Combining aesthetics and perception for display retargeting

Thèse soutenue à Rennes le 23 Octobre 2017
devant le jury composé de :

Kadi BOUATOUCH
Professeur Emérite, Université de Rennes / Président du Jury

Céline LOSCOS
Professeur, Université de Reims Champagne-Ardenne, / Rapporteur

Patrick LE CALLET
Professeur, Université de Nantes / Rapporteur

Frédéric DUFAUX
Directeur de Recherche CNRS, LTCI, Paris / Examinateur

Rémi COZOT
Maître de Conférences HDR, Université de Rennes, / Directeur de Thèse
This work has been achieved within the Institute of Research and Technology b<>com, dedicated to digital technologies. It has been funded by the French government through the National Research Agency (ANR) Investment referenced ANR-A0-AIRT-07.
Acknowledgements

I would like to express my deep and sincere gratitude to my supervisors- Rémi Cozot, Gérard Madec and Xavier Ducloux. This PhD would not have been possible without their help and the valuable time they provided amidst their busy schedules. I am indebted to them for their insightful discussions, ambitious challenges and critical feedback. It was an honor to work with some of the leading experts in HDR imaging and video engineering.

I am sincerely grateful to my thesis committee for accepting to evaluate this dissertation. Their assessment has significantly improved the quality of this manuscript and opened new avenues for future work. I would particularly like to convey my thanks to Kadi Bouatouch, who generously corrected and improved my publications during my PhD.

I am thankful to IRT b<>com for providing me a world-class research facility to pursue my doctorate. I am extremely grateful to b<>com’s management, especially my lab manager Jean-Yves Aubié, for believing in my research and industrializing our technology. I am truly blessed to have worked with such amazing peers and I thank them all for their constant encouragement and technical support that they offered to my work.

I am extremely grateful to Professor Touradj Ebrahimi for supervising my internship at the EPFL. It was a privilege for me to collaborate with an esteemed research group and expand my horizons on an international level.

Finally, I would like to thank my friends and my family for their love and support.
Abstract

This thesis presents various contributions in display retargeting under the vast field of High Dynamic Range (HDR) imaging. The motivation towards this work is the conviction that by preserving artistic intent and considering insights from human visual system leads to aesthetic, comfortable and efficient image retargeting.

To begin with, we present one of the main contributions of this thesis in retargeting of legacy content on to HDR displays. We propose a novel style-aware retargeting approach that considers image aesthetics and takes into account fundamental perceptual observations in viewing HDR content.

Furthermore, with the recent boom in HDR displays on the consumer market, a number of tonal inconsistencies are observed in displaying HDR content. We address this issue by investigating tone compatibility between HDR displays. A series of perceptual studies are presented considering content aesthetics and display characteristics.

In addition, we demonstrate the benefits of considering perception and propose a comfort-based retargeting for HDR displays. Based on subjective tests we develop a brightness preference computational model that mitigates brightness discomfort while retargeting for HDR displays.

Finally, we present an exploration into the benefits of HDR retargeting for omnidirectional imaging. We introduce a novel dual stimulus subjective methodology for Head Mounted Displays (HMDs) to conduct an evaluation of various HDR retargeting techniques. The experiment concludes that there is a slightly improved perceptual quality using HDR processing. The study also serves as a platform for next generation 360° HDR retargeting.
# Table of contents

## List of figures

- List of tables

## List of tables

- Nomenclature

## 1 Introduction

1. The Retargeting Process
2. Importance of Perception
3. Introducing Aesthetics in the pipeline
4. Goal and Overview of the Thesis
5. Contributions and Measurable Results
   - 1.5.1 Publications
   - 1.5.2 Industrial Contributions and Merits
7. Outline

## 2 Retargeting for High Dynamic Range Imaging

2.1 HDR Fundamentals
   - 2.1.1 Light in HDR imaging
   - 2.1.2 Color and Color Spaces
2.2 Human Visual System
   - 2.2.1 Photoreceptor Model
   - 2.2.2 Visual Sensitivity
   - 2.2.3 Color Appearance Models
2.3 HDR Capture and Display
2.4 Display Retargeting: Tone Mapping
   - 2.4.1 Global TMOs
   - 2.4.2 Local TMOs
   - 2.4.3 Tone Mapping for HDR video

2.4.4 Subjective Evaluations of TMOs ........................................ 23
2.5 Display Retargeting: Tone Expansion .................................. 24
  2.5.1 Global EOs .......................................................... 24
  2.5.2 Local EOs .......................................................... 28
  2.5.3 Subjective Evaluations of EOs .................................... 32
2.6 Aesthetic Considerations in HDR imaging .......................... 34
  2.6.1 A Brief Review of Image Aesthetics ............................. 35
  2.6.2 Tone Mapping as an Artform ..................................... 36
  2.6.3 Aesthetic Concerns in Tone Expansion ......................... 36
  2.6.4 Aesthetic Challenges in HDR production ....................... 37
2.7 HDR Ecosystem ......................................................... 38

3 Aesthetic HDR Retargeting ............................................... 41
  3.1 Background ........................................................... 42
    3.1.1 Aesthetics and its Impact on Tone Expansion ............... 42
    3.1.2 Color Appearance and Reproduction in HDR ............... 44
  3.2 Style Aware Tone Expansion ....................................... 45
    3.2.1 Lighting Style Classification and Source Content ........ 46
    3.2.2 User Study on Stylized Video ............................... 47
    3.2.3 Style Adaptive Gamma Correction ........................... 50
    3.2.4 Temporal Coherency in Tone Expansion .................... 51
  3.3 Color Correction for Tone Expansion ................................ 53
    3.3.1 User Study 3: Color Fidelity ................................ 53
  3.4 Tone Expansion Framework ......................................... 55
  3.5 Results ............................................................... 55
  3.6 Conclusion and Future Work ....................................... 60

4 Retargeting between HDR Displays .................................. 63
  4.1 Background ........................................................... 65
  4.2 Simple Retargeting Experiments ................................... 67
    4.2.1 Clipping for Tone Mapping ................................... 68
    4.2.2 Remapping for Tone Expansion .............................. 68
  4.3 Gamma as a Dynamic Range Converter ............................. 68
    4.3.1 Experiments .................................................... 69
    4.3.2 Results ........................................................ 70
  4.4 Conclusion and Future Works ..................................... 72
# Table of contents

5 Comfort-based HDR Retargeting ........................................ 75
   5.1 Background ............................................................ 76
      5.1.1 Perceptual Studies .............................................. 76
      5.1.2 Brightness Control .............................................. 77
   5.2 Subjective Evaluation ............................................... 78
      5.2.1 Stimuli and Pre-Processing .................................... 78
      5.2.2 Participants and Apparatus .................................... 79
      5.2.3 Procedure ........................................................ 80
   5.3 Results Analysis .................................................... 80
      5.3.1 Analysis of Experiments ....................................... 80
      5.3.2 Modeling Brightness Preference ............................... 81
   5.4 Conclusion and Future Works ...................................... 85

6 Exploring HDR Retargeting for Omnidirectional Imaging .......... 87
   6.1 Background ............................................................ 88
      6.1.1 HDR Acquisition ................................................ 88
      6.1.2 Tone Mapping Operators ....................................... 89
      6.1.3 Quality evaluation for HDR and omnidirectional imaging . 90
   6.2 Experiment Setup .................................................... 91
      6.2.1 Equipment ........................................................ 91
      6.2.2 Content Description ............................................. 92
      6.2.3 Content preparation ............................................ 93
      6.2.4 Content Selection ............................................... 94
   6.3 Methodology .......................................................... 97
      6.3.1 Pair Comparison approaches .................................. 97
      6.3.2 Toggling .......................................................... 98
      6.3.3 Experiment design .............................................. 99
   6.4 Results ................................................................. 101
      6.4.1 Subjective scores ............................................... 101
      6.4.2 Post-questionnaire answers ................................... 103
      6.4.3 Toggling locations analysis ................................... 105
   6.5 Conclusion and Future Works ...................................... 107

7 Conclusion ............................................................... 109
   7.1 Summary .............................................................. 110
   7.2 Open Issues and Perspectives ...................................... 111
References 113

Résumé en Français 129

Appendix A  Publications and Patents 135
   A.1 International Journal 135
   A.2 International Conferences 136
   A.3 Patents 138

Appendix B  Standardization Contributions 139
   B.1 Conversion from SDR/BT.709 to PQ10/HLG10 139
## List of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The retargeting process</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Perceptual illusions</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Artistic intent</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Limitations of standard imaging using single exposure</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>The electromagnetic spectrum</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Threshold Versus Intensity</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>Multi-camera bracketing</td>
<td>19</td>
</tr>
<tr>
<td>2.5</td>
<td>Examples of TMO’s</td>
<td>23</td>
</tr>
<tr>
<td>2.6</td>
<td>Linear EO</td>
<td>25</td>
</tr>
<tr>
<td>2.7</td>
<td>Dataset for Gamma Correction EO</td>
<td>27</td>
</tr>
<tr>
<td>2.8</td>
<td>Expand Map EO</td>
<td>30</td>
</tr>
<tr>
<td>2.9</td>
<td>Hallucination EO</td>
<td>31</td>
</tr>
<tr>
<td>2.10</td>
<td>‘The Brig’ by Gustave Le Gray</td>
<td>34</td>
</tr>
<tr>
<td>2.11</td>
<td>Examples of Image Aesthetics</td>
<td>35</td>
</tr>
<tr>
<td>2.12</td>
<td>Aesthetic TMO</td>
<td>37</td>
</tr>
<tr>
<td>2.13</td>
<td>Wider color gamuts for future displays on the CIE-1931 diagram</td>
<td>39</td>
</tr>
<tr>
<td>3.1</td>
<td>Lighting styles</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparing EOs with stylized content</td>
<td>44</td>
</tr>
<tr>
<td>3.3</td>
<td>A fuzzy c-means clustering of lighting styles of videos</td>
<td>46</td>
</tr>
<tr>
<td>3.4</td>
<td>Aesthetics in choice of gamma by lighting style</td>
<td>49</td>
</tr>
<tr>
<td>3.5</td>
<td>Fidelity in choice of gamma by lighting style</td>
<td>50</td>
</tr>
<tr>
<td>3.6</td>
<td>Adaptive gamma curve</td>
<td>51</td>
</tr>
<tr>
<td>3.7</td>
<td>Temporal Coherency in Tone Expansion</td>
<td>52</td>
</tr>
<tr>
<td>3.8</td>
<td>Fidelity in choice of saturation by style dependent γ value</td>
<td>54</td>
</tr>
<tr>
<td>3.9</td>
<td>File format conversions for Tone Expansion</td>
<td>56</td>
</tr>
<tr>
<td>3.10</td>
<td>Color Correction for Tone Expansion</td>
<td>57</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.11</td>
<td>MOS score evaluation of aesthetics of algorithms.</td>
<td>58</td>
</tr>
<tr>
<td>3.12</td>
<td>MOS score evaluation of fidelity of algorithms.</td>
<td>58</td>
</tr>
<tr>
<td>3.13</td>
<td>Comparing the results of various EOs</td>
<td>59</td>
</tr>
<tr>
<td>4.1</td>
<td>A Tone Reproduction Operator using a forward and reverse transform.</td>
<td>65</td>
</tr>
<tr>
<td>4.2</td>
<td>Clipping for Tone Mapping</td>
<td>67</td>
</tr>
<tr>
<td>4.3</td>
<td>Remapping for Tone Expansion</td>
<td>69</td>
</tr>
<tr>
<td>4.4</td>
<td>Box-and-Whiskers plots per lighting style</td>
<td>71</td>
</tr>
<tr>
<td>5.1</td>
<td>A subset of 9 screen content images used in the subjective experiments.</td>
<td>79</td>
</tr>
<tr>
<td>5.2</td>
<td>Boxplot diagram of brightness preference</td>
<td>82</td>
</tr>
<tr>
<td>5.3</td>
<td>Examples of the segmentation</td>
<td>83</td>
</tr>
<tr>
<td>5.4</td>
<td>HSV and CIE-Lab colorspace representations</td>
<td>84</td>
</tr>
<tr>
<td>5.5</td>
<td>Brightness Preference as a function of \texttt{BrightPixels}</td>
<td>85</td>
</tr>
<tr>
<td>6.1</td>
<td>SDR and HDR imaging workflow</td>
<td>89</td>
</tr>
<tr>
<td>6.2</td>
<td>Multi-exposure picture of the \texttt{Lake} content</td>
<td>93</td>
</tr>
<tr>
<td>6.3</td>
<td>Tone-mapped and exposure fusion pictures of the \texttt{Lake} content</td>
<td>93</td>
</tr>
<tr>
<td>6.4</td>
<td>Sample Images of HDR 360° dataset</td>
<td>95</td>
</tr>
<tr>
<td>6.5</td>
<td>Variation in key value of the 360° HDR content</td>
<td>97</td>
</tr>
<tr>
<td>6.6</td>
<td>Examples of side-by-side Pair-Comparison methodologies</td>
<td>98</td>
</tr>
<tr>
<td>6.7</td>
<td>MOSs and CIs analysis</td>
<td>102</td>
</tr>
<tr>
<td>6.8</td>
<td>Post-questionnaire analysis</td>
<td>103</td>
</tr>
<tr>
<td>6.9</td>
<td>Toggling locations across contents</td>
<td>106</td>
</tr>
<tr>
<td>B.1</td>
<td>Sample Live Workflow with Mixed Format Source Content</td>
<td>141</td>
</tr>
</tbody>
</table>
List of tables

2.1 Classification of EOs .................................................. 33
2.2 UHD Alliance Premium Specification .......................... 40
4.1 Tone Reproduction Methods ...................................... 64
4.2 Summary of the Experiments on Tone Compatibility between HDR Displays 72
5.1 Results of Visual Discomfort Questionnaire ............. 81
6.1 Description of various operators used in our experiment. 90
6.2 Characteristics and global statistics over the entire 360° image 96
Nomenclature

Acronyms / Abbreviations

ACR  Absolute Categorical Rating
AMC  Advance Media Coding
ANOVA  ANalysis Of VAriance
ARIB  Association of Radio Industries and Businesses
BEF  Brightness Enhancement Function
BK  Bright Key
CAM  Color Appearance Models
CI  Confidence Interval
CIE  Commission Internationale d'Eclairage
DCI  Digital Cinema Initiatives
DK  Dark Key
DR  Dynamic Range
DVB  Digital Video Broadcasting
DVE  Digital Video Effects
EBU  European Broadcasting Union
EO  Expansion Operator
EOTF  Electro-Optical Transfer Function
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPFL</td>
<td>École Polytechnique Fédérale de Lausanne</td>
</tr>
<tr>
<td>EV</td>
<td>Exposure Value</td>
</tr>
<tr>
<td>HD</td>
<td>High Definition</td>
</tr>
<tr>
<td>HDR</td>
<td>High Dynamic Range</td>
</tr>
<tr>
<td>HFR</td>
<td>High Frame Rate</td>
</tr>
<tr>
<td>HK</td>
<td>High Key</td>
</tr>
<tr>
<td>HLG</td>
<td>Hybrid Log-Gamma</td>
</tr>
<tr>
<td>HMD</td>
<td>Head Mounted Display</td>
</tr>
<tr>
<td>HSV</td>
<td>Hue Saturation Value</td>
</tr>
<tr>
<td>HVS</td>
<td>Human Visual System</td>
</tr>
<tr>
<td>IBC</td>
<td>International Broadcasting Convention</td>
</tr>
<tr>
<td>IBL</td>
<td>Image based Lighting</td>
</tr>
<tr>
<td>IEC</td>
<td>International Electrotechnical Commission</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>JPEG</td>
<td>Joint Photographic Experts Group</td>
</tr>
<tr>
<td>LCD</td>
<td>Liquid Crystal Display</td>
</tr>
<tr>
<td>LED</td>
<td>Light Emitting Diode</td>
</tr>
<tr>
<td>LFR</td>
<td>Low Frame Rate</td>
</tr>
<tr>
<td>LK</td>
<td>Low Key</td>
</tr>
<tr>
<td>MK</td>
<td>Medium Key</td>
</tr>
<tr>
<td>MMSPG</td>
<td>Multimedia Signal Processing Group</td>
</tr>
<tr>
<td>MOS</td>
<td>Mean Opinion Score</td>
</tr>
<tr>
<td>MPEG</td>
<td>Motion Picture Experts Group</td>
</tr>
<tr>
<td>NAB</td>
<td>National Association of Broadcasters</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>OETF</td>
<td>Opto-Electronic Transfer Function</td>
</tr>
<tr>
<td>OLED</td>
<td>Organic Light Emitting Diode</td>
</tr>
<tr>
<td>PC</td>
<td>Pair Comparison</td>
</tr>
<tr>
<td>PQ</td>
<td>Perceptual Quantizer</td>
</tr>
<tr>
<td>QoE</td>
<td>Quality of Experience</td>
</tr>
<tr>
<td>SAMVIQ</td>
<td>Subjective Assessment Methodology for Video Quality</td>
</tr>
<tr>
<td>SDR</td>
<td>Standard Dynamic Range</td>
</tr>
<tr>
<td>SI</td>
<td>Spatial Information</td>
</tr>
<tr>
<td>SMPTE</td>
<td>Society of Motion Picture &amp; Television Engineers</td>
</tr>
<tr>
<td>SS</td>
<td>Single Stimulus</td>
</tr>
<tr>
<td>STB</td>
<td>Set Top Box</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machines</td>
</tr>
<tr>
<td>TMO</td>
<td>Tone Mapping Operator</td>
</tr>
<tr>
<td>TV</td>
<td>Television</td>
</tr>
<tr>
<td>UHD</td>
<td>Ultra High Definition</td>
</tr>
<tr>
<td>VDP</td>
<td>Visual Difference Predictor</td>
</tr>
<tr>
<td>VSF</td>
<td>Video Service Forum</td>
</tr>
<tr>
<td>WCG</td>
<td>Wide Color Gamut</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

From the consumer television to the mobile phone to wearable displays, display technologies evolve over time. The increasing variety of consumer display devices requires adapting visual media to match the capabilities of these diverse display technologies. This process is known as display retargeting. With the explosion of new multimedia formats along with an abundance of existing content, the display retargeting problem is more relevant now than ever.

High Dynamic Range (HDR) imaging is the latest multimedia format that is being deployed in consumer market. With advances in acquisition and display technologies, the HDR imaging pipeline is able to capture, transmit, and display potentially the full range of light information of a scene. This characteristic makes it overcome several physical and perceptual limitations of Standard Dynamic Range (SDR) imaging systems. Over the last three decades, retargeting for HDR imaging has been extensively researched. HDR retargeting exists in two forms: 1) Tone Mapping for retargeting HDR content on SDR displays, and 2) Tone Expansion for retargeting SDR content on HDR displays. Although both retargeting methods seem reciprocal at first, on an algorithmic level both methods are far apart mostly due to differences in content characteristics, display capabilities and perceptual challenges, hence justifying the continuous quest amongst researchers to explore new retargeting methods in HDR.

This thesis presents contributions on different aspects of display retargeting under the large thematic of HDR imaging. Although the contributions are diverse, they are driven by the conviction that preserving artistic intent and considering perceptual insights from the human visual system are essential for aesthetic and visually comfortable display retargeting. This chapter provides a brief overview of the relevant topics and the contributions of this PhD thesis.
Fig. 1.1 The retargeting process in all its stages. Scene light is captured using a camera and stored in raw source format. This content is then rendered on a target monitor to generate the final mastered version. The targeting process could possibly require human assistance for providing artistic intent. Finally the mastered content could be visualized on a variety of displays thanks to in-built display retargeting mechanisms present in each display system shown.

1.1 The Retargeting Process

For most display technologies, retargeting is an inevitable step. This is mainly due to the fact that an image or video is adapted to a target display in terms of resolution, aspect ratio, color gamut, dynamic range, frame rate, etc. Retargeting is the additional processing required for the content to be viewed on a display that has even the slightest difference in characteristics than the target display [20]. Figure 1.1 outlines the retargeting process in all its stages. Retargeting is a large domain of work that cannot only be considered as a simple characteristic mapping from one display to the other. Some common retargeting techniques are briefly described below:

Spatial Retargeting

Spatial retargeting is required when a content needs to be resized for a particular display. Simple image resizing methods, scaling or cropping, do not always produce desired results, especially if they do not consider regions of interest of the image content. These methods are well defined in a survey by Vaquero et al. [208]. From computer graphics literature, we know
that non-homogeneous warping [215], seam carving [15][186] and the scale-and-stretch method [211] are popular spatial retargeting techniques. Super resolution algorithms are a special case of spatial retargeting dedicated to generating a high resolution image from one or more low resolution images. This field has also garnered a lot of research interest [162][73][220]. New spatial retargeting techniques are also being explored in omnidirectional imaging for mapping spherical content on view ports of head mounted displays [121][48].

**Temporal Retargeting**

Temporal Retargeting is the process of converting videos from low frame rate (LFR) to high frame rate (HFR) or vice versa. Blur reduction is a key issue for LFR to HFR conversion methods. Typical consumer displays double their frame rate by creating additional frames by using interpolation based on optical flow methods [114]. More advanced perception driven conversions are continuously being explored [52] [200].

**3D Retargeting**

3D retargeting has always been an exciting field of research and will continue to be so in the future. Many works have explored 2D to 3D retargeting in depth [109] [122]. Stereoscopic disparity retargeting methods have also been studied taking special precaution to avoid visual discomfort [137] [54]. A detailed description of retargeting a 3D scene for holographic displays can be found in Gilles’ PhD thesis [77].

**Luminance and Color Retargeting**

Retargeting luminance and color can be considered a science of its own when it comes to HDR imaging and is thus, at the heart of this thesis dissertation. A number of books on this subject have been published in the last few years [176] [21] [60] and have been referenced systematically during the course of this thesis. HDR retargeting will be discussed in greater detail in the next chapter.

### 1.2 Importance of Perception

Perceptually motivated retargeting methods require modeling the human visual system (HVS) in some form or the other. By exploiting the limitations of the HVS, perceptual retargeting can overcome the limitations of the display system. A number of methods have already been mentioned in previous examples. Wetzstein et al. [213] introduced the notion
of computational perception for display retargeting. The authors propose the following advantages of perceptually driven retargeting:

- Produces high visual quality beyond the limits of the display
- Increases visual comfort
- Reduces computation cost

Visual challenges are not only limited to the light emitted by the display device. Surrounding lighting conditions play a major role in perception of light and colors and vary significantly whether we are viewing the content in the cinema, at home or in a head mounted display. Hence, retargeting algorithms must also adapt with the changing viewing conditions for the best possible user experience. Figure 1.2 shows the importance of brightness and color perception and discrepancies between actual physical and perceptual reality. These illusions deceive our visual systems on standard displays and we can expect greater challenges on other display systems especially HDR monitors. A number of existing methods and new contributions dealing with perceptual HDR retargeting will be discussed in the coming chapters.

![Fig. 1.2 A few perceptual illusions are demonstrated. In (a) and (b) the tiles A and B have the same physical value. Both illusions highlight that context is important for luminance and color retargeting even if a light meter gives the same readings. Image (c) is an image that went viral on the Internet with people debating if the dress was white/gold or blue/black. Research suggests that people who prefer a cool illuminant perceive white/gold, while others who prefer a warm illuminant see blue/black [115]. Image courtesy: (a) [5], (b) [171] and (c) [115].](image-url)
1.3 Introducing Aesthetics in the pipeline

In philosophy it is common to tie aesthetics to perception [152] especially since the word aesthetics is derived from the greek word “aesthemi” which means “perception.” However, this definition has evolved over time along with our changing concepts of aesthetics. Italian philosopher, Paolo D’Angelo, argues that aesthetics can not only be appreciated by physical sensation but also as an experience of human imagination [45]. It is with this spirit that we would like to introduce aesthetics to the display retargeting pipeline.

The entire chain of the artistic work-flow can be seen in Figure 1.1 before the mastered content files have been generated. The choice of shooting locations and camera settings play a major role in creating intent. Furthermore, artificial light sources are often used to give dramatic or stylized feel to the scene. Finally, the content is adapted to a target display where the artist can add the final touches. The entire creative process is best explained by Virginia Wisslar in her book [214]. Once the final version has been mastered it is ready for deployment and will be retargeted by various display devices.

While there is a significant body of research focused on making images look more plausible on single target display, to the best of our knowledge, little to no attention has been paid to preserve artistic intent during retargeting. Unlike natural images, stylized content has very different characteristics and image statistics, thus requiring unique processing for retargeting. Furthermore, special effects are added to bring out a particular artistic intent as seen in Figure 1.3. Often, dedicated algorithms treat these stylized effects as artifacts and

![Film Grain](image1.png)  ![Blockiness as Art](image2.png)

**Fig. 1.3** We present two examples where artistic intent is sometimes considered as artifacts. In (a) we observe a film grain noise intentionally added to give the shot an old feel similar to analog film grain. The image in (b) is part of the collection “American Pixels” of the photographer Jörg Colberg. The photograph is the result of a modified jpeg-based compression algorithm designed to give artistic results instead of saving bits. The blockiness is part of the photographers intent. Image courtesy: (a) Mario Morales and (b) Jörg Colberg.
as a result the content loses its intended aesthetics. In this thesis, we aim to draw attention to artistic intent. We make an effort to go beyond the traditional content-aware retargeting approaches and introduce the notion of style-aware retargeting.

1.4 Goal and Overview of the Thesis

The goal of this thesis is to improve the retargeting process by combining our knowledge of human visual perception and image aesthetics for display retargeting. All our retargeting methods and experiments are within the realms of HDR imaging. For the remainder of this manuscript we will interchangeably use the terms display retargeting, HDR retargeting or simply retargeting all in the context of HDR imaging. We believe that there are several unsolved problems in HDR retargeting that can benefit from our holistic approach. Throughout this thesis we push further in this direction, searching for perceptually and aesthetically viable solutions to fundamental problems in retargeting for HDR imaging. A short overview of this thesis is listed below:

**Retargeting for High Dynamic Range Imaging**

An in depth literature review of HDR imaging is presented. The end-to-end HDR pipeline is considered including a background on fundamentals, the human visual system, acquisition and display technologies, tone mapping, tone expansion, aesthetics in content creation and the HDR broadcasting standards.

**Aesthetic HDR Retargeting**

We apply our knowledge on aesthetics and perception to a special case of retargeting in the form of SDR to HDR tone expansion. This study is done in the light of the fact that most existing video content today is in SDR format and there is a growing necessity to upscale this content for HDR displays. We demonstrate that current state-of-the-art tone expansion algorithms do not preserve artistic intent when dealing with content of various lighting style aesthetics. Furthermore, we present a series of subjective user studies evaluating user preference for various lighting styles as seen on HDR displays. This study shows that tone expansion of stylized content takes the form of gamma correction and we propose a novel retargeting method that adapts the gamma value to the intended style of the video. We also propose a simple color correction method that can be applied after tone expansion to emulate the intended colors in HDR. We validate our method through a perceptual evaluation against existing methods. In addition to this, this study targets 1000 nits HDR displays and we
1.4 Goal and Overview of the Thesis

present a framework aligning our method in conformance with existing SDR standards and the latest HDR TV standards.

Retargeting between HDR Displays

We investigate a study towards a generic retargeting framework between two HDR displays. With consumer displays with a peak brightness of 1000 nits and professional displays with a peak brightness of 4000 nits, retargeting between different HDR displays is an essential problem to be addressed. It is expected that dynamic range adaptation between HDR displays uses similar techniques as found with tone mapping and tone expansion operators. We survey simple retargeting techniques to map 1000 nits content to 4000 nits displays and 4000 nits content to 1000 nits displays. To challenge these retargeting methods we cherry pick stylized content that has been professionally mastered. We conclude that the best tone reproduction technique between HDR displays strongly depends on the lighting style of the content.

Comfort-based HDR Retargeting

We address the issue of brightness comfort for HDR displays. To retarget legacy content on HDR displays, a scaling is applied on the content to peak brightness of the display which leads to visual discomfort. To study this, we conduct a perceptual experiment for brightness preference considering the special case of screen content which has unique image statistics and has been created through some form of artistic intervention. The experiment concludes that brightness preference is highly content dependent. Using the data from the subjective study, we propose a new feature that analyses the source content to appropriately control the brightness of the HDR display. Based on this feature, we propose a brightness control retargeting algorithm that adjusts the luminance of the display depending on the number of bright pixels in the content.

Exploring HDR Retargeting for Omnidirectional Imaging

This study can be considered to be the first step in attempting HDR retargeting for virtual reality applications. We begin this exploration by assessing the quality of tone mapped 360° or omnidirectional HDR content. This study is an attempt to answer various questions in this direction. As a minimum, a new data set for 360° HDR content is proposed and a new methodology is designed to assess subjective quality of 360°HDR content on a head mounted display after retargeting with various tone mapping operators. The results are then analyzed and conclusions are drawn.
1.5 Contributions and Measurable Results

1.5.1 Publications

The main contributions of this thesis have already been published (including one international journal publication, five international conferences and two patents):

**International Journal**


**International Conferences**


- X. Ducloux, and C. Bist, “SDR to HDR Tone Expansion for Broadcast” in Broadcast Engineering and Information Technology Conference, 2017 [59].


**Patents**

• Bist C., Cozot R., Madec G. Procédé de traitement d’une image numérique, dispositif équipement terminal et programme d’ordinateur associés. 2015. French Patent No. 1562446

1.5.2 Industrial Contributions and Merits

The thesis resulted in several contributions in the industrial domain as well. These contributions along with some other merits are listed below:

**Industrial Contributions**

Our contribution on aesthetic retargeting is a technology that has been licensed to various actors in the industry. While developing this method we participated in various standardization activities. This included the Motion Picture Experts Group (MPEG), Video Service Forum (VSF), Digital Video Broadcasting (DVB) and the UHD Forum. Furthermore, we drafted the UHD Forum Phase A Guidelines: Section 8.2 Conversion from SDR/BT.709 to PQ10/HLG10. A number of technological demonstrations were presented in various trade-shows and events including National Association of Broadcasters (NAB) 2016 and 2017, International Broadcasting Convention (IBC) 2016 and the European Broadcasting Union (EBU) Production Technology Seminar 2017. Leading experts in video broadcast have evaluated our solution with other competitive industrial propositions and we have received encouraging positive feedback. On April 26 2017, IRT b<>com was awarded the prestigious NAB Technology Innovation award for “demonstration of SDR to HDR video up-conversion technology vital to the transition to next generation television broadcasting.” This technology also received positive reviews from the local and international press [85] [11].

**Research Visits**

One research visit was carried out during this PhD at the Multimedia Signal Processing Group (MMSPG) at the École Polytechnique Fédérale de Lausanne (EPFL) under the guidance of Professor Ebrahimi. This visit was funded mainly by IRT b<>com. Based on this, a successful collaboration started between the Advanced Media Coding (AMC) lab of IRT b<>com and the MMSPG lab of the EPFL. This partnership resulted in a successful publication in an international conference and also established plans for future projects between both organizations.
Awards and Recognition

- NAB 2017 Technology Innovation Award to IRT b<>com for SDR to HDR up conversion technology
- Le Doctorant, IRT b<>com’s best PhD student award 2016
- Amongst the top 5 papers at Graphics Interface 2016 for the work on aesthetic HDR retargeting (proposed for an extended version and submission to the Computers & Graphics journal, the extension was accepted)
- 2nd place for IRT b<>com’s best patent award 2015

1.6 Outline

This dissertation is organized as follows: Chapter 2 gives background information on the fundamentals of HDR imaging and retargeting. Chapter 3 proposes a style-aware retargeting method for SDR to HDR conversion. In Chapter 4 we extend our knowledge of aesthetics and perception for retargeting between HDR displays. Chapter 5 introduces the concepts of comfort-based retargeting for HDR displays. We explore HDR retargeting for virtual reality applications in Chapter 6. We conclude this dissertation and give an outlook for future work in Chapter 7. Appendix A lists the publications and patents with abstracts and Appendix B present our contribution to the UHD Forum.
Chapter 2

Retargeting for High Dynamic Range Imaging

Reproducing accurate visual appearance is the ultimate goals of retargeting in HDR imaging. Yet, the complexity of the human visual system and the perceptual attributes of visual appearance make the HDR retargeting an ongoing research problem. Along with knowledge in vision science, understanding the physical nature of light and color is key to HDR retargeting. Finally, the know-how of the advantages and limitations of engineering hardware such as camera and display systems are also fundamental to HDR retargeting.

In this chapter, we will cover the main elements of HDR retargeting related to this thesis. We begin with a brief introduction to HDR imaging detailing into fundamentals of light and color in Section 2.1. Followed by Section 2.2, where we discuss the human visual system and its role in HDR imaging. Section 2.3 looks into acquisition and display technologies for HDR. Tone Mapping and Tone Expansion, the main retargeting technologies in HDR are explained in Section 2.4 and 2.5 respectively. A discussion on aesthetic considerations for HDR imaging is followed in Section 2.6. Finally, we conclude the chapter with Section 2.7 giving an overview of current HDR ecosystem.

2.1 HDR Fundamentals

Real word luminance levels range from extremely low levels such as starlight to very ones as in the case of direct sunlight. The human visual system can easily adapt to these changes in luminance as it has complex adaptation mechanisms to process the incoming light from the scene. Traditional digital imaging technologies, notably SDR, fail to process the light information like the HVS. The source of the problem derived from limited dynamic range
Fig. 2.1 Images captured with increasing exposure time are shown from left to right. Single exposures having a limited dynamic range sensor cannot properly capture the entire light information of the scene and is far from emulating the HVS.

of both the camera sensor and the target display leading to clipping the signal in the bright and dark regions. An example of this is explained in figure 2.1. Using a short exposure time (left image) the camera captures details bright parts of the scene but under exposes the rest of the scene resulting in a dark image. The right image is captured using a long exposure time, detailing the dark areas but losing information in the bright regions. The image in the center has an optimized exposure time, in between short and long exposure times, and results in clipping the dark and bright areas of the scene. In all three cases, the resultant image is a compromise on the light information.

High Dynamic Range imaging techniques overcome the limitations of standard imaging systems in several ways. A key property of HDR imaging is to capture and store potentially the entire range of luminance levels of a scene. This is possible because HDR images are often stored in floating point precision. SDR images are typically stored in 8 bits per pixel format and eventually lose important light information. Furthermore, SDR images are mostly meant to be sent to a target display which must be calibrated for this purpose. This makes the SDR imagining pipeline display referred. This process takes the form of a gamma encoding, \( L = Y^{1/\gamma} \), where \( L \) is the encoded luminance (luma), \( Y \) is captured light luminance and \( \gamma \) value typically ranges from 1.8 to 2.6. The \( \gamma \) correction was initially applied to match the response curve for CRT monitors. Coincidently, this correction also conformed with our perception of brightness [169]. This \( \gamma \) correction has also found its way into digital display systems. It is often exploited by consumer display manufacturers to give more contrast resulting in tonal incompatibility between different displays which is a major inconvenience of display referred systems.

On the other hand, HDR imaging systems are scene referred. This means it contains the physical values of light of the scene making it an attractive imaging format for various applications including image based lighting (IBL), medical imaging, broadcast and multimedia etc. HDR imaging also explores representing light information on a display to achieve the closest reproduction to the HVS. The process of rendering HDR scene light to SDR
display system is known as tone mapping. It is one of the most in-depth studied forms of HDR display retargeting found in literature. The other field in HDR display retargeting is known as tone expansion. It deals with expanding the tonal capabilities of display referred SDR content onto HDR displays. In order to dwell deeper into these concepts, we introduce the basic concepts in visible light and color science which are essential to understand HDR imaging systems.

2.1.1 Light in HDR imaging

Light energy travels in space and interacts with materials in three possible ways: transmission, absorption, and reflection. When light interacts with the human eyes, it creates a perceptual sensation corresponding to a wavelength as seen in Figure 2.2. The human eye can only see a small part of the electromagnetic spectrum, from red to violet. The science of measuring the physical nature of light in any portion of the electromagnetic spectrum is known as radiometry. Irradiance and radiance are two radiometric quantities are used to quantify light in HDR imaging. Irradiance is defined as the quantity of light as a power per unit area (\(W m^{-2}\)) while radiance defined as the power emitted in a particular direction in steradian (\(W sr^{-1}m^{-2}\)). Photometry is the science of measuring visible light in units that are weighted according to the spectral responses of the standard observer defined by the Commission Internationale d’Eclairage (CIE)[203]. The photometric equivalent of irradiance and radiance

![Electromagnetic Spectrum](image_url)

Fig. 2.2 Visible light has an electromagnetic spectrum of 400 nm and 700 nm. Figure from wikimedia commons.
are known as illuminance and luminance. Illuminance is measured as luminous power per unit area also known as lux while luminance is the photometric quantity of light arriving at the human eye and is measured in \( cd \ m^{-2} \) or nits. For the remainder of this manuscript, we will use the unit nits to quantify luminance.

### 2.1.2 Color and Color Spaces

Colorimetry is the field of color science concerned with assigning code values of physically defined stimuli so that stimuli with the same code value look alike [217]. A color space is an abstract mathematical description for representing colors. Most color spaces are represented with three primary colors such as RGB or XYZ. Color spaces are designed considering attributes of the HVS and also the capabilities of the sensor or display technologies. The term gamut is used to define the complete range of colors for a capture or display device.

In color science there are two types of color spaces: device dependent and device independent [78]. The classic example of a device dependent color space is sRGB [10] used in most imaging systems. This is because for a given sRGB code value, for example RGB = [76 126 201], the perceived color can change significantly by simply changing the brightness or contrast of the display. Clearly, the sRGB color space is dependent on the display system being used, and is thus considered as a device dependent. A device independent color space is characterized by the code values used for a particular color will produce the same color regardless of where they are applied. Popular example of a device independent color space is the CIE Lab color space as it is based on the human visual system. Similarly, the HSV color space is another device independent color space used for fast computations and approximations to the HVS in computer graphics. Device independent color spaces are mostly used in applications that require imitating the HVS but cannot address a display directly. In this thesis, we will exploit both device dependent and independent color spaces.

In HDR imaging the RGB color space is commonly used. The XYZ color space is also popular in HDR but mainly for calculating operations on the Y channel as it closely models the perceived luminance. Fairchild and Chen [67] present two new color spaces: hdr-CIELAB and hdr-IPT based on physiological experiments. Although these color spaces show promise for HDR applications, the models require more testing and refinement. For HDR broadcast, Lu et al.[124] propose the ITP color space designed to preserve the HDR signal during video compression. Similarly, Poynton et al. [170] present the Y”u”v” 4:2:0 color encoding and subsampling scheme for HDR deployment. Apart from these, there has been few works in literature that attempt to model color spaces specific to HDR imaging making this an open area for research.
2.2 Human Visual System

Understanding human perception is very important in HDR imaging. Specific aspects and models of the HVS used in HDR retargeting algorithms play a major role in determining the quality of the final results. The human eye does not perceive light linearly the same way a camera captures light. This is because if the number of photons hitting the sensor of a camera is doubled, then twice the signal value is recorded. This is why we refer to light captured from the sensor as linear light. On the other hand, if we double the amount of light entering the human eye, it is observed only a fraction brighter. We refer to this non-linearity as the lightness perception of the HVS. We shall now discuss the major concepts involving our perception of light and color used in this thesis.

2.2.1 Photoreceptor Model

Photoreceptors are cells present in the retina of the human eye that are sensitive to light. These cells transmit light information through the optical nerve to the brain which interprets this signal to form an image. Over the years, researchers have attempted to model the photoreceptor to understand our perception of light. The photoreceptor response mathematically modeled by Naka and Rushton [151] is described as:

\[ \frac{R}{R_{\text{max}}} = \frac{I^n}{I^n + \sigma^n} \]

where \( R \) is the photoreceptor response, \( R_{\text{max}} \) is the maximum response, \( I \) is the light intensity, \( \sigma \) is the semi-saturation constant and \( n \) is the sensitivity control exponent. This response resembles to an S-shaped sigmoidal curve on a log-linear plot and is known as the Naka-Rushton equation. However, this model cannot cover all the wide range of luminance that the human eye can perceive. This is due to the fact that we cannot perceive the entire HVS range of \( 10^{-6} \) to \( 10^6 \) nits simultaneously and requires adaptation to the surround environment.

There are two types of photoreceptors that help in adapting to different lighting conditions: cones and rods [176]. Cones are concentrated in the center of the retina also known as fovea. They are responsible for our color vision, visual acuity and ability to capture rapid changes. Cones operate in bright light conditions (photopic vision ranging from \( 0.01 \) to \( 10^6 \) nits) and do not work in dark environments. Rods are the photoreceptors responsible for low light vision and are located outside fovea. They operate on dark environments (scotopic vision ranging from \( 10^{-6} \) to 10 nits). These photoreceptors also have no role in our color vision and are the reason we experience a lack of color at night. Cones and rods operate simultaneously in slight dark environments (mesopic vision ranging from \( 0.01 \) to 10 nits).
The range of light adaptation by the photoreceptors strongly influences our color perception. The change in color appearance when illumination changes to low illumination levels is known as the Purkinje effect [25]. On the other hand, the Hunt effect states that an increase in luminance level results in an increase in perceived colorfulness [88]. Several methods in HDR retargeting, as shown later on, incorporate such psychophysical phenomena in their algorithms.

### 2.2.2 Visual Sensitivity

In 1934, Weber conducted psychophysical experiments introducing the notion of perceived noticeable difference using hand held weights. The experiment showed that if we add 0.1 kg on to a person carrying a 1 kg weight, the person will perceive a difference in weight. However, if the person is carrying a weight of 10 kg, adding of 0.1 kg will be hardly noticeable. This relationship is known as Weber’s law. Experiments in perception extend upon this concept in perception and show that the smallest noticeable luminance difference, $\Delta L$, on a uniform background is linearly proportional to the background luminance $L$:

$$\frac{\Delta L}{L} = k$$

![Fig. 2.3 Threshold Versus Intensity (TVI) functions. Image courtesy Massof et al. [139].](image-url)
where $k$ is a constant called the Weber fraction and $\Delta L$ is called just noticeable difference (JND). Fechner [71] extend the work of Weber by claiming relationship between stimulus and perception is logarithmic. Their combined effort is known as the Weber–Fechner law. The plot of JND against luminance in loglog scale is called Threshold Versus Intensity (TVI) functions (see Figure 2.3). TVI has been furthered explored in works by Barten [26] and Blackwell et al.[35].

The JND is not only dependent on the background luminance level but also other important factors such as the adaptation level of the HVS, spatial frequency etc. This makes Weber–Fechner law difficult to model real world stimuli. Stevens and Stevens [199] proposed a perceptual model to relate the brightness perceived by test subjects to the actual luminance of the light. This study found that we perceive an increase in local perceptual brightness contrast due to the increase in luminance level. This is known as the Stevens effect. The experiment concluded by modeling brightness to the luminance using a power function, with an exponent value of $1/3$. The model later became known as the lightness response of the HVS, symbolized by $L^*$, and was standardized by the CIE in 1976. It is also the known as the lightness channel in the CIE Lab color space. This concept is fundamental not only to HDR but the imaging pipeline as a whole as it addresses the concept of perceptual uniformity in digital imaging. In fact, the various concepts explored in this section, notably Naka-Rushton’s sigmoid, Weber-Fechner’s logarithm and Stevens’ power function, are all estimates of perceptual uniformity of the HVS.

### 2.2.3 Color Appearance Models

In Sections 2.2.1 and 2.2.2 we described mostly the brightness and contrast perception of the HVS. We shall now focus on the color perception aspects of human color vision. The perceived color at each point of scene is dependent on the relationship between light signals from that point and it surrounding environment. Color Appearance Models (CAM) are used to model this relationship. CAMs are defined by Sharma et al. [192] as “any model with some form of a chromatic adaptation transform and includes predictors of at least the relative color-appearance attributes of lightness, chroma, and hue.” To this effect, these models should be able to model contrast and brightness dependent phenomena such as the Stevens effect or the Hunt effect. Given this definition, the CIE Lab color space could also be considered as a CAM.

The main goal of CAMs is to predict the appearance of a given image under different viewing conditions taking into account information about the image, the original scene, and the viewing environment [66]. Most CAMs are designed using psychophysical experiments with simple stimuli. Spatially localized adaptation is considered in models such as s-CIE
2.3 HDR Capture and Display

Methods for creating HDR content is being explored since the early 1990s. These methods include Ward’s physically based rendering engines [212] to create synthetic HDR scenes and reconstruction from multiple exposure bracketing[126] [51] for real word capture. Alternatives to exposure bracketing includes work by Hristova et al[86] who propose a method for creating HDR images from only two images — flash and non-flash images. Transition from static to video has been mostly straight forward for rendering engines to generate HDR computer graphics, but has been a challenge for camera captured HDR video capture. Tocci et al.[201] demonstrated a multi-sensor bracketing using a beam-splitter to divide the light, which resulted in capturing different exposures with multiple sensors embedded. Various methods have proposed HDR capture using SDR video cameras by alternating exposure times between frames. Kang et al. [102] and Mangiat et al. [125] both proposed an HDR video reconstruction method using optical flow while Kalantari et al.[101] proposed an algorithm using patch-based reconstruction. Gryaditskaya et al. [80] presented another patch-based method using a new motion-aware bracketing scheme. Significant advancements were made in computational photography during last decade for native in a single camera. Froehlich et al. [75] employ a multi-camera bracketing which is described in Figure 2.4. This method combines two cameras with a rig and different neutral gray filters to generate the different exposures. These methods are able to generate 18-24 f-stops of dynamic range. Recent improvements in sensor capabilities have also seen the rise of professional HDR video cameras such as Arri Alexa XT and the Sony F-65 ranging up to 14-16 f-stops.

HDR displays have a greater range of luminance and colors than an SDR monitors. One of the first HDR TV technologies was showcased by Seetzen et al. [190] by combining two light modulation devices: a modulated back-light with an LCD panel. The method was then used by a university spin-off, SunnyBrook Technologies (later known as BrightSide), which made displays such as the DR37-P with a peak luminance of around 4000 nits. Early works in HDR imaging often cite this display. The dual modulation method was further
incorporated into what is now known as the Dolby Pulsar which is also has a peak luminance up to 4000 nits. There are only about 40 of Pulsar displays in the world mostly being used by post production facilities in the cinema industry [55]. Another display manufacturer, SIM2, also uses dual modulation technology in their HDR47 series [193] and has recently released a prototype HDR display that peaks up to 10,000 nits [209]. Although dual modulation LCD displays have the capability to reach high brightness levels they often don’t perform to well when displaying dark content. OLED technology was predicted to have significant potential for HDR displays in [138], especially for its ability to shows details in the deeper blacks. An example of this is the Sony BVM-X300 OLED display [198] which is currently the reference display in the industry with peak brightness of 1000 nits with a contrast ratio of a million-to-one. HDR LCD’s having brighter whites and HDR OLED’s providing deeper blacks makes it a difficult choice on HDR display to pick for content providers, consumers and researchers in display retargeting.

In comparison to these prototype HDR displays, a SDR TV has a peak luminance of 200-300 nits, with rare high-end systems peaking up to 500 nits. With recent advances in LCD and OLED display technologies and strong interest from consumer display manufacturers, a large number of HDR TV sets have flooded the consumer market during 2015-2017. However, these display range from 500 to slightly more than 1000 nits in peak brightness and are still far from the capability of the Dolby Pulsar, Sim2 HDR47 and Sony BVM-X300 displays. The arrival of these new HDR consumer and professional displays (LCD/OLED) and the wide abundance of legacy SDR displays, poses a new challenge to the budding HDR ecosystem and requires imminent attention for HDR display retargeting.
2.4 Display Retargeting: Tone Mapping

Tone mapping is described as the transformation of an HDR image to a display device with limited dynamic range. In most cases, tone mapping is a scene referred process involving converting the linear scene referred luminance on to a display. Scene referred tone mapping can be implemented for both HDR and SDR displays as both are incapable of reproducing actual scene light values. Tone mapping can also take the form of display referred retargeting when mapping content from an HDR display to an SDR display. Most works in literature focus of the extreme case of dynamic range compression i.e. mapping linear scene referred HDR signal on to typical displays with limited dynamic range. This will also be the focus of this section on tone mapping.

The process of tone mapping is executed by a tone mapping operator (TMO). The development of tone mapping operators has been an active field of research in HDR retargeting for many years. Simple TMOs are based on operations such as, scaling, clipping or a gamma correction. While more complex TMOs take into account the characteristics of the scene or properties of the human visual system in order to reproduce the best visual experience on the SDR display. TMOs are broadly classified into global and local operators. Global operators apply a single mapping function to every pixel in the image while local operators apply unique mapping per pixel in an image depending upon its neighboring pixels. In this section, we give a short review of the vast field of tone mapping.

2.4.1 Global TMOs

The simplest global TMO used to display HDR images is by linear scaling, followed by clipping and a final gamma encoding. However, as we already know, the HVS has a non-linear response and scaling linear light will not be seen in a perceptually uniform manner. This gives rise to non-linear mapping of luminance. This is done by applying an exponential function in the linear domain or a scaling in the logarithmic domain. Tumblin and Rushmeier [204] were first to employ exponential mapping based on the Stevens’ power function. Drago et al. [57] used a similar concept for adaptive logarithmic compression of luminance values, imitating the HVS using Weber-Fechner’s logarithm. Schlick [188] argued the use of a sigmoid tone curve (similar to the Naka-Rushton equation) claiming that they are computationally less expensive than exponential or logarithmic functions. The sigmoid function was also used by Pattanaik et al.[163] on the premise of the photoreceptor model.

Global tone curves can also be formulated by using histograms. Ward et al. [117] were first to use this method. This method modifies classical histogram equalization and addresses several aspects of the HVS such as glare, loss of acuity, and color sensitivity. Histogram-
based tone mapping has evolved over the years to address different issues such as minimizing contrast artifacts in tone mapping [129], noise-aware tone mapping [63] and clustering for tone mapping [160].

In general, Global TMOs are computationally inexpensive and rarely introduce artifacts, however they usually but tend to produce lower contrast images and often lose details in very dark and very bright areas.

### 2.4.2 Local TMOs

The most basic local TMO could be considered the Gaussian filter. It seems to be an ideal way to characterize an HDR image’s illuminance, however it is prone to produce visible halo artifacts because the filter will be influenced by typical large gradient edges found in HDR images. Gaussian filters also were also used in [172] and [99] seeking inspiration from the retinex theory used in perception [116].

To overcome the limitations of the Gaussian filter Pattanaik et al. [163] use a Laplacian pyramid. The pyramidal filtering approach for local tone mapping was further explored by Ashikhmin et al. [13] based on an iterative Gaussian filter. However, halo artifacts continue to be an issue along with computational cost with such methods. The Gaussian pyramid is also utilized by Reinhard et al.[180] in the popular Photographic TMO. This method is inspired by photographic practices of dodging and burning [4]. This TMO consists of two steps: 1) applying a global tone curve based on the sigmoid function 2) using a Gaussian kernel to enhance local contrast. The idea of applying a global non-linear tone curve followed by a local step for contrast enhancement is also shown by Ferradans et al. [72].

The bilateral filter was introduced to HDR tone mapping by Durand and Dorsey [61]. The bilateral filter can separate the HDR image into a detail and a base layer. The detail layer consists of a high frequency image while the base layer is a low frequency image with preserved edges. The base layer is compressed with the detail layer added later to produce the final image. The main disadvantage of the bilateral filter is it computational complexity. Furthermore, this approach may also result in exaggerated edges in high gradient areas and might also over-smooth ramp edges.

Overall, local TMOs produce good images if implemented with care. However, they are computationally more expensive than global TMOs and often produce artifacts around edges such as halos.
2.4.3 Tone Mapping for HDR video

A large majority of TMOs in literature operate on still images. Applying them directly to video frame-by-frame basis could lead to temporal incoherency, flickering and other artifacts. In addition to this, a large number of frames are required to be processed requiring computational efficiency along with the knowledge of HDR capturing devices.

Global tone mapping algorithms are the obvious choice for HDR video tone mapping. This is mainly due to their simplicity, low computational cost, and ease of applying temporal coherency. A method of using a temporal filter was proposed by Kiser et al. [108] by averaging the TMOs parameter value of the previous frame with the current frame. Similarly, Ramsey et al. [174] adapted parameter values from multiple previous frames. A two-pass procedure was suggested in [129]. This method involved a first pass where the tone curve for each frame was calculated following a second pass where the tone curve was adjusted in a temporally coherent manner to be mapped on the HDR video.

Adapting local TMOs are not so trivial for videos since they are spatially varying. Aydin et al. [18] successfully show a spatio-temporal filter utilizing the optical flow. The method used is based on the bilateral filter. Another HDR video TMO that employs the bilateral filter is proposed by Eilertsen et al. [64]. This method applies a noise-aware tone curve on the base layer similar to the one in [129] and separates the detail layer using a novel fast edge-stopping bilateral filter designed to mitigate ringing artifacts that result from such processing.

Tone mapping HDR video could also take the form of offline post processing. One of these methods is the brightness coherence approach [37] which applies a global TMO to given video sequence. The temporal coherency is dealt with by adapting all the video frames to the anchor frame by a scaling operation. The anchor frame is automatically chosen and is effectively the best exposed frame in the video. Although this method is simple it is prone to local fluctuations. The zonal brightness coherency method [38] improves upon this by segmenting the video frames in brightness zones and processing them separately. Both these methods and the state-of-the-art in video tone mapping is well documented in Boitard’s PhD thesis [36].

The current boom in HDR video has seen HDR video tone mapping find place in multimedia applications. Video broadcasting applications have embraced global TMOs mostly for their simplicity and easy integration into existing hardware. Works by Cyriac et al. [43] employing a perceptually adaptive gamma curve and Lenzen et al. [123] using a sigmoidal tone curve, solve use-cases at different part of the broadcasting pipeline. Although local TMOs, such as [18], have found their way in post-production houses in the cinema
industry, most tone mapping for Hollywood films is preferred to be done manually by a colorist.

### 2.4.4 Subjective Evaluations of TMOs

Several studies in the past have evaluated the visual quality of tone mapping operators. Ledda et al. [120] conducted subjective evaluation of tone mapping operators with a reference HDR display while Yoshida et al. [222] presented a study evaluating tone mapped images with real-world scenes. Kuang et al. [111] did a pairwise comparison experiment between TMOs on a single SDR display. Evaluation of tone mapping for HDR video has also been well covered by Eilertsen et al. [65] [62] and by Melo et al. [142]. Despite the vast study done on this topic, the underlying conclusion amongst existing works is that preference of TMO is very subjective. An example is shown in Figure 2.5.

![Different TMOs give different results.](image)

**Fig. 2.5** Different TMOs give different results. The preference of TMO strongly depends on the aesthetic preference of the observer. The TMO’s shown (from left to right and top to bottom) are: [134] [57] [61] [70] [164] [180]. Image from Debevec and Malik[51].
2.5 Display Retargeting: Tone Expansion

Tone expansion is the process of transforming an SDR image into an HDR image. This conversion problem is often called inverse or reverse tone mapping but we use the term tone expansion in this manuscript as defined by [21]. Expansion operators (EOs) are used to process the SDR content in the best possible way so as to be displayed on an HDR monitor, in other words requiring display referred processing. There are a few methods that attempt to extract scene referred data of the actual HDR radiance values from the SDR image. As a result, they attempt to solve a typical ill-posed inverse problem where the additional dynamic range available in HDR displays is missing in SDR format. As mentioned earlier, this information is lost during acquisition when using an SDR camera in the form of sensor saturation. Most EOs attempt to solve this problem on strong a priori assumptions that lead to positive or negative results based on specific conditions. Just like TMOs, EOs are also classified into global and local operators. In this section, we overview some of the main expansion operators.

2.5.1 Global EOs

Similar to tone mapping, global expansion operators apply the same transformation function to every pixel of the SDR image. The two main global operators were proposed by Akyuz et al. [7] and Masia et al. [135]. Akyuz et al. [7] have presented a series of psychophysical experiments to conclude that a trivial linear operation is sufficient to provide good HDR tone expansion. This method works well in optimum conditions such as for uncompressed well-exposed images. For the special case of over-exposed content, Masia et al. [135] have found that a nonlinear tone expansion, in the form of an adaptive gamma (\( \gamma \)) correction, is sufficient for HDR displays. In real world scenarios, particularly in the case of live broadcasting, these global methods are practical for real time applications. We develop them in detail below:

Linear Scaling EO

In [7], Akyuz et al. explored the following question: Do HDR displays support SDR content? To answer this, the authors conducted a psychophysical experiment to investigate the expansion of single exposure images in comparison to the HDR images on the DR37-P HDR monitor. In order expand the single exposure image, they employed the following equation:

\[
Y_{HDR} = L_{max} \times \left( \frac{Y_{SDR} - Y_{SDR, min}}{Y_{SDR,max} - Y_{SDR, min}} \right)^\gamma
\]  

(2.3)
where, $Y_{HDR}$ is the HDR luminance, $Y_{SDR}$ is the SDR luminance, $L_{max}$ is the peak brightness of the target display, $Y_{SDR,max}$ and $Y_{SDR,min}$ are the maximum and minimum values of the SDR luminance, and $\gamma$ was used as a non-linear scaling term having the values 0.45, 1, and 2.2. The gamma curves are shown in figure 2.6. The subjects were required to rank the images which were looking best to them. The results showed that brighter images with higher mean luminance were preferred. This was true when expanded single exposure image had higher mean luminance than the HDR image and vice-versa. The authors concluded that higher mean luminance is preferable over contrast when viewing any content on an HDR display. The result also concluded that linear scaling, $\gamma = 1$, was the most favored setting for expansion. This method works well in ideal conditions such as for uncompressed, high resolution and artistically captured well-exposed images. However, it does raise concern when dealing with compressed or noisy content. Furthermore, having a high mean luminance may not always preserve artistic intent or display a visually comfortable image. We will discuss some of these issues in later chapters.

![Gamma curves](image)

Fig. 2.6 Three $\gamma$ curves used for expanding the dynamic range of an SDR image by Akyuz et al. [7]. We can observe that when $\gamma > 1$ the image becomes darker and when $\gamma < 1$ the image becomes lighter. Image courtesy of Banterle et al. [21].
Gamma Correction EO

Masia et al. [135] also proposes an EO through psychophysical experiments. Unlike Akyuz et al. [7], the authors considered a variety of exposure settings in order to find the expansion $\gamma$. The first experiment compares the performance of three EOs on expanding SDR images with increasing exposures. The EOs considered are the linear EO and two local EOs by Banterle et al. [23] and Rempel et al. [184], which will be described later. The results of this experiment show that the selected EOs decreased in performance as exposure of the source content increased. The authors observed that spatial artifacts that were part of the source SDR images re-appeared in the HDR display after expansion. They concluded that the perceived quality of HDR images depends more on the presence of disturbing spatial artifacts rather than the exact light intensities. Masia et al. [135] also notice that for over exposed content a large portion of the image has been saturated to white. Based on this, they propose an expansion method which preserves the remaining details while not boosting the saturated regions.

The simplest way to implement this is by a $\gamma$ correction. The authors conducted a pilot study asking users to manual adjust the $\gamma$ value for each image in the dataset shown in figure 2.7. The chosen $\gamma$ values were fitted to the key value of the image. The key value is a popular metric in HDR imaging to indicate the average brightness of a scene [181] as seen in the equation below:

$$
Key\ value = \frac{\log Y_{avg} - \log Y_{min}}{\log Y_{max} - \log Y_{min}}
$$

(2.4)

where $\log Y_{avg}$ is the log geometric mean of luminance of the image, $Y_{max}$ and $Y_{min}$ are the maximum and minimum scene referred luminance values clipped at 0.1 and 99.9th percentile. $\log Y_{avg}$ is calculated using the equation:

$$
\log Y_{avg} = \exp\left(\frac{1}{N} \sum_{n=1}^{N} \log (\delta + Y_{HDR})\right)
$$

(2.5)

where $\delta$ is a small value to avoid singularity, $Y_{HDR}$ is the HDR scene referred luminance and $N$ is the number of pixels in the image. The $\gamma$ and the key value, $k$, was empirically fitted using linear regression as follows:

$$
\gamma = 10.44 \times k - 6.282
$$

(2.6)

One of the major drawbacks of this regression is that it was designed using a small dataset and mostly considered over exposed content with a high key value. For dark content (with a low key value), this regression resulted in a negative $\gamma$ value which results in unnatural images.
In addition to this, the data collected by the authors suggests that the $\gamma$ value increases as the image gets brighter. From Figure 2.6, we learn that increasing the $\gamma$ value makes the image darker. This may also account for the unnatural results. In 2017, Masia et al. [136] revised their analysis and proposed an alternative robust regression:

$$\gamma = 2.4379 + 0.2319 \times \log Y_{\text{avg}} - 1.1228 \times k + 0.0085 \times p_{\text{ov}}$$  \hspace{1cm} (2.7)$$

where $\log Y_{\text{avg}}$ is the log geometric mean of luminance of the image, $k$ is the key value and $p_{\text{ov}}$ is the percentage of over exposed pixels (number of pixels having code value greater than 254). This new $\gamma$ expansion avoids negative $\gamma$ values that resulted from Equation 2.6. However, this regression also models the same data presented in the first publication [135]. Hence, it also implies that $\gamma$ value will increase for brighter images and thus result in unnatural HDR images. We shall revisit some of these issues in Chapter 3.
2.5.2 Local EOs

Local EOs are different from local TMOs as they are normally based on a form of image segmentation. These EOs segment the SDR image into different regions (such as diffuse and specular highlights) and expands the segments accordingly. Meylan et al.’s [145] work detects specular highlights and diffuse parts of the image and applies a separate linear tone curve to both regions. Didyk et al. [53] also use EOs based on segmentation but require manual assistance. Wang et al. [210] proposed another EO that requires user input to add missing detail into areas that have been over or under exposed. In a practical system, an automated approach would be more appropriate. Hence, Kuo et al. [112] builds upon this hallucination scheme by proposing an automated method. The expand map model is another approach that has been investigated in various works [184] [23] [24] [113]. In these techniques, the expansion process is guided by an expansion map which gives the magnitude of expansion per pixel. Finally, the last category of local EOs is based on modeling the HVS without the need of segmentation or expands maps. This was proposed by Huo et al. [89] where the authors expands SDR pixels by inverting the local retina response based on the local contrast perception of the HVS. These EOs are discussed below.

Segmentation EOs

The first segmentation EO was proposed by Meylan et al. [145]. This method is focused on the representation of highlights found in SDR images on HDR monitors. This is EO segments the diffuse and the specular region in the image and linearly expands the two regions separately. The segmentation is done by finding a threshold parameter $\omega$ of diffuse luminance value by implementing a low pass filter on the luminance channel. Once $\omega$ is calculated the piecewise expansion take the following form:

$$Y_{HDR} = \begin{cases} s_1 Y_{SDR} & \text{if } Y_{SDR} \leq \omega \\ s_1 \omega + s_2 (Y_{SDR} - \omega) & \text{otherwise} \end{cases}$$

where $s_1 = \frac{\rho \omega}{\omega}$ and $s_2 = \frac{1-\rho}{1-\omega}$

where $\rho$ is the percentage of the diffuse part of the image which was allocated to the HDR luminance. The authors conducted a subjective study for determining the optimal $\rho$ value using the DR37-P HDR monitor. The study found that users allotted more dynamic range to the diffuse signal for outdoor scenes and vice-versa for indoor images. The authors finally concluded that $\rho = 66\%$ is an optimal compromise using this method. This EO works well to reproduce specular highlights on HDR screens, but in certain cases prone to quantization artifacts around the enhanced highlights.
2.5 Display Retargeting: Tone Expansion

Didyk et al. [53] extend the segmentation EO to video. The main concept behind this EO is to segment each frame to three classes: diffuse, reflections, and light sources, and then to preserve the diffuse regions while enhancing the reflections and light sources. The method begins with several pre-processing steps such as calculating optical flow, detecting over exposed pixels and a number of geometrical and statistical features of each frame. Consequently, the method relies on a training set of 2000 manually classified regions to classify the three segments. In the case, where the algorithm wrongly classifies a region the method requires user input to correct the error. For the brightness enhancement the EO applies a bilateral filter separating the base and detail layer, applying a non-linear adaptive tone curve based on histogram equalization to the base layer. The final HDR image is generated by combining the detail and expanded base layer. The method is considered semi-automatic and hence not suitable for real-time video retargeting applications especially since user input is required.

**Expand Map EOs**

Expand Map EOs were introduced by Banterle and a detailed description can be found in his thesis [19]. Expand map EOs generally follow a three step process: 1) using a global expansion curve to expand the image 2) generating an expand map to enhance highlights and mitigate artifacts and 3) combining the expanded image and expand map to generate the final HDR image. Banterle et al. [23] propose a framework involving linearizing the SDR image followed by using a non-linear expansion curve which inverts the Reinhard’s photographic TMO using Equation 2.9:

\[
Y_{HDR} = \frac{1}{2} L_{Max} L_{White} \left( Y_{SDR} - 1 + \sqrt{(1 - Y_{SDR})^2 + \frac{4}{L_{White}^2} Y_{SDR}} \right) 
\]  

(2.9)

where \( L_{Max} \) is the peak brightness of the display in nits and \( L_{White} \) is contrast varying parameter. Once the expansion is applied, the authors then propose to generate the expand map. The expand map is calculated using a density estimation using median-cut sampling [49]. The expand map and expanded image are combined using linear interpolation. This process is explained in figure 2.8. The authors also provide methods to extend this EO for video.

Rempel et al. [184] also exploit the expand map concept for their EO. This method begins with a filtering to remove compression artifacts followed by a linearizing the SDR image. A linear global operator is applied to the image to match the display capabilities of the DR37-P HDR monitor. A Brightness Enhancement Function (BEF) which acts as the expand map is applied to the saturated regions. This BEF starts with thresholding the linear
Retargeting for High Dynamic Range Imaging

Inverted Reinhard TMO
Expand Map
Linear SDR image
Linearly interpolated HDR image
Expand Map
Expanded Image

Fig. 2.8 An example of an Expand map EO proposed by Banterle et al. [23] [24]. Figure from [19].

SDR image at a value of 0.92. This is followed by applying a Gaussian filter and an edge stopping function to increase contrast around the edges. Once the BEF is constructed it is combined with the linearly expanded HDR image resulting in the final HDR image. The most recent expand map-based EO was proposed by Kuo et al. [113]. In this method, the expansion is implemented by inverting Schlick’s sigmoidal TMO. The tone expansion curve adapts to different contents with the help of Support Vector Machines (SVM) scene classifier. The expand map is calculated using an erosion operation followed by bilateral filtering.

All three Expand Map EOs presented in this section offer innovative approaches to tone expansion. A common trait amongst these publications is that they conclude their study by comparing their method using an HDR objective metric from the HDR - Visual Difference Predictor (VDP) [130] family. Over the years, HDR objective metrics have evolved. More recent objectives include HDR-VDP 2 by Mantiuk et al. [131], HDR VDP 2.2 by Narwaria et al. [157] and HDR Visual Quality Metric (VQM) also proposed by Narwaria et al. [153]. However, there is a general consensus that the human visual system remains the gold standard for qualitative evaluation for HDR imaging. We will examine this is in Section 2.5.3.

Hallucination EOs

Wang et al. [210] proposed a method to recover lost information in SDR images by in-painting and coined it as hallucination. The authors demonstrate that by using the information of neighboring pixels, it is possible to fill in under and over exposed regions followed by luminance boosting to create the final HDR image. After linearizing this method, the image is decomposed into a base layer and detail layer using a bilateral filter. The over exposed regions of the base layer are in-painted using the neighboring pixels by using the method
2.5 Display Retargeting: Tone Expansion

Fig. 2.9 An example of a Hallucination EO proposed by Wang et al. [210]. The missing details (left image) can be reconstructed from the neighboring pixels resulting in the hallucinated image (right). Figure from [210].

in [79] which involves linear interpolation by using an elliptical Gaussian kernel. This step could benefit from user intervention as it’s similar to the brush stroke tool in Adobe Photoshop [6]. The detail layer is hallucinated using intervention using the in-painting method presented in [29]. The method draws similarities to the stamp and healing tools in Adobe Photoshop [6] which involves using a stroke-based interface needed to indicate the source region and target region, for pixel transfer. An example of Hallucinations using this method is seen in figure 2.9. We can observe that this method gives good results, however the user interaction factor makes it difficult to extend to video.

In [112], Kuo et al. extend their expand map method described in [113] by adding the final step to permit hallucination. Once the expand map is generated, the authors employ the well-known GrabCut method [185] to segregate the overexposed part of the image in the expand map. The automatic in-painting method used in this work combines several exiting techniques including the exemplar-based in-painting algorithm [42] and multi-level in-painting [219]. The in-painting is applied on to the expand map which is then combined with the expanded image in order to create the final HDR result. This method shows promising visual results, however an extension to video may result in temporal incoherency due to the nature of the in-painting. Furthermore, a case where the algorithm fails is not discussed. Automatic in painting could pose an issue for images that have intentionally over exposed regions or images with white text.
HVS-based EOs

The last class of local EO consists of the method proposed by Huo et al. [89]. This EO proposes to expand the SDR pixels by inverting the local retina response using the equation below:

\[
Y_{\text{HDR}} = \left( \frac{Y_{\text{SDR}} \times \sigma^n}{Y_{\text{SDR,max}} - Y_{\text{SDR}} + \delta} \right)^{1/n}
\]  

where \(n\) is defined as constant value set by the authors to 0.9 and \(\sigma\) is the output of the bilateral filter applied to \(Y_{\text{SDR}}\). The method is validated using the objective metric based in an earlier version of HDR-VDP [16]. The EO is light weight in terms of computational efficiency but a more reliable qualitative evaluation is required to compare it with existing technologies.

Classifications of EOs

From the discussion above we can classify these state-of-the-art EOs as seen in Table 2.1. The classification of the EOs is based on the suggestions made by the authors in their respective publications. Some of the EOs do not directly mention an extension to video, so we have placed * symbol do denote an application to video tone expansion seems plausible. It should be noted that EOs by Akyuz et al. [7], Masia et al. [135] and Meylan et al. [145] are the only ones to have conducted a perceptual user study to develop their EO. The remaining EOs validate their expansion methods using objective metrics, many of which are outdated. With a growing number of EOs and the limited scope of existing objective metrics there is a strong requirement to conduct rigorous subjective evaluation of EOs. This is described in the next section.

2.5.3 Subjective Evaluations of EOs

There have been three major publications on the subjective evaluation of EOs. The first psychophysical evaluation was presented by Banterle et al. [22]. The selected EOs were Akyuz et al. [7], Meylan et al. [145], Banterle et al. [23], Rempel et al. [184] and Wang et al. [210]. The SDR images were generated by clipping a selection of eight HDR images. The evaluation methodology required subjects to choose from a pair of expanded HDR images while comparing to a reference HDR image. The display used was the DR37-P HDR monitor. The results of the study suggested that more complex EOs such as Banterle et al. [23], Rempel et al. [184] and Wang et al. [210] outperform simpler EOs i.e. Akyuz et al. [7] and Meylan et al. [145]. Amongst the better performing EOs, the authors suggest ones using
Table 2.1 Classification of EOs

<table>
<thead>
<tr>
<th>EO</th>
<th>Class</th>
<th>Expansion</th>
<th>Automatic</th>
<th>Video</th>
<th>Hallucination</th>
<th>User study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akyuz et al. [7]</td>
<td>⊙</td>
<td>−</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Masia et al. [135]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Meylan et al. [145]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Banterle et al. [23]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Rempel et al. [184]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Didyk et al. [53]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Wang et al. [210]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Huo et al. [89]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Kuo et al. [113]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Kuo et al. [112]</td>
<td>⊙</td>
<td>∼</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>

✓ = yes, × = no, ⊙ = global, ⊖ = local, − = linear, ∼ = non-linear, ⋆ = potential

A non-linear expansion (Banterle et al. [23] and Wang et al. [210]) supersedes the linear expansion methods (Rempel et al. [184]). The work also advocates that computationally expensive methods are better for tone expansion as they correctly reproduce the visual attributes of the original HDR image.

However, the study by Banterle et al. [22] only considers still images. Another study, conducted by De Simone et al. [47] evaluates EOs for video applications. The EOs studied were chosen to represent different tone expansion approaches including Akyuz et al. [7], Meylan et al. [145], Rempel et al. [184] and Huo et al. [89]. Six different uncompressed videos were used for the evaluation and the test was conducted on the SIM2 HDR47 display. The evaluation methodology used the ITU Recommendation P.910 [98] for Absolute Categorical Rating (ACR) having 5-levels discrete quality scale. The experiment concluded that the linear operator works better than the other EOs. It also identifies that global operators are temporally coherent, easy to implement and are more practical for video broadcast.

A new study by Abebe et al. [1] evaluates EOs and HDR color appearance models through psychophysical tests. The EOs selected are Akyuz et al. [7], Meylan et al. [145], Banterle et al. [23] and Rempel et al. [184]. The CAMs selected are Kim et al. [106] and Reinhard et al. [179]. The SAMVIQ (Subjective Assessment Methodology for Video Quality) methodology was used to evaluate 10 images on the SIM2 HDR monitor. The best performing EO was once again the linear operator by Akyuz et al. [7] while the better performing CAM was by Kim et al. [106]. In comparison CAMs did not score as well as the linear EO as they were not designed for expansion. Furthermore, the local EOs have the lowest scores indicating that more complex solutions are often not best choice for tone expansion. The work concludes by suggesting that global expansion leads to better results.
and reproduction of contrast and luminance is more important than color in tone expansion. The evaluations presented in this section will be often recalled in this thesis.

### 2.6 Aesthetic Considerations in HDR imaging

We have now discussed the two main retargeting operators in HDR imaging: TMOs and EOs. However, there is very little insight on aesthetic considerations for HDR retargeting. This is surprising especially since the origins of HDR imaging is drawn from photography for purely aesthetic reasons. Figure 2.10 shows one of the first HDR photograph that was shot by Gustave Le Gray [14] in 1856 titled ‘The Brig’. The photographer combined two different exposures to avoid over/under exposure, resulting in an aesthetic monochrome photograph. Over the years, research in HDR imaging has mostly focused on retargeting light signal to a given display and rarely considers aesthetics. We should not forget that HDR technology is not only an attractive option for consumer experience but also for artistic purposes. This is why capturing the entire light information of a scene on an HDR image vital as it gives artists unprecedented possibilities of light and colors to create their intent.

![The Brig by Gustave Le Gray](image)

Fig. 2.10 ‘The Brig’ by Gustave Le Gray, shot in 1856, would become one of the first HDR images created by exposing one negative for the sky and another one the sea, and combining the two into one positive picture. This technique is similar to methods found in modern day digital photography. Figure from wikimedia commons.
2.6 Aesthetic Considerations in HDR imaging

2.6.1 A Brief Review of Image Aesthetics

To have a better understanding of this section, let’s briefly review the fundamentals of image aesthetics. When an artist conceives an image, s/he wishes to convey an intention, emotion, or a message in their art form. They attempt to control various elements, to achieve their purpose. These elements can be described in terms of composition, framing, light, color harmony, sharpness, depth of field, etc. All these elements correspond to conscious technical choices made from the desired aesthetic style of the creator. A state-of-the-art review on these artistic elements can be found in Cozot’s “Habilitation à la Direction de Recherches" thesis [41]. In the context of HDR retargeting, we focus only on the control of light and color to deliver intent.

In a recent SIGGRAPH course, Pouli and Pines [168], present several historical examples to claim that tone mapping done by artists “is the secret underlying technical basis for the entire history of visual arts." Distributing light and color has always been a challenge for artists as their canvas or print photographs have very limited dynamic range (contrast ratio of 1:100). However, over many centuries, artists have utilized this limited container to give us some of the greatest works in art.

The artist is responsible for the lighting configuration of the scene. S/he chooses the position, orientation and other characteristics of the light sources. This part in image creation is probably one of the most important in giving character to the scene. Not only does it determine the brightness and the contrast of the objects in the image but also the overall aesthetic of the light in the image [119]. A serious topic of debate amongst art historians is

Fig. 2.11 Examples of various image aesthetics are presented. Figures (a)-(c) are from wikimedia commons and (d)-(f) are Inez/Vinoodh, Le Turk/ Frédéric Rozot, and Charles Mehling respectively from the vimeo community.
36 Retargeting for High Dynamic Range Imaging

while comparing the use of light in paintings by Rembrandt and Vermeer [158]. A Rembrandt painting often has main light source is in the frame. It produces an image with high contrast. A Vermeer painting uses a diffuse light source outside the frame making the lighting in the image is soft. Similar to Rembrandt, Italian painter Caravaggio uses the chiaroscuro effect [44] to creates sharp contrasted scenes with exaggerated artificial lighting. These three painting styles can be seen in Figure 2.11 (a)-(c).

Stylization using the positioning of light sources can be found in photography and cinematography. Several tools and techniques are used to fine-tune the illumination received by the various areas to compose the image. To balance the light information in a photograph was described by Ansel Adams [2] [3]. It defines an aesthetic of light such that all parts of the image are well-exposed and evenly balanced. This style of aesthetics is called medium key. Two other known classes can be found literature: low key and high key. Low key aesthetics represents an image that appears dark with high contrast while high key aesthetics represents an image that appears bright with low contrast. Examples of these styles are shown in Figure 2.11 (d)-(f). One must remember that artistic is not only limited to portray a style but also to convey an emotion or to set a mood. We shall revisit these concepts in chapter 3.

2.6.2 Tone Mapping as an Artform

In [9], Akyuz et al. share a similar opinion as Pouli and Pines [168]. The authors reject the opinion that a single TMO could solve the tone mapping problem. They argue that TMO preference is a personal choice that is influenced by the artistic taste of the observer. This is why different painters paint the same scene with their personal artistic choices. The authors further extend this concept to tone mapping considering each TMO as an artistic choice and we also observe this in Figures 2.5 and 2.12. This work then proposes an algorithm that learns tone mapping preferences amongst artist on a small set of calibrated images and applies them to any given image. Another work in literature relating aesthetic to tone mapping is by Aydin et al. [17]. The authors present a metric that evaluate images based on fundamental aesthetic qualities such as sharpness, depth, clarity, tone and colorfulness. They validate their metric on various TMOs and conclude that each TMO has its own “aesthetic signature”. Other than these, there are few works in tone mapping that directly address the potential of aesthetics in HDR imaging.

2.6.3 Aesthetic Concerns in Tone Expansion

Even though none of the studies mentioned for tone expansion in Section 2.5 considers artistic intent, they do give insights into the importance on aesthetic considerations. For example,
2.6 Aesthetic Considerations in HDR imaging

Masia et al. [135] base their method on images with varying exposure conditions. Over or under exposing are aesthetic choices made by the photographer. This studies inherently models an adaptive $\gamma$ correction with respect to the exposure. We explore, in depth, the consequences of this in the Chapter 3. In [107], Kinoshita et al. propose an EO by inversing the Reinhard’s photographic TMO [180]. They conduct an experiment where they tone map an HDR image using a number TMOs including Reinhard’s photographic TMO. They find that their EO best matches the HDR image when their EO has been applied to an image tone mapped using Reinhard’s method. At first this may seem obvious but it very important in showing that choice of TMO may strongly influence the EO’s output. This is true for any content mapped on to a SDR container either by automated or manual tone mapping as it will have its own artistic characteristics. Furthermore, neglecting artistic intent could have severe consequences in applications such as live broadcast, which requires automated conversion from HDR to SDR and vice-versa at different points in the transmission chain. For example, a proprietary TMO could be used in the production end while a display manufacturer will have an in-built EO for its panel. This leads to loss of artistic intent and also poor video quality when non-matching TMO-EO pairs are used. We touch upon this subject in Chapter 4.

2.6.4 Aesthetic Challenges in HDR production

A large majority of film directors and technician have been trained to use SDR. Several aesthetic choices are created by uniquely by exploiting the limited dynamic range of SDR production work-flows. We will discuss these aspects in later chapters. However, a number of movies and TV series are transitioning to HDR production for content creation. There are a number of challenges during this phase. The biggest one is technically training the film crew to exploit HDR. An example of this is that cameramen are used to clipping information from
SDR cameras and often set their HDR cameras to do the same. In general, the cameramen should expose the HDR camera to capture the largest dynamic range in scene and do any necessary clipping in post-production.

HDR is strongly related to contrast and this creates several perceptual phenomena that may require artistic consideration. The main visual gain of having higher contrast is better sharpness [140]. This means that noise is easily perceived in HDR. A change in camera setting or de-noising has to be applied in post-production. Furthermore, with HDR we also perceive better contrast detail. This implies that we will be able to see details which were hidden in SDR scenes. For example, when shooting a horror movie [194], which requires low lighting and a dark setting, the hidden monster may be completely visible when captured in HDR. Other choices such as make up, clothing, set design and back drops will have to be re-considered when shooting for HDR. High contrast will accentuate back-lights, jewelry, windows etc. In fact, any bright object will be much brighter. HDR also increases perception of judder or video jerkiness [103], this means panning will also have to be done more carefully when filming for HDR. Even though new lessons will be learned during this transition phase, it will also evoke more opportunity for creativity in the long run for artists and technicians alike. Furthermore, contents such as Star Wars [100] and Game of Thrones [28] are being filmed in HDR and this is generating a lot of anticipation amongst consumers to see how artists will exploit the larger HDR container for multimedia entertainment.

### 2.7 HDR Ecosystem

Recently, many new HDR displays and cameras have been showcased at various technology shows such as National Association of Broadcasters (NAB), Consumer Electronics Show (CES), International Broadcasting Convention (IBC) etc. There is also strong effort amongst standardization bodies to define a telecommunication system for distribution of HDR content. As HDR technology is being standardized, it is clear that video transmission chain will be similar to that of the SDR transmission chain. The SDR video transmission is governed by an OETF-EOTF (Opto-Electronic Transfer Function - Electro-Optical Transfer Function) combination defined in BT.709 [95] and BT.1886 [96] (encoding with a gamma of 2.2 and 2.4 respectively). The current SDR TV systems were not designed to handle HDR content and hence an alternative OETF-EOTF is needed to capture the dynamic range of the HDR video. The current status of the industry sees the Society of Motion Picture & Television Engineers (SMPTE) that has standardized the ST 2084 or Perceptual Quantizer (PQ) [197] and the Association of Radio Industries and Businesses (ARIB) that has standardized the STD-B67 or Hybrid Log-Gamma (HLG) [12]. Both offer higher dynamic ranges than the
2.7 HDR Ecosystem

SDR BT.709/BT.1886. In 2016, PQ and HLG have been both considered by the International Telecommunication Union (ITU) standard the BT.2100 [94] combining both standards to be used in HDR broadcast for production and international programme exchange.

Wide Color Gamut is another video characteristic which goes hand-in-hand with HDR for next generation displays. Current SDR video systems based on BT.709 color gamut cannot support the wide range of colors that the human eye can perceive. To support this need, the ITU has specified BT.2020 [91] color space for next generation TV sets and Digital Cinema Initiatives (DCI) has specified the P3 color space [195] for digital movie projection. Figure 2.13 shows these wider color gamuts in the CIE 1931 chromaticity diagram. Although most consumer TV sets cannot yet produce the entire BT.2020 gamut, they always cover a wider gamut than BT.709. Moreover BT.2020 is the color space to be used in distribution of HDR/WCG content [206].

A number of displays in the consumer space claim to support HDR. However not all are truely BT.2100 and BT.2020 compliant. The UHD Alliance (UHDA) [205] announced the HDR specifications for next generation displays as seen in Table 2.2. This gives a good indication of the diversity in display technology and requirements for consumer TV’s available in the market. It is also points to the various challenges need for a content to retarget
Table 2.2 UHD Alliance Premium Specification

<table>
<thead>
<tr>
<th>Specification</th>
<th>LCD HDR Displays</th>
<th>OLED HDR Displays</th>
<th>Mastering Displays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>3840x2160</td>
<td>3840x2160</td>
<td>3840x2160</td>
</tr>
<tr>
<td>Bit Depth</td>
<td>Minimum 10-bit</td>
<td>Minimum 10-bit</td>
<td>Minimum 10-bit</td>
</tr>
<tr>
<td>Gamut</td>
<td>More than 90% of P3</td>
<td>More than 90% of P3</td>
<td>Minimum 100% of P3</td>
</tr>
<tr>
<td>EOTF</td>
<td>SMPTE ST 2084</td>
<td>SMPTE ST 2084</td>
<td>SMPTE ST 2084</td>
</tr>
<tr>
<td>Peak Brightness</td>
<td>More than 1000 nits</td>
<td>More than 540 nits</td>
<td>More than 1000 nits</td>
</tr>
<tr>
<td>Black Level</td>
<td>less than 0.05 nits</td>
<td>less than 0.0005 nits</td>
<td>Less than 0.03 nits</td>
</tr>
</tbody>
</table>

for different TV systems with various characteristics and specifications, highlighting the importance of display retargeting in HDR imaging.
Chapter 3

Aesthetic HDR Retargeting

Among the multiple new dimensions of Ultra High Definition (UHD), there is a wide consensus that High Dynamic Range is the most valuable, based on its perceived “wow factor” and modest upgrade cost compared with SDR. An HDR-ready ecosystem is currently being built, ranging from new cameras and production infrastructure to consumer displays and set top boxes (STBs). In view of this, the HDR and WCG video pipeline was standardized in 2016 in the form of the ITU document BT. 2100 [94]. There is a growing need amongst content providers to use this standard and exploit the benefits of HDR technology to deploy better services to their customers. In order to support the migration toward HDR broadcast, an SDR to HDR tone expansion will be a key requirement for playout servers, broadcast converters and encoders in the coming decade, given the existing amount of SDR content assets. This chapter explores retargeting in this context of SDR to HDR tone expansion for video broadcast.

We may recall from Section 2.6.3, that no study has yet considered aesthetics in the tone expansion process. In this chapter, we will re-examine aesthetic styles and their effect on EOs. We will also consider the subjective evaluation by De Simone et al. [47] and Abebe et al. [1] presented in Section 2.4.4. These works suggest that amongst the many EOs in the literature, a simple linear operator gives the best results. However, as shown later on, such an approach does not always preserve artistic intent particularly for complex lighting styles characterized by extreme variation in luminance and contrast.

The main aim of this chapter is to develop a retargeting method to display SDR content on an HDR monitor exploiting the available dynamic range and color gamut while fully preserving the artistic intent. Furthermore, unlike other EOs, our experiments have been implemented on a professional 1000 nits HDR display instead of prototype displays. We kept in mind that initial commercial HDR displays in the market are limited to 1000 nits and have oriented our work towards it. The main contributions of this chapter are:
Aesthetic HDR Retargeting

• A review of aesthetics and its implication in tone reproduction,
• User studies giving insights on human preference on viewing HDR content,
• A tone expansion technique that preserves artistic intent and is also faithful to reality,
• HDR color reproduction and correction for tone expansion,
• Industrial framework for tone expansion on commercial HDR displays compliant with emerging standards for HDR video.

This chapter is structured as follows. Section 3.1 reviews the background on aesthetics, EOs and HDR color reproduction. Section 3.2 describes of our tone expansion method with detailed user studies. In Section 3.3, we propose a color correction method specific to our EO. Section 3.4 integrates our method in an industrial framework. We present our results in Section 3.5 and conclude the chapter in Section 3.6.

3.1 Background

In this section we look into lighting style aesthetics and its impact on existing EOs. Based on this study, we position our approach with respect to the current state-of-the-art tone expansion models. We also explore some color correction techniques important in tone reproduction.

3.1.1 Aesthetics and its Impact on Tone Expansion

It is necessary to clearly define aesthetics, in order to design an EO that preserves artistic intent. Let us recall from Section 2.6.1 where we classify aesthetics by painting styles and cinematography styles based on lighting techniques. In this chapter, we consider lighting styles aesthetics that are used in cinematography for artistic purposes in order to portray a particular mood of a scene [214]. For example, a low key image usually captures a dark and dramatic mood while high key image is typically associated with a upbeat feelings. This effect is created in low key and high key photography techniques by creating an image with a low luminance with high contrast and bright tones with low contrast respectively. We may recall that medium key images are often associated to medium contrast and luminance and is the most common lighting style. According to style analysis and statistics, we notice that these styles do not address bright luminance with high contrast and a dark image with low contrast. That is why in addition to these existing lighting styles, we introduce two new classes to cover a wider distribution of luminance and contrast in image aesthetics- dark key and bright key. All five classes are defined below:
Fig. 3.1 Lighting styles can be distinguished by their luminance and contrast. We introduce two new classes highlighted in red - dark key and bright key. Image courtesy Dominik Muench, Tsalparoglou Spiros, Javier Lopez and Georges Antoni from the vimeo community.

1. Dark Key (DK): Lighting with low contrast and low luminance. Typically found in horror or murder mystery movies used to build fear or suspense.

2. Low Key (LK): Lighting with high contrast and low luminance. This lighting style is used to convey intense and tension filled atmosphere. A good example of low key cinematography is the film *Sin City* [147].

3. Medium Key (MK): Lighting with average contrast and moderate luminance. Most video content comes under this category - news broadcasting, television drama series, etc.

4. Bright Key (BK): Lighting with high contrast and high luminance. This style is commonly found in an exterior environment when shooting nature on a bright sunny day.

5. High Key (HK): Lighting with low contrast and high luminance. This lighting style is often used to express a light mood and can be found in fashion and advertisement videos.

Figure 3.1 summarizes the five lighting styles considered in this chapter and gives examples of each style. It must be noted that our study is based on the hypothesis that we
Aesthetic HDR Retargeting

Fig. 3.2 The SDR input image on the left has a dark key lighting style. In the center image the Linear Scaling EO by Akyuz et al. [7] is applied while the right image is processed using the Gamma Correction EO by Masia et al.[135]. In both cases the intended aesthetic of the SDR image is lost. It should be noted that most of the tone mapped HDR content to be seen on the printed version of this thesis is done using display adaptive tone mapping by Mantiuk et al. [129] unless specified.

are working with stylized SDR content coming out of cinematic color grading workflow. In other words, dark key is not the equivalent of under exposed and high key does not mean over exposed. This is a significant distinction of our work from Masia et al.’s [135] which considers under and over-exposed source content. We assume that this content is driven by pure artistic intent made by the cinematographer and not wrong choices made by the cameraman. This is a reasonable hypothesis since existing video catalogs of professional content rarely has a wrongly exposed shot unless it is absolutely intentional. Furthermore, usage of various lighting styles provide a good metric to challenge existing tone expansion techniques especially in terms of preserving artistic intent. An example of EOs failing to preserve artistic intent can be see in Figure 3.2.

In our study we shall consider global operators as recommend in previously mentioned subjective evaluations [47] [1]. The main reason we prefer a global approach is because local methods do not offer substantial gains in terms of video quality or reduced complexity. A general global model is based on the equation below:

\[ Y_{HDR} = L_{max} \times Y_{SDR}^{\gamma} \]  

where \( Y_{HDR} \) is the HDR luminance, \( L_{max} \) is the maximum luminance that can be supported by the display, \( Y_{SDR} \) is the SDR luminance and \( \gamma \) is an adaptive parameter that varies per frame. We align our work on similar user studies that model the choice of \( \gamma \) [178] [43] considering not only well-exposed or over-exposed images but a variety of lighting style aesthetics.

3.1.2 Color Appearance and Reproduction in HDR

Color reproduction has been extensively studied on devices with limited dynamic range and color gamut in the context of tone mapping and gamut mapping [149] [167] [133]. However,
very little is known on correctly expanding tone and gamut of legacy content on to next generation displays. From color appearance studies we know that a number of factors affect our color perception such as image size [159], luminance and color of background and viewing environment [177] [179]. Perhaps various psychophysical phenomena could such as the hunt effect or Stevens effect could play an important role in accurate expansion of tones and color. However, work by Abebe et al. [1] did point out that HDR color appearance models such as works by Kim et al. [106] and Reinhard et al. [179] are not always suitable for tone expansion. Thus in this chapter, we refrain from using appearance models as they are used to predict perceived colors of a given stimulus under different viewing environments [66]. In our case, we attempt to reproduce cinematic colors [191] which are colors mastered by a professional color grading suite.

Discarding appearance models as a possible color reproduction solution we seek inspiration from methods used in tone mapping for color correction. In tone mapping, the basic approach for color processing was introduced by Schlick [188] by replacing color channels as shown below:

\[ R_{out} = \left( \frac{R_{in}}{L_{in}} \right)^s L_{out} \] (3.2)

where \( R \) represents a color channel (red, green, or blue), \( L \) is the luminance channel, \( s \) is a color saturation parameter, and in/out indicates the channel before and after tone reproduction. It is known that mapping using only the luminance channel in the RGB color space with saturation at \( s = 1 \), often results in over-saturated images in tone mapping. The case when \( s \neq 1 \), may also lead to shifts in luminance. To avoid this effect, an alternative saturation adjustment was proposed by Mantiuk et al. [133]:

\[ R_{out} = \left( \frac{R_{in}}{L_{in}} - 1 \right)s + 1 L_{out} \] (3.3)

This formulation preserves luminance however it creates slight changes in hue [175]. To use these methods the operator must be global, hence making it an ideal match to a global EO and also an attractive alternative to color appearance models. In this work, we look into such correction strategies and apply them in the opposite direction, i.e., tone expansion instead of tone mapping.

## 3.2 Style Aware Tone Expansion

This section presents our tone expansion operator. We start by developing a method to classify different lighting styles and discuss the source content selection. We then explore
3.2.1 Lighting Style Classification and Source Content

We can broadly visualize each video style mentioned in Section 3.1.1 in Figure 3.3. This technique can be easily applied on images as well. For each video sequence style, every frame was analyzed in the CIE-Lab domain. Each point on the map represents the statistics of one frame. We define the luminance of a style as the median of the lightness channel $L_{\text{median}}^*$ of the frame and the contrast as the standard deviation $\sigma_{L^*}$ of the lightness channel of the frame. We have selected the $L_{\text{median}}^*$ over the $L_{\text{mean}}^*$ as it does not vary significantly over time when there are slight changes in lighting of a scene and hence gives better representation of the mood. We discuss this phenomenon in detail in Section 3.2.4.

The data in Figure 3.3 has been extracted from a set video scenes of 4-5 seconds which were considered to be a good representation of each style. However, a normal video sequence may undergo a number of lighting style changes over a period of time. To visualize smooth transitions over different style clusters, we allow the clusters boundaries to overlap...

---

Fig. 3.3 A fuzzy c-means clustering of lighting styles of videos. Clusters with overlapping boundaries account for the change in lighting style of video over time, thus an image can be part of more than one style.
3.2 Style Aware Tone Expansion

in the form of a fuzzy c-means algorithm. This soft classification can be visualized by the contour map which represents the position of points of equal membership.

For this chapter, our source content could be classified using one or more lighting styles. The video sequences were in 1280 × 720p or 1920 × 1080p, 8 bits and of BT.709 colorimetry. The sequences consisted of various frame rates: 24, 25, 30, 50 and 60 fps. A lot of our stylized content are available on the Internet and have been compressed using VP9 and AVC codecs.

3.2.2 User Study on Stylized Video

This section investigates the human preference for \( \gamma \) at different video lighting styles. The choice of \( \gamma \) is essential to the tone expansion process in Equation 3.1. In this study, an HDR video of a particular style are shown to the user using different \( \gamma \) values (1.0-2.2) and we attempt to answer the following questions:

1. At what \( \gamma \) is the HDR video aesthetically pleasing?
2. At what \( \gamma \) does the HDR video best represent the style of the SDR video?

The first question is non-reference user study, where the user is required to choose the best \( \gamma \) without a reference SDR video. Hence, the user is required to choose the most aesthetically pleasing \( \gamma \) without knowing the style of the source content. The second question is a reference user study, where the user matches the fidelity of the lighting style of the HDR content using the SDR video as reference. We aim to use both tests not only to give us a better understanding of choice of \( \gamma \) but also human preference on viewing different lighting styles in HDR.

Testing Environment and Methodology

Both tests were conducted using ITU-R recommendations BT.500-13 [97], in a room with ambient luminance of approximately 25 lux which is the accepted luminance levels for a dark environment. The HDR display used is the Sony BVM-X300 while the SDR display is the Panasonic BT-LH310P. Both displays are 32 inch professional monitors which have minimal internal processing. It should be noted that the Sony is an OLED while the Panasonic is a LCD display and the Sony gives better picture quality in SDR mode than the Panasonic. For the experiments, both displays were placed side by side 0.5 m apart from each other. The distance between the participants and the display was around 1.5 m. For this work, we have used a high performance video player that can display UHD (3840 x 2160) uncompressed video content at high frame rates. It is able to play 10 bit video in Y’CbCr 4:2:2 format. Both
the Sony and the Panasonic displays can play UHD video while requiring a connection with 4 3G-SDI cables. This allowed us to concatenate 4 HD videos in UHD format and feed each display with a single SDI cable to view content in HD. This was very useful in side by side subjective testing.

We worked with the professional Sony BVM-X300 OLED display [198] because its is equipped with standardized EOTFs. We chose to work with the PQ transfer function since at the time of the experiment the HLG firmware update was not available on the Sony. The Sony display is different from the Sim2 HDR47 or the BrightSide DR-37P monitors used in previous HDR studies as it can be interfaced with industry standards and is compliant with BT.709 and BT.2020 colorimetry. In addition to this, the BVM-X300 is a professional display so we expect no internal processing in comparison to consumer displays and hence an ideal system to test our algorithms.

Since both these studies are conducted to determine the choice of \( \gamma \) in Equation 3.1, it was preferred to carry out this pilot study with a few video experts. We considered three videos each of different lighting styles: LK, DK, MK, BK and HK. Hence a total of 15 videos of varying 8-15 seconds. Videos of the same style were not displayed one after the other. Each video was displayed with different \( \gamma \) values (1.0-2.2) in an incremental order and the participant was asked to choose the preferred \( \gamma \). The user was given the option to view the sequence multiple times before they made their final choice. The objective of the experiment was given to the experts prior to the test. This was followed by a short training on a Medium Key sequence which was not used in our results. Each user study lasted between 45 minutes and 1 hour. Hence, both tests were conducted with prolonged breaks between tests and sometimes different days depending on the availability of the video expert.

**User Study 1: Aesthetics**

The result of the non-reference test is shown in Figure 3.4. It can be seen that DK and LK sequences have a preference for \( \gamma \) ranging from 1.9 to 2.1. While MK and HK sequences have a variety of \( \gamma \) preferences and BK sequences are closer to \( \gamma = 1 \). This result suggests that changing \( \gamma \) or contrast does play an important aesthetic role for videos with different lighting styles. Experiments conducted by Akyuz et al. [7] and De Simone et al. [47] were done for correctly exposed or medium key images so a \( \gamma = 1 \) could be a reasonable outcome. However, for LK and DK contents, \( \gamma = 1 \) results in expanding compression artifacts and banding. Our user study shows a higher \( \gamma \) value avoids such artifacts. This is mostly due to the non-linear signal expansion in which the \( \gamma \) expands the mid-tones and highlights more than dark regions. It should be noted that current video compression algorithms have been designed for SDR content, and are efficient at hiding artifacts just below the visible...
threshold [182]. These artifacts are more easily detected in HDR displays upon expansion. Using a high gamma value in LK and DK content can help to suppress these artifacts and preserve the style at the same time. For this non-reference study, we also observe that for HK content there is a huge variation in $\gamma$. This is primarily because the participants were trying to increase the contrast to find missing regions in the over-saturated white regions (saturated for artistic reasons). It could also possible explain the user study conducted by Masia et al. [135] which suggested a higher $\gamma$ value for over exposed images. Hence, a more reliable $\gamma$ could be chosen by performing a test using the SDR video as reference.

**User Study 2: Fidelity**

The result for the reference test can be seen in Figure 3.5. We once again observe higher $\gamma$ values for DK and LK similar to the previous test. MK has a $\gamma$ value in the mid range while BK and HK have low $\gamma$ values. It can be noticed that the LK boxplot has a larger deviation from the mean. This is because a high $\gamma$ value exhibits nonlinearity and this results in a desaturated appearance. Some participants attempted to lower the $\gamma$ to achieve matching color saturation levels while compromising lighting style fidelity. Nevertheless, the range of each boxplot is limited as compared to the non reference test which makes this a more precise model to select the $\gamma$. The most important findings of both these user studies suggest that adjusting contrast in the form of $\gamma$ correction is essential for a global tone
expansion model as it improves both aesthetics and fidelity of the HDR video. Furthermore, the fidelity study shows a decrease in the $\gamma$ value as the lighting of the style increases. This is fundamentally different from what is proposed by Masia et al. [135] which suggests that the $\gamma$ value increases as the image gets brighter (or over exposed).

### 3.2.3 Style Adaptive Gamma Correction

User studies conducted in Section 3.2.2 gives us an indication on how $\gamma$ preference varies with the lighting style. However, in a practical scenario, it is not possible to manually tweak $\gamma$ values according to the style. In this section we propose a frame by frame style adaptive $\gamma$ correction process.

As previously mentioned, the results of the fidelity test show a steady decline in $\gamma$ as the luminance of the style increases (Figure 3.5). Considering our video style classification in Figure 3.3, we can characterize a video’s style by the $L^\text{median}_{s}$ statistic of each frame. Thus, the referenced subjective tests have shown that the ideal $\gamma$ for tone expansion can be modeled by the $\tilde{L}^s$ of the SDR image in Figure 3.6 and is fitted to the following equation:

$$\gamma = 1 + \log_{10}\left(\frac{1}{\tilde{L}^s}\right)$$

(3.4)
3.2 Style Aware Tone Expansion

where:
\[ \bar{L}^* = \frac{L^*_{\text{median}}}{100} \]
and has been clipped in the interval [0.05();0.95] to make the regression less sensitive to outliers [177].

Hence, our tone expansion model takes the form:

\[ Y_{HDR} = L_{\text{max}} \times Y_{SDR}^{\gamma} \tag{3.5} \]

where:
- \( Y_{HDR} \) is the \( Y \) component of XYZ colorspace in HDR domain
- \( Y_{SDR} \) is the \( Y \) component of XYZ colorspace in SDR domain
- \( L_{\text{max}} \) is the maximum display luminance (1000 nits)
- \( \gamma \) is a style adaptive parameter defined by \( \gamma = 1 + \log_{10}\left(\frac{1}{\bar{L}^*}\right) \)

### 3.2.4 Temporal Coherency in Tone Expansion

In this work, we base our EO on the Equation 3.5. A concern using such an adaptive \( \gamma \) correction is that the video analysis for lighting style is spatially global, so in a video that has lighting variations in different parts of the scene will likely exhibit luminance artifacts.

\[ \gamma = 1 + \log_{10}\left(\frac{1}{\bar{L}^*}\right) \]

Fig. 3.6 Adaptive gamma curve.
We resolve this issue by only using the $L^*_{\text{median}}$ to calculate the $\gamma$ value. In this section we explain this choice.

Our analysis in Section 3.2.1 classifies lighting styles from high-level information in the form of $L^*_{\text{median}}$ and $\sigma_{L^*}$. From Figure 3.3, we observe that for a given video sequence there are larger spreads of data points along the $\sigma_{L^*}$ axis than the $L^*_{\text{median}}$ axis. This observation demonstrates the temporal stability of the $L^*_{\text{median}}$. Through this we identify that including $\sigma_{L^*}$ to determine the $\gamma$ value for the global model may result in a temporal incoherent video. In addition to this, we consider that $L^*_{\text{median}}$ is sufficient to indicate the style of the video since there is significant overlap in choice of $\gamma$ for different styles as seen in Figure 3.5 in User Study 2. It is more important to maintain higher $\gamma$ values for darker styles and lower $\gamma$ values for brighter styles with perceptually stable transitions. Moreover, these perceptually stable transitions are inherent to the the changes in lightness channel of CIE-Lab which represent perceptually uniform changes of same visual importance. Hence, we propose that this is best represented by the $L^*_{\text{median}}$.

A related point to consider is that both works by Akyuz et al. [7] and Masia et al. [135] were designed for tone expansion of images and thus do not account for the temporal aspect.

![Temporal Coherency of the Median](image)

Fig. 3.7 This graph shows the $L^*_{\text{median}}$ and the $L^*_{\text{mean}}$ of several frames of an SDR video transitioning several times in between DK, LK and MK. Looking closely at frames: 350-500, 600-650, and 900-1000, it can be observed that the temporal transitions of $L^*_{\text{median}}$ are far more stable than the $L^*_{\text{mean}}$. Besides the fact that using the $L^*_{\text{mean}}$ to calculate the $\gamma$ can lead to global flickering artifacts, these jumps in $\gamma$ may result in change of lighting style as well. Furthermore, it is well known that an increase of the retinal illuminance increases perceptual sensitivity to flicker [104], hence it is more sensible to use the $L^*_{\text{median}}$ for tone expansion.
In case of Akyuz et al. [7], $\gamma = 1$ thus no temporal inconsistency is expected. Masia et al.’s [135] method calculates the $\gamma$ using the key which is related to the geometric mean of the input luminance. In Figure 3.7, we see that using the $L^*_{\text{median}}$ over the $L^*_{\text{mean}}$ in Equation 3.4 also has significant benefits.

Using only the $L^*_{\text{median}}$ for our EO and not the $\sigma_{L^*}$ does make an argument against our classification method used in Figure 3.3. When we initially began investigating tone expansion we believed both contrast and lightness both would play a role in computing the HDR image. However, as previously mentioned, only the $L^*_{\text{median}}$ is necessary for the final gamma value. Nevertheless, we present our classification method for lighting style believing its a useful scientific contribution which came about as a side result of this research.

3.3 Color Correction for Tone Expansion

Our EO presented in Section 3.2.3 is based on a $\gamma$ correction that preserves the style of the frame. However, in User Study 2 we observed that for high $\gamma$ values the resultant HDR video had a desaturated color appearance with respect to the SDR video. Equations 3.2 and 3.3 from Section 3.1.2 provide means of correcting this issue. The main difference between both color correction formulations is that Equation 3.2 is a hue preserving color saturation method while Equation 3.3 is a luminance preserving color saturation method. Recalling that luminance and contrast reproduction are more important in tone expansion than color [1], we compromise on hue preservation and choose Equation 3.3 to color correct our EO. The main reason we abstain from using Equation 3.2, is that not only does it modify saturation, but also the luminance of HDR video. Our EO has been designed to determine the best possible style preserving $Y_{\text{HDR}}$ and we cannot compromise this luminance to preserve hue. In the following sections we present a user study investigating the evolution of the saturation with respect to the $\gamma$ of the tone curve.

3.3.1 User Study 3: Color Fidelity

The choice of saturation is required to determine the color corrected channel $R_{\text{out}}$ in Equation 3.3. We aim to model this saturation value with respect to the tone curve, i.e., $\gamma$ value of the frame. To do this we perform a user study asking users: *At what saturation value does the HDR image best model the color appearance of the SDR image?* This study is conducted using the testing conditions and methodology presented in Section 3.2.2. It is very similar to User Study 2 as an expert user is required to match the color fidelity of the HDR content using an SDR reference. The main difference in this study is that the HDR content
displayed to the user is already \( \gamma \) corrected using our EO and the user is required to choose the saturation value (varying from 0.8-1.6) that best matches the SDR input in terms of color fidelity. Also, in this study we are working with still images that cover all \( \gamma \) values 1.0-2.2 (hence, encompassing all five lighting styles).

The results for this test is shown in Figure 3.8. It can be seen that the saturation value shows no direct trend with respect to the \( \gamma \) non-linearity of the curve. Most noticeably, the DK (high \( \gamma \) values) and HK (low \( \gamma \) values) require very little to no changes in saturation. We find this behavior is mostly due to the fact there is very little color in DK and HK content and varying the \( s \) effectively re-saturates the levels of black or white. The effect of the nonlinearity of our EO on the color de-saturation is more apparent in LK, MK and BK videos. We see that highest color saturation is chosen for \( \gamma = 1.6 - 1.8 \) and significantly high values chosen for \( \gamma = 1.3 - 1.5 \). This is because these image styles have a larger abundance of color. However, if we discard HK and DK contents and consider only more colorful images using \( \gamma = 1.4 - 1.8 \), we observe and increasing trend in \( s \) with respect to \( \gamma \). This shows the the output image colorfulness is indeed affected by the \( \gamma \). Thus, this user study suggest that saturation parameter for tone expansion is dependent on a combination of both the EO tone curve and the colorfulness of the source SDR content.

In agreement with the study by Abebe et al. [1] and through our own observations, applying a correct \( \gamma \) is far more critical to tone expansion than choosing the right saturation
value [1]. Keeping this in mind, we cannot completely ignore this de-saturation for certain cases, hence we suggest $s = 1.25$ as the recommended re-saturation constant if this correction is to be applied in an automated application. This value is derived from several pilot studies as it provides acceptable results for all lighting styles to the extent that it can emulate the source SDR colors to an appropriate saturation level on the Sony BVM-X300 display. Further studies, however, should investigate expanding an entire color volume, taking into account the colorfulness of source content and the applied $\gamma$ value in a temporally coherent manner.

### 3.4 Tone Expansion Framework

Now all pieces are in place to position our tone expansion operator in an industrial framework. To apply our EO to a video, in SDR format to be displayed on an HDR television, requires certain format conversions and signal pre-processing. We define the SDR to HDR up-conversion process chain in Figure 3.9. This exact framework was used for the user studies and in evaluating our results in Section 3.5.

At the heart of this framework is our EO which is placed on the right hand side of Figure 3.9. This operator is applied in the linear domain and packaging the output of our linear EO into a standardized HDR format has its consequences especially in terms of color gamut. This is primarily because Y’CbCr has its shortcomings in containing an HDR signal with saturated primaries [76]. Our source content natively is within the BT.709 gamut and applying our recommended saturation parameter $s = 1.25$ during the color correction process often pushes the colors outside BT.709 container. Consequently, out of gamut colors are inevitably clipped and result in visible color bursting artifacts as seen in Figure 3.10 (c). To prevent this, we map the output of our EO to a BT.2020 container. By doing so, we prevent color artifacts and also format the HDR video to wider color space which is an essential characteristic of HDR TVs.

### 3.5 Results

To evaluate our method, we compare our EO with existing global operators by Akyuz et al. [7] and Masia et al. [135]. The evaluation will be a subjective study that will take into account the fidelity of the HDR content and the aesthetic pleasingness of the results. The environmental setup and test methodology was similar to the user studies conducted with experts.

We asked 23 users to evaluate 3 videos of each style. We start with a test without SDR reference to evaluate the aesthetic value of the three methods and a test with SDR reference
Aesthetic HDR Retargeting

Fig. 3.9 File format conversions for Tone Expansion is defined on the left hand side: The input is an SDR (R'G'B') video which is a display referred nonlinear signal encoded by an OETF (gamma for BT.709/BT.1886). The content is 8 bit and hence has [0:255] code values. The SDR video input is normalized and then linearized using the BT.709 OETF −1 decoding. The result is a linear RGB signal that passes through an Expansion Operator. The output from the EO block is an HDR RGB signal to which we apply a gamut mapping to BT.2020 using the BT.2087 ITU recommendation. This signal is then encoded using an HDR OETF (PQ) and converted again into R'G'B' but in HDR format. We then change the colorspace to Y'CbCr used by a video player and quantize it 10 bit with video levels of [64:940]. The Chroma is sub-sampled from 4:4:4 to 4:2:2 (Y210 format). We use a 10 bit signal as the PO curve exhibits perceptual uniformity at 10 bits [148]. The 10 bit Y'CbCr 4:2:2 HDR signal is the final format displayed on the HDR TV. Our tone expansion method is presented on the right hand side: The input is a linear SDR RGB video which is then converted to XYZ in linear domain. We extract the Y component and calculate \( \gamma \) as defined in Equation 3.4. We then apply the tone expansion model using Equation 3.5. Once the \( Y_{HDR} \) is calculated, we can determine the RGB values in HDR domain by subsequently applying the color correction from Equation 3.3.
3.5 Results

Fig. 3.10 The original SDR BT.709 gamut (a) is compared with the resultant color gamut at different stages of our EO. The HDR chromacity is shown in (b) where $s = 1$. The colors remain well within the BT.709 gamut but perceptually the image looks desaturated. In (c) we set $s = 1.25$ and image appears correctly saturated. However, certain regions in the image (zoom on tyres) have color artifacts. We notice a number of colors are at the edge of the gamut. These colors are mapped on the wider BT.2020 gamut in (d). This step prevents color artifacts, packages our EOs HDR video into BT.2020 and allows us to vary the saturation parameter.

To assess the fidelity, the participants were 23 to 55 years old with 20 males and 3 females. 10 had previously seen HDR videos. 11 of them wear glasses and all have normal color vision. First, the user was asked to do the non-reference test and give a score to each method on a Five-point scale (5-excellent, 4-good, 3-acceptable, 2-poor, 1-bad). Second, the user had to evaluate the fidelity of the HDR rendering of the different EOs to the SDR content using the same scoring method. For this evaluation choice of $\gamma$ was automated using Equation 3.5 and saturation factor was set to $s = 1$.

Figures 3.11 and 3.12 show the Box-and-Whiskers plots for each method. Both plots identify that our method gives significant gains in terms of aesthetics and fidelity over existing methods. We see that our method has a highest MOS (Mean Opinion Score) for both types of scores while the method by Masia et al. [135] has the lowest MOS. Our method also has less deviation than Akyuz et al. [7] in both types of scores.
Fig. 3.11 MOS score evaluation of aesthetics of algorithms.

Fig. 3.12 MOS score evaluation of fidelity of algorithms.

Through the non-reference study we see that users preferred our method purely in terms in aesthetics. However, our method performed even better in the reference study highlighting the importance of preserving artistic intent and the need for aesthetic retargeting.
3.5 Results

We can also visualize the results in Figure 3.13. The tone-mapped images seen in this figure also show that lighting style is not well preserved by Akyuz et al. [7] especially in LK and DK sequences, where they amplify quantization and compression artifacts and camera noise. Furthermore, experiments by Masia et al. [135] were conducted for a very specific set of over-exposed images and thus give an unnatural result when dealing with DK, LK, HK and sometimes even with MK videos. For the sequences where Masia et al.’s [135] does seem to give a natural appearance (mostly BK) we observe constant flickering. This is because their method adapts the \( \gamma \) with respect to the mean and not the median and thus results in temporal incoherency explained in Section 3.2.4.

Overall, our method adapts very well to the lighting style of the source content. It does not amplify coding artifacts which may exist in dark regions or exhibit any flickering when compared to other methods. However, there were some color discrepancies while comparing SDR-HDR fidelity for DK and LK videos. The users claimed the videos were of cooler color temperature. This can be corrected using a saturation factor of \( s = 1.25 \). Furthermore, the mood of the scene is miscalculated in case of sequences with Digital Video Effects (DVE). For example, if a MK video splits half of its screen with a drop-down black colored

<table>
<thead>
<tr>
<th></th>
<th>DK</th>
<th>LK</th>
<th>MK</th>
<th>BK</th>
<th>HK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SDR</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td><strong>Our</strong></td>
<td>(3.78, 3.98)</td>
<td>(4.11, 4.63)</td>
<td>(4.02, 4.64)</td>
<td>(4.30, 4.32)</td>
<td>(4.54, 4.36)</td>
</tr>
<tr>
<td><strong>Akyuz</strong></td>
<td>(2.25, 2.16)</td>
<td>(2.20, 2.36)</td>
<td>(2.20, 2.63)</td>
<td>(2.21, 2.00)</td>
<td>(1.32, 1.23)</td>
</tr>
<tr>
<td><strong>Masia</strong></td>
<td>(1.00, 1.00)</td>
<td>(1.00, 1.00)</td>
<td>(1.66, 1.00)</td>
<td>(1.88, 1.84)</td>
<td>(1.00, 1.00)</td>
</tr>
</tbody>
</table>

Fig. 3.13 Comparing the results of various EOs: The aesthetic and fidelity scores for each algorithm are presented below each result in the form: (Fidelity, Aesthetics). These results are far more visually convincing when seen through an HDR display. However, certain discrepancies of other algorithms can be easily spotted. Image courtesy Krisztian Tejfel, Dominik Muench, Rustykicks and Dmitry Arn from the vimeo community.
advertisement, our method will be misled to calculate a high $\gamma$ value adapted to an LK or DK sequence and the intended mood of the MK video will be lost.

The results of this subjective evaluation were produced using the Sony BVM-X300. To verify that our results were not “display specific” we visualized the videos (with $L_{\text{max}} = 1000$ nits) on the Sim2 HDR47 monitor and the Samsung JS9500 1000 nits commercial display. Through a candid crosscheck, we once again observed our method performs better than the others.

3.6 Conclusion and Future Work

Previous works on global tone expansion models have mostly dealt with well-exposed and over-exposed images. These methods provide substantial knowledge and lay the foundations for EOs. However, legacy SDR content not only varies in exposure but also offers a variety of lighting style aesthetics which are the result of artistic choices. Existing EOs such as the one by Akyuz et al. [7] tend to boost compression artifacts while Masia et al.’s [135] $\gamma$ correction mostly works with over-exposed images. Both strategies work within their respective use-cases but no study has yet been considered using lighting style.

We performed two pilot studies regarding $\gamma$ correction, one being a non-reference test evaluating the effect of $\gamma$ correction on the HDR aesthetics while the other evaluates the same in terms of HDR fidelity to source content. This study shows that lighting style of content is strongly related to the $\gamma$ value used in tone expansion. We present a simple global tone expansion model where the $\gamma$ adapts to the style of the content. We then conducted a third user study based on color fidelity, studying the effect of the $\gamma$ correction on the color saturation. Through this, we propose a simple color correction method for our EO. Finally, we demonstrated that our EO outperforms existing methods through a subjective evaluation. This work has been oriented for 1000 nits HDR displays that have slowly started infiltrating the consumer market. We also presented a framework that places our EO in conformance with existing SDR standards and the latest HDR TV standards.

For future work, we would like to improve certain shortcomings of our EO. One such weaknesses is in terms of colorimetry, we have recommended a quick fix by using a saturation parameter of $s = 1.25$. Further studies are required to correctly relate the $s$ factor to the colorfulness of the frame. Hence, a more elegant approach has to be targeted for color volume mapping for tone and gamut expansion. In addition to this, there are special cases where a global $\gamma$ correction fails such as a change in lighting style in background and foreground as well as the case for DVE. More effort is needed to alleviate these discrepancies. Our EO also mitigates compression artifacts in dark regions by choosing an higher $\gamma$ value. However, the
types of compression and banding artifacts depend on the choice of codec and the bit rates chosen for transmission. Separate processing is necessary to reduce such artifacts. Finally, our EO is based on the hypothesis that our source content has been generated from a stylized professional workflow. In the case of user generated unprofessional video, a style-aware algorithm may not always be necessary. Alternative methods must be explored for such situations.

In conclusion, in this chapter we combined concepts in aesthetics and perception to demonstrate a novel HDR retargeting method for tone expansion that could be used for video broadcast. The method unique and it is designed to cater to artistic intent by using a simple $\gamma$ correction. However, this method has been applied for a fixed peak brightness of 1000 nits. It would be interesting to extend our retargeting method into a display independent model. In other words, adapting displays of any dynamic range and color gamut. This will be the topic of the next chapter.
Chapter 4

Retargeting between HDR Displays

An obvious difference a customer can spot when walking into a consumer electronics store is that the same video content looks slightly different on various TV sets. One of the main reasons for this discrepancy is that not all consumer TV’s are configured with the same system $\gamma$. A $\gamma$ of 2.2 fits well for ambient lighting conditions while a $\gamma$ of 2.4 gives slightly more contrast. This $\gamma$ correction has a direct effect on the tone curve and hence the same video looks slightly different on such displays. Tone compatibility is a common problem amongst many commercial SDR displays since they are a display referred system (recall Section 2.1). This issue is far more critical for HDR television. This is because current work flows in HDR video are also following the display referred paradigm.

Traditionally, SDR content has been graded to 100 nits and have been targeted for displays with similar peak brightness. For HDR video content providers professionally grade their HDR content on a mastering display ranging from 1000 to 4000 nits. Potential mastering displays that could be used are the Dolby Pulsar, Sim2 HDR47 or the Sony BVM X-300. The mastered content would be distributed to consumer displays ranging from 500-1000 nits (recall Section 2.2). These displays would be required to implement a TMO to compress the dynamic range. Furthermore, with increasing trends in peak brightness and lower black levels, we can expect future mastering displays with higher dynamic range. This would also mean that over a period of time a given video library would have HDR content graded at 1000 nits, 4000 nits or more and legacy video graded at 100 nits. When this content is transmitted to a consumer HDR TV of a given brightness, the display compresses or expands the dynamic range of the content to exploit the capabilities of the display. This eventually distorts the content to the extent that the viewer experiences the content differently for a given HDR display. This is a major issue especially when preserving artistic intent of the original content and hence underlines the need of good display retargeting.
To the best of our knowledge, no study has been presented comparing reproduction of dynamic range between HDR displays. In order to retarget between HDR displays, we are required to implement both tone mapping and tone expansion. We define different combination of these processes in Table 4.1. In addition to this, few studies have considered the significance of aesthetics in tone reproduction in HDR imaging. Therefore in this chapter, we set out to identify basic tone reproduction techniques in order to retarget between HDR displays taking into account artistic intent and put in context of the complex HDR video ecosystem.

<table>
<thead>
<tr>
<th>Display</th>
<th>SDR Legacy Display</th>
<th>1000 nits Display</th>
<th>4000 nits Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDR Legacy Video</td>
<td>—</td>
<td>Tone Expansion</td>
<td>Tone Expansion</td>
</tr>
<tr>
<td>1000 nits Video</td>
<td>Tone Mapping</td>
<td>—</td>
<td>Tone Expansion</td>
</tr>
<tr>
<td>4000 nits Video</td>
<td>Tone Mapping</td>
<td>Tone Mapping</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 4.1 Tone Reproduction Methods: This table explains tone compatibility terminologies between content and displays. The lines “—” represent no compatibility issues, as content has been mastered for this display. Tone Mapping is performed to reduce dynamic range of 1000 or 4000 nits content on SDR displays. Tone Expansion is used to expand the dynamic range of SDR content for HDR displays. The focus of this chapter is exploring Tone Mapping and Tone Expansion methods highlighted in blue for tonal compatibility between HDR displays.

The aim of this study is to extend the work of the previous chapter in combining aesthetics and perception by evaluating basic retargeting methods between HDR monitors. The evaluation consists surveying simple tone mapping methods for 4000 nits color-graded content for 1000 nits HDR displays and also investigating tone expansion strategies when content graded in 1000 nits is displayed on 4000 nits monitors. We evaluate these tone reproducing methods on content of various lighting styles, which we have already studied, such as low key, medium key and high key. Based on the results of these experiments we identify critical issues that need to be resolved for accurate retargeting between HDR displays.

This chapter is organized as follows. Section 4.1 presents a relevant background on retargeting between displays. A description of simple retargeting experiments are exposed in Section 4.2. We explore $\gamma$ correction for dynamic range conversion in Section 4.3 and conclude in Section 4.4.
4.1 Background

This section discusses related work on dynamic range tone reproduction specific to this topic of retargeting between HDR displays. We briefly describe the experiments that we will conduct and source content used.

We have explored different methods in tone mapping and tone expansion in Section 2.4 and 2.5 respectively. In the case of retargeting between HDR content and HDR displays both with different dynamic range we look for common ground in existing methods in tone mapping and expansion. It must be noted that we are not compressing or expanding high dynamic range ratios seen in previous studies but dealing with medium dynamic range compression or expansion.

![Tone Reproduction Operator](image)

Fig. 4.1 A Tone Reproduction Operator using a forward and reverse transform.

Reinhard et al. [178] first anticipated the current day scenario with compatibility issues between HDR displays. This work suggested a tone reproduction framework consisting of a forward and reverse transform as seen in Figure 4.1 similar to the ones used in color appearance models [66]. The authors in [178] define the forward transform using a sigmoidal tone curve based on the photoreceptor model:

\[
V = \frac{L_i^n}{L_i^n + (\bar{L}_i/k)^n}
\]  

(4.1)

where \( V \) is the perceived luminance, \( L_i \) is the input display luminance, \( \bar{L}_i \) is geometric average of the image, \( n \) is a constant which is either fixed or set differently per image and \( k = 0.18 \) is a user parameter set by the authors (which corresponds to the key value of the scene). It should be noted that any operation done on the signal \( V \) is effectively done in the Tone Reproduction Operator.
Space as seen in Figure 4.1. To map this signal $V$ onto any given display system the reverse transform has to be defined as follows:

$$L_o = \frac{V \times L_{i,mean}}{1 - V}$$

(4.2)

where $L_o$ is the output display luminance and $L_{i,mean}$ is the mean of the output display luminance. This study concluded that applying a sigmoid in the forward and backward step amounts to a $\gamma$ correction. This can be proved by substituting Equation 4.1 in Equation 4.2 as shown:

$$L_o = \frac{L_i^n}{L_i^n + (L_i/k)^n} \times L_{i,mean}$$

(4.3)

$$L_o = \frac{L_i^{n/m}}{(L_i/k)^{n/m}} \times L_{i,mean}$$

(4.4)

$$L_o = c \times L_i^{n/m}$$

(4.5)

where $c = \frac{L_{i,mean}}{(L_i/k)^{n/m}}$ is a constant and $m$ is an adjustable display related parameter. Equation 4.5 makes a solid case for using $\gamma$ correction to map between HDR displays. This also encourages to experiment with other simple methods to retarget between HDR displays. In this chapter, we will be conducting experiments for retargeting between two HDR monitors: a 4000 nits Sim2 HDR47 display and a 1000 nits Sony BVM X-300. The Sim2 HDR47 has a higher peak brightness while the Sony has deeper black levels. Both display have a contrast ratio of almost a million to one. The different characteristics of the display technology makes it an interesting case-study in order to investigate retargeting methods between both panels. Taking this into account, we will consider three simple HDR retargeting techniques:

- Clipping: a method to tone map 4000 nits content on to a 1000 nits display
- Remapping: a method to map 1000 nits content to 4000 nits display
- Gamma as a dynamic range converter: use $\gamma$ correction to map between displays

From the Chapter 3, we learn that using stylized content is a good metric for evaluating tone reproduction especially to see if it preserves artistic intent. In this experiment, we only exploit 3 classes of lighting styles notably: low key, medium key and high key. This is because there is limited high quality stylized HDR video content available in bright key and dark key styles. We used medium key and low key content from the MPEG test set [202] and some high key content from Froehlich et al. [75]. The 1000 nits content have been extracted
from a footage taken during the French stage of the Volvo Ocean Race. This was done under a collaborative project by French Multimedia companies [166] [30]. It should be noted that for a given content we did not have both 4000 nits and 1000 nits grade.

4.2 Simple Retargeting Experiments

In the following sections we touch upon some basic methods of tone reproduction. The first experiment looks into clipping the signal in the case of tone mapping from 4000 to 1000 nits content. The second experiment explores remapping 1000 nits content into a 4000 nits container for tone expansion.

Fig. 4.2 Clipping for Medium dynamic range compression leads to loss in detail of bright regions. From this figure we see loss in detail: (a) in the flame in low key image (b) along the walls and doors in medium key image and (c) face texture and clothing in high key image. For the high key image, the certain regions seem to be blended with the background.
4.2.1 Clipping for Tone Mapping

When working in an HDR post production environment we often observe that content creators do not necessarily exploit the entire dynamic range available to them. Even when grading on 4000 nit monitors, we identify that artists don’t master indoor scenes in HDR any differently than in SDR. In the case of outdoor scenes, they grade bright whites more brighter than indoor scenes and also preserve a number of specular highlights. However, the major bulk of the scene’s histogram is still SDR. It could be argued that reduction of highlights in the form of clipping can suffice for medium dynamic range compression. From our observation we learn that clipping an image is very much dependent on the type of scene. For a low key scene clipping from 4000 to 1000 nits does make sense since most of the content is within the 1000 nits containers. However, for medium key or high key scenes clipping results in loss of details and sometimes large over saturated regions. Figure 4.2 shows the effect of clipping on different stylized content. From this we can deduce that clipping for medium dynamic range compression does not seem like a viable solution especially for brighter lighting styles.

4.2.2 Remapping for Tone Expansion

The UHD forum guidelines for HDR broadcast [206] introduces the notion of remapping SDR content into an HDR container. Remapping is the process of repackaging the SDR content in HDR containers without changing the color gamut or the dynamic range of the content. This permits content providers to preserve artistic intent as the pixels are mapped to the equivalent color and luminance values. This also avoids the step of tone expansion which is not preferred by all artists. We attempted a few tests while remapping 1000 nits content on a 4000 nits SIM 2 monitor. In terms of artistic intent the video remains unchanged. However, when the video is played sequentially in between two 4000 nits videos this results in poor user experience as explained in Figure 4.3. Hence, it’s likely that remapping may not be the ideal solution to preserve artistic intent as display manufacturers would like to have a constant user experience which requires to adapt the content to tonal range of the display.

4.3 Gamma as a Dynamic Range Converter

From Section 4.1, we learn that a $\gamma$ correction has is simple method used in tone mapping and tone expansion. Here we explore it to use it as a dynamic range converter in the form below:

$$Y_{Target} = L_{max} \times Y_{Master}^\gamma$$  \hspace{1cm} (4.6)
Fig. 4.3 Playing videos sequentially from left to right, it is observed that remapping medium or lower dynamic range content along with native HDR content results in a poor user experience.[82].

where $L_{Target}$ is target luminance and $L_{Master}$ is mastering luminance, $L_{max}$ is maximum target display luminance and $\gamma$ is the dynamic range conversion value. The target display can be of 4000 nits or 1000 nits and vice versa. Same for the mastering display. The $\gamma$ value controls this conversion process. We may recall that the work presented in Chapter 3 shows that tone reproduction could take the form of an adaptive gamma correction which is content dependent. From MPEG, contributions by Lasserre et al. [118] also describes the HDR grading process as power function for display referred tone reproduction that is very much content dependent. This Section looks into gamma correction as a viable tone reproduction operator using lighting style as an indicator of content dependence.

### 4.3.1 Experiments

For this experiment, we follow an analogous methodology presented in Chapter 3. Our aim to answer a similar question: *At what gamma value does the output video best represent the style of the input video?* This study is in two parts. The first part tests tone compression with 4000 nits input video and 1000 nits output. The second examines tone expansion with 1000 nits input video and 4000 nits output. In both studies, the user is requested to choose a gamma value to get a style match.

The test set-up is also akin to Chapter 3. We use the ITU-R recommendations BT.500 [97], in a dark environment of approximately 25 lux. The content was of LK, MK and HK lighting styles. The resolution was 1080p and we considered a variety of frame rates up to 24-60 fps. Short videos up to 5-10 seconds were used. Since we hope to get information
about the choice of $\gamma$, we conducted tests with the same video experts who participated in the experiments in Chapter 3. We preferred conducting the test sequentially instead of side by side since the brighter display tends to derive more attention from the user. It also prevents them from using the $\gamma$ correction for brightness matching which is not the goals of the test. For example, in the case of tone mapping, we first display the original 4000 nits content on the Sim2 HDR47 display, followed by a $\gamma$ corrected version on the 1000 nits Sony display. After viewing several $\gamma$ corrected sequences the user chooses the best gamma value at which they style of the original content was preserved. The user can repeat the test several times for a particular sequence until s/he is satisfied with an appropriate $\gamma$ value.

4.3.2 Results

For tone compression tests we allow the user to choose $\gamma$ values between 0.5 to 1.2 while for tone expansion the user has the choice between 0.8 and 2.2. These gamma values increment on a step size of 0.1. The results of both experiments can be seen in Figure 4.4. For tone compression, we observe that when moving from low key to high key there is a tendency for the $\gamma$ value to increase. While for tone expansion, we observe the $\gamma$ value decreases as the brightness of the lighting style increases. Both results strongly suggests that for ideal tone reproduction, the $\gamma$ value is pretty much style dependent.

One direct observation we can make from these graphs is when the $\gamma$ choice is 1, which essentially amounts to linear scaling. For majority of the content we see that $\gamma = 1$ is not chosen. In the case of tone compression, only a few user preferred linear scaling for HK video, while most users preferred a $\gamma$ correction. This is because a simple linear scaling quantizes small perceptual differences to the same value on the display resulting in loss of detail. In addition to this, viewers claimed a linear operator was too dim in our tone mapping studies especially for LK and MK content. For compressing the dynamic range of HK content, majority of the scene is very bright and there is very little dark regions thus the loss of detail caused by linear scaling is relatively low as opposed to LK and MK video. Thus for a HK video, $\gamma \approx 1$, could be acceptable.

Similarly, the tone expansion results also shows no preference for linear scaling. Viewers claimed LK and MK content looked too bright and creative intent was clearly lost. This explains the deceeding trend in $\gamma$ as the brightness of the style increases and is also in accordance with results of the previous chapter. This is also contradictory to the work by Masia et al. [135] which suggests to increase $\gamma$ value for over-exposed or brighter images. Other interesting findings of 1000 to 4000 nits tone expansion of HK content was a tendency amongst users to use higher gamma values in an attempt to reduce the brightness of the scene as many found 4000 nits HK video uncomfortably bright. This could be linked to the
same perceptual phenomena that caused users to choose higher $\gamma$ in the works by Masia et al. [135]. Furthermore, we observed that users preferred a dimmer HK sequence as opposed to style preserving 4000 HK nits video. This is in contradiction with works by Hanhart et al. [82] which claims that most viewers prefer content with higher peak brightness. Our tone expansion study suggests that a brighter video is not always preferable to the user, especially in HK lighting style.

For both studies, we observe that for all three styles there are cases where there is an overlap in $\gamma$ values suggesting a single $\gamma$ value can work on different lighting styles. The
main reason for this is lack of number of data points to give statistically plausible results. Lack of data points also suggests that there is very little stylized HDR content available and artist are yet to get comfortable with grading LK and HK content for HDR. Furthermore, the ideal case would be to have the same content for graded in 1000 nits and 4000 nits to find a $\gamma$ correction technique that is invertible. This would mean modeling a single equation that could map content of a given dynamic range on to a display of any dynamic range.

### 4.4 Conclusion and Future Works

Previous works on HDR imaging deals with converting SDR to HDR or HDR to SDR, however there is limited work on tone compatibility between HDR displays. This chapter is based on existing literature and industrial trends in tone mapping and tone expansion and sets out to find plausible tone reproduction techniques between HDR Displays. We take into account lighting style aesthetics as a testing parameter for stylized content. We found that basic methods such as clipping and remapping do not work as well as $\gamma$ correction. We conducted an expert user study using a $\gamma$ correction approach and found the choice of gamma strongly depends on the style of the content for best tone reproduction. Through this user study we also saw that linear scaling is not appropriate for HDR retargeting and a form of $\gamma$ correction is often preferred by most users. We summarize our finding in table 4.2.

Retargeting between HDR displays is an important issue to be considered for both content providers and display manufacturers in the near future. Our user studies show that best tone

<table>
<thead>
<tr>
<th>Tone Reproduction Operation</th>
<th>Tone Mapping 4000 nits to 1000 nits</th>
<th>Tone Compression 1000 nits to 4000 nits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clipping</td>
<td>—</td>
<td>Loss of detail in bright regions especially for MK and HK styles</td>
</tr>
<tr>
<td>Remapping</td>
<td>Poor user experience when played sequentially along side Native HDR content</td>
<td>—who was not preserved. Also, users found 4000 nits HK content an uncomfortably bright experience.</td>
</tr>
<tr>
<td>Linear Scaling</td>
<td>Loss in detail for LK and MK content. LK and MK content looked dimmer and creative intent was lost.</td>
<td>LK and MK content looked brighter and style was not preserved. Also, users found 4000 nits HK content an uncomfortably bright experience.</td>
</tr>
<tr>
<td>Gamma Correction</td>
<td>Different $\gamma$ values for different styles in tone reproduction can best preserve lighting style aesthetics.</td>
<td>—</td>
</tr>
</tbody>
</table>
reproduction technique could be simply done using a $\gamma$ operation. To successfully reproduce artistic intent, we performed a scene by scene user-controlled gamma correction for tone reproduction. Although this method is time consuming and requires extensive man power, more effort is required to resolve this problem. For real-time applications, we find that the current state-of-the-art techniques for retargeting between HDR displays are not sufficient. For future works, it is recommended to find real time solution based on modeling a temporally coherent $\gamma$ correction curve. This model should preferably be invertible for tone expansion and compression. Furthermore, it would be ideal to extend these tests with same stylized content graded at 4000 nits, 1000 nits and possibly in SDR as well.

In conclusion, this chapter combines aesthetics and perception to shed light on the topic of retargeting between HDR displays. This study also points to an important direction of visual comfort which has not been well-addressed in HDR imaging. Viewing tone expanded HK content on a 4000 nits was described as visually unpleasant by most users. This poses an interesting research problem and is the basis of the next chapter.
Chapter 5
Comfort-based HDR Retargeting

Statistical studies on multimedia trends show that on an international average, people spend around 400 minutes every day looking at screens [141]. This includes consuming content in the form of watching TV, working on a computer monitor, spending time on smartphones and tablets, etc. Most content viewed on screens is not necessarily camera-captured content but may also contain computer graphics and text, often mixed with camera content. Graphics-based screen content is often characterized by no sensor noise, repeating patterns and few saturated colors [218]. Screen content may not have artistic intent in the aesthetic sense (as defined in Section 2.6.1). However, it is used for personal browsing, advertisements, to communicate information for news feeds or even to convey a corporate message. With the arrival of new display technology, adapting existing screen content puts forth a new challenge especially since previous work on display retargeting mostly focuses on camera-captured content or natural images.

Amongst next generation displays, it is widely accepted that HDR displays are the most valuable in terms of Quality of Experience (QoE) [154]. However, none of the existing tone expansion methods detailed in Chapter 2 deal with the case of screen content nor consider the visual comfort aspect of QoE.

In this chapter, we address the issue of brightness comfort in HDR imaging more in terms of perception than purely considering aesthetics. We explore the human preference for brightness, specific to tone expansion, and propose a novel brightness control model. As already mentioned, state-of-the-art methods including the EO presented in Chapter 3, scales the SDR content to the maximum luminance provided by the display. Through our experiments in Chapter 4, we have also observed that scaling often leads to an uncomfortably bright image, especially when applied to high key content. We can assume this would result in an even worse experience for screen content which may have a completely white background. In this chapter we put into evidence that, when retargeting legacy content, brightness scaling is
highly content dependent. We propose a brightness control model that attempts to mitigate brightness discomfort. Our objective is to improve upon existing expansion techniques with a new method of brightness control. In summary, our contributions are:

- A user study giving insight on human preference for brightness when looking at HDR content,
- a brightness control method for retargeting screen content that adheres to human visual comfort,
- a design of a novel feature that facilitates brightness retargeting for HDR monitors.

This chapter is organized as follows. Section 5.1 reviews the related works in psychophysics and brightness control. In Section 5.2 we present the subjective evaluation methodology. Section 5.3 details the analysis of the results and the brightness model. Finally in Section 5.4, we conclude the chapter.

5.1 Background

This section introduces the existing literature on perceptual studies related to brightness comfort and bright color perception. We also discuss methods of brightness control and standardization efforts in brightness regulation.

5.1.1 Perceptual Studies

Various subjective experiments, in the literature, have considered peak brightness as a user preference. In [81], Guterman et al. suggest user preference plateaus as maximum luminance level increases. This trend is confirmed by Hanhart et al. [82] for HDR displays claiming that human preference increases logarithmically with the increase in peak brightness of the HDR content displayed. Both studies use natural images or cinematic contents in their experiments and do not consider screen contents. Freysinner at al. [74] have conducted a study on brightness preference on signage displays reporting that the signage panel brightness was dependent on surround lighting. Similarly, Mantel et al. [128] have explored the impact of peak white and ambient light on video quality and proposed test cases for adapting both factors.

Visual disturbance is an important perceptual factor when considering brightness preference. The World Health Organization (WHO) defines visual disturbance as a high degree of visual discomfort typically occurring after prolonged visually intense activity, and the
5.1 Background

Symptoms include fatigue, pain around the eyes, blurred vision, dry eyes or headache [27]. Among studies in HDR imaging, Rempel et al. [183] have shown that viewing HDR content, even in dark environments, may not lead to visual fatigue. Na and Suk [150] have investigated visual disturbance caused by viewing content on mobile phones in low light conditions. They separate brightness comfort into two parts: 1) initial viewing comfort (reaction to brightness by turning on display) 2) continuous viewing comfort (reaction to brightness after prolonged content viewing). In this chapter, we focus on initial viewing comfort of screen content on HDR displays, and thus not consider modeling visual fatigue caused by brightness discomfort in continuous viewing.

Various studies in color science have attempted to model our perception of chromatic objects including highly saturated colors. The most pertinent psychophysics phenomena related to bright color perception is the Helmholtz-Kohlrausch effect which states that the perceived lightness increases as the chroma increases for colors of equal luminance [56]. In 1957, Sanders and Wyszecki [187] [216] described a bright color perception model by matching color patches with achromatic patches. Fairchild and Pirrotta [69] presented a rigorous experiment modeling this effect as an extension to CIE-Lab lightness channel also known as \( L^{*} \). As far as we know, this phenomenon has not been explored to HDR displays or highly saturated colors in screen content. The Helmholtz-Kohlrausch effect, as seen later on, will also play an important role in our experiments involving brightness preference.

5.1.2 Brightness Control

Automatic brightness control is a prominent feature in most modern display technology today. Merrifield and Silverstein [143] define the general method of brightness control depending on the ambient light of the surroundings. This method has been improved over the years. More recently, Schuchhardt et al. [189] proposed a context-aware approach for mobile phones where screen brightness is determined as a function of ambient light, user preference, location, sun angle, battery level etc.

Brightness control is strongly correlated with TV power consumption. The International Electrotechnical Commission (IEC) have published the standard IEC 62087 [90] for monitoring display energy consumption and signifies importance of brightness control. With the introduction of HDR displays, power or brightness management has become even more necessary. To this effect, the Society of Motion Pictures and Television Engineers have released SMPTE ST 2086 [196] which defines mastering metadata for HDR video including Maximum Content Light Level (MaxCLL) describing maximum brightness and Maximum Frame-Average Light Level (MaxFALL) describing the maximum value of the frame-average luminance of the content. These parameters are applicable to control the brightness of HDR.
Comfort-based HDR Retargeting displays on the consumer end [226]. Efforts in the Green Metadata standard amongst MPEG specifies methods for reduction of power consumption on consumer devices in the form of brightness reduction [58]. However, to the best of our knowledge, no existing guidelines specify the management of excessively high brightness. This makes it a challenge for tone expansion and encourages the exploratory aspect of this study on HDR brightness preference.

5.2 Subjective Evaluation

The aim of our study is to investigate human brightness preference considering brightness comfort for special case of retargeted HDR screen content. The independent variable is the screen content and the measured variable is the brightness preference of the user. Through this experiment, we intend to answer the following question: what is the preferred brightness in terms of initial viewing comfort for a given retargeted screen content? To this end, the user has been presented with screen content images on an HDR display and was instructed to vary the brightness level with a graphical slider. Details of the experiment such as stimuli, pre-processing, participants, apparatus and procedure are described in this section.

5.2.1 Stimuli and Pre-Processing

The dataset for this experiment contains 47 screen content images. This content varies significantly in appearance, style and image statistics. This included signage images, computer graphics, text documents, web browsers, desktop wallpapers, powerpoint slides, natural images with text, etc., as seen in Figure 5.1. The stimuli used in this experiment and the processed HDR images are available upon request to IRT b<>com. All images are in 8 bit Full HD (1920x1080) resolution and in sRGB colorspace. We have processed these 8 bit images in this order: de-quantizing, converting to double precision, linearizing by applying an inverse gamma and extracting the SDR luminance component $Y_{SDR}$. Recall the previously defined tone expansion formulation below:

$$Y_{HDR} = L_{Max} \times Y_{SDR}^{\gamma} \quad (5.1)$$

where $Y_{HDR}$ is the HDR luminance, $L_{Max}$ is a brightness scaling factor and $\gamma$ is a style adaptive parameter used in the expansion operator provided in Chapter 3. Finally, the resulting $Y_{HDR}$ is used to determine the RGB values in HDR domain by subsequently performing an operation...
5.2 Subjective Evaluation

![A subset of 9 screen content images used in the subjective experiments.](image)

Fig. 5.1 A subset of 9 screen content images used in the subjective experiments.

of luminance replacement:

\[
\begin{bmatrix}
R_{\text{HDR}} \\
G_{\text{HDR}} \\
B_{\text{HDR}}
\end{bmatrix}
= Y_{\text{HDR}} \begin{bmatrix}
R_{\text{SDR}} \\
G_{\text{SDR}} \\
B_{\text{SDR}}
\end{bmatrix}
\]

(5.2)

For this experiment, we vary the \( L_{\text{Max}} \) from 100 nits to 4000 nits with a step size of 100. It should also be noted that we have kept the saturation value used in Chapter 3 set to \( s = 1 \) in these experiments.

5.2.2 Participants and Apparatus

10 males and 6 females have participated in the experiment. The average age of the users was 24.2 and ranged from 19-48 years. 9 had previously seen HDR videos. 6 wore glasses and all have normal color vision. The experiments were conducted on the SIM2 HDR47 display, which has Full HD resolution, peak brightness of 4000 nits and contrast ratio higher than \( 4 \times 10^6 \). The test has been conducted in a room within ambient luminance of approximately 5 lux and observers were positioned at a distance 3 times the height of the display. In this study, we purposely use a dark surround environment as it represents the worst case scenario, since the human visual system is more susceptible to brightness discomfort in a dark setting.
5.2.3 Procedure

The experiments began with a 5-10 minute introduction to the participant to HDR technology with few HDR images and videos samples. This step also permitted their eyes to adapt to the surround lighting conditions. This was followed by a short training session using images similar to the ones presented in [87]. The images displayed were scaled to $L_{\text{Max}} = 4000$ nits and the user was asked if the image was visually comfortable or too bright. If the image was too bright, the participant was requested to re-adjust the brightness to a comfortable setting and if not, they were asked to move on to the next image. Our choice for using this process is based on pilot studies, starting at $L_{\text{Max}} = 100$ nits or $L_{\text{Max}} = 1000$ nits did not create the necessary initial visual discomfort as compared to $L_{\text{Max}} = 4000$ nits. Once the user was comfortable in this methodology, they were asked to repeat the same with the testing screen content dataset.

The preferred brightness was selected by the user using a graphical slider representing an unlabeled continuous scale of 100 – 4000 nits with a step size of 100 as mentioned in Section 5.2.1. This slider was placed on the bottom middle part of display, ensuring it had no impact on user’s vision. The knob of the slider was set to the extreme right ($L_{\text{Max}} = 4000$). The participants had to shift the knob to the left to reduce the brightness to their preference, after which they were instructed to click on the next button that recorded their chosen value and also allowed them to move on to the next image. A black image was placed in between images for a waiting time of 5 seconds, in order to simulate effective initial viewing comfort when viewing the retargeted image. Users were given around 10 seconds to choose their preferred brightness, however most users made their choices much quicker. Finally, we have concluded the study by asking the participants to fill in a questionnaire regarding visual comfort as in previous studies [47] [183].

5.3 Results Analysis

In this section we analyze the results of the experiment, develop a model for preferred brightness using a new feature and validate the model.

5.3.1 Analysis of Experiments

The boxplot in Figure 5.2 shows the dispersion in preferred brightness. It should be noted that we represent preferred brightness (measured in absolute luminance) in the logarithmic domain. We observe that, for some content, setting $L_{\text{MAX}}$ to the maximum brightness of the target display is visually comfortable while for others it is not the case. There was a
tendency amongst users to reduce the brightness of the display when the source content had large regions of white pixels or highly saturated bright colors. This phenomena is a direct result of the Helmholtz-Kohlrausch effect for bright colorful pixels on the SIM2 display. A limitation of the SIM2, is that if more than 40% of the back-light LEDs of the SIM2 display are at full power, the global power of all LEDs is lowered and for a full white image, the peak brightness is only 2300 nits. In such scenarios, we observed that the users reduced the peak brightness far below this limit. For example, in Figure 5.1e, on average the preferred brightness was 1700 nits. There were also cases where the users inquired if brightness could be scaled beyond the 4000 nits limit. These images are mostly the left plateau region of plot in Figure 5.2. The result of the questionnaire on visual comfort differed significantly from [183] as users perceived uncomfortable sensations such as burning/pricking sensation in the eyes, tearing/watery eyes and pain around the eyes. This result highlights the importance of brightness control for retargeting screen contents.

5.3.2 Modeling Brightness Preference

The previous section shows that scaling to peak brightness of the display is not always preferred. We believe that in order to control this brightness, $L_{\text{MAX}}$, we can derive information from the SDR image statistics. A variety of statistics have been considered in order to model our data of preferred brightness. This includes first order statistic of the luminance of the image such the mean, variance, skewness and kurtosis. HDR based first order statistics using the image key [181] has been taken into account. Furthermore, color related statistics such as lightness and colorfulness [83] have also been considered. Our first step was to compute a regression between preferred brightness and one of these predictors. Unfortunately, none of the regressions tested using a single variable resulted in $R^2 > 0.65$. The next approach, consisted of finding the best combination of variables, that provide the best fit in the form

<table>
<thead>
<tr>
<th>Table 5.1 Results of Visual Discomfort Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom</td>
</tr>
<tr>
<td>Double vision</td>
</tr>
<tr>
<td>Problems in focusing</td>
</tr>
<tr>
<td>Burning/pricking sensation in the eyes</td>
</tr>
<tr>
<td>Tearing/Watery eyes</td>
</tr>
<tr>
<td>Pain around the eyes</td>
</tr>
<tr>
<td>Headache</td>
</tr>
<tr>
<td>Image floating</td>
</tr>
<tr>
<td>Color change</td>
</tr>
</tbody>
</table>

1 = I did not perceive this symptom - 5 = I perceived this symptom a lot
Fig. 5.2 A boxplot diagram of the results. Black line: 5% truncated mean, Red horizontal lines = median values; Blue boxes = inter quartile range; Dotted whiskers: adjacent values; red crosses = outliers. It can be observed that setting $L_{MAX}$ to the maximum luminance of the target display is not always preferred for brightness comfort.

Based on our observations, we have noticed that the user brightness preference decreases as number of bright pixels on the HDR monitor increases. We define the term bright pixels as a combination of white pixels and pixels with a high colorfulness value. It must be noted that bright color perception has to be taken into account as a direct correlation with the Helmholtz-Kohlrausch effect. We propose a feature that enables us to find the ratio of $Bright\text{Pixel}_s$ with respect to the total number of pixels, as defined in Equation 5.3:

$$Bright\text{Pixel}_s = \frac{1}{N} \sum_{k=1}^{N} White\text{Pixel}_s[k] + Colorful\text{Pixel}_s[k]$$  \hspace{1cm} (5.3)

where $N$ is the total number of pixels and $k$ represents the pixel coordinates $(i,j)$. In order to get to Equation 5.3, we follow a simple approach based on our knowledge of color appearance models. The process begins with converting RGB SDR images into the HSV colorspace and separating the $S[k]$ (saturation) and $V[k]$ (value or lightness) color channels. Although HSV is not a perceptually linear color space, it is chosen mostly because of its fast computation. Any other perceptually linear color space that can separate saturation and lightness, e.g. CIE-Lab,
can also be considered. Using the $S[k]$ and $V[k]$, we determine the Chroma of the image as follows:

$$V_T[k] = \begin{cases} 
1 & \text{if } V[k] > \alpha \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (5.4)

$$Chroma[k] = S[k] \cdot V_T[k]$$  \hspace{1cm} (5.5)

The operation in Equation 5.5 is analogous to the calculation of the Chroma used in color appearance models which is defined as the product of saturation and lightness in [66]. We purposely threshold the lightness channel using $\alpha = 0.8$ to preserve the brightest part of the image. Finally, to determine the $White_{Pixels}[k]$ and $Colorful_{Pixels}[k]$, we can segment the Chroma in the following manner:

$$White_{Pixels}[k] = \begin{cases} 
1 & \text{if } Chroma[k] < \beta \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (5.6)

$$Colorful_{Pixels}[k] = \begin{cases} 
1 & \text{if } Chroma[k] > \theta \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (5.7)

where $\beta < \theta$. We have chosen the values $\beta = 0.1$ and $\theta = 0.8$. The values of $\alpha$, $\beta$ and $\theta$ have been set empirically based on visual observations. These parameters could be adjusted

Fig. 5.3 A few examples of the segmentation results are shown. The white regions in the second row are identified as the brightest regions which cause visual discomfort. Brightness of the pixels can be represented as a percentage through the operation: $Bright_{Pixels} \times 100$. These values for the images above are: (a) Weather 14% (b) Basketball 58% and (c) Ski 29%. Clearly, images with a large number of bright pixels should be scaled appropriately.
depending on the choice of the colorspace. We can visualize the effect of applying the bright pixels feature on the entire sRGB gamut in HSV and CIE-Lab colorspace in Figure 5.4. Figure 5.3 shows the bright pixel feature used on screen content images.

We model our preferred brightness data to the proposed bright pixels feature of the image. This yields our brightness control algorithm defined with the equation below:

\[
\log_{10}(L_{\text{Max}}) = -0.2118 \times \log_{10}(\text{BrightPixels}) + 3.28 \tag{5.8}
\]

The \( R^2 \) measure of goodness of fit is \( R^2 = 0.9267 \). Figure 5.5 shows this equation in a graphical form. We can generalize Equation 5.8 to absolute luminance by the expression:

\[
L_{\text{Max}} = \frac{L_{\text{Critical}}}{\text{BrightPixels}^\mu} \tag{5.9}
\]

where \( L_{\text{Critical}} \) is the minimum preferred absolute brightness for a very bright content (when \( \text{BrightPixels} > 0.95 \)) and \( \mu \) is a decay factor. Our experiments on the SIM2 HDR display resulted in \( L_{\text{Critical}} = 1905 \) nits and \( \mu = 0.2118 \). It should be noted that we permit our model to predict preferred brightness beyond the measured values (\( L_{\text{Max}} > 4000 \) or \( \log_{10}(L_{\text{Max}}) > 3.6 \)). This is because we believe certain images can be perceived as visually comfortable.

Fig. 5.4 We can visualize the sRGB gamut on the HSV and CIE-Lab colorspace in (a) and (c) respectively. Applying the bright pixels feature on the sRGB gamut in (b) HSV and (d) CIE-Lab colorspace identifies the white and highly saturated colorful pixels that can potentially cause visual discomfort on high brightness displays.
5.4 Conclusion and Future Works

In this chapter, we proposed a brightness control method for retargeting legacy contents on to HDR displays. This method was developed with a help of a perceptual experiment evaluating user brightness preference. We intentionally have focused on screen content as stimuli due to their unique properties in image statistics. The experiment resulted in emphasizing that brightness preference is strongly content dependent. We found that existing image statistics beyond the 4000 nits limitation of the SIM2 HDR monitor. In order to avoid $L_{Max} \to \infty$ for the case where $Bright_{Pixels} \to 0$, we clip $Bright_{Pixels}$ within the range $[0.001, 1]$. Finally, our model for $L_{Max}$ defined in Equation 5.9 could be substituted into the original global operator in Equation 5.1, resulting in a method for tone expansion with brightness control as follows:

$$Y_{HDR} = \frac{L_{Critical}}{Bright_{Pixels}^\mu} \times Y_{SDR}^\gamma$$  \hspace{1cm} (5.10)

Fig. 5.5 Brightness Preference as a function of $Bright_{Pixels}$.
were ineffective to model brightness preference and thus designed our own feature. This feature is based on a simple segmentation scheme that identifies the bright pixels in the image that may cause visual discomfort. Based on this metric, we have proposed a brightness control algorithm that adjusts luminance of display depending on the amount of bright pixels in the content.

Our work is an effort towards improving QoE in terms of brightness comfort for SDR to HDR display retargeting. However, we feel this study is not only applicable for brightness control in tone expansion. As mentioned previously, most existing HDR video content is camera captured and little is known on generating HDR graphics for HDR TV sets. This work gives insight in creating HDR screen contents. Particularly, the segmentation technique, which identifies tentative regions that cause visual discomfort, could potentially assist content producers. This study also brings out the need of brightness regulation for avoiding excessive use of high brightness. We could envisage a future brightness model in the lines of existing loudness models in audio. Lastly, this work could also be used to devise new brightness comfort based metrics for QoE.

Given the novelty of this field of research, there is ample opportunity and requirement for future work. A limitation of our study is that we only consider initial viewing comfort. We plan to extend our work to a continuous viewing comfort and develop a brightness control model that alleviates visual discomfort as well as fatigue. Extending this model for continuous viewing also requires adapting to the temporal aspect of video, thus the final model should be free of temporal artifacts. A change in brightness also affects the color appearance, which is another important factor to be considered. For future works, an experiment on the modeling of the Helmholtz-Kohlrausch effect on HDR displays can also be considered. Moreover, our brightness control feature only counts the number of bright pixels without considering spatial distribution of these pixels (i.e. dispersed or contiguous). This is a promising direction for future improvements. Furthermore, this work has been performed in a dark testing environment. A more rigorous evaluation calls for testing in different ambient lighting conditions for a more adaptable brightness model. It would also be interesting to explore whether our methodology could be extended to other display systems such as smartphones and tablets. Additional research could also be conducted amongst different age groups. It is well known that brightness comfort changes with age and it would be interesting if different targeting models are required for younger and older users. Finally, this chapter deals with controlling brightness as a function of visual comfort. In real world scenarios, the display controls brightness based on energy consumption, image aesthetic quality and ambient lighting of surround. An ideal brightness model for displays should consider all these aspects as well.
Chapter 6

Exploring HDR Retargeting for Omnidirectional Imaging

Omnidirectional or 360° imaging is an emerging format in the field of immersive multimedia. 360° cameras allow the capture of the entire field of view that covers a full sphere. This content is usually visualized using near-eye Head Mounted Displays (HMD) allowing the viewer to freely change the direction of their sight across the omnidirectional scene. However, to the best of our knowledge, most omnidirectional systems today are based on SDR imaging systems i.e. captured with a single exposure and viewed on a display with limited dynamic range. This makes current 360° imaging more vulnerable to the limitation of SDR systems especially from loss of information due to under and over exposure, as we have seen in Section 2.1. This in turn could lead to a less immersive experience. HDR imaging seems to be a potential technology to resolve this issue. Therefore, in this chapter, we set out to identify whether there is a qualitative benefit of applying existing HDR retargeting methods the current SDR 360° imaging pipeline. We focus on the perceptual factors involved with omnidirectional HDR retargeting and provide insight on potential aesthetic retargeting for future work.

To investigate this, we generate our own omnidirectional HDR images with a commercially available 360° camera using classical techniques based on multiple exposure bracketing [126]. As the 360° HDR images have a higher dynamic range than that of typical head mounted display, it requires the additional retargeting step of tone mapping for the HMD. We shall consider well known off-the-shelf tone mapping operators for this work. The basis of this study is a qualitative experiment comparing various tone mapped images with a single exposure 360° reference image. In order to conduct a rigorous evaluation, a novel dual stimulus methodology designed specifically for pairwise comparison on an HMD is proposed. The conclusion of the experiment is that there is slightly improved perceptual
quality of using a multiple exposure workflow for omnidirectional imaging. We identify this result as limitation of using consumer 360° camera. Based on our experiments, we also identify certain drawbacks with existing TMOs, and propose requirements for more aesthetic, robust and accurate tone mapping dedicated to 360° HDR. In addition to this, the novel dual stimulus approach of this study also gives new insights for estimation and extraction of regions of interest in omnidirectional images.

The remainder of the chapter is outlined as follows. Section 6.1 presents the background on HDR capture, tone mapping and subjective evaluation methodologies. The setup of the experiment is described in Section 6.2 including equipment used, acquisition of a new dataset, pre-processing required and content selection. The details of the evaluation methodology together with the experiment design are presented in Section 6.3. Section 6.4 exposes in depth analysis of the results. Finally, we conclude the paper in Section 6.5.

6.1 Background

In this section, the state-of-the-art literature related to this chapter is reviewed. We begin with a discussion on HDR acquisition methods and related work done on omnidirectional imaging. This is followed by a brief review on tone mapping operators highlighting the lack of methods available for omnidirectional tone mapping. Lastly, we discuss previous work on subjective evaluation methodologies in both HDR and 360° imaging which provides the necessary background for our proposed methodology.

6.1.1 HDR Acquisition

Let us now recall some concepts previously visited in Chapter 2. High end DSLR cameras (e.g., Nikon D5 and Canon 1DX Mark II) and specialized professional video cameras (e.g., Arri Alexa XT and Sony F65) are some of the state-of-the-art HDR capable capturing technologies available. However, most consumer cameras only capture 8 bit images and are far from capturing the full range of luminance. To overcome this limitation, multiple single photographs are taken at different exposures in order to capture details from the darkest to brightest regions [126]. In this study, we consider a traditional single exposure workflow and two workflows based on multiple exposures as seen in Figure 6.1.

Although omnidirectional HDR imaging is still not yet to be seen on consumer multimedia platforms, it is very much present for Image Based Lighting (IBL) applications. IBL is a 3D rendering method used in computer graphics for illuminating 3D objects with real world illumination [50]. This often requires capturing an omnidirectional image (typically using
Fig. 6.1 SDR and HDR imaging workflows are shown. The SDR workflow is based on capturing a single exposure image. The example in (a) shows the process by capturing the mid-exposure of the scene. The HDR workflow in (b) describes the typical HDR pipeline consisting of bracketing, HDR reconstruction and tone mapping. The reconstruction technique used in this chapter is based on the work of Debevec and Malik [51]. After reconstruction we employ the tone mapping step. The exposure fusion workflow as seen in (c) is based on the ‘HDR mode’ we find on mobile devices. This process is akin to the work by Mertens et al. [144] which consists of blending multiple exposures into a single image by skipping the HDR reconstruction step.

6.1.2 Tone Mapping Operators

The development of tone mapping operators has been an active field of research for many years, as we have already reviewed in Section 2.4. For this study, we have chosen commonly used TMOs that have been freely available on commercial software [132] for many years and are fundamental to HDR imaging. A description of various TMOs along with the exposure fusion method is described in Table 6.1.

The TMOs considered in this study are based on 2D imaging. This is mostly because very little has been done in omnidirectional HDR tone mapping. Amongst the few works in literature, Hausner and Stamminger [84] present and extension of the Photographic TMO
Table 6.1 Description of various operators used in our experiment.

<table>
<thead>
<tr>
<th>Operator</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear TMO</td>
<td>Global</td>
<td>Simple linear scaling involves normalizing the HDR radiance values between [0 1] followed by applying a gamma correction of 2.2.</td>
</tr>
<tr>
<td>Photographic TMO [181]</td>
<td>Global</td>
<td>Based on the work by Reinhard et al.[181], this TMO relies on photographic principles. In our experiment, apply only the global component of this algorithm mainly to compress high luminance values.</td>
</tr>
<tr>
<td>Display adaptive TMO [129]</td>
<td>Global</td>
<td>Display adaptive tone mapping preserves the contrasts of the HDR image taking into account the characteristics of an output display. We set the output display specification similar to that of an HMD, i.e. up to 100 nits in peak brightness and surround environment of 0 lux.</td>
</tr>
<tr>
<td>Detail preserving TMO [134]</td>
<td>Local</td>
<td>This TMO is based on contrast domain processing algorithm designed to preserve details in the HDR image.</td>
</tr>
<tr>
<td>Exposure fusion [144]</td>
<td>Local</td>
<td>This technique fuses multiple exposures into a single image giving similar results to that of a TMO. However, since it never generates an HDR image it is not considered an HDR technique.</td>
</tr>
</tbody>
</table>

which adapts to the user’s central field of view by using tracking information. Yang et al. [221] and Mikamo et al. [146] display two different tone mapped images of the same HDR input image on each eye of a binocular display (such as an HMD). When seen through a binocular display, the fused image presents more visual richness and detail than both tone mapped versions. These works are potential directions for research in omnidirectional HDR imaging for HMDs but they are out of the scope of this study. Hence, we will only consider existing TMOs part of existing 2D HDR workflows for qualitative assessment against SDR 360° workflows.

6.1.3 Quality evaluation for HDR and omnidirectional imaging

As already mentioned, several studies have evaluated the visual quality of tone mapping operators [120] [222] [111] [65] [62] [142]. There is a wide consensus that preference of TMOs is very subjective. In a subjective evaluation more aligned to our work, Narwaria et al. [156] compared tone mapped images with single exposure images. The study concluded that
observers saw no significant differences between tone mapped and single exposure content.
The authors defend this unexpected result citing a number of possibilities including details in
the bright or dark areas, unnatural colors, overall contrast, naturalness of the scene, etc. We
will also investigate these perceptual factors for our experiments.

Unlike HDR imaging, limited work exists on qualitative assessment of omnidirectional
imaging. Yu et al. [223] proposed two objective metrics for omnidirectional video comparing
different panoramic projections. Zakharchenko et al. [224] proposed a quality metric
which remaps omnidirectional images to a crazer parabolic projection to compare different
geometrical projections. Upenik et al. [207] introduced a test bed for single stimulus
subjective evaluations of 360° images on head mounted displays. We build upon this test bed
to introduce new dual stimulus evaluation method for omnidirectional imaging in order to
conduct our experiments.

## 6.2 Experiment Setup

This section describes the experimental setup. This includes equipment used, proposition of
a new dataset, pre-processing required and content selection.

### 6.2.1 Equipment

The capture device for our experiments was the Ricoh S 360-degree panoramic pocket
camera. This consumer camera provides 8 bit, 5376x2688 (14MP), RGB pictures. It has
two 190-degree field of view camera which enables efficient stitching to generate a single
360-by-360 degree sphere. The Ricoh theta S allows the use of a shutter speed ranging
from 1/6400 to 1/8 seconds and an ISO sensitivity ranging from 100 to 1600 with a wide
lens’s aperture (F2.0) and a focal length of 1.31. An automatic exposure control enables the
acquisition of images with EV from -2 to 2 by steps of 1/3. Each content is stored on an
equirectangular projection compressed version. The used compression is an DCF2.0 and Exif
ver. 2.3 compliant JPEG process, with a quality factor of 95. To ensure the stability of the
camera while acquiring the multi-exposure pictures, a tripod was used, when not exploiting
the environment itself to hold the camera.

The rendering and evaluation of the content produced are based on the subjective evaluation
testbed for omnidirectional images proposed by Upenik et al. [207]. This testbed enables
test sessions proceeding as follows: textual instructions, a training session and the subjective
evaluation using a single-stimulus methodology consisting of displaying a stimulus, then a
scoring menu after interaction via a one touch button and finally the selection of the score
before continuing to the next stimulus. The testbed is able to acquire information from every stimulus including the set of each subject’s scores, the tracking of the direction of view and the duration of the stimulus visualization. The direction of view is recorded by three coordinates, yaw, pitch and roll, representing the angles formed with the normal, lateral and longitudinal axis, respectively. The frequency of acquisition is about 60 Hz and the precision of timestamps is $10^{-7}$ seconds. The equipment required to perform the evaluation is a hand-held device (e.g. iOS mobile) combined with a HMD.

We have used the iPhones 6 and 6S, compatible with the testbed previously described. The devices are 4.7 inch diagonal with an HD resolution (326 ppi). Their maximum brightness is 500 nits and their contrast ratio is 1400:1. The color space representation of those displays are full sRGB. There is no difference between the two devices displays and the followed calibration process was strictly the same.

As far as we know, there is no peak brightness recommendations or standards for HMDs. Therefore, to insure the subjects comfort as well as an optimal visualization setup, it was decided to set the display luminance to 100 nits. This decision is based on a few pilot tests as well as on recommended brightness settings for SDR TV in broadcast [92]. The peak brightness was determined using a white screen and by manually adjusting the iphone’s brightness slider. Measurements were taken using the x-rite i1 Display Pro\textsuperscript{1} and the i1 Profiler Software 1.1.1.

The experiments were conducted using Merge virtual reality\textsuperscript{2} headset. This HMD is compatible with Android and iOS smartphones. Its simple double input interfaces (referred to as buttons in the following) facilitate and expand possible interactions within the testbed. The adjustable lenses lead to a more comfortable experience as they are designed to fit one’s specific eye distance. This HMD guarantees a field of view of 90 degrees. The entire system has a precision of 8.3 pixels per degree.

### 6.2.2 Content Description

The 360° HDR dataset is composed of 43 contents consisting of a set of 5 multi-exposures pictures, an HDR reconstruction as well as tone-mapped and exposures fusion versions of a scene. The multi-exposure pictures were captured using the consumer camera Ricoh Theta S, previously described, with EV ranging from -2 to 2 by step of 1. As an example, Figure 6.2 presents the five multi-exposure images of the scene Lake. As a result, an attempt to capture large range of luminance of the scene is made. Figure 6.3 illustrates the SDR images resulting from the HDR and Exposure fusion workflows described in Section 6.1.1.

\textsuperscript{1}http://www.xrite.com/categories/calibration-profiling/i1display-pro

\textsuperscript{2}https://mergevr.com/goggles
6.2 Experiment Setup

The captured contents are classified in three content types, namely Indoor (9 contents), Outdoor (24 contents) and Night (10 contents). This dataset aims at the evaluation of a broad collection of scenes in addition of capturing a wide variety of shooting conditions and environments e.g. with various overall contrast and brightness, diffuse and specular reflections, deep shadows and bright highlights. Although the capturing process included a variety of lighting conditions, no specific attempt was made to add artistic intent during the photo sessions. The main goal was to capture the entire dynamic range of natural scenes. Furthermore, specific attention was drawn on having as few as possible of multi-exposure fusion impairments such as ghost effects and noise. The introduced dataset is publicly available\(^3\). It should be noted that even though 5 exposures enable the capture of a large part of the dynamic range of a scene, the sensor properties of the Ricoh Theta S consumer camera limit the order of magnitude of the acquisition of the dynamic range. We shall revisit this issue in Section 6.2.4.

6.2.3 Content preparation

After capturing different exposures, there are several ways to generate an HDR image. Based on the study of Akyuz et al. [8], state-of-the-art reconstruction algorithms are not statistically superior to each other in terms of accuracy. Hence, we simply apply the classical method proposed by Debevec and Malik [51] using a triangular weighting function for our study. For all the TMOs, we applied the recommended settings suggested by the authors except for the Display Adaptive TMO where we applied the surround lightings setting specific to our HMD.

\(^3\)[http://mmspg.epfl.ch/360hdr-consumercamera]
It must be noted that all pre-processing were done on the equirectangular projection of the contents.

A special attention should be paid to processing equirectangular projections with Detail preserving TMO and Exposure fusion. This is because both operators implement a local processing which implies that the left and right sides of the projected image will have a different processing and as a consequence a vertical line can appear on the HMD, clearly depicting the region where the stitching is applied. This is a major drawback of working in the equirectangular projection and creates a reoccurring artefact for all local processing potentially affecting users judgment during subjective quality assessments. We resolved this issue by doing additional processing for local operators. This involves three steps: 1) concatenating the columns of left part of the projection to the right part and vice versa, 2) applying the local operator and 3) removing the additional columns and retrieve the locally processed image.

6.2.4 Content Selection

In order to conduct a reliable subjective study, the test images should be containing a large variety of stimuli. In HDR imaging, we often consider scenes with extreme variations in lighting levels. Keeping this in mind, we have selected 8 different HDR images (shown in Figure 6.4 along with log_2 histograms) which are representative of a diverse set of content with indoor, outdoor and night scenes with varying lighting conditions. The selection criteria also relied on finding images with minimal artifacts (ghosting, noise, over/under exposure etc). In addition to this, we selected 3 more images for the training set: Bridge, Trail and Sculpture.

We also attempted to capture and classify our photographs as described in previous chapters such as low key, medium key, high key etc. However, this is not so obvious in omnidirectional imaging. For example, in the Room content, we have a bright key view of the scene when looking outside the window and a medium key scene when we look into the bedroom. Similarly, in the Cellar image, we have a low key environment across most the image however if place our viewport in direction of the light source we lose the dark ambiance of the scene. Introducing aesthetics in omnidirectional imaging is a challenge however, as we can see there is plenty of scope of future work.

A number of different global statistics were considered over the entire sphere for each image. The statistics can be found in Table 6.2. The definitions of the statistics are described as follows.
6.2 Experiment Setup

Fig. 6.4 The equirectangular projections and the histograms of the images used in the experiments. The linear TMO has been applied for each of the images above. The histograms are computed over the entire sphere of the omnidirectional image, showing the relative frequency of the $\log_2$ luminance of the pixels.

1. Dynamic Range (DR):

$$DR = \log_2 \left( \frac{L_{\text{max}}}{L_{\text{min}}} \right),$$  \hspace{1cm} (6.1)

where $L_{\text{max}}$ and $L_{\text{min}}$ are the maximum and lowest scene referred luminance values clipped at 0.1 and 99.9th percentile to make the metric robust against outliers. This is the most popular unit for expressing dynamic range and is measured in f-stops.

2. Key Value:

As we have seen in Chapter 2, the key value is a popular metric in HDR imaging to indicate the average brightness of a scene [181]. Let’s recall from Equation 2.4 and Equation 2.5, the definition of the key value:

$$\text{Key value} = \frac{\log L_{\text{avg}} - \log L_{\text{min}}}{\log L_{\text{max}} - \log L_{\text{min}}},$$  \hspace{1cm} (6.2)
where \( \log L_{\text{avg}} \) is the log geometric mean of luminance of the image, \( L_{\text{max}} \) and \( L_{\text{min}} \) are the maximum and minimum scene referred luminance values clipped at 0.1 and 99.9th percentile.

3. Spatial Information (SI):

The SI is an indicator of spatial details in an image. This was determined by first applying the Sobel filter on the linear TMO image followed by calculating the output image’s standard deviation.

In Table 6.2, we see that various statistics don’t show a clear indication that the chosen images contain a large variety. One of the main concerns of the dataset is that the dynamic range of the images varies from 7.59 – 12.27 f-stops. We believe this limitation to be due to the Ricoh Theta S sensor at extreme EVs. In fact, one can also observe over-exposed and under-exposed regions in the reconstructed linear HDR scenes which is atypical in HDR imaging. Nevertheless, this dataset remains useful as it is the best possible dynamic range that can be captured using the current consumer omnidirectional cameras which is the use-case we are addressing.

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Content type</th>
<th>Exposure time min</th>
<th>Exposure time max</th>
<th>Key value</th>
<th>DR</th>
<th>SI</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trail</td>
<td>Outdoor</td>
<td>1/6400</td>
<td>1/400</td>
<td>7.92</td>
<td>0.46</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Bridge</td>
<td>Outdoor</td>
<td>1/6000</td>
<td>1/125</td>
<td>7.59</td>
<td>0.46</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Sculpture</td>
<td>Outdoor</td>
<td>1/4000</td>
<td>1/200</td>
<td>7.93</td>
<td>0.59</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>Indoor</td>
<td>1/200</td>
<td>1/30</td>
<td>9.88</td>
<td>0.54</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Indoor</td>
<td>1/30</td>
<td>1/8</td>
<td>9.70</td>
<td>0.54</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>Cellar</td>
<td>Indoor</td>
<td>1/180</td>
<td>1/30</td>
<td>7.86</td>
<td>0.55</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Cafeteria</td>
<td>Indoor</td>
<td>1/1500</td>
<td>1/40</td>
<td>8.22</td>
<td>0.41</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Vase</td>
<td>Night</td>
<td>1/30</td>
<td>1/10</td>
<td>10.59</td>
<td>0.52</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Berlin</td>
<td>Night</td>
<td>1/60</td>
<td>1/8</td>
<td>12.27</td>
<td>0.42</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Lake</td>
<td>Outdoor</td>
<td>1/5000</td>
<td>1/160</td>
<td>7.97</td>
<td>0.44</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Rolex</td>
<td>Outdoor</td>
<td>1/6400</td>
<td>1/500</td>
<td>9.58</td>
<td>0.57</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

This is not to say that content is not diverse. To put into perspective the diversity of the selected omnidirectional content, we refer to the work in video tone mapping by Boitard et al. [37]. If we consider the viewport as video frame and the movement of the observer as the motion of the camera, it is possible to study the changed key value as the user navigates across the 360° image. To simulate this, we extract the viewports along the center of the
omnidirectional images and calculate the key value per viewport as seen in Figure 6.5. It must be noted that the brightness not only varies between the images but also changes significantly within the entire sphere in each image.

![Change in key value while moving horizontally across the scene](image)

Fig. 6.5 The variation in key value of the omnidirectional HDR content used in the experiment. The key value is calculated using Equation 6.2 for a given viewport across the entire yaw angle of the scene while fixing the pitch and roll to 0°. This graph shows the content selected was diverse with varying luminance levels to challenge TMOs.

### 6.3 Methodology

In this section is discussed the design of the subjective test, from the design of a methodology for omnidirectional content pair-comparison to the definition of our test and the creation of a comprehensive questionnaire.

#### 6.3.1 Pair Comparison approaches

To the best of our knowledge, only the single stimulus (SS) methodology has been used in previous subjective evaluations for omnidirectional content. A single stimulus evaluation consists in the presentation of a single content assessed according to a predetermined scale. In pair-comparison (PC) methods, the relation between two images or image sequences is evaluated. The set of stimuli is usually presented juxtaposed or sequentially. Even though those are two preferential judging methods, as emphasized by the ITU-R recommendation BT.500 [97], the choice of using SS or PC is based on the context of the analysis and the aim of the experiment. SS sorts the assessed impairments, algorithms in an absolute way while PC permits to discriminate one content, by determining their relative quality. As we conduct here an analysis on the relative improvement of the perceptual quality of multi-exposures over single exposure workflows, a PC methodology is more suitable.
Fig. 6.6 Approaches considered to reproduce Side-by-side Pair-Comparison methodology in an omnidirectional environment are presented. Using a split screen as seen in (a), forces the user to evaluate different parts of the same content which violates the construct validity of the experiment. Similarly the butterfly comparison in (b) and (c) will result in very unnatural environments which may bias the assessment. Thus, the toggling approach for the pair-comparison evaluation is proposed.

A number of methods of pair comparisons methodologies were envisioned. This included well known side-by-side implementations such as butterfly and split screen. The main disadvantage of these approaches are lack of naturalness of the scene especially when seen through a HMD. Moreover, such representation may also increase the likelihood on sickness and further introduce bias. In Figure 6.6, we demonstrate the limitations of various side-by-side implementations for the Room image. Hence, we propose an alternate PC approach based on image toggling.

6.3.2 Toggling

A pair-comparison toggling approach involves the user being able to visualize both contents by switching between them. We build upon the test bed described in Section 6.2.1 to introduce a new double stimulus evaluation methodology for omnidirectional content. The proposed evaluation starts by displaying the test stimulus first. By pressing the right button of the merge HMD, one can display the reference stimulus for the same view port. By pressing the right button multiple times, the user can toggle between the test and the reference stimulus. Each scene is labeled either as ‘T’ for test or ‘R’ for reference at the center of the view port, indicating the subject with which stimulus is currently being displayed. In our experiments, it was mandatory to toggle at least twice before scoring. No maximum limit of the number of toggling was set. The vote is enabled only on the test stimuli so that the last visualized content is the one to be assessed. The left button of the HMD is used for the voting process. By first pressing the left button a stationary vote menu is displayed. With the help of a red cursor at the center of the view port, users can select their score on the voting menu by head movement. Once the cursor has been correctly positioned over the preferred score, the user can cast his/her vote by pressing again on the left button.
pair-comparison evaluation starts. The order of the presentation of pair-comparison stimuli are randomized in such a way that never the same content is assessed successively.

The proposed double stimulus test bed also collects similar data to the single-stimulus test bed such as scores and view direction tracking data. It also has the feature of recording toggling information by storing the timestamps and the view port directions when users toggles. The analysis of such information can be used to investigate subject’s comparison process and observing the regions of interest for HDR omnidirectional contents. This could potentially help identify subjects behavior during an assessment, gather data for design and implementation of future omnidirectional TMOs, conceive and validate new subjective quality assessment methodologies and objective metrics.

### 6.3.3 Experiment design

An adjectival categorical rating methodology was selected on a 5-point grading scale to score pair comparison results. The assessment scale stands as follows: 1: T worse than R ; 2: T slightly worse than R ; 3: T same as R ; 4: T slightly better than R ; 5: T better than R. T refers to the test stimulus (TMO or exposure fusion image) while R is the reference stimulus (single exposure image). The rating scale is displayed on the voting menu of the testbed. The reference image is chosen as the single exposure image with EV = 0. This choice is based on photographic principles as the mid-exposure usually permits the acquisition of bright and dark areas without favoring neither highlights nor shadows.

Additionally, we have created a post-questionnaire investigating the following criteria:

1. **Criteria considered during the evaluation:**
   HDR images are evaluated according to various criteria, such as image contrast, naturalness, colorfulness and overall brightness. In view of recommendations for the development of omnidirectional TMOs as well as for future questionnaires appraising 360 HDR contents, an investigation of the criteria impacting the assessment is conducted. Narwaria et al. [155] reviewed the criteria on which is based the differentiation of tone mapped images. The recurrent and most used criteria were selected for our questionnaire and are the following: *Details in the bright areas, Details in the dark areas, Unnatural colors, Ghosting, Noise, Overall brightness, Overall contrast* and *Naturalness of the scene*. Additional empty spaces are added in case subjects want to make propositions. These criteria were ranked from the most considered (1) to the least considered (8) when making the evaluation decision.

2. **Interest and appreciation of the content:**
   The assessment of the interest and appreciation of the content is a self-made question-
naire investigating if a content is Boring, Interesting, Colorful, Aesthetic, Familiar, Of quality, Pleasurable and Immersive. The subjects select the attributes when they find it suitable. Those dimensions are selected based on the information we want to gather to explain the possible discrepancies in the results we foresaw and are inspired from the work of Mansilla [127]. An additional open question inquiring the dislike or like of the content was added in order to get further insight in the reasons of content preference.

3. Sickness:
In order to accurately evaluate the degree and type of sickness generated by our experiment, the widely used simulator sickness questionnaire by Kennedy et al. [105] is included in our questionnaire. The symptoms General discomfort, Fatigue, Headache, Eye strain, Difficulty focusing, Increased salivation, Sweating, Nausea, Difficulty concentrating, “Fullness of the head”, Blurred vision, Dizziness with eyes open, Dizziness with eyes closed, Vertigo, Stomach awareness and Burping are evaluated on a 4-points scale from none (1) to severe (4).

4. Virtual reality experience:
The aim of such questions is to acquire sufficient knowledge on the experience to improve the next design of the evaluation methodology. An investigation of the overall appreciation of subject’s experience in terms of Immersion, Isolation, enjoyment/valence, arousal/excitation and the Enjoyment due to the novelty of the visualization is conducted. A 5-point scale from strongly disagree to strongly agree is used when asking subjects if they experienced the above-mentioned attributes. The immersion and isolation attributes evaluates the interest in using an HMD for the evaluation, the valence and arousal attributes come from the Self-Assessment Manikin [39] measuring the emotions generated by the experiment. Also, the level of experience of subjects regarding their use of HMD is gathered. An additional open question also asks them to describe the VR experience in a few words.

The study is carried out on eight contents, selected from the dataset, processed using four TMOs and an exposure fusion operator described in Section 6.2.2. Overall, 40 pair comparison stimuli are evaluated. A single test session of about 20 minutes is needed for the assessment of all pair comparison stimuli.

Prior to the test, subjects were screened for correct color vision and vision accuracy using Ishiara and Snellen charts. As recommended in the simulator sickness questionnaire [105], subjects were asked if they are in a normal state of health. Any violation of the previous conditions results in the rejection of the subject from the experiment. Overall, 25 (17 male and 8 female) subjects participated in tests and were 29.5 years old in average,
ranging from 19 to 55 years old. After being instructed with the study process, subjects were requested to read the information and consent forms and the post-questionnaire, as well as an explanation sheet, describing complex terminologies used in the questionnaire. Any question was answered to before starting the experiment.

In order to have subjects familiarized with the assessment procedure as well as to reduce across-subjects grading discrepancies, instructions are again provided by the testbed and a training session is performed in advance of the test session. Three pair comparison stimuli were rendered during the training session in the following order: The photographic tone mapped Bridge content, assessed as worse than R (1), the display adaptive tone mapped Trail, evaluated as equivalent to R (3), and the exposure fusion of Sculptures, assessed as better than R (5). Different contents were used in the training session when compared to test session in order to prevent the introduction of any bias. The test session starts right after the completion of the training session. As mentioned previously, subjects had to toggle at least twice for each pair comparison and could only vote on the test stimulus T. No time-constraints were defined per stimulus or per test session. Once the test session is over, subjects are required to fill the post-questionnaire. The test was carried out in a calm environment, free from any disturbances. A rotatable chair is used during the assessment for subjects’ comfort as well as to ease their 360° navigation within the omnidirectional content.

6.4 Results

The analysis covers three types of information, namely, the subjective scores, the post-questionnaire data and the tracking of the direction of view combined with toggling information.

6.4.1 Subjective scores

An outliers detection, based on ITU-R recommendations BT.500 [97], did not reject any subjects of this study. The MOS and 95% CI were computed as depicted figure 6.7. Figure 6.7a shows the variations across contents, for each operator considered while Figure 6.7b presents a comparison matrix depending on contents and operators.

Overall, the results do not show a clear preference of the processed content over a single exposure content. Indeed, MOS mainly range from 2 (T slightly worse than R) up to slightly above 4 (T slightly better than R), indicating a moderate improvement or deterioration of the perceived quality.
Based on the Figure 6.7b, the exposure fusion and the Linear TMO show better performance when compared to the other operators. The extent of the performance is a similar or slightly improved perceived quality. On the other hand, the Photographic, Display Adaptive and Detail Preserving TMOs do not perform as well. This result does not come as a surprise and can be explained through the limitations of the consumer camera. Let us recall that the DR measure of the dataset spanned from 7.59 - 12.27 f-stops. These TMOs are designed for content with a DR greater than 16 f-stops. Thus, this results suggests that TMOs require further evaluation using professional 360° HDR content having higher DR. As a result the dataset is ideal for the exposure fusion algorithm which works well for images consisting of equally spaced EVs. As this method avoids the HDR reconstruction step and is thus independent of the DR of the scene. Also, it is well known in HDR imaging that a linear TMO often results in a dark image and loss of detail [178]. The fact that we can clearly visualize the Figure 6.2a and observe little loss of detail also indicates the lack of DR in the content.

In addition to this, the differences across contents are significant, based on several non-overlapping CIs. This illustrates the variety of the chosen contents. All operators, except the Detail preserving TMO, are particularly preferred over the reference for the contents Lab and Room, two indoor contents with an outside view through a window. No trend in the statistics was found in Table 6.2 and Figure 6.5 seems to justify this behavior. This demonstrates the need of a precise questionnaire investigating specific criteria.

The non-normality of scores distributions prevents us to run a repeated measure ANOVA, especially considering our sample size (=25), not sufficient to overcome the violation of assumptions.
6.4 Results

### (a) Rank of evaluation criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details in bright areas</td>
<td>1</td>
</tr>
<tr>
<td>Overall contrast</td>
<td>2</td>
</tr>
<tr>
<td>Naturalness of the scene</td>
<td>3</td>
</tr>
<tr>
<td>Details in dark areas</td>
<td>4</td>
</tr>
<tr>
<td>Overall Brightness</td>
<td>5</td>
</tr>
<tr>
<td>Unnatural colors</td>
<td>6</td>
</tr>
<tr>
<td>Noise</td>
<td>7</td>
</tr>
<tr>
<td>Glare</td>
<td>8</td>
</tr>
</tbody>
</table>

### (b) Content type

- Lab
- Room
- Content

### (c) Sickness symptoms

- None
- Slight
- Moderate
- Severe
- Extremely severe

### (d) VR experience

- Agree
- Undecided
- Disagree
- Strongly disagree

**Fig. 6.8 Post-questionnaire analysis**

#### 6.4.2 Post-questionnaire answers

**Criteria considered during the evaluation**

In order to identify which criteria have most impact on the assessment, the Borda count method [46] was applied. The results are presented in Figure 6.8a. From the latter we see that the details in bright areas, overall contrast and naturalness of the scene are the most important factors for the subjects. This information could be considered for the design of future TMOs for and also content creators for generating aesthetic 360° images.

**Interest and appreciation of the content**

The Figure 6.8b illustrates the subjects choices of suitable adjectives describing the contents. It has to be noted that the contents showing the best results, Lab and Room were both assessed as boring by more than 50% of the subjects. This is an interesting finding that could be further investigated regarding subjective tests content selection. When considering the
content Lake, its MOS results indicate a similar perceived quality from T to R stimuli for all TMOs except for the Photographic TMO. The content was assessed by more than 80% of subjects as colorful and the above-mentioned operator introduced unnatural colors. Even if the unnaturalness of colors criteria was among the least considered, we can conclude that this criterion is still of importance when tone mapping omnidirectional HDR content. The Lake image was also considered highly aesthetic and pleasurable by the subjects. Although no particular effort was made to introduce aesthetics in the acquisition or selection process, we did find this content particularly pleasant.

Sickness

Various works in literature have reported that a virtual environment can generate sickness [40]. A subjective evaluation is prone to bias if users are in discomfort or tired. The questionnaire aims to certify the validity of the test. The figure 6.8c reports the degree of severity of the sickness effect of the subjects. We can observe that the extent of sickness is from none to slight.

In addition, 48% of our subjects reported not having experienced sickness. These two facts confirm the proper design of the test for pair-comparisons, concerning the sickness effect. Overall, when considering the severity and occurrence of sickness symptoms, Nausea and General discomfort characterize the sickness experienced by our subjects.

Virtual Reality Experience

The investigations carried out on subjects’ experiences and feedbacks are reported here. Prior to the test, 36% of the subjects’ had never used an HMD while 16% had used it more than 5 times. We can conclude that our population sample is representative of the population when considering their VR experiences.

The evaluations of subjects experience show a moderate immersion, isolation and enjoyment. However, the subjects acknowledged a lack of arousal. Those results are mainly explained in the open question which asks for summarizing subjects experience, at the end of the post-questionnaire. The justification of the above comment was provided by more than a quarter of the subjects mentioning that their excitation level was counter-balanced by their perception of the low level of quality of the images, especially when pixelation was perceived, or by the experience of sickness.

The reported feedbacks from the subjects in the open question indicate discrepancies in the appreciation of enjoyment and perceived quality. Spontaneously, nine subjects reported a funny or great experience while five did not fully enjoyed the experience. Regarding the
quality of the display, two subjects mentioned a good quality of images while three found the quality (or resolution) not as expected or not sufficient to provide a truly immersive experience. These findings explain the modest experience of enjoyment reported by our MOSs.

Subjects complained that it was tiring to have the two hands lifted to their face in order to control the two buttons of the HMD. This complaint emphasizes the high likelihood of tiredness and lack of attention from subjects. Controlling the HMD with a controller (e.g. game-pad) could tackle this issue. Subjects also gave indications to improve the immersion and/or isolation experience: one mentioned than he was “experiencing contents as a ghost instead of being physically there”, another stated that “it would be more immersive to zoom in and out”. Despite of the soundness of these comments, several are out of the scope of this study.

Some subjects indicated that the duration of the test was correctly set as they were starting to get bored and/or annoyed when visualizing the two or three last stimuli. Considering this comment, a test session (including explanations and training, if applicable) should not exceed 20 minutes as outlined in several similar subjective assessments. One subject mentioned his approach carried out during the experiment, especially regarding the visualization: “At the beginning, I have tended to move in all directions and then, I have progressively focused on areas of interest, typical of each content” and had the “tendency to navigate more horizontally than vertically, and then to balance in both directions”.

A few subjects claimed that choosing between the test and reference was a challenge. They found cases where they preferred the test content over the reference content in the bright regions while not so in the dark regions and vice-versa. This suggests that further work is needed in 360° TMO design to overcome such challenges.

6.4.3 Toggling locations analysis

Let us recall from Section 6.3.2, the test bed is able to collect data on the viewing direction and regions where the users toggled. We attempt to use this information to find the regions of interests that may have influenced the decision of the users during the voting. In Figure 6.9 we see the toggling locations across various contents. Each point in the image represents the center of the viewport where the toggling took place.

The inconvenience of this representation is that it is based on the assumption that an effective comparison is operated exclusively where subjects have toggled and this may not be true in all cases. Another limitation of this representation is that it only considers the head movement positions. As a future extension of our work, more accurate fixations maps can
be computed based on the prediction of the eye gaze fixation from head fixation locations, applying the work of Rai et al. [173].

Nevertheless, we can still draw some interesting conclusions from Figure 6.9. The first interesting result is that there is no influence of the initial view-port voting process. We observe this on the contents Cellar, Lake, Rolex and Vase where there is few or no toggling locations at the initial view-port location, indicating it is not an area of interest. In some cases, such as in Berlin, Cafeteria, Lab and Room images, the initial viewport is one of the areas of interests. Secondly, the locations emphasized in the results are mostly longitudinally distributed at a latitude close to the equator. Additionally, the main areas of toggling are also mainly centered at the equator latitude. In fact, the first quartile, average and third quartile of the set of pitch coordinates are 28.6, 1.9 and -23 respectively. This range of latitudinal head positions seems comfortable and natural for the users.

Secondly, while analyzing the areas of toggling, it seems that subjects are assessing their preference in terms of loss of information. This is seen in several contents, such as the outside information through the windows of the Lab and Room images and the building or the sky over the grass field in the Lake content, are clear toggling areas. Their toggling locations correspond to areas over-exposed in SDR contents and which are replaced by more details in the HDR contents. However, in some cases the over-exposed areas in SDR contents are not taken in consideration. For example the reflection of the light on the lake or windows in the Lake and Rolex contents respectively. Also, the objects that serve as light sources such as the light bulbs in the Cafeteria and Cellar images and lit rooms in the buildings seen on the night images were not considered by users. In such cases the loss of information is usually considered by expert viewers in order to have a faithful representation of reality while evaluating TMOs. A possible explanation is that for naive observers, the importance of the lost information of specular and light emitting areas is not significant compared to the other areas in the omnidirectional scene.

![Fig. 6.9 Toggling locations across contents](image-url)
Lastly, one can also notice the trend whereby observers focus more often on distant details than on close ones. In the Cellar content, the two clear areas of interest are the two extremities of the cellar room. Overall, in all contents, the main areas of toggling coincide with the locations of the furthest objects. The subjects behavior for comparison is very interesting and can lead to new visual attention models as well as provide clues for the development of 360° dedicated TMOs.

### 6.5 Conclusion and Future Works

In this chapter, we assessed the importance of developing HDR for omnidirectional contents as well as proposed directions for 360°-dedicated retargeting. An evaluation of an end-to-end HDR pipeline using consumer equipment has been carried out. We introduce a publicly available dataset composed of 43 multi-exposure images, acquired with a consumer camera, also including the HDR reconstructions and the SDR versions resulting from four well-known off-the-shelf TMOs and one exposure fusion algorithm. The acquisition and database generation process provided interesting lessons for aesthetic 360° capture. Eight of the contents from the database were carefully selected and used to conduct evaluation tests. A new subjective test methodology, enabling pairwise comparisons for omnidirectional content is introduced. The toggling permits to switch between the assessed stimuli and the reference. Our results show that none of the evaluated operators show a substantial increase of perceived quality. However the exposure fusion, as well as the linear TMO to a lesser extent, show promising results as being assessed as slightly better than the single exposure reference. In addition, the discrepancies across operators and contents leads to the identification of a need for dedicated TMOs for omnidirectional contents. As future work, an evaluation of TMOs with professional content can be considered. The analysis performed on our self-made questionnaire emphasized that Details in bright areas, Overall contrast and the Naturalness of the scene are important criteria to consider during the assessment. Some subjects reported slight sickness effects mostly felt as Nausea and General discomfort. This fact confirms the proper design of our methodology. Furthermore, we analyzed the toggling information to draw important conclusions regarding human perception while comparing tone mapped omnidirectional images on a head mounted display. This work aims to lay a basis for the future development of HDR imaging for omnidirectional representations.
Chapter 7

Conclusion

In this thesis we have presented a variety of contributions on different aspects of display retargeting in HDR imaging. The main conclusion of our work highlights the importance of preserving artistic intent and considering our knowledge of human perception for aesthetic and visually comfortable retargeting.

We are convinced that preserving artistic intent should play a major role in the design of future algorithms in all fields of display retargeting. We have presented in this thesis a number of examples where aesthetic knowledge was leveraged to improve the viewing experience. More specifically, by developing our computational models of aesthetics (Chapter 3) or by conducting simple experiments (Chapter 4). Similarly, we also believe considering perceptual effects and limitations of the human visual system is fundamental for display retargeting. We demonstrate the benefits of perceptual retargeting by enabling comfort-based retargeting (Chapter 5) and by investigating novel perceptual cues for retargeting in HMDs (Chapter 6).

Despite the vast amount of research on human perception, our visual system is extremely complex and is an area of study with growing importance. Furthermore, considering aesthetics for retargeting adds a layer of cognitive challenges to the HDR retargeting problem. As a result representing comprehensive computational models combining both perception and aesthetics is a novel direction for retargeting. This introduces new hurdles and requires balancing a number of trade-offs involving mathematical and algorithmic formulations, computational complexity and the limitations of camera and display hardware. Therefore, integrating efforts from various domains of display retargeting is essential. In this chapter we present a summarized compilation of the specific conclusions affiliated to each Chapter in Section 7.1. We then discuss open issues and prospects raised by this thesis in Section 7.2.
7.1 Summary

Chapter 3 presented our work on aesthetic retargeting. We realized that aesthetics was rarely considered in HDR imaging especially in the field of tone expansion. As we already know, most existing video content today is in SDR format and over the decades a tremendous amount of effort has been made by artists to exploit this SDR container to convey intent. Thus, we begin by studying aesthetics and introduce a classification of artistic intent in the form of lighting styles—dark key, low key, medium key, bright key and high key. We demonstrate that current state-of-the-art tone expansion algorithms do not preserve the lighting style intended by the artist. In addition to this, we present a series of perceptual experiments studying the user preference for various lighting styles as seen on HDR displays. This evaluation shows that tone expansion of stylized content is best preserved by gamma correction that adapts to the intended style of the video. Using this, we design an novel SDR to HDR retargeting algorithm which includes an additional color processing step. We validate our algorithm through a subjective study against existing operators. We find that our method performs better not only in terms of subjective quality but also by preserving artistic intent. A number of directions were identified for future work. For example, special attention is needed in the special case of digital video effects where a $\gamma$ correction may fail. Furthermore, we identify the need to have a display independent tone expansion model, going beyond the 1000 nits limit.

Chapter 4 is devoted to retargeting between HDR displays. This topic is of utmost important in the current HDR ecosystem as mastering displays have peak brightness ranging from 1000 to 4000 nits while consumer displays have their peak brightness range from 500-1000 nits. This tonal incompatibility between displays in the transmission and receiver lead to loss of artistic intent. Hence, we apply our knowledge on aesthetics and perception to develop a generic retargeting framework between two HDR displays. We experiment with simple tone mapping and tone expansion operators to map 1000 nits content to 4000 nits displays and 4000 nits content to 1000 nits displays. To challenge these retargeting methods we specifically choose content of various lighting styles. A perceptual study was conducted and we found that a scene by scene user-controlled gamma correction is preferred for retargeting between HDR displays. This method is time consuming and requires extensive man power and is not practical for real-time applications. Thus for future work, it is recommended to model a temporally coherent $\gamma$ correction curve.

Chapter 5 is dedicated to comfort-based retargeting for HDR displays. This area has been mostly unexplored and we have proposed a unique study measuring brightness comfort for HDR retargeting. We have seen in the previous chapter we had found that users felt high key content on a 4000 nits display uncomfortably bright. This propelled us to investigate into
modeling brightness comfort for the special case of legacy screen content. Screen content has unique statistical properties making it a challenge for retargeting. We conduct a perceptual study measuring the brightness preference for each content on a 4000 nits Sim2 HDR47 display. We find that the brightness preference is highly content dependent. Based on this we design a novel feature which measures the amount of bright pixels in an image. We correlate this feature to the subjective scores to propose a brightness control retargeting algorithm that controls the peak brightness of the display. This chapter leads to a number of open ended research questions including the affect of brightness on visual fatigue, color appearance ambient lighting, different age groups, display technologies etc.

Chapter 6 steers this thesis in a new direction involving retargeting for HMDs. Omni-directional or 360° imaging has seen a recent boom in the consumer electronic industry. The goal of this study is to explore if an HDR workflow can benefit this popular display technology. We begin by generating a 360° HDR dataset using a consumer camera and applying the classical HDR reconstruction techniques based on exposure bracketing. During this process, we learn key lessons on omnidirectional aesthetic acquisition. We consider 4 well known TMOs and an exposure fusion algorithms to evaluate them against a single exposure SDR image. We propose a novel dual stimulus methodology to conduct a pairwise comparison on a virtual reality head set. We find that there is slightly improved perceptual quality by using HDR processing. Through the experiment, we find certain limitations of existing TMOs. Based on the perceptual evaluation we suggest improvements for the next generation of TMOs dedicated to omnidirectional imaging.

7.2 Open Issues and Perspectives

The advancements made in HDR technology during the course of this thesis have been remarkable. Over the last three years, we have seen advances in camera technologies, arrival of state-of-the-art displays and new standards for HDR video transmission. This has made it a very exciting yet competitive environment for both academic and industrial researchers. On a personal note, adapting to changing technologies and evolving standards has been a real challenge but also a pleasure in learning and innovating.

During the course of this thesis, we have opened the doors to a number research problems. We recognize the need for creating aesthetic HDR content for exploratory experiments in retargeting. This could lead to the creation of dedicated aesthetic models for HDR content. It could also give us new perspective on content creation with possible new stylization techniques. Furthermore, the arrival of LCD and OLED technology on consumer displays have significantly changed the landscape for display retargeting algorithms. This is because
the same video stream is rendered differently on both display systems even when using same retargeting technology. More attention has to be given not only to the type of display but also the black levels, peak brightness and color gamut of the display. This leads us to the importance of dedicated color processing for HDR and the need of a perceptually linear colorspace which can handle both SDR and HDR content. Having a device independent colorspace indicating key perceptual features is crucial to any imaging technology. Similarly, quality metrics play an important role in different part of the imaging pipeline. The human visual system serves as best metric in HDR imaging but it is also essential to have reliable and robust HDR metrics independent of dynamic range. Furthermore, aesthetics-based metrics should also be considered in future research. Lastly, the HDR retargeting problem exists in the forms of tone mapping and tone expansion. For future work, it would be ideal to have a retargeting framework that incorporates both tone mapping and tone expansion independent of displays.

With the work presented in this thesis, we demonstrates that HDR retargeting is far from a solved problem and greater hurdles await researchers. But one must not forget to acknowledge the importance of combining aesthetic and perception for improved HDR retargeting for various display technologies.
References


References


Résumé en Français

Introduction

De la télévision grand public au téléphone mobile en passant par les écrans des ordinateurs portables, les technologies d’affichage évoluent sans cesse. La diversité croissante des caractéristiques des écrans numériques grand public nécessite l’adaptation des images aux différentes technologies d’affichage. Ce processus est connu sous le nom d’adaptation à l’affichage ou display retargeting. Avec l’émergence de nouveaux formats multimédias et l’abondance de contenu existant, le problème de display retargeting est plus pertinent que jamais.

L’imagerie HDR (pour High Dynamic Range imaging en anglais) est le dernier format multimédia déployé sur le marché grand public. Grâce aux progrès réalisés dans les technologies d’acquisition et d’affichage, le pipeline d’imagerie HDR est capable de capter, transmettre et afficher potentiellement toute la gamme de luminance d’une scène. Cette caractéristique permet de surmonter les principales limitations physiques et perceptuelles des systèmes existants dits SDR (Standard Dynamic Range en anglais). Au cours des trois dernières décennies, le display retargeting pour l’imagerie HDR a été largement étudié.

Le display retargeting HDR existe sous deux formes: 1) Tone Mapping pour afficher le contenu HDR sur les écrans SDR et 2) Tone Expansion pour afficher le contenu SDR sur les écrans HDR. Bien que les deux méthodes de d’adaptation (Tone Mapping et Tone Expansion) semblent proches, d’un point de vue algorithmique, les deux méthodes sont très éloignées principalement en raison des différences dans : les caractéristiques du contenu, les capacités des écrans et les limites perceptuelles de l’œil humain. Cela justifie le fait que les chercheurs explorent de nouvelles méthodes de display retargeting en HDR.

Cette thèse présente des contributions sur différents aspects du display retargeting dans le cadre de l’imagerie HDR. Bien que les contributions soient diverses, elles sont motivées par notre conviction que la préservation de l’intention artistique et la prise en compte de caractéristiques en termes de perception du système visuel humain sont essentielles pour un display retargeting esthétiquement et visuellement confortable. Nous pensons
que plusieurs problèmes non résolus dans display retargeting peuvent bénéficier de notre approche holistique. Tout au long de cette thèse, nous poussons plus loin notre réflexion dans cette direction, en cherchant des solutions perceptuelles et esthétiquement viables aux problèmes fondamentaux de display retargeting pour l’imagerie HDR. Ce résumé présente brièvement les sujets pertinents et les contributions de cette thèse.

**Display retargeting pour l’imagerie HDR**

Nous commençons la thèse par une analyse approfondie de la littérature sur l’imagerie HDR. Cette analyse couvre : l’ensemble du pipeline HDR, le système visuel humain, les technologies d’acquisition et d’affichage, le tone mapping, le tone expansion, l’esthétique dans la création de contenu et les normes de transmission pour la vidéo HDR.

**Retargeting esthétique en HDR**

Nous appliquons nos connaissances sur l’esthétique et sur la perception visuelle humaine au cas spécifique du retargeting et nous proposons une nouvelle approche de la problématique du tone expansion (c’est à dire à la conversion SDR vers HDR). Notre motivation vient du fait que la plupart des contenus vidéo existants sont aujourd’hui en format SDR et il est de plus en plus nécessaire d’adapter ce contenu aux écrans HDR. L’adaptation consiste en une conversion SDR vers HDR, c’est à dire à l’utilisation d’une technique de Tone Expansion. Nous démontrons que les algorithmes actuels de Tone Expansion de l’état d’art ne préservent pas l’intention artistique lorsqu’ils adaptent un contenu ayant une esthétique de lumière très marquée, comme l’esthétique dite « noire » (low key en anglais) des films noirs par exemple. En complément, nous présentons une série d’études subjectives où les utilisateurs évaluent, sur les écrans HDR, l’esthétique et la qualité visuelle de contenus ayants des styles d’éclairage différents. Ces études montrent qu’un opérateur de Tone Expansion qui préserve le style esthétique du contenu peut prendre la forme d’une correction gamma, si la valeur de gamma dépend du style esthétique de la vidéo. Nous proposons donc une nouvelle méthode de retargeting qui adapte la valeur gamma au style de la vidéo. Nos contributions sont : une méthode d’analyse automatique du style/esthétique des vidéos, une fonction qui donne la valeur de la correction gamma en fonction du style considéré et une méthode de Tone Expansion (correction gamma). Nous proposons également une méthode simple de correction des couleurs qui peut être appliquée après le Tone Expansion pour restituer correctement les couleurs en HDR. Nous validons notre méthode et la comparons aux méthodes existantes, par une série d’évaluations subjectives. Ce travail cible des écrans HDR de 1000 nits et nous
présentons un cadre mettant en conformité notre méthode avec les normes SDR existantes et les dernières normes de la télévision HDR.


Cette contribution a donné lieu à un dépôt de brevet, et celui-ci a été licenciée à divers acteurs de l’industrie de broadcast. Un certain nombre de démonstrations technologiques ont été présentées dans divers salons et événements tels que National Association of Broadcasters (NAB) 2016 et 2017, International Broadcasting Convention (IBC) 2016 et European Broadcasting Union (EBU) Production Technology Seminar 2017. Plusieurs experts en broadcast ont évalué notre solution avec d’autres propositions industrielles concurrentielles et nous avons reçu des commentaires encourageants et positifs. Le 26 avril 2017, IRT b <> com a reçu le prestigieux prix de Technology Innovation Award de NAB pour "la démonstration de la technologie de conversion vidéo SDR à HDR vitale pour la transition vers la prochaine génération de radiodiffusion télévisuelle".

**Retargeting entre les écrans HDR**

Nous avons également travaillé sur la problématique du retargeting générique entre deux écrans HDR. Les écrans HDR grand public ont une luminosité maximale de 1000 nits, tandis que les écrans professionnels atteignent une luminosité maximale de 4 000 nits. Dans ce contexte, le retargeting entre différents écrans HDR est un problème essentiel à résoudre. Il est probable que l’adaptation de la gamme dynamique entre les écrans HDR utilise des techniques similaires à celles obtenues avec les opérateurs de tone mapping ou de tone expansion. Nous avons analysé et testé des techniques de retargeting issues de l’état de l’art pour adapter des contenus de 1000 nits à des affichages de 4000 nits et aussi pour l’adaptation de contenu à 4000 nits vers les écrans à 1000 nits. Afin de trouver les limites des méthodes
de retargeting classiques, nous avons sélectionné un ensemble de contenus professionnels ayant des styles et des esthétiques variées. Nous concluons que la meilleure technique de reproduction de tonalité entre les écrans HDR dépend fortement de l’esthétique du contenu. Cette étude a été publiée dans une conférence internationale mentionnée ci-dessous.


**Retargeting HDR basé sur le confort visuel**

Dans le cadre de cette thèse, nous avons également abordé la question du confort visuel en fonction de luminosité pour les écrans HDR.

Afin d’afficher des images dites screen content sur les écrans HDR, comme pour les images dites naturelles, une adaptation, **Tone Expansion**, est appliquée au contenu. Dans un certain nombre de cas, le contenu adapté à l’écran HDR apparaît comme trop lumineux et provoque un inconfort visuel. Pour étudier ce phénomène, nous avons effectué une série d’expériences perceptuelles dans laquelle les utilisateurs devaient choisir leur luminosité préférée pour chaque image testée. Dans la base d’image testée nous avons incorporé des images naturelles avec des styles de lumières et des palettes de couleurs variés et des images de **screen content**. Ces dernières sont caractérisées par des statistiques singulières. L’expérience conclut que la préférence de luminosité dépend fortement du contenu. En utilisant les données de l’étude subjective, nous proposons une nouvelle fonctionnalité qui analyse le contenu source afin de contrôler la luminosité de l’écran HDR de manière à éviter l’inconfort visuel. Sur la base de cette fonctionnalité, nous proposons un algorithme de retargeting qui ajuste la luminance de l’affichage en fonction du nombre de pixels fortement colorés et brillants dans le contenu. Ce travail a été publié dans la conférence international suivante.

Exploration du Retargeting HDR pour l’imagerie omnidirectionnelle

Cette étude peut être considérée comme une première étape dans la problématique spécifique du retargeting HDR pour les applications de réalité virtuelle. Nous commençons cette exploration en évaluant la qualité subjective de différents algorithmes de tone mapping sur du contenu HDR 360° également appelé HDR omnidirectionnel. Cette étude tente de répondre à diverses questions et notamment : les algorithmes de Tone Mapping utilisés pour des images fixes HDR sont ils adaptés à un contenu HDR 360°? Comme premiers résultats de ce travail, nous avons mis à disposition de la communauté internationale une nouvelle base de données pour le contenu HDR 360° et nous avons proposé une nouvelle méthodologie pour évaluer la qualité subjective du contenu HDR 360° sur un affichage visiocasque après retargeting. Les résultats sont ensuite analysés et des conclusions ont été tirées. Cette étude est publiée dans une conférence internationale.


Conclusion


En conséquence, les modèles computationnels combinant à la fois la perception visuelle et l’esthétique des images s’avèrent une nouvelle direction de recherche dans le domaine
du *display retargeting*. Cela lève de nouveaux obstacles et nécessite d’équilibrer un certain nombre de compromis impliquant des formulations mathématiques et algorithmiques complexes.

Les progrès réalisés par les chercheurs à travers le monde dans la technologie HDR, au cours de cette thèse, ont été remarquables. Au cours des trois dernières années, nous avons connu des progrès dans les technologies de caméras, l’arrivée des écrans HDR grand public et des nouvelles normes pour la transmission vidéo HDR. Avec le travail présenté dans cette thèse, nous démontrons que le retargeting HDR vers SDR et réciproquement est loin d’être un problème résolu. De grands défis attendent les chercheurs. Nous avons montré que la combinaison, de la qualité visuelle avec le confort visuel et l’esthétique des images, est l’un des défis clés. Nous avons apporté un ensemble de premières contributions à ce défi.
Appendix A

Publications and Patents

A.1 International Journal


Abstract – High Dynamic Range (HDR) is the latest video format for display technology and there is a strong industrial effort in deploying an HDR capable ecosystem in the near future. However, most existing video content today are in Standard Dynamic Range (SDR) format and there is a growing necessity to upscale this content for HDR displays. Tone expansion, also known as inverse tone mapping, converts an SDR content into an HDR format using Expansion Operators (EOs). In this paper, we show that current state-of-the-art EOs do not preserve artistic intent when dealing with content of various lighting style aesthetics. Furthermore, we present a series of subjective user studies evaluating user preference for various lighting styles as seen on HDR displays. This study shows that tone expansion of stylized content takes the form of gamma correction and we propose a novel EO that adapts the gamma value to the intended style of the video. However, we also observe that a power function-based expansion technique causes changes in terms of color appearance. To solve this problem, we propose a simple color correction method that can be applied after tone expansion to emulate the intended colors in HDR. We validate our method through a perceptual evaluation against existing methods. In addition to this, our work targets 1000 nits HDR displays and we present a framework aligning our method in conformance with existing SDR standards and the latest HDR TV standards.
A.2 International Conferences


Abstract – The vast majority of video content existing today is in Standard Dynamic Range (SDR) format and there is a strong interest in upscaling this content for upcoming High Dynamic Range (HDR) displays. Tone expansion or inverse tone mapping converts SDR content into HDR format using Expansion Operators (EO). In this paper, we show that current EO’s do not perform as well when dealing with content of various lighting style aesthetics. In addition to this, we present a series of perceptual user studies evaluating user preference for lighting style in HDR content. This study shows that tone expansion of stylized content takes the form of gamma correction and we propose a method that adapts the gamma value to the style of the video. We validate our method through a subjective evaluation against state-of-the-art methods. Furthermore, our work has been oriented for 1000 nits HDR displays and we present a framework positioning our method in conformance with existing SDR standards and upcoming HDR TV standards.


Abstract – High Dynamic Range (HDR) is the latest trend in television technology and we expect an influx of HDR capable consumer TV’s in the market. Initial HDR consumer displays will operate on a peak brightness of about 500-1000 nits while in the coming years display peak brightness is expected to go beyond 1000 nits. However, professionally graded HDR content can range from 1000 to 4000 nits. As with Standard Dynamic Range (SDR) content, we can expect HDR content to be available in variety of lighting styles such as low key, medium key and high key video. This raises concerns over tone-compatibility between HDR displays especially when adapting to various lighting styles. It is expected that dynamic range adaptation between HDR displays uses similar techniques as found with tone mapping and tone expansion operators. In this paper, we survey simple tone mapping methods of 4000 nits color-graded HDR content for 1000 nits HDR displays. We also investigate tone expansion strategies when HDR content graded in 1000 nits is displayed on 4000 nits HDR monitors. We conclude that the best tone reproduction technique between HDR displays strongly depends on the lighting style of the content.


Abstract – This paper proposes a brightness control method for retargeting legacy content on High Dynamic Range (HDR) displays. This method is based on a perceptual experiment that evaluates the brightness preference of users. We study the special case of screen content to better understand the effect of image statistics on human preference for brightness. The experiment concluded that brightness preference is highly content dependent. Using the data of the subjective study, we propose a new feature that analyses the source content to appropriately control the brightness of the HDR display. Based on this feature, we propose a brightness control algorithm that adjusts the luminance of display depending on the number of bright pixels in the content.


Abstract – Among the multiple new dimensions of Ultra High Definition, there is a wide consensus to consider High Dynamic Range (HDR) as the most valuable, both for its perceived "wow factor" and its modest upgrading cost compared to Standard Dynamic Range (SDR). An HDR-ready eco-system is currently being built from new cameras and production infrastructure to consumer displays and set top boxes. In order to support this migration towards HDR broadcast, an SDR to HDR tone expansion will be a key feature for playout servers, broadcast converters and encoders in the coming decade, in view of the existing SDR assets (both content and equipment). This paper presents the theory and results to a novel approach for tone expansion which preserves the intended style of the original video. It then discusses different use cases in where tone expansion will provide substantial benefits as well as certain limitations of the technology.


Abstract – Although HDR content processing, coding and quality assessment have been largely addressed in the last few years, little to no work has been concentrating on how to assess quality in HDR for 360° or omnidirectional content. This paper is an attempt to answer to various questions in this direction. As a minimum, a new data set for 360° HDR content is proposed and a new methodology is designed to assess subjective quality of HDR 360° content when it is displayed on SDR HMD after applying various tone mapping operators. The results are then analyzed and conclusions are drawn.
A.3 Patents


Appendix B

Standardization Contributions

We present our standardization contribution which is currently part of the UHD Forum Guidelines.

B.1 Conversion from SDR/BT.709 to PQ10/HLG10

In multiple areas of the production chain, it is anticipated that a requirement exists to “mix” or “switch” SDR/BT.709 sources and PQ10/HLG10 sources when constructing a Real-time Program Service. Mixing of this kind may take many forms as it does today in SDR only environments (practicing Rec. ITU-R BT.1886 gamma and BT.709 colorimetry). To make it possible to combine such elements, SDR/BT.709 sources must be converted into the PQ10 HLG10 domain with respect to both an industry compliant transfer function (e.g., Draft Rec. ITU-R BT.[HDR-TV] HLG, SMPTE ST 2084 PQ) and color space (i.e., Rec. ITU-R BT.709 primaries “mapped” into the Rec. ITU-R BT.2020 container). Such conversions can utilize:

- Remapping: SDR/BT.709 content is decoded and repackaged in the appropriate PQ10/HLG10 containers, but it does not change the color grading or dynamic range of the content; the content is simply mapped across to the same color and brightness values. When remapping SDR to HDR for PQ and HLG, the level of diffuse white should be considered. Diffuse white will typically map to approximately the middle of the coded range for both HLG and PQ. It should also be estimated that for default HLG (system gamma 1.2) nominal peak luminance is specified as 1,000 nits. Blind (not supervised) upconversion can lead to undesirable results so care should be used and any algorithms or equipment should be carefully evaluated.

- Up-conversion: SDR/BT.709 is decoded and then enhanced/modified to emulate PQ10/HLG10 and repackaged as above.
Both methods may be preferable for different circumstances. If the service provider intends to simulcast the UHD Phase A service in SDR/BT.709 for backward compatibility, then remapping may be preferable, because content that was originally SDR/BT.709 will remain exactly as intended in the legacy version of the service. Conversely, the service provider may prefer that all the segments of the Real-time Program Service look as uniform as possible, and thus up-conversion may be appropriate. For example, up-conversion may be preferred for live production mixing SDR/BT.709 and PQ10/HLG10 cameras, high quality SDR content, and legacy content like advertisements.

Under some circumstances, (e.g. viewing in a darkened room) HDR displays and content can cause discomfort if consideration is not given to the viewer’s vision adapting to the average light level on the screen at any given moment. For instance, if a feature program has a low average light level such as a night scene and the picture abruptly cuts to a high average luminance scene in an interstitial, the viewer may experience discomfort similar to that experienced with changing light conditions in nature. When advertisements are inserted into content, consideration should be given with respect to transitions from dark to light.

The described conversions typically will be performed by dedicated devices using a combination of 1d and 3d LUTs or other appropriate technology. Devices of this type may be used for both SDR/BT.709 to PQ10/HLG10 or vice versa as the production and the capability of the equipment in use requires.

A real-time dedicated conversion device is essential for some use cases, which may be encountered throughout the production chain, such as:

• Mix of SDR/BT.709 and PQ10/HLG10 live sources
  Broadcasting of live events (typically sports) in PQ10/HLG10 may require a relatively high number of cameras and it is probable that all these cameras will not be HDR-capable. In that situation, SDR/BT.709 cameras can be utilized if the conversion process is implemented

• SDR/BT.709 interstitials in a PQ10/HLG10 program
  With SDR/BT.709 interstitials, the interstitial content will likely come from a playout server. In this case the conversion process has to be implemented either at ingest, at the output of the playout server, or at the input of the mixer.

• Use of SDR/BT.709 content
  Extensive libraries of SDR/BT.709 content may be used, for example a live sports match production that includes archive footage from previous matches; such material needs to be converted to PQ10/HLG10 (using a near real-time file-to-file converter) before entering the video pipeline
Converting content to a common set of HDR and WCG technologies can occur at different points in the workflow from Production to the Consumer premises.

At the Production stage, multiple versions of the content may be produced. In the case of Prerecorded content, producers have the opportunity to create multiple versions of the content applying creative judgement to each version. This is recommended for all Pre-recorded content and especially for interstitial material that may be inserted into a variety of Real-time Program Services, which may be operating with different HDR/WCG configurations. Live content may also be output in multiple formats; however, time constraints may prevent highly detailed artistic input. Automated conversion technologies can be “tuned” by the content creator to make the conversion the best it can be for the given program.

At the Broadcast Center or Service Provider stage, content can be converted to the provider’s chosen HDR/WCG configuration using automated tools. This may not include program-specific creative input; however, professional equipment may produce acceptable results.

At the Consumer premises, tone mapping may be possible for the purpose of backward compatibility. Also, display devices may “map” content internally to best suit the display characteristics. Both of those processes operate on a content stream with one, constant dynamic range and color gamut container. Real-time Program Service providers should not expect consumer equipment to render seamless transitions between segments of content that have different transfer function or color gamut containers.