ANNÉE 2014

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

École doctorale Matisse

présentée par

Ronan BOITARD

préparée à l'unité de recherche IRISA
Rennes Bretagne Atlantique

Temporal Coherency in Video Tone Mapping

Thèse soutenue à Rennes le 16 Octobre 2014

devant le jury composé de :

Rafał MANTIUK
Ass. Professor, Univ. of Bangor / *Rapporteur*

Sumanta PATTANAIK
Ass. Professor, Univ. of Central Florida / *Rapporteur*

Alan CHALMERS
Professor, Univ. of Warwick / *Examinateur*

Frédéric DUFAUX
Research Director, Télécom ParisTech / *Examinateur*

Luce MORIN
Professor, INSA Rennes / *Examinateur*

Rémi COZOT
Ass. Professor, Univ. of Rennes 1 / *Examinateur*

Dominique THOREAU
Researcher, Technicolor / *Examinateur*

Kadi BOUATOUCH
Professor, Univ. of Rennes 1 / *Directeur de thèse*

Abstract

One of the main goals of digital imagery is to improve the capture and the reproduction of real or synthetic scenes on display devices with restricted capabilities. Standard imagery techniques are limited with respect to the dynamic range that they can capture and reproduce. High Dynamic Range (HDR) imagery aims at overcoming these limitations by capturing, representing and displaying the physical value of light measured in a scene. However, current commercial displays will not vanish instantly hence backward compatibility between HDR content and those displays is required. This compatibility is ensured through an operation called tone mapping that retargets the dynamic range of HDR content to the restricted dynamic range of a display device. Although many tone mapping operators exist, they focus mostly on still images. The challenges of tone mapping HDR videos are more complex than those of still images since the temporal dimensions is added. In this work, the focus was on the preservation of temporal coherency when performing video tone mapping. Two main research avenues are investigated: the subjective quality of tone mapped video content and their compression efficiency.

Indeed, tone mapping independently each frame of a video sequence leads to temporal artifacts. Those artifacts impair the visual quality of the tone mapped video sequence and need to be reduced. Through experimentations with HDR videos and Tone Mapping Operators (TMOs), we categorized temporal artifacts into six categories. We tested video tone mapping operators (techniques that take into account more than a single frame) for the different types of temporal artifact and we observed that they could handle only three out of the six types. Consequently, we designed a post-processing technique that adapts to any tone mapping operator and reduces the three types of artifact not dealt with. A subjective evaluation reported that our technique always preserves or increases the subjective quality of tone mapped content for the sequences and TMOs tested.

The second topic investigated was the compression of tone mapped video content. So far, work on tone mapping and video compression focused on optimizing a tone map curve to achieve high compression ratio. These techniques changed the rendering of the video to reduce its entropy hence removing any artistic intent or constraint on the final results. That is why, we proposed a technique that reduces the entropy of a tone mapped video without altering its rendering. Our method adapts the quantization to increase the correlation between successive frames. Results showed an average bit-rate reduction under the same PSNR ranging from 5.4% to 12.8%.

Acknowledgments

First of all, I would like to thank my three supervisors: Kadi Bouatouch, Rémi Cozot and Dominique Thoreau. Each of you helped me with different aspects of this thesis and were complimentary of each other. Thank you for the support, fruitful discussions and most importantly all the time you spent advising me.

I would also like to thanks my colleagues from Technicolor and the FRSense team to provide a work place that is more than just a place to work. I will definitely miss the coffee breaks, lunch breaks and any type of breaks we had. I am really glad to have met you all.

Many thanks as well to Adrien Gruson, Mickaël Ribardière and Ricardo Marquès for designing and rendering the computer generated sequences so badly needed for all the tests.

A special thanks to all the members and participants of the COST IC1005 action. It has been a pleasure to attend to all those meetings, training schools and workshops and to meet so many wonderful people.

Finally, I would like to thank all my friends for all the week-ends, trips, music festival and Amaryllis nights. Although Mondays were tough sometimes, these activities allowed me to keep the balance between work and social life.

Contents

Chapter 1

Introduction

Since the dawn of the field of photography, photographers and scientists have striven to solve two main issues:

- how to capture a scene with the highest fidelity?
- how to best reproduce the perception of a human observer when the capture of a scene is projected on a targeted display?

To achieve both of these goals, the field of photography has been in constant evolution since the first picture taken by Nicéphore Niépce in the mid-1820s. Nevertheless, current capture and display technologies are still limited in regard to the dynamic range that they can achieve as illustrated in Figure 1.1.

The first observation that we can make from this figure is that a camera adapts the captured dynamic range to a scene. Indeed, in cameras, a parameter **eV** (exposure Value) allows us to tune the exposure of a captured scene. Second, displays lack such an adaptation and can only achieve luminances lying in a fixed range. Finally, most cameras capture a higher dynamic range than can be directly reproduced on a display. From these three observations, we understand that a scene captured by a camera needs to be adapted to the targeted display characteristics.

In digital photography, pixel values correspond to a relative value of a standard representation. The conversion of physical values, measured by sensors, to this standard representation is performed in the camera and is defined by a camera response function. The resulting image can then be displayed on any commercial display, its rendering depends on the display. If the dynamic range of a camera is wider when compared to what the chosen standard pixel format can represent, then information gets lost. To summarize, the traditional digital imaging pipeline suffers from several shortcomings:

- during the capture, information gets lost as a camera cannot record all the dynamic range present in a scene,
- during the storing, information captured by the sensor's camera is adapted to the standard representation capabilities,

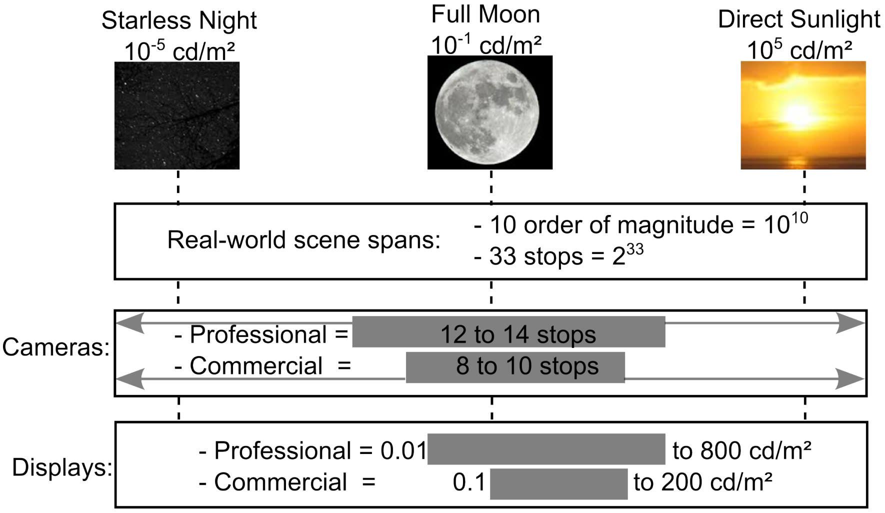


Figure 1.1: Dynamic range available for capture and display compared with physical scene. A stop correspond to a standard power-of-2 exposure step, that is to say 1 stop more means that the double amount of light is recorded.

- during the rendering, pixels represented by relative values are interpreted by the targeted display, consequently the perception of an image can greatly vary from one display to another.

To overcome these issues, new techniques aiming at capturing, representing and displaying a scene have been developed during the last decade. These techniques encompass what is referred to as High Dynamic Range (HDR) imaging. HDR imaging solves the aforementioned issues by:

- capturing most of all the luminance information present in a scene through bracketing techniques,
- storing the recorded values in absolute physical units to prevent relative interpretation,
- displaying those pixels on HDR monitors that emit this absolute physical quantity.

The main concepts of HDR imaging are presented in Chapter 2.

1.1 Tone Mapping

It is in this context of transition between two digital imaging techniques (HDR and LDR) that the tone mapping (also written as tonemapping) field takes its roots. On the one hand, we have the HDR imaging which represents all physical values of light that the human visual system can perceive. On the other hand, we have LDR imaging which can only represent a small fraction of the visible color gamut and store perceptually encoded values that correspond to a standard of representation. The backward compatibility between HDR content and Low Dynamic Range (LDR) displays is ensured by a tone mapping operation.

Tone mapping an HDR image amounts to retargeting physical values, with a virtually unlimited bit-depth, to a constrained space (2^{2n} color hue over 2^n tonal level, n being the targeted bit-depth). This limited bit-depth means that many similar HDR values will be tone mapped to the same LDR one. Consequently, contrast between neighboring pixel as well as spatially distant areas will be reduced. Furthermore, LDR displays have a low peak luminance value when compared to the luminance of a real scene. Consequently, captured color information will have to be reproduced at different luminance level. To summarize, tone mapping an HDR image amounts to finding a balance between the preservation of details, the spatial coherency of the scene and the fidelity of reproduction. This balance is usually achieved by taking advantage of the many weaknesses of the human visual system. Finally, the reproduction of a scene is sometimes constrained by an artistic or application dependent intent. That is why, a lot of Tone Mapping Operators (TMOs) have been designed with different intents: from simulating the human vision to achieving the best subjective quality.

Due to the lack of HDR video content and the assumption that TMOs developed for images would behave correctly for videos, the tone mapping field has for a long time only focused on still images. However, with the rising interest of both the digital cinema industry (ACES workflow) and the television broadcasters (demos at NAB, IBC, etc.) in HDR video content, tone mapping HDR video sequence has received a great deal of attention lately. That is why, the goal of this thesis is to assess the maturity of the video tone mapping field especially in regard to the mass-distribution of HDR content to the end-user. More precisely, we want to know if the current tone mapping techniques are robust enough to tone map HDR videos without user-interaction. Furthermore, those tone mapped content will need to be distributed to the end-consumer before being displayed. Indeed, uncompressed video content (HDR or LDR) are represented by a too large amount of data to fit the storage or broadcast requirements of current video processing pipelines. Work has been performed on tone mapping and video compression, but focused on optimizing the tone map curve to achieve high compression ratio. These techniques modify the mapping performed by a TMO to reduce the entropy of a sequence. However, this modification alters the visual perception of this content and can impair any artistic intent or desired rendering. We would like to increase the compression efficiency of tone mapped video content without altering their rendering.

1.2 Structure of the Thesis

The first chapter of this thesis provides the necessary knowledge to understand how HDR imaging techniques capture, represent and display more luminance levels than traditional LDR imagery. A special attention is given to the tone mapping operation.

In Chapter 3, we first verify the assumption regarding the application of a TMO designed for still images to an HDR video. We generated several HDR synthetic video sequences to test the behavior of the different types of TMO present in the literature. Those experiments led us to identify six types of temporal artifact that occur when applying separately a TMO to each frame of an HDR video sequence: Global and Local Flickering Artifacts (**FA**), Temporal Noise (**TN**), Temporal Brightness Incoherency (**TBI**), Temporal Object Incoherency (**TOI**) and Temporal Hue Incoherency (**THI**).

We then provide a description of the few existing video tone mapping techniques, namely techniques that rely on other frames than the current one to perform the tone mapping. Those methods solve only three out of the six types of described artifact: **Global and Local FA and TN**. Finally, we identify two additional types of temporal artifact caused by video tone mapping techniques.

Through the study performed in Chapter 3, we observe that three types of temporal artifact are not yet accounted for. Consequently, we present in Chapter 4 a technique to deal with these three types of artifact. However, a subjective evaluation reported that our method was only efficient when those temporal artifacts are of global nature. We then modified, in Chapter 5, our technique to make it local. Finally, we conducted a subjective evaluation to evaluate whether reducing those artifacts increases the subjective quality of tone mapped video content.

In Chapter 6, we propose to analyze the relationship between video tone mapping and video compression. More precisely, we study the relation between the compression efficiency of tone mapped video content and the preservation of temporal coherency in video tone mapping. We show that the choice of a TMO greatly influences the compression ratio that a codec can achieve and hence the quality of decoded content for targeted bit-rates.

However, changing the TMO to achieve higher compression efficiency amounts to changing the rendering of the tone mapped video. That is why, we propose a technique to increase the compression efficiency of any tone mapped video content without altering its rendering. This technique can be adapted to any TMO and is presented in Chapter 7.

Finally, a summary as well as a discussion on future work related to this thesis are presented in Chapter 8.

1.3 Contributions

The work presented in this thesis brings the following contributions to the video tone mapping field:

- A state of the art on HDR imagery.
- A description of the different types of artifact occurring when applying a TMO, designed for still images, to an HDR video.
- A survey of video tone mapping.
- A corpus of synthetic video sequences specifically designed to test TMOs.

- Two post-processing techniques to reduce temporal artifacts in video tone mapping.
- A survey of tone mapping and video compression.
- A technique to increase the compression efficiency of tone mapped video content without altering its rendering.
- An analysis of most of the current HDR video sequences publicly available.

1.4 Publications

Most of the work presented in this thesis is published in the following papers:

International Journals

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal Brightness Coherency for Video Tone Mapping," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 229-246, 2014.

International Conferences

- **R. Boitard**, K. Bouatouch, R. Cozot, D. Thoreau and A. Gruson, "Temporal Coherency for Video Tone Mapping," in *Proc. SPIE, Applications of Digital Image Processing XXXV*, 2012.
- **R. Boitard**, D. Thoreau, R. Cozot and K. Bouatouch, "Impact of Temporal Coherence-Based Tone Mapping on Video Compression," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013.
- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Motion-Guided Quantization for Video Tone Mapping," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2014.
- A. Le Dauphin, **R. Boitard**, D. Thoreau, Y. Olivier, E. Francois and F. LeLéannec, "Prediction-Guided Quantization for Video Tone Mapping," *Proc. SPIE, Applications of Digital Image Processing XXXVII*, 2014.

International Workshops

- **R. Boitard**, D. Thoreau, K. Bouatouch, and R. Cozot, "Temporal Coherency in Video Tone Mapping , a Survey," in *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, 2013.
- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Survey of temporal brightness artifacts in video tone mapping," in *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014.

- D. Touzé, Y. Olivier, S. Lasserre, F. Leléannec, **R. Boitard** and E. Francois, "HDR Video Coding based on Local LDR Quantization," in *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014.

Chapter 2

Background in High Dynamic Range Imaging

High Dynamic Range (HDR) imaging aims at capturing, representing and reproducing the physical value of light [Reinhard et al., 2010, Myszkowski et al., 2008, Banterle et al., 2011, McCann and Rizzi, 2011]. Apart from being able to represent more data with a virtually unlimited bit-depth, HDR imaging will most likely reduce the difference of rendering between displays. For all those reasons, both academic (COST HDRi 1005 action) and industrial (Technicolor, Dolby, Philips, etc.) scientists strive to bring this technology to the end-consumer. Although HDR concepts allow us to capture and reproduce all the color gamut and luminance perceived by a human observer, the limitations of both capture and display technologies prevent a one to one correspondence between a real scene and its display on a monitor.

Furthermore, capture and reproduction are not the only processing required to enjoy a multimedia experience. Figure 2.1 illustrates the many types of signal processing that HDR contents undergo before reaching the consumer. Content creation corresponds to either the capture of a real-word scene or the generation of a synthetic image or video. Post-processing encompasses all the creation processes to add artistic intent on content, for example color grading or addition of virtual effects. Encoding allows to compress the amount of data required to represent any content, the encoding can be lossless or lossy. The distribution to the consumer can be performed by either storing content on a storage disc (DVD, Blu-Ray, etc.) or by broadcasting it over a network (Digital Television Broadcast, Internet Streaming, etc.). Decoding is the reverse operation of the encoding and consists in reconstruction the video from the bit-stream stored on a storage disc or broadcasted over a network. Finally, the display is the interpretation by a display of the pixel values.

This chapter describes briefly basic concepts of HDR imaging and is structured as follows:

- Section 2.1. **Fundamentals**: introduces some basic concepts required to understand HDR imaging.

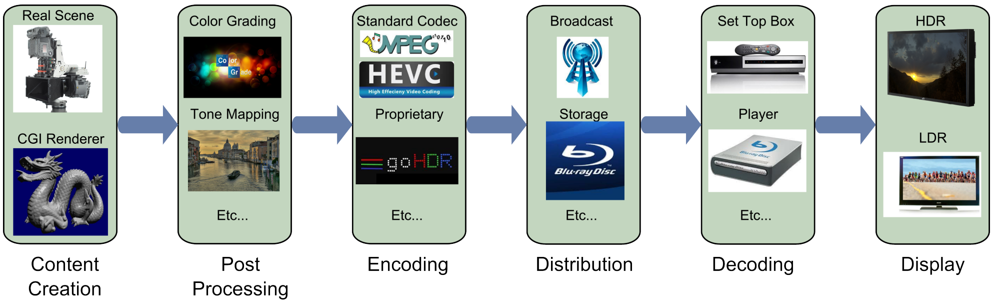


Figure 2.1: HDR pipeline from content generation to the end-consumer's display

- Section 2.2. **HDR Imaging**: describes the motivations behind HDR imaging as well as how to generate and store HDR content.
- Section 2.3. **Display of HDR/LDR Video Content**: details the technique needed to prepare HDR or LDR video content for LDR or HDR displays.
- Section 2.4. **Video Compression**: gives an overview of compression standard and techniques to deal with the storage and broadcast of HDR/LDR content.

2.1 Fundamentals

HDR imaging deals with the capture of physical real values of light and color as well as their representation on a display to achieve the closest reproduction from a human observer point of view. This section gives an introduction to these concepts in order to better understand the purpose of HDR imaging.

2.1.1 Light

Light is radiant energy that is measured in **Joules**. Light can be represented as a mix of electromagnetic radiations, with each radiation corresponding to a wavelength. The light emitted by the sun is characterized by a spectrum of wavelengths. The human eye can only perceive a small part of this spectrum, from red to violet. Figure 2.2 represents the full visible spectrum and its associated wavelength range. Two radiometric quantities are used to quantify light: **irradiance** and **radiance**. Irradiance describes the quantity of light incoming upon a unit area (i.e. $m^2 \dots$) while radiance describes the power (flux) emitted in a direction (radiance is expressed in $W/m^2/sr$, Watt per square meter per steradian). The term **luminance** represents the photometric quantity of light arriving at the human eye. When multiplying the radiance by the standard luminous efficacy of equal energy white light (in lumen per watt, lm/W), we obtain the luminance (represented by candela per square meter, cd/m^2).

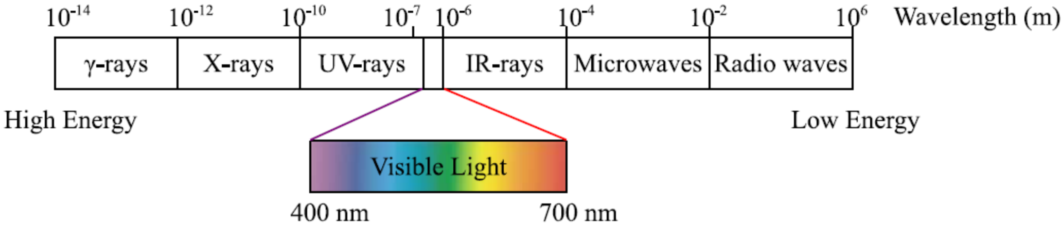


Figure 2.2: Spectrum of the sun light and the full visible spectrum

2.1.2 Human Vision System

Photoreceptors The retina of the human eye includes 130 million cells sensitive to light and called photoreceptors. Photoreceptors, via a process called phototransduction, convert luminous energy (photons) into an electric signal. This signal is transmitted thought the optic nerve and allows the brain to interpret the viewed scene. The photoreceptors respond to the number of incoming photons as well as their wavelengths. [Naka and Rushton, 1966] fitted the photoreceptors response as:

$$\frac{R}{R_{max}} = \frac{C^n}{C^n + C_{50}^n} + b, \tag{2.1}$$

where R and R_{max} are the neuronal response and maximal response of the photoreceptors to the grating stimuli C, C_{50} represents the semi-saturation constant, meaning the value for which half of the upper bound is reached, n is the photoreceptor's sensitivity and b the background discharge. This response corresponds to an S-shaped curve on a log-linear plot, with saturation equal to 0 (because it is impossible to have negative occurrence of a phenomenon) and a maximal intensity.

However, unless by drastically decreasing the sensitivity (i.e. steepness of the slope), this response cannot cover all the wide range of luminance that we can perceive. This is due to the fact that although we can perceive luminance ranging from 10^{-6} to 10^6 cd/m^2, we cannot perceive this range simultaneously. Instead, we need to adapt to our environment.

In order to adapt, two types of photoreceptor cells exit: cones (5 million) and rods (125 million) [Reinhard et al., 2010]. Cones are principally gathered in the center of the retina, also called the fovea. They are responsible for the perception of color, finer details and rapid changes. Three kinds of cone with different peak wavelength sensitivities exist: Long-564 nm (red), Medium-533 nm (green) and Short-437 nm (blue) [Fairchild, 2005]. The colors that we perceive are obtained thanks to the combination of these three types of cones. The difference in peak wavelength sensitivity means that two packets of light with different spectral power distributions can produce the same response, which is called metamerism [Hunt, 2005]. Cones can only operate in daylight (photopic vision ranging from 0.01 to 10^6 cd/m^2) and stop functioning when there is not enough light entering the eye.

Rods are highly sensitive to light (i.e. number of photons), they are scattered over the outside fovea and are used in peripheral vision. They are responsible for low light levels (scotopic vision ranging from 10^{-6} to 10 cd/m^2). As there is only one type of rod sensitive around the blue green wavelength, we experience a lack of color at night. This property, namely the changing of color appearance when illumination changes, is known as the Purkinje effect (a.k.a the Purkinje shift) [Barlow, 1957].

Cones and rods can operate simultaneously in low but not quite dark lighting situations (i.e. when the luminance level ranges between 0.01 and 10 cd/m^2). This is called mesopic vision regime.

Threshold versus Intensity (TVI) Back in 1834, Weber hinted that when adapted to a uniform background, the perception of a variation of intensity is usually linearly proportional to the background intensity. Consider a person carrying a 1 kg weight, if we add 0.1 kg on top of it, the person will perceive the difference of weight. However, if the initial weight is 10 kg, the adding of 0.1 kg will be hardly noticeable. This relationship is known as Weber's law.

This relation has been applied to the field of light perception and several experiments were conducted to verify it [Blackwell, 1981, Barten, 1992]. These experiments are usually made by increasing the intensity of a gray patch on a uniform gray background. When the patch becomes visible, the Just Noticeable Difference (JND) is reached. These experiments shaped the Threshold Versus Intensity (TVI) functions that proved that Weber's law is only true in the range where the photoreceptors are fully responsive (see Figure 2.3). Two equations fit the experimental results:

$$log_{10}T_s(\mathbf{L_w}) = \begin{cases} -2.86 & log_{10}(\mathbf{L_w}) \leq -3.94 \\ log_{10}(\mathbf{L_w}) - 0.395 & log_{10}(\mathbf{L_w}) \geq -1.44 \\ (0.405 log_{10}(\mathbf{L_w}) + 1.6)^{2.18} - 2.86 & \text{otherwise} \end{cases} \tag{2.2}$$

$$log_{10}T_p(\mathbf{L_w}) = \begin{cases} -0.72 & log_{10}(\mathbf{L_w}) \leq -2.6 \\ log_{10}(\mathbf{L_w}) - 1.255 & log_{10}(\mathbf{L_w}) \geq 1.9 \\ (0.249 log_{10}(\mathbf{L_w}) + 0.65)^{2.7} - 0.72 & \text{otherwise} \end{cases} \tag{2.3}$$

where T_s (respectively T_p) represents the TVI response function in the scotopic (respectively photopic) vision regime and $\mathbf{L_w}$ the luminance [Ferwerda et al., 1996].

2.1.3 Colorimetry

Colorimetry is the field of assigning code values to perceived colors. There are several ways to represent perceived colors using color spaces. A color space is an abstract mathematical representation designed to describe the way color can be represented as a combination of code values (i.e. color components or color channels). Furthermore, the term gamut is used to designate a complete range or scope of a color space.

Many color spaces exist and have different purposes. We distinguish two types of color space: full gamut and display-dependent. In 1931, the "Commission Internationale de l'éclairage" (CIE) defined the standard CIE 1931 XYZ color space which includes

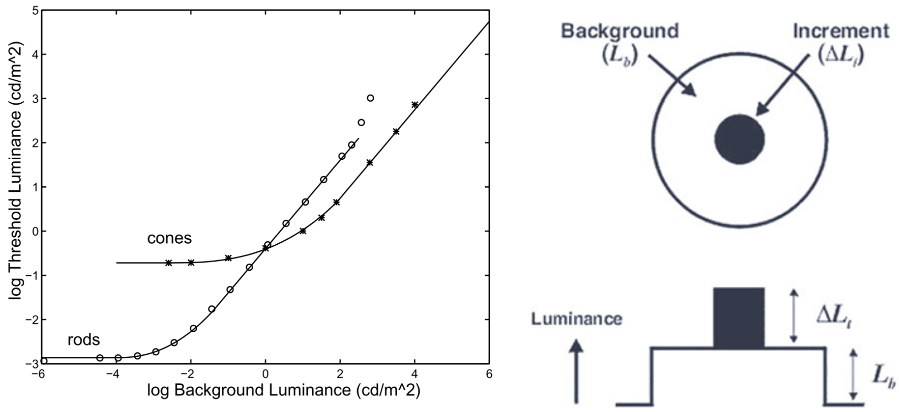


Figure 2.3: Left: plot of TVI functions for scotopic and photopic (taken from [Ferwerda et al., 1996]). Right: example of threshold versus intensity detection patch.

all of the visible gamut [Smith and Guild, 1931]. In 1976, the CIE defined two other color spaces that are approximately perceptually uniform, meaning that a difference in the value will correspond to the same difference in perception. They are the CIE 1976 L*,a*,b* color space (commonly referred to as CIELAB) and the CIE 1976 L*,u*,v* color space (commonly referred to as CIELUV). These color spaces are used for computations but do not address a display directly. To achieve that, we need to convert these values to a display-dependent color space.

Display-dependent color spaces allow the representation of color on a display accordingly to a standard. They use a limited bit-depth to represent the light intensity and color information. The most common display-dependent color space is described by the standard ITU-R Recommendation BT.709 (also known as Rec.709) [ITU, 1998]. Figure 2.4 illustrates the proportion of the full visible gamut that the Rec.709 color space covers. When these standards were defined, they were based on a maximum luminance intensity of the current display technology (i.e. 100 cd/m^2). They used an 8 bits integer quantization to sample the color space, allowing to represent 256 levels of gray (2^8 values ranging from 0 to 255). Another color space like AdobeRGB covers a wider gamut while having a coarser sampling of this gamut. In a nutshell, there is a trade-off between the coverage of the used gamut and the distance between two colors of slightly different hues.

2.2 HDR Imaging

This section provides a small history of digital HDR imaging. We first recall the motivations behind HDR imaging before describing the file format used. Finally, we detail current techniques used to generate HDR content.

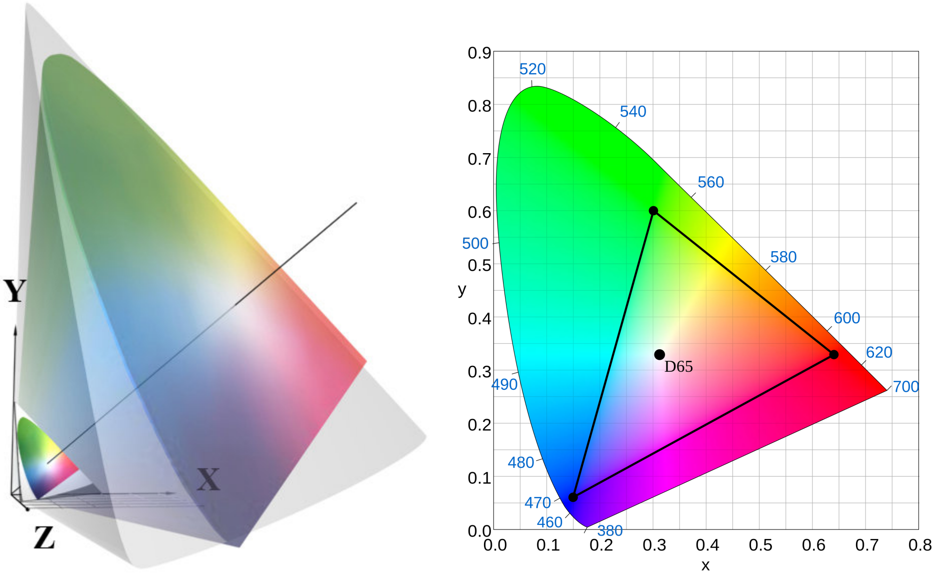


Figure 2.4: Left: 3-D representation of the CIE 1931 XYZ color space [Smith and Guild, 1931]. Right: CIE 1931 xy chromaticity diagram. The BT.709 color space along with the location of its primary colors are represented in the triangle. BT.709 uses Illuminant D65 as white point [ITU, 1998].

2.2.1 Computer Graphics Content

Not all images or video content are captured from a real-world scene, some are computer generated. This is the field of Computer Graphics (CG) which creates virtual images based on a 3-Dimensional (3-D) model of a scene. By simulating the propagation of light, one can compute the contribution of all the light sources to each pixel of a computer generated image. The results is a physically-based representation of a 3-D model under a set of lighting conditions.

To accurately represent light propagation, computer graphics need a large range of values that represent physical light values with minimal quantization steps. For that reason, all computations are performed in floating point that represent either the radiance or the luminance in the scene. The main problem with floating values is that their storing as uncompressed data result in using 96 bits per pixel (bpp). This is four times the amount of space needed for a standard images. In the following section, a format that addresses the problem of encoding and storing this information will be described.

2.2.2 Radiance File Format

In 1991, Greg Ward [Ward, 1991] created the RGBE image format to store HDR images. This format was designed to store HDR images generated by the Radiance rendering software [Ward, 1994b]. By assuming that colors, when represented by **RGB** tri-stimulus values, are highly correlated, Ward proposed to share the exponent between the three color channels. Each pixel is stored in 32 bits, with 8 bits for each of the three color mantissas: $\mathbf{R_m}$, $\mathbf{G_m}$ and $\mathbf{B_m}$ and another 8 bits for the common exponent $\mathbf{E}$:

$$\mathbf{R_m} = \left\lfloor \frac{256\mathbf{R}}{2^{\mathbf{E}-128}} \right\rfloor, \mathbf{G_m} = \left\lfloor \frac{256\mathbf{G}}{2^{\mathbf{E}-128}} \right\rfloor, \mathbf{B_m} = \left\lfloor \frac{255\mathbf{B}}{2^{\mathbf{E}-128}} \right\rfloor, \text{ with} \tag{2.4}$$

$$\mathbf{E} = \lceil log_2(max(\mathbf{R}, \mathbf{G}, \mathbf{B}) + 128\rceil. \tag{2.5}$$

This format allows to cover 76 orders of magnitude (i.e. the number of powers of 10 that can be represented). However, this representation does not cover the full gamut of color. Converting an image to the XYZ color space before computing the shared exponent allows to solve this issue. This format is then referred to as the XYZE format. The RGBE or XYZE format covers a huge range of luminance, more than what the human eye can perceive.

2.2.3 OpenEXR File Format

By redistributing the quantization steps in a more human restricted dynamic range, it is possible to achieve a finer quantization over a more restrictive range. That solution is provided by the half floating point format which is part of the specification of the OpenEXR file format [Magic, 2008]. The OpenEXR file format relies on the IEEE 754 16-bit float standard [Hough, 1981] and is defined as:

$$H = \begin{cases} 0 & \text{if } (M = 0 \bigwedge E = 0) \\ (-1)^S 2^{E-15} + \frac{M}{1024} & \text{if } E = 0 \\ (-1)^S 2^{E-15} \left(1 + \frac{M}{1024}\right) & \text{if } 1 \leq E \leq 30 \\ (-1)^S \infty & \text{if } (E = 31 \bigwedge M = 0) \\ NaN & \text{if } (E = 31 \bigwedge M > 0) \end{cases} \tag{2.6}$$

where S is the sign, M the mantissa and E the exponent. This representation allows to cover around 10.7 orders of magnitude while using 48 bpp ($3 \cdot 16$ bits: $S = 1$ bit, $E = 5$ bits and $M = 10$ bits). The half-float pixel representation is used in the new standard Academy Color Encoding System (ACES) which is currently undergoing industry-wide production trials by the major Hollywood studios and standardization by the Society of Motion Picture and Television Engineers (SMPTE).

Other file formats can be used to store HDR values such as the LogLuv encoding [Ward Larson, 1998] that allows to encode HDR values within a TIFF image.

2.2.4 Real-World HDR capture

Recall that HDR images can be generated using a renderer to generate CG content. However, how to capture an HDR image of a real-world scene? Current commercially available sensors are limited with respect to the amount of light that they can record. It is usually not enough to recover all the light information present in an outdoor scene.

In addition, the resulting images use a display-dependent color space that does not represent the absolute physical values of light.

To recover physical values from a camera, one can calibrate a camera. Calibration consists in measuring the Camera Response Function (CRF), that is to say the code value assigned by a camera to a recorded physical value [Mann and Picard, 1994]. Once the CRF is known, it is possible to invert it to obtain the physical values (luminance) of the image.

The second limitation is the inability of current sensors to record all the light information present in a scene. Indeed, sensors cannot capture information smaller or higher than a certain threshold. The exposure Value (**eV**) setting defines these thresholds and pixels with a luminance lying outside them are clipped. In, photography, **exposure** is the amount of light per unit area reaching a photographic film sensor, as determined by shutter speed, lens aperture and scene luminance. Exposure is measured in lux seconds, and can be computed from the **eV** and scene luminance in a specified region. In photography jargon, exposure generally refers to a single capture for a given shutter speed, lens aperture and scene luminance.

To overcome this limitation, it is possible to capture the same scene several times with different exposures. Using the CRF, one can combine those exposures to obtain an HDR image with all the information present in the scene. This processing is called bracketing (Figure 2.5). Several types of bracketing techniques exist:

- Temporal bracketing consists in taking images with different exposures one after another [Debevec and Malik, 1997, Mitsunaga and Nayar, 1999].
- Spatial bracketing relies on a neutral filter superimposed onto a sensor to have spatially varying exposure [Schoberl et al., 2012].
- Multi-sensor bracketing uses beam-splitter to divide the light, which results in having different exposures with multiple sensors embedded in a single camera [Tocci et al., 2011].
- Multi-camera bracketing combines several cameras (usually two) with a rig and different neutral gray filters to generate the different exposures.

All these techniques have their pros and cons and no solution currently stands out. The common point is that these techniques capture the scene with different exposures in order to create an HDR content. However, with the improvement of sensor capabilities, it is highly probable that in the near future, sensors will be able to capture the wide range of visible light and color gamut without having to resort to bracketing techniques.

2.3 Display of HDR/LDR Content

CG renderers or bracketing techniques allow to generate HDR content that cover the wide range of physical luminance values of a real world scene. However, to address a display, content needs to be represented in a display-dependent color space, potentially

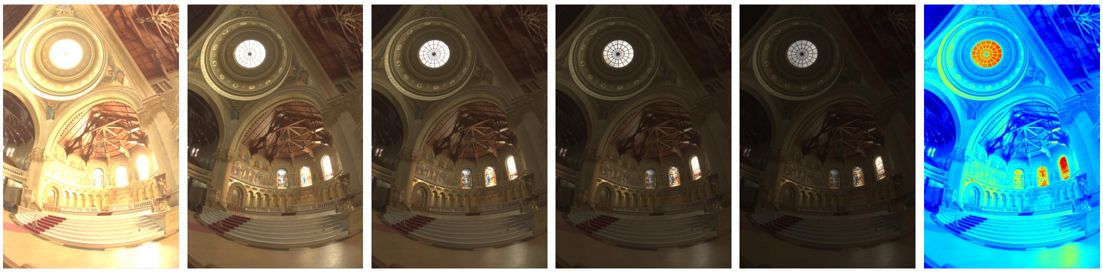


Figure 2.5: 1^{st} to 5^{th} images: different exposures of the same scene, two successive images are separated by 4 f-stops. Rightmost image: false color representation of the generated HDR image.

with a much more restricted dynamic range than an HDR scene. Creating HDR content raises two questions:

- How to reproduce HDR content on current commercial displays of more restricted capabilities?
- Assuming that displays capabilities increase, how to display LDR content on future HDR displays?

This section provides the necessary information to comprehend how to retarget HDR/LDR content to LDR/HDR ones. First, we describe current commercial display capabilities. Then, we detail the HDR to LDR conversion using Tone Mapping Operators (TMOs). Section 2.3.3 describes the future generation of display: HDR displays. Finally, the LDR to HDR conversion is presented in the last section.

2.3.1 Display Capabilities

As stated before, current displays technology cannot represent the full color gamut and luminance range existing in the real world. Most of those displays use only 8 bits for each of the three color components and can therefore represent 65,025 colors over 255 levels of gray (16,777,216 code values). The way these 3 bytes interact to shape color is defined through a display-dependent color space (Section 2.1.3).

Peak Luminance and Contrast In addition to the limited code values available to encode different colors or gray levels, displays have two other limitations: their peak luminance and their contrast. The peak luminance indicates the maximum amount of brightness that can be perceived by an observer while staring at the display. It is achieved when displaying the maximum code value represented by the tuple (R = 255, G = 255 and B = 255), also known as the white point of a display. Similarly, the black point of a display is represented by the tuple (R = 0, G = 0 and B = 0). On an LCD-based display, the black point is never truly 0 cd/m^2 because black is simulated by the crystals of the panel being completely shut, along with a polarized layer behind the crystals, to prevent light from the backlight to go through. Yet precisely because the

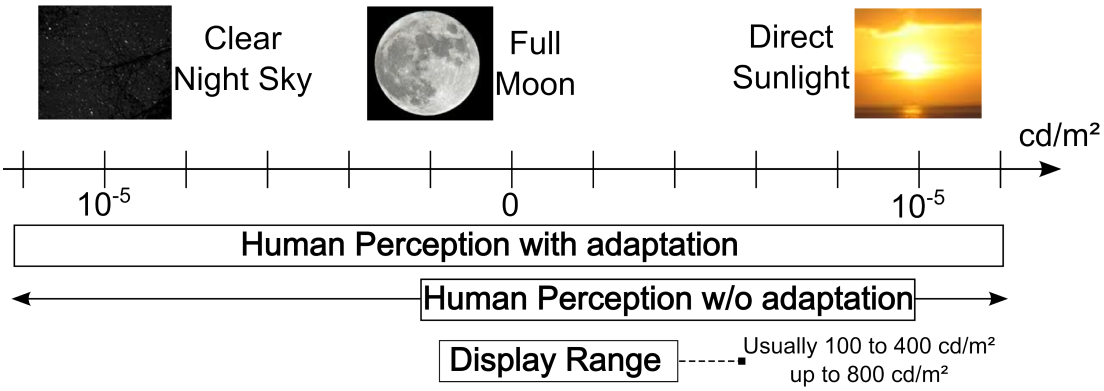


Figure 2.6: Display capabilities of commercial displays compared to the human perception of light. Note that recent displays can achieve up to 800 cd/m^2 while most of them range between 100 and 400 cd/m^2.

backlight is still on and the structure of the crystal array is not perfect, some amount of light will leak through the crystal and be seen. The contrast of a display is the ratio between the white and black point of a display. Figure 2.6 illustrates the range of commonly achieved luminance displayable in comparison to what the human eye can see.

Gamma Correction Apart from these limitations, Cathode Ray Tube (CRT) displays have also a non linear relationship γ between the input voltage V and the corresponding output light L_v:

$$\mathbf{L_v} = k\mathbf{V}^{\gamma}, \tag{2.7}$$

In order to correct this non-linearity, a gamma correction is applied, in the display, to each color channel:

$$\mathbf{R'} = \mathbf{R}^{\frac{1}{\gamma}}, \mathbf{G'} = \mathbf{G}^{\frac{1}{\gamma}}, \mathbf{B'} = \mathbf{B}^{\frac{1}{\gamma}}. \tag{2.8}$$

Although this non-linearity should not be present in current LCD displays, to preserve backward compatibility, they reproduce this effect.

Gamma Encoding The human eye does not perceive light linearly, as a camera would perceive. Indeed, if the number of photons hitting the sensor of a camera is doubled, then twice the signal output is recorded. On the other hand, doubling the amount of light would be perceived by an observer as just a fraction brighter, this fraction increasing for higher light intensity. Gamma encoding is a perceptual encoding of digital color pixels. By applying a gamma encoding function, we can fit a value stored in a file so that it corresponds more closely to what we would perceive as twice brighter, thus optimizing and equally distributing the range at our disposal.

Gamma encoding is not to be confused with gamma correction explained before. However, due to an odd bit of engineering luck, the native gamma of a CRT is 2.5, consequently, the gamma correction is not required anymore if the gamma encoding is designed such as it compensates the non-linearity of a display [Poynton, 1996].

2.3.2 HDR to LDR conversion

During the transition period between HDR and LDR imaging, captured HDR content will still need to be displayable on LDR monitors. Indeed, traditional display will not vanish instantly and hence backward compatibility is mandatory. The conversion from HDR content to LDR ones is performed by Tone Mapping Operators (TMOs). In HDR imaging, the pixels represent the physical scene luminance (expressed in cd/m^2) stored as floating point values. In the case of LDR imaging, the pixels are assigned code values corresponding to a display-dependent color space. Furthermore, the term luminance is replaced by luma, which corresponds to the weighted sum of gamma-encoded $\mathbf{R'G'B'}$ components of a color video.

In the introduction, we outlined the different challenges of tone mapping that are:

- the mapping of a wide range of luminances with unlimited bit-depth to a limited amount of code values (255 for 8 bits),
- the mapping of all colors perceivable by a human observer to a limited bit-depth (65,024 colors for each luma level with 8 bits),
- the reproduction of this color at a different luminance level,
- the preservation of spatial details (contrast) on a limited bit-depth,
- the preservation of any artistic intent imposed on the HDR content,
- the usage of this content for a specific application other than visualization.

Figure 2.7 illustrates the three steps that compose a TMO:

- **The mapping operation**: compresses HDR luminance values to fit in the range [0-1].
- **The gamma encoding**: redistributes the tonal level closer to how our eyes perceive them (see §Gamma Encoding in Section 2.3.1).
- **The quantization**: converts floating point values to integer code values corresponding to the used bit-depth (i.e. $[0; 2^n - 1]$ for n bits). This operation consists in scaling the gamma encoded values to the maximum value desired (i.e. $2^n - 1$) and then rounding them to the nearest integer.

The last two steps (gamma encoding and quantization) can be considered fixed although the γ value is tunable. That is why when describing a TMO, these two processes are usually left out. Table 2.1 summarizes the main notations that we use for the different steps of tone mapping.

The results of TMOs can be quite different since they depend on the targeted applications. Over the last two decades, a multitude of TMOs have been developed

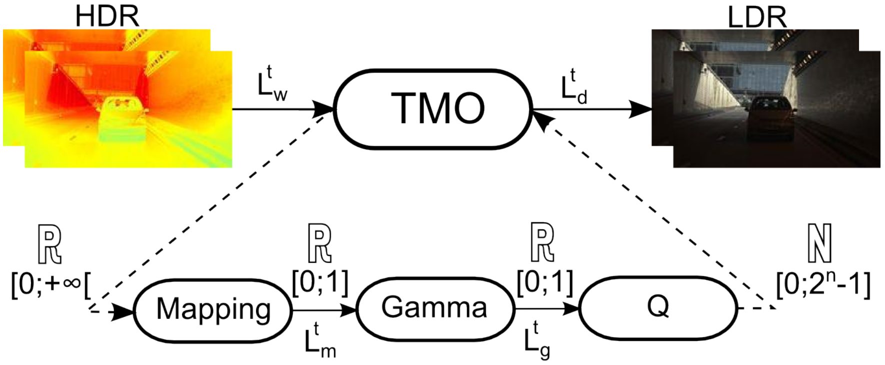


Figure 2.7: Workflow of the three steps needed to perform a tone mapping operation

$\mathbf{L_w}$	HDR luminance (cd/m^2) $[0; +\infty]$
$\mathbf{L_m}$	Tone mapped LDR luminance $[0; 1]$
$\mathbf{L_g}$	Gamma encoded LDR luminance $[0; 1]$
$\mathbf{L_d}$	Tone mapped LDR luma $[0; 2^n - 1]$
$\mathbf{I_{w,m,g,d}}$	3 Color channel image

Table 2.1: Used notations for the different steps of tone mapping.

[Reinhard et al., 2010, Banterle et al., 2011]. In this section, we propose to classify TMOs into 5 categories:

- **Global operators:** compute a monotonously increasing tone map curve [Tumblin and Rushmeier, 1993], [Ward, 1994a], [Reinhard et al., 2002], [Mantiuk et al., 2008].
- **Local operators:** map a pixel depending on information from its spatial neighborhood [Chiu et al., 1993], [Pattanaik et al., 1998], [Li et al., 2005].
- **Edge-Aware operators:** compress separately the edges and the image's background (low frequency component) [Tumblin, 1999], [Durand and Dorsey, 2002], [Fattal et al., 2002], [Farbman et al., 2008].
- **Human Visual System (HVS) operators:** simulate the behavior of the HVS [Pattanaik et al., 2000], [Ledda et al., 2005].
- **Color Appearance Model (CAM) operators:** reproduce, to the closest, colors of a scene [Fairchild, 2004], [Kuang et al., 2007a, Reinhard et al., 2012].

We propose for each category a brief description of operators that have proved to perform well after subjective evaluation [Yoshida, 2005], [Ledda et al., 2005, Kuang et al., 2007a]. We first present two global operators: [Mantiuk et al., 2008] and [Reinhard et al., 2002]. Then, as the latter exists in a local version, we also describe it. Two edge-aware operators are also detailed: [Fattal et al., 2002], [Durand and Dorsey, 2002] along with a generic implementation of edge-aware filters for tone mapping. [Ledda et al., 2004] operator is detailed as part of the HVS operator. Then we describe the iCAM06 [Kuang et al., 2007a] as part of CAM operators. Finally we present several subjective evaluations that assess the quality of TMOs. Note that all the presented TMOs have been designed for HDR images, their application to video sequences as well as video TMOs are described in Chapter 3.

Display Adaptive Tone Mapping: The main limitation of LDR display is their inability to reproduce high luminance values because their peak luminance ranges from 100 to 800 cd/m^2. In addition, displays can use different color gamuts and dynamic ranges. For all those reasons, [Mantiuk et al., 2008] proposed a TMO that provides the least perceptually distorted LDR image on a targeted display. This TMO adaptively adjusts the rendering of an HDR image based on the display's characteristics.

Figure 2.8 depicts the workflow of this TMO. First, the HDR image is tone mapped using the default parameters of the TMO. In the same time, the TMO computes the response R_{orig} of the Human Visual System (HVS) to the HDR image. Note that this response may be computed on an enhanced version (denoised, sharpened etc.) of the HDR image. Then the TMO applies an inverse display model on the resulting LDR image and computes the response R_{disp} of the HVS. This TMO computes a piece-wise tone map curve where the location of the curve's nodes are refined to minimize an error metric between R_{orig} and R_{disp} responses.

To inverse the display model, several characteristics of the targeted display are needed: the gamma value γ, the peak luminance display l, the black level b, the screen reflectivity k and the ambient illumination a. Figure 2.9 illustrates the rendering of 3 predefined display types: *LCD Office, LCD Bright* and *CRT*. The characteristics of the predefined displays can be found on the documentation page of *pfsTMO* [Grzegorz Krawczyk, 2007]. This TMO is among the best rated TMO in several subjective evaluations. As it uses a monotonously increasing tone map curve, it is classified as a global TMO.

Photographic Tone Reproduction Another global TMO well rated by subjective evaluations is the Photographic Tone Reproduction algorithm [Reinhard et al., 2002]. This TMO is based on photographic techniques and allows to choose the exposure of a tone mapped image. It uses a system designed by Adams [Adams, 1981] to rescale HDR images at a defined exposure:

$$\mathbf{L_s} = \frac{\alpha}{k}\mathbf{L_w}\,, \tag{2.9}$$

$$k = \exp\left(\frac{1}{n_p}\sum_{x=1}^{n_p}\log(d+\mathbf{L_w}(x))\right)\,, \tag{2.10}$$

where α is the chosen exposure, $\mathbf{L_w}$ the HDR luminance image and $\mathbf{L_s}$ the scaled luminance image. The geometric mean k (a.k.a. the key value) is an indication of an

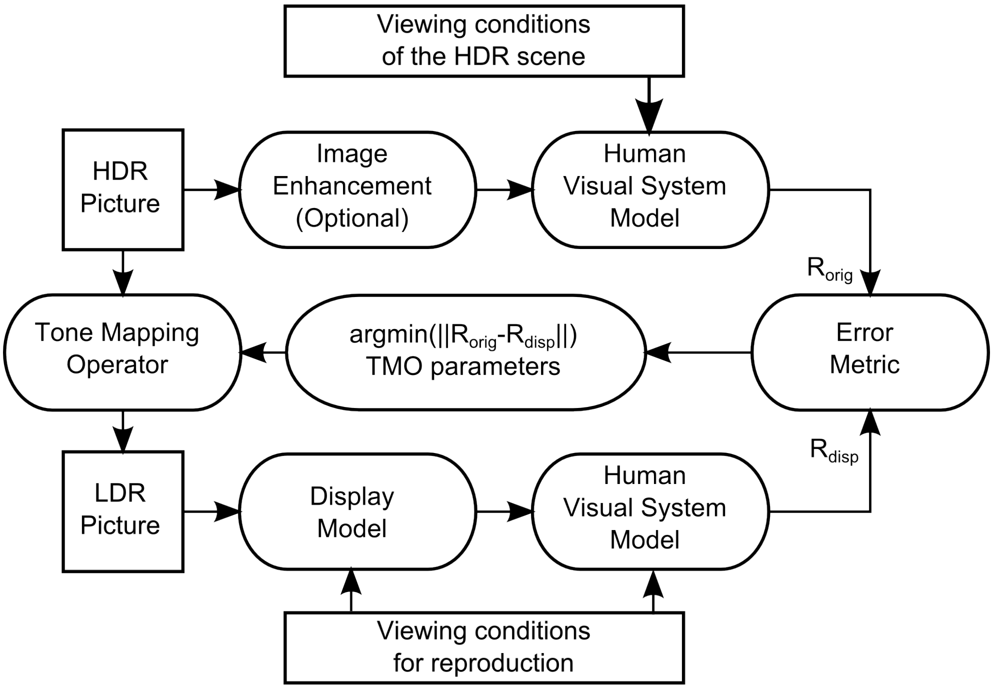

Viewing conditions of the HDR scene
HDR Picture
Image Enhancement (Optional)
Human Visual System Model
R_orig
Tone Mapping Operator
argmin(||R_orig-R_disp||) TMO parameters
Error Metric
R_disp
LDR Picture
Display Model
Human Visual System Model
Viewing conditions for reproduction

Figure 2.8: Workflow of the Display Adaptive operator [Mantiuk et al., 2008].

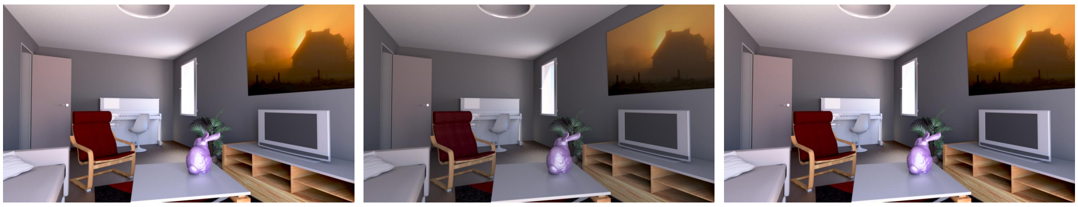


Figure 2.9: Comparison of three settings of [Mantiuk et al., 2008] operator. From left to right: LCD Office, *LCD Bright and CRT* setting.

image's overall brightness. It is computed using Equation 2.10, where d is a small value (i.e. $10^{-6} cd/m^2$) to avoid singularity and n_p the number of pixels in the image. The tone map curve is a sigmoid function given by :

$$\mathbf{L_m} = \frac{\mathbf{L_s}}{1+\mathbf{L_s}}\left(1+\frac{\mathbf{L_s}}{\omega^2}\right) , \tag{2.11}$$

where ω is used to burn out areas with high luminance value and $\mathbf{L_m}$ is the tone map LDR luminance. Two parameters (α and ω) are necessary to perform the tone mapping. In [Reinhard et al., 2002], these parameters are set to $\alpha = 18\%$ and ω to the maximum luminance value of $\mathbf{L_s}$.

A local version of this TMO also exists. Local operators usually compute an adaptation luminance $\mathbf{L_a}$ to adapt the mapping of a pixel to its spatial neighborhood. Most of them compute $\mathbf{L_a}$ using a Gaussian pyramid. Recursive application of a Gaussian

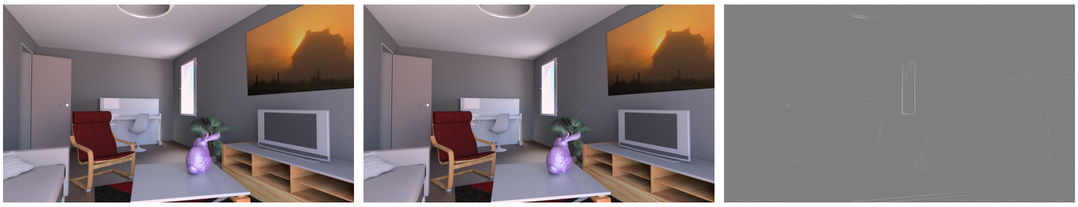


Figure 2.10: Comparison of the global and local version of [Reinhard et al., 2002] operator. From left to right: global, local and difference in luma.

filter decomposes an image into several low-frequency subbands with different cut-off frequency [Burt, 1981]. Local TMOs usually choose one layer or combine all the layers of a Gaussian pyramid.

This operator however, chooses, for each pixel, the layer of the pyramid that best approximates its neighborhood. To achieve that, the difference between successive layers is normalized and a threshold allows to select the right layer for each pixel. Equation 2.11 is modified to include $\mathbf{L_a}$:

$$\mathbf{L_m} = \frac{\mathbf{L_s}}{1+\mathbf{L_a}}\left(1+\frac{\mathbf{L_s}}{\omega^2}\right). \tag{2.12}$$

Figure 2.10 illustrates both global and local tone mapping results of the same HDR image.

Gradient Domain Compression In 2002, a new trend in tone mapping appeared: edge-aware tone mapping. Starting from the fact that we are more sensitive to contrast than to absolute values, this type of TMO compresses edges and background differently. The Gradient Domain Compression algorithm [Fattal et al., 2002] performs the tone mapping in the gradient domain. A gradient field $\mathbf{\Delta H}$ is computed at each level of a Gaussian pyramid. A scaling factor is then determined for each pixel of each layer based on the magnitude of the gradient:

$$\varphi_{\mathbf{k}} = \frac{\alpha}{\|\mathbf{\Delta H_k}\|}\left(\frac{\|\mathbf{\Delta H_k}\|}{\alpha}\right)^{\beta}, \tag{2.13}$$

where $\|\mathbf{\Delta H_k}\|$ is the gradient field of the layer k while α determines which gradient magnitude remains unchanged. Gradients of larger magnitude than α are attenuated (assuming that $\beta < 1$), while gradients of smaller magnitude are magnified. In [Fattal et al., 2002], these parameters are set to 0.1 times the average gradient magnitude for α, and to a value between 0.8 and 0.9 for β. The scaling factors are propagated and accumulated from level to level in a top-down fashion. Although these scaling factors are computed using a Gaussian pyramid, they are used only to manipulate the gradients of the finer resolution to prevent halo artifacts. As the modified gradient field may not be integrable, the LDR image is computed by finding the output image whose gradient is the closest to the modified gradient field (in a least mean square fashion).

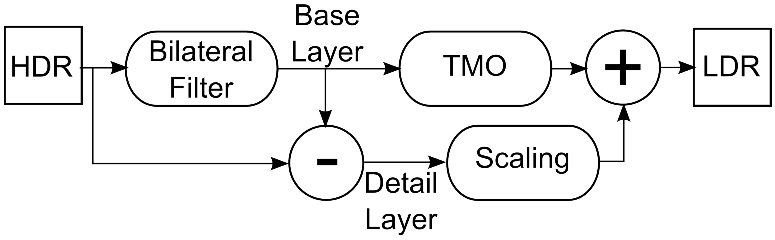


Figure 2.11: Possible workflow of the bilateral filter when used for tone mapping. A different implementation scales only the base layer and leave untouched the detail layer.

Bilateral Filtering Another edge-aware operator separates high frequency subband (detail layer) from the low frequency subband (base layer or background) using the bilateral filter [Tomasi and Manduchi, 1998]. Recall that a Gaussian filter weights the pixels of a spatial neighborhood based on their spatial proximity to the central pixel [Burt, 1981]. The bilateral filter adds another dimension, the photometric distance (difference in intensity). The closer in intensity a pixel's value is, the more it should contribute to the weighted result.

To apply this filter to tone mapping, an HDR image is separated into a base layer (low frequency subband) and a detail layer (high frequency subband). Figure 2.11 illustrates the workflow to tone map an HDR image using the bilateral filter [Durand and Dorsey, 2002]. Once the filter is applied to the HDR image, the base layer is removed from the original image to obtain the detail layer. The base layer is fed to a TMO while the detail layer is scaled to sharpen/smooth the resulting LDR image. Any TMO can be used, even a simple scaling, while the detail layer scaling factor is provided as a parameter. The two layers are then summed to obtain the LDR image.

Generic Edge-Aware TMOs Following the bilateral filter success, many other edge-aware filters were used as TMO [Choudhury and Tumblin, 2005, Farbman et al., 2008, Gastal and Oliveira, 2011, He et al., 2013]. By interchanging the different filters, it is possible to design a generic implementation for any edge-aware filter to tone mapping. This implementation decomposes an HDR image into a multi-scale pyramid. The coarser level corresponds to the base layer while each other level represents a detail layer of different granularity (similarly to a Laplacian pyramid [Burt and Adelson, 1983]). One can formally compute the resulting LDR image $\mathbf{I}_m$ from the HDR image $\mathbf{I}_w$ using L detail layers by:

$$\mathbf{I_m} = \alpha \cdot \mathbf{I_w} \frac{TMO(\mathbf{B_l}) + \sum_{l=0}^{L} \beta_l \mathbf{D}_l}{\mathbf{L_w}}, \tag{2.14}$$

where $\mathbf{L_w}$ is the HDR luminance, $\mathbf{B_l}$ the base layer of coarser granularity (i.e. the most filtered HDR luminance). $\mathbf{D}_l$ is the detail layer of the l^{th} level ($\mathbf{D}_l = \mathbf{B_{l-1}} - \mathbf{B_l}$ with $\mathbf{B_0} = \mathbf{L_w}$). The two images $\mathbf{I_w}$ and $\mathbf{I_m}$ correspond to a 3 color channel (RGB) image while the base and detail layers to the luminance channel. The β_l are user-defined parameters that allow to smooth/sharpen the edges at different levels of the pyramid. Finally α is a gain used to tune the exposure (overall brightness) of the final image. This

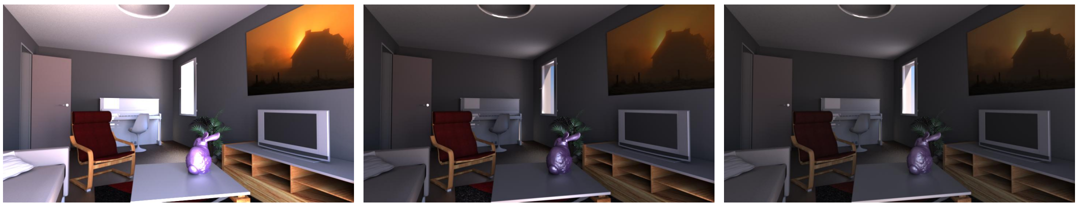


Figure 2.12: Comparison of Edge-Aware TMOs using only one detail layer. From left to right: bilateral filter [Durand and Dorsey, 2002], weighted least square filter [Farbman et al., 2008] and recursive filter [Gastal and Oliveira, 2011].

gain often adapts to the image to tone map (i.e. normalization factor). Note that using only one detail layer corresponds to using the tone mapping technique described for the bilateral filter (Figure 2.11). Figure 2.12 provides results given by different edge-aware TMOs using only one detail layer.

A Local Model of Eye Adaptation Human Visual System (HVS) operators aim at reproducing, on a display with limited capabilities, the perception of a human observer when looking at a real world scene. The goal here is not to provide the best subjective quality but rather to simulate perceptual concepts. Take for example viewing the same scene under a bright or a dim illumination. The color perception of the scene changes when the illumination changes (under dim illumination, perceived colors shifts toward the blue end of the color spectrum due to the rod's peak sensitivity) [Barlow, 1957]. This is called the Purkinje shift and is a perceptual phenomenon, the hue captured by the sensor is not changed . Consequently, when a TMO maps the HDR values to LDR ones, this shift in color is lost.

Another example is how our perception adapts to the wide range of luminance present in the real world. Consider someone walking outdoors on a bright day and entering a dim environment (e.g. a theater). At first, as he is adapted to the outdoor ambient illumination, he perceives nothing inside the theater. After a while, as he begins to adapt to the dim ambient illumination, his perception of the theater will increase until he is fully adapted. This is called the dark adaptation and the same effect also applies for light adaptation. four main operators deal with these aspects: Visual Adaptation Model [Ferwerda et al., 1996], Time-Dependent Visual Adaptation [Pattanaik et al., 2000], Local Model of Eye Adaptation [Ledda et al., 2004] and Perceptually Based Tone Mapping [Irawan et al., 2005]. Figure 2.13 describes the workflow of [Ledda et al., 2004] operator. First the HDR image is separated into CIE photopic ($\mathbf{Y}$ [Vos, 1978]) and scotopic ($\mathbf{Y'}$ [Crawford, 1949]) luminance given by:

$$\mathbf{Y} = 0.256\mathbf{R} + 0.67\mathbf{G} + 0.065\mathbf{B}, \tag{2.15}$$

$$\mathbf{Y'} = 0.702\mathbf{R} + 1.039\mathbf{G} + 0.433\mathbf{B}, \tag{2.16}$$

where $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ are the three color components of the HDR image. A bilateral filter is then applied separately on $\mathbf{Y}$ and $\mathbf{Y'}$ to determine the ambient illumination. The

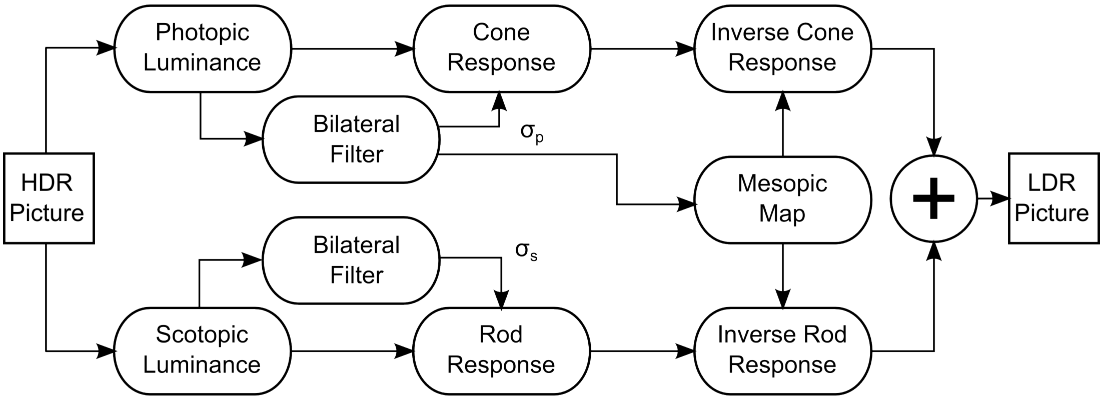


Figure 2.13: Workflow of [Ledda et al., 2004] operator.

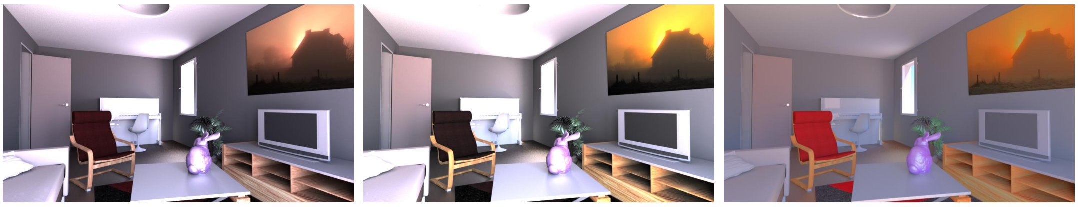


Figure 2.14: Comparison of HVS TMOs. From left to right: [Ferwerda et al., 1996], [Pattanaik et al., 2000] and [Ledda et al., 2004].

cone and rod responses are computed using both the actual luminance value and the corresponding ambient illuminations σ_p and σ_s. The photoreceptor responses are then reversed using the black and white references of the display. Finally, a mesopic map is computed to weight the contribution of each photoreceptor response in the intermediate range between photopic and scotopic.

Regarding the dark and light adaptations, the time-course of adaptation is often approximated by a reciprocal exponential function [Hood et al., 1986]. However, adaptation is not only a function of light intensity change as the pre-adaptation duration is fundamental [Mote, 1951]. That is why, given an initial and a final states, this operator interpolates the ambient illumination as a function of both time and preadaptation luminance.

Figure 2.14 depicts the results provided by 3 HVS operators.

iCAM In most of the existing TMOs, the color reproduction is usually ignored. However in 2002, [Fairchild, 2004] developed an image Color Appearance Model (iCAM) which has been improved by [Kuang et al., 2007a] in 2006.

The workflow of the iCAM06 operator ([Kuang et al., 2007a]) is described in Figure 2.15. The general idea is to use the best matching color space for every process in that workflow. The tone compression distinguishes photopic from scotopic range and is based on the Michaelis-Menten equation [Michaelis and Menten, 1913] for the photopic

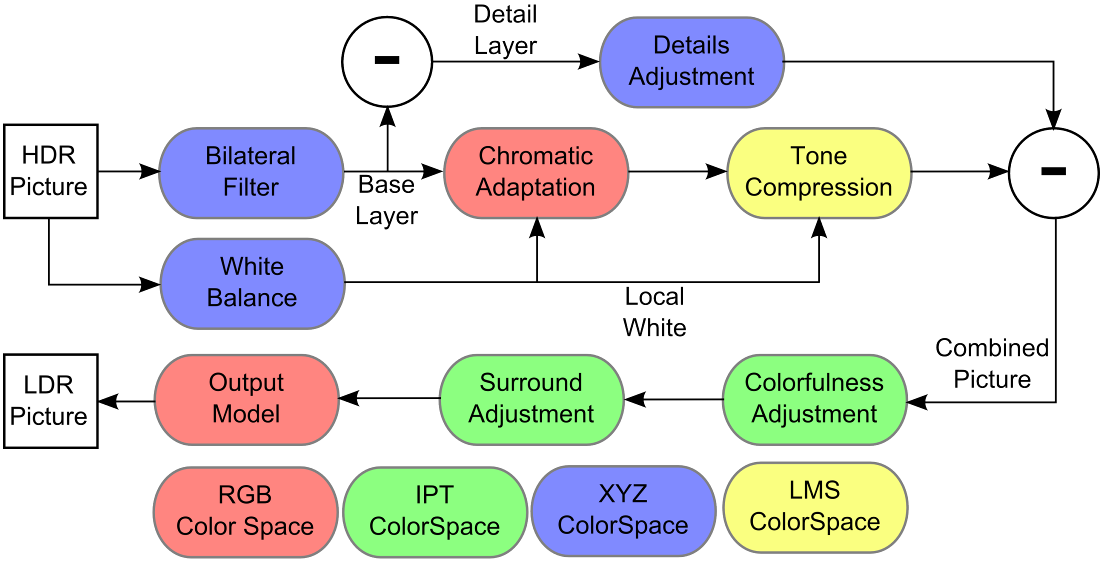


Figure 2.15: Workflow of the iCAM06 operator. The best matching color space for every process is used (blue for XYZ, red for RGB and green for IPT).

and Hunt's model for the scotopic [Hunt, 2005]. Both the chromatic adaptation and the tone compression are performed in the spectral sharpened RGB space (Hunt-Pointer-Estevez fundamentals, LMS color space). In parallel, a detail adjustment is applied to predict the Stevens effect, i.e. an increase in luminance results in an increase of local perceptual contrast. The detail and base layer are then combined before performing two adjustments: the colorfulness and the surround adjustments. The colorfulness adjustment is based on Hunt's effect, which predicts that an increase in luminance level results in an increase in perceived colorfulness. It is computed in the IPT color space. The surround adjustment states that the perceived image contrast should increase when the image surround is changed from dark to dim. It is also computed in the IPT color space. Finally, the image is reverted back to the XYZ color space before being converted to a device-dependent color space.

Evaluating TMOs Assessing the quality of TMOs has been an active field in the last decade. Despite the large number of contributions addressing this topic, no standard methodology yet exists. However, we distinguish three kinds of method to evaluate TMOs: the fidelity with reality, the fidelity with HDR reproduction and non-reference methods [Eilertsen et al., 2013].

In the fidelity with reality method, an observer compares several tone mapped images with a real scene [Yoshida, 2005, Ashikhmin and Goyal, 2006, Yoshida et al., 2006, Čadík et al., 2008]. It is useful to assess the naturalness of the reproduction of this scene. However, to achieve such an evaluation, one requires a setup with fixed illumination conditions and without motion in the scene.

For the fidelity with HDR reproduction, the reference is shown on an HDR display and an observer chooses the TMO that provides the closest reproduction of the HDR image [Ledda et al., 2005, Kuang et al., 2010]. This evaluation is easier to set up and allows to test TMOs that preserve an artist intent (color grading, contrast enhancement, etc.).

Finally the non-reference method compares several tone mapped images that are ranked by an observer [Drago et al., 2003a], [Kuang et al., 2007b], [Čadík et al., 2008], [Petit and Mantiuk, 2013]. This method is the easiest to set up and amounts to choosing the preferred result without knowing the HDR reference.

No method is free of problems and most of the work performed so far focused only on the subjective quality of the tone mapped content. Furthermore, the result of a TMO is highly correlated to the targeted display as outlined in [Mantiuk et al., 2008]. That is why different studies provide varying results although a general trend can be outlined. The lack of reliable objective quality metrics for tone mapped content increases the need for subjective evaluation. Unfortunately, the development of objective metrics is usually validated by user's study experiment, hence subjective evaluation of TMO and objective metrics are still an open field.

2.3.3 HDR Displays

TMOs allow to convert an HDR image into an LDR one displayable on legacy monitor. This process results in a loss of quality, usually in term of contrast and color reproduction. To avoid using techniques that reduce the quality of the displayed HDR images, Seetzen et al. proposed in [Seetzen et al., 2003, Seetzen et al., 2004] a design to create HDR displays by combining two light modulation devices: a modulated backlight with an LCD panel. The effective dynamic range of the image created is going to be a product of the dynamic range of both modulators. Seetzen et al. proposed two types of backlight that can be used for creating an HDR display. The first approach employed a digital light processing (DLP) projector producing a modulated backlight that later falls onto the back of an LCD panel [Seetzen et al., 2003].

In the second design, an hexagonal matrix of individually controlled white Light Emitting Diodes (LEDs) is used to produce the backlight for the LCD panel. A university spin-off company, SunnyBrook Technologies (later known as BrightSide), further developed the technology. The company built a small amount of displays, mostly for the purpose of research and for advancing the technology. Their two most well known displays were the LED based DR37-P and a projector-based SBT1.3. The DR37-P model used 1,395 controlled LEDs to provide a backlight for a 37" LCD display with a resolution of 1920x1080, with an effective contrast of 200,000 : 1 [Seetzen et al., 2004].

This technique has been used by Sim2 to build the first commercially available HDR display, the SIM2-HDR47E that can achieve 4,000 nits (nit is a unit of luminance equivalent to one cd/m^2). An evaluation of current HDR displays has been proposed in [Wanat et al., 2012].

2.3.4 LDR to HDR Conversion

The development of HDR displays allows to reproduce a wider variety of scenes with greater fidelity. However, even HDR displays have limitations (peak luminance, quantization of luminance step, etc.). By matching the physical luminance recorded in a scene to the luminance displayed on a device, disparities between the reproduction of different displays should be reduced.

However to display an LDR image or video sequence on an HDR display, its dynamic range needs to be expanded. We distinguish two types of such techniques: Expand Operator (EO) and inverse Tone Mapping Operators (iTMO). EO represents the expansion of LDR content when no information of prior tone mapping has been performed (i.e. without knowing if the content was originally HDR). On the contrary, an iTMO reconstructs an HDR image or video sequence by performing the inverse operation performed by a TMO. For example, consider an HDR video that has been tone mapped using a TMO. By using information related to the application of the TMO, the iTMO will be able to reconstruct the HDR video sequence with greater fidelity than the EO. Note that there is no consensus on those two acronyms, it is not unusual to find articles where both terms have the same meaning. We provide these two definitions because several aspects explained afterwards requires such a differentiation.

2.3.4.1 Expand Operators (EO)

An example of an EO is proposed by [Akyüz et al., 2007] where the expansion is computed by:

$$\mathbf{L_w} = L_p \left(\frac{\mathbf{L_d} - L_{d,min}}{L_{d,max} - L_{d,min}} \right)^{\gamma}, \tag{2.17}$$

where L_p is the peak luminance of the HDR display, γ is a non-linear scaling factor and $\mathbf{L_w}$ and $\mathbf{L_d}$ are the HDR luminance and LDR luma respectively. $\mathbf{L_{d,min}}$ and $\mathbf{L_{d,max}}$ are the image minimum luma and maximum luma respectively and fitting experiments provide γ values of 1, 2.2 or 0.45.

Another EO was designed by conducting two psychophysical studies to analyze the behavior of an EO across a wide range of exposure levels [Masia et al., 2009]. The authors then used the results of these experiments to develop a gamma expansion technique applied on each LDR color channel $\mathbf{C_d}$ to obtain the HDR color channel $\mathbf{C_w}$:

$$\mathbf{C_w} = \mathbf{C_d}^{\gamma}, \tag{2.18}$$

where γ is computed by:

$$\gamma = ak + b = a\frac{log(L_{d,H}) - log(L_{d,min})}{log(L_{d,max}) - log(L_{d,min})} + b, \tag{2.19}$$

where a and b were fitted by experimentation ($a = 10.44$ and $b = -6.282$) and $L_{d,H}$ is the LDR geometric mean. One of the major drawbacks of this expansion technique is that it fails to utilize the entire dynamic range of the targeted display.

EO techniques reconstruct data that were not recorded by the camera. To recover the lost information in saturated areas, EOs use expansion maps to represent the low frequency version of an image in areas of high luminance [Banterle et al., 2007].

2.3.4.2 Inverse Tone Mapping Operators

iTMOs aims at reconstructing HDR content from LDR ones but with the knowledge of the way the tone mapping was performed. For example, consider a commercial camera taking an image of a real-world scene. Most of current cameras will provide an image with 8-bits integer per color channel. By inverting the Camera Response Function (CRF in Section 2.2.4), one can retrieve the physical value (i.e. the amount of light in cd/m^2) that fell on the sensor. The resulting image can then directly be displayed on an HDR monitor.

As will be seen in Section 2.4.2, iTMOs are useful to design backward compatible HDR compression scheme. Note that the loss of information, when performing a TMO-iTMO process, is only due to the quantization step. Without this step and with enough metadata, any TMO is fully invertible without any loss.

2.4 Video Compression

Uncompressed video content (HDR or LDR) require too much data to fit the storage or broadcast requirements of current video processing pipelines. Table 2.2 summarizes the bit-rates required to store uncompressed LDR or HDR videos as well as the storage capacity of legacy storage discs and targeted digital broadcast bit-rates. The ratio between the broadcaster's bandwidth and uncompressed data bit-rate is of the order of 200. That is why video codecs (coder-decoder) compresses the size of video data to fit in the broadcast bit-rates or storage disc capacity.

We distinguish two main types of codec: lossless and lossy. Lossless encoders remove redundant information that can be derived at the decoder side in such a way that no information gets lost. However their compression ratio is quite low (usually below 10). Lossy encoders remove information based on a rate-distortion criterion, say a trade-off between the distortion and the bit-rate. Lossy encoders usually achieve high compression ratio (above 100).

This section aims at giving an overview of different codecs and techniques used to compress LDR or HDR video content. First we provide a description of the ITU-T H.265 / MPEG-H Part 2 'High Efficiency Video Codec' (HEVC) [Sullivan et al., 2012]. Then we present HDR backward compatible compression techniques.

2.4.1 HEVC

HEVC is the successor of the ITU-T H.264 / MPEG-4 Part 10 'Advanced Video Coding' (AVC) codec [Wiegand et al., 2003]. Developed by the Joint Collaborative Team on Video Coding (JCT-VC), it was released in January 2013 and is reported to double AVC compression ratio. The HEVC test Model (HM) is currently in its version 14.0.

Type	bpp	Resolution	Bit-rate
LDR	24	HD @25 fps	≈ 1.25 Gb/s
HDR	96	HD @25 fps	≈ 5 Gb/s
EXR	48	HD @25 fps	≈ 2.5 Gb/s
RGBE	32	HD @25 fps	≈ 1.65 Gb/s

Storage	Size	Bit-rate
Blu-Ray	25 Gb	/
DVD	4.7 Gb	/
Broadcast	/	≈ 8 Mb/s

Table 2.2: Left: LDR and HDR bit-rates required for uncompressed movie in High Definition (HD, resolution = 1920x1080) at 25 frame per seconds (fps). EXR designates the OpenEXR half-float format and RGBE the Radiance format. Right: Storage capacity of Blu-Ray, DVD and bit-rates considered in television broadcast.

HEVC is a block-based codec that exploits both spatial and temporal correlations between the code values of the pixels to achieve a high compression ratio. Figure 2.16 gives an overview of the different processes and their relationship inside a block-based encoder. To exploit these correlations, blocks are predicted using two types of prediction: Intra or Inter. Intra prediction relies on spatial correlation to predict the current block using blocks already decoded in the current frame. Inter prediction exploits the temporal correlation by predicting the current block using blocks from a set of previous/subsequent decoded frames. Using a rate-distortion function, the best predictor is selected among the Inter and Intra prediction. The predicted block is then subtracted from the original block to obtain the residual blocks.

To encode the residuals, the current block is first converted to the frequency domain using a frequency transform (T, i.e. discrete cosine transform or discrete sine transform). The resulting harmonic coefficients are then quantized (Q) before being fed to the entropy coder. The bitstream is composed of the encoded residuals and the prediction side-data to reproduce the same prediction on the decoder side. Note that in oder to use the same pixels value on the decoder and encoder side, the blocks are reconstructed on the encoder side to be used for the prediction process of future blocks.

As this thesis deals with temporal coherency, a detailed explanation of the Inter-prediction is given hereafter. To predict the current block, a block-based motion estimation is performed to find the motion vector pointing to the best temporal predictor. The best predictor is the block that minimizes the distortion with the current block to be encoded. An example of distortion metrics is the Sum of Absolute Differences (SAD) or the Mean Square Error (MSE). This motion estimation is performed for different block-sizes (from 64 to 4 organized in a quad-tree in the HM). The selected motion vector is the one that minimizes a rate-distortion function. The Inter-prediction results from performing a motion compensation on a reference frame using the selected motion vectors. In video compression, it is generally considered that the closest the prediction, the more efficient the compression.

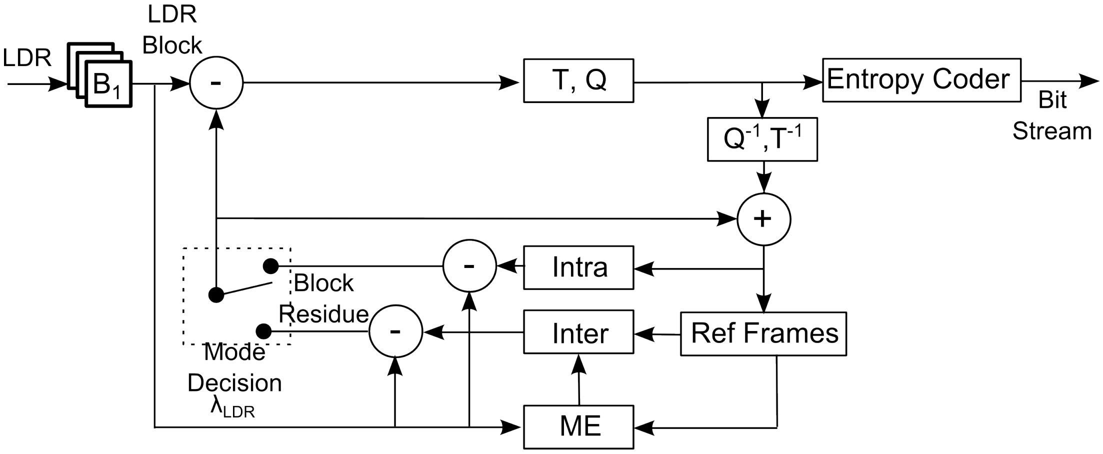


Figure 2.16: Block-based encoder workflow. ME stands for Motion Estimation, λ_{LDR} represents the rate-distortion function, T the frequency transform and Q the quantization of the harmonic coefficients.

2.4.2 HDR Backward Compatible Video Compression

If compressing LDR content is a mandatory operation for storage and transmission, it is even more important for HDR content. Indeed, current HDR file formats require two to four times the size needed by LDR content. In addition, due to the small amount of available HDR displays, backward compatibility with legacy displays is required. To that end, scalable backwards compatible techniques have been designed [Mantiuk et al., 2006a, Mai et al., 2011].

To achieve backward-compatibility, the codec first tone maps the HDR content to obtain an LDR version. The LDR sequence is encoded using a codec to obtain the LDR stream which is internally decoded to obtain the reconstructed sequence. The reconstructed sequence is then inverse tone mapped and used as a prediction for the HDR sequence. The residuals thus obtained are quantized and then encoded using another codec. The second stream obtained is called the residuals stream or the HDR stream. Figure 2.17 depicts the workflow of such a technique.

These techniques optimize the pair TMO / iTMO (inverse Tone Mapping Operator) to obtain the best reconstruction. The resulting tone map curve is computed so as to preserve the maximum level of information on a per frame basis, usually at the expense of the LDR sequence quality. Several important points have to be considered when choosing a TMO for the encoder stage. First, the subjective quality of the LDR results depends obviously on the TMO. Then, the compression efficiency of the LDR streams depends on the level of decorrelation achievable in the tone mapped video sequence. In addition, the reversibility of the TMO as well as the loss of information due to the quantization influence the amount of data that require compression in the residual stream. Finally, the TMO needs to be fully automatic as tuning the parameters is not practical when encoding any content.

Since the TMO needs to be fully automatic, the content creator has no control on

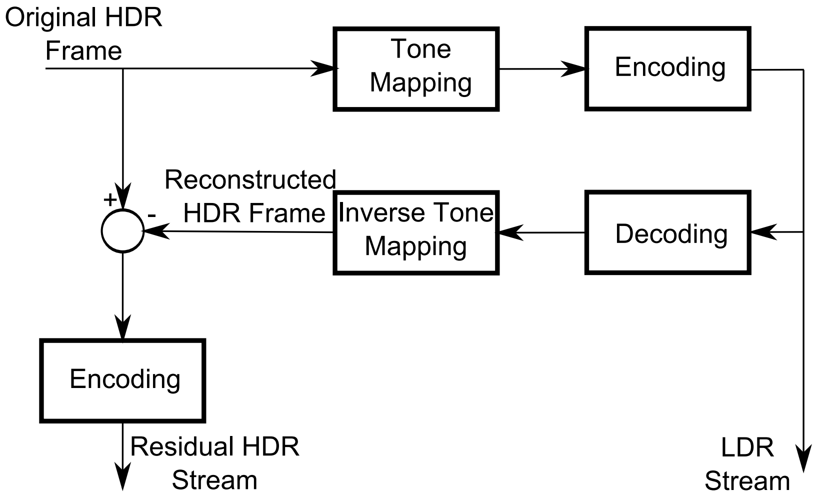


Figure 2.17: Example of the workflow of a backward compatible scalable HDR video codec.

the artistic intent of the tone mapped content. [Mantiuk et al., 2006a] proposes a technique that takes as input both an HDR version and a color-graded LDR version of the video. The TMO is computed so as to minimize the distortion with the LDR reference provided. This model allows to regain some sort of control on tone mapped content, but it is less optimal in term of compression efficiency. That is why, the new trend in HDR video compression is based on a perceptual curve, also called inverse Electro-Optic Transfer Function($EOTF^{-1}$), that transforms HDR sequence (floating-point data) into a high bit-depth representation (integer data with quantization step not visible by the human eye) [Mantiuk et al., 2004, Mantiuk et al., 2006b, Motra and Thoma, 2010, Miller et al., 2013, Kunkel et al., 2013]. The generated video is directly compressed using a single high-bit depth codec. At the decoder side, the decoded sequence is computed by the inverse of the perceptual curve to reconstruct the HDR sequence. Backward compatibility is achieved by simulcast, say providing two streams without joint compression.

2.5 Summary

In this chapter, we first introduced the fundamentals to understand how the human visual system reacts to light (Section 2.1). We then described in Section 2.2 the motivations behind HDR imaging and how both bracketing techniques and CG renderers can capture/generate more information than that can be displayed on current display devices. In Section 2.3, we presented the current technology to build HDR display along with techniques that transform HDR/LDR content to address LDR/HDR displays. Finally in Section 2.4, we presented general video compression schemes, both for HDR and LDR content, required to distribute these video to the end-users. To sum up, we described all the mandatory steps of an HDR pipeline from the content generation to the end-consumer's display (Figure 2.1).

Chapter 3

Video Tone Mapping

In the previous chapter, we introduced tone mapping and different TMOs designed to tone map HDR images (Section 2.3.2). Due to the lack of high quality HDR video content, the temporal aspect of tone mapping has been dismissed for a long time. However, with recent developments in the HDR video acquisition field [Tocci et al., 2011, Kronander et al., 2013], more and more HDR video content are now available. That is why we propose in this chapter to evaluate the temporal behavior of the different types of TMO designed for still images. We will show that naively applying a TMO to each frame of an HDR video sequence leads to different types of temporal artifact that we classify in six categories. Then we will describe video TMOs, that is to say TMOs that rely on information outside the current frame to tone map it. We will show that only three out the six types of artifact encountered can be solved by these techniques. Finally, we describe two new types of temporal artifact that are introduced by video TMOs. This chapter is structured as follows:

- Section 3.1. **Temporal Artifacts**: describes the different types of temporal artifact along with their main causes.
- Section 3.2. **Video TMOs**: presents techniques to reduce temporal artifacts by extending TMOs to more than one frame.
- Section 3.3. **Temporal Artifacts caused by Video TMOs**: details how these techniques, intended to reduce artifacts, entail other types of temporal artifact.

3.1 Temporal Artifacts

In this section, we describe temporal artifacts encountered when applying naively a TMO to each frame of an HDR video sequence. We propose to classify these artifacts into several categories:

- Section 3.1.1. **Global and Local Flickering Artifacts**,
- Section 3.1.2. **Temporal Noise**,

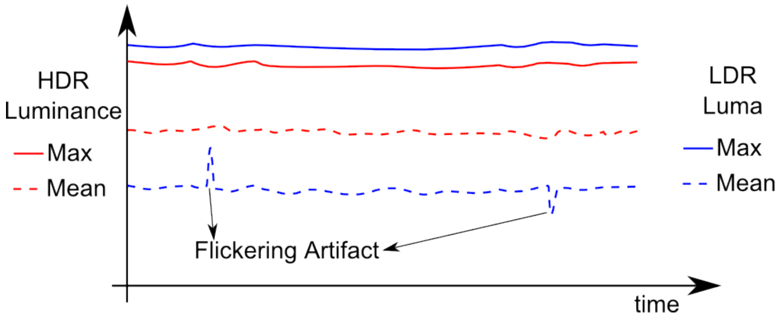


Figure 3.1: Illustration of global flickering artifacts. The overall brightness in the HDR sequence is stable over time while two abrupt variations occur in the LDR sequence. As luminance and luma have different dynamic range, they have been scaled to obtain a meaningful comparison.

- Section 3.1.3. **Temporal Brightness Incoherency**,
- Section 3.1.4. **Temporal Object Incoherency**,
- Section 3.1.5. **Temporal Hue Incoherency**.

This section provides a description of those artifacts along with some examples. Note that all the results are provided using TMOs that do not handle time-dependency, namely TMOs that rely only on statistics of the current frame.

3.1.1 Flickering Artifacts

The main type of temporal incoherency that has been investigated is **Flickering Artifacts (FA)**. A flickering artifact occurs when the mapping changes abruptly in successive frames, that is to say similar HDR luminance are mapped to different LDR luma. These artifacts appear because TMOs adapt their mapping using image statistics that tend to be unstable over time. Consequently small changes in the image may greatly alter the tone mapping. These artifacts can either be global or local, usually depending on the type of TMO used.

Global Flickering Artifacts are characterized by an abrupt change of the overall brightness in successive frames of a tone mapped video sequence. An analysis of the geometric mean (Equation 2.10) over time is usually sufficient to detect those artifacts. Recall that the geometric mean is commonly considered as an indication of the overall brightness of an image. Figure 3.1 illustrates the occurrence of two global flickering artifacts by plotting the geometric mean of both the HDR and LDR sequences. These artifacts appear because one of the TMO's parameter, that adapts to each frame, varies over time. To summarize, global flickering artifacts mostly occur with TMOs that rely on content adaptive parameters that are unstable over time. Figure 3.2 illustrates such an artifact occurring in two successive frames of a tone mapped video sequence. The

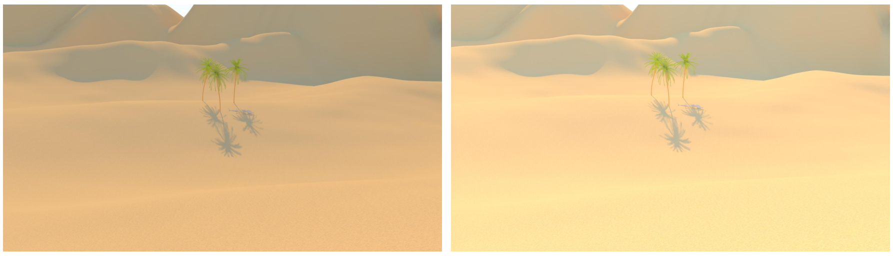


Figure 3.2: Global flickering artifacts due to the use of the 99% percentile on two successive frames of the *Desert* sequence ([Farbman et al., 2008] operator).

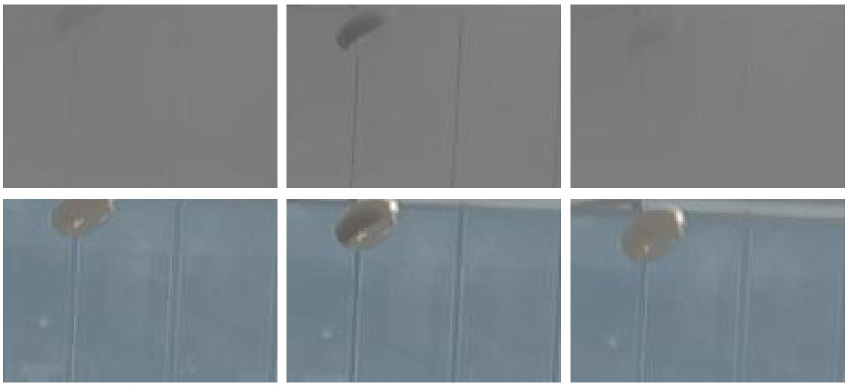


Figure 3.3: Example of local flickering artifacts when applying an edge-aware TMO to 3 consecutive frames. Top row: base layer computed using the bilateral filter. Bottom row: corresponding tone mapped result.

overall brightness has changed because the relative area of the sky in the second frame is smaller, hence reducing the chosen normalization factor (99^{th} percentile).

Local Flickering Artifacts correspond to the same phenomenon as its global counterpart but on a reduced area. They appear mostly with TMOs that map a pixel based on its neighborhood. Small changes of this neighborhood in consecutive frames may result in a different mapping. Edge-aware TMOs are particularly prone to such artifacts as they decompose an HDR image into a base layer and one or more detail layers. As each layer is tone mapped independently, a difference in the filtering in successive frames results in local flickering artifacts. The top row of Figure 3.3 represents a zoom on a portion of the computed base layer of 3 successive frames. Note how the edges are less filtered out in the middle frame compared to the other two. Applying the bilateral filter [Tomasi and Manduchi, 1998, Durand and Dorsey, 2002] operator results in a local flickering artifact in the tone mapped result (bottom row).

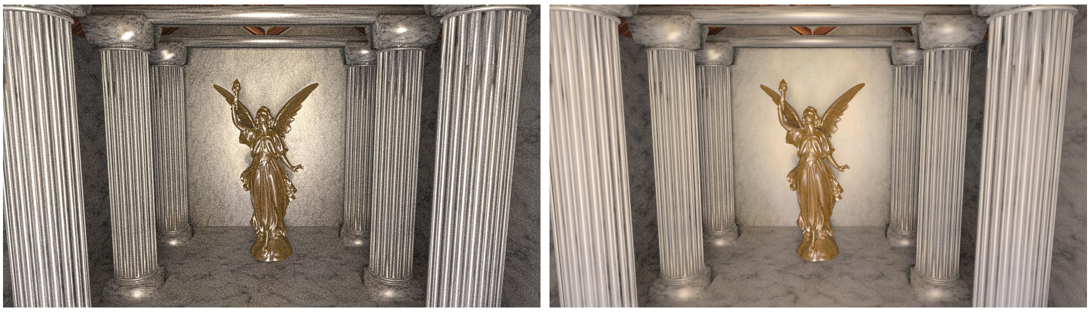


Figure 3.4: Example of temporal noise amplification due to the application of an edge-aware TMO (left, [Gastal and Oliveira, 2011]) compared to a global TMO (right, [Drago et al., 2003b]).

3.1.2 Temporal Noise

The second type of temporal artifact presented here, namely **Temporal Noise (TN)**, is a common artifact occurring in digital video sequences. Noise in digital imaging is mostly due to the camera and is particularly noticeable in low light conditions. On images, camera noise has a small impact on the subjective quality, however, for video sequences, its variation over time makes it more noticeable. It is the reason why denoising algorithms are commonly applied to video sequences to increase their subjective quality.

As most TMOs aim at reproducing minute details, they struggle to distinguish information from noise. Consequently most of current TMOs increase the noise rather than reducing it. Edge-aware TMOs are particularly prone to such artifacts as they enhance details while compressing more harshly a filtered version of the original image. An example of temporal noise enhanced by the application of an edge-aware TMOs is illustrated in Figure 3.4.

3.1.3 Temporal Brightness Incoherency

Temporal Brightness Incoherency (TBI) artifacts occur when the relative brightness between two frames of an HDR sequence is not preserved during the tone mapping. As a TMO uses for each frame all its available range, the temporal brightness relationship between frames is not preserved throughout the tone mapping operation. Consequently, frames perceived as the brightest in the HDR sequence are not necessarily the brightest in the LDR one.

For example, TBI artifacts appear when a change of illumination condition in the HDR sequence is not preserved during the tone mapping. Consequently, temporal information (i.e. the change of condition) is lost, which changes the perception of the scene (along with its artistic intent). Figure 3.5 illustrates a TBI artifact by plotting the geometric mean of both HDR and LDR sequences. Note that, although the geometric mean greatly varies in the HDR sequence, it remains stable in the LDR one. This is due to the fact that a TMO searches for the best exposure for each frame. As it

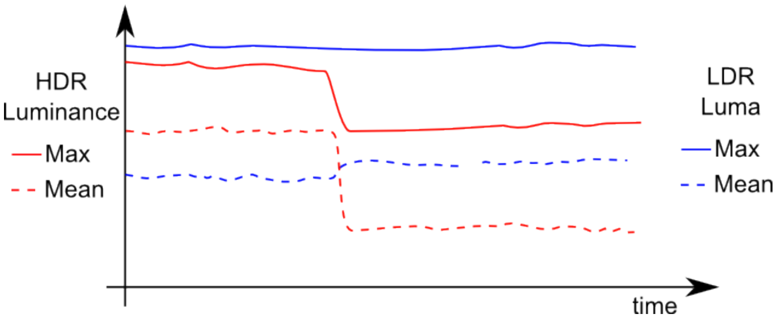


Figure 3.5: Illustration of a temporal contrast incoherency. The change of illumination condition (represented by the mean value) in the HDR sequence is not preserved in the tone mapped result.

has no information on temporally close frames, the change of illumination is simply dismissed and the best exposure is defined independently (usually in the middle of the available range). Figure 3.6 illustrates a TBI occurring in consecutive frames of a tone mapped video sequence. The top row displays the HDR luminance of these frames in false color. The transition of illumination conditions occurs when the disco ball light source is turned off. When applying a TMO, this change of illumination condition is lost (bottom row).

TBI artifacts can appear even if no change of illumination condition occurs when the tone mapping adapts to the content. When this adaptation occurs abruptly on successive frames, it gives rise to Flickering Artifacts (FA) as seen in the previous section. However when this adaptation is smoother, say over a longer range of time, the brightness relationship between the HDR and LDR sequence will be slowly disrupted. These artifacts are similar to those occurring when commercial cameras adapt their exposure during a recording [Farbman and Lischinski, 2011]. Such an artifact is shown in Figure 3.7 as the brightest HDR frame (rightmost) is the dimmest one in the LDR sequence. This second cause of TBI artifacts is also a common cause of Temporal Object Incoherency (TOI) artifact presented in the next section.

3.1.4 Temporal Object Incoherency

Temporal Object Incoherency (TOI) occurs when an object's brightness, stable in the HDR sequence, varies in the LDR one. Figure 3.8 plots the HDR and LDR values of a pixel. Note that the HDR pixel's value is constant over time while the recorded scene changes. As the TMO adapts to the current frame, the LDR pixel's value changes, resulting in a TOI artifact. Figure 3.7 illustrates visually such an artifact. When looking at the false color representation of the HDR luminance (top row, Figure 3.7) the level of brightness of the downside of the bridge appears stable over time. However, after applying a TMO (bottom row), the bridge, that appears relatively bright at the beginning of the sequence, is almost dark at the end of this sequence. The temporal coherency of the bridge in the HDR sequence has not been preserved in the LDR one.

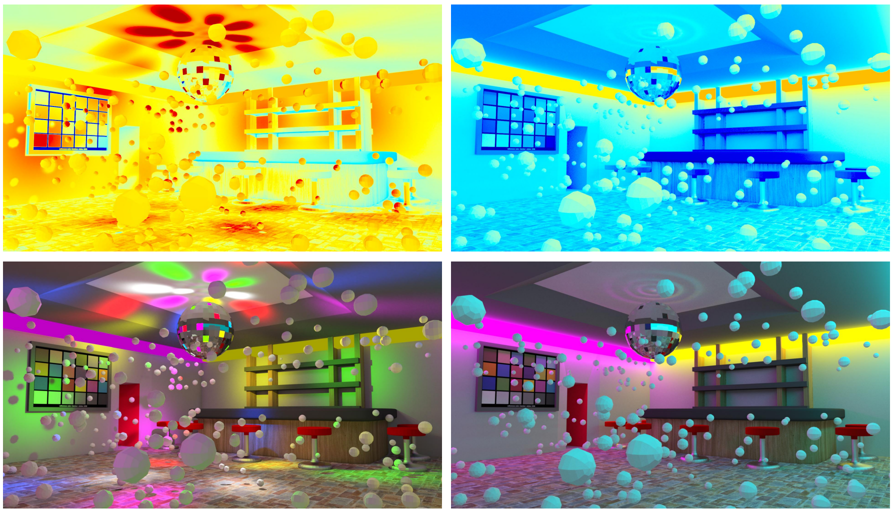


Figure 3.6: Example of a temporal contrast incoherency when a change of illumination occurs. False color luminance (top row) and tone mapped result using [Reinhard et al., 2002] operator (bottom row). Both frames appear at the same level of brightness although the false color representations indicate different levels.

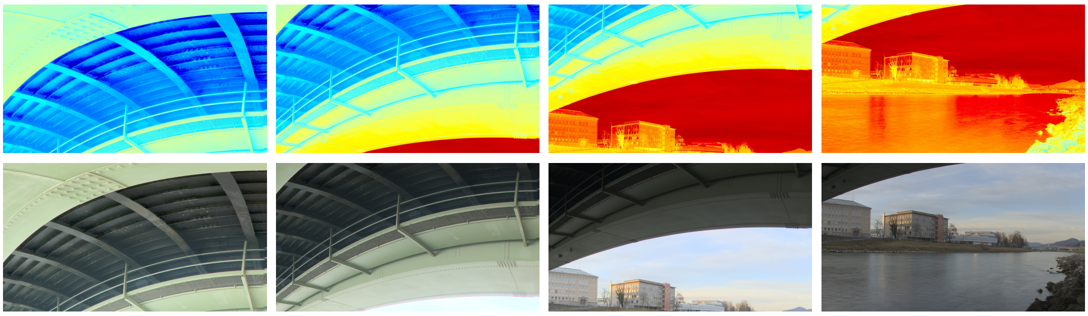


Figure 3.7: Example of TBI and TOI artifacts. False color luminance (top row) and tone mapped result using [Reinhard et al., 2002] operator (bottom row). The TBI artifact is represented by the overall brightness of each frame that is not coherent between the HDR and LDR frames. The TOI artifact is represented by the pixels' values of the downside of the bridge similar in the HDR sequence while greatly varying in the LDR one.

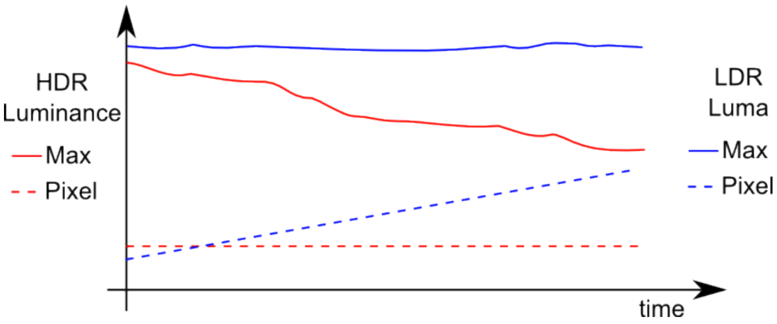


Figure 3.8: Illustration of temporal object incoherency. A pixel's value that is constant in the HDR sequence varies greatly in the LDR one.

The adaptation of a TMO to a scene is source of TBI and TOI artifacts. However, TBI artifacts are of global nature (difference of overall brightness between frames) while TOI artifacts are of local nature (difference of brightness between a reduced area over time).

3.1.5 Temporal Hue Incoherency

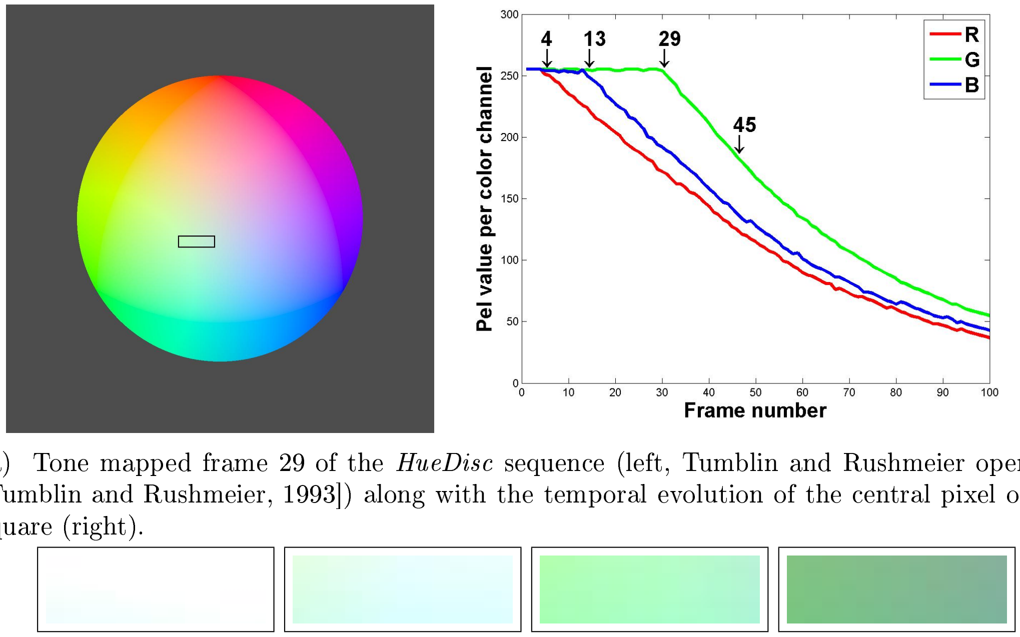
The last type of artifact presented in this section is the **Temporal Hue Incoherency (THI)**. This artifact is closely related to TBI artifact as it corresponds to the variation of the color perception of an object rather than its brightness. These artifacts occur when the balance between tristimulus values in successive frames is not temporally preserved by the tone mapping. The main reason of this imbalance is color clipping.

Color clipping corresponds to the saturation of one or more of the tone mapped color channels (e.g. red, green or blue). Color clipping is a common artifact inherent in tone mapping still images when one aims at reproducing to the best the HDR color [Xu et al., 2011, Pouli et al., 2013]. When considering color clipping as a temporal artifact, it is not the difference between the HDR and LDR reproduction that is important but rather the LDR coherency from frame to frame. Indeed, variations in the tone mapping may saturate one color channel of an area which was not in the previous frame.

To illustrate such an artifact, we generated an HDR sequence, called HueDisc (see Appendix A.3), with the following characteristics:

- a disc area of constant luminance (500 cd/m^2) with a wheel color,
- a neutral gray border area with a temporally varying luminance ranging from 50 to 5000 cd/m^2.

Figure 3.9 illustrates a temporal hue incoherency due to the clipping of one or more color channels by a TMO. Note the shift in color illustrated both on the three color channel pixel value (Figure 3.9a, right) and a zoom on a portion of the tone mapped frames (Figure 3.9b).



(a) Tone mapped frame 29 of the *HueDisc* sequence (left, Tumblin and Rushmeier operator [Tumblin and Rushmeier, 1993]) along with the temporal evolution of the central pixel of the square (right).

(b) Zoom on the area outlined by the rectangle in frames 4, 13, 29 and 45 (from left to right).

Figure 3.9: Example of temporal hue incoherency due to color clipping. Each color channel desaturates at different temporal positions.

3.2 Video TMOs

Applying a TMO naively to each frame of a video sequence leads to temporal artifacts. The aim of Video TMOs is to prevent or reduce those artifacts. Video TMOs rely on information outside the current frame to perform their mapping. So far, Video TMOs are techniques that extend or post-process TMOs designed for still images to tone map HDR video sequences. These techniques can be divided in 3 categories:

- **Section 3.2.1:** Global Temporal Filtering,
- **Section 3.2.2:** Local Temporal Filtering,
- **Section 3.2.3:** Detection and Reduction of Artifacts by Post-Processing.

For each category, we provide a description of the general technique along with different state of the art references.

3.2.1 Global Temporal Filtering

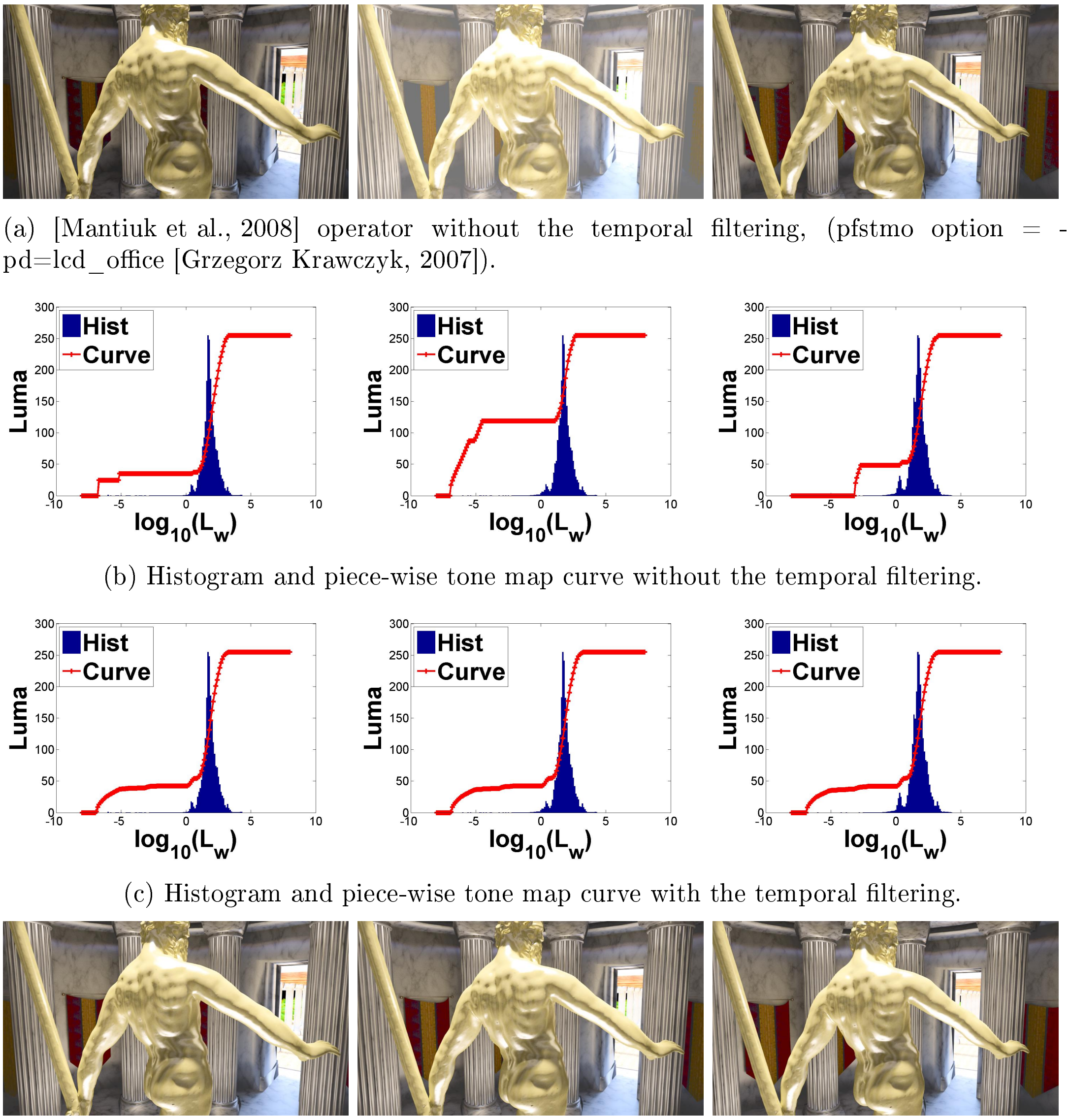
Global temporal filtering aims at reducing global flickering artifacts (Section 3.1.1). Two main approaches have been formulated so far: filtering the tone map curve or the parameters of a TMO.

Filtering the Tone Map Curve Recall that global operators compute a monotonously increasing tone map curve (Section 2.3.2). Abrupt changes of this curve in successive frames results in flickering artifacts. By applying a temporal filter to these tone map curves, global flickering artifacts can be reduced. This filtering is usually performed in two passes: one to compute a tone map curve per frame and a second to filter those curves temporally.

The Display Adaptive [Mantiuk et al., 2008] operator is able to perform such a temporal filtering on the nodes of a computed piece-wise tone map curve. The efficiency of this filtering is illustrated in Figure 3.10. The top row in Figure 3.10 provides the independently tone mapped version of three successive frames of an HDR video sequence. The second row displays the corresponding piece-wise tone map curve on top of their histogram. Note how the tone map curve of the middle frame is different from the other two, resulting in a change of overall brightness (global **FA**) in the tone mapped result. The third row shows the temporally filtered version of the piece-wise tone map curves. Finally, the bottom row provides the tone mapped frames after reducing the global flickering artifact.

Filtering TMO's Parameters Most global TMOs rely on parameters to compute their tone map curve. These parameters usually correspond to image statistics that tend to be unstable over time (e.g. the 99^{th} percentile, the geometric mean, etc.).

For example, the Photographic Tone Reproduction [Reinhard et al., 2002] operator relies on the geometric mean (also called key value) to scale an HDR image to the best exposure. One temporal extension of this operator filters this key value along a set



(a) [Mantiuk et al., 2008] operator without the temporal filtering, (pfstmo option = -d pd=lcd_office [Grzegorz Krawczyk, 2007]).

(b) Histogram and piece-wise tone map curve without the temporal filtering.

(c) Histogram and piece-wise tone map curve with the temporal filtering.

(d) [Mantiuk et al., 2008] operator with the temporal filtering.

Figure 3.10: Reduction of global flickering artifacts by temporally filtering the tone mapping curve.

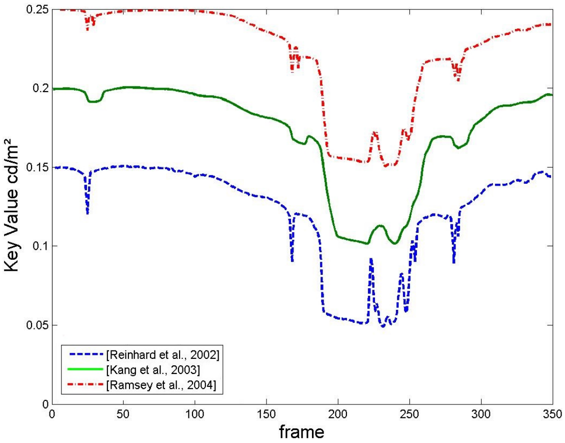


Figure 3.11: Evolution of the frame key value computed for every frame of a video sequence. An offset is added to avoid an overlap between the curves, the smoothing effect of both techniques ([Kang et al., 2003] and [Ramsey et al., 2004]) are compared to [Reinhard et al., 2002] operator.

of previous frames [Kang et al., 2003]. As a consequence, this method smooths abrupt variations of the frame key value throughout the video sequence. This technique is capable of reducing flickering for sequences with slow illumination variations. However, for high variations, it fails because it considers a fixed number of previous frames. That is why, [Ramsey et al., 2004] proposed a method that adapts dynamically this number. The adaptation process depends on the variation of the current frame key value and that of the previous frame. Moreover, the adaptation discards outliers using a min/max threshold. This solution performs better than [Kang et al., 2003] and for a wider range of video sequences. The computed key values for both of these techniques and the original algorithm are plotted in Figure 3.11. The green curve ([Kang et al., 2003]) smooths every peak but also propagates them to successive computed key values. The red curve however ([Ramsey et al., 2004]) reduces the abrupt changes of the key value without propagating it.

Another temporal extension of the Photographic Tone Reproduction operator has been proposed in [Kiser et al., 2012]. It first modifies the scaling of the HDR frame (Equation 2.9) to perform the temporal filtering:

$$\mathbf{L_s} = \frac{\epsilon \cdot 2^{2(B-A)/(A+B)}}{k}\mathbf{L_w} = \frac{a}{k}\mathbf{L_w} \tag{3.1}$$

where $A = L_{max} - k$ and $B = k - L_{min}$ with L_{max} and L_{min} the maximum and minimum value of $\mathbf{L_w}$. k corresponds to the key value (a.k.a. the geometric mean). The temporal filtering consists of a leaky integrator applied to the three variables (a, A and B) used to tune the tone mapping:

$$v_t = (1 - \alpha_v)v_{(t-1)} + \alpha_v v_t \tag{3.2}$$

where v_t represents any of the three variables at time t and α_v is a time constant giving the strength of the temporal filtering. The tone mapping equation remains the same (Equation 2.11 in Section 2.3.2).

Many other TMOs filter their parameters temporally including [Pattanaik et al., 2000], [Durand and Dorsey, 2000], [Ledda et al., 2004], [Irawan et al., 2005] and [Van Hateren, 2006]. Most of them either aim at simulating the temporal adaptation of the HVS or at reducing global flickering artifacts.

3.2.2 Local Temporal Filtering

Global temporal filtering cannot apply to local TMOs as such operators rely on a spatially varying mapping function. As outlined in Section 3.1.1, local changes in a spatial neighborhood cause local flickering artifacts. To prevent these local variations of the mapping along successive frames, video tone mapping operators can rely on a pixel-wise temporal filtering.

For example, the Gradient Domain Compression operator [Fattal et al., 2002] has been extended by [Lee and Kim, 2007] to cope with videos. Recall that this operator computes the LDR result by finding the output image whose gradient field is the closest to a modified gradient field. [Lee and Kim, 2007] propose to add a regularization term which includes a temporal coherency relying on a motion estimation:

$$\sum_{x,y} \|\Delta \mathbf{L_d}(x,y,t) - \mathbf{G}(x,y)\|^2 + \lambda \sum_{x,y} \|\mathbf{L_d}(x,y,t) - \mathbf{L_d}(x+\delta x, y+\delta y, t-1)\|^2 \quad (3.3)$$

where $\mathbf{L_d}$ is the output LDR luma at the precedent or current frame ($t-1$ or t) and $\mathbf{G}$ is the modified gradient field. The pairs (x, y) and $(\delta x, \delta y)$ represent respectively the pixel location of a considered pixel and its associated motion vectors. The parameter λ balances the distortion to the modified gradient field and to the previous tone mapped frame.

Another operator [Ledda et al., 2004] performs a pixel-wise temporal filtering. However the goal with this operator is to simulate the temporal adaptation of the human eye on a per-pixel basis. Besides increasing the temporal coherency, pixel-wise temporal filtering also has denoising properties. Indeed, many denoising operators rely on temporal filtering to reduce noise [Brailean et al., 1995]. Performing such a filtering during the tone mapping allows to keep the noise level relatively low.

3.2.3 Detection and Reduction of Artifacts by Post-Processing

The techniques presented so far in this section focus on preventing temporal artifacts (mostly flickering) when tone mapping video sequences. These a priori approaches consist in either preprocessing parameters or modifying the TMO to include a temporal filtering step. Another trend analyzes a posteriori the output of a TMO to detect and reduce temporal artifacts.

One of these techniques ([Guthier et al., 2011]) first detects flickering artifacts by assuming that they are of global nature. An artifact is detected if the overall brightness

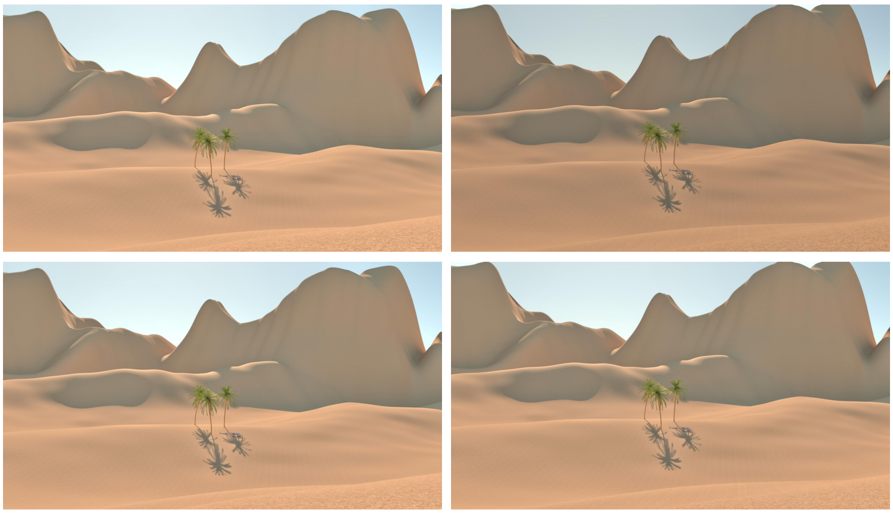


Figure 3.12: Results of [Farbman et al., 2008] operator without (top row) and with (bottom row) [Guthier et al., 2011] post-processing.

difference (a.k.a. the geometric mean) between successive frames of a video sequence is greater than a brightness threshold (defined using either Weber's law [Ferwerda, 2001] or Steven's power law [Stevens and Stevens, 1963]). As soon as an artifact is located, it is reduced using an iterative brightness adjustment until reaching the brightness threshold. Note that this technique performs an iterative brightness adjustment using floating point values to avoid loss of signal due to clipping and quantization. Consequently, the TMO's implementation needs to embed the flickering reduction operation before the quantization step. This technique relies only on the output of the TMO and hence can be applied to any TMO. Figure 3.12 illustrates the reduction of a global flickering artifact when applying the post-processing proposed in [Guthier et al., 2011].

3.3 Temporal Artifacts caused by Video TMOs

In the previous section, we presented solutions to reduce temporal artifacts when performing video tone mapping. These techniques target mostly the flickering artifact as it is considered one of the most disturbing one. However, these techniques can generate new temporal artifacts: **Temporal Contrast Adaptation (TCA)** and **Ghosting Artifacts (GA)**.

3.3.1 Temporal Contrast Adaptation

To reduce global flickering artifacts, many TMOs rely on global temporal filtering. Depending on the used TMO, the filter is either applied to the computed tone map curve [Mantiuk et al., 2008] or to the parameter that adapts the mapping to the image [Ramsey et al., 2004, Kiser et al., 2012]. However, when an actual change of illumination occurs, as presented in Section 3.1.3, it also undergoes temporal filtering. Conse-

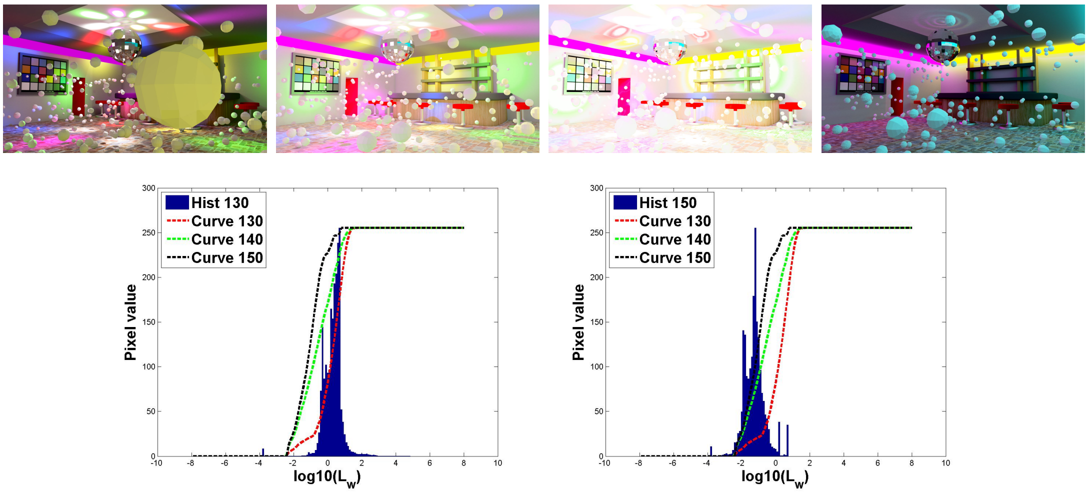


Figure 3.13: Example of temporal filtering of the tone map curve when a change of illumination occurs. Top row: tone mapped result (frame number 130, 140, 149 and 150) using [Mantiuk et al., 2008] operator with the temporal filtering active (pfsTMO implementation [Grzegorz Krawczyk, 2007]). Bottom row: histograms of frames 130 (left) and 150 (right) along with the corresponding tone map curve for frames 130, 140 and 150.

quently, the resulting mapping does not correspond to any of the conditions but rather to a transition state. We refer to this artifact as **Temporal Contrast Adaptation (TCA)**.

Figure 3.13 shows how the temporal filtering of the tone map curve causes this artifact during a change of illumination. Note how the tone map curve, plotted on top of the histograms, shifts from the first illumination condition (frame 130) toward the second state of illumination (frame 150, see Figure 3.6 for the false color luminance). As the tone map curve has anticipated this change of illumination, frames in the illumination transition are tone mapped incoherently.

These artifacts also occur when performing a post-processing to detect and reduce artifacts as presented in Section 3.2.3. Indeed, [Guthier et al., 2011] technique only relies on the LDR results to detect and reduce artifacts. If one has no information related to the HDR video then a change of illumination suppressed by a TMO cannot be anticipated nor predicted.

3.3.2 Ghosting Artifacts

Similarly to global temporal filtering, local temporal filtering generates undesired temporal artifacts. Indeed, pixel-wise temporal filtering relies on a motion field estimation which is not robust to change of illumination conditions and object occlusions. When the motion model fails, the temporal filtering is computed along invalid motion trajectories which results in **Ghosting Artifacts (GA)**.

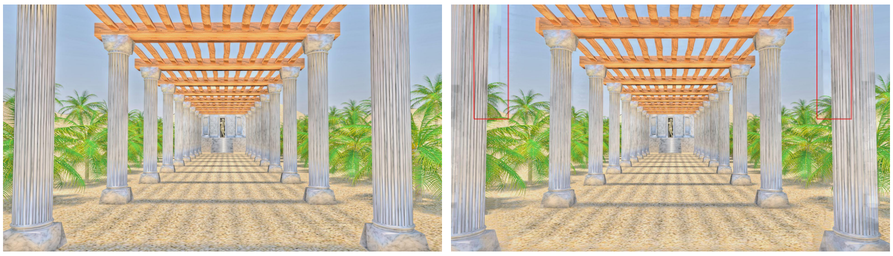


Figure 3.14: Ghosting Artifacts appearing on two successive frames. It is most noticeable around the two forefront columns (red squares).

Figure 3.14 illustrates a ghosting artifact in two successive frames resulting from the application of [Lee and Kim, 2007] operator. This artifact proves that pixel-wise temporal filtering is only efficient for accurate motion vectors. When a motion vector associates pixels without temporal relationship, ghosting artifacts will occur. Those "incoherent" motion vectors should be accounted for to prevent ghosting artifacts as these have been shown to be the most disturbing type of artifact [Eilertsen et al., 2013].

3.4 Summary

In this chapter, we proposed a classification in six types of temporal artifact that occur when applying separately a TMO to each frame of an HDR video sequence. Only two type of artifact had been outlined and addressed in state of the art work (global and local flickering). Then, we described Video TMOs that extend TMOs design for still images to the temporal domain. We have shown that these techniques could generate two new types of temporal artifact.

Table 3.1 gives an overview of the temporal artifacts presented in this chapter. All the solutions provided to reduce temporal artifacts focus on flickering. Furthermore, techniques that rely on temporal filtering can generate other types of artifact. The analysis of problems occurring in video tone mapping, along with the solutions presented in this chapter, has been published in:

- **R. Boitard**, D. Thoreau, K. Bouatouch, and R. Cozot, "Temporal Coherency in Video Tone Mapping , a Survey," in *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, 2013.

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Survey of temporal brightness artifacts in video tone mapping," in *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014.

Furthermore, recent subjective evaluations have been performed on video tone mapping [Petit and Mantiuk, 2013, Melo et al., 2013, Eilertsen et al., 2013]. They all draw the same conclusion that no solution yet exists and that further research in video tone mapping is required.

Temporal Artifact	**Cause**	**Solution**
FA (Global Flicker)	Temporal unstability of Parameters	Global temporal filtering
FA (Local Flicker)	Different spatial filtering in successive frames	Pixel-wise temporal filtering
TN (Noise)	Camera noise	Spatial and/or temporal filtering (pixel-wise)
TBI (Brightness)	Change of illumination Adaptation of the TMO	
TOI (Object)	Adaptation of the TMO	
THI (Hue)	Saturation of color channel (due to clipping)	
TCA (Contrast)	Due to global temporal filtering	
GA (Ghosting)	Due to pixel-wise temporal filtering	

Table 3.1: Summary of temporal artifacts and their causes.

Chapter 4

Global Brightness Coherency

In the previous chapter, we described why applying naively a TMO to an HDR video sequence leads to temporal artifacts. As summarized in Table 3.1, technical solutions to reduce temporal artifacts in video tone mapping focus mostly on flickering artifacts. That is why, we propose, in this chapter, a technique to address the reduction of **Temporal Brightness Incoherency (TBI)** artifacts. Recall that TBI artifacts occurs when the temporal brightness relationship between frames of an HDR video sequence is not preserved throughout the tone mapping operation. This chapter is organized as follows:

- Section 4.1. **Temporal Contrast and Number of Tonal Levels**: outlines the differences between the HDR and LDR spatial/temporal contrast and sampling along with a description of LDR techniques to optimize the spatio-temporal sampling.

- Section 4.2. **Brightness Coherency Post-Processing**: describes our technique to reduce TBI artifacts when performing video tone mapping.

- Section 4.3. **Results**: presents some results regarding the preservation of brightness and object coherencies.

- Section 4.4. **Limitations of the Brightness Coherency Post-Processing**: describes when the BC post-processing fails to preserve temporal brightness coherency.

4.1 Temporal Contrast and Number of Tonal Levels

We define the temporal contrast as the ratio between the highest and the lowest luminance value in a video sequence. However this metric is not very meaningful if the number of different tonal values between the highest and lowest luminance values is not accounted for.

For HDR imaging, the contrast and the associated number of tonal level is limited by the used representation format (RGBE, half-float, etc.). The spatial contrast of

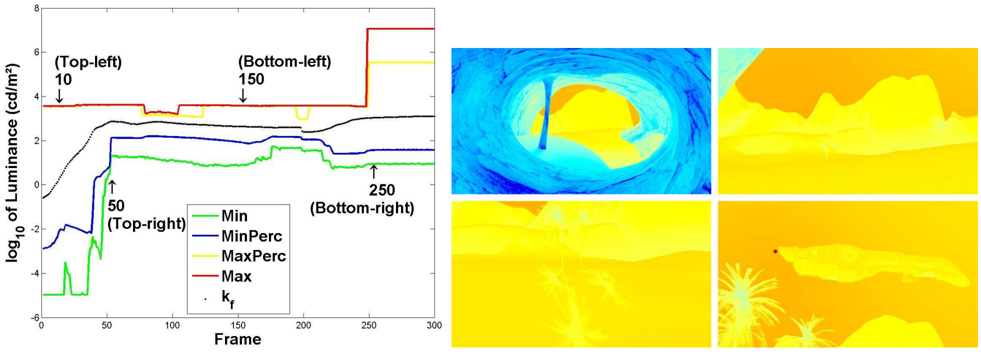


Figure 4.1: Left: spatial and temporal contrasts in the *Desert* sequence. MinPerc and MaxPerc stand for minimum and maximum percentile, set at 1 % and 99 % respectively. k_f is the key value computed for each frame. Right: frames 10, 50, 150 and 250 of the corresponding sequence in false color representation.

each frame is the ratio between the maximum and minimum luminance values of this frame, while the temporal contrast of a video is the ratio between the maximum and minimum luminance values of all the frames of the video. Obviously the temporal contrast is always greater or equal to the spatial contrast. Figure 4.1 illustrates the difference between temporal and spatial contrasts in a video sequence. In this figure, the maximum spatial contrast is 6.4 orders of magnitude while the temporal one is 8.4 orders of magnitude. Note that we computed those contrasts using the 1 % and 99 % percentile (to remove outliers values).

For LDR imaging, the spatial and temporal contrasts for a given display is fixed and depends on the white and black points of the display (Section 2.3.1). The number of tonal levels depends on the bit-depth n, which corresponds to the size of a quantization bin q (the width of each tonal value):

$$q = \frac{L_{d,max} - L_{d,min}}{2^n - 1}, \tag{4.1}$$

where $L_{d,max}$ and $L_{d,min}$ correspond to the white and black points of the display respectively. When capturing an image with a wide range of luminance, the contrast limitation results in over-exposed (too much light hitting the sensor) and/or under-exposed (not enough light) pixels. The amount of over/under exposed pixels depends on the exposure (**eV** setting) of a camera. For still images, this parameter is set to maximize the number of well-exposed pixels, that is to say to have the finest granularity for the biggest amount of pixels. Additionally, the exposure is also often used to produce artistic effects such as all white or all black backgrounds. In other words, setting the exposure can be seen as a form of tone mapping of a real scene to get an LDR image.

When capturing a video, setting the exposure is more difficult as it is hard to foretell how the temporal contrast will evolve. Commercial cameras possess an automatic mode, which adapt the exposure to the scene before filtering it temporally (to prevent flickering). This technique is similar to the global temporal filtering presented in Section 3.2.1. As the exposure adapts to the scene, this type of acquisition is particularly

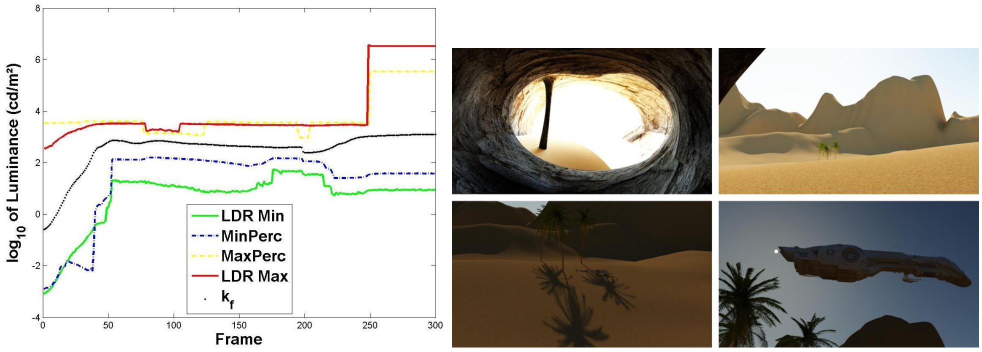


Figure 4.2: Left: example of camera's exposure adaptation, the results were obtained using the camera TMO [Petit and Mantiuk, 2013]. The LDR Min and Max value have been mapped back to HDR value to illustrate the dynamic range that the camera TMO covers. k_f correspond to the frame's key value computed using Equation 4.2. Right: tone mapped frame 10, 50, 150 and 250 of the corresponding sequence.

prone to TBI artifacts with the only difference that the mapping is performed between the real scene and the LDR picture. To better understand this artifact, we simulated a camera trying to capture the *Desert* sequence (Appendix A.3) using the CameraTMO [Petit and Mantiuk, 2013]. We plotted the used dynamic range in Figure 4.2 along with several frames of the tone mapped result. We observe that the variations of the computed key values are high and that the perception of the scene is not at all consistent over the video (going from over exposed in frame 10 to really dark in frame 150). Temporal brightness incoherency usually appears only with user-generated content that use per default the automatic mode.

For professional content, the camera's exposure is fixed for a specific shot, which ensures temporal brightness coherency. If the temporal contrast is too high, insertion of "cuts" provides a transition between two exposure settings. Furthermore, over/under exposed pixels undergo color grading at the post-production stage. This process can be compared to manually tone mapping each frame of an HDR sequence to achieve the best chosen representation of the video.

Although temporal exposure adaptation or even fixing the exposure settings per scene can lead to satisfying results, they are not well suited to extreme changes in illumination as the capture is still a limitation. However, capturing an HDR video with a high temporal contrast is not an issue thanks to bracketing techniques but its tone mapping is still an issue since the targeted number of tonal levels and contrast are too limited to well depict it. Existing TMOs mimic the exposure adaptation of LDR capture device, hence resulting in temporal brightness incoherency. As a TMO uses all the available range for each frame, the HDR brightness coherency is not preserved throughout the tone mapping operation. Consequently, frames perceived as the brightest in the HDR sequence are not necessarily the brightest in the LDR one. To solve this issue, we propose a technique, called Brightness Coherency (BC), that aims at preserving the temporal brightness coherency while still benefiting from the increased detail reproduc-

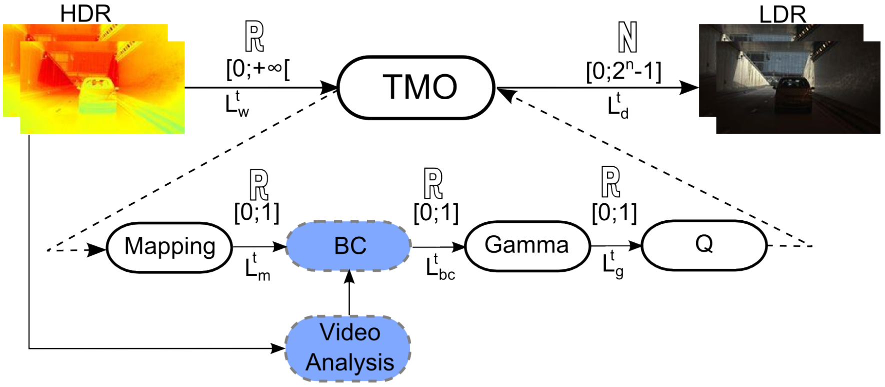


Figure 4.3: Modified workflow to include the Brightness Coherency (BC) algorithm in any TMO. The different notation of the luminance is indicated along with their range and type of data.

tion capability of TMOs. Our goal is to provide results closer to those that could be achieved when using fixed exposure but with the enhanced reproduction capabilities of any TMO. We present this technique in the next section.

4.2 Brightness Coherency Post-Processing

The intended goal of the Brightness Coherency (BC) post-processing is to preserve the temporal brightness coherency of the HDR sequence in the tone mapped one. To achieve this, our technique adds two processing steps to the usual TMO workflow: a video analysis and a post-processing operation. The video analysis component computes characteristics of a video to determine an anchor frame, which serves as a frame of reference brightness for each other frame. The post-processing operation modifies each resulting LDR frame to preserve the brightness coherency between the anchor frame and each other frame. Figure 4.3 illustrates the modified workflow implementing the BC post-processing in any TMO.

The video analysis consists in finding an anchor frame that will be considered as the best exposure. To achieve this, we compute the geometric mean k_f^t of each frame at time index t by:

$$k_f^t(\mathbf{L^t}) = \exp\left(\frac{1}{n_p}\sum_{x=1}^{n_p}\log(d+\mathbf{L^t}(x))\right) , \tag{4.2}$$

where $\mathbf{L^t}$ is the luminance of the t^{th} frame, d a small value to avoid singularity and n_p the number of pixels. Choosing the anchor frame is analogous to choosing the exposure in LDR photography. To simplify the choice of the anchor frame, we propose a user-defined parameter χ that selects the anchor frame depending on its geometric mean. Three choices are predefined, the frame with the highest ("Max"), median ("Median") or lowest ("Min") geometric mean. The influence of χ is detailed in Section 4.3.1.

We denote $k_v(\mathbf{L_w})$ the geometric mean of the HDR anchor frame and $k_v(\mathbf{L_m})$ the geometric mean of the corresponding tone mapped frame. Note that the video analysis is performed prior to the tone mapping and we need to compute the key value of the anchor frame in the LDR domain $k_v(\mathbf{L_m})$ during the video analysis (which is a preprocessing step).

Now that both k_v (the anchor's geometric means) are known, we perform the BC post-processing on each frame of the tone mapped sequence. As we want to preserve the relative difference of brightness between each frame and the anchor, the post-processed LDR luminance should satisfy the following equation:

$$\frac{k_f^t(\mathbf{L_w^t})}{k_v(\mathbf{L_w})} = \frac{k_f^t(\mathbf{L_{bc}^t})}{k_v(\mathbf{L_{bc}})} \tag{4.3}$$

To make this ratio hold, we post-process the tone mapped luminance $\mathbf{L_m^t}$ to obtain the post-processed tone mapped luminance $\mathbf{L_{bc}^t}$:

$$\mathbf{L_{bc}^t} = \left(\zeta + (1-\zeta)\frac{k_f^t(\mathbf{L_w^t}) \cdot k_v(\mathbf{L_m})}{k_v(\mathbf{L_w}) \cdot k_f^t(\mathbf{L_m^t})}\right)\mathbf{L_m^t} = s^t \cdot \mathbf{L_m^t} \tag{4.4}$$

where s^t represents the scale ratio of the t^{th} frame and ζ is a user-defined parameter to avoid low scale ratio (default value 0.1). The influence of ζ is detailed in Section 4.3.1. The remaining of the tone mapping is performed as before (gamma encoding, scaling and quantization to integers). Note that although the $\mathbf{L_m^t}$ values are supposed to be in the range [0;1], values lying outside this range are not clipped until the gamma encoding. The consequence is that, for scale ratio below 1, pixels that would have been clipped during the tone mapping still contain useful information. Similarly, well-exposed pixels may be clipped when the scale ratio is above 1.

4.3 Results

The BC technique aims at preserving the overall temporal brightness coherency of the HDR sequence in the LDR one. This attribute allows our technique to handle artistic effects such as fading that other TMOs cannot cope with. We experimented with different HDR sequences (*Desert*, *UnderBridgeHigh*, *MtStMichelWhite* and *MtStMichelBlack* detailed in Appendix A) to assess the efficiency of our method in terms of:

- reduction of **Temporal Brightness Incoherency (TBI)** artifacts,
- reduction of **Temporal Object Incoherency (TOI)** artifacts,
- preservation of fading effects.

We will discuss each of these in detail in the following sections.

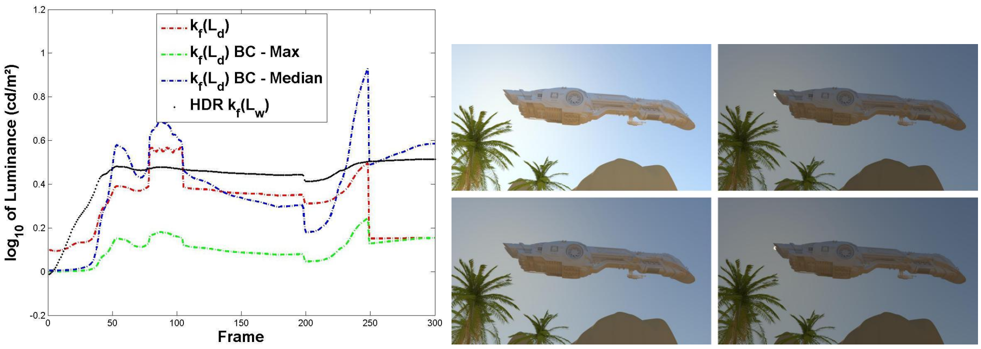


Figure 4.4: Left: geometric means of the HDR and tone mapped video sequence. Right, upper row: global flickering artifact occurring between frame 248 and 249 due to the appearance of the sun. Right, lower row: global flickering artifact reduced using the BC post-processing.

4.3.1 Temporal Brightness Coherency

Preserving the relative levels of brightness throughout a video is the main goal of our technique. We have applied the [Drago et al., 2003b] operator to tone map the *Desert* sequence with and without the BC post-processing. Figure 4.4 plots the geometric mean of the HDR and tone mapped video sequences. Improved temporal brightness coherency is achieved especially between the beginning (which should be very dim) and the end (which should be bright) of the sequence. Note that a global flickering artifact occurs between frame 248 and 249, which is partly reduced thanks to the BC post-processing.

Figure 4.5 illustrates the influence of χ by providing the tone mapped results of the four frames presented in Figure 4.1 and 4.2. Recall that χ is a user-defined parameter to select the anchor frame depending on its geometric mean. The HDR geometric mean k_f^t of those frames, from left to right, is: 0.9205, 630, 440 and 1070 cd/m^2. When the BC post-processing is not used, all frames appear of similar overall brightness. Furthermore, the dimmest frame (right frame) is actually the brightest one in the HDR sequence. When the BC is used with χ='Max', the tone mapped brightest HDR frame is not changed while all the other frames are dimmer. Coherency is improved, however only few details are visible in the leftmost frame. When using the median geometric mean as anchor (χ='Median'), more details appear in the leftmost frame but some pixels are over-exposed in the 2^{nd} and last frame.

The other parameter of the BC post-processing is ζ, which prevents low scale ratio. Its effect are illustrated in Figure 4.6, where its value ranges from 0.01 to 1. Too low a ζ will result in low details reproduction for dim HDR frames while a too high one will lessen the effect of the BC post-processing. Effectively, this parameter controls the trade-off between reproduction of spatial details and temporal brightness coherency.

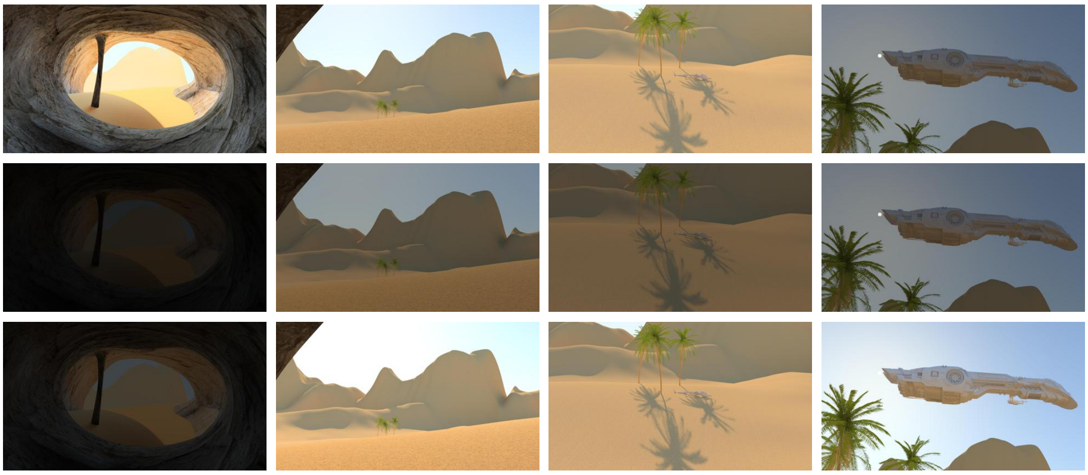


Figure 4.5: Example of BC post-processing when tone mapping the *Desert* sequence. Tone mapped results using [Drago et al., 2003b] operator without (top-row) and with the BC post-processing ($\zeta = 0.05$, 2^{nd} row χ='Max' and bottom-row χ='Median'). By using the BC post-processing, the overall brightness coherency is preserved.

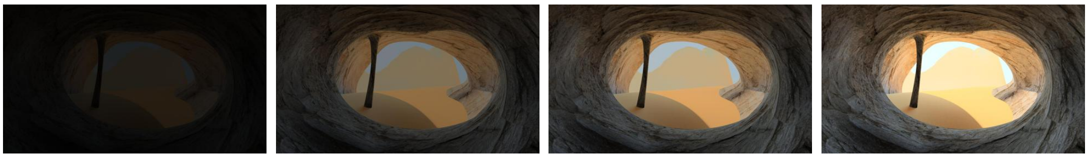


Figure 4.6: Influence of ζ from left to right: 0.01, 0.2, 0.5 and 1.

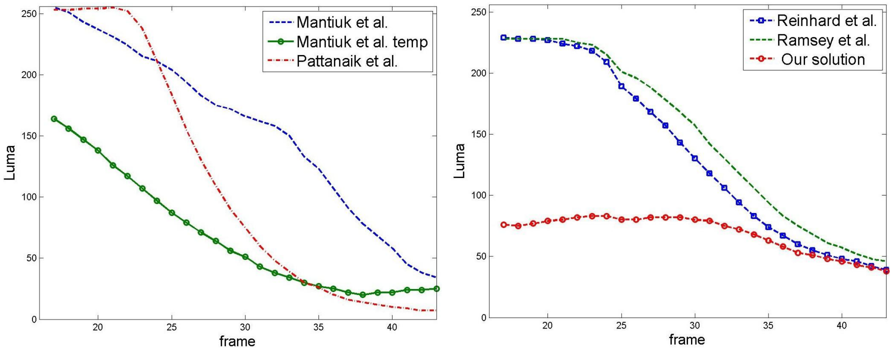


Figure 4.7: Evolution of a pixel throughout the LDR video sequence for different TMOs. Left: [Pattanaik et al., 2000] and [Mantiuk et al., 2008] with and without temporal filtering. Right: [Reinhard et al., 2002], [Ramsey et al., 2004] without and with the BC post-processing.

4.3.2 Temporal Object Coherency

Temporal Object Incoherency (TOI) occurs when an object's brightness, which was stable in the HDR sequence, varies in the LDR one. If we consider a small area visible in successive frames with constant HDR luminance value, the luma of the tone mapped area should have very low variations for the perception of the object to be coherent.

To assess the performance of the BC post-processing, we tone mapped the *UnderBridgeHigh* sequence using different TMOs. We then considered a pixel with the same HDR luminance value along 28 frames. Figure 4.7 shows the evolution of this pixel for all the TMOs tested. The pixel's luma variations are high with all the used TMOs while our technique keeps these variations very low. Figure 4.8 displays the area surrounding the pixel studied in Figure 4.7. The visual perception of the underside of the bridge is completely different between frame 25 (upper row) and 35 (lower row). Only our solution preserves the overall appearance of the bridge over time.

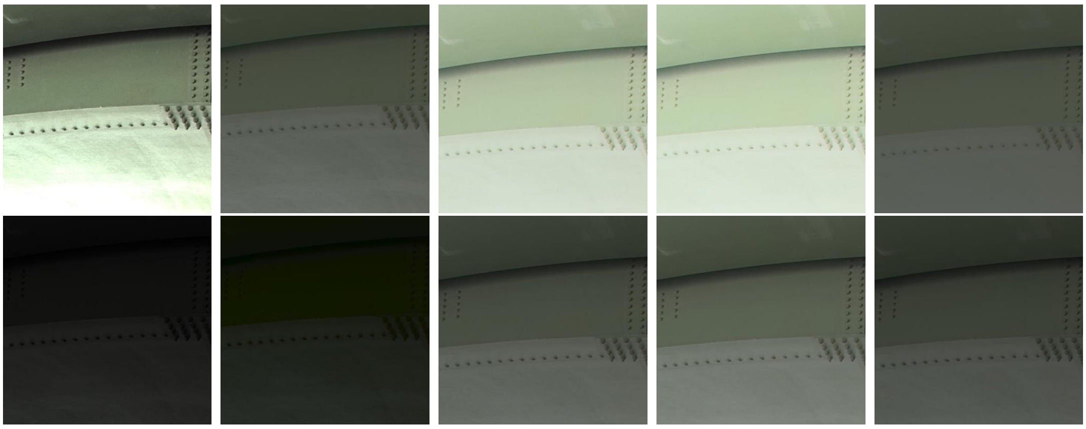
4.3.3 Video Fade

A video fade occurs when a shot gradually fades to (or from) a single color, usually black or white. In LDR video processing, a fade to black is computed using an alpha-blending:

$$\mathbf{R^t} = \alpha_t \mathbf{O^t} + (1 - \alpha_t)\mathbf{F}, \tag{4.5}$$

where $\mathbf{O^t}$ is the t^{th} original frame of the sequence, $\mathbf{F}$ a black image and $\mathbf{R^t}$ the t^{th} resulting fade to black frame. α_t ranges from 1 to 0 and controls the speed of the fading. We consider the speed of the fading as the number of intermediate frames n_t needed to satisfy $\mathbf{R^t} = \mathbf{F}$.

Although video fade effects are well known in LDR video processing, HDR video fade has not been addressed yet. This is due to the fact that HDR luminance represents



[Pattanaik et al., 2000] [Mantiuk et al., 2008] [Reinhard et al., 2002] [Ramsey et al., 2004] [Ramsey et al., 2004] with BC

Figure 4.8: Area surrounding the pixel studied in Figure 4.7 with the considered TMOs. The upper image corresponds to frame 25 while the lower one to frame 35.

a photometric quantity expressed in cd/m^2, while LDR luma represents a code value relative to a display-dependent color space. However, a black or white point is not defined in HDR. For this reason, we define the white point as the maximum HDR luminance value of the image and the black point its minimum. To make the video fade effect perceptually linear, we apply Equation 4.5 in the log domain.

To preserve video fade effects, the parameter ζ (Equation 4.4) must be set to 0. To deal with fade to black or fade to white, the anchor must be defined as the brightest (χ="Max") or the dimmest (χ="Min") frame respectively. A fading can also be detected during the video analysis using a fade detection technique [Alattar, 1997]. Video TMOs cannot handle video fade effects as they do not analyze the video before performing the tone mapping.

In this section, we show some results obtained when applying the BC algorithm on video fade sequences. Figure 4.9 illustrates the *MtStMichelBlack* sequence in false color as well as after applying [Ramsey et al., 2004] operator with and without the BC post-processing. Without post-processing, the TMO fails to reproduce the fade to black effect. Figure 4.10 shows the results of tone mapping a fade to white sequence. Similarly to the fade to black, the video fade effect is preserved only if the BC post-processing is applied.

4.4 Limitations of the BC Post-Processing

The BC post-processing was evaluated in [Eilertsen et al., 2013, Melo et al., 2013] and was proven to have less temporal artifacts than other TMOs. However, these evaluations also reported that the provided results were often too dim which lessened their subjective quality. Indeed, the range associated with each frame corresponds to a portion of the

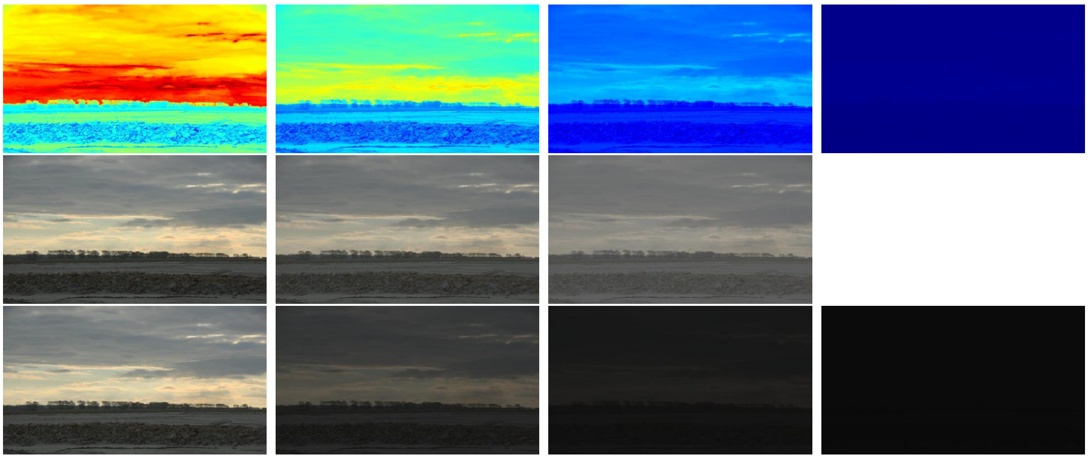


Figure 4.9: *MtStMichelBlack* video sequence represented in false color representation (top-row) and after applying [Ramsey et al., 2004] operator without (middle-row) and with the BC post-processing (bottom-row). Black fade effect illustrated through frames 1, 33, 66 and 99 (from left to right).

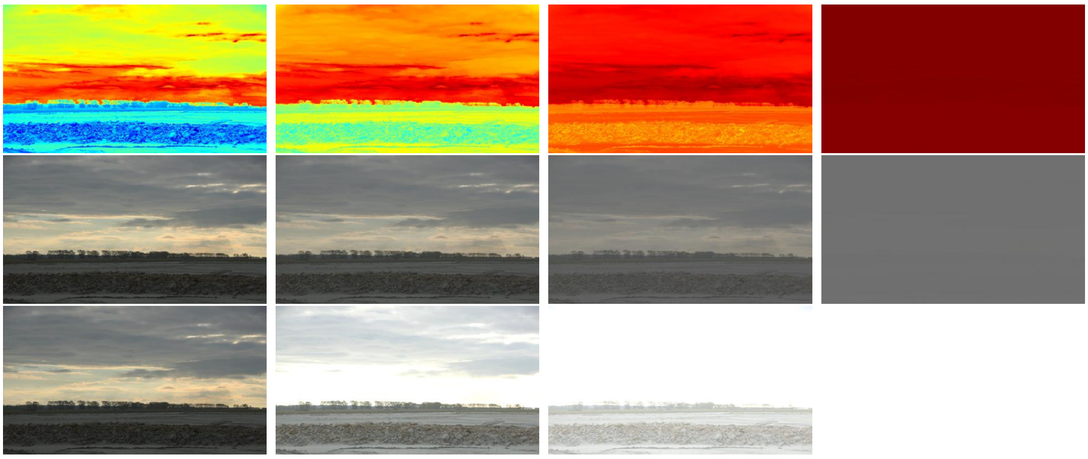


Figure 4.10: *MtStMichelWhite* video sequence represented in false color representation (top-row) and after applying [Ramsey et al., 2004] operator without (middle-row) and with the BC post-processing (bottom-row). White fade effect illustrated through frames 1, 33, 66 and 99 (from left to right).

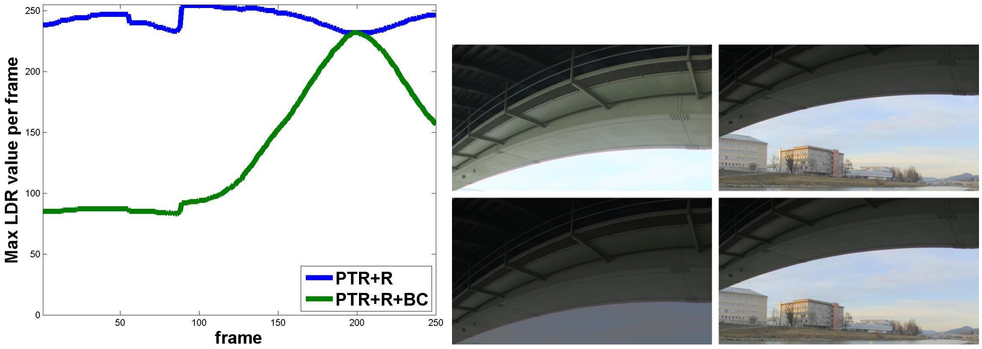


Figure 4.11: Reduced dynamic range and loss of contrast due to the BC algorithm. Left: maximum LDR luma value of the *UnderBridgeHigh* sequence with and without using the BC post-processing. Right: reduced contrast and TOI artifact (sky's brightness) due to the global scaling. Frames 120 and 160 tone mapped with the TMO alone (top row, [Ramsey et al., 2004] operator) and with (bottom row) the BC post-processing.

available dynamic range (this portion being defined by the scaling). Figure 4.11 plots the used dynamic range of a tone mapped sequence with and without using the BC post-processing (left). Without the BC, the maximum LDR luma remains close to its maximum value (i.e. 255). On the other hand, the BC technique reduces the used dynamic range up to 2/5 of its maximum value. The difference of results between the TMO alone (top frames on the right) and the TMO associated with the BC post-processing (bottom frames on the right) illustrates this reduced contrast. The rightmost frames correspond to the anchor frame used, which is not affected by the BC post-processing. On the contrary, as the BC only preserves the overall brightness coherency, the sky's brightness in the bottom left frame is scaled down although it is the brightest element in the video sequence. From this example, we see that preserving only the global brightness coherency not only reduces the spatial contrast but also impairs the reduction of **Temporal Object Incoherency (TOI)** artifacts.

4.5 Summary

In this chapter, we presented a post-processing method to preserve the overall HDR brightness coherency in a tone mapped sequence. Our method relies on an anchor based on the geometric mean to scale each frame of a tone mapped video sequence. Two parameters tune the post-processing, χ to choose the anchor and ζ to prevent a too low scale ratio. We provided results regarding the reduction of **TBI** and **TOI** artifacts as well as the preservation of artistic effects such as fading. This technique has been published in:

- **R. Boitard**, K. Bouatouch, R. Cozot, D. Thoreau and A. Gruson, "Temporal Coherency for Video Tone Mapping," in *Proc. SPIE, Applications of Digital Image Processing XXXV*, 2012.

The BC post-processing was evaluated in [Eilertsen et al., 2013, Melo et al., 2013] and was proven to have less temporal artifacts than other TMOs. However, these evaluations also reported that the provided results were often too dim which lessened their subjective quality. Indeed the BC post-processing trades off spatial contrast to increase the temporal brightness coherency. As this technique deals only with the overall brightness, each pixel of each frame is scaled down without discrimination. In other terms, no pixel, apart from those belonging to the anchor frame, is mapped to the maximum of the dynamic range. Consequently, the range associated with each frame corresponds to a portion of the available dynamic range (this portion being defined by the scaling).

Chapter 5

Zonal Brightness Coherency

In the previous chapter, we presented the Brightness Coherency (BC) post-processing that aims at preserving the temporal brightness coherency. Similar to the method proposed in [Farbman and Lischinski, 2011], the BC technique performs well when the brightness fluctuations in a scene change in a global way. However, for local fluctuations, this technique scales similarly each pixel of a frame, resulting in a lack of spatial contrast due to the reduced dynamic range as pointed out in [Eilertsen et al., 2013]. To overcome this problem, we propose to make the BC local by relying on a zonal system. The resulting Zonal Brightness Coherency (ZBC) post-processing is presented in this chapter that is organized as follows:

- Section 5.1. **Global and Local Post-Processing**: describes the cause of the reduced spatial contrast when using the BC and proposes a solution to overcome it.
- Section 5.2. **Zonal Brightness Coherency (ZBC) Post-Processing**: describes the modification required to make the BC post-processing local.
- Section 5.3. **Results**: presents results regarding the preservation of temporal brightness object and hue coherency, it also reports a subjective evaluation to assess the difference in subjective quality between no post-processing or the use of the BC/ZBC.

5.1 Global and Local Post-Processing

In the BC post-processing, the global brightness coherency is ensured by matching the LDR key value ratio with the HDR one (Equation 4.4). However using only one key value per frame is not always enough as shown in Figure 5.1. In this example, the left image can be segmented into three spatial areas (two green and one blue) while in the luminance domain, the histogram (right plot in Figure 5.1) suggests only two segments (leftmost and rightmost of the histogram). A single key value (green line in Figure 5.1 right) is not sufficient to accurately represent the mentioned zones. With this example

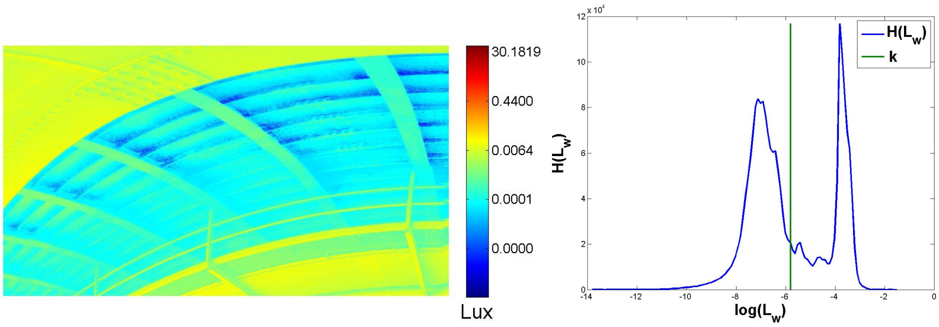


Figure 5.1: Spatial areas and luminance segments for frame 50 of the *UnderBridgeHigh* sequence. The false color luminance (left) suggests 3 spatial areas (top green bridge, blue separation and bottom green bridge) while the luminance histogram (right) only two luminance segments. The vertical green line indicates the location of the frame's key value (a.k.a. geometric mean).

in mind, we propose in the next section, to preserve the local brightness coherency using key values per zone rather than per frame.

5.2 Zonal Brightness Coherency Post-Processing

In the previous section, we outlined the issues related to the BC algorithm. They are due to the use of only one key value k_f per frame that represents only an indication of a frame's overall brightness. However, preserving the overall brightness coherency does not preserve local brightness coherency.

In order to improve this aspect, we propose to apply the BC algorithm to zones rather than to a whole frame. To this end, a histogram-based segmentation algorithm divides each frame into segments in the luminance domain. As the segment's boundaries change from frame to frame, flickering artifacts may appear as detailed in Section 5.2.1. To prevent flickering, we compute video zones based on the segments' key values. We define each video zone by two luminance values kept constant throughout the video sequence. The frame segmentation as well as the video zones determination are performed during the video analysis step. To apply the BC per zone, we compute a scale ratio per zone instead of per frame. As in the BC algorithm, the ZBC scale ratio is used to post-process any TMO's output.

The proposed decomposition of each frame in segments is similar to [Krawczyk et al., 2004] which divided an image in frameworks based on the its histogram. However, [Krawczyk et al., 2004] was looking for the brightest area in each framework to have a reference white and did not deal with videos. The steps of our solution are described in detail in the following sections.

5.2.1 Frame Segmentation

Segmenting a frame is usually done either in the spatial or the luminance domain. We chose to segment the HDR frames in the luminance domain for the following reasons. First, we want to preserve the spatial brightness coherency. By using the luminance, we make sure that two spatially distant areas of same luminance have the same scale ratio. Second, as spatial segmentation is never perfect, a scale ratio, not related to the luminance of a given pixel, could be used because of its spatial neighborhood. Third, spatial areas cannot be kept coherent throughout the video, especially when objects appear or disappear. That is why the segmentation is performed in the luminance domain and consists of several successive operations:

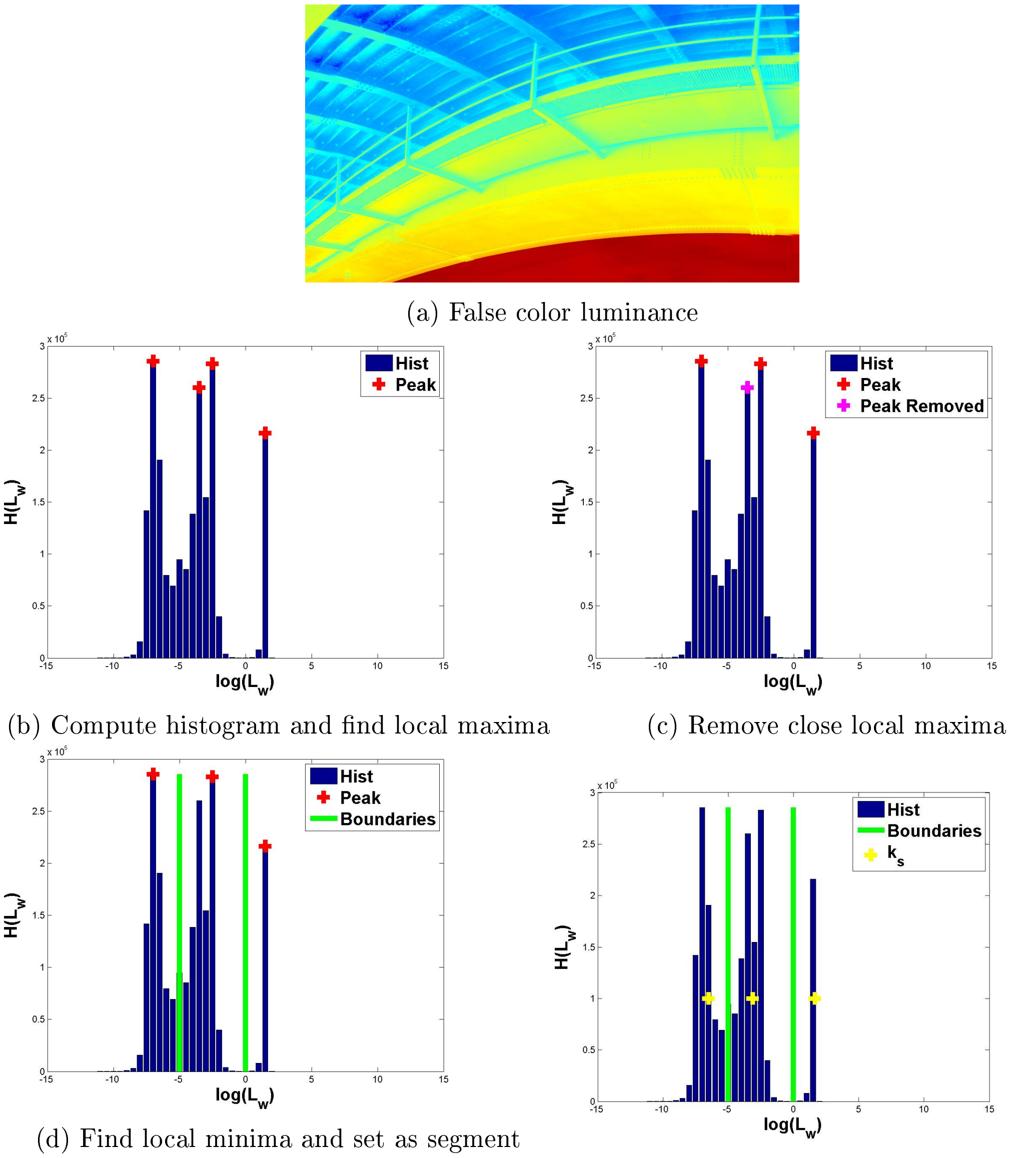
1. Compute the luminance histogram of each frame (bin width w_b in the log domain)
2. Find local maxima in the histogram (higher than threshold th)
3. Remove local maxima that are too close to each other (distance ρ)
4. Find local minima between successive local maxima
5. Define local minima as segment's boundaries
6. Compute the key value for each segment (Eq. 2.10)

Three parameters, namely w_b, th and ρ, allow to tune the segmentation. w_b corresponds to the width of each bin of the histogram (expressed in the log domain). Each local maximum of the histogram must be higher than the threshold th. Finally, ρ corresponds to the minimum distance between two local maxima (in log scale). The effects of these parameters on the segmentation are detailed in Section 5.2.4. The whole frame segmentation process is summarized in Figure 5.2. In this example, the algorithm efficiently divides the HDR frame into three separate luminance segments. Figure 5.3 presents, in false color, the corresponding luminance segments computed in Figure 5.2.

We now have several segments per frame whose boundaries change from frame to frame. Figure 5.4 plots the video analysis of the *UnderBridgeHigh* sequence, showing the results obtained for one key value k_f per frame (left) and each j^{th} segment's key value k_s^j per frame (right). However, the change of a segment's boundaries throughout the video is a source of flickering artifacts. Large variations of segment boundaries between successive frames result in different key values and hence different scaling factors. Figure 5.5 illustrates a change of boundaries between frames 127 and 128 of the *UnderBridgeHigh* sequence. To avoid this problem, we compute video zones that have constant boundaries (in term of luminance values) throughout the video.

5.2.2 Video Segmentation

Frame segmentation results in several key values k_s per frame. To prevent the change of boundaries in successive frames, we propose to determine video zones based on the segments' key values. We choose to use the segments' key values because they provide



(a) False color luminance

(b) Compute histogram and find local maxima

(c) Remove close local maxima

(d) Find local minima and set as segment boundary

(e) Compute key value per segment

Figure 5.2: Example of histogram-based segmentation with frame 108 of the *UnderBridgeHigh* sequence. (a) illustrates this frame in false color luminance while (b) to (e) summarize all the segmentation steps.

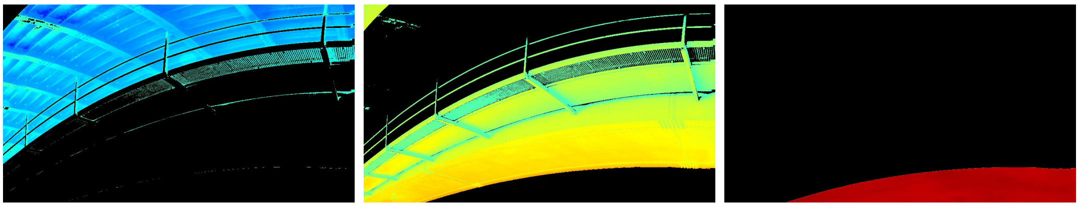


Figure 5.3: Segmentation of frame 108 of the *UnderBridgeHigh* sequence from left to right: dim, medium and bright segments. Images are displayed in false color luminance. Spatially close pixels may appear in different segments if their luminance is close to a boundary (e.g. cyan pixels in dim and medium segment).

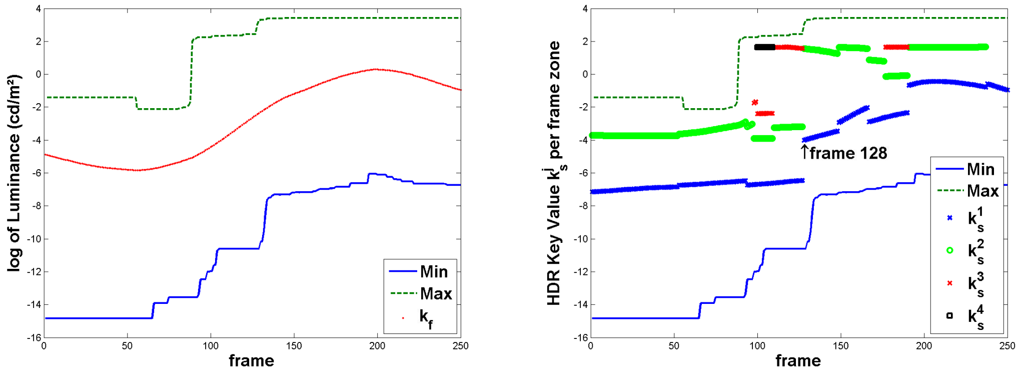


Figure 5.4: Left: min, max and key value k_f for each frame of the *UnderBridgeHigh* sequence. Right: min, max and key value k_s^j of the j^{th} segment. Up to 4 segments are detected (around frame 100). The discontinuity at frame 128 is illustrated in Figure 5.5.

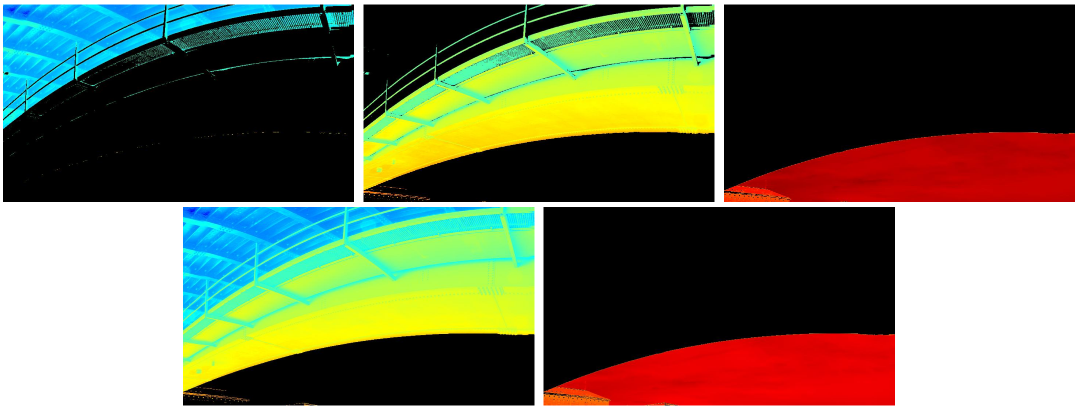


Figure 5.5: Example of change of segment's boundaries. The top row represents the three resulting segments obtained using the histogram-based segmentation of frame 127. When applied to frame 128, only two segments remain (bottom row). This change of boundaries corresponds to the discontinuity of the green and black curves in Figure 5.4.

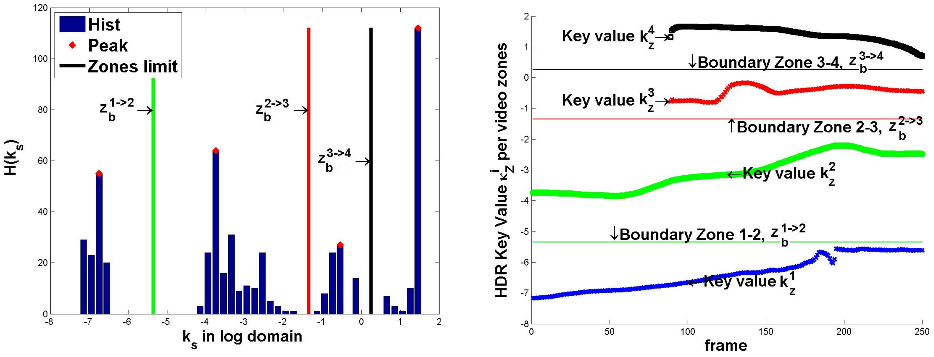


Figure 5.6: Video segmentation of the *UnderBridgeHigh* sequence. Left: segmented key value histogram where $z_b^{j->j+1}$ represents the video zone boundary between zones j and j+1. Right: j^{th} zone key value k_z^j for each frame and video zone's boundaries.

information on the distribution of the luminance values of each frame. Computing video zones on a histogram containing all the pixels of each frame of the video sequence would not take into account this information.

We want to compute video zones that best fit the distribution of the segments' key values of all the frames. We choose to use the same histogram-based segmentation as in Section 5.2.1, the luminance values being replaced by the segments' key values k_s of each frame. Figure 5.6 plots the results of the segmentation of the key values histogram for the *UnderBridgeHigh* sequence. The left plot represents the segmented key values histogram, z_b being the zones' boundaries delimiting the video zones to which we apply the BC algorithm. The figure on the right plots again the video zones' boundaries along with the video zones' key values k_z for each frame. We see that the key value within each zone varies continuously, which ensures temporal coherency.

Now that we have video zones, we apply the BC algorithm to each video zone of each frame. We consider two cases: a pixel luminance lies within a video zone or on the boundary.

5.2.3 Applying BC to video zones

After computing a key value for each video zone of each frame, we apply the BC algorithm to each zone. For each pixel located within a video zone (say, its luminance lies in-between the boundaries of a video zone), we apply the scale ratio defined in Equation 4.4. In this case, the key value $k_f^t(\mathbf{L_w})$ of the t^{th} HDR frame is replaced by the key value of the j^{th} zone of the t^{th} HDR frame $k_z^{t,j}(\mathbf{L_w})$. Similarly, the highest video zone's key value of the HDR sequence is denoted $k_{vz}(\mathbf{L_w})$ while $k_{vz}(\mathbf{L_m})$ denotes its corresponding LDR key value.

At each video zone's boundary, we define a blending zone (b_z) containing all the pixels whose luminance is close to this boundary up to a distance δ. To each pixel luminance within a b_z, we assign a scale ratio s which is a combination of those of the video zones sharing this boundary. This weighted scale ratio prevents abrupt scale ratio

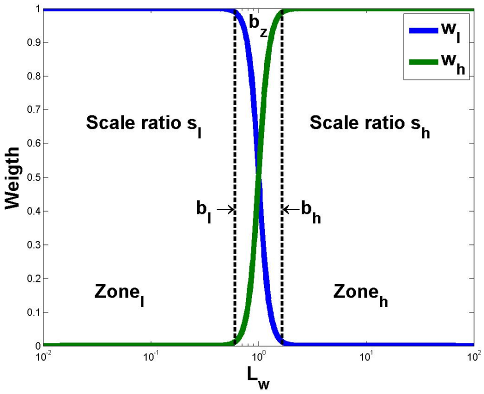


Figure 5.7: Weight distribution function of the blending zone. $\mathbf{w_l}$ corresponds to the weight of $Zone_l$ scale ratio and $\mathbf{w_h}$ the weight of $Zone_h$ scale ratio.

changes that could cause high spatial contrast not present in the HDR frame. δ is a user-defined parameter in the log domain. The weight $\mathbf{w_l}$ of the lower scale ratio s_l assigned to a pixel x lying in a blending zone is defined as follows:

$$\mathbf{w_l}(x) = \frac{exp\left(-\frac{(\log(\mathbf{L_w}(x))-\log(b_l))^2}{2\sigma^2}\right)}{s^h + s^l}, \tag{5.1}$$

where the luminance b_l represents the lower bound of b_z. s^h and s^l are the scale ratios, used in Equation 4.4, of the two video zones sharing the boundary and are used to normalize each weight. To compute the weight $\mathbf{w_h}$ of the higher scale ratio s_h, b_l is replaced by the higher bound b_h in Equation 5.1. As for σ, it corresponds to the width of the weighting function support:

$$\sigma = \frac{\delta}{2\sqrt{2\log(3)}} \tag{5.2}$$

Finally the scale ratio $\mathbf{s^{l,h}}$ of a pixel x is given by:

$$\mathbf{s^{l,h}}(x) = s^l\mathbf{w_l}(x) + s^h\mathbf{w_h}(x), \tag{5.3}$$

where δ is the boundary width expressed in the log domain (set to 1 for all our experiments). Figure 5.7 plots the weight distribution function in a blending zone b_z. Outside the zone delimited by b_l and b_h, the pixels are scaled by only one scale ratio. Inside this zone, the scale ratio is computed using Equation 5.3.

5.2.4 Parameters

Our solution relies on six parameters (χ, ζ, δ, b_w, th and ρ). χ and ζ are the same parameters as the ones used in the BC algorithm. χ selects the anchor zone depending on its geometric mean. ζ prevents low scale ratio to trade off brightness coherency for a better preservation of details in dim frames. We found that setting it to 0.1 for every tested sequence provides good results.

Regarding δ, it is used to prevent abrupt changes of scale ratio on zone's boundaries. Too low a value results in such abrupt changes while a too high one increases the number of pixels that lie on a boundary. As mentioned in Section 5.2.3, we found that setting it to 1 provides good results for all the tested sequences.

The three last parameters (namely b_w, th, ρ) allow to determine the number of video zones. As all three parameters are highly correlated, it is hard to set each of them separately. That is why, we propose a way which facilitates the tuning of these parameters. We express the bin's width b_w so as to achieve the same quantization as for 8-bit LDR imagery (256 gray levels corresponding to 8 stops in dynamic range):

$$b_w = \theta \frac{8}{256} \tag{5.4}$$

where θ is a parameter used to increase or decrease width of the quantization step (binning). A high θ provides a coarse quantization thereby reducing the precision of the histogram. From b_w, we can derive the number of bins n_b used for each segmentation as:

$$n_b = \frac{r}{b_w}, \tag{5.5}$$

where r is the dynamic range of the HDR video sequence.

Recall that a local maximum of the histogram must be higher than the threshold th. Consequently, th must be coherent with the number of bins:

$$th = \tau \frac{n_p}{n_b}, \tag{5.6}$$

where n_p is either the number of pixels in the image for the frame segmentation or the number of key values in the histogram for the video segmentation. If τ is lower than 1, then the number of selected peaks increases otherwise it decreases. Experiments showed that a good setting for all the tested sequences is $\theta = 1$ and $\tau = 2$.

Finally, we express the distance between two peaks ρ directly in the log domain. More peaks are discarded with a high ρ, reducing then the number of segments/zones used. The number of video zones directly influences the locality of the ZBC technique. We found that a value ranging from 0.5 to 1.5 offers a good compromise to tune the locality of our algorithm. Note that using only one video zone amounts to use the BC algorithm.

To summarize, we recommend to set $\delta = 1$, $\theta = 1$, $\tau = 2$ and tune only ρ for the locality (default value = 0.65) and ζ for the details/brightness coherency trade off (default value = 0.1).

In conclusion, like the BC method, the ZBC resorts to a video analysis step prior to the tone mapping operation. This analysis relies on a histogram-based segmentation as shown in Figure 5.8. Performing the segmentation on a per frame basis results in several segments per frame. The key value of the j^{th} segment of each frame is noted $k_s^j(\mathbf{L_w})$ and $k_s^j(\mathbf{L_m})$ respectively for the HDR and LDR frames. Using the segments'

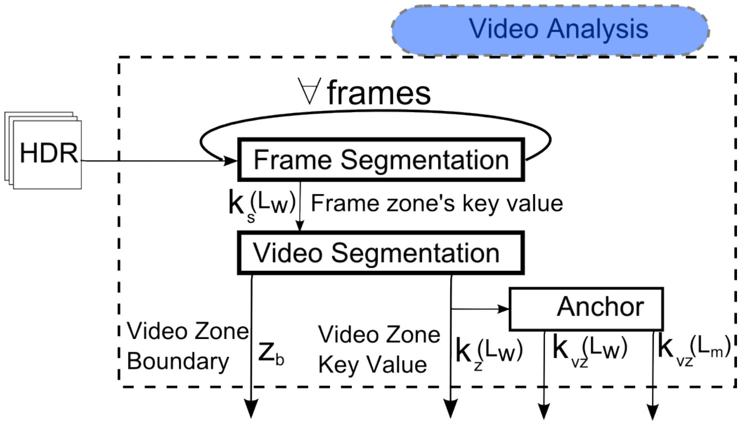


Figure 5.8: Details on the video analysis. The *Frame Segmentation* function segments each frame of the sequence and computes each segment's key value. The *Video Segmentation* determines the video zone's boundaries and their corresponding key values. The *Anchor* function determines the anchor zone in the HDR sequence $k_{vz}(\mathbf{L_w})$ and computes its corresponding LDR key values $k_{vz}(\mathbf{L_m})$.

key value histogram, a video segmentation is performed to provide zone boundaries. We then compute the key value $k_z^{t,j}(\mathbf{L_w})$ of each zone j of each frame t. Therefor a key value can be associated with either a frame (k_f), a segment (k_s) or a zone (k_z). $k_v(\mathbf{L_w})$ and $k_{vz}(\mathbf{L_w})$correspond to the anchor frame's key value and the anchor video zone's key value of the HDR sequence ($k_v(\mathbf{L_m})$ and $k_{vz}(\mathbf{L_m})$ in the LDR case).

Once the video analysis has been performed, a scale ratio is applied to each video zone. More precisely, for each video zone j of each frame t, we compute an LDR zone key value ($k_z^{t,j}(\mathbf{L_m})$) and use it to determine the associated scale ratio s^j (Equation 4.4). We then compute the blending scale ratio $\mathbf{s^{j,j+1}}$ (Equation 5.3) for each pixel belonging to a blending zone b_z. The whole workflow of our method is given in Figure 5.9.

Our ZBC algorithm performs better than the BC algorithm as will be seen in the next section.

5.3 Results

In Section 4.4, we showed that the BC post-processing suffers from two issues: lack of temporal local brightness coherency and loss of spatial contrast. The ZBC algorithm overcomes these issues, as shown by the obtained results presented in this section. As our technique is a post-processing, we can apply it to any TMO. In our experiments, we used 3 different TMOs: a flickering removal operator ([Ramsey et al., 2004]), a global operator (Exponential Mapping [Banterle, 2011]) and a local operator ([Li et al., 2005]). We experimented with different HDR sequences (*UnderBridgeHigh*, *GeekFlat* and *Tunnel* detailed in Appendix A.

In Section 5.3.1, we present results concerning the preservation of contrast and spatial brightness coherency. The temporal local brightness coherency, namely the reduc-

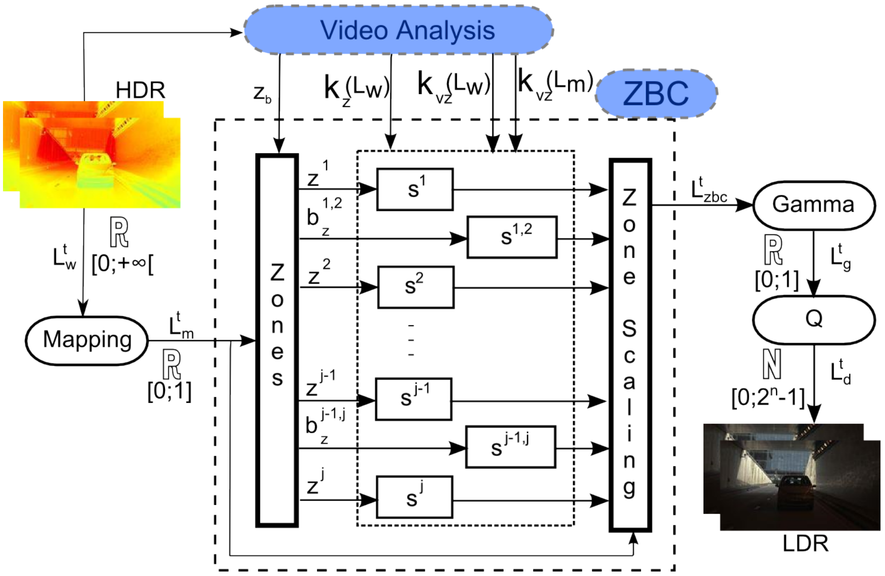


Figure 5.9: Complete ZBC workflow with details on the scaling phase. The *Zones* function determines, for each tone mapped frame and each pixel, the corresponding video zone z^j as well as the video blending zone $b_z^{j,j+1}$. Their respective scaling ratios s^j and $\mathbf{s^{j,j+1}}$ are computed. The *Zone Scaling* function applies the scale ratios to the tone mapped frames.

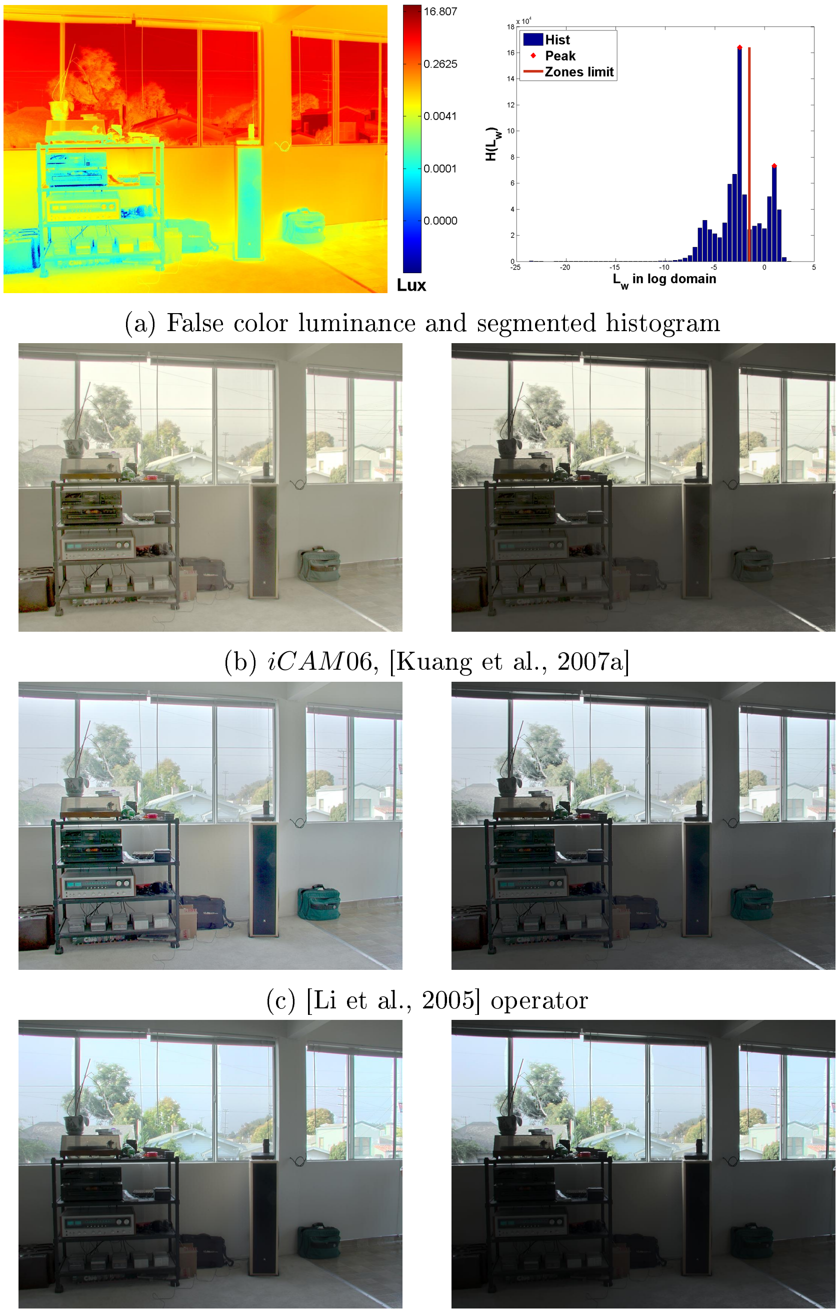
tion of **TOI** artifacts, is addressed in Section 5.3.2. Section 5.3.3 presents some results to assess the reduction of **Temporal Hue Incoherency (THI)** artifacts. Finally, we report the results of a subjective evaluation that assessed the increase of subjective quality brought by the BC and ZBC post-processing.

5.3.1 Spatial Contrast and Brightness Coherency

One of the goals of the ZBC algorithm is to preserve spatial brightness coherency and contrast. This property is valid for still images as well as video sequences. Figure 5.10 presents an HDR image tone mapped with several local TMOs with and without our algorithm. Local TMOs fail to preserve spatial brightness coherency (leftmost images in Figures 5.10): the room seems as bright as outdoors. On the contrary, the ZBC preserves the indoor/outdoor contrast (rightmost images in Figure 5.10).

For video sequences, the BC algorithm scales each frame except the anchor. As the ZBC uses as anchor a zone rather than a frame, all frames are scaled. Figure 5.11 illustrates such a case where the tone mapped frame is the brightest one of the *UnderBridgeHigh* sequence (BC scale ratio is equal to 1 when χ is set to 'Max'). Note that no change is noticed when the BC is used (Figure 5.11c). However, when applying the ZBC, a higher contrast is achieved resulting in a frame with a sharper look.

Another advantage of the ZBC is that it preserves details that otherwise would be clamped by a TMO. Indeed, many TMOs choose to burn areas in order to associate a



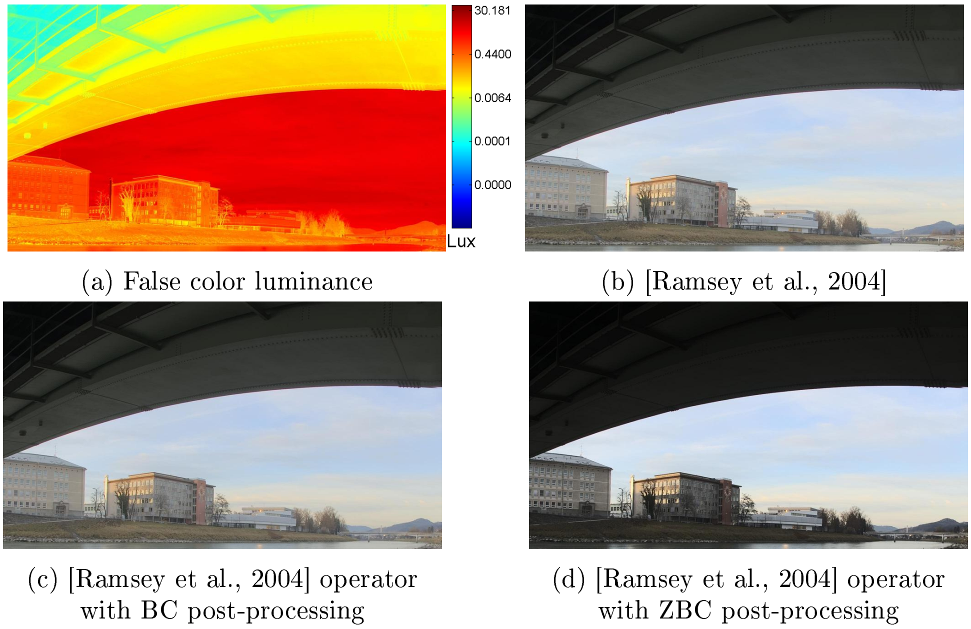
(a) False color luminance and segmented histogram

(b) $iCAM06$, [Kuang et al., 2007a]

(c) [Li et al., 2005] operator

(d) [Reinhard et al., 2002] operator

Figure 5.10: Local TMOs without (left) and with the ZBC technique (right). The ZBC preserves the indoor/outdoor spatial contrast and coherency in the tone mapped results.



(a) False color luminance

(b) [Ramsey et al., 2004]

(c) [Ramsey et al., 2004] operator with BC post-processing

(d) [Ramsey et al., 2004] operator with ZBC post-processing

Figure 5.11: Tone mapping of *UnderBridgeHigh* sequence, frame 160. Improvement brought by the ZBC when the BC has no effects. Note that to increase the contrast, some details are lost in the downside of the bridge.

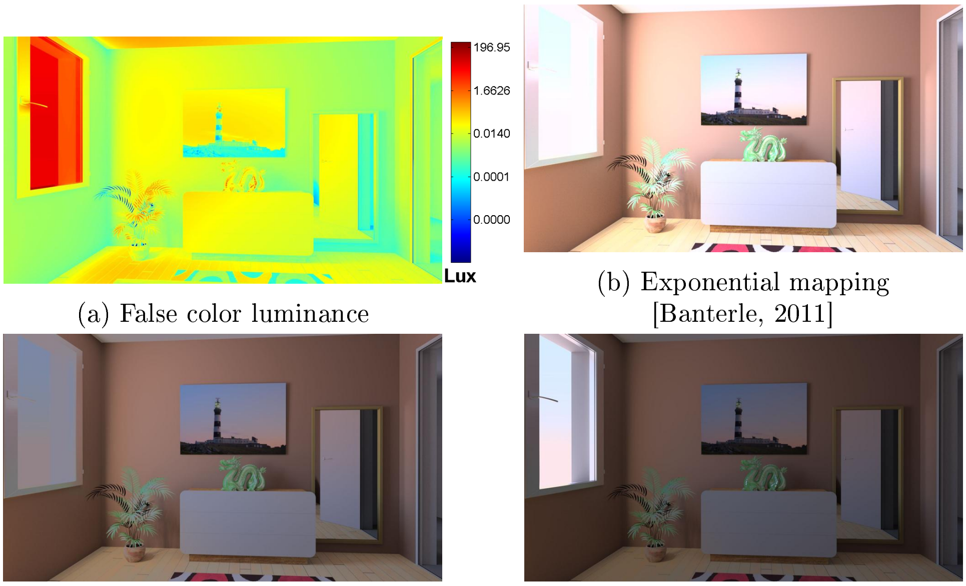
higher dynamic range with mid-tones areas. As the ZBC scales differently each zone of each frame, details ignored by a TMO can be recovered (Figure 5.12). Note the lack of contrast through the window in Figures 5.12b and 5.12c while it is preserved with our solution (Figure 5.12d). Another interesting feature of this image is the specular highlight located on the dragon. This highlight is present in each image, however only the ZBC preserves the original contrast of the HDR image (Figure 5.12a).

We showed that the ZBC preserves spatial brightness coherency and contrast even when tone mapping a still image. Concerning video sequences, the results presented in this section focused on each frame of a sequence individually. In the next section, we provide results regarding the temporal evolution of local brightness throughout a video.

5.3.2 Local Brightness Coherency

The second goal of the ZBC technique is to preserve temporal local brightness coherency using video zones. We assess the preservation of local brightness coherency by comparing the HDR and LDR evolutions of the video zones' key values throughout a video sequence.

Figure 5.13 plots such an evolution for the *UnderBridgeHigh* sequence. Figure 5.13a represents the HDR video zones' key value ($k_z^j(\mathbf{L_w})$). Figure 5.13b plots the tone mapped video zones' key value ($k_z^j(\mathbf{L_m})$) after applying [Ramsey et al., 2004] operator. Note how the TMO fails to preserve the brightness ratios (spatial coherency) between each zone. Moreover, the key values of dim zones (blue and green curves) at



(a) False color luminance

(b) Exponential mapping [Banterle, 2011]

(c) Exponential mapping [Banterle, 2011] with the BC post-processing

(d) Exponential mapping [Banterle, 2011] with the ZBC post-processing

Figure 5.12: Tone mapping of the *GeekFlat* sequence, frame 26. When combining the Exponential TMO [Banterle, 2011] and the ZBC algorithm, details lost in 5.12b reappear in 5.12d.

the beginning of the sequence and the brightest ones at the end of the sequence have approximately the same value, which is not coherent. When using the BC technique (Figure 5.13c), this incoherency is reduced thanks to the scaling, but the brightness ratio is still not preserved. With its zonal preservation system, the ZBC algorithm (Figure 5.13d) solves both issues. Note that albeit the coherency is preserved between zones, it is not necessarily true inside the dimmest zones. This is due to the ζ parameter that prevents low scale ratio. For example, in the *UnderBridgeHigh* sequence, a low bound ζ is assigned to the scale ratio to prevent all dark frames under the bridge. This bounding explains the difference in brightness of the downside of the bridge between the beginning and the end of the sequence. In other words, the ζ parameter trades off temporal brightness coherency to preserves detail in dimmer parts of a video.

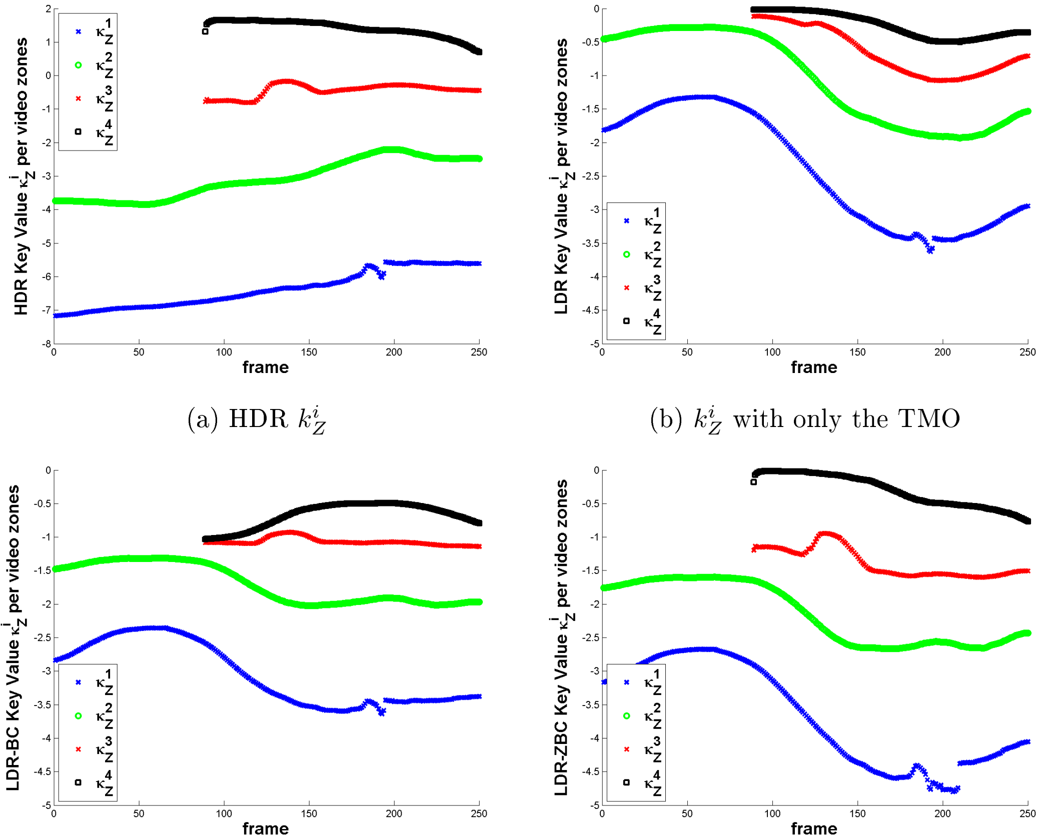
Preserving the temporal local brightness coherency amounts to preserving temporal object coherency as described in Section 3.1.4. Figure 5.14 shows frames 80, 100, 120 and 140 of the *UnderBridgeHigh* sequence tone mapped with and without using the BC or the ZBC. Without any post-processing (top-row), the lower side of the bridge in frame 80 (that is a dim zone) is as bright as the sky in frame 140. With the BC technique (3^{rd} row), the bridge brightness appears more stable, however the sky's brightness is significantly reduced. Only the ZBC (bottom row) preserves both aspects, providing frames with higher contrast and brightness coherency. However, differences are still noticeable in the brightness of the downside of the bridge, due to the use of ζ that prevents a too low scale ratio. Effectively the BC technique allows to preserve the downside of the bridge temporal coherency but fails on the sky's coherency. On the other hand, the ZBC preserves both local areas thanks to its zonal system providing a good compromise between preservation of temporal brightness and object coherency while avoiding loss of spatial contrast.

Results regarding the *Tunnel* sequence are presented in Figure 5.15. Note how the TMO alone (2^{nd} row) fails to reflect the variations of the overall brightness of the HDR sequence. In addition, the appearance of the outside of the tunnel is almost burnt out. By applying the BC technique (3^{rd} row), more overall brightness coherency is achieved, however the outside of the tunnel still lacks details. The ZBC technique (bottom row) preserves both the coherency and the appearance of the outside of the tunnel.

5.3.3 Hue Coherency

In Section 3.1.5, we described the **Temporal Hue Incoherency (THI)** artifacts that occurs when the balance between tristimulus values in successive frames is not temporally preserved. The main reason of this imbalance is color clipping that desaturates color channels at different temporal positions. As the ZBC post-processing preserves the temporal object coherency, it will also preserve the temporal hue coherency either by always or never clipping the color channels. By studying Figure 3.9 and Figure 5.16, we notice that temporal hue coherency of the studied pixel and patch is preserved when the ZBC is used.

We demonstrated that our technique reduces **TBI**, **TOI** and **THI** artifacts. However, reducing artifacts does not always result in greater subjective quality, especially



(a) HDR k_Z^i

(b) k_Z^i with only the TMO

(c) k_Z^i with TMO+BC

(d) k_Z^i with TMO+ZBC

Figure 5.13: Video zone key value (k_z^j) for the HDR (5.13a), [Ramsey et al., 2004] alone (5.13b), with the BC (5.13c) and with the ZBC (5.13d). The HDR brightness ratio between the two higher curves (red and black) is only preserved when using the ZBC (5.13d). Concerning the two lower curves (blue and green), both the ZBC and the BC solve the temporal brightness coherency problem. The lack of coherency of each curve along the sequence is due to the ζ parameter that prevents low scale ratio.

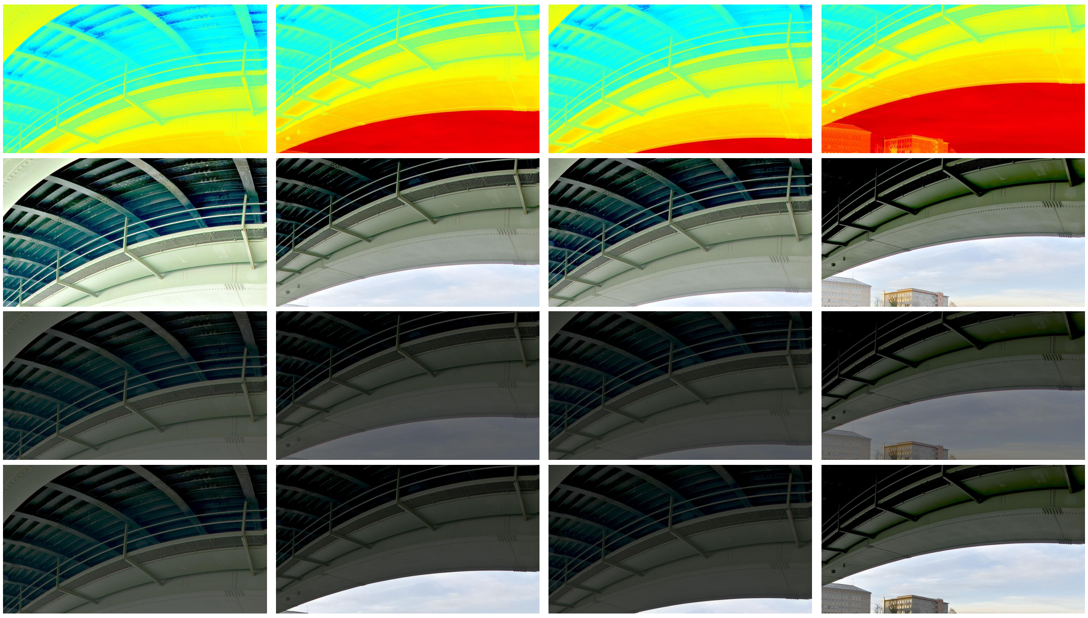


Figure 5.14: *UnderbridgeHigh* video sequence represented in false color representation (top row) and after applying [Li et al., 2005] operator alone (2^{nd} row), with the BC (3^{rd} row) and with the ZBC (bottom row). Frames number, from left to right: 80, 100, 120 and 140.

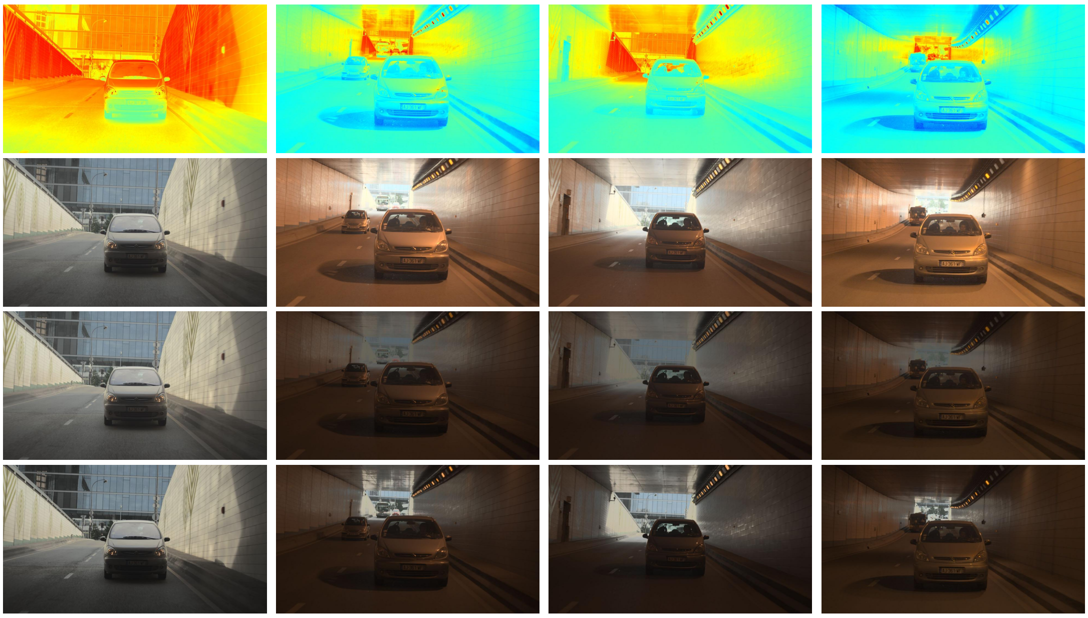
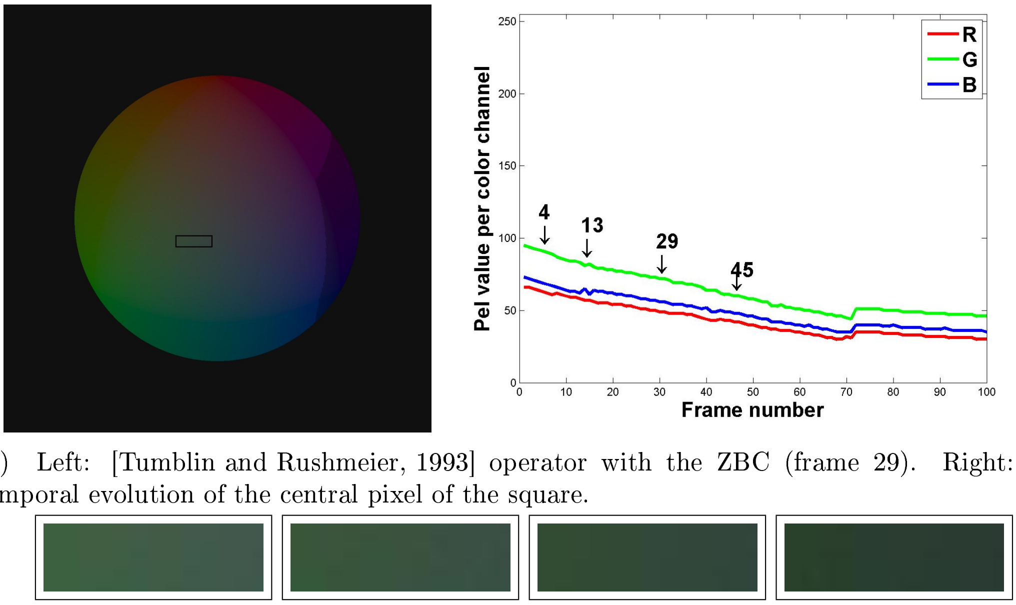


Figure 5.15: *Tunnel* video sequence represented in false color representation (top row) and after applying [Ramsey et al., 2004] operator alone (2^{nd} row), with the BC (3^{rd} row) and with the ZBC (bottom row). Frames number, from left to right: 1, 100, 200 and 300.



(a) Left: [Tumblin and Rushmeier, 1993] operator with the ZBC (frame 29). Right: the temporal evolution of the central pixel of the square.

(b) Zoom on the area outlined by the rectangle in frame 4, 13, 29 and 45 (from left to right).

Figure 5.16: Example of preserved temporal hue coherency thanks to the ZBC. The result without using the ZBC is provided in Figure 3.9.

if the overall brightness is altered. Consequently, to validate our ZBC technique, we performed a subjective evaluation presented in the next section.

5.3.4 Subjective Evaluation

To evaluate whether preserving spatio-temporal brightness coherency improves the subjective quality, we conducted a subjective experiment using the non-reference method (§ Evaluating TMOs in Section 2.3.2). For the comparison, we chose a forced-choice pairwise comparison as it provides the most accurate results as detailed in [Mantiuk et al., 2012]. The experiments were run in a dark room containing a 4K display device (TVlogic 56" Lum560W). The monitor was divided into 4 parts, the two HD videos to be compared were displayed on the upper-left and upper-right parts of the screen.

We experimented with three TMOs ([Ramsey et al., 2004] which rely on global temporal filtering, Exponential mapping [Banterle, 2011] which is a global TMO and [Li et al., 2005] which is a local TMO) and three HDR sequences *UnderBridgeHigh*, *TunnelHD* and *GeekFlat*). We performed the test for three conditions: the TMO alone (TMO), the TMO with the Brightness Coherency (BC) and the TMO with the Zonal Brightness Coherency (ZBC). The observers were asked to choose the video that they preferred. Each combination of two conditions was tested twice (inverting the left-right order of the showed videos). A choice is considered as coherent if the same video is

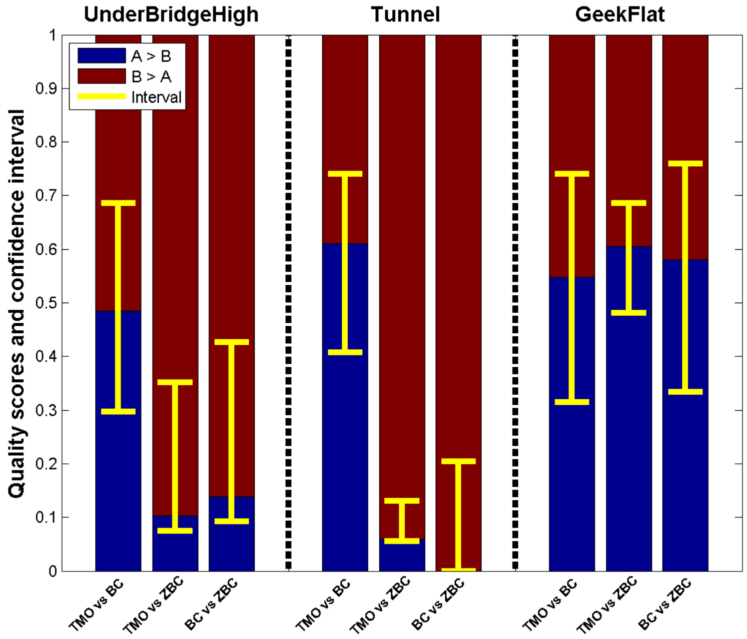


Figure 5.17: Results of the subjective evaluation per video. The quality score and confidence interval were computed using [Lee et al., 2011]. When the confidence interval (yellow segment) includes the quality score 0.5, no conclusion can be drawn.

chosen for both left-right orders. The experiment consisted of 54 comparisons (3 TMOs · 3 Sequence · 3 Conditions · 2 Combinations). 18 people (aged from 23 to 59) participated in the experiment, among them 14 are computer scientists without knowledge regarding HDR and tone mapping while the last four had prior experience with TMOs.

The results were fitted in a Bradley-Terry model using Branderson's equation [Lee et al., 2011]. The confidence intervals were computed using a non-coherent choice as a tie, while coherent choices were reported as win/lose results. Figure 5.17 plots the quality scores of each technique as well as their confidence intervals for each tested video. These results show that for a given TMO, applying the BC post-processing does not improve the subjective quality. On the contrary, when comparing the ZBC with either the TMO+BC or the TMO alone, the ZBC post-processed video is preferred for the *UnderBridgeHigh* and *Tunnel* sequence while no conclusion can be drawn for the *GeekFlat* sequence. This is mostly due to the fact that only two video zones were determined using the default parameters in the case of the *GeekFlat* sequence.

This subjective evaluation confirmed the improvement brought by our ZBC method.

5.4 Summary

In this chapter, we proposed an improvement of the Brightness Coherency (BC) algorithm so as to overcome its limitations. Our modified post-processing technique aims at preserving the local brightness coherency as well as the contrast in each frame. To this end, we divide each frame into segments using a histogram-based segmentation of HDR images. We then define video zones according to the resulting frame's segments. Finally, we apply the BC algorithm to each video zone independently. This Zonal Brightness Coherency (ZBC) algorithm better preserves both the temporal local brightness coherency and the spatial contrast in each frame. Furthermore, our technique handles, better than the BC, temporal artifacts that no other solution addressed before: **Temporal Brightness Incoherency (TBI)**, **Temporal Object Coherency (TOI)** and **Temporal Hue Incoherency (THI)**. Finally, we conducted a subjective evaluation to evaluate whether our technique increase the subjective quality of the tone mapped video content. This technique has been published in:

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal Brightness Coherency for Video Tone Mapping," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 229-246, 2014.

In this chapter, we saw that preserving temporal coherency leads to higher subjective quality content and we proposed post-processing techniques to address temporal artifacts not considered before. In addition to the subjective quality of a video, another aspect can benefit from temporal coherency: the compression efficiency. Consequently, we propose in the next part to study temporal coherency with respect to the compression efficiency.

Chapter 6

Compression of Tone Mapped Video Contents

In the previous chapters, we presented the different types of temporal artifact that impair the subjective quality of tone mapped video content. We proposed two post-processing techniques to reduce the Temporal: Brightness, Object and Hue Incoherency artifacts (**TBI**, **TOI** and **THI**). A subjective evaluation reported that post-processing video content with the ZBC increases their subjective quality. However, no such conclusion could be drawn when using the BC meaning that the efficiency of this technique highly depends on the content. However compression also affects the subjective quality of tone mapped video contents. Indeed, we mentioned in Section 2.4 that codecs are required to compress the size of video data to fit in the broadcast bit-rates or storage disc capacity. This chapter proposes a study on the relationship between video tone mapping and video compression. It is organized as follows:

- Section 6.1. **Tone Mapping and Video Compression**: describes the need of tone mapping to compress HDR and LDR video content.
- Section 6.2. **Temporal Coherency and Inter-Prediction**: analyses the impact of preserving temporal coherency on the quality of the Inter-prediction.
- Section 6.3. **Compression of Tone Mapped Video Sequences**: presents some results of compression efficiency.

6.1 Video Tone Mapping and Video Compression

In LDR digital imagery, all the signal processing operations are performed on perceptually encoded integer values (that is to say linear in regard to the perception). However, HDR data represent linear physical quantities in floating point values. Consequently, legacy video processing such as video compression are not optimized for HDR content. That is why tone mapping is required to transform HDR content into integer values before any video is being fed to a codec. In this section, we address the compression of

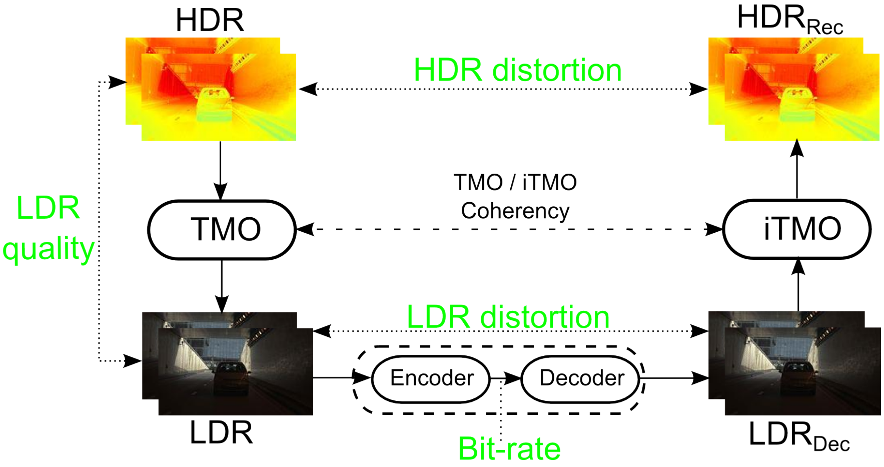


Figure 6.1: Generic compression scheme to compress HDR, LDR or both types of video contents.

two types of contents: tone mapped LDR and native HDR video content. Figure 6.1 presents a generic scheme that can be optimized to compress only the LDR, only the HDR or both LDR and HDR video content.

In the case of compressing tone mapped LDR video content only, the inverse Tone Mapping Operator (iTMO) is not needed and only three criteria drive the optimization:

- the quality of the tone mapped video content; note that this quality can correspond to an artistic intent, the best subjective quality or the fidelity to the HDR video.
- the bit-rate to encode the tone mapped video,
- the distortion between the uncompressed and the decoded video (named respectively LDR and LDR_{Dec}).

The distortion associated with a compression scheme is computed with an objective metrics such as the Peak Signal to Noise Ratio (PSNR):

$$PSNR = 10 \cdot log_{10}\left(\frac{(2^n)^2}{MSE}\right), \tag{6.1}$$

where n is the used bit-depth to represent the video sequence and the Mean Square Error (MSE) is computed as:

$$MSE = \frac{1}{M \cdot N}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\mathbf{F_o}(x,y) - \mathbf{F_d}(x,y)]^2, \tag{6.2}$$

where $\mathbf{F_o}$ (respectively $\mathbf{F_d}$) is the original frame (respectively the decoded frame) and the pair (M, N) the video spatial resolution.

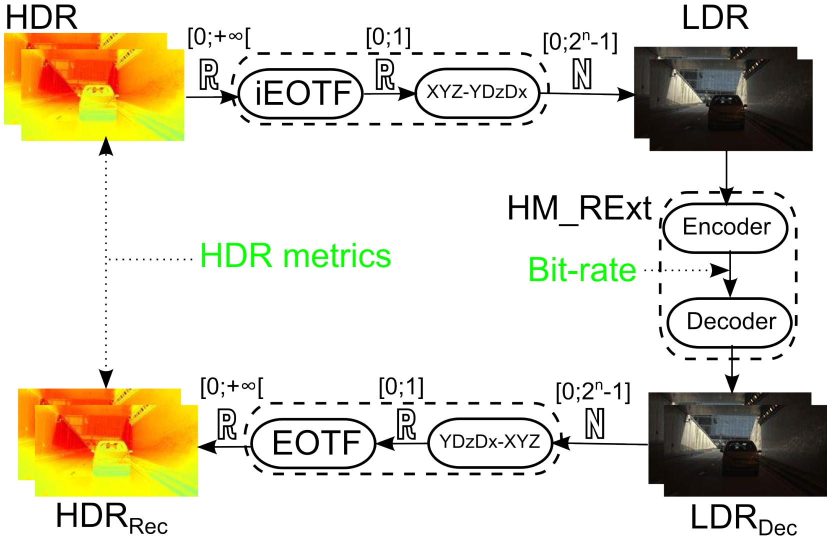


Figure 6.2: Encoding natively HDR video sequences using standard codec. The HDR values that are linear physically are first encoded perceptually using an inverse EOTF. Then the conversion from XYZ values to YD_zD_x is performed and the signal is fed to the encoder. On the decoder side, the YD_zD_x values are first converted to XYZ values and then to linear physical values using the corresponding EOTF.

When compressing only HDR video content, the TMO's characteristics are transmitted as meta-data to reconstruct the HDR sequence using inverse tone mapping. Only the pair bit-rate/distortion needs to be optimized and the latter is assessed through an HDR perceptual metrics (ΔE_{00}, HDR-VDP, PQ-PSNR, etc.). This scenario is currently considered by the MPEG standardization committee (ad hoc MPEG group on High Dynamic Range and Wide Color Gamut Content Distribution [Luthra et al., 2014]). In their experiments, the TMO/iTMO considered is an inverse Electro-Optic Transfer Function (iEOTF) / EOTF pair followed by a quantization into a color difference encoding using either 10 or 12 bits. Three types of color difference encodings have been investigated so far: YD_zD_x [SMPTE, 2014], YC_bC_r using the BT.2020 recommendation [ITU, 2012] and the logLuv [Ward Larson, 1998]. This pair is optimized for video compression purposes and the pseudo-LDR layer might not be visualized on LDR displays. Figure 6.2 illustrates the considered workflow, note that iEOTF/EOTF are constant mapping function (not adapted to the content) and are by definition temporally coherent.

Finally, when compressing HDR content to provide both an LDR and HDR output, a balance between the four presented criterion must be achieved. Table 6.1 summarizes the different broadcasting conditions associated with Figure 6.2 along with their corresponding criterion. As existing video codecs take as input integer values which is not efficient to represent linear physical ones, tone mapping is required before the compression stage. For this reason, we propose in the next section to study the relationship between video compression of tone mapped video content and the preservation

Broadcasting condition	**LDR quality**	**bit-rate**	**LDR distortion**	**HDR distortion**
Only LDR	✓	✓	✓	
Only HDR		✓		✓
LDR and HDR	✓	✓	✓	✓

Table 6.1: Summary of broadcasting conditions for compression scheme illustrated in Figure 6.1 along with their associated criterion.

TMO	D = 1	D = 2	D = 4	D = 8
[Reinhard et al., 2002]	32.85	29.37	26.37	23.35
[Reinhard et al., 2002]. + BC	38.37	34.91	31.89	28.73
[Ramsey et al., 2004]	32.84	29.36	26.35	23.38
[Ramsey et al., 2004] + BC	38.33	34.87	31.87	28.69

Table 6.2: PSNR of the Inter-predicted frame resulting from motion estimation. D represents the distance (in display order) between the reference and the predicted frame.

of temporal coherency in video tone mapping.

6.2 Temporal Coherency and Inter-Prediction

Thanks to the high level of correlation existing between successive frames of a video sequence, codecs achieve high compression ratio via the Inter-prediction. However, independently tone mapping frames of an HDR video sequence disrupts this temporal coherency. In this section, we propose to assess the impact that the preservation of temporal coherency in video tone mapping has on motion estimation.

In LDR video compression, the quality of the motion estimation, and hence the inter-prediction, is evaluated by computing the PSNR between the predicted frame and the reference one. For our study, we tone mapped the *UnderBridgeHigh* sequence using 4 different configurations of a TMO: [Reinhard et al., 2002] operator alone, with flickering reduction ([Ramsey et al., 2004] operator), with the BC post-processing and with both the flickering reduction and the BC. We evaluated the performance of each TMO by computing the PSNR between the predicted frame resulting from the motion estimation and the frame to predict. The motion estimation was performed at temporal distances corresponding to the hierarchical Group Of Pictures (GOP) structure specified in the main configuration profile of HEVC [Bross et al., 2013]. Table 6.2 reports the average PSNR of the predicted frames at temporal distances 1, 2, 4 and 8 for a block-size of 8. Results show that using the BC post-processing increases the quality of the predicted frame. Note that reducing flickering artifacts does not improve the Inter-prediction for the used video sequence.

It is usually reckoned in video compression that improving the prediction results in a higher compression efficiency. We verify the validity of this assumption in the next

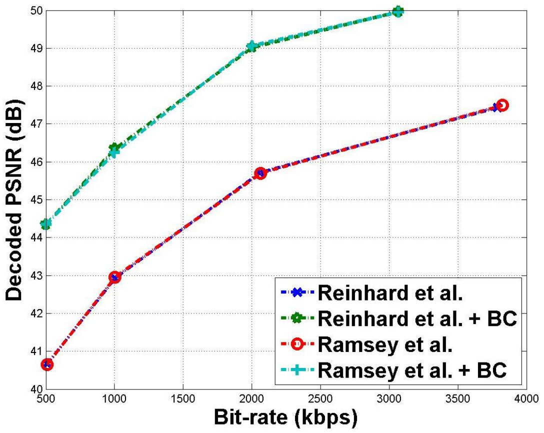


Figure 6.3: Rate-distortion curves of the *UnderBridgeHigh* and *GeekFlat* sequences for targeted bit-rates 500 kilo bits per second (kbps), 1, 2 and 4 Mega bps (Mbps). The top curves correspond to the use of the BC post-processing.

section.

6.3 Compression of Tone Mapped Video Sequence

Quality of the LDR Decoded Video The main objective of TMOs is to achieve the best subjective quality. However, when tone mapped contents are compressed using a codec, this quality is impaired. That is why, this section addresses the evaluation of the quality of decoded tone mapped LDR sequences. For our evaluation, we compared the same configurations for the TMO used in the previous section for four HDR sequences (*UnderBridgeHigh*, *GeekFlat*, *Sun* and *Tunnel*). Figures 6.3 and 6.4 plots the rate-distortion curves obtained using the main configuration profile with the HEVC test Model (HM) 10.0 [Bross et al., 2013]. As expected, the increased quality of the inter-prediction obtained with the BC post-processing results in a PSNR gain ranging from 1 to 4 dB.

In addition, we propose to illustrate the types of artifact that are generated by compressing video contents. Figure 6.5 illustrates a zoom on a frame of the *Tunnel* sequence for each of the targeted bit-rates. Notice the appearance of compression artifacts around the car for bit-rates 250 and 125 kbps.

Results show that reducing flickering artifacts does not influence the compression efficiency while using the BC post-processing does. This is explained by the fact that the BC post-processing increases temporal coherency and hence the quality of the Inter-prediction Furthermore, the distortion generated by the codec can cause artifacts that degrade the subjective quality of tone mapped video content. To sum up, achieving the best subjective quality, when performing video tone mapping, does not guarantee high quality video when it is distributed.

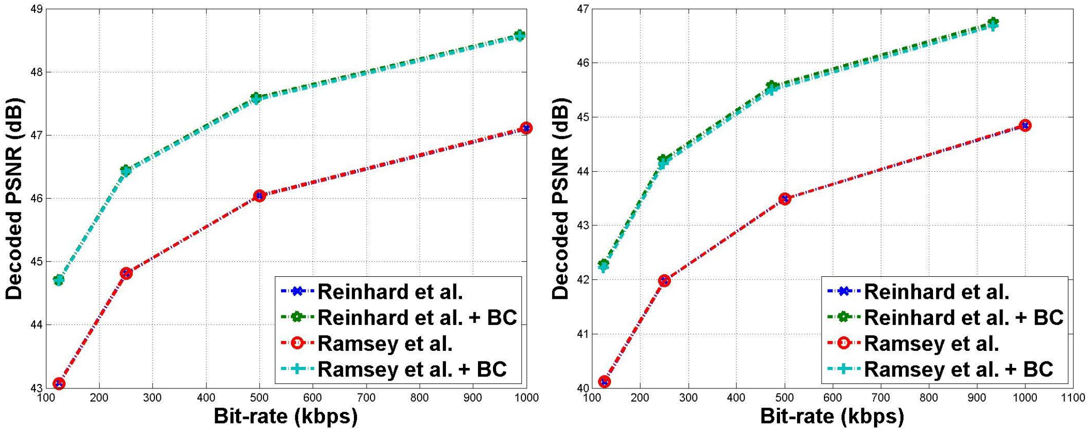


Figure 6.4: Rate-distortion curves of the *Sun* and *Tunnel* sequences for targeted bit-rates 125, 250, 500 and 1000 kbps.

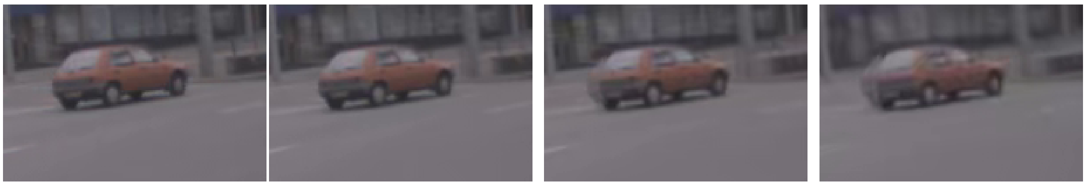


Figure 6.5: Zoom on the decoded frame 2 of the *Tunnel* sequence (TMO = [Ramsey et al., 2004]) at bit-rates 1000, 500, 250 and 125 kbps(from left to right). Artifacts around the red car appear for bit-rates 250 and 125 kbps. Corresponding PSNR with the original frame 1000 kbps = 40.82 dB, 500 kbps = 40.20 db, 250 kbps = 38.85 dB and 125kbps = 37.71 dB.

Quality of the HDR Reconstructed/Decoded Video We have demonstrated that preserving the temporal coherency using the BC post-processing allows to decrease the distortion due to the codec (higher PSNR) when broadcasting tone mapped video sequences. However, decoded LDR content can be used to address an HDR display as illustrated in Figure 6.1. That is why, we assess in this section the impact of the BC post-processing on the quality of the HDR reconstructed sequence on the decoder side.

To reconstruct an HDR video sequence from the decoded LDR video sequence, we use an inverse Tone Mapping Operator (iTMO). As all the used configurations are based on [Reinhard et al., 2002] operator (see Section 2.3.2), we apply the same iTMO to reconstruct an HDR video sequence. The different steps of the iTMO are explained as follows:

- convert n bits integer to floating point values $L_g^t = \frac{float(L_d^t)}{2^n - 1}$,
- invert the γ encoding $L_{bc}^t = {L_g^t}^\gamma$,
- (optional) invert the BC scaling (Equation 4.4): $L_m^t = \frac{L_{bc}^t}{s^t}$,

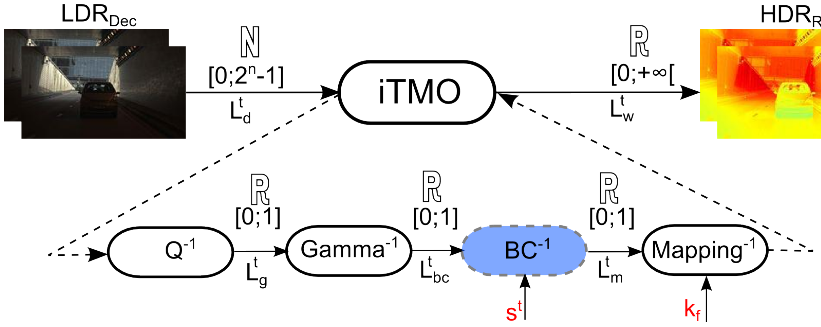


Figure 6.6: Workflow of the proposed iTMO, the BC^{-1} is only required if the BC post-processing was applied during the tone mapping. k_f and s^t need to be transmitted and decoded to reconstruct the HDR sequence. The different notations of the luminance are indicated along with their range and type of data.

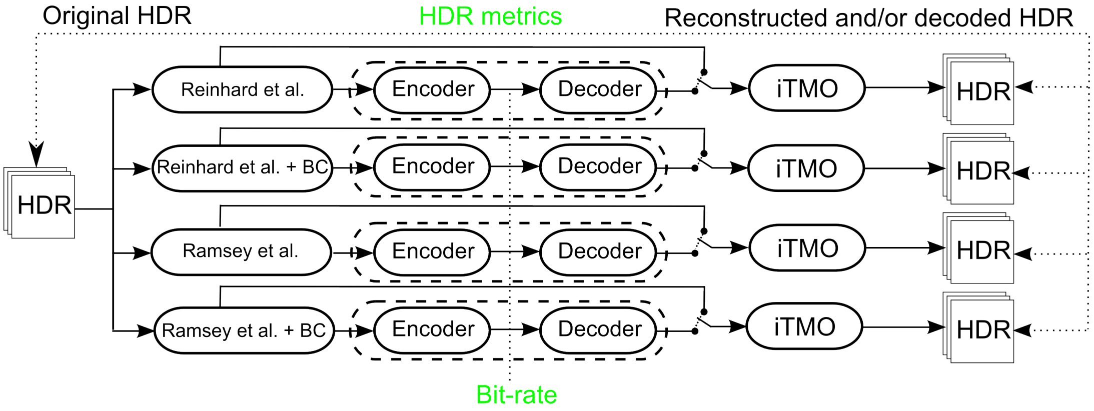


Figure 6.7: Proposed workflow to evaluate the quality of the different tone mapped sequences. An example of HDR metrics is the MSE computed in the log domain.

- invert sigmoidal compression (Equation 2.11): $L_s = \frac{L_m^t}{1-L_m^t}$,
- invert HDR frame's scaling (Equation 2.9): $L_w^t = L_s^t \frac{k_f}{\alpha}$.

α is the same parameter as in Equation 2.9 and k_f the key value of either the original HDR frame for [Reinhard et al., 2002] operator or the temporally filtered key value for [Ramsey et al., 2004] operator. To reconstruct the HDR frame, we need the key value and γ on the decoder side. We use an HEVC Supplemental Enhancement Information (SEI) message to encode this value [Bross et al., 2013]. In the case of the BC post-processing, we also need to encode the scale ratio s^t used per frame. The iTMO workflow is illustrated in Figure 6.6 with the different notations of the luminance L^t.

To evaluate the impact of both the TMO and the codec, we reconstruct two HDR video sequences, one before applying the codec and another after. Figure 6.7 represents the workflow of the test. Tables 6.3 and 6.4 report the average distortion (MSE in the log domain) between a reconstructed HDR sequence and the original one. Results

TMO	500 kbps	2 Mbps	4 Mbps	Original
Reinhard et al.	15.34	13.55	11.04	3.28
Reinhard et al. + BC	27.65	21.79	18.82	4.00
Ramsey et al.	15.39	13.63	11.12	3.38
Ramsey et al. + BC	25.51	21.80	18.81	4.18

Table 6.3: Average distortion for the *UnderBridgeHigh* sequence using the MSE metrics in the log domain (for clarity purposes, all values are multiplied by 100).

TMO	125 kbps	500 kbps	1 Mbps	Original
Reinhard et al.	9.55	7.25	6.66	2.33
Reinhard et al. + BC	11.30	8.50	7.94	2.71
Ramsey et al.	9.54	7.26	6.68	2.29
Ramsey et al. + BC	11.25	8.48	7.94	2.70

Table 6.4: Average distortion for the *Tunnel* sequence using the MSE metrics in the log domain (for clarity purposes, all values are multiplied by 100).

show that using the BC post-processing reduces the quality of the reconstructed HDR sequence even without applying the codec. This can be explained as follows. TMOs assign several HDR values to a same LDR value. This results in a loss of information regarding the HDR sequence. Contrary to other TMOs that use the full available range, the BC post-processing assigns a variable and reduced range to each frame to preserve the brightness coherency. As a result, a BC-based reconstructed HDR video sequence is more distorted due to these reduced ranges.

To illustrates the meaning of the MSE in the log domain, we show for different bit-rates a zoom on the reconstructed HDR frame 139 of the *UnderBridgeHigh* video sequence. Those HDR images are represented in false color luminance in Figure 6.8. When only the TMO is applied, only small differences on the chimney are noticeable (top row). When adding the compression at high bit-rates, banding artifacts begin to appear in the sky (2^{nd} row). Finally when lowering the bit-rates, banding artifacts increase in strength and details on the building's windows disappear (bottom row).

Results presented in this section show that reducing flickering artifacts has no effect either on the quality of the LDR video or on the reconstructed HDR video sequence. On the contrary, the BC post processing preserves temporal brightness coherency while achieving a higher compression efficiency. However, it distorts the reconstructed HDR video sequence more than other methods. Finally, for all configurations, the codec generates more distortion than the tone mapping.

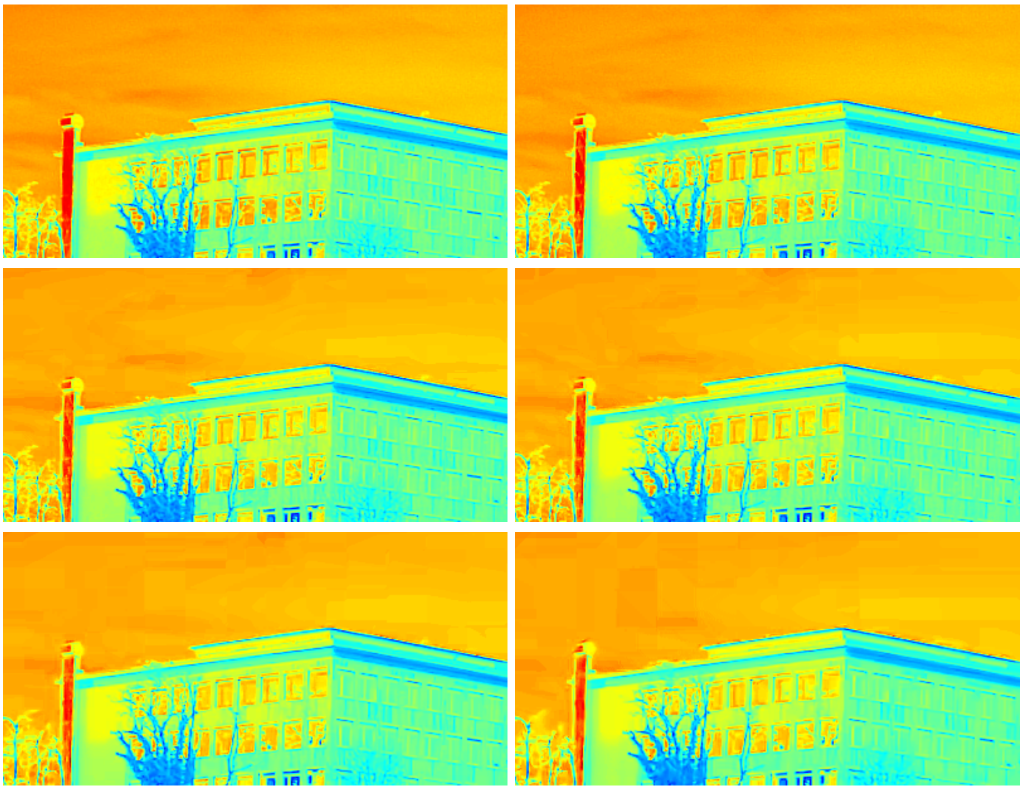


Figure 6.8: Zoom on the reconstructed HDR frame 139 of the *UnderBridgeHigh* video sequence represented in false color luminance. Top row represents the original HDR frame (left) and the reconstructed HDR frame without compression (right, MSE = 0.0620). 2^{nd} row represents the decoded/reconstructed HDR frame for bit-rates 4000 kbps (left, MSE = 0.1093) and 2000 kbps (right, MSE = 0.1130). Bottom row represents the decoded/reconstructed HDR frame for bit-rates 1000 kbps (left, MSE = 0.1269) and 500 kbps (right, MSE = 0.1416).

6.4 Summary

In this chapter, we studied the relation between the compression efficiency of tone mapped video content and the preservation of temporal coherency in video tone mapping. Two criteria were evaluated in regard to video compression: the distortion of the LDR decoded video and that of the HDR reconstructed video. Experiments were performed to assess the influence of reducing flickering artifacts and/or preserving temporal brightness coherency. Results showed that using the BC post-processing increases the compression efficiency of tone mapped video content. Furthermore, the subjective evaluation performed in Section 5.3.4 reported that applying the BC post-processing does not improve nor decrease the subjective quality. Consequently, using the BC post-processing allows to reduce the future distortion caused by a video codec while maintaining the same level of subjective quality.

Regarding the HDR reconstruction, using the BC post-processing or the flickering reduction technique alter marginally the distortion. However, when combined with a codec, the BC post-processing generates higher distortion even at high bit-rates. Finally, results showed that the main cause of distortion is compression whether we use the BC or not. This study has been published in:

- **R. Boitard**, D. Thoreau, R. Cozot and K. Bouatouch, "Impact of Temporal Coherence-Based Tone Mapping on Video Compression," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013.

To go further, we believe that these experiments should be performed for a wider range of HDR video sequences and TMOs. In addition, the ZBC post-processing should be included as it was reported to increase the temporal coherency and to improve the subjective quality as well. The only limitations of the ZBC with respect to compression efficiency is that we would need to transmit the scale ratio corresponding to each pixel in order to invert the ZBC scaling. Included so many information would more than likely impair the compression efficiency when the reconstruction of the HDR content is required.

However, one limitations of using the BC or ZBC post-processing is that this latter alters the rendering of a TMO. If this rendering was considered as an artistic intent performed previously in the pipeline, then the gain in compression is not enough to change this intent. Consequently, using the BC or ZBC is beneficial only when the LDR video sequence is not constrained to a certain rendering. That is why, we propose in the next chapter a technique to increase the compression efficiency of tone mapped video content without altering its rendering of a TMO.

Chapter 7

Motion-Guided Quantization for Video Tone Mapping

In the previous chapter, we showed that preserving temporal coherency allows to improve the compression efficiency of tone mapped video content. Other techniques aim at optimizing the tone map curve of a TMO to increase the quality of decoded content [Mai et al., 2011, Koz and Dufaux, 2012]. However, these techniques optimize the rendering of the video to achieve high compression ratio and hence do not preserve artistic intent nor achieve the best subjective quality.

However, the tone map curve is not the only aspect that can influence the compression efficiency. Indeed, the last operation performed by a TMO consists in quantizing floating point values to integer ones. We believe that by making this quantization temporally coherent a higher compression efficiency can be achieved. To this end, we propose to adapt the quantization to increase the temporal correlations between successive frames. This chapter presents this technique and is organized as follows:

- Section 7.1. **Quantization in Video Tone Mapping**: details the traditional quantization performed in video tone mapping.
- Section 7.2. **Motion-Guided Quantization**: describes our technique to increase the compression efficiency of tone mapped video content without altering their rendering.
- Section 7.3. **Results**: assesses the efficiency of our technique for compressing and denoising tone mapped video content.

7.1 Quantization in Video Tone Mapping

In this chapter, the term quantization refers to converting a floating point value to an integer one. When quantizing a value, one has only three choices: floor ($\lfloor \cdot \rfloor$), ceil ($\lceil \cdot \rceil$) or round ($\lfloor \cdot + 0.5 \rfloor$). This quantization is different from the adaptive quantization, usually considered in imagery, that consists in optimizing the distribution of the quantization bins in regard to an image's cumulative distributive function [Yang and Wu, 2012].

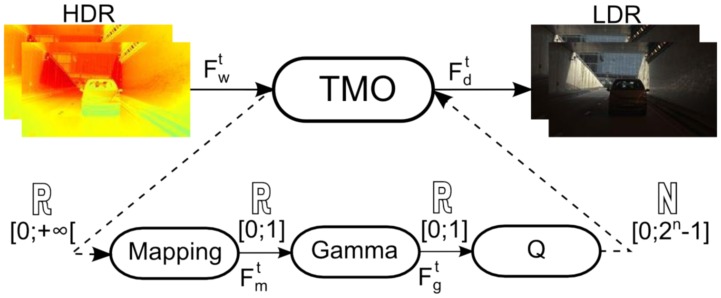


Figure 7.1: Workflow of the three steps needed to perform a tone mapping operation. F^t is the t^{th} frame of the video sequence.

Recall that in HDR imaging, the pixels represent the physical scene luminance (expressed in cd/m^2) stored as floating point values. In the case of LDR imaging, the pixels are assigned code values corresponding to a display-dependent color space (Section 2.1.3). Figure 7.1 illustrates the HDR to LDR conversion as presented in Section 2.3.2. First, the mapping operation, which is the core of a TMO, compresses HDR values to fit in the range [0-1]. Secondly, the gamma encoding redistributes the tonal level closer to how our eyes perceive them (usually $\gamma = 1/2.2$). Finally, the quantization step converts floating point values to integer code values corresponding to the used bit-depth (i.e. $[0; 2^n - 1]$ for n bits). This operation consists in scaling the gamma encoded values of the current frame ($\mathbf{F^t_g}$) to the maximum value desired (i.e. $2^n - 1$) and then rounding them to the nearest integer:

$$\mathbf{F^t_d} = Q(\mathbf{F^t_g}) = \lfloor (2^n - 1)\mathbf{F^t_g} + 0.5 \rfloor == \lfloor \mathbf{F^t_s} + 0.5 \rfloor \tag{7.1}$$

where $Q(\cdot)$ represents the quantization operation, n the used bit-depth and $\lfloor \cdot \rfloor$ the rounding to the nearest lower integer. $(2^n - 1)\mathbf{F^t_g}$ (respectively $\mathbf{F^t_d}$) represents the unquantized (respectively quantized) gamma encoded frame. For the sake of clarity, the scaled gamma encoded frame $\mathbf{F^t_s}$ will be preferred to the notation $(2^n - 1)\mathbf{F^t_g}$. The quantization is performed for each channel of the frame $\mathbf{F^t_g}$ separately, regardless of its representation, i.e. RGB, YC_bC_r etc.

7.2 Motion-Guided Quantization

Preserving temporal coherency in video tone mapping increases the compression efficiency of a tone mapped content. However, the only step of a TMO that has been made temporally coherent is the mapping function. Furthermore, by modifying the tone map curve, these techniques alter the rendering of tone mapped content. That is why we propose a technique which aims at increasing the compression efficiency of tone mapped video content without altering its rendering. We believe that by making the quantization step temporally coherent, we could achieve such a goal.

As codecs greatly rely on the temporal correlation between successive frames, we propose to adapt the quantization using the previously tone mapped frame. Figure 7.2

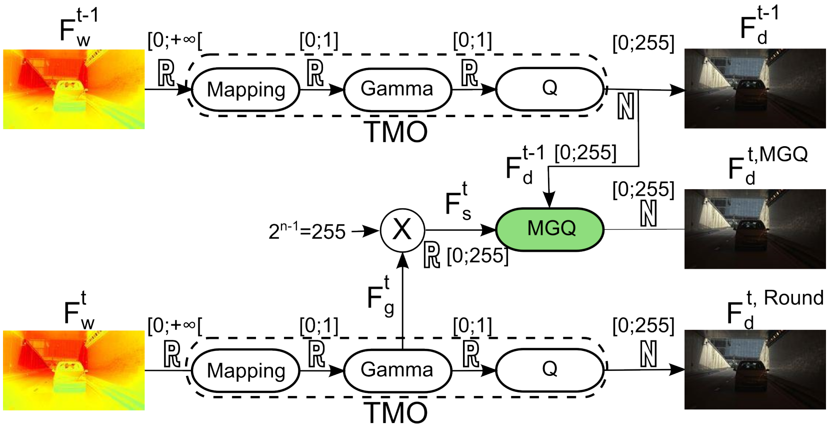


Figure 7.2: Tone mapping two consecutive frames with and without applying the MGQ. The range of each input/output as well as the type of data is indicated ($\mathbb{N}$ = uint, $\mathbb{R}$ = float).

illustrates the tone mapping of two successive frames of an HDR video sequence with and without using our technique. The following section details how our Motion-Guided Quantization (MGQ) technique adapts the quantization.

7.2.1 Our Approach

Recall that the aim of our technique is to increase the compression efficiency of tone mapped content by adapting the quantization operation. As mentioned in Section 2.4.1, the Inter-prediction relies on a motion estimation/compensation operation to remove redundant data between frames of a video sequence. That is why we first perform a Motion Estimation (ME) between $\mathbf{F_d^{t-1}}$ and $\mathbf{F_s^t}$ to obtain, for each pixel location (x, y), a motion vector $(\delta x, \delta y)$. We then compute the Motion Compensation (MC) which provides the Inter-predicted frame $\mathbf{F_p}$:

$$\mathbf{F_p}(x, y) = \mathbf{F_d^{t-1}}(x + \delta x, y + \delta y) \tag{7.2}$$

To be consistent with the prediction process used in HEVC, the motion estimation is only performed on the luma channel and the resulting motion vectors are used for each channel of a YC_bC_r frame. Our technique uses the predicted frame $\mathbf{F_p}$ to adapt the quantization of the current frame $\mathbf{F_s^t}$:

$$\mathbf{F_d^{t,MGQ}} = GQ(\mathbf{F_s^t}) = \begin{cases} \lfloor \mathbf{F_s^t} \rfloor & \text{if } \mathbf{F_s^t} - \mathbf{F_p} \geq 0 \\ \lceil \mathbf{F_s^t} \rceil & \text{if } \mathbf{F_s^t} - \mathbf{F_p} < 0 \end{cases} \tag{7.3}$$

where $GQ(\cdot)$ represents the Guided Quantization operation while $\lfloor \cdot \rfloor$ (respectively $\lceil \cdot \rceil$) represents the rounding to the nearest lower (respectively higher) integer. Recall that $\mathbf{F_s^t}$ is expressed with floating point values while $\mathbf{F_p}$ with integer ones. Both frame's values range from 0 to $2^n - 1$. The *MGQ* is applied to each channel of the frame

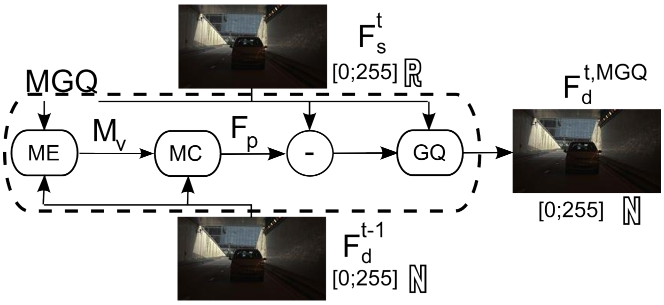


Figure 7.3: Details on the MGQ. GQ stands for Guided Quantization (Equation 7.3 or Equation 7.4). $\mathbf{F_d^{t-1}}$ is the tone mapped reference frame, $\mathbf{F_s^t}$ the tone mapped current frame before quantization. The MGQ provides the quantized tone mapped frame $\mathbf{F_d^{t,MGQ}}$.

$\mathbf{F_s^t}$ separately. The workflow of the MGQ technique is illustrated in Figure 7.3. Our method efficiently increases the quality of the Inter-prediction by reducing the distortion between the predicted frame $\mathbf{F_p}$ and the current frame $\mathbf{F_d^{t,MGQ}}$.

However, with our technique the distortion between $\mathbf{F_s^t}$ and $\mathbf{F_d^{t,MGQ}}$ is always higher than or equal to the rounding quantization. To tune the trade-off between the quantization distortion and the Inter-prediction efficiency, we add a parameter δ that enables our technique to adapt to the difference between $\mathbf{F_p}$ and $\mathbf{F_s^t}$:

$$\mathbf{F_d^{t,MGQ}} = \begin{cases} \lfloor \mathbf{F_s^t} \rfloor & \text{if} \quad 0 \leq \mathbf{F_s^t} - \mathbf{F_p} < \delta \\ \lceil \mathbf{F_s^t} \rceil & \text{if} \ -\delta < \mathbf{F_s^t} - \mathbf{F_p} < 0 \\ \lfloor \mathbf{F_s^t} + 0.5 \rfloor & \text{otherwise} \end{cases} \tag{7.4}$$

To better understand the way this trade-off behaves, let us consider three cases: $\delta = 0$, $\delta = 1$ and $\delta = \infty$. When $\delta = 0$, the GQ behaves as a rounding quantization while for $\delta = \infty$ it corresponds to Equation 7.3. For $\delta = 1$, we define Ω as the set of pixels to which the GQ has been applied, say those that satisfy $\|\mathbf{F_p} - \mathbf{F_s^t}\| < \delta$. After applying the quantization, we obtain $\mathbf{F_d^{t,MGQ}(\Omega)} = \mathbf{F_p(\Omega)}$ since the distortion was lower than δ (i.e. 1). Consequently, when predicting $\mathbf{F_d^{t,MGQ}(\Omega)}$ using $\mathbf{F_p(\Omega)}$ the resulting residuals are equal to 0. All the other pixels are quantized using the rounding operation. Table 7.1 illustrates different quantizations corresponding to different pixels conditions. Table 7.2 summarizes the trade-off between the distortion to the original unquantized values $\mathbf{F_s^t}$ and that of the predicted values $\mathbf{F_p}$.

To summarize, fixing δ allows the user to balance the number of pixels quantized using the MGQ or the rounding, based on the distortion between the unquantized values and the prediction. A higher δ means a higher distortion between $\mathbf{F_d^{t,MGQ}}$ and $\mathbf{F_s^t}$ as well as a higher quality of the prediction $\mathbf{F_p}$, thereby reducing the amount of residuals to encode.

$\mathbf{F_s^t}$	7.2	30.2	67.8	130.7	236.3
$\mathbf{F_p}$	8	28	67	127	238
$\mathbf{F_s^t} - \mathbf{F_p}$	-0.8	2.2	0.8	3.7	-1.7
$\mathbf{F_d^{t,MGQ}}, \delta = 0$	7	30	68	131	236
$\mathbf{F_d^{t,MGQ}}, \delta = 1$	8	30	67	131	236
$\mathbf{F_d^{t,MGQ}}, \delta = \infty$	8	30	67	130	237

Table 7.1: Example of the different quantization techniques. $\mathbf{F_s^t}$ is the unquantized tone mapped frame, $\mathbf{F_p}$ the predicted frame and $\mathbf{F_d^{t,MGQ}}$ the current tone mapped frame quantized using different values of δ (cf. Equation 7.4).

	$\|\mathbf{F_s^t} - \mathbf{F_d^{t,MGQ}}\|$	$\|\mathbf{F_d^{t,MGQ}} - \mathbf{F_p}\|$
$\mathbf{F_d^{t,MGQ}}, \delta = 0$	1.2	10
$\mathbf{F_d^{t,MGQ}}, \delta = 1$	2.4	8
$\mathbf{F_d^{t,MGQ}}, \delta = \infty$	3.2	6

Table 7.2: Sum of distortions resulting from the different quantization techniques of Table 7.1.

7.3 Results

Our technique aims at increasing the compression efficiency while adapting to any TMO without altering its intent. In this section, we show that the distortion obtained with our technique and the rounding quantization are very close. Finally, we report the compression efficiency of tone mapped content with and without using our technique.

7.3.1 Quantization Loss

Integer quantization assigns several floating point values to the same integer. This process obviously results in a loss of information in the quantized signal. We assess the loss due to the quantization by computing the PSNR between the unquantized current frame $\mathbf{F_s^t}$ and the quantized one $\mathbf{F_d^{t,MGQ}}$. Table 7.3 reports the PSNR using three different quantizations: Rounding, MGQ with $\delta = 1$ (noted MGQ_1) and MGQ with $\delta = \infty$ (named MGQ_{Inf}). As mentioned earlier, our quantization technique provides a slightly more distorted sequence than the rounding quantization. This distortion is no greater than one code value for all the quantized pixels (contrary to the rounding quantization that entails a maximum distortion of half a code value). For comparison, the distortion due to a lossy codec is always greater than or equal to one code value.

Quantization	Sun	Tunnel	Students	TunnelHD
Rounding	58.93	58.91	58.92	58.92
MGQ_1	56.65	56.74	57.13	56.75
MGQ_{Inf}	55.68	55.53	56.51	56.15

Table 7.3: PSNR in dB when quantizing with and w/o using our technique (59 dB corresponds to a MSE of 0.081).

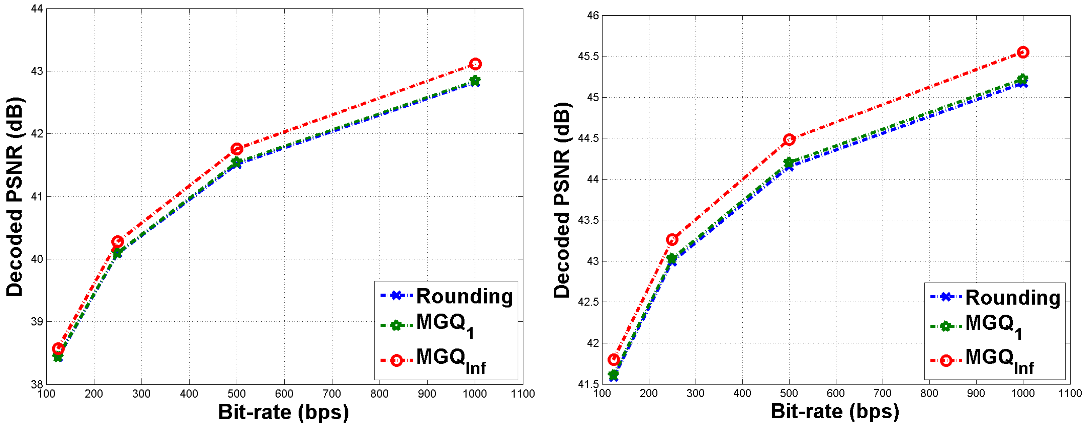


Figure 7.4: Rate-distortion results for the *Tunnel* (left) and *Sun* (right) sequences. The points represent measurements while the curves have been fitted to the experimental data.

7.3.2 Compression Efficiency

For our experiments on compression efficiency, we used the HM 12.0 with the Random Access Main Profile configuration. To assess the compression efficiency, one usually compares the PSNR between the input video and its decoded counterpart. This comparison can be performed in two different ways.

First, each input video is encoded at targeted bit-rates. A direct comparison of the PSNR allows to assess the increased quality of the content for these bit-rates. Figure 7.4 plots the results with and without using the MGQ quantization. The two sequences *Sun* and *Tunnel* are encoded at targeted bit-rates 125, 250, 500 and 1000 kbps. We used [Ramsey et al., 2004] operator to tone map both HDR sequences. Results show that we achieve a higher quality of reconstruction (between 0.15 dB and 0.4 dB gain) at the decoding stage using the MGQ_{Inf}. We can also notice that the higher the bit-rate, the higher the gain. The case MGQ_1 provides only a small improvement over the rounding operation. The trade-off between distortion and compression efficiency is illustrated through Table 7.3 and Figure 7.4. Note that by tuning the δ parameter, one can shift the MGQ_δ curve from the Rounding to the MGQ_{Inf} curve.

The second technique computes the average percentile bit-rate reduction under the same PSNR. Table 7.4 reports the Bjontegaard Distortion rate [Bjontegaard, 2008] (BD-

Sequence	Y	U	V
Sun	-12.8%	-40.1%	-40.6%
Tunnel	-10.4%	-31.7%	-32.7%
Students	-5.6%	-18.8%	-17.5%
TunnelHD	-5.4%	-21.7%	-24.5%
Average	-8.5%	-28.1%	-28.8%

Table 7.4: Average percentile bit-rate reduction under the same PSNR when comparing the Rounding and the MGQ_{Inf} quantization techniques. The BD-rate is computed using piece-wise cubic interpolation.

rate) for the tested video sequences. The sequence *TunnelHD* has been also tone mapped using [Ramsey et al., 2004] TMO. The *Students* sequence however has been tone mapped using [Farbman and Lischinski, 2011] operator. The results show that for the same quality, the MGQ_{Inf} provides an average bit-rate reduction of 8.5% for all the test sequences. Note that the *Sun* and *Tunnel* sequences perform better than the other two. This is due to the fact that these sequences are relatively noisy and our quantization technique reduces some of the temporal noise.

7.3.3 Improving Performances

Our method has some limitations, the main one being its computational complexity due to the motion estimation. Furthermore, the proposed implementation is sub-optimum with respect to the prediction process performed in HEVC. Because, first our technique does not follow the GOP hierarchical pattern that is used in a codec. Indeed, it only uses motion compensation between successive frames. Second, the Intra-prediction process should also benefit from our quantization. Third, a block-based codec uses a rate-distortion cost function to select the best predictor for each block while our solution only relies on a distortion metrics.

To summarize, the MGQ technique, if implemented in the coding loop instead of being a pre-processing, should provide an even higher compression efficiency. It would allow to tune the quantization separately for each of the available prediction modes. The selected mode and its associated quantization would depend on the codec's rate-distortion function rather than solely on the distortion. Regarding the trade-off parameter δ, it could be linked to the rate-distortion function to achieve a higher compression efficiency while reducing the quantization distortion. Furthermore, the computational complexity would no longer be an issue as the motion estimation is already performed for the Inter-frame prediction. However, to implement our approach inside the coding loop of a block-based encoder (say HEVC), the computations within the codec should be performed with floating point values rather than integer. Finally, our method should be tested with the compression of native HDR sequence using perceptual encoding (Section 6.1).

We argue that our method can adapt to any application where tone mapping is

Quantization	Sun	Tunnel	Students	TunnelHD
Rounding	48.52	45.58	47.78	49.15
MGQ_{Inf}	49.59	46.56	48.98	50.58

Table 7.5: Average PSNR between a denoised version of a tone mapped video sequence and this sequence quantized either with a rounding quantization or the MGQ.

required provided that the right test value is chosen (e.g. $\mathbf{F_p}$ for compression). To support this claim, we propose to test its efficiency on denoising tone mapped video sequences.

7.3.4 Denoising Application

As mentioned above, our method performs better for the *Sun* and *Tunnel* sequences because it reduces the temporal noise. When compression is not the targeted applications, our method can reduce the noise present in a tone mapped video sequences. Indeed, in the previous section, the MGQ was guided by the value of the difference between $\mathbf{F_s^t} - \mathbf{F_p}$. Instead of adapting to the Inter-predicted frame $\mathbf{F_p}$, we adapt our quantization to a denoised frame $\mathbf{F_n}$. The way $\mathbf{F_n}$ is computed is not relevant to this paper and any existing denoising technique can be used [Brailean et al., 1995]. For our experiments, we will consider a simple temporal-filtering with motion compensation:

$$\mathbf{F_n^t}(x, y) = \sum_{l=-N}^{M} \frac{\mathbf{F^{t-l}}(x - \delta x^{t,t-l}, y - \delta y^{t,t-l})}{w(l)} \tag{7.5}$$

where $(\delta x^{k,k-l}, \delta y^{k,k-l})$ is a motion vector obtained through a motion estimation between frames $\mathbf{F^{t-l}}$ and $\mathbf{F^t}$. N (respectively M) represents the number of non-causal (respectively causal) extents of the averaging window and $w(l)$ are the weights or the filter coefficients. Note that causal frames are expressed with integer values while non-causal ones with floating point values (including the current one which is in our case $\mathbf{F_s^t}$).

For our experiments we used only two frames in the filter bank: the previous one $\mathbf{F_d^{t-1}}$ and the current one $\mathbf{F_s^t}$. We tested our method on the same set of sequences and TMOs as in Section 7.3.2. To assess the performance of our method when compared to the rounding quantization, we compute the PSNR between the quantized frame (either $\mathbf{F_d^{t,Round}}$ or $\mathbf{F_d^{t,MGQ}}$) and the desired denoised frame $\mathbf{F_n}$. We report those PSNR in Table 7.5. For all test sequences, we achieve at least 1 dB of gain using the MGQ technique when compared to the rounding technique. The main advantage of using our technique rather than performing a denoising after the tone mapping lies in the fact that our technique does not introduce additional artifacts to the sequence. Indeed, denoising usually results in a smoothing which is source of problems when performed on edges. However, the amount of denoising that our technique can achieve is limited to the quantization which is not sufficient for really noisy sequences.

7.4 Summary

In this chapter, we pointed out that, when performing tone mapping, rounding quantization is not efficient. We chose a quantization that aimed at improving the compression efficiency of tone mapped video content. Our technique relies on the motion compensation between two successive frames of a sequence to adapt the quantization during the tone mapping. Results showed an average bit-rate reduction under the same PSNR ranging from 5.4% to 12.8%. The proposed method allows a trade-off between compression efficiency and quantization distortion of the original video. We also applied our technique to denoising to show that our method can be generalized to deal with any applications where tone mapping is needed. The work presented in this chapter has been published in:

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Motion-Guided Quantization for Video Tone Mapping," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

As mentioned in Section 7.3.3, the main limitation of our technique is its computational complexity. However, if implemented in the loop of a codec, not only the complexity issue would no longer exist, but the efficiency would be increased.

Chapter 8

Conclusion

In this thesis, we first began by outlining the limitations of the standard imagery pipeline. To overcome these limitations, we described the emergent High Dynamic Range imaging techniques from capture to rendering in Chapter 1. In the HDR imaging pipeline, we focused on the role of tone mapping that ensures backward compatibility between future HDR content and current LDR displays. Although tone mapping has been an active research field for more than one decade, few researchers have investigated the temporal aspect of tone mapping.

From this fact, we first conducted several experiments to test state of the art TMOs with several HDR video sequences. From these experiments, we described the source of six types of temporal artifacts occurring when applying naively a TMO to an HDR video sequence. We then listed and experimented with different techniques that aimed at coping with temporal artifacts. Results showed that only three out of the six types of artifact are dealt with. Furthermore, these techniques can generate two new types of temporal artifact. The descriptions of temporal artifact and video tone mapping was detailed in Chapter 3 and presented at two workshops:

- **R. Boitard**, D. Thoreau, K. Bouatouch, and R. Cozot, "Temporal Coherency in Video Tone Mapping , a Survey," in *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, 2013.

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Survey of temporal brightness artifacts in video tone mapping," in *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014.

After showing that several temporal artifacts were not accounted for in video tone mapping, we drafted the requirements for an efficient and future proof technique. These requirements were the reduction of the aforementioned artifacts as well as its adaptability to any TMO. Indeed, since the tone mapping field is in constant evolution, adapting to any TMO allows technique to be future-proof. From these two requirements, we chose to rely on an analysis of the HDR video sequence in order to post-process any TMO's output. Our post-processing consists in preserving the temporal coherency between the brightest frame of the sequence and all the other ones. The Brightness Coherency (BC) method has been presented in Chapter 4 and in an international conference:

- **R. Boitard**, K. Bouatouch, R. Cozot, D. Thoreau and A. Gruson, "Temporal Coherency for Video Tone Mapping," in *Proc. SPIE, Applications of Digital Image Processing XXXV*, 2012.

However, when evaluated in a subjective study, our technique was reported to have too low spatial contrast when the temporal contrast was high. Indeed, as this technique preserves only the overall frame brightness coherency, it fails when the temporal contrast is of local nature. Consequently, we modified the BC post-processing to make it local. The Zonal Brightness Coherency (ZBC) technique determines constant video zones over the whole sequence, each zone being bounded by two luminance values. It preserves then the temporal coherency between the brightest zone of the sequence and all the other zones. The ZBC post-processing was evaluated in a subjective evaluation which reported that it always preserves or increases the subjective quality of the tone map video. The ZBC technique has been presented in Chapter 5 and resulted in an international journal:

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal Brightness Coherency for Video Tone Mapping," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 229-246, 2014.

Reducing artifacts allows to increase the subjective quality of a video. However, as videos are bound to be distributed to the end user, they will undergo compression. Compressing a video sequence for mass distribution inevitably results in a loss of information which leads to a decrease of the subjective quality. As storage disc capacity and broadcaster's bandwidth are limited, compressed content bit-rates need to at least be below this limitation. Consequently, the level of entropy that any content can achieve is as important as the compression efficiency of a codec. That is why, we proposed, in Chapter 6, a survey on the benefits of preserving temporal coherency in video tone mapping before the compression stage. We demonstrated that even if the tone mapping achieves the best subjective quality envisioned, it is more than likely that the compression stage will impair the quality of this content. We also showed that optimizing the TMO to increase the compression efficiency can remove any artistic effect in the tone map result. This survey has been published in an international conference:

- **R. Boitard**, D. Thoreau, R. Cozot and K. Bouatouch, "Impact of Temporal Coherence-Based Tone Mapping on Video Compression," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013.

Through this survey, we understood that it is possible to modify the mapping performed by a TMO to alter the entropy of a tone mapped video sequence. However, this modification changes the visual perception of this video and can impair any artistic intent or desired rendering. We wanted a technique that reduce the entropy of a video sequence while preserving its rendering. The last operation performed by a TMO is the quantization that consists of a simple rounding to the nearest integer. As this quantization is by nature not visible (the number of quantization bins is higher than the number of Just Noticeable Difference in an 8 bits video), modifying this quantization will not alter the rendering. Consequently, we designed a Motion-Guided Quantization (MGQ) technique that adapts the quantization operation of any TMO to minimize the distortion between successive frames through a motion compensation. Thanks to the MGQ, the Inter-prediction performed in any codec will be improved, hence less residuals will need to be encoded. Results showed an average bit-rate reduction under the same PSNR ranging from 5.4% to 12.8%. The proposed method allows a trade-off between compression efficiency and the distortion between the unquantized and quantized tone mapped video. We also applied our technique to denoising to show that it can be generalized to deal with any application where tone mapping is required. The work presented in Chapter 7 was published in an international conference:

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Motion-Guided Quantization for Video Tone Mapping," Accepted in *IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

8.1 Future Work

We suggest three research of avenues as future work, two related to the work presented in this thesis and one more exploratory.

Reduction of Temporal Artifacts In Chapter 3, we described two types of temporal artifacts generated by video TMOs: **Ghosting Artifacts** and **Temporal Contrast Adaptation**. Those issues have not been dealt with in this thesis and we believe that designing generic solutions to reduce those two types of artifact is mandatory to develop the field of video tone mapping.

Ghosting Artifacts appear because temporal filtering put in correspondence pixels without relation in successive frames. To minimize these mismatch associations, motion estimation and compensation are performed to align pixels. However, foolproof motion estimation does not exist and hence mismatch motion vectors will remain. When performing temporal filtering along the motion vectors, these mismatch vectors create artifacts highly visible as they introduce edges in smooth areas. To reduce these effects, we need an efficient way to detect mismatch vectors. As long as these mismatch vectors cannot be detected accurately, video tone mapping operators that rely on spatio-temporal filtering will be prone to ghosting artifacts.

Regarding **Temporal Contrast Adaptation** artifacts, they are caused by TMOs that rely on global temporal filtering. Global temporal filtering is an efficient technique to reduce global flickering artifacts but creates artifacts in case of change of illumination. Adapting the global temporal filtering by removing outliers as in [Ramsey et al., 2004] solves this issue but only if the change of illumination is of global nature. For local changes of illumination, a pixel-wise comparison of change of illumination in the HDR and LDR sequence can be used to detect artifacts. If a change of illumination occurs in the HDR sequence and not in the LDR one, then an artifact is present. Once an artifact is detected, a post-processing operation, similar to [Guthier et al., 2011], but on a per pixel basis, could be used to alleviate it.

Adaptive Quantization The second research avenue corresponds to the implementation of the Motion-Guided Quantization in the loop of an encoder. As stated in Chapter 7, the current pre-processing method does not take into account the group of picture hierarchy used in a codec, nor does it adapt the quantization to the Intra-prediction. Furthermore, the only criterion to adapt the quantization is the distortion while in the codec it could be adapted to the Rate Distortion Optimization (RDO).

Another lead could be to use the adaptive quantization when encoding HDR video content perceptually. Indeed, current standardization bodies are investigating perceptual transform to convert HDR values (floating point values that are physically linear) to perceptually uniform values (represented by integer on a limited bit-depth). We believe that using the adaptive quantization in this context could also improve the compression efficiency of HDR content.

HDR Color Encoding The last research avenue is related to the perceptual encoding of HDR values to perceptually linear ones. Two main transforms are of interest for video compression: the Electro-Optic Transfer Function (EOTF) and the color difference encoding. EOTF is a constant transformation that encodes luminance values to luma values [Mantiuk et al., 2004, Miller et al., 2013, Kunkel et al., 2013, Touzé et al., 2014].

Color difference encoding separates luma from chroma channels. This type of representation allows to subsample chroma as we are less sensitive to color difference than luma difference. Evaluating the minimum bit-depth required to encode HDR content with those different EOTFs and color difference encoding would allow to discriminate perceptually relevant information from quantization noise.

Afterword All those research avenues along with the work presented in this manuscript demonstrate that video tone mapping is not a solved problem. Furthermore, with the recent public announcements of HDR commercial displays and cameras, it is likely that more issues related to video tone mapping will arise in the near future.

Résumé en Français

Depuis les origines du domaine de la photographie, les professionnels de l'imagerie tentent de résoudre deux principaux problèmes :

- comment capturer une scène réelle avec la plus grande fidélité?
- comment reproduire l'apparence de cette scène sur un écran aux caractéristiques techniques limitées?

Pour essayer de résoudre ces deux questions, la photographie analogique puis numérique n'a cessé d'évoluer. Cependant, les technologies de captation et d'affichage sont toujours très limitées en termes de capacité d'affichage comme illustré sur la Figure 8.1.

Grâce à cette figure, nous constatons que pour capturer une partie de la gamme de luminance qui existe dans une scène réelle, une caméra a besoin de s'adapter. En effet, les caméras possèdent un facteur d'exposition (**eV**) qui permet de modifier l'exposition d'une scène. Cependant, les dispositifs d'affichage (TV, écrans, etc..) ne peuvent reproduire que des luminances appartenant à un intervalle fixe donné (celui-ci dépendant du dispositif). Enfin, une caméra peut en général capturer plus que ce que l'on peut reproduire sur un écran. A partir de toutes ces observations, on comprend qu'une scène capturée par une caméra a besoin d'être adaptée aux capacités de l'écran ciblé.

En imagerie numérique, cette adaptation se fait au niveau du capteur de la caméra qui transforme la luminance mesurée en valeur de pixel standardisée. Les images ainsi obtenues peuvent donc être directement affichées sur un écran traditionnel. Cependant, tous les écrans qui utilisent un même standard de représentation ont des caractéristiques techniques différentes, le rendu du contenu varie donc d'un écran à un autre. De plus, comme la caméra peut capturer une dynamique de valeurs supérieure à celle qu'un écran peut représenter, une partie de l'information sera forcément perdue durant cette transformation.

Pour résumer, l'imagerie telle que nous la représentons traditionnellement pose plusieurs problèmes :

- pendant la captation, de l'information est perdue puisque la caméra ne peut pas capturer toute la gamme de luminance d'une scène.
- pendant la conversion en valeur de pixel standardisée, l'information capturée par la caméra est adaptée aux capacités de représentation du standard avant d'être quantifiée.

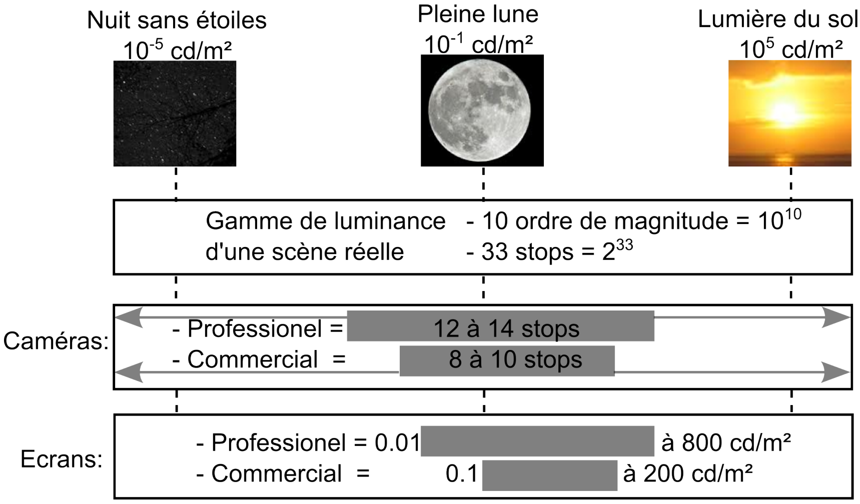

Nuit sans étoiles
10-5 cd/m²
Pleine lune
10-1 cd/m²
Lumière du soleil
105 cd/m²
Gamme de luminance d'une scène réelle
- 10 ordre de magnitude = 1010
- 33 stops = 233
Caméras:
- Professionel = 12 à 14 stops
- Commercial = 8 à 10 stops
Ecrans:
- Professionel = 0.01 à 800 cd/m²
- Commercial = 0.1 à 200 cd/m²

Figure 8.1: Capacité de captation et d'affichage des technologies actuelles comparées à une scène réelle.

- pendant l'affichage, le rendu obtenu dépend de l'écran ciblé et varie donc d'un écran à un autre.

En raison de tous ces problèmes, les professionnels de l'imagerie numérique ont développé de nouvelles techniques pour capturer, représenter et afficher une scène. Cet ensemble de techniques est communément appelé imagerie HDR (pour *High Dynamic Range imaging* en anglais). L'imagerie HDR résout les problèmes mentionnés plus haut en :

- capturant pratiquement toute la gamme de luminance d'une scène grâce à des techniques de *bracketing* (c'est à dire la fusion de plusieurs images capturées à des expositions différentes),

- représentant les valeurs ainsi capturées directement en valeurs physiques, évitant ainsi une interprétation relative,

- affichant les pixels sur des écrans HDR qui reproduisent la valeur physique (luminance en candela par mètres carrés cd/m^2) sur l'écran.

Les principaux concepts de cette nouvelle imagerie numérique sont présentés dans le Chapitre 2.

Cependant, les technologies d'affichage actuelles ne vont pas disparaitre instantanément et la rétrocompatibilité entre les contenus HDR et les écrans LDR traditionnels (*Low Dynamic Range*) se doit d'être assurée. L'opération qui permet de convertir des contenus HDR en contenus LDR est appelée *tone mapping*. C'est dans ce contexte de *tone mapping* que cette thèse a été initiée.

Travaux de la thèse

Durant cette thèse, nous avons en premier lieu défini les limitations de l'imagerie traditionnelle et évalué les aspects de l'imagerie HDR qui permettent de dépasser ces limitations. Une fois les différences entre HDR et LDR bien maitrisées, nous nous sommes intéressés au *tone mapping* qui permet la rétrocompatibilité entre les contenus HDR et les écrans LDR. L'état de l'art du *tone mapping* nous a permis de comprendre qu'une multitude de Tone Mapping Operators (TMOs) existait avec des buts différents : de la simulation du système visuel humain au plus beau rendu possible. Ces méthodes nécessitent souvent d'affiner le réglage de paramètres en fonction du contenu et aucune technique ne permet d'avoir un rendu satisfaisant pour un large panel d'images. De plus, toutes ces méthodes n'ont été testées que pour des images et non pour des vidéos HDR. La raison étant que très peu de contenu vidéo HDR existait jusque-là et que l'affinage de paramètres est fastidieux pour une vidéo.

Nous avons ainsi étudié le comportement des TMOs lorsqu'ils sont utilisés pour transformer une vidéo HDR. Grâce à cette étude, nous avons isolé six types d'artefacts temporels qui apparaissent lorsqu'un TMO est appliqué indépendamment à chaque image d'une vidéo. Quelques techniques ont ainsi été développées pour tone mapper des images successives d'une vidéo HDR de façon cohérente. Ces techniques, appelées vidéo TMOs, ont été testées sur le même corpus de séquence et nous avons constaté qu'elles ne permettaient de résoudre que trois des six types d'artefacts. De plus, ces vidéo TMOs pouvaient également créer deux nouveaux types d'artefacts temporels. Le détails des différents types d'artefacts temporels ainsi que la description des vidéo TMOs sont présentés dans le Chapitre 3 ainsi que dans deux articles de *workshop* :

- **R. Boitard**, D. Thoreau, K. Bouatouch, and R. Cozot, "Temporal Coherency in Video Tone Mapping , a Survey," in *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, 2013.
- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Survey of temporal brightness artifacts in video tone mapping," in *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014.

L'état de l'art du *video tone mapping* nous a permis d'identifier trois types d'artefacts temporels non-résolus par les méthodes existantes. Deux choix étaient possibles pour réduire ces artefacts: créer un nouveau vidéo TMO ou bien proposer une méthode générique qui permet de réduire ces artefacts lorsque l'on utilise n'importe quel TMO. Comme le domaine du *tone mapping* est en perpétuelle évolution, nous avons choisi la seconde solution pour que notre méthode puisse s'adapter aux futurs TMOs développés. Nous avons donc proposé une technique de correction de vidéo tone mappée en se basant sur une analyse de la vidéo HDR. Un traitement a posteriori préserve la cohérence temporelle de la vidéo tone mappée en s'assurant que le ratio de brillance entre l'image la plus brillante de la séquence HDR et les autres images soit préservé dans la vidéo LDR. Notre technique baptisée *Brightness Coherency* (BC) est présentée dans le Chapitre 4 et a donné lieu à un article de conférence internationale :

- **R. Boitard**, K. Bouatouch, R. Cozot, D. Thoreau and A. Gruson, "Temporal Coherency for Video Tone Mapping," in *Proc. SPIE, Applications of Digital Image Processing XXXV*, 2012.

Cependant, lorsque notre méthode a été utilisée lors d'une évaluation subjective, les participants ont constaté qu'elle produisait un résultat sous-exposé lorsque le contraste spatio-temporel était élevé. En effet, comme notre technique ne se base que sur la brillance globale d'une image, elle n'arrive pas à gérer un contraste de nature locale. Nous avons donc modifié notre traitement a posteriori pour le rendre local. Cette nouvelle méthode baptisée *Zonal Brightness Coherency* (ZBC) détermine des zones vidéos définies par deux valeurs de luminance constantes durant toute la séquence. Nous utilisons ces zones vidéos pour préserver la cohérence temporelle de brillance entre la zone la plus brillante et toutes les autres zones. Cette technique a également été testée lors d'une évaluation subjective et les résultats indiquent qu'elle préserve toujours ou améliore la qualité subjective pour les différentes séquences et TMOs testés. Le traitement a postériori ZBC a été présenté dans le Chapitre 5 et a été publié dans un journal international :

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal Brightness Coherency for Video Tone Mapping," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 229-246, 2014.

Réduire les artefacts temporels permet d'améliorer la qualité visuelle d'un contenu. Cependant, ces contenus vidéo sont voués à être distribués au grand public. Or les moyens de distribution traditionnels sont limités en termes de capacité de stockage ou de bande passante. Ces contenus vont donc devoir être compressés avant de pouvoir être distribués. La compression vidéo entraine inéluctablement une perte d'information qui diminue la qualité de la vidéo décodée. Dans ce contexte de compression de contenu tone mappée, deux caractéristiques principales permettent d'améliorer la qualité du contenu décodé. La première est l'efficacité du codec (*coder-decoder*) utilisé tandis que la seconde correspond à l'entropie du signal. Comme le *tone mapping* consiste en une transformation d'une vidéo HDR, le choix du TMO influence l'entropie de la vidéo LDR à encoder. Nous avons donc étudié dans le Chapitre 6 les bénéfices de préserver la cohérence temporelle lors du *tone mapping*, sur la qualité de la vidéo décodée. Nous avons démontré que l'utilisation du traitement a posteriori BC permet d'améliorer la qualité d'une vidéo reconstruite. De plus, les résultats indiquent que même si la meilleure qualité visuelle possible est obtenue après le *tone mapping*, l'étape de compression va très probablement dégrader cette qualité. Cette étude nous a montré qu'un équilibre entre la qualité visuelle du *tone mapping* et l'efficacité de la compression est nécessaire pour éviter qu'un traitement ne dégrade la qualité de l'autre. Cette étude a donné lieu à un article dans une conférence internationale :

- **R. Boitard**, D. Thoreau, R. Cozot and K. Bouatouch, "Impact of Temporal Coherence-Based Tone Mapping on Video Compression," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013.

Cette étude sur la relation entre le *video tone mapping* et la compression nous permet de constater qu'il est possible de modifier la transformation effectuée par un TMO pour améliorer l'entropie de la vidéo à encodée. Cependant, cette modification altère le rendu de cette vidéo, changeant par conséquence une quelconque intention artistique qui pouvait être imposée sur le rendu final. A partir de cette observation, notre objectif a été de développer une méthode qui améliore l'entropie d'une vidéo tone mappée sans altérer son rendu. La dernière opération d'un TMO est la quantification des valeurs flottantes en entier sur une profondeur de bits définie. Cette quantification consiste en un arrondi au plus proche entier et n'altère pas le rendu d'un contenu. En effet, le pas de quantification utilisé en imagerie LDR est plus petit que notre seuil de détection d'une différence (*Just Noticeable Difference* en anglais). A partir de cette observation, nous avons implémenté une méthode qui adapte la quantification de n'importe quel TMO pour minimiser la distorsion entre images successives en utilisant une compensation de mouvement. Cette méthode a été baptisée *Motion-Guided Quantization* (MGQ) et permet d'améliorer la qualité de la prédiction Inter images qui est réalisée au sein d'un codec. Par conséquence, la quantité de résidus à encoder est moindre et l'efficacité de compression améliorée. Les résultats montrent une réduction moyenne du débit pour le même Peak Signal to Noise Ratio (PSNR) allant de 5.4% to 12.8%. De plus, la méthode proposée permet un compromis entre le gain en efficacité de compression et la distorsion avec la vidéo originale non-quantifiée grâce à un paramètre de régularisation. Notre technique a également été utilisée pour débruiter des vidéos tone mappées et nous avons démontré qu'elle peut être généralisée à n'importe quelle application qui nécessite une opération de *tone mapping.* Ces travaux ont été présentés dans le Chapitre 7 et ont été publiés dans une conférence internationale :

- **R. Boitard**, R. Cozot, D. Thoreau, and K. Bouatouch, "Motion-Guided Quantization for Video Tone Mapping," Accepted in *IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

Contributions

Les travaux réalisés dans le cadre de cette thèse ont permis d'apporter les contributions suivantes au domaine du *video tone mapping*:

- Un état de l'art sur l'imagerie HDR.
- Une description des différents types d'artefacts temporels qui apparaissent lorsque l'on applique un TMO créé pour des images fixes à une vidéo HDR.
- Un état de l'art du *video tone mapping.*
- Un corpus de séquences synthétiques conçu pour mettre à mal les TMOs.
- Deux techniques de traitement a postériori pour réduire les artefacts temporels lors d'une opération de *video tone mapping.*

- Un évaluation de l'impact de la préservation de la cohérence temporelle lors du *video tone mapping.*
- Une technique pour améliorer l'entropie d'une vidéo tone mappée sans en altérer son rendu.
- Un analyse de la plupart des vidéos HDR actuellement disponibles.

Appendix A

HDR Sequences

In this appendix, we provide an analysis of the HDR sequences used throughout our experiments. The analysis of these sequences is presented by plotting their minimum and maximum luminance value per frame. We added the 1^{st} and 99^{th} percentile luminance value per frame to better assess the dynamic range of the sequence. The key (geometric mean) which represents an indication of the overall brightness is also plotted for each frame. These sequences were captured using different techniques and we propose to classify them by the research laboratory or project that captured or generated them. The different spatial resolutions of these sequence are listed in Table A.1.

A.1 Max Planck Institute fur Informatiks (MPII) Sequences

Two sequences from the Max Planck Institute fur Informatiks (MPII) have been made publicly available.

Sun The *Sun* sequence is composed of 860 VGA frames and illustrates a drive on a highway in the vicinity of Saarbruecken. Temporal contrast is achieved though the sun that is either in full view or obscured by trees. Figure A.1 (left) plots its analysis.

Tunnel The *Tunnel* sequence is composed of 640 VGA frames at the frequency of 25 fps. It represents a car entering a tunnel in broad daylight. Figure A.1 (right) plots its analysis.

A.2 OpenFootage Sequence

OpenFootage is a website that provides HDR panoramas of high resolution (10,000 by 5,000 pixels). Video sequences can be extracted from the panorama through panning in the high resolution image.

Full Name	Acronym	Width	Height
Video Graphics Array	VGA	840	640
quater High Definition	qHD	960	540
High Definition Ready	HD Ready or 720p	1280	720
Full High Definition	Full HD or 1080p	1920	1080

Table A.1: Spatial resolution of the sequences presented in this appendix. "p" in 1080p and 720p stand for progressive scanning.

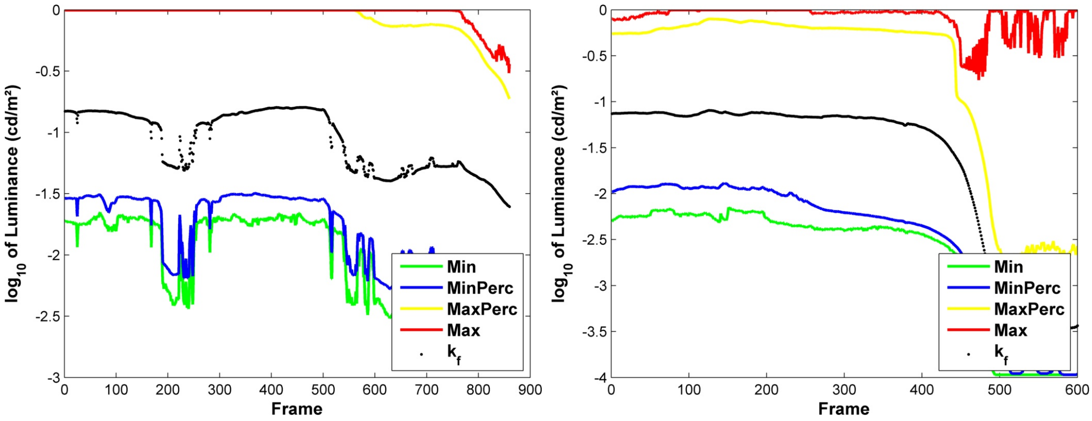


Figure A.1: Analysis of the *Sun* (left) and *Tunnel* (right) sequences.

UnderBridgeHigh The *UnderBridgeHigh* video is composed of 250 Full HD frames captured at the frequency of 25 fps. We generated these frames using a 10 pixels vertical traveling inside a high resolution HDR image. Consequently, two successive frames are similar apart from the ten lower rows which correspond to new information due to the panning. Figure A.2 (left) plots its analysis.

A.3 IRISA Sequences

In order to test the different artifacts presented in Chapter 3, we created several HDR sequences. Those synthetic sequences were rendered using Mitsuba [Jakob, 2010] and the modeling was performed with Blender.

ColorCheckBall The *ColorCheckBall* sequence is composed of 100 qHD frames generated at the frequency of 25 fps. It represents a ball moving in front of a ColorChecker (checkerboard array of 24 scientifically prepared colored squares in a wide range of colors). Part of the ColorChecker is under a bright lamp while the rest of it lies in shadows. The ball moves from the dim part to the bright one. Figure A.2 (right) plots its analysis.

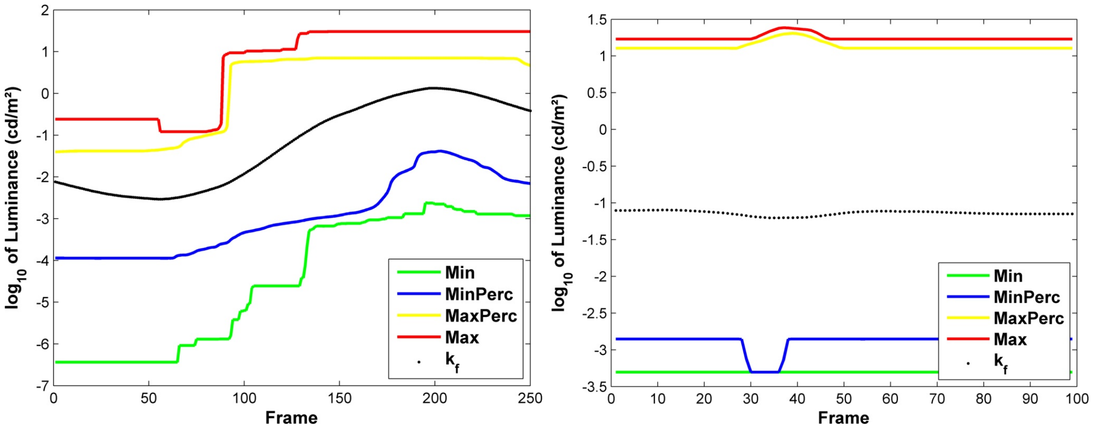


Figure A.2: Analysis of the *UnderBridgeHigh* (left) and *ColorCheckBall* (right) sequences.

HueDisc The *HueDisc* sequence is composed of 100 HD Ready frames generated at 25 fps. Contrary to the other sequences, this one was created with Matlab. It has the following characteristics:

- a disc area of constant luminance (500 cd/m^2) with a wheel color,
- a neutral gray border area with a temporally varying luminance ranging from 50 to 5000 cd/m^2.

Its analysis is plotted in Figure A.3 (left).

ModernFlat The *ModernFlat* sequence is composed of 75 HD Ready frames generated at the frequency of 25 fps. It browses a living with a window where a bright outside day can be seen. The light coming from the window has been artificially boosted to have a high but slowly variation of the key value. Its analysis is plotted in Figure A.3 (right).

Temple The *Temple* sequence is composed of 260 Full HD frames generated at the frequency of 25 fps. It represents a camera traveling outdoors in the desert under a pergola, before entering a dim temple. The end of the sequence is rather noisy and allowed us to see the impact of TMO on noise amplification. Its analysis is plotted in Figure A.4 (left).

GeekFlat The *GeekFlat* sequence is composed of 400 Full HD frames generated at the frequency of 25 fps and represents a camera moving in a 3-room flat. Two well illuminated rooms are separated by a dim corridor. This sequence is a good test for brightness coherency as some TMO render the dim corridor brighter than the two rooms. Its analysis is plotted in Figure A.4 (right).

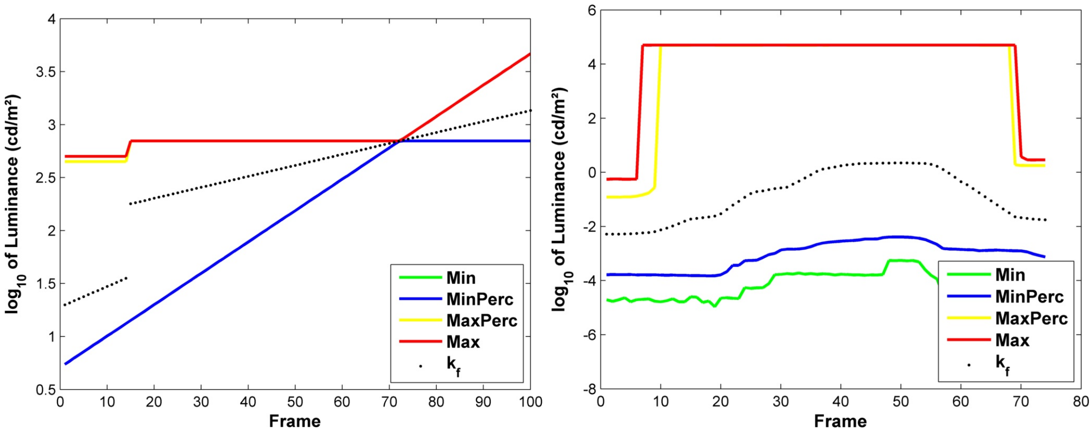


Figure A.3: Analysis of the *HueDisc* (left) and *ModernFlatAnalysis* (right) sequences.

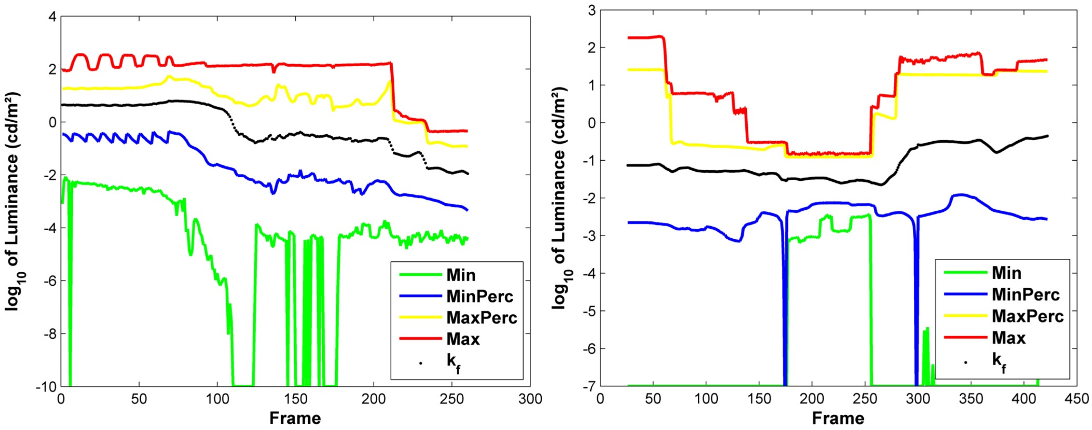


Figure A.4: Analysis of the *Temple* and *GeekFlat* sequences.

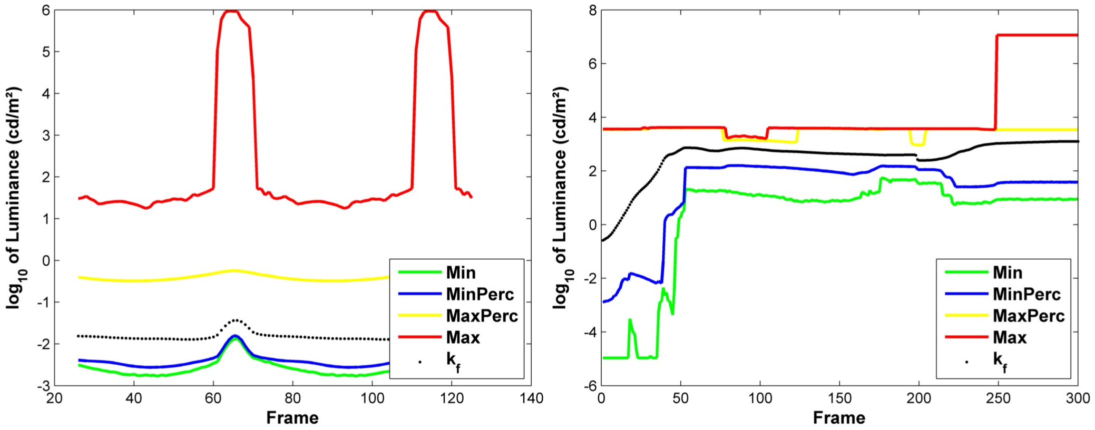


Figure A.5: Analysis of the *LightHouse* (left) and *Desert* (right) sequences.

LightHouse The *LightHouse* sequence is composed of 100 Full HD frames generated at 25 fps. A lighthouse in the fog illuminates the landscape thanks to two rotating lamps. The lamps project directly their light in the camera when moving in front of it. Its analysis is plotted in Figure A.5 (left).

Desert The *Desert* sequence is composed of 250 Full HD frames generated at 25 fps. It simulates a walkthrough from a dim cave to an open field desert. At the end of the video, a spaceship shadows the sun to create a global change of illumination. Its analysis is plotted in Figure A.5 (right).

Disco The *Disco* sequence is composed of 250 Full HD frames generated at 25 fps. This sequence represents a disco ball lighting an underground dance floor with balls flying all over. At the middle of the video, the disco ball is turned off and only small neons illuminate the room. Its analysis is plotted in Figure A.6 (left).

Soccer The *Soccer* sequence is composed of 250 Full HD frames generated at 25 fps. This video takes place in a football stadium where two figures pass a ball before scoring. All of the stadium is in shadows while the other half is under direct sunlight. Its analysis is plotted in Figure A.6 (right).

A.4 Nevex Sequences

TunnelHD The *TunnelHD* sequence is composed of 250 Full HD frames captured at the frequency of 25 fps. A car is entering a tunnel during a sunny day. The change of illumination between the two conditions is slow but at one point we can still see the outside of the tunnel while being completely accommodated to the inside. Its analysis is plotted in Figure A.7 (left).

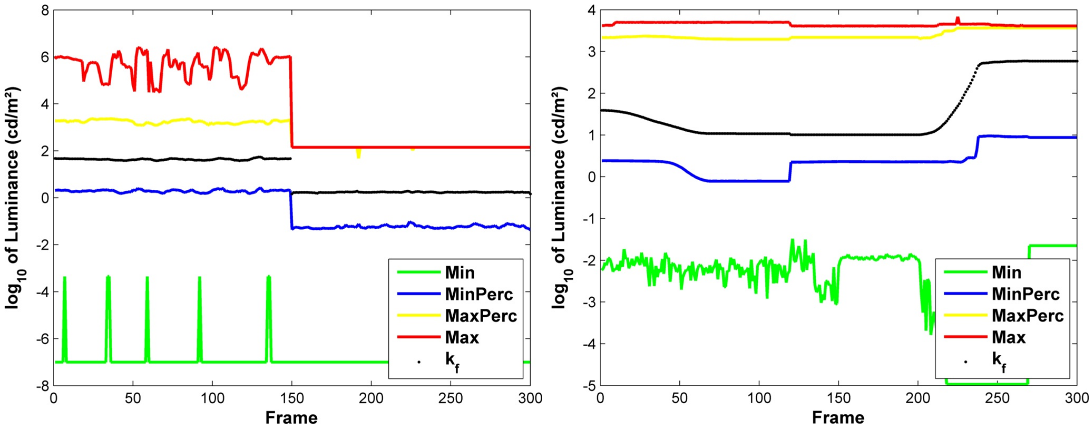


Figure A.6: Analysis of the *Disco* and *GeekFlat* sequences.

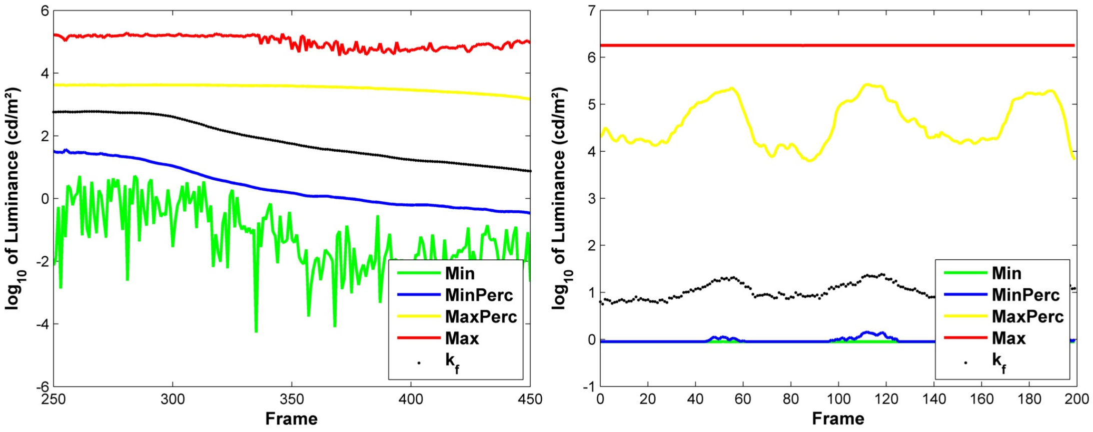


Figure A.7: Analysis of the *TunnelHD* and *FireEater2* sequences.

FireEater2 The *FireEater2* sequence is composed of 200 Full HD frames at the frequency of 25 fps. It records a spectacle at night with the contrast between the spectators that are in the dark and the performers that are blowing fire. Its analysis is plotted in Figure A.7 (right).

A.5 Linköping Sequences

The sequences were captured using an system developed in collaboration between a camera manufacturer (SpheronVR) and the Visual Computing Laboratory at Linköping University in Sweden. The camera is a multi-sensor imaging system that captures HDR images with a resolution of 2336x1752 pixels at up to 30 fps with a dynamic range of over 24 f-stops. The high bandwidth HDR data stream can be viewed in real time through GPU processing, and is written without compression to a file on an external storage unit. A detailed overview of the imaging hardware setups and the image reconstruc-

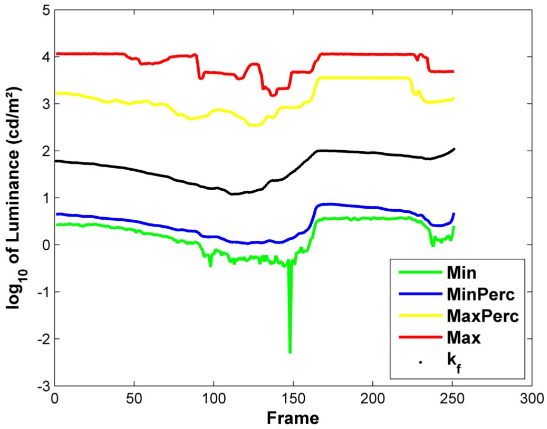


Figure A.8: Analysis of the *Students* sequence.

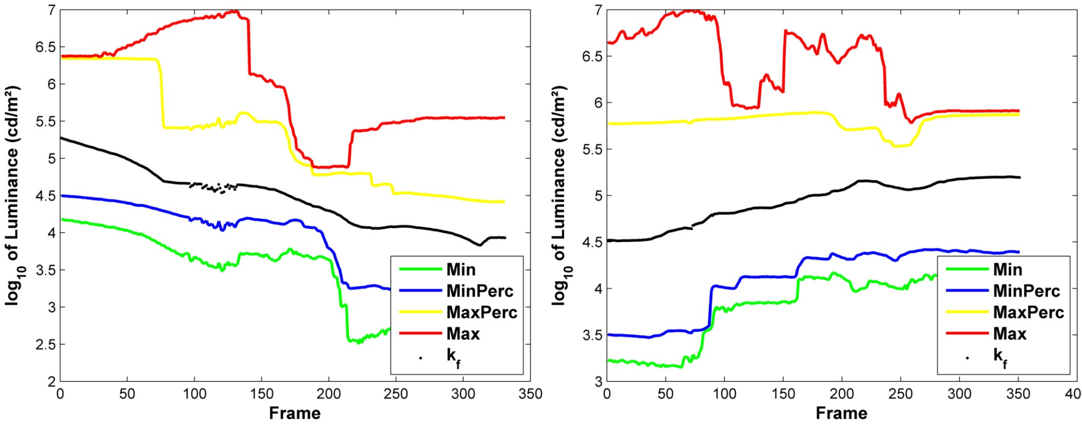


Figure A.9: Analysis of the *Hallway2* (left) and *Hallway3* (right) sequences.

tion algorithms can be found in [Kronander et al., 2013], [Kronander et al., 2014]. The public version of these footages have been down-sampled from the Full HD resolution 7 to HD Ready.

Students The *Students* sequence is composed of 250 HD Ready frames captured at the frequency of 25 fps Several students walking in a dim corridor with a view on the outside through a window on the background. The analysis is plotted in Figure A.8.

Hallway2 and Hallway3 The *Hallway2* sequence is composed of 330 HD Ready frames captured at the frequency of 25 fps. The *Hallway3* sequence is composed of 350 HD Ready frames captured at the frequency of 25 fps. They both illustrate a walk-through in a corridor of the Linköping university. Their analysis are plotted in Figure A.9.

Bibliography

[Adams, 1981] Adams, A. (1981). *The Print: The Ansel Adams Photography Series 3*. Little, Brown and Compagny.

[Akyüz et al., 2007] Akyüz, A. O., Fleming, R., Riecke, B. E., Reinhard, E., and Bülthoff, H. H. (2007). Do HDR displays support LDR content? *ACM Transactions on Graphics*, 26(3):38.

[Alattar, 1997] Alattar, A. (1997). Detecting fade regions in uncompressed video sequences. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 4, pages 3025–3028 vol.4.

[Ashikhmin and Goyal, 2006] Ashikhmin, M. and Goyal, J. (2006). A reality check for tone-mapping operators. *ACM Transactions on Applied Perception*, 3(4):399–411.

[Banterle, 2011] Banterle, F. (2011). Exponential TMO available in the HDR toolbox.

[Banterle et al., 2011] Banterle, F., Artusi, A., Debattista, K., and Chalmers, A. (2011). *Advanced High Dynamic Range Imaging: Theory and Practice.* AK Peters (CRC Press), Natick, MA, USA.

[Banterle et al., 2007] Banterle, F., Ledda, P., Debattista, K., Chalmers, A., and Bloj, M. (2007). A framework for inverse tone mapping. *The Visual Computer*, 23(7):467–478.

[Barlow, 1957] Barlow, H. B. (1957). Purkinje Shift and Retinal Noise. *Nature*, 179(4553):255–256.

[Barten, 1992] Barten, P. G. J. (1992). Physical model for the contrast sensitivity of the human eye. In Rogowitz, B. E., editor, *SPIE 1666, Human Vision, Visual Processing, and Digital Display III,*, pages 57–72.

[Bjontegaard, 2008] Bjontegaard, G. (2008). Improvement of the bd-psnr model. ITU-T SG16, VCEG-AI11.

[Blackwell, 1981] Blackwell, H. (1981). An analytical model for describing the influence of lighting parameters upon visual performance, volume 1: Technical foundations. *Comission Internationale De L'Eclairage*, 11.

[Brailean et al., 1995] Brailean, J., Kleihorst, R., Efstratiadis, S., Katsaggelos, A., and Lagendijk, R. (1995). Noise reduction filters for dynamic image sequences: a review. *Proceedings of the IEEE*, 83(9):1272–1292.

[Bross et al., 2013] Bross, B., Han, W.-J., Ohm, J.-R., Sullivan, G. J., Wang, Y.-K., and Wiegand, T. (2013). High efficiency video coding (hevc) text specification draft 10 (for fdis & consent). ISO/IEC JTC1/SC29/WG11 JCTVC-L1003.

[Burt, 1981] Burt, P. J. (1981). Fast filter transform for image processing. *Computer Graphics and Image Processing*, 16(1):20 – 51.

[Burt and Adelson, 1983] Burt, P. J. and Adelson, E. H. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540.

[Chiu et al., 1993] Chiu, K., Herf, M., Shirley, P., Swamy, S., Wang, C., Zimmerman, K., et al. (1993). Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface*, pages 245–245. Canadian Information Processing Society.

[Choudhury and Tumblin, 2005] Choudhury, P. and Tumblin, J. (2005). The trilateral filter for high contrast images and meshes. In *ACM SIGGRAPH 2005 Courses on - SIGGRAPH '05*, page 5, New York, New York, USA. ACM Press.

[Crawford, 1949] Crawford, B. H. (1949). The Scotopic Visibility Function. *Proceedings of the Physical Society. Section B*, 62(5):321–334.

[Debevec and Malik, 1997] Debevec, P. E. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97*, pages 369–378, New York, New York, USA. ACM Press.

[Drago et al., 2003a] Drago, F., Martens, W. L., Myszkowski, K., and Seidel, H.-P. (2003a). Perceptual evaluation of tone mapping operators. In *Proceedings of the SIGGRAPH 2003 conference on Sketches & applications in conjunction with the 30th annual conference on Computer graphics and interactive techniques - GRAPH '03*, page 1, New York, New York, USA. ACM Press.

[Drago et al., 2003b] Drago, F., Myszkowski, K., Annen, T., and Chiba, N. (2003b). Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Computer Graphics Forum*, 22(3):419–426.

[Durand and Dorsey, 2000] Durand, F. and Dorsey, J. (2000). Interactive tone mapping. In *Proceedings of the Eurographics Workshop on Rendering*. Springer Verlag.

[Durand and Dorsey, 2002] Durand, F. and Dorsey, J. (2002). Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257.

[Eilertsen et al., 2013] Eilertsen, G., Wanat, R., Mantiuk, R. K., and Unger, J. (2013). Evaluation of Tone Mapping Operators for HDR-Video. *Computer Graphics Forum*, 32(7):275–284.

[Fairchild, 2004] Fairchild, M. D. (2004). iCAM framework for image appearance, differences, and quality. *Journal of Electronic Imaging*, 13(1):126.

[Fairchild, 2005] Fairchild, M. D. (2005). *Color Appearance Models*. Wiley-IS&T, Chichester, UK.

[Farbman et al., 2008] Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM SIGGRAPH 2008 papers on - SIGGRAPH '08*, page 1, New York, New York, USA. ACM Press.

[Farbman and Lischinski, 2011] Farbman, Z. and Lischinski, D. (2011). Tonal stabilization of video. *ACM Transactions on Graphics*, 30(4):1.

[Fattal et al., 2002] Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3).

[Ferwerda, 2001] Ferwerda, J. (2001). Elements of early vision for computer graphics. *IEEE Computer Graphics and Applications*, 21(4):22–33.

[Ferwerda et al., 1996] Ferwerda, J. A., Pattanaik, S. N., Shirley, P., and Greenberg, D. P. (1996). A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96*, pages 249–258, New York, New York, USA. ACM Press.

[Gastal and Oliveira, 2011] Gastal, E. S. L. and Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics*, 30(4):1.

[Grzegorz Krawczyk, 2007] Grzegorz Krawczyk, R. M. (2007). Display adaptive pfstmo documentation.

[Guthier et al., 2011] Guthier, B., Kopf, S., Eble, M., and Effelsberg, W. (2011). Flicker reduction in tone mapped high dynamic range video. In *Proc. of IS&T/SPIE Electronic Imaging (EI) on Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications.*

[He et al., 2013] He, K., Sun, J., and Tang, X. (2013). Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1397–409.

[Hood et al., 1986] Hood, D., Finkelstein, M., Boff, K., Kaufman, L., and Thomas, J. (1986). Handbook of perception and human performance. *Sensory Processes and Perception, chapter Sensitivity to Light. Wiley.*

[Hough, 1981] Hough, D. (1981). Applications of the proposed ieee 754 standard for floating-point arithetic. *Computer*, 14(3):70–74.

[Hunt, 2005] Hunt, R. (2005). *The Reproduction of Colour*. The Wiley-IS&T Series in Imaging Science and Technology. John Wiley & Sons.

[Irawan et al., 2005] Irawan, P., Ferwerda, J. A., and Marschner, S. R. (2005). Perceptually Based Tone Mapping of High Dynamic Range Image Streams. *Image (Rochester, N.Y.)*.

[ITU, 1998] ITU (1998). Recommendation ITU-R BT.709-3: Parameter values for the HDTV standards for production and international programme exchange. International Telecommunications Union.

[ITU, 2012] ITU (2012). Recommendation ITU-R BT.2020: Parameter values for ultra-high definition television systems for production and international programme exchange. International Telecommunications Union.

[Jakob, 2010] Jakob, W. (2010). Mitsuba renderer. http://www.mitsuba-renderer.org.

[Kang et al., 2003] Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2003). High dynamic range video. *ACM Trans. Graph.*, 22(3):319–325.

[Kiser et al., 2012] Kiser, C., Reinhard, E., Tocci, M., and Tocci, N. (2012). Real-time Automated Tone Mapping System for HDR Video. In *Proceedings of the IEEE International Conference on Image Processing*, pages 2749–2752.

[Koz and Dufaux, 2012] Koz, A. and Dufaux, F. (2012). A comparative survey on high dynamic range video compression. In Tescher, A. G., editor, *SPIE 8499, Applications of Digital Image Processing XXXV*, page 84990E.

[Krawczyk et al., 2004] Krawczyk, G., Myszkowski, K., and Seidel, H.-p. (2004). Lightness Perception in Tone Reproduction for High Dynamic Range Images. *Computer Graphics Forum (Proc. of Eurographics)*, 24(3):635–645.

[Kronander et al., 2013] Kronander, J., Gustavson, S., Bonnet, G., and Unger, J. (2013). Unified HDR reconstruction from raw CFA data. *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9.

[Kronander et al., 2014] Kronander, J., Gustavson, S., Bonnet, G., Ynnerman, A., and Unger, J. (2014). A unified framework for multi-sensor HDR video reconstruction. *Signal Processing: Image Communication*, 29(2):203–215.

[Kuang et al., 2010] Kuang, J., Heckaman, R., and Fairchild, M. D. (2010). Evaluation of HDR tone-mapping algorithms using a high-dynamic-range display to emulate real scenes. *Journal of the Society for Information Display*, 18(7):461.

[Kuang et al., 2007a] Kuang, J., Johnson, G. M., and Fairchild, M. D. (2007a). iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414.

[Kuang et al., 2007b] Kuang, J., Yamaguchi, H., Liu, C., Johnson, G. M., and Fairchild, M. D. (2007b). Evaluating HDR rendering algorithms. *ACM Transactions on Applied Perception*, 4(2):9–es.

[Kunkel et al., 2013] Kunkel, T., Ward, G., Lee, B., and Daly, S. (2013). HDR and wide gamut appearance-based color encoding and its quantification. In *2013 Picture Coding Symposium (PCS)*, pages 357–360, San Jose. IEEE.

[Ledda et al., 2005] Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H. (2005). Evaluation of tone mapping operators using a high dynamic range display. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 640–648, New York, NY, USA. ACM.

[Ledda et al., 2004] Ledda, P., Santos, L. P., and Chalmers, A. (2004). A local model of eye adaptation for high dynamic range images. In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa - AFRIGRAPH '04*, page 151, New York, New York, USA. ACM Press.

[Lee and Kim, 2007] Lee, C. and Kim, C.-S. (2007). Gradient domain tone mapping of high dynamic range videos. In *2007 IEEE International Conference on Image Processing*, volume 3, pages III – 461. IEEE.

[Lee et al., 2011] Lee, J.-S., De Simone, F., and Ebrahimi, T. (2011). Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding. *IEEE Transactions on Multimedia*, 13(5):882–893.

[Li et al., 2005] Li, Y., Sharan, L., and Adelson, E. H. (2005). Compressing and companding high dynamic range images with subband architectures. *ACM Transactions on Graphics*, 24(3):836.

[Luthra et al., 2014] Luthra, A., Francois, E., and Husak, W. (2014). Draft Requirements and Explorations for HDR / WCG Content Distribution and Storage. In *ISO/IEC JTC1/SC29/WG11 MPEG2014/N14510*, Valencia, Spain. IEEE.

[Magic, 2008] Magic, I. L. . (2008). Openexr.

[Mai et al., 2011] Mai, Z., Mansour, H., Mantiuk, R., Nasiopoulos, P., Ward, R., and Heidrich, W. (2011). Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE Transactions on Image Processing (TIP)*, 20(6):1558–1571.

[Mann and Picard, 1994] Mann, S. and Picard, R. (1994). Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. Technical Report 3223, M.I.T. Media Lad Perceptual Computing Section, Boston, Massachusetts. Also appears, IS&T's 48th Annual conference, Cambridge, Massachusetts, May, 1995.

[Mantiuk et al., 2008] Mantiuk, R., Daly, S., and Kerofsky, L. (2008). Display adaptive tone mapping. *ACM Transactions on Graphics*, 27(3):1.

[Mantiuk et al., 2006a] Mantiuk, R., Efremov, A., Myszkowski, K., and Seidel, H.-P. (2006a). Backward compatible high dynamic range MPEG video compression. *ACM Transactions on Graphics*, 25(3):713.

[Mantiuk et al., 2004] Mantiuk, R., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2004). Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics*, 23(3):733.

[Mantiuk et al., 2006b] Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2006b). Lossy compression of high dynamic range images and video. In *Proc. of Human Vision and Electronic Imaging XI*, volume 6057 of *Proceedings of SPIE*, page 60570V, San Jose, USA. SPIE.

[Mantiuk et al., 2012] Mantiuk, R. K., Tomaszewska, A., and Mantiuk, R. (2012). Comparison of Four Subjective Methods for Image Quality Assessment. *Computer Graphics Forum*, 31(8):2478–2491.

[Masia et al., 2009] Masia, B., Agustin, S., Fleming, R. W., Sorkine, O., and Gutierrez, D. (2009). Evaluation of reverse tone mapping through varying exposure conditions. *ACM Transactions on Graphics*, 28(5):1.

[McCann and Rizzi, 2011] McCann, J. J. and Rizzi, A. (2011). *The art and science of HDR imaging.* Sons, John Wiley.

[Melo et al., 2013] Melo, M., Bessa, M., Debattista, K., and Chalmers, A. (2013). Evaluation of HDR video tone mapping for mobile devices. *Signal Processing: Image Communication*, pages 1–10.

[Michaelis and Menten, 1913] Michaelis, L. and Menten, M. L. (1913). Die kinetik der invertinwirkung. *Biochem. z*, 49(333-369):352.

[Miller et al., 2013] Miller, S., Nezamabadi, M., and Daly, S. (2013). Perceptual Signal Coding for More Efficient Usage of Bit Codes. *SMPTE Motion Imaging Journal*, 122(4):52–59.

[Mitsunaga and Nayar, 1999] Mitsunaga, T. and Nayar, S. (1999). Radiometric self calibration. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pages 374–380. IEEE Comput. Soc.

[Mote, 1951] Mote, F. A. (1951). The Effect of Varying the Intensity and the Duration of Preexposure upon Foveal Dark Adaptation in the Human Eye. *The Journal of General Physiology*, 34(5):657–674.

[Motra and Thoma, 2010] Motra, A. and Thoma, H. (2010). An adaptive Logluv transform for High Dynamic Range video compression. In *2010 IEEE International Conference on Image Processing*, pages 2061–2064. IEEE.

[Myszkowski et al., 2008] Myszkowski, K., Mantiuk, R., and Krawczyk, G. (2008). *High Dynamic Range Video*, volume 2. Morgan & Claypool.

[Naka and Rushton, 1966] Naka, K. and Rushton, W. (1966). An attempt to analyse colour reception by electrophysiology. *The Journal of physiology*, 185:556–586.

[Pattanaik et al., 1998] Pattanaik, S. N., Ferwerda, J. A., Fairchild, M. D., and Greenberg, D. P. (1998). A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*, pages 287–298, New York, New York, USA. ACM Press.

[Pattanaik et al., 2000] Pattanaik, S. N., Tumblin, J., Yee, H., and Greenberg, D. P. (2000). Time-dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 47–54, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

[Petit and Mantiuk, 2013] Petit, J. and Mantiuk, R. K. (2013). Assessment of video tone-mapping : Are cameras ' S-shaped tone-curves good enough? *Journal of Visual Communication and Image Representation*, 24:1020–1030.

[Pouli et al., 2013] Pouli, T., Artusi, A., and Banterle, F. (2013). Color Correction for Tone Reproduction. In *Color and Imaging Conference Final Program and Proceedings*, pages 215–220.

[Poynton, 1996] Poynton, C. A. (1996). *A Technical Introduction to Digital Video.* Sons, John Wiley &, New York, NY, USA.

[Ramsey et al., 2004] Ramsey, S. D., Johnson III, J. T., and Hansen, C. (2004). Adaptive temporal tone mapping. In *Computer Graphics and Imaging - 2004*, pages 3–7.

[Reinhard et al., 2010] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K. (2010). *High Dynamic Range Imaging, 2nd Edition: Acquisition, Display, and Image-Based Lighting.* Morgan Kaufmann.

[Reinhard et al., 2012] Reinhard, E., Pouli, T., Kunkel, T., Long, B., Ballestad, A., and Damberg, G. (2012). Calibrated image appearance reproduction. *ACM Transactions on Graphics*, 31(6):1.

[Reinhard et al., 2002] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267–276.

[Schoberl et al., 2012] Schoberl, M., Belz, A., Seiler, J., Foessel, S., and Kaup, A. (2012). High dynamic range video by spatially non-regular optical filtering. In Sampat, N. and Battiato, S., editors, *2012 19th IEEE International Conference on Image Processing*, pages 2757–2760. IEEE.

[Seetzen et al., 2004] Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L. A., Trentacoste, M., Ghosh, A., and Vorozcovs, A. (2004). High dynamic range display systems. *ACM Transactions on Graphics*, 23(3):760.

[Seetzen et al., 2003] Seetzen, H., Whitehead, L. A., and Ward, G. (2003). A High Dynamic Range Display Using Low and High Resolution Modulators. *SID Symposium Digest of Technical Papers*, 34(1):1450.

[Smith and Guild, 1931] Smith, T. and Guild, J. (1931). The c.i.e. colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73.

[SMPTE, 2014] SMPTE (2014). YDzDx Color-Difference Encoding for XYZ Signals. Society of Motion Picture and Television Engineers SMPTE ST 2085.

[Stevens and Stevens, 1963] Stevens, J. C. and Stevens, S. S. (1963). Brightness Function : Effects of Adaptation. *Journal of the Optical Society of America*, 53(3):375.

[Sullivan et al., 2012] Sullivan, G. J., Ohm, J.-R., Han, W.-J., and Wiegand, T. (2012). Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668.

[Tocci et al., 2011] Tocci, M. D., Kiser, C., Tocci, N., and Sen, P. (2011). A versatile HDR video production system. *ACM Transactions on Graphics*, 30(4):1.

[Tomasi and Manduchi, 1998] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846. Narosa Publishing House.

[Touzé et al., 2014] Touzé, D., Lasserre, S., Olivier, Y., Boitard, R., and François, E. (2014). HDR Video Coding based on Local LDR Quantization. In *HDRi2014 - Second International Conference and SME Workshop on HDR imaging (2014),*.

[Tumblin, 1999] Tumblin, J. (1999). LCIS: A boundary hierarchy for detail-preserving contrast reduction. *Proceedings of the 26th annual conference on.*

[Tumblin and Rushmeier, 1993] Tumblin, J. and Rushmeier, H. (1993). Tone reproduction for realistic images. *Computer Graphics and Applications IEEE.*

[Van Hateren, 2006] Van Hateren, J. H. (2006). Encoding of high dynamic range video with a model of human cones. *ACM Transactions on Graphics*, 25(4):1380–1399.

[Čadík et al., 2008] Čadík, M., Wimmer, M., Neumann, L., and Artusi, A. (2008). Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics*, 32(3):330–349.

[Vos, 1978] Vos, J. J. (1978). Colorimetric and photometric properties of a 2 fundamental observer. *Color Research & Application*, 3(3):125–128.

[Wanat et al., 2012] Wanat, R., Petit, J., and Mantiuk, R. (2012). Physical and Perceptual Limitations of a Projector-based High Dynamic Range Display. In Carr, Hamish and Czanner, S., editor, *TPCG*.

[Ward, 1991] Ward, G. (1991). Real pixels. In *Graphics Gems*, volume 2, pages 15–31. Morgan Kaufmann.

[Ward, 1994a] Ward, G. (1994a). A contrast-based scalefactor for luminance display. *Graphics gems IV*.

[Ward, 1994b] Ward, G. J. (1994b). The radiance lighting simulation and rendering system. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '94, pages 459–472, New York, NY, USA. ACM.

[Ward Larson, 1998] Ward Larson, G. (1998). LogLuv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools*, 3(1):815–30.

[Wiegand et al., 2003] Wiegand, T., Sullivan, G., Bjontegaard, G., and Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560–576.

[Xu et al., 2011] Xu, D., Doutre, C., and Nasiopoulos, P. (2011). Correction of clipped pixels in color images. *IEEE transactions on visualization and computer graphics*, 17(3):333–44.

[Yang and Wu, 2012] Yang, L. and Wu, D. (2012). Adaptive quantization using piecewise companding and scaling for Gaussian mixture. *J. Vis. Comun. Image Represent.*, 23(7):959–971.

[Yoshida, 2005] Yoshida, A. (2005). Perceptual evaluation of tone mapping operators with real-world scenes. In *Proceedings of SPIE*, volume 5666, pages 192–203. SPIE.

[Yoshida et al., 2006] Yoshida, A., Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2006). Analysis of Reproducing Real-World Appearance on Displays of Varying Dynamic Range. *Computer Graphics Forum*, 25(3):415–426.