i is in ascending order. We seek a monotonically increasing function $\hat{m}()$ that minimizes

$$\sum_{i=1}^n (Y_i - \hat{m}(X_i))^2 \quad (5.8)$$

subject to

$$\hat{m}(X_{(1)}) \leq \hat{m}(X_{(2)}) \leq \dots \leq \hat{m}(X_{(n)}). \quad (5.9)$$

The PAVA algorithm (originally presented in (Härdle, 1992), which optimally solves this problem, works as follows:

1. Sort the data according to X (Y is reordered according to the X) [X_i is in monotonically increasing order but Y_i may not be] $\{\hat{m}(X_{(i)})\}_{i=1}^n$.
2. Starting from the leftmost of the function $Y_{(1)}$ move to the right and stop if the pair (Y_i, Y_{i+1}) violates the monotonicity constraint we seek: $Y_i > Y_{i+1}$.
3. Pool $Y_{(i)}$ and the adjacent $Y_{(i+1)}$ together and replacing them both by their average, $Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i)} + Y_{(i+1)})/2$.
4. Next check that $Y_{(i-1)} \leq Y_{(i)}^*$. If not, pool $\{Y_{(i-1)}, Y_{(i)}, Y_{(i+1)}\}$ into one average. Continue to pool to the left unless the monotonicity requirement is satisfied.
5. Proceed to the right and keep repeating from step 2 until an monotonically increasing solution is derived.

In the worst case, the computational complexity of PAVA is $O(n)$ (very expensive if every pixel has a unique intensity!). This worst case scenario of PAVA happens in step 4 when the algorithm has to search back to our first data point. However, in practice (for our application) the worst case is never encountered (the expected complexity of PAVA can be shown to be low).

We also (slightly) modify PAVA so the procedure runs very rapidly and for the problem at hand (tone-mapping of images) also produces pleasing results (the default PAVA can result in a loss of detail). We demonstrate this in the following two sections.

Reducing the complexity of PAVA with fixed quantization levels

Suppose we have $n + 1$ quantization levels of $X : q_n, q_n - 1, \dots, 0$. If the minimum \log -value is M then let $q_i = \frac{i}{n}M$. For each quantization level there may be many different output values. But, the complexity of PAVA is bounded by the n quantization levels (say 32, compared with the millions of pixels in the original image). We calculate PAVA only for $X_i = q_i$ and the corresponding output Y (a single quantization level can

have many different output values). Of course our input HDR image is not quantized thus we calculate the outputs by interpolation (a linear interpolation is used here). For an arbitrary X (a brightness in the input HDR image whose brightness is between quantile levels u and $u+1$) we calculate the position of this brightness between the appropriate quantization levels:

$$\alpha = \frac{X - q_u}{q_{u+1} - q_u} \quad (5.10)$$

We assume that the output Y is the same linear combination of the outputs for these quantization levels:

$$\hat{m}(X) = (1 - \alpha)\hat{m}(q_u) + \alpha\hat{m}(q_{u+1}) \quad (5.11)$$

Smoothed PAVA

Unfortunately, PAVA can produce long flat levels (step function) [see Figure 5.3]. Such tone-curves (though least-squares optimal) will not lead to good looking images. The visual meaning of the flat part of a tone-curve is that a range of input values are all mapped to the same output values (with a potential loss of detail as a result). To avoid this flattening behavior, we simply smooth the PAVA solution which is mathematically equivalent to finding an optimal smooth curve as part of the general optimization (Friedman and Tibshirani, 1984). But, post-smoothing has the advantage that it is carried out on a small number of regression points (e.g. for the 32 quantization levels) which leads to much faster processing than smoothing the set of actual mapping points. The robust loess estimation procedure (Cleveland, 1979; Cleveland and Devlin, 1988) is used to smooth our data.

An example of PAVA and its robust smoothed version (PAVAs) along with a scatter plot of pixel distribution are shown in Figure 5.3. There is a flat area occurred in the highlight of the original PAVA function which is smoothed out.

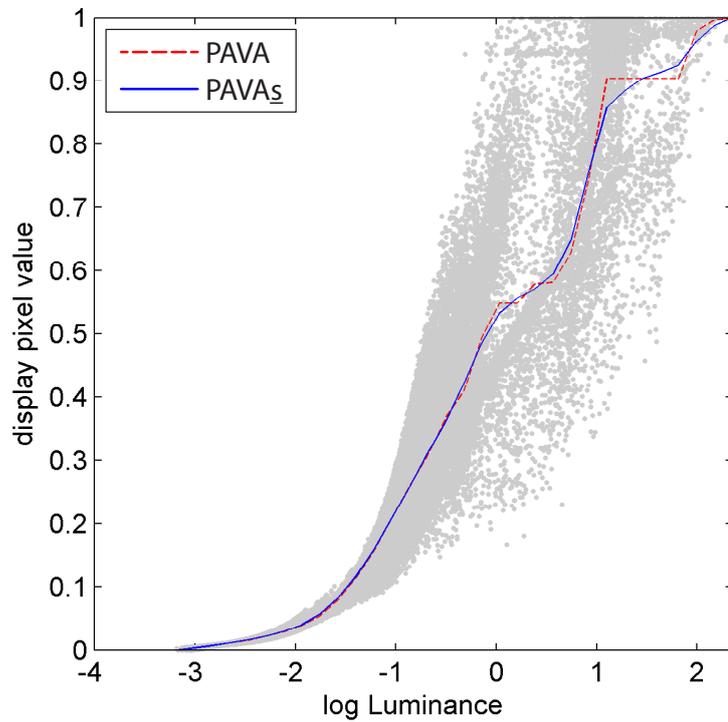


Figure 5.3: Tone-curves resulted from PAVA and its robust smoothed version (PAVA_S).

Even though the basic PAVA solution is optimal in terms of Root Mean Square Error (RMSE) [refer to Section 2.8.1], flat regions of tone-curve can produce poor image outputs. Figure 5.4 illustrates this issue. There we see a smoother tone rendering in PAVA_S than in PAVA. Significantly, we have found that the PAVA_S curve has almost the same RMSE. For the data fit shown in Figure 5.3, the optimal PAVA output captures 95.02% of the spatially varying TMOs output (RMSE of 0.0498) whereas the smoothed version (PAVA_S) is almost as good 94.94% (RMSE of 0.0506). Similar smoothing results were found for all images we tested. It is true that the smoothed function is not an optimal solution, however having the optimal solution does not mean having a visually pleasing image. Smoothed PAVA produces visually more pleasing output at the cost of a very small decrement in the data fit.



Figure 5.4: Compare PAVA output (left) and PAVA_s output (right) (we see a smoother tone rendering in the PAVA_s than the original PAVA).

Discussion

Tone-curve manipulations tend to stretch contrast in some image areas and compress in others. When the derivative of the tone curve is less than 1 detail is being compressed and when it is greater than 1 there is an increase in contrast. Thus, while tone-mapped images generate by Smoothed PAVA are often similar to the spatially-varying outputs they can look very flat (typically in the highlight region area, the best global tone-curve has a <1 derivative). Simple unsharp masking (as shown in the workflow of Figure 5.2) can often ameliorate this problem.

However, unsharp masking does not always work. Spatially-varying TMOs such as Retinex manipulate images in a very local manner. Indeed, one of the problems of Retinex is that it can introduce artifacts such as halo around high contrast edges. Although, often problematic the processes that make halos also add a contrast boost (a ‘punch’) to local areas of images which is often preferred results. Of course, the more locally an image is processed, the less well a global tone-curve can approximate the outputs. Figure 5.5 gives an example of this poor *visual fit*. Arguably the Retinex output on the left has an almost ‘hyper’ realism. But, the G-TMO output on the right looks unnaturally flat.



Figure 5.5: An example of a poor fit of our tone-curve approximation. Left, the reference image resulted from Retinex has higher local contrast compared to our mapping result on the right.

5.2.2 Detail Recovery

Thus, in common, with previous works on G-TMOs, we must recover the detail that is missing in the G-TMO reproduction. Here, the detail recovery process is introduced. The idea is to use an edge preserving smoothing technique called cross-bilateral filtering. Denoting the output of the smooth PAVA tone-curve as L_{approx} , the output of the proposed operator is:

$$L_{out} = L_{in} + BF(L_{in}, L_{Approx} - L_{in}) \quad (5.12)$$

The advantage of (5.12) is that the BF is applied only to the gain (mapping input to output). We know it is this gain (directly related to the global tone-curve) that causes flattening. So, (5.12) alters the gain to bring back detail. Equation (5.12) is a concise summary of the exact computation needed to bring back the detail. However, mathematically the approaches of (5.12) and (5.1) through (5.6) are equivalent).

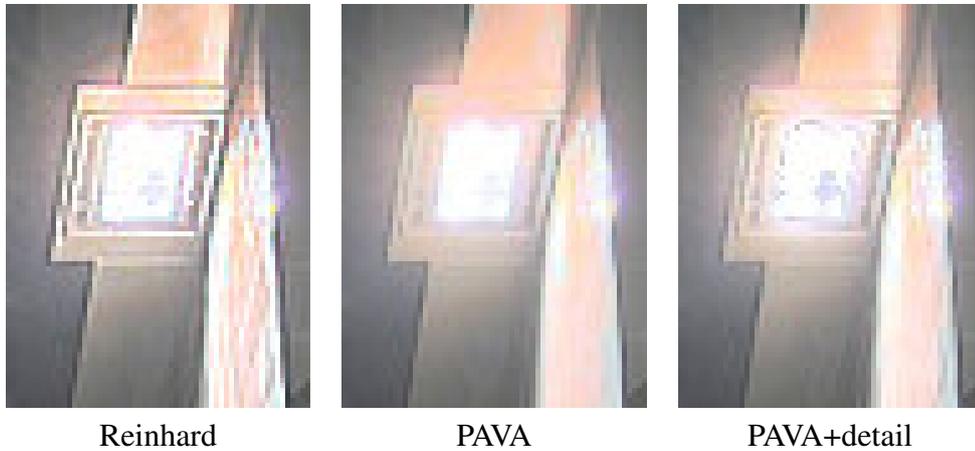


Figure 5.6: Details recovered from the cross-bilateral of the proposed operator. Images were cropped from images shown in middle row of Figure 5.7.

To add back the colour, Equation 2.18 (mentioned in Chapter 2) is used. Figure 5.7 shows the outputs of SV-TMOs (left panels) and our G-TMO approximation (center column). It is clear a G-TMO can produce a very good approximation to a spatially

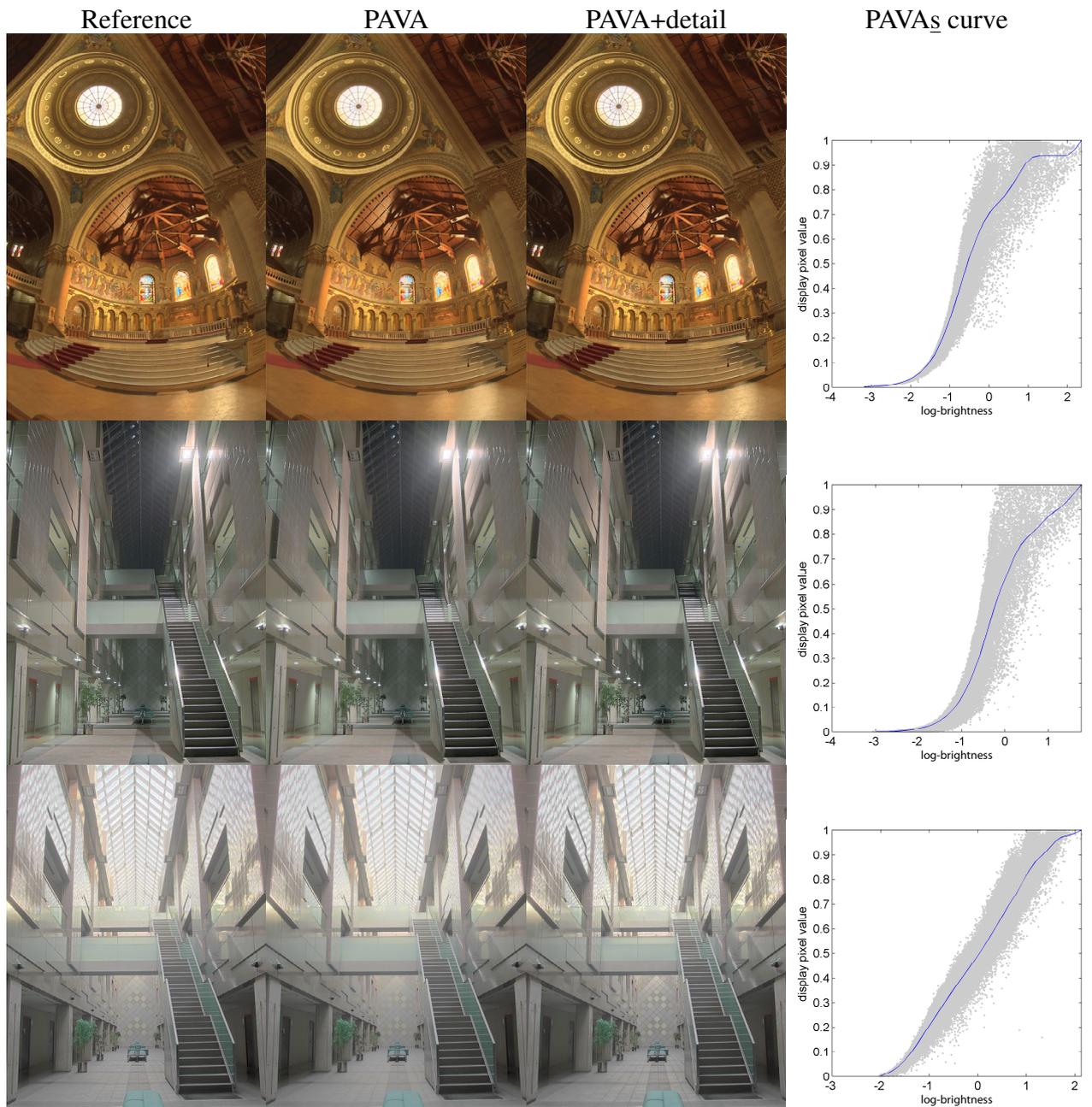


Figure 5.7: A comparison chart of HDR pixel encodings. Results of our approximation to the three tone-mapped images that we test. Left column, the reference tone-mapped image. Our approximation output before and after the detail recovery process are given in middle left and middle right columns, respectively. Right column, approximated tone-curve (blue) and the scatter plot showing the correlation between each pixel of the LDR and HDR brightness values. Note that: Although the regression is done in log-log space, we found that it is more intuitive to visualize it in log-linear space.

varying algorithm. The images in the rightmost column are the outputs from the G-TMO after applying a detail recovery procedure.

5.3 Preference Experiment

To test image preference of the optimal approximation operator against spatially-varying TMOs, a paired comparison psychophysical experiment (Engeldrum, 2000) was conducted. The purpose of this experiment was to compare all the test SV-TMO images and to ensure that in general the G-TMO images derived from the smooth-PAVA optimization are comparable.

Five HDR images were used in our experiments: Figure 5.8 shows the image dataset. Tone-mapped images used in the experiment were generated by applying the optimal approximation operator to tone-mapped images of 3 well-known spatially-varying TMOs. Table 5.1 lists TMOs that have been used in the experiment (noted, the labels we use to index the results in Table 5.2 and Figure 5.9). To avoid unfaithful implementation and bias from parameter selection problems, we use the images available from a previous study (Drago et al., 2010).

There are 6 tone-mapped images per HDR image (3 reference tone-mapped images, and 3 created using smooth-PAVA in the context of the processing workflow shown in Figure 5.2). Thus, there are 30 tone-mapped images in total. These images were evaluated by 21 observers (9 males and 12 females) with normal color vision, naive to the goal of the experiment. Images were viewed under controlled experimental conditions (ISO 3664:2009, 2009).

There were a total of 75 pairwise comparisons in the experiment (5 images * 15 pairs of algorithms). Observers were asked to make judgements of the TMOs based on overall appearance. For each image pair, observers were instructed to observe the two tone-mapped images and select the one they preferred. Images were shown randomly

on the left or the right of the screen. All algorithms were shown roughly an equal number of times left and right. The whole procedure per participant took approximately 8-10 minutes. In order to evaluate the results, preference scores are generated using Thurstone's Law of Comparative Judgement Case V (Thurstone, 1927).

5.3.1 Algorithm Parameters

For PAVA, the number of quantization points that we use is 32. For bilateral filter parameters, we set the spatial closeness (σ_s) to 2% of the image size and the photometric similarity (σ_r) to 0.6 *log* units since these two values performed consistently well for all test images. All processing is carried out in the brightness domain. The color output is created according to Equation 5.

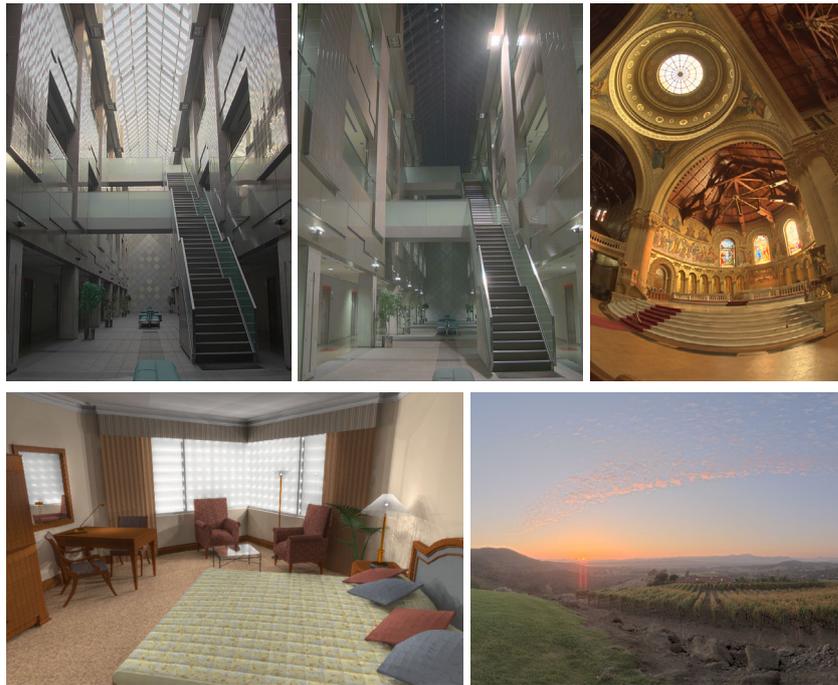


Figure 5.8: 5 well-known images used in the experiment. From left to right AtriumMorning, AtriumNight, HotelRoom, Memorial, and NapaValley. Image courtesy of Frédéric Drago, Paul Debevec, Simon Crone, and Spheron AG.

Name	Label	Category
A Tone Mapping Algorithm for High Contrast Images (Ashikhmin, 2002)	A	S
Photographic Tone Reproduction (Reinhard et al., 2002)	P	
Retinex adapted to tone-mapping (Drago et al., 2002)	R	
Optimal Global Approximation to A	\hat{A}	G
Optimal Global Approximation to P	\hat{P}	
Optimal Global Approximation to R	\hat{R}	

Table 5.1: Tone mapping operators used in the experiment together with their labels and categories used in the paper. For the category, S stands for spatially-varying TMOs, G stands for the optimal global approximation operator.

5.3.2 Experimental Results

The average preference scores of 5 test scenes from the 21 subjects was given in Figure 5.9 (the x-axis shows the operators). The actual scores are given in Table 5.2. The numbers in Table 5.2 are the number of times a particular algorithms is preferred. The last 2 columns compared the pooled algorithm performance: SV-TMO vs G-TMO. Using statistical assumptions (Thurstone’s law of comparative judgement Case V) we can turn these raw preference numbers into a preference score with confidence intervals. See (Green and MacDonald, 2002) for a full discussion of how this is done.

In Figure 5.9, The 95% confidence interval (error bars) are shown in a normalised preference score interval. The y-axis here (preference score) can be interpreted as a z-score: if one operator were strongly preferred it could, in theory have a score of 2 and if an operator was strongly unpreferred its normalised preference could be as low as -2. In cases where algorithms do not deliver strong preference, the preference scores tend to cluster around 0 on this normalised scale. If error bars do not overlap then one algorithm is better than another at the 95% confidence interval.

The results show that on average among six TMOs, TMO P (Reinhard’s photographic tone reproduction operator) was most preferred but that the G-TMO \hat{P} produced images which were not significantly different (in terms of preference). We can also

clearly see that A (Ashikhmin) has the lowest score whereas its global approximating version \hat{A} have a significant higher score (the score difference exceeds the confidence interval) and in fact ranked the third overall. This indicates that TMO which generate too much detail may be judged poorer than its global approximating version.

The *pooled* average scores for each category (S for spatially-varying and G for the proposed global approximation) are given on the right of the Figure. Since the two scores fall in to the same interval scale, it is reasonably to conclude that in general, the perceived quality of the optimal approximation operator is similar to its spatially-varying operators.

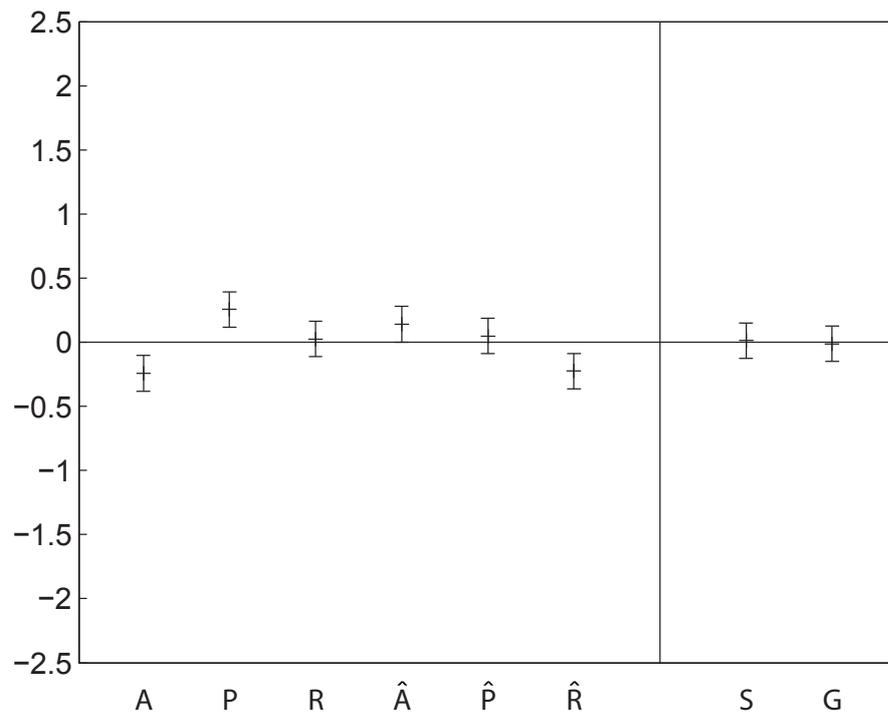


Figure 5.9: Overall preference scores of 6 different TMOs (6 scores on the left). The average score of spatially-varying TMOs and the average score of the proposed operator are shown on the right.

A	P	R	\hat{A}	\hat{P}	\hat{R}	S	G
213	315	267	291	273	216	795	780

Table 5.2: The number of times each TMO is preferred over the others.

5.4 Conclusions

We demonstrate that many spatially-varying TMOs can be visually approximated by the optimal global tone-curve approximation so long as care is taken to preserve local detail. A psychophysical experiment validates our method. Our experiment demonstrates that for three of the most widely used spatially-varying TMOs, their optimal global tone-curve approximation generates images that were equally preferred. This is a significant result as our global operator does not suffer from spatial artifacts (such as halos) and can be implemented to run extremely rapidly (only a fractional of second is required compared to several seconds of the spatially-varying TMOs). Indeed the global tone-curve itself can be *learned* using only a small input thumbnail.

Chapter 6

Gamma Adjustment for HDR Post Processing

The intention of the work presented in this chapter is to introduce an additional post-processing step that is simple, fast, and automated to improve the image appearance of the tone-mapped images (specifically, the optimal global approximation TMO proposed in Chapter 5). The optimal global approximation TMO although optimal in a least-square sense does not guarantee the optimal in term of visual appearance since the operator only attempts to generate a similar output to the reference spatially-varying TMO. So, there are many cases that the reference TMO results in either a brighter or darker image than it should be, e.g. due to wrong parameter setting. One way to solve this problem is to use the automatic gamma adjustment proposed in Chapter 4. Thus, in this chapter a combined process of the optimal global approximation TMO coupled with an automatic gamma adjustment is proposed. The proposed combination is straightforward, it consists of two sub-processes: (1) tone-mapping, and (2) gamma adjustment. Specifically, the first process is the G-TMO proposed in Chapter 5. The second process is the automatic gamma adjustment presented in Chapter 4. We call this a “G-TMO γ ”. To evaluate the image preference of the proposed operator, a paired comparison exper-

iment is conducted. The observer's task is to choose the image that is more visually pleasing. Our hope is that G-TMO γ images are preferred over the simpler G-TMO approach. This chapter is structured in three parts. The revised gamma prediction is given in Section 6.1. Section 6.2 details the experimental set up. Section 6.3 discusses the results.

6.1 Revisiting the Gamma Adjustment Equation

When we thought about using gamma adjustment as a post-processing step, we visually considered the possibility that a linear regression might not be the best way to fit the data that is the scatter plot relating actual gamma adjustments to image entropy (Figure 4.2) was only roughly linear. So, the linear regression (Equation 4.9) is replaced by a sigmoid (S-shape) in order to better fit the data (The correlation coefficient (R^2) of the sigmoid is improved to 0.91, in contrast, the linear one has R^2 of 0.86). Figure 6.1, illustrates the sigmoidal regression (red solid line) along with the linear regression (blue dashed line). G-TMO γ (as the extension of Equation 5.7) is summarised as:

$$L_{out} = DR(G - TMO(L_{in}))^\gamma, \quad (6.1)$$

$$\gamma_{optimal} = 0.1434 + \frac{2.005}{\left(1 + e^{-\frac{\gamma - 1.071}{0.1827}}\right)} \quad (6.2)$$

where $\gamma_{optimal}$ is the optimal gamma that maximise the entropy of image an can be calculated as in Equation 2.20.

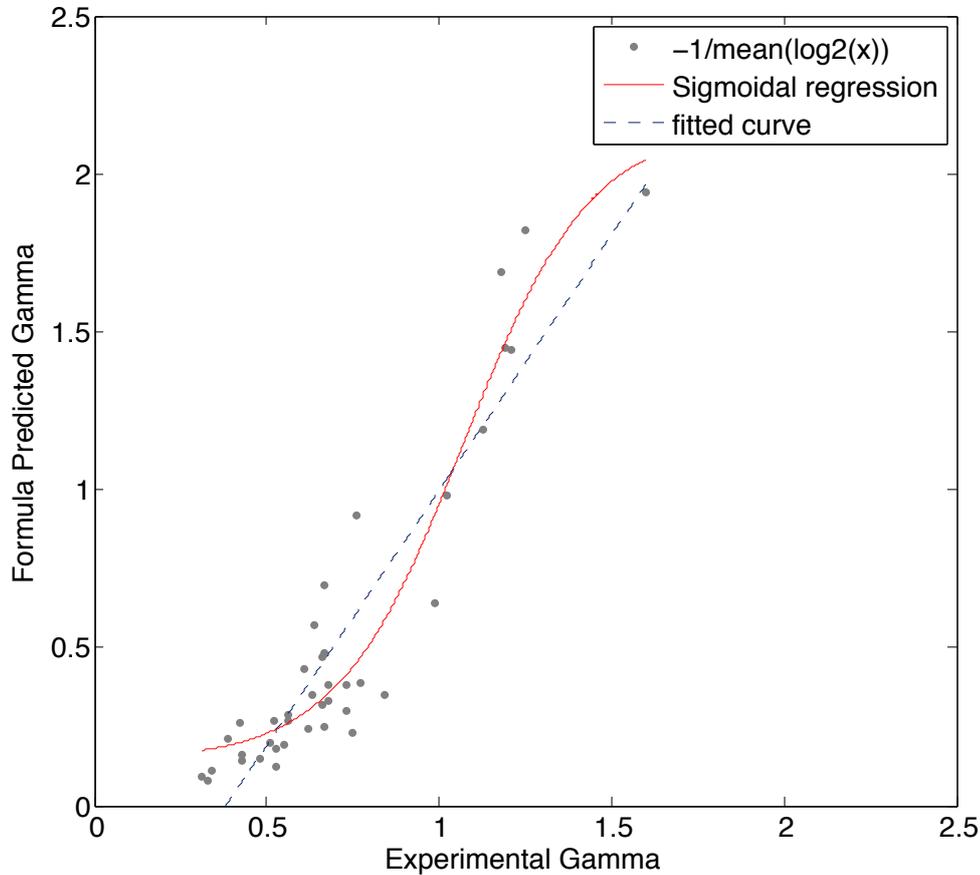


Figure 6.1: Sigmoidal regression versus linear regression of the experimental results conducted in Chapter 4.

6.2 Preference Experiment

A paired comparison psychophysical experiment was conducted. The purpose of the experiment was to evaluate the perceived image quality of the $G\text{-TMO}_\gamma$ compared with TMO alone. In other words, given a tone-mapped image and its gamma-adjusted counterpart, which one is more visually pleasing? In this experiment, only two methods are compared: G-TMO and $G\text{-TMO}_\gamma$.

The HDR images used in the experiment are the same as ones used in the experiment

conducted in Chapter 5. There are five HDR scenes (see Figure 5.8), three test TMOs as listed in Table 5.1 were applied per each scene, with two global renditions (G-TMO and G-TMO γ). Thus, there are 15×2 pairs of images. With two repetitions of each image pairs, there were a total of 30 comparisons in the experiment. These images were evaluated by 24 observers (twelve males and twelve females) with normal color vision, naïve to the goal of the experiment under the control experimental environment. The whole experiment took approximately few minutes per observer. Figure 6.2 shows some of the test images.

6.3 Results and Discussions

Figure 6.3 shows average preference scores for five scenes. The error bar in the plot represent the 95% confidence interval for 24 observers which has the value of 0.2829. From the figure, it is clear that, on average, GTMO γ is broadly preferred over G-TMO alone (the score difference exceeds the error bar). Figure 6.4 shows the scores obtained for individual test scenes. The results show that, among all five test scenes, there are three of images 6.4b 6.4c 6.4e that G-TMO γ has preference scores which are higher and which exceed the confidence intervals. For another image (6.4a), the scores are equal. Only the image shown in Figure 6.4d where the tone-mapping alone is preferred but again this is within the confidence interval.

6.4 Conclusion

The goal of our study in this chapter was to investigate the role of gamma adjustment on the final image appearance in the case of HDR imaging pipeline. A psychophysical paired comparison experiment was conducted where we compared tone-mapped images of the proposed TMO alone and the combination of this TMO with the gamma



Figure 6.2: (left column) images were obtained with the optimal global approximation TMO. (right column) images were obtained with the combination process proposed here (generated by applying an automatic gamma adjustment to the right column images).

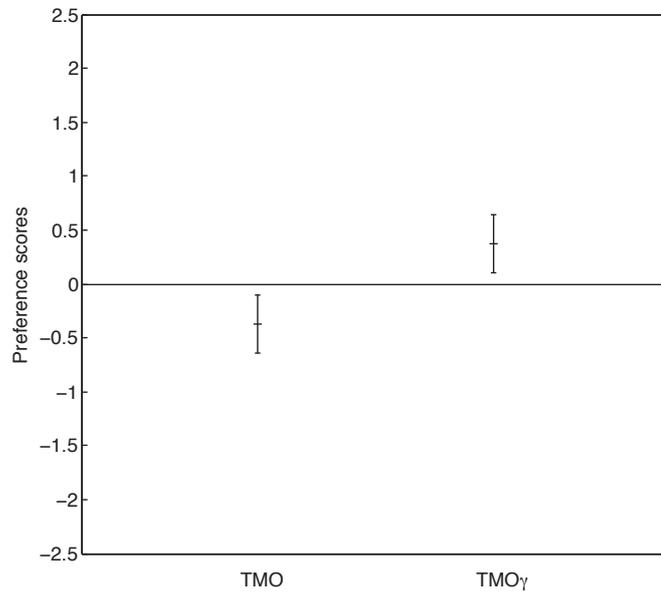
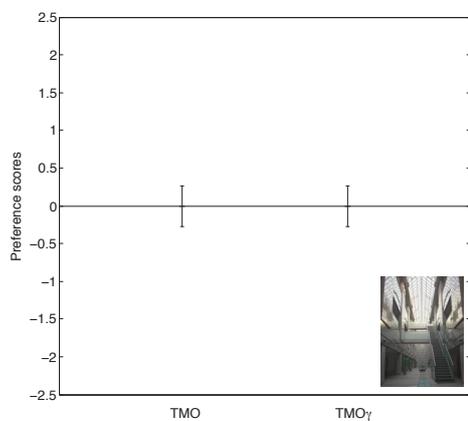


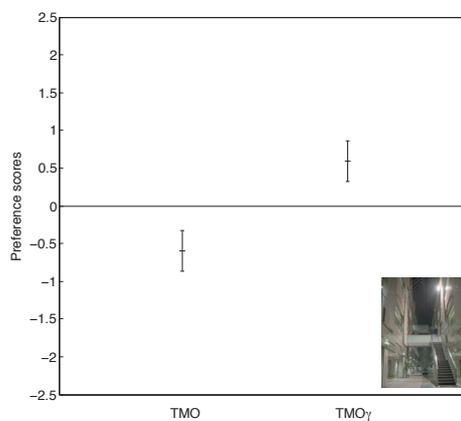
Figure 6.3: Overall preference scores over five test scenes. The operators are labeled as TMO and TMO γ for the tone mapping alone and the combined processed proposed in this chapter (tone mapping coupled with automatic gamma adjustment), respectively.

adjustment.

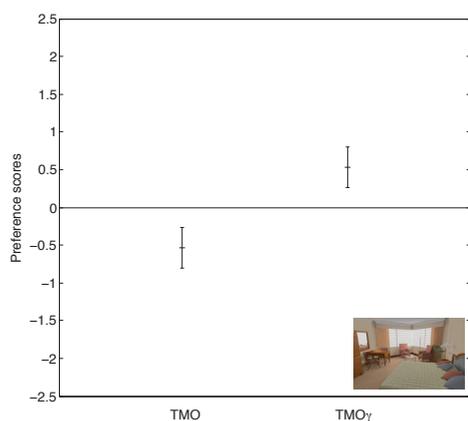
The experiment demonstrates that for the test scenes, the use of the TMO proposed in Chapter 5 followed by the post-processing process of gamma adjustment derived from the relationship between the information theoretically inspired optimal gamma and the gamma adjustment made by observers as proposed in Chapter 4, results in more preferred tone-mapped images which implies that the proposed gamma predictor in Chapter 4 is suitable for HDR post-processing.



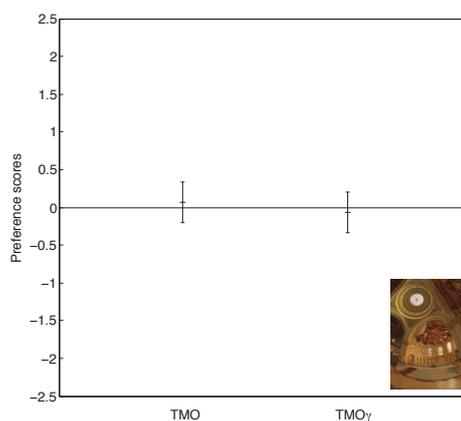
(a)



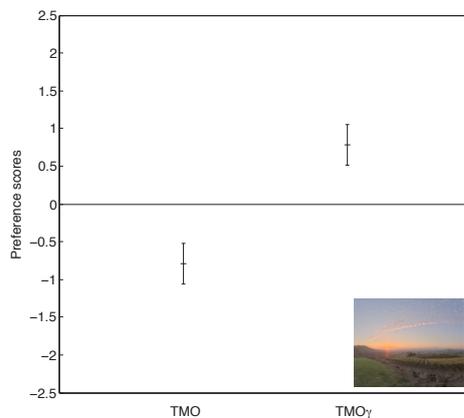
(b)



(c)



(d)



(e)

Figure 6.4: Preference scores for each scene. (a) to (e) preference scores of each test scene.

Chapter 7

Conclusion and Future works

7.1 Conclusion

The aim of this thesis was to simplify the HDR imaging pipeline in the context of digital photography. To achieve this aim, several parts of the pipeline are needed to be simplified. In this thesis, three parts of the pipeline have been investigated: HDR image encoding, the tone-mapping operator, and gamma-adjustment as a post-processing step.

In Chapter 3, HDR image encoding has been investigated. In digital photography application, the RGBE radiance encoding is widely used. However, it is not always possible or easy to use the RGBE format since not all conventional imaging applications support this encoding and those that do take considerable time to read or write and this hinders workflow productivity. We propose a simple, fast, and practical HDR encoding that bridges the gap between the conventional and HDR imaging workflows. Our proposed method is based, directly, on the conventional 12 and 16-bit/component integer TIFF image format. We take a linear HDR image and apply a fractional gamma (a number in the range [0.33 and 0.45]) and quantize the resulting pixel values). Our implicit hypothesis is that most typical HDR images do not have extreme dynamic ranges and so a simple extension of conventional image representations will suffice. We consider

the potential of our method for the tone-mapping application by both measuring the perceptual color difference ΔE S-CIELAB of different tone-mapped representations, and by conducting a 2AFC psychophysical experiment to compare the encoding preference between the proposed encoding and RGBE. We do this by using tone mapping operators from the literature to map HDR scenes into the display range. If our new encoding suffices then observers should either prefer or not care the outputs delivered when our new encoding is used (compared with RGBE). The results show that the proposed encoding both 12 and 16-bit frameworks perform well and give equivalent results compared to the state-of-the-art RGBE. Further in terms of file storage, irrespective of the actual bit-depth required by the encoding, a lossless compress version of our encoding often needs less storage space than the compressed RGBE. Crucially, compared to RGBE our method does not suffer from the software incompatibility, suggesting that our method, used in the normal HDR imaging workflow, is a good candidate for HDR image encoding and could easily be integrated with the existing TIFF image library.

Following HDR image encoding, we focus on the HDR visualization (i.e. rendering images for display). The visualization of HDR images on typical display devices is well-known to be problematic since the dynamic range of the HDR has to be compressed. Here, we divide the visualization step into two stages: tone-mapping and the post-processing.

In Chapter 4, we investigated the post-processing process called gamma-adjustment operator. We attempt to find an automatic way to determine an appropriate gamma parameter of the operator based on image's content. A psychophysical experiment called a "double-staircase" was conducted in order to investigate the relationship between the theoretic optimal gamma (Finlayson and Xu, 2012) that maximizes image entropy and the user preferred gamma (the gamma obtained from the experiment). Interestingly, the results show that, there is a strong relationship between the two gammas, indicating that the combination of entropy calculation together with the regression line derived

from this relation effectively provides an automatic gamma adjustment. Significantly, our new auto gamma (a function of optimal gamma) better fits our data compared with similarly optimized formula found in the literature.

Our discussion continues with a discussion of tone mapping. A tone mapping operator (or TMO) maps the high dynamic range input image to the smaller display range. TMOs, broadly, fall into two classes. Global tone maps (every unique input brightness maps to a single output) and spatially varying (where this ‘functional’ constraint does not hold). The latter can produce flatter images which are lacking in high frequency detail. The latter preserve detail everywhere but can introduce artifacts such as halos around high contrast edges. Spatially-varying algorithms have the additional disadvantage that they can be very expensive computational procedures compared to global algorithms. In Chapter 5, we presented a hybrid method that combines both global and local operators that can overcome these problems. Our idea is to run a spatially varying tone mapping operator and then to find the global mapping best taking the input image to the spatially varying output. In our approach we, for the first time, adopt the PAVA monotonic optimization procedure to find the desired mapping. Not only is PAVA fast (it can be executed in real time), but the PAVA global TMO delivers surprisingly good dynamic range compression. However, as expected, some fine details are compressed (compared with the spatially varying counterpart).

The global tone-curve approach has a specific advantage as the tone-curve can be generated using a small thumbnail of the input image making the approach significantly faster as the thumbnail can be 16, 32 or more times smaller than the original image. The proposed operator results in a tone-mapped image that often visually similar to its reference spatially-varying tone-mapped image that is far more computational expensive. This indicates that some spatially-varying TMOs can be visually approximated by the optimal global tone-curve approximations.

To evaluate the proposed operator, a paired comparison psychophysical experiment

has been conducted. Here we compare the spatially varying outputs to those delivered by the global PAVA counterpart. The preference scores show that in general the perceived quality of the proposed operator is equally preferred to a range of spatially-varying tone-mapping operators. In addition, for spatially-varying TMOs that generates large amount of halo artifact such as Retinex, the optimal global approximation results in a more natural looking as the global-operator does not suffer from this artifact.

To complete the simplification of the HDR visualization. In Chapter 6, the optimal global operator proposed in Chapter 5 is combined with an automatic gamma adjustment that is introduced in Chapter 4. Here, the automatic gamma adjustment serves as a post-processing step (a final step after tone-mapping) of the imaging pipeline. We find that the application of our HDR visualization delivers more pleasing images that contain proper exposure compared to the tone-mapping alone. To evaluate the outcomes, a paired comparison psychophysical experiment was conducted. Images that are gamma corrected post-tone mapping are preferred over those that use tone mapping alone.

HDR plays an important role in many applications as in digital photography. The use of HDR imaging will become more widespread in the near future because of the rapid development of the supported hardware technology. To make the transition between LDR and HDR as smooth as possible, each stage of the HDR imaging pipeline is needed to be simplified. With this research, some of them are simplified as discussed in the preceding chapters.

7.2 Future Works

Here, we discuss some of the remaining problems that we think are important and would like to investigate in the future. These include:

For the HDR encoding proposed in Chapter 3, with the HDR displays become available, it is interesting to validate that whether the proposed encoding can directly be used

in these displays. One possible way is to conduct a preference psychophysical 2AFC experiment where the observer's task is to identify whether the proposed encoding is similarly perceived to the reference IEEE tiff float format on such display.

Another interesting topic regards to the encoding aspect is the topic of image compression. Currently, a lossless compression technique from the Joint Bi-level Image Experts Group (JBIG) (JBIG, 1993) is categorized as one of the best lossless compression technique. This could be useful if we conduct an experiment to test whether JBIG (instead of using ZIP compression that we already tested in Section 3.3.3) can further reduces file size of the proposed encoding compared with the RGBE.

For the automatic gamma adjustment proposed in Chapter 4, currently, we are conducting an experiment to test the performance of our gamma predictor to a large set of test images. In the hope to ensure that the predictor gives a consistency results across the wide range of images.

For the optimal global approximation TMO proposed in Chapter 5, our model could be improved by constraining slope of the mapping function to be larger than zero (avoids flat area of the function) instead of smoothing it as we did here in the thesis, this would result in the function that has a lower RMSE compared to PAVAS. Currently, we are studying how we can implement this in the most effective way and also planning to validate this assumption by constructing the psychophysical experiment.

Whether in terms of algorithms or experiments with people, this work is quite theoretical. There will be many steps to implement these algorithms into commercial products. In the future, my intention is to take this work and place it into the real commercial framework (Colour Lab at UEA has close commercial contacts with number of companies).

References

- Adams, A., 1971a, *Camera and Lens*, Morgan & Morgan, Dobbs Ferry.
- Adams, A., 1971b, *The Print*, Morgan & Morgan, Dobbs Ferry.
- Adams, A., 1972, *The Negative*, Morgan & Morgan, Dobbs Ferry, (1971a), *Camera and Lens*, Morgan & Morgan, Dobbs Ferry, NY. (1971b), *The Print*, Morgan & Morgan, Dobbs Ferry, NY. *Negative*, Morgan & Morgan, Dobbs Ferry, NY.
- Akenine-Möller, T., Haines, E., and Hoffman, N., 2008, *Real-Time Rendering 3rd Edition*, A. K. Peters, Ltd., Natick, MA, USA, ISBN 987-1-56881-424-7.
- Akenine-Moller, T., Moller, T., and Haines, E., 2002, *Real-Time Rendering*, A. K. Peters, Ltd., Natick, MA, USA, 2nd ed., ISBN 1568811829.
- Alessi, P. J., Fairchild, M. D., Hashimoto, K., Hunt, R. W., Luo, M. R., Mori, L., Nayatani, Y., Seim, T., Sobagaki, H., Richter, K., Tc, C., (usa, P. J. A., (usa, M. D. F., (japan, K. H., (england, R. W. G. H., (england, M. R. L., (japan, L. M., (japan, Y. N., (norway, T. S., (japan, H. S., and (germany, K. R., 1998, "The cie 1997 interim colour appearance model (simple version), ciecam97s," .
- Ashikhmin, M., 2002, "A tone mapping algorithm for high contrast images," in Gibson, S. and Debevec, P. E., eds., *Rendering Techniques*, vol. 28 of *ACM International Conference Proceeding Series*, Eurographics Association, ISBN 1-58113-534-3.
- Banterle, F., Artusi, A., Debattista, K., and Chalmers, A., 2011, *Advanced High Dynamic Range Imaging: Theory and Practice*, AK Peters (CRC Press), Natick, MA, USA, ISBN 9781568817194.
- Banterle, F., Ledda, P., Debattista, K., and Chalmers, A., 2006, "Inverse tone mapping," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, GRAPHITE '06, pp. 349–356, ACM, New York, NY, USA, ISBN 1-59593-564-9, doi:10.1145/1174429.1174489, URL <http://doi.acm.org/10.1145/1174429.1174489>.

- Barlow, R. and Brunk, H., 1972, "The isotonic regression problem and its dual," *Journal of the American Statistical Association*, vol. 67, pp. 140–147.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D., 1972, *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*, John Wiley & Sons, London-New York-Sydney.
- Bodmann, H.-W., Haubner, P., and Marsden, A., 1979, "A unified relationship between brightness and luminance," in *CIE Proc. 19th Session (Kyoto) 99-102, 1979, CIE 50-1979*.
- Bogart, R., Kainz, F., and Hess, D., 2003, "Openexr image file format," .
- Buchsbaum, G. and Gottschalk, A., 1983, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," in *Proc. of the Royal Society of London*, vol. 220, pp. 89–113.
- Burdakov, O., Sysoev, O., Grimvall, A., and Hussian, M., 2004, "An algorithm for isotonic regression problems," in *European Congress on Computational Methods in Applied Sciences and Engineering ECCOMAS*, pp. 1–9, University of Jyväskylä.
- Čadík, M., Wimmer, M., Neumann, L., and Artusi, A., 2008, "Evaluation of hdr tone mapping methods using essential perceptual attributes," *Computers & Graphics*, vol. 32, no. 3, pp. 330–349, ISSN 0097-8493, URL <http://www.cg.tuwien.ac.at/research/publications/2008/CADIK-2008-EHD/>.
- Chiu, K., Herf, M., Shirley, P., Swamy, S., Wang, C., and Zimmerman, K., 1993, "Spatially nonuniform scaling functions for high contrast images," pp. 245–253.
- CIE, 1931, *Documents concernant le film "les Galeries Levy et Cie", 1932*, Cambridge University Press, URL <http://books.google.co.uk/books?id=QaNtLwEACAAJ>, cIE (1932). Commission internationale de l'Eclairage proceedings, 1931. Cambridge University Press, Cambridge.
- Cleveland, W. S., 1979, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, ISSN 01621459, doi:10.2307/2286407, URL <http://dx.doi.org/10.2307/2286407>.
- Cleveland, W. S. and Devlin, S. J., 1988, "Locally weighted regression: An approach to regression analysis by local fitting," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 596–610, ISSN 01621459, doi:10.2307/2289282, URL <http://dx.doi.org/10.2307/2289282>.

- Cornsweet, T. N., 1962, "The staircase-method in psychophysics," *American Journal of Psychology*, vol. 75, pp. 485–491.
- David, H., 1988, *The method of paired comparisons*, Oxford University Press, New York, 2nd ed.
- DCI, 2005, "Dci, "digital cinema system specifications v1.0"," .
- DCI, 2008, "Dci, "digital cinema system specifications v1.2"," .
- de Leeuw, J., Hornik, K., and Mair, P., 2009, "Isotone optimization in r: Pool-adjacent-violators algorithm (pava) and active set methods," *Journal of Statistical Software*, vol. 32, no. 5, URL <http://www.jstatsoft.org/v32/i05>.
- Debevec, P., 2005, "Image-based lighting," in *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, ACM, New York, NY, USA, doi:10.1145/1198555.1198709, URL <http://doi.acm.org/10.1145/1198555.1198709>.
- Debevec, P. E. and Malik, J., 1997, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, pp. 369–378, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, ISBN 0-89791-896-7, doi:<http://dx.doi.org/10.1145/258734.258884>, URL <http://dx.doi.org/10.1145/258734.258884>.
- Denecker, K., Neve, P. D., Assche, S. V., de Walle, R. V., Lemahieu, I., and Philips, W., 2002, "Psychovisual evaluation of lossy cmyk image compression for printing applications," *Comput. Graph. Forum*, vol. 21, no. 1, pp. 5–17.
- Devlin, K., 2002, "A review of tone reproduction techniques," Tech. Rep. CSTR-02-005, Department of Computer Science, University of Bristol, URL <http://www.cs.bris.ac.uk/Publications/Papers/1000680.pdf>.
- Dicarlo, J. M. and Wandell, B. A., 2000a, "Rendering high dynamic range images," vol. 3965, pp. 392–401, SPIE, doi:10.1117/12.385456, URL <http://dx.doi.org/10.1117/12.385456>.
- Dicarlo, J. M. and Wandell, B. A., 2000b, "Rendering high dynamic range images," vol. 3965, pp. 392–401, SPIE, doi:10.1117/12.385456, URL <http://dx.doi.org/10.1117/12.385456>.
- Drago, F., Martens, W., Myszkowski, K., and Seidel, H.-P., 2002, "Perceptual evaluation of tone mapping operators with regard to similarity and preference," Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany.

- Drago, F., Myszkowski, K., Seidel, H.-P., and Martens, W. L., 2010, “Gallery of reference tone mapped images,” <http://www.mpi-inf.mpg.de/resources/tmo/NewExperiment/TmoOverview.html>, [Online; accessed 06-July-2011].
- Durand, F. and Dorsey, J., 2002, “Fast bilateral filtering for the display of high-dynamic-range images,” in *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 257–266, ACM, New York, NY, USA, ISBN 1-58113-521-1, doi:10.1145/566570.566574, URL <http://dx.doi.org/10.1145/566570.566574>.
- Eisemann, E. and Durand, F., 2004, “Flash photography enhancement via intrinsic relighting,” in *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pp. 673–678, ACM, New York, NY, USA, doi:<http://doi.acm.org/10.1145/1186562.1015778>, URL <http://doi.acm.org/10.1145/1186562.1015778>.
- Engeldrum, P. G., 2000, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press.
- Fairchild, M., 2005, *Color Appearance Models*, Wiley.
- Fairchild, M. D., Johnson, G. M., Kuang, J., and Yamaguchi, H., 2004, “Image appearance modeling and high-dynamic-range image rendering,” in *In APGV 2004: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, p. 159, ACM Press.
- Fattal, R., Lischinski, D., and Werman, M., 2002a, “Gradient domain high dynamic range compression,” *ACM Trans. Graph.*, vol. 21, pp. 249–256, ISSN 0730-0301, doi:<http://doi.acm.org/10.1145/566654.566573>, URL <http://doi.acm.org/10.1145/566654.566573>.
- Fattal, R., Lischinski, D., and Werman, M., 2002b, “Gradient domain high dynamic range compression,” in *Proceedings of ACM SIGGRAPH 2002*, pp. 249–256.
- Ferwerda, J. A., 1998, “Fundamentals of spatial vision,” in *In Applications of visual perception in computer graphics*.
- Finlayson, G. and Xu, R., 2012, *Gamma Adjustment for Maximizing Information in Images*, US Patent 2012/0114236 A1.
- Franzen, R., 2004, “Kodak lossless true color image suite,” URL <http://www.r0k.us/graphics/kodak/index.html>.
- Friedman, J. and Tibshirani, R., 1984, “The monotone smoothing of scatterplots,” *Technometrics*, vol. 26, pp. 243–250.

- Gescheider, G., 1997, *Psychophysics: The Fundamentals*, L. Erlbaum Associates, ISBN 9780805822816, URL <http://books.google.co.uk/books?id=gAFtxKQI1mAC>.
- Girod, B., 1993, *Digital Images and Human Vision* (A. B. Watson, ed.), Chapter What's wrong with mean-squared error, the MIT press, b. Girod. in book *Digital Images and Human Vision* (A. B. Watson, ed.), chapter What's wrong with mean-squared error, pages 207-220. the MIT press, 1993.
- Gonzalez, R. C. and Woods, R. E., 2006, *Digital Image Processing (3rd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, ISBN 013168728X.
- Green, P. and MacDonald, L., 2002, *Colour engineering: achieving device independent colour*, Wiley SID series in display technology, Wiley, ISBN 9780471486886, URL <http://books.google.com/books?id=tn09voxr6agC>.
- Härdle, W., 1992, *Applied Nonparametric Regression (Econometric Society Monographs)*, Cambridge University Press, ISBN 0521429501, URL <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0521429501>.
- Hough, D., 1981, "Applications of the proposed ieee 754 standard for floating-point arithmetic," doi:10.1109/C-M.1981.220381, computer.
- IEEE, 1985, *ANSI/IEEE 754-1985, Standard for Binary Floating-Point Arithmetic*, IEEE, New York, a preliminary draft was published in the January 1980 issue of IEEE Computer, together with several companion articles author = IEEE Task P754,.
- ISO 3664:2009, 2009, *Graphic technology and photography – Viewing conditions*, ISO, Geneva, Switzerland.
- ITU, 2002, "ITU-R BT.709-5-2002 parameter values for the hdtv standards for production and international programme exchange," <http://www.freestd.us/soft/29074.htm>.
- JBIG, 1993, "JBIG, Progressive Bi-level Image Compression, ISO/IEC International Standard 11544, ITU Recommendation T.82," .
- Jobson, D. J., Rahman, Z.-U., and Woodell, G. A., 1996, "Retinex image processing: Improved fidelity to direct visual observation," Tech. rep., NASA Langley Research Center.
- Kuang, J., Johnson, G. M., and Fairchild, M. D., 2007, "icam06: A refined image appearance model for hdr image rendering," *J. Vis. Comun. Image Represent.*, vol. 18, no. 5, pp. 406–414, ISSN 1047-3203, doi:<http://dx.doi.org/10.1016/j.jvcir.2007.06.003>.

- Land, E., 1974, "The retinex theory of colour vision," *Proceedings of The Royal Institution of Great Britain*, pp. 23–58.
- Larson, G. W., 1998, "Logluv encoding for full-gamut, high-dynamic range images," *J. Graph. Tools*, vol. 3, no. 1, pp. 15–31, ISSN 1086-7651, doi:10.1080/10867651.1998.10487485, URL <http://dx.doi.org/10.1080/10867651.1998.10487485>.
- Larson, G. W., Rushmeier, H., and Piatko, C., 1997, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, pp. 291–306.
- Larson, G. W. and Shakespeare, R., 1998, *Rendering with Radiance*, Morgan Kaufmann Publishers, ward, Larson Greg and Shakespeare, Rob.
- Levitt, H., 1971, "Transformed up-down methods in psychoacoustics," *Acoustical Society of America Journal*, vol. 49, doi:10.1121/1.1912375, URL <http://dx.doi.org/10.1121/1.1912375>.
- Magic, I. L. ., 2003, "Openexr," <http://www.openexr.org>, [Online; accessed 01-December-2012].
- Mann, S. and Picard, R., 1994, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," Tech. Rep. 323, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, also appears, IS&T's 48th annual conference, Cambridge, Massachusetts, May 1995.
- Mantiuk, R., Daly, S., Myszkowski, K., and Seidel, H.-P., 2005, "Predicting visible differences in high dynamic range images - model and its calibration," in *Human Vision and Electronic Imaging X, IS&T/SPIEs 17th Annual Symposium on Electronic Imaging (2005)*, pp. 204–214.
- Mantiuk, R., Myszkowski, K., and Seidel, H.-P., 2004, "Visible difference predictor for high dynamic range images," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 2763–2769.
- Mantiuk, R. and Seidel, H.-P., 2008, "Modeling a generic tone-mapping operator," *Computer Graphics Forum*, vol. 27, pp. 699–708, doi:10.1111/j.1467-8659.2008.01168.x.
- Meylan, L. and Susstrunk, S., 2006, "High dynamic range image rendering using a retinex-based adaptive filter," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2820–2830, URL <http://rr.epfl.ch/12/>.

- Moroney, N., 2000, "Local color correction using non-linear masking," in *Color Imaging Conference*, pp. 108–111.
- Moroney, N., Fairchild, M. D., Hunt, R. W. G., Li, C., Luo, M. R., and Newman, T., 2002, "The ciecam02 color appearance model," in *IS&T/SID 10 th Color Imaging Conference*, pp. 23–27.
- Panoscan, 2012, "Panoscan mark iii," <http://www.panoscan.com>, [Online; accessed 06-July-2012].
- Pappas, T. N. and Safranek, R. J., 2000, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, pp. 669–684, Academic Press.
- Pattanaik, S. N., Fairchild, M. D., Ferwerda, J. A., and Greenberg, D. P., 1998, "A multiscale model of adaptation and spatial vision for realistic image display," pp. 287–298.
- Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., and Toyama, K., 2004, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, pp. 664–672, ISSN 0730-0301, doi:<http://doi.acm.org/10.1145/1015706.1015777>, URL <http://doi.acm.org/10.1145/1015706.1015777>.
- Rahman, Z., Jobson, D., and Woodell, G., 1997a, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. on Image Proc.*
- Rahman, Z., Jobson, D. J., and Woodell, G. A., 1996, "A multiscale retinex for colour rendition and dynamic range compression," *SPIE International Symposium on Optical Science, Engineering and Instrumentation, Applications of Digital Image Processing XIX, Proceedings SPIE 2825*, Andrew G. Tescher, ed.
- Rahman, Z., Woodell, G. A., Jobson, D. J., and Retinex, T. M., 1997b, "A comparison of the multiscale retinex with other image enhancement techniques," pp. 426–431.
- Reinhard, E., Khan, E. A., Akyz, A. O., and Johnson, G. M., 2008, *Color Imaging: Fundamentals and Applications*, A. K. Peters, Ltd., Natick, MA, USA, ISBN 1568813449, 9781568813448.
- Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J., 2002, "Photographic tone reproduction for digital images," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques, SIGGRAPH '02*, pp. 267–276, ACM, New York, NY, USA, ISBN 1-58113-521-1, doi:<http://doi.acm.org/10.1145/566570.566575>, URL <http://doi.acm.org/10.1145/566570.566575>.

- Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P., 2005, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ISBN 0125852630.
- Robertson, T., Wright, F. T., and Dykstra, R. L., 1988, *Order restricted statistical inference*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Chichester.
- Schlick, C., 1994, "Quantization techniques for visualization of high dynamic range pictures," pp. 7–20, Springer-Verlag.
- scRGB, 2003, "Iec. 2003. 61966-2-2. extended rgb colour space – scrgb, multimedia systems and equipment - colour measurement and management - part 2-2: Colour management," .
- Seetzen, H., Heidrich, W., Stürzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., and Vorozcovs, A., 2004, "High dynamic range display systems," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 760–768.
- Shannon, C. E., 1951, "Prediction and entropy of printed english," *Bell Systems Technical Journal*, vol. 30, pp. 50–64.
- Spheron, 2012, "Spheronvr," <http://www.spheron.com>, [Online; accessed 06-July-2012].
- Stevens, J. C. and Stevens, S. S., 1963, "Brightness function : Effects of adaptation," *J. Opt. Soc. Am.*, vol. 53, no. 3, pp. 375–385, doi:10.1364/JOSA.53.000375, URL <http://www.opticsinfobase.org/abstract.cfm?URI=josa-53-3-375>.
- Stevens, S. S., 1946, "On the theory of scales of measurement," *Science*, vol. 103, no. 2684, pp. 677–680, doi:10.1126/science.103.2684.677, URL <http://www.sciencemag.org/content/103/2684/677.short>.
- Stevens, S. S. and Stevens, J. C., 1960, "Brightness function: Parametric effects of adaptation and contrast," *J. Opt. Soc. Am.*, vol. 50, 11, p. 1139.
- Stokes, M., Anderson, M., Chandrasekar, S., and Motta, R., 1996, "A standard default color space for the internet - srgb," <Http://www.w3.org/Graphics/Color/sRGB>.
- Thurstone, L. L., 1927, "A law of comparative judgment." *Psychological Review*, vol. 34, no. 4, pp. 273–286, ISSN 0033-295X, doi:10.1037/h0070288, URL <http://dx.doi.org/10.1037/h0070288>.

- Tomasi, C. and Manduchi, R., 1998, "Bilateral filtering for gray and color images," pp. 839–846, URL <http://dx.doi.org/10.1109/ICCV.1998.710815>.
- Tumblin, J., Hodgins, J. K., and Guenter, B. K., 1999, "Two methods for display of high contrast images," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 56–94.
- Tumblin, J. and Rushmeier, H., 1993, "Tone reproduction for realistic images," *IEEE Computer Graphics and Applications*, vol. 13, pp. 42–48, ISSN 0272-1716, doi:<http://doi.ieeecomputersociety.org/10.1109/38.252554>.
- Tumblin, J. and Turk, G., 1999, "Lcis: a boundary hierarchy for detail-preserving contrast reduction," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques, SIGGRAPH '99*, pp. 83–90, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, ISBN 0-201-48560-5, doi:<http://dx.doi.org/10.1145/311535.311544>, URL <http://dx.doi.org/10.1145/311535.311544>.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., 2004, "Image quality assessment: From error visibility to structural similarity," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 13, no. 4, pp. 600–612.
- Ward, G., Unknown, "High dynamic range image encodings," URL <http://www.anywhere.com/gward/hdrenc/Encodings.pdf>.
- Ward, G. and Simmons, M., 2004, "Subband encoding of high dynamic range imagery," in *Applied Perception in Graphics and Visualization*, pp. 83–90, doi:10.1145/1012551.1012566.
- Ward, G. and Simmons, M., 2005, "Jpeg-hdr: a backwards-compatible, high dynamic range extension to jpeg," in *ACM SIGGRAPH 2005 Courses, SIGGRAPH '05*, ACM, New York, NY, USA, doi:10.1145/1198555.1198708, URL <http://doi.acm.org/10.1145/1198555.1198708>.
- Weiss-AG, 2012, "Civetta 360," <http://weiss-ag.info/>, [Online; accessed 06-July-2012].
- Wetherill, G. and Levitt, H., 1965, "Sequential estimation of points on a psychometric function." *Br J Math Stat Psychol*, vol. 18.
- Whitted, T., 1979, "An improved illumination model for shaded display," in *Proceedings of the 6th annual conference on Computer graphics and interactive techniques, SIGGRAPH '79*, pp. 14–, ACM, New York, NY, USA, ISBN 0-89791-004-4, doi:10.1145/800249.807419, URL <http://doi.acm.org/10.1145/800249.807419>.

- Whitted, T., 1980, "An improved illumination model for shaded display," *Commun. ACM*, vol. 23, no. 6, pp. 343–349, ISSN 0001-0782, doi:10.1145/358876.358882, URL <http://doi.acm.org/10.1145/358876.358882>.
- Wolin, D., Johnson, K., and Kipman, Y., 1998, "The importance of objective analysis in image quality evaluation," in *NIP14: International Conference on Digital Printing Technologies*, pp. 603–606, IS&T, Toronto, Ontario, Canada.
- Wyszecki, W. S. S. G., 2000, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Wiley, 2nd ed.
- Zhang, X. and Wandell, B. A., 1996, "A spatial extension of cielab for digital color image reproduction," in *in: Proceedings of the SID Symposiums*, vol. 27, pp. 731–734.