Tone Mapping based on Natural Image Statistics and Visual Perception Models

Praveen Cyriac

DOCTORAL THESIS UPF / YEAR 2017

Thesis Director:
Marcelo Bertalmío
Department of Information and Communication Technologies

To my family

Acknowledgements

First of all, I would like to thank my supervisor Marcelo Bertalmío for his continuous support and encouragement throughout my graduate studies. Without his help and guidance this thesis would not have been possible.

I would like to thank all the co-authors of my publications, without them a substantial portion of this thesis would not have been possible: Thomas Batard, David Kane, Javier Vazquez Corral, Jesus Malo, Marina Martinez and Adrian Galdran.

Thanks to all my colleagues and friends for their help, support, and for the great time we spent together: Syed Waqas Zamir, Gabriela Ghimpeteanu, Raquel Gil Rodriguez, Jihyun Kim, Sadegh Mohammadi, Antoine Grimaldi, Jose Maria Rojano, Rashid Zulqarnain, Kamruddin Nur, Raul Parada and all my malayali friends in Barcelona.

Many thanks to Lydia Garcia, Jana Safrankova, Joana Soria, Vanessa Jimenez, Joana Clotet, Magda Castellnou, Marcel Xandri and Lluis Bosch for always being ready to help me in administrative bureaucracy, your help made my life lot more easier.

Last but not the least, I would like to thank my family, my mother Saly, my father Cyriac, my brother Anish and his wife Thushara, and my sister Aiswarya for their love, support and encouragement throughout writing this thesis.

This work has been supported by the Department of Information and Communication Technologies (DTIC) PhD fellowship (2013-16), Universitat Pompeu Fabra and the European Research Council, Starting Grant ref. 306337.

Abstract

High Dynamic Range (HDR) imaging techniques potentially allow for the capture and storage of the full information of light in a scene. However, common display devices are limited in terms of their contrast and brightness capabilities, thus HDR images must be tone mapped before presentation on a display device to ensure that the original appearance of the scene is reproduced.
In this thesis, we take two approaches to the tone mapping problem. First, we develop a general framework for improving any tone mapped image by reducing the distance with the corresponding HDR image in terms of a non-local perceptual metric. The distance is minimized by means of a gradient descent algorithm. Second, we develop a real-time Tone Mapping Operator (TMO) that is well suited to the statistics of natural scenes, and is in keeping with new psychophysical findings and neurophysical data. We determine the adequate non-linear adjustments needed for our tone mapping results to look best in different viewing conditions through a psychophysical experiment and develop an automatic method that can predict the experimental data. Our TMO produces results that look natural, without any spatio-temporal artifacts. User preference tests show that our method outperforms state of the art approaches. The TMO is fast and could be implemented on camera hardware. It can be used for on-set monitoring of HDR cameras on regular displays, as a substitute for gamma correction, and as a way of providing the colorist with content that is both natural looking and has a crisp and clear appearance.

Resum

Les tècniques d'imatge d'alt rang dinàmic (HDR) potencialment permeten la captura i l'emmagatzematge de tota la informació de llum en una escena. No obstant això, els dispositius comuns de visualització són limitats en termes de les seves capacitats de contrast i brillantor, per tant, les imatges HDR han de ser mapejades tonalment abans de presentar-les en un dispositiu de visualització per assegurar que es reprodueix l'aspecte original de l'escena. En aquesta tesi, es prenen dos enfocaments del problema de mapeig tonal. En primer lloc, es desenvolupa un marc general per a la millora de qualsevol imatge mapejada tonalment mitjançant la reducció de la distància a la corresponent imatge HDR en termes d'una mètrica perceptiva no local. La distància es redueix al mínim per mitjà d'un algoritme de descens de gradient. En segon lloc, es desenvolupa un operador de mapeig tonal (TMO) en temps real que s'adapta bé a les estadístiques d'escenes naturals, i concorda amb els nous descobriments psicofísics i dades neurofísiques. Determinem les correctes adaptacions no lineals necessàries per als nostres resultats de mapeig tonal per tal d'obtenir l'aparença òptima en diferents condicions de visualització, a través d'experiments psicofísics i desenvolupar un mètode automàtic per poder predir dades experimentals. El nostre TMO produeix resultats d'aspecte natural, sense cap tipus d'artefactes espacials o temporals. Els tests de preferència dels usuaris mostren que el nostre mètode obté millors resultats en comparació amb les tècniques més recents. El TMO és ràpid i podria ser implementat en el hardware de la càmera. Pot ser utilitzat per al monitoratge de càmeres HDR en pantalles regulars, com a substitut de la correcció gamma, i com una manera de proporcionar al colorista amb contingut que té alhora un aspecte natural i una aparença nítida i clara.

Contents

Glossary

CCD	charge-coupled device.
CFA	color filter array.
CGI	computer-generated imagery.
CMOS	complementary metal-oxide- semiconductor.
CSF	contrast sensitivity function.
DCT	discrete cosine transform.
DLP	digital light processing.
DMD	digital micro-mirror device.
DR	dynamic range.
DRIM	dynamic range independent metric.
EOTF	electro-optical transfer function.
HDR	high dynamic range.
HLG	hybrid log gamma.
HVS	human visual system.
JND	just noticeable difference.
JPEG	joint photographic experts group.
LCD	liquid crystal display.
LCoS	liquid crystal on silicon.
LDR	low dynamic range.
LED	light emitting diode.
LGN	lateral geniculate nucleus.

LUT	look up table.
MAD	maximum differentiation.
MOS	mean opinion score.
MSE	mean square error.
OETF	opto-electrical transfer function.
OOTF	opto-optical transfer function.
PQ	perceptual quantizer.
PSF	point spread function.
PSNR	peak signal-to-noise ratio.
QD	quantum dot.
SDR	standard dynamic range.
TMO	tone mapping operator.
TMQI	tone mapping quality index.

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Light intensities in the real world span from around $10^{-6}cd/m^2$ for starlight to $10^9cd/m^2$ for direct sunlight covering a large range. Moreover, the dynamic range (ratio between the darkest and brightest values) within a single scene can easily go above four orders of magnitude. Even though the simultaneous Dynamic Range (DR) of the Human Visual System (HVS) is only about three to four orders of magnitude, nonetheless it is capable of handling intensities from that vast range [Pattanaik et al., 2000], due to the fact that it has multiple adaptation mechanisms that continuously adjust according to the light in any viewed scene.

Traditional imaging technologies fail in capturing and storing all the details and contrast that we perceive with the naked eye. The problem comes from the limited dynamic range captured by cameras and stored by standard formats, see Figure 1.1. With a short exposure time setting a camera captures well the details in the brightest parts of a high dynamic range scene, but fails to render details in the dark regions. On the other hand, with a long exposure time setting we can capture information of the dark regions but at the cost of clipping the bright intensity values. And there is no single exposure value that allows to properly capture the entire scene.

High Dynamic Range (HDR) imaging techniques overcome this limitation by capturing and storing potentially the full range of light information of a scene. The capturing techniques include: using HDR-capable imaging sensors or us-

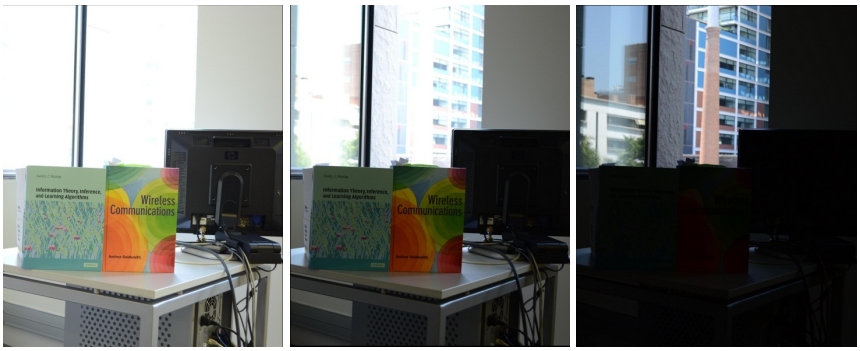


Figure 1.1: Camera output with decreasing exposure time from left to right.

ing standard cameras and capturing multiple exposures of the same scene and then merging them [Debevec and Malik, 1997] [Gil Rodríguez et al., 2015]. The captured values are generally stored in floating point numbers, providing much higher accuracy that even exceeds the capabilities of the HVS [Reinhard et al., 2010]. The RAW camera sensor data, even though it may not capture the entire dynamic range of a scene, provides a wider dynamic range and higher bit precision (12-14 bits) when compared to the final default output of a camera.

Common display devices are unable to reproduce the range of luminances found in the real world, their dynamic range is only about two orders of magnitude. Even though currently available HDR displays provide a significant improvement in this direction, their capabilities are still far from matching the real world appearance. Therefore, when HDR or RAW images are linearly represented on display devices the details and contrasts are not accurately reproduced (see Figure 1.2), so an operation called tone mapping is performed on the image to scale down the dynamic range to match the display capability. A successful tone mapping algorithm should produce in anyone watching the display a perception of details and color that matches as much as possible to the perception of the real world scene [Larson et al., 1997].

In digital cameras, tone mapping is achieved by applying automatic non-linear transformations to the RAW sensor data. The non-linearity applied in most digital cameras is well approximated by a simple power law, and is called gamma correction. While this may perform well on average, in general when dealing

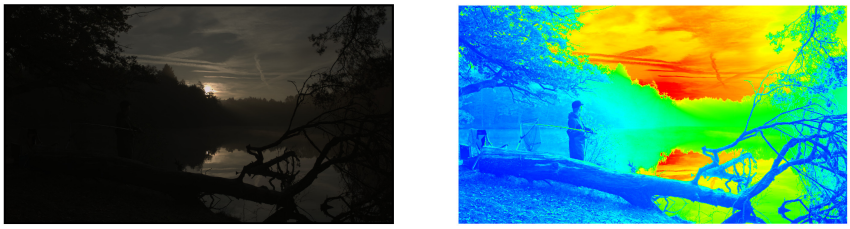


Figure 1.2: Linear representation of an HDR image (left), false color image of the HDR image (right).

Figure 1.3: Comparison between default camera output (left) and manual grading (right). Image courtesy Harald Brendel, ARRI.

with HDR scenes it is sub-optimal [Petit and Mantiuk, 2013]. In professional filmmaking, a skilled technician (colorist) non-linearly corrects the HDR image or the RAW sensor data in post-production to optimize image appearance for a specific viewing condition. Figure 1.3 compares the default camera output produced by a simple gamma correction with that of the result of manual grading. Even though both operations receive the same image content as input, the default camera operation over-saturates the outdoor regions resulting in loss of details, whereas the manual grading preserves well the overall appearance of the scene.

The literature provides a number of image-dependent automatic and semi-automatic non-linearities, some based on models of the HVS, see [Ferradans et al., 2011] and references therein. Additionally, there are more complex, local tone mapping algorithms that could, in theory, transform the sensor data, but these tend to be computationally more expensive, sensitive to image fluctuations, and harder to tune for a specific application. As a result, manual grading is always preferred

in professional filmmaking even though it is a time-consuming process.

The metrics that compare an HDR image with its tone mapped version allow us to estimate the degree of distortion introduced by tone mapping operators (TMOs). Figure 1.4 shows the results of different TMOs with their pixel-wise distortion map computed with the perceptually-based metric of [Aydin et al., 2008]. The color red indicates a reversal of contrast, green is for loss of visible contrast and blue represents amplification of invisible contrast. The distortion maps in this example clearly indicate that the TMOs introduce a certain amount of distortion. The quality of the tone mapped images can be improved by minimizing the detected distortions.

The advancement in display technologies in the past decade resulted in a whole new variety of displays in the market. They greatly vary in terms of contrast and brightness. For example, in the low end there are LCD displays that can produce a peak brightness of 100 cd/m^2 with a contrast ratio of 100:1, and in the high end there are currently available HDR displays that can produce a peak brightness of more than 5000 cd/m^2 with a contrast ratio greater than 5000:1. Also, with the arrival of portable display devices, the surround and the viewing environment in which the display is watched greatly vary. For instance, a tablet can be viewed in a sun-lit outdoor environment or in a dark room. This variation in the display capabilities and viewing conditions significantly affects the perceived image quality. Therefore, for an optimal visual experience, the input image content should be transformed by taking into account the viewing condition and display capabilities.

The main aim of this thesis is to develop a real-time tone mapping algorithm that ensures that the displayed content appears realistic and appealing to the viewers in terms of detail and contrast visibility. In addition, the algorithm should be of low complexity so that it can easily be implemented with limited hardware resources, for example in a camera. To achieve this aim, we look into neural and psychophysical models of the visual system, natural image statistics, and also the effect of display capabilities and the viewing environment on perception.

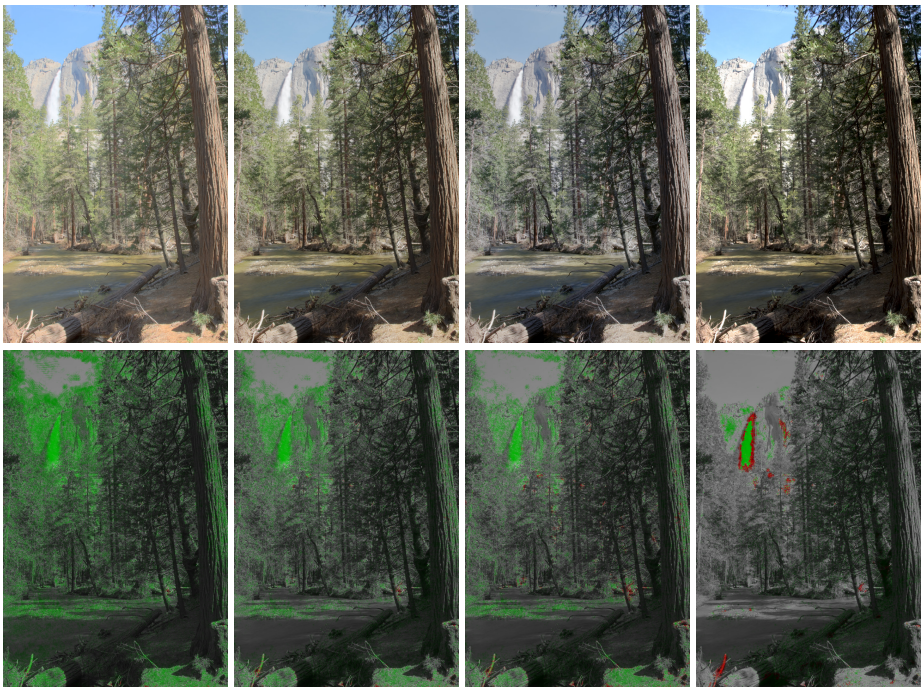


Figure 1.4: Tone Mapping results (top row) and distortion maps (bottom row) for different operators, from left to right: [Drago et al., 2003a], [Reinhard and Devlin, 2005], [Ferradans et al., 2011], [Mantiuk et al., 2008].

1.1 Contributions

The main contributions of this thesis are listed below.

- Optimizing tone mapped images. The evaluation with the perceptually-based metric [Aydin et al., 2008] showed that even the best-performing TMOs introduce a certain amount of distortion. Considering this problem, we develop a general framework for optimizing tone mapped images with respect to a non-local perceptual metric. The approach reduces the distance between the HDR image and its tone mapped version by means of a gradient descent algorithm. We test our method with the metric of [Aydin et al., 2008] for different tone mapped images provided by several TMOs. Results show that our method improves the tone mapped images tested by

substantially reducing their perceptual distance with the HDR source image. On average, an improvement of more that 25% is obtained for several state of the art TMOs.

- We develop a TMO based on visual perception models that is well suited to the statistics of natural scenes, and is in keeping with new psychophysical findings and neurophysical data. The method finds, for a given image, a transform which ensures that the output values are more evenly distributed over the available range. The TMO consists of two stages, a global adaptive non-linear transform followed by a local contrast enhancement step. We optimize the parameters of the first stage through psychophysical experiments in an office surround setup. The method produces results that look natural, without any spatio-temporal artifacts. Psychophysical tests based on users' preference confirm that our approach outperforms other state of the art algorithms. In addition, the method has low complexity and can potentially be executed in real time.

- Tone mapping dependent on viewing condition. The viewing conditions and the display's contrast and brightness capabilities can significantly affect the perceived image quality. As a result, the image that appears optimal in one display and surround setup may not look best in some other setup. We conduct a psychophysical experiment to determine the adequate non-linear adjustments needed for the images to look best in different viewing conditions. We also develop a mathematical model to predict the experimental data from some luminance measurements of the display. The non-linear adjustment predicted by the model allows us to automatically map the results produced by our TMO on to any given display by taking into account the viewing environment.

The application of the contributions covers the following areas:

- the proposed TMO can be implemented in-camera as a substitute for gamma correction.

- our TMO can provide the colorist with a fast and automatic graded content that is both natural looking and has a crisp and clear appearance.

- the developed TMO can also be used as an off-line tone mapping method for converting HDR images into LDR ones, with applications to cinema shoots (on-set use of LDR monitors with an HDR camera), television broadcast (making HDR signals compatible with LDR equivalents), and rendering in computer graphics (for video games, 3D animation, the integration of CGI onto real footage, etc.)

- the model developed for tone mapping can be used for dehazing application. The results compare well with that of the state of the art approaches in the case of standard images, whereas our results out-perform the state of the art in the case of challenging images by keeping the artifacts to minimum.

- the developed tone mapping model is used as the first layer of a four-layer vision model, with application to image quality assessment. The results show that the model correlates well with the psychophysical data on subjective distance.

1.2 Chapter Overview

This thesis is organized as follows: Chapter 2 discusses some fundamentals of human visual perception which are relevant to the topics in this thesis. Chapter 3 summarizes the characteristics of digital cameras and describes various techniques involved in HDR image capture and storage. In Chapter 4 we describe various display technologies, explain the tone mapping literature and look into different tone mapping evaluation approaches. We present a general framework to optimize any tone mapped image by taking into account an evaluation metric in Chapter 5. In Chapter 6 we explain our proposed TMO that is based on visual perception models, natural image statistics and psychophysical data. We detail the proposed viewing condition dependent mapping that optimizes the perceived image quality of the tone mapped image in each viewing condition in Chapter 7. In Chapter 8 we show some applications of the proposed model. We conclude the thesis and suggest some future works in Chapter 9.

1.3 Publications and Patent

Journals

- P. Cyriac, T. Batard, M. Bertalmío, A non local variational formulation for the improvement of tone mapped images, SIAM Journal on Imaging Sciences (SIIMS) 2014.
- J. Vazquez-Corral, A. Galdran, P. Cyriac, M. Bertalmío, A fast image dehazing method that does not introduce color artifacts, Transactions on Image Processing (TIP) 2016 (in review).

Conferences

- P. Cyriac, D. Kane, M. Bertalmío, Automatic, viewing-condition dependent contrast grading based on perceptual models, SMPTE Annual Technical Conference 2016.
- P. Cyriac, D. Kane, M. Bertalmío, Optimized tone curve for in-camera image processing, IS&T Electronic Imaging 2016. (Best Student Paper Award)
- P. Cyriac, D. Kane, M. Bertalmío, Perceptual dynamic range for in-camera image processing, British Machine Vision Conference (BMVC) 2015.
- P. Cyriac, M. Bertalmío, D. Kane, J. Vazquez-Corral. A tone mapping operator based on neural and psychophysical models of visual perception. In SPIE/IS&T 2015 Human vision and electronic imaging XX., 2015.
- P. Cyriac, T. Batard, M. Bertalmío, A variational method for the optimization of tone mapping operators, Pacific-Rim Symposium for Image and Video Technology (PSIVT), pages 505-516, 2013.

Patent

- M. Bertalmío, D. Kane, P. Cyriac, Method, system and device for generating more natural images. European patent. Application no. 15154172.9-1906.

Chapter 2

HUMAN VISUAL SYSTEM

This chapter introduces the basic physical and perceptual properties of the human visual system. The specific aspects and models of the visual system incorporated in the tone mapping algorithms greatly determine the quality of their results.

2.1 Light

Light is electromagnetic radiation with wavelength (λ) between 380 nm and 740 nm. The flow of light (radiant energy) measured per unit time is radiant flux. Irradiance is the radiant flux per unit area. Spectral irradiance ($I(\lambda)$) describes the properties of a light source by giving the irradiance for each wavelength (λ). The reflectance ($R(\lambda)$) describes the light absorption properties of a surface. Radiance is the measure of flux per unit area per unit solid angle. The spectral radiance of a surface with reflectance $R(\lambda)$, illuminated by a light source of spectral irradiance $I(\lambda)$, is given by:

$$E(\lambda) = I(\lambda) \times R(\lambda) \tag{2.1}$$

2.2 The Eye

The human eye is an organ of approximately spherical shape in a protective envelope. The protective envelope has three pairs of extrinsic muscles that allow the eye to move in the eye socket providing a change of the point of fixation in the external field [Wyszecki and Stiles, 1982]. Figure 2.1 shows the physical

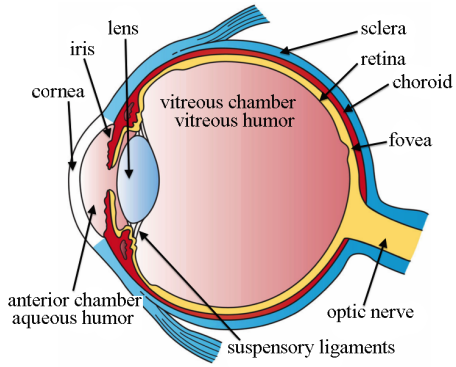


Figure 2.1: The general plan of the human eye. Image from Wikimedia Commons.

construction of the eye which can be divided into two main parts: optical and sensory. The optical part mainly consists of cornea, pupil, lens, aqueous humor, and vitreous body. The cornea is a transparent outermost layer through which the light enters the eye. The pupil is an opening in the eye that contracts and relaxes to regulate the amount of light entering the eye. The lens focuses the light entering the eye onto the fovea in the retina. The accommodation process adjusts the shape of the lens, allowing it to focus on various distances. The aqueous humor, a transparent liquid, fills the chamber between the lens and the cornea. The space between the lens and the retina is filled with a transparent jelly called the vitreous body.

The sensory part of the eye consists of the retina (see Figure 2.2). The retina is a thin multi-layered neural tissue that lines the eye, it can be subdivided into three layers based on the functional properties of the neural network: 1) layer of photoreceptor cells, 2) layer of intermediate neurons and 3) layer of ganglion cells [Wyszecki and Stiles, 1982]. Several neighboring photoreceptor cells are connected to a common bipolar cell and also to a common horizontal cell in the intermediate neurons layer. The horizontal cell is further connected to a number of photoreceptor cells in other areas of the retina, but it is not connected to the ganglion cells. Whereas, the bipolar cells are further connected to the ganglion cells. The amacrine cells connect the neighboring bipolar and ganglion cells governing lateral process. The ganglion cells provide the only retinal output signals.

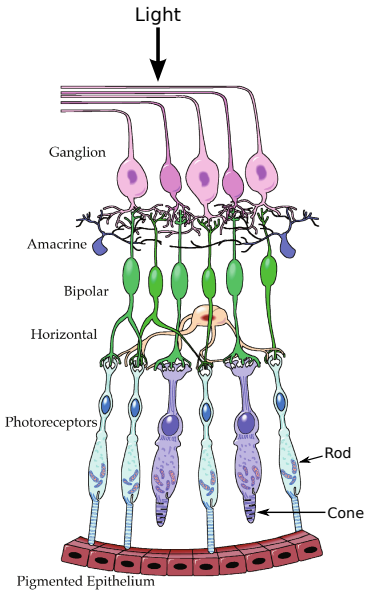


Figure 2.2: Human retina. Figure from [Fairchild, 2013].

The photoreceptor cells, both rods and cones, contain the light-sensitive pigments that when absorb photons, evoke neural signals. These signals are passed through the intermediate layer of the retina and result in a coded signal emerging from the ganglion cells. The axons of the ganglion cells form an optic nerve that is connected to the brain [Wyszecki and Stiles, 1982]. About 90% of the fibers in the optic nerve pass through the lateral geniculate nucleus (LGN) where they relay neural signals to the primary visual cortex (V1). However, the LGN most certainly functions also as a center for visual processing since it receives numerous reciprocal input from the visual cortex that could serve as an adaptation feedback mechanism [Fairchild, 2013].

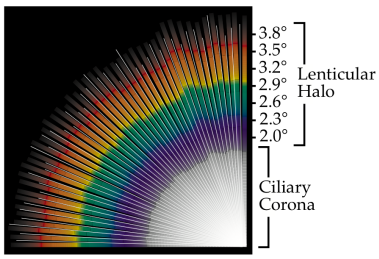


Figure 2.3: The ciliary corona and lenticular halo for a small white light source. Figure from [Spencer et al., 1995].

2.2.1 Optics of the Eye

The optical part of the human eye, like any other optical systems, is imperfect, thereby distorts the light that passes through it. The distortion can be due to the absorption and scattering in the optic media, and the reflections at the surfaces between the media, and results in a glare effect. Glare effect is composed of two major components: flare and bloom. The flare is mainly caused by the lens and is observed as radial streaks (ciliary corona) originating from the center of the light source, and as a set of colored, concentric rings (lenticular halo) surrounding the light source [Spencer et al., 1995] (Figure 2.3). Bloom, also referred to as veiling luminance, is caused by the light scattering from the cornea, the lens, and the retina. It is observed as a glow around the bright light sources resulting in a contrast loss or blur. The blurring effect can be modeled as a point spread function (PSF). This function depends on the angular distance, eccentricity (distance from the foveal region) and pupil size of the eye. An example of PSF is shown in Figure 2.4.

2.2.2 Rods and Cones System

Rods are functional from very low light levels to around 10 cd/m^2. On the other hand, cones are insensitive at very low light levels; the functional range is from 10^{-1} to 10^{8} cd/m^2. This operational characteristic of rods and cones determines three vision modes of the HVS: when only rods are active is called scotopic, when only cones are active is called photopic, and when both rods and cones are active is called mesopic (see Figure 2.5).

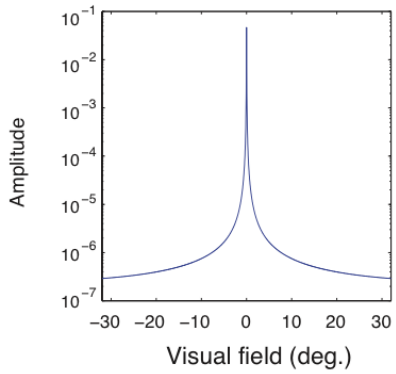


Figure 2.4: An example of the Point Spread Function. Figure from [Kane and Bertalmio, 2015].

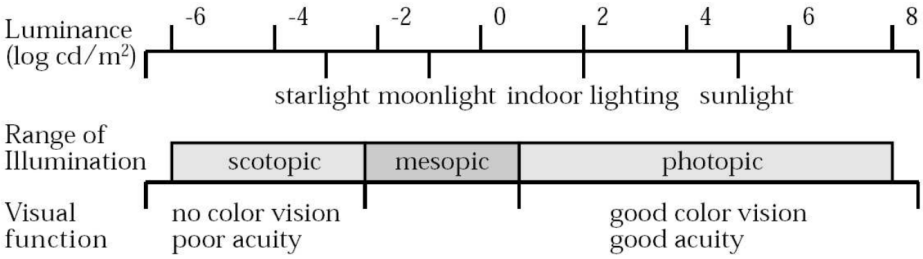


Figure 2.5: The luminance range of the real world and associated visual parameter. Figure from [Ferwerda et al., 1996].

The scotopic vision mode occurs under dark and dim illumination conditions. At this mode, only rods are active because the cones are insensitive to very low light levels. The rods cannot distinguish colors, so in scotopic mode, vision is achromatic. Also, the visual acuity is reduced with the decrease in the illumination as the neural response is collected over larger groups of rods to increase the effective sensitivity of the visual system.

The vision is in photopic mode under well-lit illumination conditions. Only cones are active at this mode because the rods get saturated by too strong illuminations. The photopic mode allows good color vision and higher visual acuity

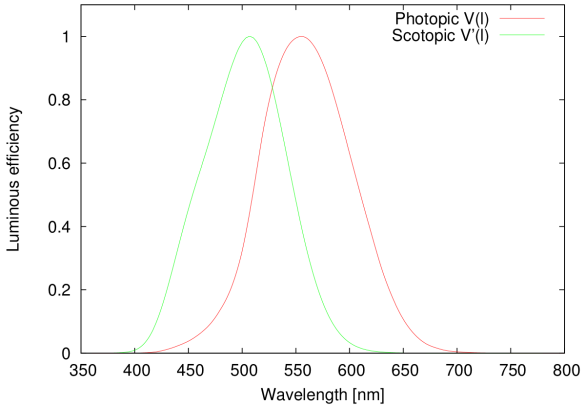


Figure 2.6: CIE spectral luminous efficiency curve for photopic and scotopic vision. Figure from http://www.cvrl.org/.

than in scotopic mode.

The mesopic vision mode is a transition mode between scotopic and photopic vision and occurs under dim lighting conditions. At this mode, both rods and cones are active, allowing for color vision.

The spectral sensitivity of rod and cone photoreceptors varies for different wavelengths of light. Rods are most sensitive at around $505\ nm$, whereas the sensitivity of the cone system peaks at around $555\ nm$. The luminous efficiency function shows the relation between the wavelength of a light and its relative efficiency as a visual stimulus. The sensitivity differences between rod and cone photoreceptors lead to separate luminous efficiency functions in scotopic and photopic vision mode. Figure 2.6 shows the normalized scotopic and photopic luminous efficiency curves $V(\lambda)$ developed by CIE.
Luminance is the integral of the spectral radiance $E(\lambda)$ over the visible spectrum with $V(\lambda)$ as follows:

$$Y = \int_{380}^{740} V(\lambda)E(\lambda)d\lambda \tag{2.2}$$

and is measured in cd/m^2 (or nits). This measurement of light is the most used in terms of imaging systems, especially for high dynamic range imaging.

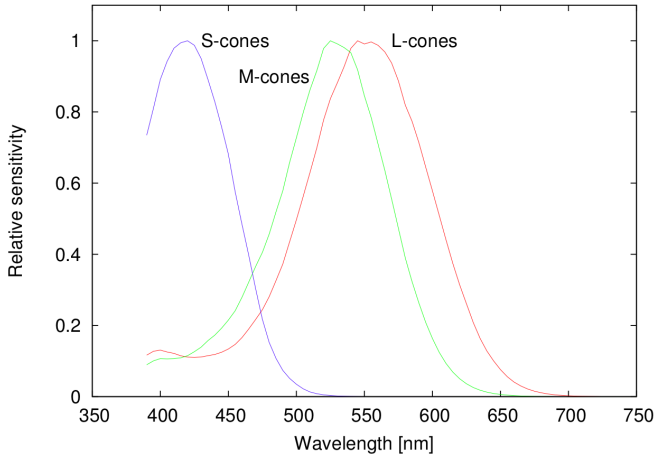


Figure 2.7: Normalized cone sensitivity. Figure from [Stockman and Sharpe, 2000].

The perception of color in human vision is possible because of the three types of cone photoreceptors: L, M and S. The spectral sensitivity of these cones varies for different wavelength of light and were first experimentally measured by König [König and Dieterich, 1892]. It is found to be as shown in Figure 2.7. The L-cones are most sensitive to long wavelengths, M-cones to medium wavelengths and S-cones to short wavelengths. With the colorimetric approach, the response of cones to a light spectrum of spectral power distribution $\phi(\lambda)$ is determined by a triplet of values, called the tristimulus values, given by:

$$\begin{aligned} L &= \int_{380nm}^{740nm} l(\lambda)\phi(\lambda)d\lambda \\ M &= \int_{380nm}^{740nm} m(\lambda)\phi(\lambda)d\lambda \\ S &= \int_{380nm}^{740nm} s(\lambda)\phi(\lambda)d\lambda, \end{aligned} \tag{2.3}$$

where $l(\lambda), m(\lambda)\ and\ s(\lambda)$ are the spectral sensitivity of L, M and S cones, respectively.

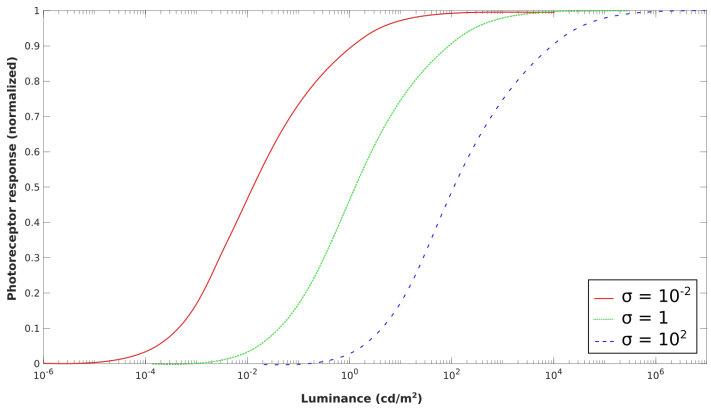


Figure 2.8: Photoreceptor response curves for several values of the semi-saturation constant σ.

2.2.3 Photoreceptor Response

Photoreceptors transform the incident light into neural signals. However, this transformation is non-linear [Hunt, 1995] and can be approximated by the Michaelis-Menten equation:

$$\frac{R(Y)}{R_{max}} = \frac{Y^n}{Y^n + \sigma^n}. \tag{2.4}$$

This equation was devised from a neuroscience experiment on a single photoreceptor response to pulses of light with intensity Y on a uniform background. The above equation is called Naka-Rushton equation when $n = 1$. The semi-saturation value σ is the value of Y at which the response R is half of the maximum response. Figure 2.8 shows the modeled response of photoreceptors for different semi-saturation values. The photoreceptors respond successfully to a luminance range of about four orders of magnitude under constant ambient illumination. However, the luminance range in nature can go from 10^{-6} to 10^9 cd/m^2. The operational range of the photoreceptors is changed according to the average light level facilitating visual adaptation.

2.2.4 Receptive Fields

The receptive field is the area in the sensory space to which a neuron responds. For example, the receptive field of a photoreceptor is the area in the visual field

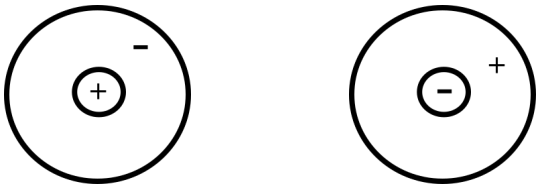


Figure 2.9: Illustration of center-surround receptive fields: on-center (left) and off-center (right).

in which light alters the response of that cell. In the case of a ganglion cell, the receptive field is formed by all the photoreceptors that connect with this ganglion cell through bipolar cells, horizontal cells and amacrine cells. The receptive field of ganglion cells is organized into a center-surround arrangement as illustrated in Figure 2.9. There are mainly two types of ganglion cell receptive fields: on-center (Figure 2.9 left), and off-center (Figure 2.9 right). An on-center cell fires when the center of its receptive field is stimulated, whereas it is inhibited when the surround is stimulated. An off-center cell has the opposite behavior. These receptive field arrangements of the ganglion cells allow them not to respond to uniform fields but to changes in the visual world. Thus, the contrast information is transmitted. Ganglion cells, in addition to spatial opponency (Figure 2.9), often show spectral opponency as illustrated in Figure 2.10. These cells can be categorized into two groups: one that responds to the difference between the firing of L and M cones (red-green opponency), the other that responds to the difference between the firing of S cones and combined L and M cones (yellow-blue opponency).

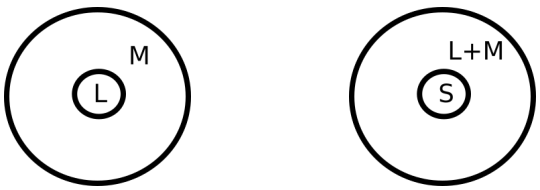


Figure 2.10: Illustration of spectral opponency in receptive fields: reg-green opponency (left) and yellow-blue opponency (right).

The LGN cells have receptive fields similar to those of the ganglion cells. The excitation and inhibition within the LGN receptive field allow to enhance the contrast across the borders and increase the dynamic range of the message relayed to the visual cortex [Martinez et al., 2014]. The receptive fields of visual cortex neurons are much more complex as they combine and compare various LGN cells and also interact with many other cortical regions. In the early 1990s, divisive normalization was proposed to model the activity of neurons in the V1 area of the cortex. According to the model, the neurons compute a linear weighted average of the input (both in space and time) and the result is further normalized by the response of nearby neurons [Carandini and Heeger, 2012]. There are cells in V1 that are found to respond selectively to the stimuli based on its orientation, spatial frequency, temporal frequency, and spatial location.

The visual processing becomes increasingly more complex as the signal moves along the visual pathway. This makes it extremely difficult to explain the behavior of a single cell at the cortical level. In addition, understanding the function of a single cell may not explain completely the visual response and perception, because collections of cells interact with each other to form the response [Fairchild, 2013]. The following section describes the psychophysical properties of the visual system.

2.3 Psychophysical Studies of Visual Sensitivity

In perception, the foundational psychophysical experiments were conducted by E. H. Weber in the mid-nineteenth century. His experiments with hand-held weights showed that the minimum amount of weight that makes a noticeable difference when added to the initial weight is proportional to the initial weight. Later visual experiments showed a fairly similar relationship, i.e., the smallest noticeable luminance difference ΔL on a uniform surround is linearly proportional to the luminance of the surround L:

$$\frac{\Delta L}{L} = k, \tag{2.5}$$

and is called Weber's law, where k is a constant called the Weber fraction and ΔL is called just noticeable difference (JND). The plot of JND versus L in log-log scale is called a threshold versus intensity (t.v.i.) curve (see Figure 2.11).

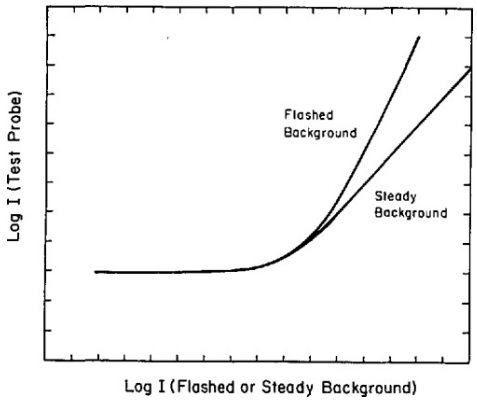


Figure 2.11: Threshold vs intensity curve. Figure from [Massof et al., 1988].

The above model does not hold for all conditions. In order to overcome the inconsistency in low luminance values, G. Fechner proposed a correction,

$$\frac{\Delta L}{L+m} = k, \tag{2.6}$$

where m is a small positive value. By adding the new term, the slope of the t.v.i. curve becomes close to zero at low-intensity values and that matches with the experimental data. Eq. 2.6 is now known as Weber-Fechner's law.

Stevens law: Weber's law is based on a detection threshold experiment and may not generalize to real world stimuli, where contrast is mostly higher than the detection threshold. In [Stevens and Stevens, 1963], the authors proposed a model to relate the perceived magnitude (brightness) to the luminance based on a magnitude estimation experiment, in which subjects rate the brightness of the stimuli on a 0-10 scale. The model relates the brightness to the luminance by a power function, with an exponent close to $\frac{1}{3}$.

The detection threshold or JND not only depends on the background luminance level but also on several other factors like the state of adaptation of the HVS, and the spatial frequency and orientation of the signal. Following sections look into these aspects.

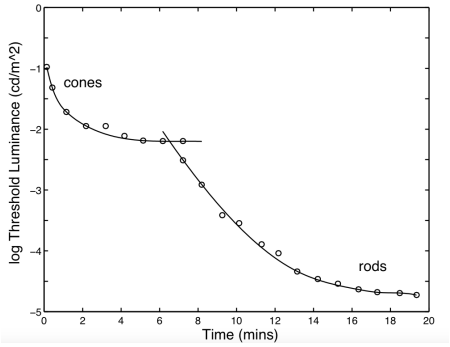
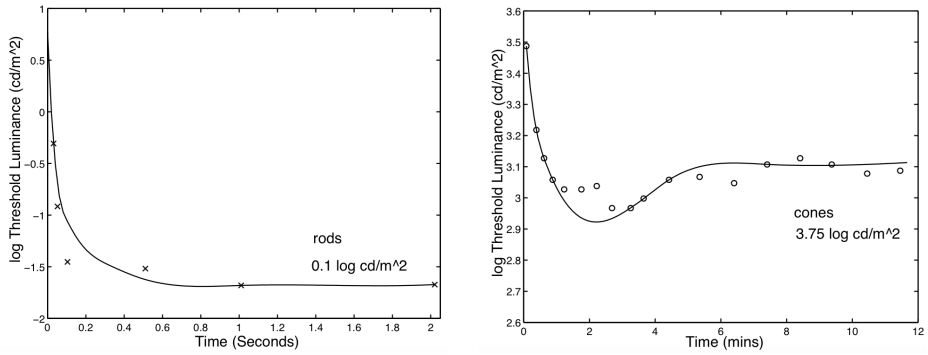


Figure 2.12: The time course of dark adaptation both in the rod and cone systems. Figure from [Ferwerda et al., 1996].

2.3.1 Dark and Light Adaptation

Visual adaptation is achieved by a collection of mechanisms that serves to optimize the visual response to a particular illumination condition. These mechanisms include: bleaching and regeneration of the photo-pigments in the photoreceptors, difference in the sensitivity and operational range of rod and cone photoreceptors, neural processes, and the contraction and dilation of the pupil [Ferwerda et al., 1996]. The adaptation due to neural processes is very fast and happens in a few seconds. On the other hand, the chemical processes of pigments in the photoreceptors are much slower and take several minutes. Threshold experiments can be used to measure the time course of visual adaptation. In such experiments, the observers are initially exposed to a particular light level until they are completely adapted. Then, the background luminance is suddenly changed to another fixed level and from that moment forward, the JND is measured. Figure 2.12 and 2.13 show the measured time course of light and dark adaptations of both rod and cone systems. The detection threshold is initially high with the sudden change in illumination and decreases asymptotically over time. Figure 2.12 shows the time course of dark adaptation as measured by [Hecht, 1934]. According to the plot, in the first five minutes, the detection threshold decreases rapidly and then remains at a relatively high level. At this



(a) (b)

Figure 2.13: The time course of light adaptation in the rod system (a) and in the cone system (b). Figures from [Ferwerda et al., 1996].

stage, the cones reach its maximum sensitivity whereas the rods have still not recovered from the initial strong illumination. Then, after about seven minutes, the detection threshold starts to drop further until the minimum threshold is reached. Figure 2.13 shows the time course of light adaptation of both rod and cone systems. The light adaptation in the scotopic range is extremely fast and about 75 % of the process happens in the first 200 ms because rods adapt quickly to light. Cones adapt relatively slower to light and take around 3 minutes to reach the minimum threshold. After about 4 minutes, the threshold level is slightly increased [Ferwerda et al., 1996].

2.3.2 Contrast Sensitivity Function

The contrast sensitivity function (CSF) describes the sensitivity of the HVS as a function of spatial frequency [Barten, 1999]. The CSF peaks at medium spatial frequencies at around 3 -5 cycles per degree and decreases for higher and lower spatial frequencies(see Figure 2.14). This effect results in increasing the detection threshold for low and high frequencies and decreasing the detection threshold for medium frequencies.

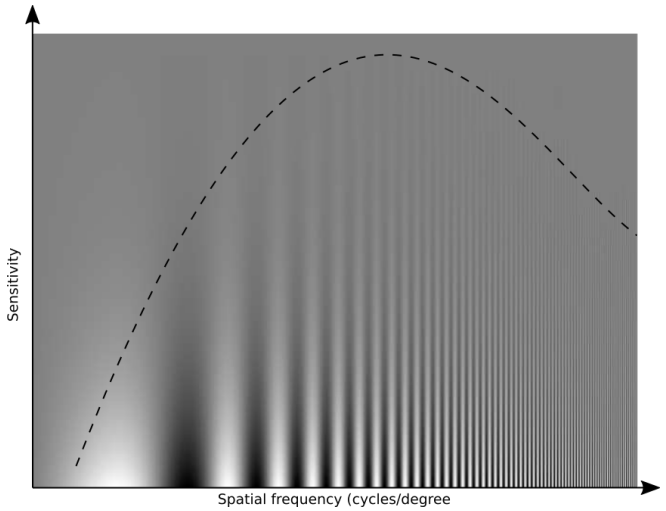


Figure 2.14: An illustration of change in contrast sensitivity with spatial frequency. Contrast sensitivity chart after [Campbell and Robson, 1968].

2.4 Limitation of Vision Models

There is a substantial gap in our knowledge about the operations in the vision system. As pointed out in [Olshausen and Field, 2005], approximately 85% of the V1 functions are still unexplained. Our knowledge of the the visual system is mainly derived from experiments using reduced stimuli (e.g., spots, white noise, or sine wave gratings). The rationale behind the use of such stimuli is that the experimenter wants to (a) have complete control over the stimulus characteristics and (b) wants to isolate a given property of the visual system and study this without the influence of other visual processes. But this poses problems, first because the use of real world stimuli in experiments is difficult, as defining even the most basic characteristics of such a stimulus such as the orientation requires assumptions, and second because the visual system is capable of using multiple different cues to solve a problem and these cues may be difficult or impossible to isolate.

It is an open question to what extent visual models developed using simple stimuli generalize to complex real world scenes of the type that the visual system is

exposed to in natural environments and has evolved to cope with. For a debate on the advantages and disadvantages of using simplistic stimuli, refer to [Olshausen and Field, 2005] and [Rust and Movshon, 2005]. In image processing, when these visual models are used, one should consider that there remain some doubts as to whether models developed using simple stimuli do generalize to complex scenes.

2.5 Dynamic Range

The dynamic range of a scene is defined as the ratio between its highest and the lowest luminance. Figure 2.5 shows the range of luminances that can exist from a starlit scene to bright sunlight. So the maximum dynamic range that can occur is $10^{14} : 1$, even though it is very unlikely that such extremes co-exist in a scene. The range can go easily over a ratio of $10^4 : 1$ in a scene. Dynamic range is often specified in orders of magnitude by taking the logarithmic base 10 of the above ratio. So the ratio $10^4 : 1$ is 4 orders of magnitude.

The dynamic range over which the HVS can operate at any given time is called simultaneous dynamic range. This range is much lower than the range of luminance values that can exist in a scene due to several reasons: light scattering in the optics of the eye; limit of the photoreceptors action potential; limited bandwidth of the neurons in the visual pathway [Kunkel and Reinhard, 2010]. In the literature, there is no clear agreement about the actual simultaneous dynamic range of the HVS. It varies from 2 orders [Myers, 2003], 3 orders [Purves and Lotto, 2003], 3.5 orders [Normann and Perlman, 1979] and even more than 3.7 orders of magnitude [Kunkel and Reinhard, 2010].

2.6 Color and Color Space

Colorimetry is the field of color science concerned with assigning numbers to the colors of physically defined stimuli such that stimuli with the same specification look alike, (i.e., are in complete color match) [Wyszecki and Stiles, 1982]. The foundation of colorimetry is the principle of trichromacy, i.e., any color can be matched by mixing three fixed primary stimuli and adjusting each of its amounts. The amount of each of the three primaries was determined through color matching experiments [Guild, 1932] [Wright, 1929]. In the experiment, the observers

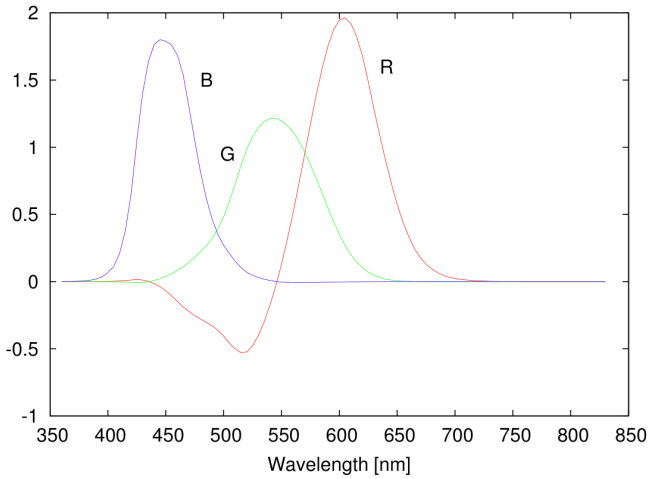


Figure 2.15: CIE RGB color matching function. Figure from http://www.cvrl.org/.

were asked to match the color of a monochromatic light by adjusting the intensity of the three monochromatic lights: red, green, and blue. The experiment was repeated over different wavelengths of test light and the intensity of each of the matching red, green, and blue primaries was collected and is given in Figure 2.15. The CIE RGB primaries of a light source with power spectral distribution $E(\lambda)$ is computed with the color matching functions $r(\lambda), g(\lambda)$, and $b(\lambda)$ as:

$$\begin{aligned} R &= \int_{380}^{740} r(\lambda)E(\lambda)d\lambda \\ G &= \int_{380}^{740} g(\lambda)E(\lambda)d\lambda \\ B &= \int_{380}^{740} b(\lambda)E(\lambda)d\lambda \end{aligned} \tag{2.7}$$

The drawback of the color matching function is that it produces negative values. CIE XYZ primaries were proposed to overcome this problem by introducing new color matching functions, $x(\lambda), y(\lambda)$ and $z(\lambda)$, which are obtained by a linear combination of $r(\lambda), g(\lambda)$, and $b(\lambda)$ with some constraints. Mainly, the color matching function should be positive and $y(\lambda)$ should match with the CIE standard luminosity function $V(\lambda)$.

The XYZ primaries of a light source with power spectral distribution $E(\lambda)$ is computed as:

$$\begin{aligned} X &= \int_{380}^{740} x(\lambda)E(\lambda)d\lambda \\ Y &= \int_{380}^{740} y(\lambda)E(\lambda)d\lambda \\ Z &= \int_{380}^{740} z(\lambda)E(\lambda)d\lambda \end{aligned} \tag{2.8}$$

Since there is a linear relationship between the CIE RGB and CIE XYZ color matching functions, a 3-by-3 matrix multiplication converts RGB colors into XYZ colors.

The primaries of CIE XYZ are not physically realizable (they would require light of "negative" intensity) therefore they are called virtual primaries and used mainly as a reference in color conversion formulas.

To provide a convenient two-dimensional representation of colors, chromaticity coordinates are often used. The transformation from (X, Y, Z) to the chromaticity coordinates are given by:

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z}. \end{aligned} \tag{2.9}$$

Figure 2.16 shows the CIE xy chromaticity coordinates.

The CIE XYZ color space is not perceptually uniform, i.e., the perceived difference between two colors is not proportional to the distance between the points corresponding to the colors in the XYZ color space or in the xy chromaticity diagram. To overcome this problem, in 1976 the CIE introduced two new color spaces that are approximately perceptually uniform. They are CIE 1976 $L^*a^*b^*$ (abbreviated CIELAB) and CIE 1976 $L^*u^*v^*$ (abbreviated CIELUV), where L^* is the lightness and both a^*, b^* and u^*, v^* are the chromaticity coordinates. The coordinates a^*, b^* are designed to make the perceptual difference between two colors proportional to the distance between the points corresponding to the colors in CIELAB [Bertalmío, 2014]. The coordinates u^*, v^* are designed to make

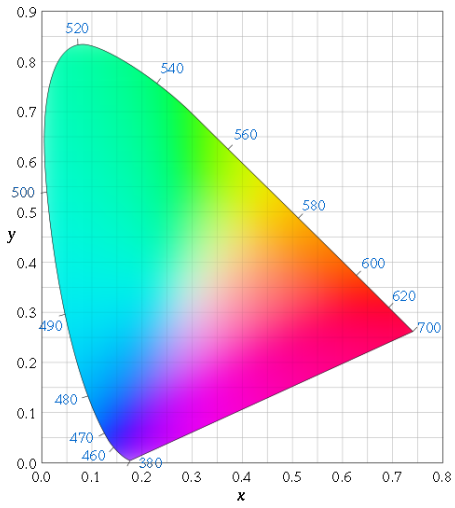


Figure 2.16: CIE xy chromaticity diagram. Figure from wikimedia commons.

the just noticeable different colors approximately equi-spaced. The color space CIELAB is defined by:

$$L^* = \begin{cases} (\frac{29}{3})^3 \dfrac{Y}{Y_n} & \text{if} \quad \dfrac{Y}{Y_n} \leq 0.008856 \\ 116 \left(\dfrac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \text{if} \quad \dfrac{Y}{Y_n} > 0.008856 \end{cases} \tag{2.10}$$

$$\begin{aligned} a^* &= 500 \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ b^* &= 200 \left(f(\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \end{aligned} \tag{2.11}$$

where

$$f(x) = \begin{cases} x^{\frac{1}{3}} & \text{if} \quad x > 0.008856 \\ \frac{1}{3}\left(\dfrac{29}{6}\right)^2 x + \frac{4}{29} & otherwise \end{cases} \tag{2.12}$$

In, CIELUV, the lightness L^* is the same as in CIELAB and the chromaticities is defined by:

$$\begin{aligned} u^* &= 13L^*(u' - u'_n) \\ v^* &= 13L^*(v' - v'_n) \end{aligned} \tag{2.13}$$

where

$$\begin{aligned} u' &= \tfrac{4X}{X+15Y+3Z} \\ v' &= \tfrac{9Y}{X+15Y+3Z} \end{aligned} \tag{2.14}$$

In these equations (X_n, Y_n, Z_n) are the tristimulus values of the reference white, u'_n and v'_n are the (u', v') chromaticity coordinates of the white point.

Chapter 3

DIGITAL CAMERAS AND HIGH DYNAMIC RANGE IMAGING

This chapter introduces the image processing steps in the camera imaging pipeline (Figure 3.1). Then we discuss various methods for capturing and encoding high dynamic range images and videos.

3.1 Camera Imaging Pipeline

A digital camera sensor is a semiconductor device formed by an array of sensor elements (pixels) that convert photons into electric signals using the photoelectric effect [Nakamura, 2005]. The sensor response is linear to the number of absorbed photons, but the proportion of absorption decreases with the wavelength of incoming light as shown in Figure 3.2. Since the sensor is sensitive up to 1100 nm, it is necessary to place an infrared filter in the optical path so that only visible light reaches the sensor. The electric charge generated in the pixels is transferred for subsequent steps through a scanning process. Based on the way the scanning is performed the sensor devices are classified into two groups: a charged coupled device (CCD) and complementary metal-oxide semiconductor (CMOS). In CCD, the pixels are scanned row-wise and charge in each of the pixels in a row is transferred sequentially and converted into voltage at one am-

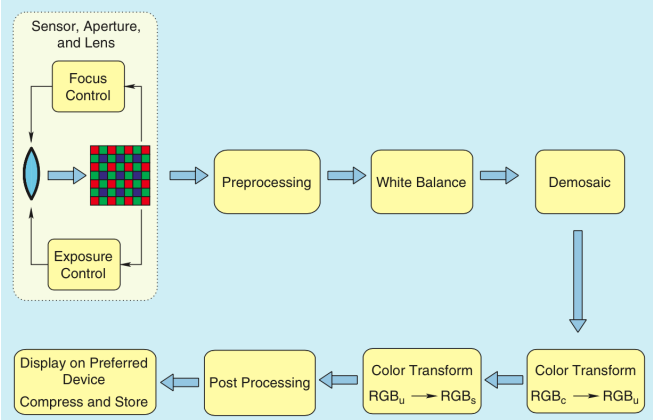


Figure 3.1: Camera imaging pipeline. Figure from [Ramanath et al., 2005].

plifier. Whereas in CMOS, at each pixel the charge is converted into voltage.

The sensors of the camera are unable to distinguish between the wavelength bands and hence cannot produce a color image. To produce a color image, a color filter array (CFA) following a Bayer pattern (see figure 3.3, left) is placed in front of the sensor [1]. With this arrangement each cell of the sensor responds only to a particular wavelength band hence measuring the amount of one of the red, green or blue components (see Figure 3.3, right).

Many cameras provide an option to record the sensor values without doing any further processing. Those are called RAW images and should be processed offline to get an image that looks visually pleasant. Therefore, both in-camera and offline the RAW sensor values must pass through a number of image processing steps to produce the final displayable image (see Figure 3.4). The offline processing allows for using more sophisticated methods to provide better results, methods that are too computationally intensive to be implemented in-camera. Camera manufacturers do not disclose the details about the in-camera processing algorithms used in their models, but in general they have the steps shown in Figure 3.1: white balance, demosaicing, color conversion, gamma correction,

[1] Other options are possible, like having three sensors and a beam-splitter, but this is much less common.

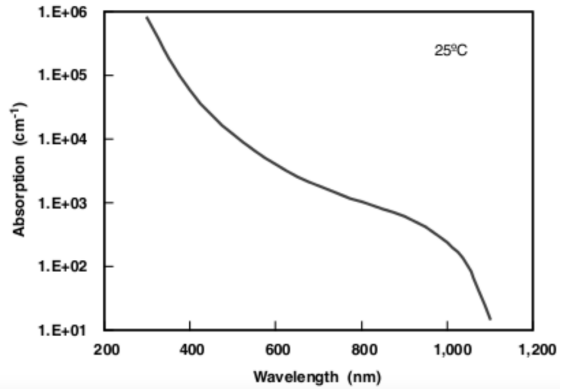


Figure 3.2: Absorption coefficient of light in silicon. Figure from [Nakamura, 2005].

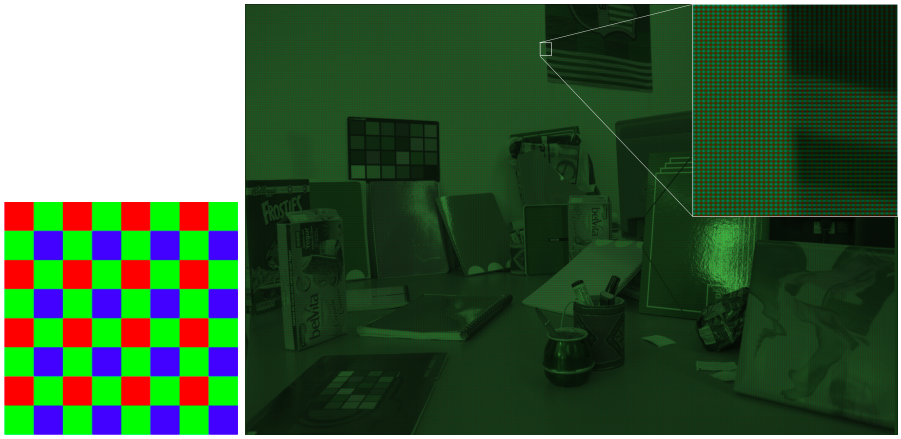


Figure 3.3: Bayer pattern (left). RAW image filtered by Bayer pattern along with a zoomed in crop (right). Image transformed for display purpose.

Figure 3.4: RAW sensor image (left) and final camera output (right).

enhancement (noise removal, contrast enhancement, etc.), and compression.

White balance: The human visual system has an important property of color constancy that allows us to perceive the color of an object to remain approximately the same under changes in illumination. White balance is the image processing technique that tries to maintain the color of an object irrespective of the changing illuminants. In camera, it is performed by an automatic process that initially estimates illumination [Buchsbaum, 1980] [Cardei and Funt, 1999] which is then used to normalize the sensor values [Lam and Fung, 2008].

Demosaicing: We have seen above that camera sensors record only one of the red, green or blue color components. In most cameras, for each pixel we only get a single color channel. The full-color image is then reconstructed by an interpolation technique called demosaicing [Menon and Calvagno, 2011].

Color conversion: The (R, G, B) sensor values are converted into (X, Y, Z) tristimulus values by multiplying it with a colorimetric matrix. The 3-by-3 colorimetric matrix is pre-computed in a least square minimization approach [Hung, 2005] during the camera calibration such that the converted (X, Y, Z) values match with actual measured (X, Y, Z) values for several color patches. The (X, Y, Z) values are further transformed into some standard color space (e.g. BT.709, DCI P3) by multiplying by another 3-by-3 matrix.

Gamma correction: Gamma correction was introduced initially for compensating the non-linear response of cathode ray tube (CRT) displays. The luminance

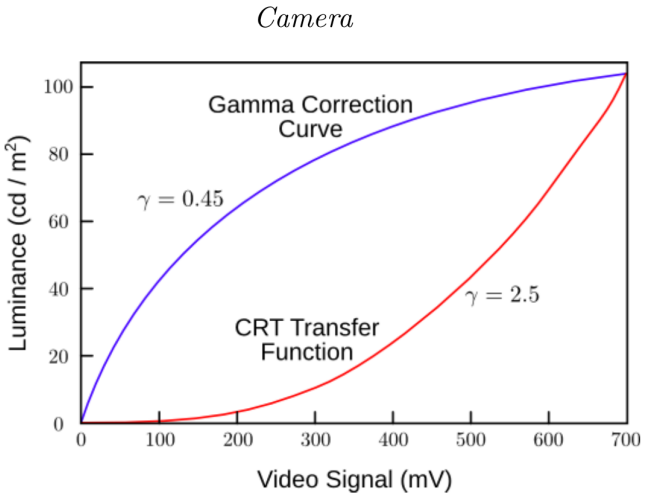


Figure 3.5: Gamma correction on a CRT. Figure from [Bertalmío, 2014].

produced by a CRT display is non-linear with respect to the device's voltage input. It can be approximated with a gamma non-linearity with a value of around 2.5 and is called decoding gamma or electro-optical transfer function (EOTF). Due to this reason, when providing a CRT with a linear input proportional to the light intensity captured by a camera, the reproduced image looks darker and unpleasing (see Figure 3.6, left). To overcome this issue the image is non-linearly corrected with an encoding gamma or opto-electrical transfer function (OETF), before feeding it to the CRT (see Figure 3.5). This process is called gamma correction and helps in producing images that look closer to the real scene (see Figure 3.6, right).

Even though CRTs became obsolete and the later display technologies are capable of producing luminance linear to the input signal, the technique of gamma correction has continued to be used by incorporating a decoding gamma similar to that of the CRT by means of a look-up table (LUT) mainly due to following reasons.

When coding signals with a limited number of bits, quantization artifacts are less likely to be visible if we code gamma corrected signals instead of linear

Figure 3.6: Linear image (left) and gamma corrected image (right).

signals. This is because, gamma correction allows to allocate more bits at dark regions where the HVS is more sensitive to luminance changes, and fewer bits to bright regions where the HVS is less sensitive to luminance changes.

The encoding gamma is not exactly the inverse of decoding gamma so that the combined effect of encoding and decoding is a gamma non-linearity and is called system gamma or opto-optical transfer function (OOTF). By setting the system gamma depending on the viewing condition we can adapt the image to give similar appearance in different viewing environment.

Image enhancement: Camera manufacturers incorporate image enhancement steps in the camera imaging pipeline to make the final image look more appealing and also to reduce the distortions introduced during the image capturing and processing step. The edge enhancement process sharpens the edges in the images which were blurred due to the optics of the camera. The most common in-camera edge enhancement technique is unsharp masking, in which a scaled edge map (the difference between the original image and its smoothed version) is added to the original image to get a sharpened image. Another enhancement step is noise reduction, in which the noise introduced during the image acquisition and processing is reduced through a denoising process. For example coring, a common in-camera denoising technique, thresholds the high spatial frequency DCT coefficients that are usually associated with noise, thus reducing the noise.

Compression: For efficient storage the images are compressed. In the case of RAW images, lossless compression is often used. The data is stored using either

Figure 3.7: Camera output with decreasing exposure time from left to right.

TIFF/EP format or using the proprietary format of the camera manufacturers (for example NEF in Nikon and CR2 in Canon). Some professional cinema formats like REDCODE and DCI X’Y’Z’ perform lossy compression based on wavelet transform. In the case of processed images and videos, lossy compression is normally used and the most common compression formats used in consumer cameras are JPEG (for images) and MPEG-4 (for videos).

3.2 High Dynamic Range Capture

The dynamic range of the camera sensor is most often measured by the peak signal-to-noise ratio (PSNR), defined as the ratio of the luminance at which the sensor saturates to the minimum luminance recorded above the noise level of the sensor. It is expressed in decibels (dB) as:

$$PSNR = 20 \times log_{10}\left(\frac{Y_{peak}}{Y_{noise}}\right). \tag{3.1}$$

The dynamic range of a conventional camera is less than 70 dB, that is around 3.5 orders of magnitude. Whereas, the dynamic range of a scene with a light source can go easily above 7 orders of magnitude. Therefore, when capturing high dynamic range scenes, based on the exposure setting, the dark and/or bright regions get clipped (see Figure 3.7). To overcome this limitation HDR capture techniques were proposed and are detailed below:

Multiple exposure fusion technique: A sequence of varying exposure images of a scene is taken so that details from the darkest to the brightest regions are captured in multiple images. If the images have been non-linearly corrected like in the processed JPEG files, then they are linearized by applying the inverse of the camera response function [Debevec and Malik, 1997]. The linear images

are further multiplied by the exposure time to get the absolute scene luminance. These images are then weighted averaged to form a single HDR image [Debevec and Malik, 1997] [Mann and Picard, 1995] [Gil Rodríguez et al., 2015]. This technique has been incorporated in many mobile phones and commercial cameras. But, in general, instead of an HDR image, they produce a single JPEG image with details from multiple exposures.
The multiple exposure fusion technique can produce some misalignment and ghosting artifacts in the final HDR image, due to camera or object movement. Additional processing like image alignment based on global homography [Tomaszewska and Mantiuk, 2007] [Granados et al., 2013] and object motion compensation based on optical flow [Zimmer et al., 2011] are used to overcome these issues.

An alternative approach is to reconstruct HDR images from dual-ISO images [Gil Rodríguez and Bertalmío, 2016] [Hajsharif et al., 2014] [Gu et al., 2010]. The dual-ISO images are captured by alternating ISO values between consecutive image rows so that the bright and dark image regions are recorded well in low and high ISO rows respectively. The dual-ISO capture is possible in some Canon camera models using the Magic Lantern (ML) [MagicLantern, 2014] camera software add-on.

Multiple sensors with beam splitters: Instead of capturing multiple exposed images at different instances of time, images are captured at once using a beam splitter that directs light to multiple sensors [Aggarwal and Ahuja, 2004]. The dynamic range of the final HDR image depends on the number of sensors used. This approach overcomes the problem of motion but requires high precision optics. The HDR video dataset of [Froehlich et al., 2014] is created using this technique. The schematic of the camera arrangement is shown in Figure 3.8. A semitransparent mirror transmits about 94% of the incoming light onto one sensor and reflects about 6% of light onto another sensor. This arrangement captures two different exposures at once and allows to create HDR content.

Solid state sensors: The sensors in this category can be of two types:

1 The sensor response is similar to that of the standard sensors but the exposure time varies spatially [Ginosar and Gnusin, 1997] [Brajovic et al., 1998].

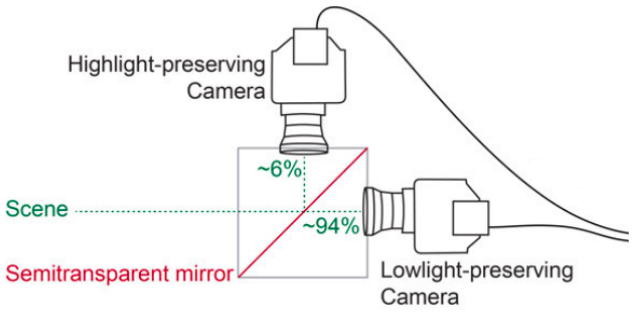


Figure 3.8: Schematic drawing of the camera arrangement in the multiple sensors with beam splitters HDR capture technique used in [Froehlich et al., 2014]. Figure from [Froehlich et al., 2014].

2 The sensor response is logarithmic and records the logarithm of the irradiance [Dierickx et al., 1996].

These sensors are used in commercial HDR video cameras (e.g., HDRC VGAx [IMS-CHIPS, 2002]) to capture dynamic scenes with strong light fluctuations. They also allow to capture wider dynamic range, but the pixel resolution tends to be low.

Computer graphics: Realistic computer graphics rendering based on modeling physics of light interaction in the rendered scenes provides a good approximation of the actual luminance distribution of the scenes [Pharr and Humphreys, 2004]. These rendered images are a good source of HDR content. The rendering process is computationally intense, but with the arrival of graphics processing units (GPU) the CGI rendering is increasingly used in real-time applications like computer games.

3.3 Encoding and Storage

HDR images encompass a wide range of luminances with high precision which would impose a huge memory demand if stored without compression. The encoding schemes try to minimize the storage demands of HDR images while preserving sufficient information for specific applications. A number of encoding

approaches mainly based on the properties of the visual system and HDR image statistics have been proposed, and they can be broadly classified into linear and non-linear encoding methods.

3.3.1 Linear Encoding Standards

OpenEXR: Is an HDR image format proposed by Industrial Light and Magic (ILM) in 2003. The files are recognized by the ".exr" extension. OpenEXR supports 16 and 32-bit floating point, and 32-bit integer. However, 16-bit floating point is generally used to encode each color channel and the formula to decode is given as:

$$H = \begin{cases} 0 & \text{if } (M = 0 \wedge E = 1) \\ (-1)^S 2^{E-15} + \frac{M}{1024} & \text{if } E = 0 \\ (-1)^S 2^{E-15} + \left(1 + \frac{M}{1024}\right) & \text{if } 1 \leqslant E \leqslant 30 \\ \infty & \text{if } (E = 31 \wedge M = 0) \\ NAN & \text{if } (E = 31 \wedge M >= 0) \end{cases} \tag{3.2}$$

where S is the sign (1 bit), M the mantissa (10 bits) and E the exponent (5 bits). This format is widely used, especially in computer graphics.

Radiance RGBE: It is an HDR image format, recognized by the ".hdr" file extension, proposed by Ward in 1998. It is a 32-bit per pixel custom floating point representation with 8-bit mantissa (R, G and B) for each color channel (r, g and b) and an 8-bit common exponent E (a common exponent is used considering the strong correlation between all color channels). The encoding is as follows:

$$E = \begin{cases} \lceil log_2 \left(max \left\{r, g, b\right\}\right) + 128 \rceil & \text{if } (r, g, b) \neq 0 \\ 0 & \text{if } (r, g, b) = 0 \end{cases} \tag{3.3}$$

$$(R, G, B) = \left\lfloor \frac{256(r, g, b)}{2^{E-128}} \right\rfloor \tag{3.4}$$

The decoding is given by:

$$(r, g, b) = \begin{cases} \frac{(R,G,B)+0.5}{256} 2^{E-128} \frac{exposure}{E_w} & \text{if } E \neq 0 \\ (0, 0, 0) & \text{if } E = 0 \end{cases} \tag{3.5}$$

The main limitation of the RGBE format is that when encoding highly saturated colors some color information is lost. This problem can be solved by encoding

in the CIEXYZ color space, using XYZE instead of RGBE.

Encoding in floating point numbers is not ideal for image compression algorithms mainly because: separately encoding mantissa and exponent requires more bits than encoding single integer values; the precision error of floating point number is not proportional to the sensitivity of the visual system. Non-linear encoding based on models of the visual system in integer numbers provides better compression.

3.3.2 Non-linear Encoding Standards

The sensitivity of the HVS to light is highly non-linear: at low luminance levels the visual system is highly sensitive to luminance changes, whereas the sensitivity is relatively low at high luminance levels. Therefore, when encoding luminance values in integer formats with limited bits, quantization error is more likely to be visible in the low luminance regions. The non-linear encoding methods consider the non-linear behavior of the visual system and transform the luminance values into an approximately perceptually uniform space in which the visual system is equally sensitive to changes throughout the entire range. As a result, banding artifacts are less likely to be visible when these transformed values are quantized into a limited bit integer format when compared to quantizing luminance values.

The quality of the encoding greatly depends on the efficiency of the non-linear transform to approximate the HVS response. The simple gamma correction that works well in encoding LDR images (with 8 bits) fails to provide reliable results with HDR images even with 12 bit representation. Figure 3.9 shows that the quantization error for 12-bit gamma encoding of luminance range of 0.01 to 10000 cd/m^2 is above the detection threshold (Schreiber threshold model mentioned in ITU BT.2246) in the low luminance region, indicating that the quantization artifacts might be visible. By encoding the logarithm of the luminance in integer format, the quantization rate is approximately proportional to the sensitivity of the HVS. Hence this encoding can provide a better representation with fewer bits. As shown in Figure 3.9, the quantization error for 12-bit logarithmic encoding is below the Schreiber threshold.

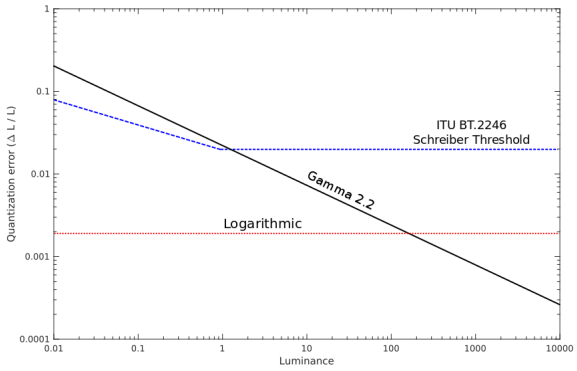


Figure 3.9: Comparison of quantization errors of logarithmic and gamma encoding approaches.

LogLuv: It uses 16-bits to encode the logarithm of the luminance and 16-bits to encode the chrominance (CIE 1976 Uniform Chromaticity Scales *u' v'*). The LogLuv encoding has a 24-bits per pixel variant with 10 bits to encode lightness and 14 bits to encode the merged u/v chrominance.

In addition, camera manufacturers introduced their own encoding quasi-log curves which include: Log C from ARRI, Canon Log from Canon, S-Log from Sony and Panalog from Panasonic.

The encoding approach proposed by [Mantiuk et al., 2004] is based on a more accurate model of the HVS response than the one used in the logarithmic encoding. The encoding of the luminance into integer values was derived from the threshold vs. intensity function [Ferwerda et al., 1996] in such a way that the quantization error is kept below the detection threshold. Chrominance components u' and v' are encoded similar to LogLuv encoding with 8 bits per channel. In [Mantiuk et al., 2005], the authors improved the above encoding by using a more accurate threshold vs. intensity function derived from the contrast sensitivity function (CSF) of [Daly, 1992].

A similar encoding approach is proposed in [Miller et al., 2013] using the CSF

developed by Barten [Barten, 2003]. Unlike the above approaches, this encoding is display referred, that is the code values correspond to the absolute luminance produced by a reference display. This allows faithful reproduction of images on displays similar to that of the reference display, but the signal has to be adapted to be viewed on other displays (detail explanation in Section 4.2.6). A modified Naka-Rushton cone response fits the proposed decoding and this functional form is called perceptual quantizer (PQ) curve:

$$Y = L\left(\frac{V^{\frac{1}{m}} - C_1}{C_2 - C_3 V^{\frac{1}{m}}}\right)^{\frac{1}{n}} \tag{3.6}$$

where Y is the display luminance; $V(0 \leqslant V \leqslant 1)$ is the input signal; $L = 10000; m = 78.8438; n = 0.1593; C_1 = 0.8359; C_2 = 18.8516; C_3 = 18.6875$. This encoding has been standardized by the Society of Motion Picture Engineers as SMPTE ST 2084 [SMPTE, 2014] and is supported by International Telecommunication Union (ITU) recommendation BT.2100. The PQ is part of two HDR formats that are supported by the recent Super Ultra High Definition (SUHD) TVs: HDR10 and Dolby vision. Both the above HDR formats include metadata in the encoded signals allowing for efficient bit allocation depending on the image content and also allowing for mapping the signals according to the display capabilities. This imposes additional complexity in the production processes as the image sequence has to be analyzed before the start of the encoding and also makes it not directly suitable for live broadcast. The experiments in [Miller et al., 2013] reported that the PQ encoding with 10 bits shows no quatization artifacts for test images, but 11 bits were needed for a JND test pattern.

The Hybrid Log Gamma (HLG) curve [Borer and Cotton, 2016], jointly proposed by BBC and NHK, is an encoding combining a gamma non-linearity for lower luminances and a logarithmic function for higher luminance values as given below:

$$E' = \begin{cases} r\sqrt{E} & \text{if } 0 \leqslant E \leqslant 1 \\ a \times log(E - b) + c & \text{if } 1 < E \end{cases} \tag{3.7}$$

where E is the signal normalized by the reference white-level and E' is the resulting nonlinear signal, $r = 0.5$, $a = 0.17883277$, $b = 0.28466892$, and $c = 0.55991073$. This is based on the observation that the HVS response has

approximately logarithmic behavior at higher luminance levels and has approximately a square root behavior at low luminance levels (De Vries-Rose Law). The use of a gamma non-linearity at the lower luminance levels makes the HLG signal compatible with standard dynamic range signals. In addition, HLG does not use metadata, making it a good candidate for encoding in live HDR broadcasts. HLG is also supported by BT.2100 with a peak display luminance of 1000 cd/m^2. The HLG encoded content is adapted for various displays by applying a system gamma depending on the peak luminance of the display. The experiments in [Borer and Cotton, 2016] reported that when encoding artificial images, in the low luminance levels, the 10 bit HLG shows slightly more banding artifacts than the 10 bit PQ. The authors also pointed out that, for natural images, the 10 bit HLG provides reliable results to the end users.

Chapter 4

DISPLAY AND TONE MAPPING

This chapter introduces various display technologies and looks into their contrast ratios (dynamic range) and the perceived display qualities. Then we discuss tone mapping techniques that allow for improving the perceived quality of the reproduction on a given display. We also show the limitations of the current tone mapping approaches and discuss some methods to evaluate the tone mapping operators (TMOs).

4.1 Display and Dynamic Range

The dynamic range or the contrast ratio of a display is generally measured by the ratio between the brightest and the darkest luminance it can produce:

$$contrast\ ratio = \frac{peak\ luminance}{minimum\ luminance} \tag{4.1}$$

and the logarithm of the ratio is expressed in orders of magnitude. Based on the way it is measured, the contrast ratio can be classified into sequential and ANSI contrast.

Sequential contrast: The minimum luminance is measured while displaying a full black image, and the peak luminance is measured while displaying a full

Figure 4.1: Display setting to measure ANSI contrast.

white image. This contrast measurement is the one usually given by manufacturers in the display specifications, but can rarely be achieved in real scenarios.

ANSI contrast: The minimum and maximum luminances are measured simultaneously while displaying the checkerboard pattern shown in Figure 4.1. This value is generally much smaller than the sequential contrast mainly because of the ambient illuminant and the reflection from the screen. This measurement gives a more reliable estimate of the perceived contrast in a real world scenario. In general, the ANSI contrast is measured by displaying 50% white content, but the amount of white content can be modified to suit the specific application.

In home TVs and computer monitors, the first major technology was the cathode ray tube (CRT). It has a vacuum tube with an electron gun and phosphor-coated screen. The electron gun focuses the beam onto the phosphors to generate photons. The maximum display luminance of a standard CRT is limited to around 100 cd/m^2 due to the limitations of the phosphorescent material and safety concerns. The minimum luminance while displaying the checkerboard pattern is around 0.5 cd/m^2 which may increase depending on the image content and ambient illuminant. This limits the ANSI contrast of the CRT display to about 2 orders of magnitude [Soneira, 2005].

Later came the Liquid Crystal Display (LCD) technology which works based on the fact that liquid crystals can transmit and change polarized light under the

effect of an electric current. That is, by controlling the voltage applied across the liquid crystal more or less light is allowed to pass and in effect produce different shades of gray (or color). The liquid crystal can only modulate the light and cannot produce light by itself, hence an external light source is required for LCD displays to produce an image. There are many techniques for the external light source arrangement, also called backlight [Baker, 2010]:

- fluorescent lamps placed at the edges of the display or an array of parallel fluorescent lamps in case of larger displays.
- light emitting diodes (LEDs) placed at the edges of the display or an array of LEDs placed behind the display.
- a digital light processing (DLP) projector.

As there is no potential limit on the amount of light the LCD can transmit, the maximum luminance it can produce depends on the backlight. In conventional LCDs, maximum luminance is limited to around 100-500 cd/m^2, mainly because more powerful single backlight considerably increases the minimum achievable luminance, also increases the power consumption and the complexity of the design. Thus the dynamic range achieved on standard LCDs is still around 2 orders of magnitude.
Better contrast ratio, about 3-4 orders of magnitude, can be achieved by separately controlling the light intensity of the LEDs in the grid or by using a DLP projector, rather than using a single backlight [Seetzen et al., 2004]. This technique is currently used in many HDR displays.

Another popular technology in HDR display is the organic light emitting diodes (OLED). OLEDs, unlike LCDs, emit light under the effect of an electric current, thus avoiding the need for an external backlight. The OLEDs can be individually controlled to produce deeper black levels and higher dynamic range in a low light condition when compared to an LCD-based HDR display [Mandle, 2010]. But their maximum achievable luminance is relatively low. Therefore, in a well-lit condition the dynamic range of OLED displays can be lower that those of LCD-HDR displays because of the black level increase with an increase in ambient lighting.

A recent advancement in home TV displays is the quantum dots (QD) technology. The QDs, depending on their size, emit light of different wavelengths under

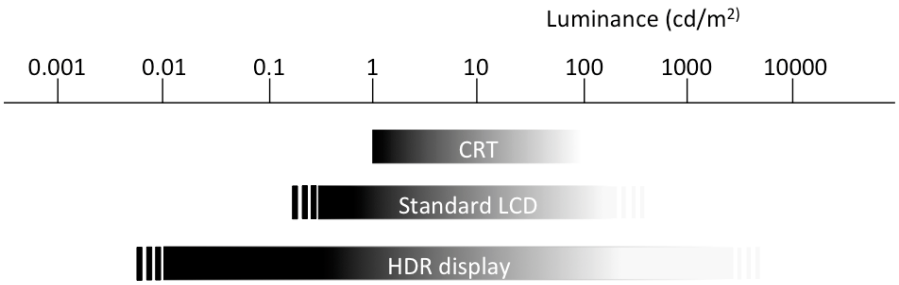


Figure 4.2: The luminance range of currently available displays.

the effect of electric current or external light source [Soneira, 2014]. The current QD displays, which are essentially LCD displays, use QD only to filter the light from the LED backlights. The QD converts the blue light from the LED into saturated red, green and blue lights so that there is no need to use a color filter like the one used in normal LED-LCD displays. Thus, avoiding the light loss at the color filter and producing displays with better peak luminance and dynamic range.

In general, the maximum ANSI contrast generated by standard displays is around 2 orders of magnitude and by improved displays is around 3 orders and in extreme cases is 4 orders of magnitude (see Figure 4.2).

The most widely used technology in digital cinema projection is Digital Light Processing (DLP). The DLP projectors use adjustable micro-mirrors to reflect the backlight. Depending on the orientation of the mirror the light is reflected either towards the screen through a lens or to a heat sink. The ratio of time between the two mirror orientations in a given time interval determines the intensity of the projection [Dudley et al., 2003]. Each mirror generally represents one image pixel so to form a complete image several million micro-mirrors are arranged in a matrix on a chip, known as Digital Micro-mirror Device (DMD).

Other light modulation technologies in projectors are based on Liquid Crystal on Silicon (LCoS) and LCD. The LCoS reflects light using liquid crystal chips on top of a mirrored back-plane. The intensity of the reflection is controlled by

the liquid crystal. LCD based projectors, similar to LCD displays, use LCD panels to modulate and transmit the backlight towards the lens.

A three-chip projection technique is usually used in digital cinema to produce color images. Each of these chips is separately illuminated by either red, green or blue primary light so that they control the intensity of one of the primary colors. The reflected light (transmitted in the case of LCD) from the chips is then recombined to form a complete color image. The minimum achievable black level by a projector is determined by its ability to minimize the internal light scattering. And the peak luminance greatly depends on the effective illumination at the chips.

The backlight illumination in the projectors can be either from a Xenon arc lamp unit, LEDs or laser emitting sources. In the case of Xenon arc lamp, the light is split using a prism and then the primary colors are directed to the particular chip and the rest are disposed of. As a result the effective illumination on the chip is considerably reduced. Laser illumination overcomes this problem by providing the backlight of specific colors. In addition, the maximum luminous flux produced by the lasers is almost double than that of the Xenon arc lamp. These factors allow laser projectors to provide higher peak brightness and better dynamic range than the lamp-based projectors [Belle, 2016]. The peak brightness can further be increased by a dual-projection technique: projecting overlapping images from two projectors. But, this setup leads to an increase in the minimum luminance so that the effective dynamic range is not further improved [Damera-Venkata and Chang, 2009]. The following section discusses the achievable contrast by projectors.

The actual dynamic range achieved in a cinema projection depends not only on the dynamic range of the projector but also on the auditorium reflectivity and the light scattering at the projector optical and port window. In [Tydtgat et al., 2015], the authors showed that a projector with a sequential contrast of 1 million to 1 can provide an ANSI contrast of only 2400:1 when displaying an image with 1.5% white content in an average theater. The ratio goes even lower to about 700:1 when displaying an image with 5% white content. Thus, the maximum achievable effective dynamic range of current cinema projection is limited to around 3 orders of magnitude.

4.2 Tone Mapping

Tone mapping is the process of mapping the contrast and brightness of an HDR image onto that of a display. We saw in the previous chapter that a linear mapping or a simple gamma correction does not provide reliable results. There are varieties of TMOs in the literature and their goals vary depending on the specific application in which they are used:

- produce in viewers a perception of details as close as possible to the one he/she would have had by observing the original scene directly.
- produce a pleasing image according to subjective preference or artistic intent.

Tumblin and Rushmeir [Tumblin and Rushmeier, 1993] formally introduced the problem of tone mapping to the computer graphics field. Since then, many TMOs have been proposed and can be broadly classified as global or local approaches. An excellent review of various TMOs can be found in [Reinhard et al., 2010] [Banterle et al., 2011].

4.2.1 Global TMOs

Most of the global TMOs consist in applying a compression curve to the image levels, based on psychovisual laws. Besides [Tumblin and Rushmeier, 1993] who use Stevens' law, the Naka-Rushton formula is used in ([Pattanaik et al., 2000], [Reinhard and Devlin, 2005], [Kuang et al., 2007]), Ferwerda's model in [Ferwerda et al., 1996], and Weber-Fechner's law in ([Ashikhmin, 2002], [Larson et al., 1997]) to name a few.

The TMO proposed in [Reinhard and Devlin, 2005] is based on the idea that tone mapping is similar to the adaptation process in the HVS, and used a modification of the Naka-Rushton equation:

$$I_d = \frac{I}{I + \sigma(I_a)}, \tag{4.2}$$

where I is the input HDR image, I_a is the adaptation level generally computed by a weighted sum of the luminance and the intensity of a color channel, and

Figure 4.3: Tone mapping result of [Reinhard and Devlin, 2005] when applied on medium DR image (left) and very HDR image (right). Original HDR image from [Fairchild, 2007].

Figure 4.4: Tone mapping result of [Drago et al., 2003a]. Original HDR image from [Fairchild, 2007].

the semi-saturation constant $\sigma(I_a) = (fI_a)^m$ depends on two parameters f and m. Figure 4.3 shows the result of the method generated with default parameters. The overall appearance of the scene is not preserved in the image shown on the right. This happens mainly because the default parameter values are not optimal. Also, a global tone curve may not preserve well the details when mapping a very high dynamic range image.

An adaptive logarithmic curve proposed in [Drago et al., 2003b] uses a collection of logarithmic functions ranging from log_2 to log_{10}, the choice of the logarithm base depending on the luminance values. It tends to over-enhance the contrast in some regions, also the image might appear unnatural (see Figure 4.4).

The TMO proposed in [Larson et al., 1997] performs a histogram adjustment

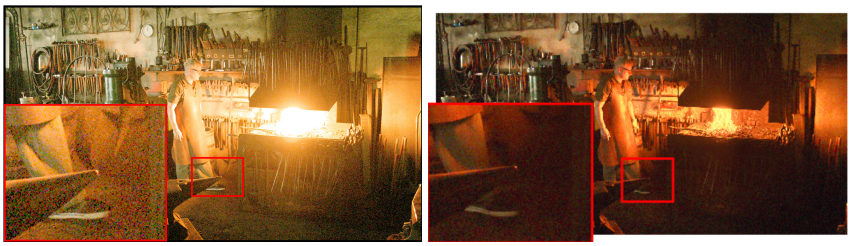


Figure 4.5: Tone mapping result of [Mantiuk et al., 2008] (left) and [Eilertsen et al., 2015] (right). Original HDR image from [Froehlich et al., 2014].

technique. The tone mapping curve is the normalized cumulative histogram of the density image (log intensities) that is modified by perceptual constraints. The method expects absolute scene luminance as the input and provides a sub-optimal result when provided with relative luminance, which is the most common situation.

A piece-wise linear tone curve is used in [Mantiuk et al., 2008] to achieve dynamic range compression. The parameters of the tone curve are chosen so as to minimize the difference between the estimated response of the HVS model for the resultant image and the original image. The method can adapt to the particularity of the display and the viewing condition. If the input HDR image is noisy, there is a high chance of noise becoming apparent in the result as the method tries to reproduce as much detail as possible (see Figure 4.5, left). In [Eilertsen et al., 2015], they try to overcome this issue by adapting the tone curve by modeling the image noise (see Figure 4.5, right).
By simplifying the visual model considered in [Mantiuk et al., 2008] a piecewise tone curve which is very similar to integrating the cube root of the image histogram is proposed in [Mai et al., 2011]. The main aim of this approach is to encode HDR images efficiently to reproduce on HDR displays and also to provide a backward compatible version which can be displayed on SDR displays. The drawback of the method is that the SDR version is not always a reliable reproduction.

A histogram equalization based tone mapping is proposed in [Duan et al., 2010]. Initially, a logarithmic function maps the input luminance into a display lumi-

nance range. Then a contrast adjustment step, depending on a weighting value, controls the image transformation between complete histogram equalization and linear mapping. There may not be a single weighting value that preserves the local image details while not making the image appear unnatural. Thus, the authors extend the method by dividing the image into blocks and applying the contrast adjustment locally on each block after imposing additional constraints to avoid halos and blocking artifacts.

In [Singnoo and Finlayson, 2010], tone mapping is achieved by means of a gamma transform (γ_{tmo}) that is derived from the gamma transform that maximizes the image entropy. The gamma value that maximizes image entropy is given by:

$$\gamma = -\frac{1}{mean(log(I))} \tag{4.3}$$

where I is the HDR image, and $\gamma_{tmo} = 0.62 \times \gamma + 0.38$.

Global TMOs are in general very fast and do not introduce halos or artifacts, but tend to produce lower contrast images.

4.2.2 Local TMOs

Local TMOs achieve dynamic range compression by modifying each pixel based on its neighborhood. They are computationally more expensive than global approaches, but they produce higher contrast images.

In [Reinhard et al., 2002] the luminance of the image is initially linearly scaled considering the log-average luminance of the image, which is similar to setting the exposure of a camera. Then it is passed through a global stage to map to the displayable range:

$$L_d(x) = \frac{L(x)\left(1 + \frac{L(x)}{L_{white}^2}\right)}{1 + L(x)}, \tag{4.4}$$

where x is the pixel position, L is the scaled luminance and L_{white} is the smallest luminance in L mapped to white. This global version is recommended in the case of medium dynamic range images, but a local TMO that enhances the contrast of the image based on the photographic "dodging and burning" [Adams, 1983] technique is used for very high dynamic range images. This local TMO

allows setting a different exposure for each part of the image. For each pixel, the operator finds the largest surrounding area (σ_{max}) without large contrast differences by iteratively increasing the size (σ) of a Gaussian kernel and computing an average until the difference between the average of consecutive iterations is greater than a threshold. Then the local operation is applied by considering the largest surround area:

$$L_d(x) = \frac{L(x)}{1 + L_{\sigma_{max}}(x)}, \tag{4.5}$$

where $L_{\sigma_{max}}(x)$ is the local average at x computed with the Gaussian kernel of width σ_{max}.

The method in [Ashikhmin, 2002] consists of three steps. First, local adaptation luminance is calculated at each point of the image by averaging luminance of nearby pixels similar to the method in [Reinhard et al., 2002]. Then, a simple function based on the linearly approximated threshold versus intensity curve is applied to these values to compress them to the displayable range. The image details that might have been lost during the above process are re-introduced by considering the detail layer given by the ratio of pixel luminance to the corresponding average luminance.

A fast bilateral filter approach for tone mapping is proposed in [Durand and Dorsey, 2002]. Bilateral filtering performs a smoothing operation that blurs an image while preserving the edge details. The operator considers two different spatial frequency layers: a base layer and a detail layer. The base layer, which encompasses high contrast edges, is compressed and added to the detail layer, which is the difference between the original image and the base layer in logarithmic scale. The bilateral filtering might result in spurious edges in high gradient and high curvature regions and also might over-smooth ramp edges and ridge-like features.

In order to overcome the limitations of bilateral filtering, a trilateral filter approach is used in [Choudhury and Tumblin, 2005]. The algorithm initially computes the logarithm of a luminance image, and then the image gradients. These gradients are smoothed and used as an indicator of the amount by which the bilateral filter [Tomasi and Manduchi, 1998] should be tilted to adapt to the local region.

A local adaptive bilateral filtering method [Wang et al., 2010] takes into account the luminance value to decide the base value, so that the contrast and details are preserved in dark areas while the luminance is compressed in highlight areas. These different bases in logarithmic domain are taken as an input for the bilateral filter.

The TMO in [Eilertsen et al., 2015] also works by initially separating the image into a base and detail layer using an edge-aware filtering. Then the base layer is split into blocks of five visual degrees each. And for each block, the method computes a piece-wise linear tone curve that minimizes the contrast distortion between the HDR and its tone mapped image, while considering the noise level in the image. The tone curve values of the neighboring blocks are interpolated to avoid discontinuity at block boundaries. Then the detail layer, after discounting for the noise visibility, is added to the transformed base layer to obtain the final tone mapped image.

The TMO in [Banterle et al., 2012] uses existing TMOs to selectively tone map different image luminance zones. The best performing TMO in each luminance zone is previously determined through a psychophysical experiment. The method segments the HDR image into luminance zones and applies the best performing TMO to each zone. Then the nearby zones are fused using Laplacian blending to get the final image.

In [Ma et al., 2015], the tone mapping is defined as an optimization problem, searching for an image in the LDR image space that optimizes the tone mapping quality index (TMQI) [Yeganeh and Wang, 2013]. The optimization is achieved by a gradient ascent method in which the initial tone mapped image is iteratively modified in the direction of optimal TMQI.

The main drawback of local TMOs apart from being computationally more expensive than global approaches is that they have the tendency to produce artifacts and halos around the edges.

4.2.3 Global + Local TMOs

Ferradans et al. [Ferradans et al., 2011] proposed a two-stage TMO. The first stage is a global operation that implements visual adaptation. It is a combination of the Naka-Rushton equation and Weber-Fechner's law:

$$c(I) = \begin{cases} \hat{k}log(I+m) + s_0 & if \quad I \leq I_M \\ \\ \frac{I^n}{I^n + I_s^n} & if \quad I > I_M \end{cases} \tag{4.6}$$

where c is the tone mapping curve, I is the intensity of the HDR image, I_s is the semi-saturation radiance, I_M is the inflection point of the radiance value (approximately $I_s * 10^2$) and s_0 is a constant to ensure continuity of c at I_M.
The second stage performs local contrast enhancement, based on a variational model inspired by color vision phenomenology. An iterative gradient descent method is applied to minimize the following energy:

$$E(I) = -\frac{1}{4}\int\int_{\Omega^2} w(x,y)|I(x)-I(y)|dxdy + \int_{\Omega}(I(x)-\mu)^2dx + \int_{\Omega}(I(x)-I_0(x))^2dx, \tag{4.7}$$

where Ω is the image domain, x and y are the pixel positions, w is a Gaussian kernel with its value decreasing with the increase in the distance between x and y, and μ is the average value of I_0, the output of first stage.

An adaptive local TM based on Retinex was proposed in [Ahn et al., 2013]. It is also a two stage operation with global and local adaptation steps. The global adaptation compresses the HDR image globally:

$$L_g(x,y) = \frac{log(\frac{L_w(x,y)}{\overline{L_w}+1})}{log(\frac{L_{wmax}}{\overline{L_w}+1})} \tag{4.8}$$

where $L_g(x,y)$ is the global adaptation output, L_{wmax} is the maximum of the input world luminance $L_w(x,y)$, and $\overline{L_w(x,y)}$ is the log-average luminance.
In the second stage, that performs the local adaptation process, an edge preserving guided filter is used. The weights of the guided filter depend not only on the Euclidean distances but also on the luminance differences. To guarantee good rendition and dynamic range compression, a contrast enhancement factor based

on luminance values of the scene is used. In addition, an adaptive non-linearity offset is introduced to deal with the strength of the logarithm function’s non-linearity.

4.2.4 Tone Mapping in Videos

Most of the TMOs in the literature are proposed to operate with still images and it is not recommended to apply them directly to video sequences on a frame-by-frame basis. This is mainly to avoid flickering or temporal inconsistencies due to the sudden change in the parameter values of the TMO between consecutive frames. Figure 4.6 shows for some TMO the parameter values over the frames of a video sequence and Figure 4.7 shows the tone mapped result of two consecutive frames.

Global tone mapping algorithms are easy to extend to videos, mainly because of their simplicity, low computational complexity, and ease of applying temporal coherency. In order to achieve temporal coherence between consecutive frames a temporal filter is applied to the parameter values of the tone curve and it can be done in mainly two ways. The first approach filters the initial parameter values

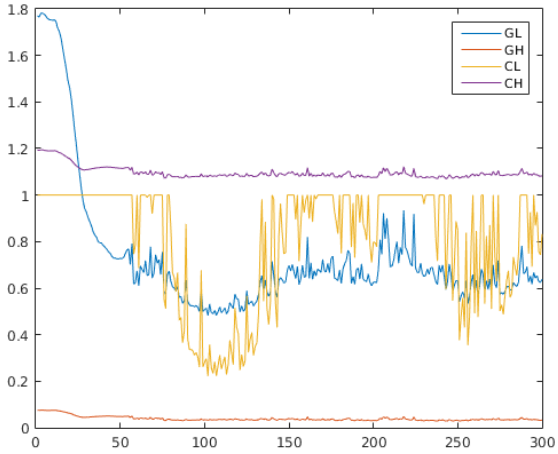


Figure 4.6: Parameter values of a TMO over the frames of a video sequence.

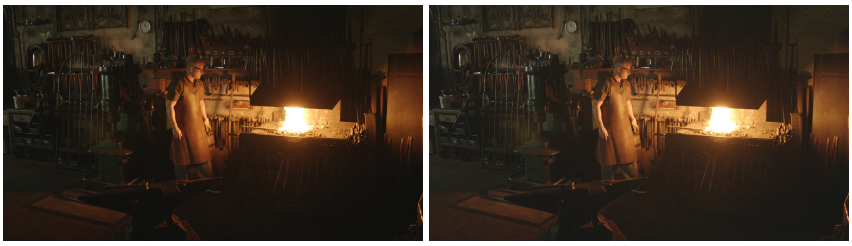


Figure 4.7: Tone mapped results of two consecutive frames after applying a TMO frame by frame. The overall brightness of the left image is less than the right one. Original HDR image from [Froehlich et al., 2014].

by averaging them with the corresponding parameter values of either a single previous frame [Kiser et al., 2012] or multiple previous frames [Ramsey et al., 2004]. The second approach is usually a two-pass procedure in which the initial parameter values are averaged with the corresponding parameter values of both previous and coming frames. For example, the method in [Mantiuk et al., 2008] is adapted for video by performing two passes. In the first pass, the nodes of the piece-wise linear tone curve are computed for all the frames, and in the second pass those nodes are adjusted to restrict any temporal variation above 0.5 Hz between the tone curves of nearby frames.

Global temporal filters as the ones just explained above cannot be used in the case of local TMOs, as they are spatially varying. A spatio-temporal domain filter can be used instead. For example, a video extension of the algorithm in [Fattal et al., 2002] is proposed in [Lee and Kim, 2007]. The method estimates a pixel-wise motion vector using block matching and tries to maintain the temporal coherence between the pixel values that represent the same object point in consecutive frames. This temporal coherence is achieved by introducing a new cost term to the original formulation so that flickering artifacts can be reduced.

In [Aydin et al., 2014], a more accurate spatio-temporal filter utilizing the optical flow is used to maintain the temporal consistency in a video. The filter decomposes each video frame into a base and a detail layer. Then, by using a graphical user interface, a user-interactive tone mapping is performed in which, the base and detail layers are adjusted and combined according to the user choice.

Due to the high computational complexity, the spatio-temporal filtering cannot be performed in real-time. Therefore, the base and detail layers of all the frames should be pre-computed before the user interactive editing. In order to allow real-time operation, [Croci et al., 2016] replaced the spatio-temporal filter with a spatial filter. Then to ensure temporal consistency, the parameter values of the tone curve applied to the base layer are smoothed in the temporal domain, as performed in [Kiser et al., 2012].

In addition, there are some post-processing techniques that allow to adapt a global TMO for video processing. One of these techniques is the brightness coherence approach [Boitard et al., 2012] that operates in two steps. In the first step the tone mapped video frames are analyzed to find an anchor frame (a best exposed frame determined based on the geometric mean of the image luminance), then in the second step the tone mapped images are adjusted to preserve the brightness coherence with the anchor frame. The brightness coherence approach is a global operation and may not handle well local fluctuations. A zonal brightness coherence method [Boitard et al., 2014] tries to overcome this limitation by dividing image frames into different brightness zones and processing each zone separately.

4.2.5 Color Reproduction in Tone Mapping

Tone mapping operators generally focus on reliably mapping the luminance of the HDR image on to the display luminance range, and then extend this mapping to color channels. This approach often causes changes in color appearance. The common approaches found in the literature for color treatment in tone mapping are:

1) Preserving the color ratio [Schlick, 1995]: The color component of a tone mapped image C_{out} is given by:

$$C_{out} = \frac{C_{in}}{L_{in}} L_{out}, \quad C = \{R, G, B\}, \tag{4.9}$$

where C_{in} represents one of the color channels of the input image, and L_{in} and L_{out} represent the luminance value before and after tone mapping respectively, that is

$$L_{out} = TMO(L_{in}). \tag{4.10}$$

Figure 4.8: Result of different color corrections in tone mapping. Left image: preserving color ratio in tone mapping (Eq. 4.9, same as Eq. 4.11 with s = 1) over-saturates the image. Right image: Eq. 4.11 with $s = 0.36$ (s computed using Eq. 4.13). Saturation can be adjusted by tuning s in Eq. 4.11 but can alter the luminance of highly saturated pixels (see the apple and grapes). Original HDR image from [Fairchild, 2007].

This simple approach tends to produce over-saturated images (see Figure 4.8, left), thus a saturation control term s is used as follows [Tumblin and Turk, 1999]:

$$C_{out} = \left(\frac{C_{in}}{L_{in}}\right)^{s} L_{out}, \quad C = \{R, G, B\} \tag{4.11}$$

When $s > 1$ the saturation is increased, and when $s < 1$ saturation is decreased. With this approach, adjusting the saturation can alter the luminance of highly saturated pixels (see Figure 4.8) and [Mantiuk et al., 2009] proposed an alternative technique to overcome this drawback:

$$C_{out} = \left(\left(\frac{C_{in}}{L_{in}} - 1\right) s + 1\right) L_{out}, \quad C = \{R, G, B\}, \tag{4.12}$$

but this method may produce hue shifts (see Figure 4.9).

In [Mantiuk et al., 2009], the authors conducted a subjective study and determined the adequate value of s needed for a contrast compression c when applied to an image:

$$s = \frac{(1 + k_1)c^{k_2}}{1 + k_1 c^{k_2}}, \tag{4.13}$$

where k_1 and k_2 are constants, and for Eq. 4.11 $k_1 = 1.6774$ and $k_2 = 0.9925$, and for Eq. 4.12 $k_1 = 2.3892$ and $k_2 = 0.8552$. The contrast compression c of

Figure 4.9: Result of color correction by [Mantiuk et al., 2009] in tone mapping (Eq. 4.12). The image has hue shifts (see the zoomed-in cropped region). Here s is computed by Eq. 4.13. Original HDR image from [Fairchild, 2007].

a tone curve is its slope in a log-log plot.

2) Independently tone map the three channels, that is

$$C_{out} = TMO(C_{in}), \quad C = \{R, G, B\} \tag{4.14}$$

where TMO can be different for different channels, although in this case the resulting image may have hue shifts (see Figure 4.10, left).

3) Applying the same tone curve to all three channels:
A single tone curve, generally derived by considering the luminance component, is applied to all three channels. This technique works well for medium dynamic range images but can produce desaturated results for very high dynamic range images (see Figure 4.10, right).

Alternatively, the TMO in iCAM06 [Kuang et al., 2007] uses an image color appearance model for color reproduction. Rather than working in the RGB color space, each operation in iCAM06 is performed in a color space that best suits that operation. Initially, the input image is converted into a base and detail layer. The CIE XYZ base image is converted to a spectrally sharp RGB image by the formula in CIE CAM02, passed through linear von Kries normalization to account for chromatic adaptation, converted into Hunt-Pointer-Estevez fundamentals using the CIE CAM02 formula, and non-linearly transformed with a generalized form of the Michaelis-Menten equation. The modified base layer is added to

Figure 4.10: Result of different color corrections in tone mapping. Independently tone map the three channels (left), applying the same tone curve to all three channels (right). Original HDR image from [Fairchild, 2007].

the adjusted detail layer and is converted into the IPT color space [Ebner and Fairchild, 1998] to perform color and surround adjustments. As a final step, an inverse output model is applied to obtain the final result.

In addition, [Preiss et al., 2014a] proposed an approach that performs tone mapping as a gamut mapping procedure in HDR color space. For this, the HDR color space HDRLAB2000HL is employed, which can simultaneously represent both the HDR and LDR images as well as the gamut boundaries of the devices. The method iteratively minimizes the distance between the original HDR image and its tone mapped (gamut mapped) image in HDRLAB2000HL space using the iCID metric [Preiss et al., 2014b] as the objective function. This method allows transforming the HDR images into the gamut of the output devices.

4.2.6 Tone Mapping for HDR Display

One might think that HDR content can be linearly reproduced on a HDR display to get reliable results. But in practice, it is never the case. Even though advanced HDR displays can provide a simultaneous contrast close to that of the simultaneous contrast of the human visual system, the input HDR content should still be tone mapped taking into account visual adaptation, viewing conditions and the brightness level the display can produce.

The PQ HDR encoding (HDR10 and Dolby Vision formats) is display referred, that is, the encoded data value corresponds to the absolute display luminance

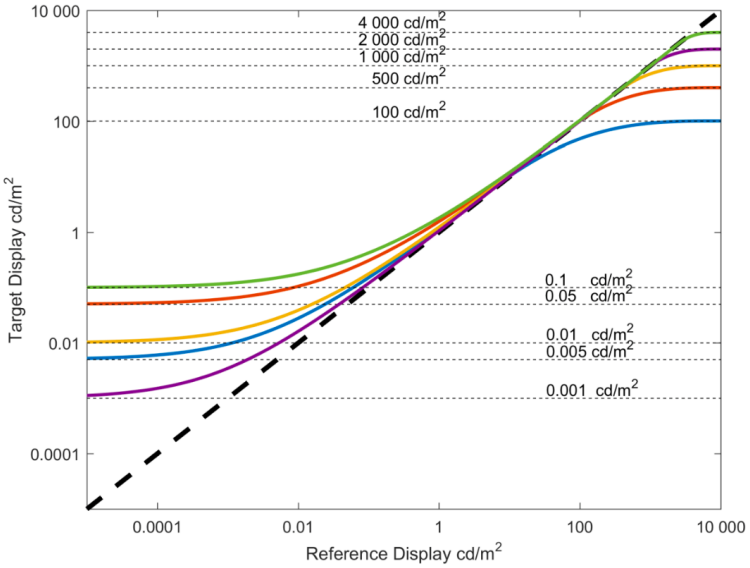


Figure 4.11: The non-linear transformations for various displays with different luminance range. Figure from [ITU, 2016].

when that image is reproduced on the target display. Currently, a target display with a peak luminance of 1000 cd/m^2 is considered in HDR10 and 4000 cd/m^2 in Dolby Vision. When those encoded contents are viewed directly on a display with different peak brightness the results are sub-optimal. A proprietary mapping technique is used in Dolby Vision whereas HDR10 uses a simple non-linear transform with a toe and knee to map the data for better visual quality. Based on the available display dynamic range the shadow and/or highlight details are 'gracefully rolled off' by the toe and knee (ITU-R BT.2390). Figure 4.11 shows the different non-linear mappings of the encoded content from a target display luminance range of 0-10000 cd/m^2 to various other ranges.

The HLG encoded HDR data is adjusted for a particular display by applying a system gamma (γ_{sys}) depending on the peak display luminance (L_w), where (γ_{sys}) is given by:

$$\gamma_{sys} = 1.2 + 0.42 \times log_{10}\left(L_w/1000\right) \tag{4.15}$$

The TMOs that take into account the display characteristics and viewing condition can also be used to map the HDR images onto the HDR displays. The TMO of [Mantiuk et al., 2008] is one such example. The display model incorporated in the method allows to determine the absolute display luminance of an image, depending on the particularities of the display and viewing conditions. The TMO adapts the tone curve so as to minimize the difference between the HVS response to the HDR image and the tone mapped image after passing through the display model. Thus, the approach can be used for tone mapping for HDR displays.

4.3 Evaluation of Tone Mapping Results

Despite being researched extensively over many years, the problem of tone mapping evaluation is still open. Some of the reasons that make the evaluation complex are:

- There isn't an image reference to which the tone mapped image should be matched.
- Tone mapping intent varies based on the application in which it is used.
- Tone mapping operators generally adapt to the input image, and it is not possible to evaluate an operator with all possible image content.
- The quality of an image is determined by multiple, sometimes contradicting, image attributes. For example, an image might look unnatural by over-enhancing the contrast.
- The complexity of the human visual system and our limited understanding of its functioning.

This latter problem can be mitigated by using the human visual system itself for the evaluation, i.e., human observers judge the quality of the tone mapped image. Then, the possible references to which the human observers could match the tone mapped image are: the real world scene (fidelity with reality), the image on an HDR display (fidelity with HDR display) or a memory image (no reference) [Eilertsen et al., 2013].

Fidelity with reality: This approach was used in a number of studies [Yoshida et al., 2005], [Ashikhmin and Goyal, 2006], [Yoshida et al., 2006], [Čadík et al., 2008] and [Cerdá-Company et al., 2016]. The experiment setup with a real world reference is extremely complex because both the physical scene and the display should be visible from the same place. This limits the range of scenes that can be considered. Also, the difference in the dynamic range, field of view, range of colors and depth information between the real scene and the tone mapped image make the task very difficult for the observers.

Fidelity with HDR display: The dynamic range produced by currently available HDR displays is considerably low when compared to the dynamic range that can exist in the real world. As a result, HDR images with dynamic range greater than that of the HDR display should be tone mapped. Also, the viewing condition of the experiment setup will also greatly affect the perception of the image on the HDR display. All these factors might introduce noise to the result. This method is less complex. Some studies that have used this approach are [Ledda et al., 2005] and [Kuang et al., 2010].

No reference: In this approach, subjects are asked to evaluate the tone mapped image without providing a reference image. The arrangement of this evaluation method is simple and easy to setup. Many studies have used this approach, e.g. [Delahunt et al., 2005], [Čadík et al., 2008], [Ashikhmin and Goyal, 2006] and [Yoshida et al., 2006]. This method is completely based on user preferences and additional measures have to be taken to deal with large inter-subject variability.

None of the above approaches is perfect. Subjective evaluation is limited in many ways: firstly, it is often time consuming and expensive; secondly, it is difficult to incorporate it in the design of tone mapping algorithms. This points out the importance of objective evaluation of TMOs. An accurate objective evaluation should mimic the subjective evaluation described above, so it requires the use of a perceptual metric between images of different dynamic range.

The first tone mapping metric was proposed in [Smith et al., 2006], where the authors estimated the local and global contrast changes present in a tone mapped image when compared to the HDR original. The global contrast change is estimated by analyzing the tone mapping curve, which may be computed by linear

regression in the brightness domain. The local contrast change is estimated by identifying the visible details in both the HDR and LDR images and by computing the magnitude of the visibility difference between them.

The Dynamic range independent metric (DRIM) [Aydin et al., 2008] compares in a perceptual manner images of any dynamic range. It aims at predicting details (contrast) changes between two images. More precisely, the purpose of this metric is to consider the perception that a viewer would have of both images relying on psychophysical data, and to estimate at each pixel the probabilities of distortions between the two images. The metric considers three types of distortions:

- Loss of Visible Contrast (LVC), meaning that contrast is visible in one image (called the reference image) and not in the second one (called test image).
- Amplification of Invisible Contrast (AIC), when details that were not in the reference image appear in the test image.
- Contrast reversal (INV), meaning that contrast is visible in both reference and test images but with different polarity (i.e. the same contour has the gradient in opposite directions in the two images.)

The metric first considers the two input images separately and estimates the pixel-wise probability of contrast to be visible ($P^{k,l}_{./v}$) and the probability of contrast not to be visible ($P^{k,l}_{./i}$) for each image for each of several bands of orientation *l* and spatial frequency *k*, using models for processes in the human visual system [Mantiuk et al., 2005, Watson, 1987, Daly, 1992].

Then the distortions LVC, AIC and INV are estimated for each pair of bands as:

$$P^{k,l}_{LVC} = P^{k,l}_{r/v} \times P^{k,l}_{t/i} \qquad P^{k,l}_{AIC} = P^{k,l}_{r/i} \times P^{k,l}_{t/v} \qquad P^{k,l}_{INV} = P^{k,l}_{r/v} \times P^{k,l}_{t/v} \times R^{k,l}. \tag{4.16}$$

The subscript *r/.* and *t/.* indicate reference and test images. If the polarity of contrast in the test and reference images is the same then R is set to 1, else to 0. The final LVC output of the metric is the probability of detecting a LVC distortion in any band, and is obtained by combining the probabilities over all the

bands, according to the formula

$$LVC = 1 - \prod_{k=1}^{N} \prod_{l=1}^{M} (1 - P_{LVC}^{k,l})$$

where N and M are the number of orientations and spatial bands, respectively. AIC and INV are computed in a similar way.
Note that for applications to tone mapping evaluation, the reference image is the luminance map of an HDR image and the test image is the luminance map of its LDR tone mapped version.
The metric also provides a "distortion map" (a color image) to visualize pixel-wise distortions with the following color code: green hue represents LVC, blue hue stands for AIC and red hue indicates INV, the saturation encoding the magnitude of the corresponding distortion, whereas the intensity corresponds to the intensity of the reference image. At each pixel, the maximum of the three distortions is computed, and the corresponding color is displayed. If the maximum is lower than 0.5, then the saturation is set to 0.
These three types of distortions estimated by the metric DRIM are illustrated on Figure 4.12. On the left column, we show two LDR gray-level images. We apply some filters to these images in order to create distortions. To the image located in the top row, we apply a Gaussian smoothing on its top-right part and unsharp masking on its top-left part in order to obtain respectively a contrast reduction and a contrast enhancement. To the image located in the bottom row, we apply some contrast reversal technique on the pattern in the chair. The images resulting from these filters are shown in the middle column. Then, we compute the metric DRIM where the reference images are the original images on the left column and the test images are their distorted versions in the middle column. The distortion maps, shown in the right column, are consistent with the artificial distortions we created.

Tone mapping quality index (TMQI) [Yeganeh and Wang, 2013] is a tone mapping metric inspired by the Structural Similarity Index (SSIM) [Wang et al., 2004]. The structural fidelity (S) measure between the HDR and its tone mapped version is computed in the same way as in SSIM, but with a modification of the contrast comparison component in order to allow comparison between images of different dynamic range; also, the luminance component is removed. The contrast comparison component penalizes only if the contrast is visible in the HDR

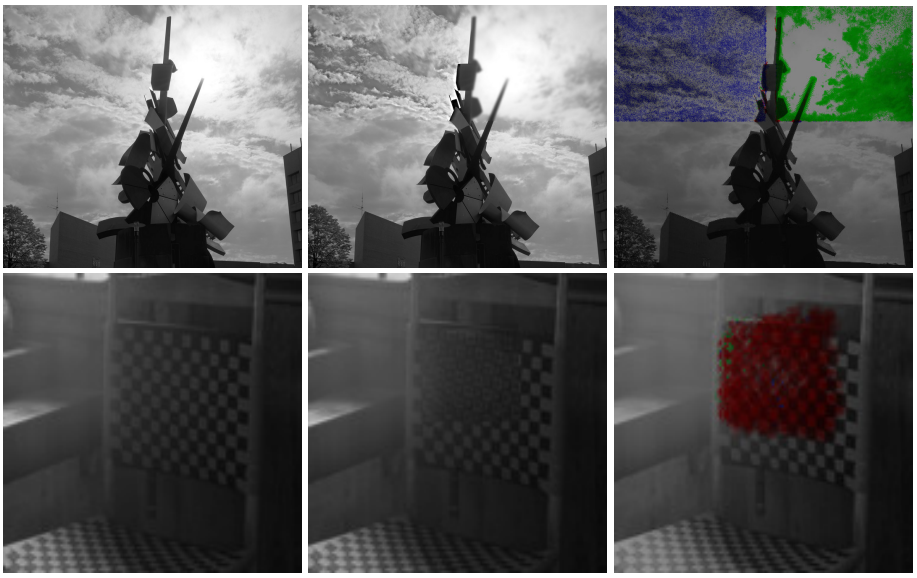


Figure 4.12: Distortion map illustration. Left: reference images. Middle: test images. Right: distortion maps.

image and is invisible in the tone mapped version, or the other way around. The metric provides a single overall quality score of the tone mapped image as well as a multi-scale quality map.

The metric also yields a naturalness measure (N) based on the similarity of the histogram of the tone mapped image with the histogram of 3,000 LDR images of a database that they consider to evaluate statistics of images. The overall quality (Q) is computed as a weighted average of S and N, thus the higher the values, the better the tone mapping result is.

4.4 Conclusion

In this chapter, we have briefly explained various display technologies and their dynamic ranges. The dynamic range of displays varies from 2 orders of magnitude in the case of standard displays up to 4 orders of magnitude for HDR displays. We then discuss about various tone mapping techniques and classify

them into global, local and global + local methods. We later describe various approaches to adapt TMOs for videos and also for faithful color reproduction. Finally, we discuss about various tone mapping evaluation approaches and their limitations.

Chapter 5

OPTIMIZING TONE MAPPED IMAGES

In this chapter we present our approach for the improvement of tone mapped images that takes into account the given metric that will be used for the evaluation. In such a way, we are able to improve an existing TMO with respect to a given metric. The aim of this work is two-fold:

1. We develop a general framework for improving any tone mapped image by reducing the distance with the corresponding HDR image. More precisely, assuming that the metric met is designed to compare images of different dynamic range, our approach for reducing the distance between an HDR image $H\colon \Omega \longrightarrow \mathbb{R}$ and its tone mapped version $L_0 \in LDR(\Omega)$ is to minimize functionals of the form

$$\underset{L\in LDR(\Omega)}{\arg\min} \int_{\Omega} \Phi(met(L,H)(x))\,dx \tag{5.1}$$

where $\Phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^+$, through a gradient descent algorithm of initial condition L_0. Then the main task to implement the algorithm is to construct a discrete approximation of the functional derivative

$$\frac{\partial met(L,H)(x)}{\partial L(y)} \tag{5.2}$$

2. We apply the general framework described in Eq. 5.1 in the context of the perceptual metric DRIM [Aydin et al., 2008]. More precisely, we consider the

following particular cases of Eq. 5.1

$$\underset{L \in LDR(\Omega)}{\arg\min} \int_{\Omega} \|DRIM(L,H)(x)\|_k \, dx \tag{5.3}$$

where k is a strictly positive constant. We evaluate our method by comparing in a two-fold way the initial condition with the steady-state of the gradient descent algorithm. Our tests show that we are able to improve the tone mapping results of several state of the art approaches by a significant amount.

This chapter is organized as follows. We initially present our general framework for improving any tone mapped image as a gradient descent algorithm associated to a non local variational problem. In particular, we give the expression of the gradient of the functional to minimize. We then consider the case of perceptual metrics and detail the computation of the discrete functional derivative Eq. 5.2 from which derives the discrete gradient of the functional Eq. 5.1. At last, we test the corresponding gradient descent algorithm on several tone mapped images dealing with the perceptual metric DRIM. The content of this chapter is derived from the work published in [Cyriac et al., 2014] [Cyriac et al., 2013].

5.1 General Framework for Improving Tone Mapping Results

5.1.1 Distance between Images as a Non local Operator

Many tasks in image processing and computer vision require a validation by comparing the result with the original data, e.g. optical flow estimation, image denoising, tone mapping. Whereas measures based on pixel-wise comparisons (e.g. MSE, SNR, PSNR) are suitable to evaluate algorithms for problems such as image denoising and optical flow estimation, they are not relevant to evaluate tone mapping results, because the ground truth is not known. Indeed, tone mapping results are evaluated by comparisons with the original HDR images, where pixel-wise comparisons are not suitable since the images compared are of different dynamic range. This leads us to the following definitions, where we first remind the concept of functional differentiation.

Definition 5.1.1 (functional derivative) *Let X be a Banach space of scalar-valued functions defined on some domain Ω. The derivative of $F\colon X \longrightarrow$*

$\mathbb{R}^n, n \geq 1$, *with respect to the function* $I \in X$ *at the point* $y \in \Omega$ *is the quantity*

$$\frac{\partial F(I)}{\partial I(y)} \colon = \lim_{\epsilon \to 0} \frac{F(I + \epsilon \delta_y) - F(I)}{\epsilon} \tag{5.4}$$

where δ_y *is the Dirac delta function concentrated at the point* y*.*

The differential of F *in the direction* $\varphi \in X$ *is then*

$$\delta F(I; \varphi) \colon = \int_\Omega \varphi(y) \frac{\partial F(I)}{\partial I(y)} \, dy \tag{5.5}$$

From now on, we consider the set of gray-level images defined on a domain $\Omega \subset \mathbb{R}^2$ as the Banach space $X \colon = C^\infty(\Omega; [0, 1])$.

Definition 5.1.2 (metric) *Let* Y *be the Banach space of vector-valued smooth functions defined on* Ω*. We call* ***metric*** *an operator* ***met***$\colon X \times X \longrightarrow Y$*.*
For $x \in \Omega$*, we consider* $F \colon X \times X \longrightarrow \mathbb{R}^n, n \geq 1$*, defined by* $F(L, H) = met(L, H)(x)$ *and the partial derivatives of* F *with respect to* L *and* H *at the point* $y \in \Omega$*, i.e.*

$$\frac{\partial met(L, H)(x)}{\partial L(y)} \colon = \lim_{\epsilon \to 0} \frac{met(L + \epsilon \delta_y, H)(x) - met(L, H)(x)}{\epsilon} \tag{5.6}$$

$$\frac{\partial met(L, H)(x)}{\partial H(y)} \colon = \lim_{\epsilon \to 0} \frac{met(L, H + \epsilon \delta_y)(x) - met(L, H)(x)}{\epsilon} \tag{5.7}$$

where δ_y *is the Dirac delta function concentrated at the point* $y \in \Omega$*.*
We say that met *is*
(i) ***pixel-wise*** *if* $\forall L, H \in X$*,* $\forall x, y \in \Omega, y \neq x$*, the quantities Eq. 5.6 and Eq. 5.7 vanish.*
(ii) ***non local*** *if* $\forall L, H \in X$*,* $\forall x \in \Omega, \exists\, \Omega \supset \mathcal{N}(x) \ni x, \mathcal{N}(x) \neq \{x\}, \exists y \notin \mathcal{N}(x)$ *s.t. the quantities Eq. 5.6 and Eq. 5.7 do not vanish.*

Definition 5.1.3 (distance) *A* ***distance*** *associated to the metric* ***met*** *is a functional* $E \colon X \times X \longrightarrow \mathbb{R}^+$ *of the form*

$$E \colon (L, H) \longmapsto \int_\Omega \Phi(met(L, H)(x)) \, dx \tag{5.8}$$

for some map $\Phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^+$.
We say that the distance E *is pixel-wise, resp. non local, if the associated metric is pixel-wise, resp. non local.*

Note that we only require the distance (Eq. 5.8) to satisfy the non-negativity property among the classical properties of a distance function (zero for $L = H$, symmetry, triangle inequality).

In this context, we can classify image quality measures into several categories. The set of pixel-wise distances includes MSE, PSNR and SNR measures, where the function Φ in Eq. 5.8 is the Euclidean norm $\| \|_2$. Indeed, the corresponding terms $met(L + \epsilon\delta_y, H)(x) - met(L, H)(x)$ and $met(L, H + \epsilon\delta_y)(x) - met(L, H)(x)$ in Eqs. 5.6, 5.7 vanish for $y \neq x$ since such quantities only depend on the values $L(x)$ and $H(x)$.
The second category gathers the metrics that are non local. This category can actually be divided into two sub-categories: the set of metrics that compare images of same dynamic range (see e.g. [Daly, 1992] for LDR images and [Mantiuk et al., 2005],[Mantiuk et al., 2011] for HDR images), and the set of metrics comparing images of different dynamic range (see e.g. [Smith et al., 2006]). The metric DRIM [Aydin et al., 2008] belongs to both sub-categories since it is independent of the dynamic range of the images it compares.
At last, there exist image quality measures that are neither metrics nor distances according to the definitions 5.1.2 and 5.1.3. This includes for instance the SSIM metric [Wang and Bovik, 2002].

5.1.2 Reduction of the Distance between an HDR Image and its Tone Mapped Version

In practice, LDR images are usually encoded in 8 bits whereas HDR images are often encoded in 16 bits. This leads us to define the functional spaces $LDR(\Omega)$ and $HDR(\Omega)$ as follows

$$HDR(\Omega) = \{f \in C^\infty(\Omega; [1/65536, 1])\} \tag{5.9}$$

$$LDR(\Omega) = \{f \in C^\infty(\Omega; [1/256, 1])\} \tag{5.10}$$

In particular, we have $LDR(\Omega) \subset HDR(\Omega)$.

We consider the following problem: given an HDR image H and a metric met, we aim at constructing the LDR image L_* minimizing the distance with H, i.e. we aim at solving the following variational problem

$$L_* : = \underset{L \in LDR(\Omega)}{\arg\min}\ E(L, H) \tag{5.11}$$

Proposition 5.1.4 *Assuming that met is continuous, bounded, and Φ is continuous, the variational problem (Eq. 5.11) has a solution.*

Proof. Under the assumption of the Proposition, the energy E is bounded since the domain Ω of an image is a compact subset of $\mathbb{R}^2$. Moreover, the set $LDR(\Omega)$ being closed, we deduce that there exists a function $L^* \in LDR(\Omega)$ (not necessarily unique) solution of the variational problem Eq. 5.11. □

Proposition 5.1.5 *Assuming that met and Φ are differentiable, the functions L which are critical points of the energy*

$$E_H : L \longmapsto E(L, H) \tag{5.12}$$

satisfy

$$\int_\Omega \delta\Phi \left(met(L,H)(x); \frac{\delta met(L,H)(x)}{\delta L(y)} \right) dx = 0 \qquad \forall y \in \Omega$$

Proof. The functional E_H being differentiable on the whole set $LDR(\Omega)$, its critical points are the functions L where its gradient ∇E_H vanishes.
Let $\psi : \Omega \longrightarrow \mathbb{R}$ be a compact support function. We compute the differential δE_H of the energy E_H at a function L in the direction ψ. We have

$$\begin{aligned}
\delta E_H(L;\psi) &= \delta E((L,H);(\psi,0)) \\
&= \int_\Omega \delta \left(\Phi \circ met\right)((L,H);(\psi,0))(x)\, dx \\
&= \int_\Omega \delta\Phi \left(met(L,H)(x); \delta\, met\left((L,H);(\psi,0)\right)(x)\right) dx \\
&= \int_{\Omega^2} \delta\Phi \left(met(L,H)(x); \frac{\delta met(L,H)(x)}{\delta L(y)} \right) \psi(y)\, dx\, dy
\end{aligned}$$

The last equality results from Eq. 5.5. At last, as ψ has compact support, $\delta E_H(L;\psi) = 0 \Longrightarrow$

$$\int_\Omega \delta\Phi\left(met(L,H)(x); \frac{\delta met(L,H)(x)}{\delta L(y)}\right) dx = 0 \qquad \forall y \in \Omega \tag{5.13}$$

□

The gradient of the functional E_H at the function L is the map

$$\nabla E_H(L)\colon y \longmapsto \int_\Omega \delta\Phi\left(met(L,H)(x); \frac{\delta met(L,H)(x)}{\delta L(y)}\right) dx \tag{5.14}$$

Due to the complexity of the metrics that compare images of different dynamic range, the non local operators met encoding the metrics lack mathematical properties like convexity. As a consequence, it is hard to prove the uniqueness or not of the solution and establish accurate numerical schemes for reaching the solution(s) of the variational problem Eq. 5.11.
For this reason, we adopt the following approach: instead of constructing the best tone mapped image(s) of a given HDR image H with respect to the metric met by constructing solution(s) of the variational problem Eq. 5.11, we improve existing tone mapping results by the use of a gradient descent algorithm where the initial condition L_0 is a tone mapped version of H.
Even if the gradient descent algorithm might converge towards critical points L_* that are not global minima, we have $E_H(L_*) < E_H(L_0)$ meaning the LDR image L_* is an improvement of the tone mapping result L_0.

The main task to compute a discrete approximation of the gradient Eq. 5.14 in order to perform the gradient descent algorithm is to compute accurate discrete approximations of the functional derivative

$$\frac{\partial met(L,H)(x)}{\partial L(y)} :\ = \lim_{\epsilon\to 0} \frac{met(L+\epsilon\delta_y,H)(x) - met(L,H)(x)}{\epsilon}$$

where δ_y is the Dirac delta function concentrated at the point $y \in \Omega$.
Our proposal is to make use of central differences of the form

$$\frac{\partial met^D(L,H)(a,b)}{\partial L(i,j)} :\ = \frac{met(L+\epsilon_1\,\delta_{(i,j)},H)(a,b) - met(L-\epsilon_2\,\delta_{(i,j)},H)(a,b)}{d(L+\epsilon_1\,\delta_{(i,j)}, L-\epsilon_2\,\delta_{(i,j)})} \tag{5.15}$$

for some well-chosen ϵ_1, ϵ_2, and where $d(L + \epsilon_1\, \delta_{(i,j)}, L - \epsilon_2\, \delta_{(i,j)})$ measures a difference between the two images $L + \epsilon_1\, \delta_{(i,j)}$ and $L - \epsilon_2\, \delta_{(i,j)}$.
A straightforward choice would be to impose ϵ_1, ϵ_2 to be constant and independent of both the intensity values $L(i, j)$ and the metric met.

5.2 Discrete Functional Derivatives in the case of Perceptual Metrics

We argue that the increments ϵ_1, ϵ_2 in Eq. 5.15 should be dependent on the nature of the metric. In particular, dealing with perceptual metrics, we claim that ϵ_1, ϵ_2 should be chosen to make the difference between the images L and $L + \epsilon_1\delta_{(i,j)}$ be perceived equally to the difference between the images L and $L - \epsilon_2\delta_{(i,j)}$.
Our proposal is to make use of the concept of Just Noticeable Difference (JND) in order to compute perceived differences between images. However, perceptual metrics do not necessarily compare images in perceptually uniform spaces, meaning that we need to express the JND into the space in which the metric is operating, which is the luminance space in most of cases.

5.2.1 Approximation of the JND in the Luminance Space

The luminance space Y is not perceptually uniform in the sense that the difference between two luminance values is not proportional to the difference of light intensity observed. In 1976, the CIE introduced a measure of the perception of the light, called lightness, as the quantity

$$L^* = \begin{cases} 903.3\,\dfrac{Y}{Y_n} & \text{if} \quad \dfrac{Y}{Y_n} \leq 0.008856 \\ \\ 116\left(\dfrac{Y}{Y_n}\right)^{1/3} - 16 & \text{if} \quad \dfrac{Y}{Y_n} > 0.008856 \end{cases} \tag{5.16}$$

where Y_n is the highest luminance value of the scene. The quantity Y/Y_n is called relative luminance. The space $\mathbf{L}^*$ is approximately perceptually uniform since a difference of 1 in the space L^* approximates pretty well 1 JND (see the book of Poynton [Poynton, 2012] for details about this and other of the following technical concepts).
Formula (Eq. 5.16) makes the lightness have a range of 0 to 100. When scaled

to the range $[0, 1]$, L^* can be approximated by the 0.4-power of the relative luminance, i.e. we have the relation

$$0.01L^* \simeq \left(\frac{Y}{Y_n}\right)^{0.4} \tag{5.17}$$

When shooting a scene, a digital camera captures the light intensity from which it encodes R, G, B values. Standard digital cameras also perform gamma correction:

$$R' = R^{\frac{1}{\gamma}} \qquad G' = G^{\frac{1}{\gamma}} \qquad B' = B^{\frac{1}{\gamma}} \tag{5.18}$$

Assuming that these values are encoded in the sRGB color space (which is the standard color space used in the broadcast and computer industries), γ is approximately 2.2.
Then the relative luminance Y/Y_n of an image displayed by a monitor is given by

$$\frac{Y}{Y_n} \simeq 0.2126\, R'^{\gamma} + 0.7152\, G'^{\gamma} + 0.0722\, B'^{\gamma} \tag{5.19}$$

From Eq. 5.19 and Eq. 5.17, we end up with the formula

$$L^* \simeq (0.2126\, R'^{\gamma} + 0.7152\, G'^{\gamma} + 0.0722\, B'^{\gamma})^{0.4} \tag{5.20}$$

(assuming that the lightness is normalized to the range [0,1]).

Finally, using the formula (Eq. 5.20) and the perceptual uniformity of the lightness L^*, we can express the JND for the light intensity of a color (R', G', B') displayed at a screen. An alternative approach to approximate the JND in luminance space can be found in [Reinhard et al., 2010] (Chapter 10, Section 10.4.2)

5.2.2 Computation of the terms ϵ_1, ϵ_2, d

We assume that the metric met in Eq. 5.15 is operating in the luminance space (e.g. the metric DRIM [Aydin et al., 2008]).
In order to relate the increments ϵ_1, ϵ_2 as well as the measure d in Eq. 5.15 to the JND, we have to face the following issue: the JND is determined on a uniform background whereas expression (Eq. 5.15) deals with image pixels. In order that the computation of the JND at an image pixel makes more sense, our proposal is to compute it using a smoothed version of the image L

$$\tilde{L} := L * G_\sigma \tag{5.21}$$

where G_σ is the Gaussian kernel associated to some variance σ. Indeed, convolving an image with a Gaussian kernel reduces its variations, making the smoothed image be locally closer to a uniform patch than the original image.
We can then express the JND for the light intensity perceived at the pixel (i, j) of the image L using the Eq. 5.17 and the smoothed version $\tilde{L}$ of L as follows. As a difference of 0.01 in the (normalized) lightness space approximates pretty well 1 JND, we aim at solving the two functional equations

$$(\tilde{L} + \epsilon_1)^{0.4} - \tilde{L}^{0.4} = 0.01 \tag{5.22}$$

$$\tilde{L}^{0.4} - (\tilde{L} - \epsilon_2)^{0.4} = 0.01 \tag{5.23}$$

whose solutions are

$$\epsilon_1 = (\tilde{L}^{0.4} + 0.01)^{2.5} - \tilde{L} \tag{5.24}$$

$$\epsilon_2 = \tilde{L} - (\tilde{L}^{0.4} - 0.01)^{2.5} \tag{5.25}$$

However, defining ϵ_1, ϵ_2 as in Eq. 5.24, 5.25 yields an extra issue: we might have $L(i, j) + \epsilon_1(i, j) > 1$ or $L(i, j) - \epsilon_2(i, j) < 0$. In such a case, the image $L+\epsilon_1(i, j)\delta_{i,j}$ or $L-\epsilon_2(i, j)\delta_{i,j}$ and consequently Eq. 5.15 would not be defined. We then relax this perceptual uniformity paradigm in order to stay in the range $[0, 1]$. We end up with the following two extensions of Eq. 5.22 and Eq. 5.23

$$(\tilde{L} + \epsilon_1)^{0.4} - \tilde{L}^{0.4} = \min(0.01, 1 - \tilde{L}^{0.4}) \tag{5.26}$$

$$\tilde{L}^{0.4} - (\tilde{L} - \epsilon_2)^{0.4} = \min(0.01, \tilde{L}^{0.4}) \tag{5.27}$$

whose solutions are

$$\epsilon_1 \colon = (\tilde{L}^{0.4} + \min(0.01, 1 - \tilde{L}^{0.4}))^{2.5} - \tilde{L} \tag{5.28}$$

$$\epsilon_2 \colon = \tilde{L} - (\tilde{L}^{0.4} - \min(0.01, \tilde{L}^{0.4}))^{2.5} \tag{5.29}$$

Hence, the image $L + \epsilon_1(i, j)\delta_{i,j}$ never exceeds 1 since $\tilde{L} \leq L$ by construction. For the same reason, $L - \epsilon_2(i, j)\delta_{i,j}$ never gets negative.
Finally, we set

$$d = min(0.01, 1 - \tilde{L}^{0.4}) + min(0.01, \tilde{L}^{0.4}) \tag{5.30}$$

where $min(0.01, 1 - \tilde{L}^{0.4})$ can be viewed as a perceptual distance between the images L and $L + \epsilon_1(i,j)\delta_{i,j}$ and $min(0.01, \tilde{L}^{0.4})$ a perceptual distance between the images L and $L - \epsilon_2(i,j)\delta_{i,j}$. The summation of the two terms makes the quantity d represent a perceptual distance between the images $L + \epsilon_1(i,j)\delta_{i,j}$ and $L - \epsilon_2(i,j)\delta_{i,j}$.

5.3 Tests with the Perceptual Metric DRIM

As discussed in Section 4.3, the metric DRIM [Aydin et al., 2008] compares in a perceptual manner images of any dynamic range. It aims at predicting details (contrast) changes between two images. More precisely, the purpose of this metric is to consider the perception that a viewer would have of both images relying on psychophysical data, and to estimate at each pixel the probabilities that distortions between the two images appear.
As the output of the metric DRIM encodes at each pixel the probabilities of detecting the distortions LVC, AIC, INV between two images, we reinterpret the metric DRIM as follows.

Definition 5.3.1 (Dynamic range independent perceptual metric) *Let $L, H : \Omega \longrightarrow \mathbb{R}$ be two images of any dynamic ranges, the perceptual metric DRIM between L and H is the map*

$$\begin{array}{rccc} & \Omega & \longrightarrow & [0,1]^3 \\ DRIM(L,H): & x & \longmapsto & (LVC(L,H)(x), AIC(L,H)(x), INV(L,H)(x)) \end{array} \tag{5.31}$$

From the perceptual metric (Eq. 5.31) we define a set of perceptual distances between images of any dynamic range.

Definition 5.3.2 (Dynamic range independent perceptual distances) *Let k be a strictly positive number. The perceptual distance $E_k(L,H)$ between L and H is*

$$E_k(L,H) := \|DRIM(L,H)\|_{L^k} \tag{5.32}$$

In others words, we have

$$E_k(L,H) = \int_\Omega (LVC^k(L,H)(x) + AIC^k(L,H)(x) + INV^k(L,H)(x))^{1/k} \; dx \tag{5.33}$$

Note that both definitions of metric and distance are coherent with the definitions 5.1.2 and 5.1.3 in Section 5.1.1.
Moreover, the distances E_k are symmetric since $LVC(L,H) = AIC(H,L)$ and $INV(L,H) = INV(H,L)$ according to Eq. 4.16.

5.3.1 Reducing a Dynamic Range Independent Perceptual Distance between an HDR Image and its Tone Mapped Version

In this Section, we detail the gradient descent algorithm mentioned in the end of Section 5.1.1 in the context of the distances (Eq. 5.32).

Expression of the gradient of the distance

We first show the existence of a solution of the minimization problem related with the distances (Eq. 5.32).

Proposition 5.3.3 *Let H be an HDR image. Assuming that the metric DRIM of definition 5.3.1 is continuous, the variational problem*

$$\underset{L \in LDR(\Omega)}{\arg\min} \; E_k(L,H) \tag{5.34}$$

has a solution.

Proof. Each distortion being in the range $[0,1]$, the L^k norms being continuous, and the space $LDR(\Omega)$ being closed, Proposition 5.3.3 appears to be a particular case of Proposition 5.1.4. □

Proposition 5.3.4 *Assuming that H is non constant and the metric DRIM is differentiable on the space $LDR(\Omega)$, the critical points of the energy*

$$E_{kH} \colon L \longrightarrow E_k(L,H) \tag{5.35}$$

satisfy

$$\int_{\Omega} \frac{LVC^{k-1}\frac{\partial LVC}{\partial L(y)} + AIC^{k-1}\frac{\partial AIC}{\partial L(y)} + INV^{k-1}\frac{\partial INV}{\partial L(y)}}{(LVC^k + AIC^k + INV^k)^{1-1/k}}\, d\Omega = 0 \qquad (5.36)$$

$\forall y \in \Omega$.

For the sake of brevity, we drop the terms in the parenthesis (i.e., LVC instead of $LVC(L, H)$) from here on.
The proof of the Proposition relies upon the following postulate: the metric DRIM only vanishes when the inputs are two identical constant images.
Even if the postulate is counter-intuitive, it holds because the metric DRIM first treats the two images separately. More precisely, it estimates the probabilities that the contrast is visible or not in each image and then combines the results in order to determine the distortions LVC, AIC, INV between the two images. Hence, even if the images are identical, the fact that the images are first treated separately makes the distortions be 0 only if the probability that the contrast is visible is 0 at each pixel of both images, i.e. if both images are constant. All the experiments we have run confirm this analysis.

Proof. The energy E_{kH} is differentiable on the whole set $LDR(\Omega)$ since the metric DRIM does not vanish for non constant images (postulate). Then, according to Prop. 5.1.5, the critical points of Eq. 5.35 satisfy

$$\int_{\Omega} \delta\| \, \|_k \left(DRIM(L,H)(x); \frac{\delta DRIM(L,H)(x)}{\delta L(y)} \right) dx = 0 \qquad \forall y \in \Omega \qquad (5.37)$$

Finally, expression (Eq. 5.36) follows from Eq.5.37 and

$$\delta\| \, \|_k((u_1,u_2,u_3);(v_1,v_2,v_3)) = \frac{u_1^{k-1}v_1 + u_2^{k-1}v_2 + u_3^{k-1}v_3}{\|(u_1,u_2,u_3)\|_k{}^{k-1}}$$

□

The term $\nabla E_{kH}(L) \colon y \longmapsto$

$$\int_{\Omega} \frac{LVC^{k-1}\frac{\partial LVC}{\partial L(y)} + AIC^{k-1}\frac{\partial AIC}{\partial L(y)} + INV^{k-1}\frac{\partial INV}{\partial L(y)}}{(LVC^k + AIC^k + INV^k)^{1-1/k}}\, d\Omega \qquad (5.38)$$

is the gradient of the energy E_{kH} at the function L.

The discrete gradient descent algorithm

The discrete approximation of the gradient descent algorithm in the context of the distances (Eq. 5.32) is of the form

$$L_{n+1} = L_n - \alpha_n \nabla^D E_{kH}(L_n), \quad L_0 = TMO(H) \tag{5.39}$$

In what follows, we detail the expressions of the discrete gradient $\nabla^D E_{kH}$, the initial condition L_0 and the steps α_n.

The discrete gradient operator We express the discrete gradient $\nabla^D E_{kH}$ as

$$\nabla^D E_{kH}(L) \colon (i,j) \longmapsto \sum_{(a,b)\in\Omega} \frac{LVC^{k-1}\frac{\partial LVC^D}{\partial L(i,j)} + AIC^{k-1}\frac{\partial AIC^D}{\partial L(i,j)} + INV^{k-1}\frac{\partial INV^D}{\partial L(i,j)}}{(LVC^k + AIC^k + INV^k)^{1-1/k}} \tag{5.40}$$

Note that we have omitted the dependence of the terms on the pixels (a, b) for the sake of brevity.
The discrete functional derivatives of the distortions LVC, AIC, INV with respect to L at the pixels (i, j) are computed using Eq. 5.15. For instance, dealing with the distortion LVC, we have

$$\frac{\partial LVC^D}{\partial L(i,j)} \colon = \frac{LVC(L + \epsilon_1(i,j)\,\delta_{(i,j)}, H) - LVC(L - \epsilon_2(i,j)\,\delta_{(i,j)}, H)}{d(L + \epsilon_1(i,j)\,\delta_{(i,j)}, L - \epsilon_2(i,j)\,\delta_{(i,j)})} \tag{5.41}$$

and the values $\epsilon_1, \epsilon_2, d$ are determined by Eq. 5.28, Eq. 5.29 and Eq. 5.30 respectively.

The step The images produced at each iteration of the gradient descent (Eq. 5.39) are not necessarily LDR images, but as the metric DRIM is designed for images of any dynamic range, the numerical scheme is well defined if the values of L_{n+1} are in the range $[0, 1]$ at each iteration which can be achieved if α_n is chosen small enough.

We then determine α_n through a line search strategy that will force the values of the image to be in the range $[0, 1]$. The pseudocode for the computation of α_n is the following

Pseudocode 1 computation of α_n

Initializations: $\lambda = 0, E_{new} = E_{kH}(L_n), E_{old} = E_{kH}(L_n) + 1,\ L_{temp} = L_n$
while $E_{new} < E_{old}$ and $L_{temp}(i, j) \in [0, 1]\, \forall (i, j) \in \Omega$ **do**
$\lambda = \lambda + 0.001$
$E_{old} = E_{new}$
$L_{temp} = L_n - \lambda \nabla^D E_{kH}(L_n)$
$E_{new} = E_{kH}(L_{temp})$
end while
$\alpha_n = \lambda - 0.001$

Initial condition and output image of the gradient descent As the perceptual metric involved is the metric DRIM, the function L_0 should be the light intensity produced by a screen when displaying the tone mapped image TMO(H), which can be obtained as follows.
Denoting by R_0, G_0, B_0 the inverse gamma corrected components (see Section 5.2.1) of the tone mapped image TMO(H) in the sRGB color space, we transform TMO(H) into XYZ color coordinates with the formula

$$\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{pmatrix} \begin{pmatrix} R_0 \\ G_0 \\ B_0 \end{pmatrix} \tag{5.42}$$

As mentioned in Section 5.2, the component Y_0 is an approximation of the light intensity produced when displaying a color image on a monitor. Hence, we take Y_0 as the initial condition L_0 of the gradient descent (Eq. 5.39).

In order to make to final output of our algorithm be an LDR image, we project the steady-state L_* of the gradient descent onto the discrete space of LDR images $LDR(\Omega)$ as follows.

Assuming that the domain Ω of the HDR image H is of size $M \times N$, we define the discrete spaces $HDR(\Omega)$ and $LDR(\Omega)$ as matrices of the form

$$HDR(\Omega) = \{A \in \mathcal{M}(M,N),\ A_{i,j} = \frac{k}{65536},\ k \in [1, 65536] \cap \mathbb{N}\} \quad (5.43)$$

$$LDR(\Omega) = \{A \in \mathcal{M}(M,N),\ A_{i,j} = \frac{k}{256},\ k \in [1, 256] \cap \mathbb{N}\} \quad (5.44)$$

Then, the projection P^D onto the set $LDR(\Omega)$ consists, for $A \in \mathcal{M}(M,N)$ with values in the range $[0,1]$, in clipping and quantizing into the usual 8 bits domain as

$$P^D A_{\,i,j} = \begin{cases} \dfrac{1}{256} & \text{if} \quad A_{i,j} \leq \dfrac{1}{256} \\ \dfrac{\lfloor 256 \times A_{i,j} \rfloor}{256} & \text{if} \quad \dfrac{1}{256} \leq A_{i,j} \leq 1 \end{cases} \quad (5.45)$$

where $\lfloor\ \rfloor$ is the floor function.
Finally, the output of the algorithm is the color image of components (R_*, G_*, B_*) defined by

$$\begin{pmatrix} R_* \\ G_* \\ B_* \end{pmatrix} = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{pmatrix}^{-1} \begin{pmatrix} X_0 \\ P^D L_* \\ Z_0 \end{pmatrix}$$

Preprocessing

To increase the chance that the gradient descent algorithm (Eq. 5.39) does not stop at a local minimum of the energy E_{k_H} too close to the initial condition L_0, we apply a preprocessing on L_0 in order to get an initial condition L_{new} of the algorithm (Eq. 5.39) that is closer (in terms of perceptual distance) to H than L_0. The method we propose relies on the intuition that high values of LVC might be reduced by application of local sharpening whereas high values of AIC might be reduced by local Gaussian blurring. Hence we perform local Gaussian blurring and unsharp masking (Eq. 5.46) (see [Adams et al., 2012] for details) to L_0

depending on the values of the function $LVC(L_0, H) - AIC(L_0, H)$. More precisely, denoting by L_{smooth} a blurred version of L_0 and defining L_{sharp} as

$$L_{sharp} = L_0 + \alpha\,(L_0 - L_{smooth}) \tag{5.46}$$

for some constant α, we define the image L_{new} as

$$L_{new}(i,j) = \begin{cases} L_{sharp}(i,j) & \text{if} \quad LVC(L_0,H) - AIC(L_0,H)(i,j) > 0 \\ (1-\beta)L_0(i,j) + \beta L_{smooth}(i,j) & \text{if} \quad LVC(L_0,H) - AIC(L_0,H)(i,j) < 0 \end{cases} \tag{5.47}$$

Experiments on different tone mapped images described in the next Section show that the distance with H is indeed reduced as well as the errors in the distortions map.

Numerical scheme

We summarize in this Section the different steps of our algorithm.

1 Let H be an HDR image and L_0 be a tone mapped version of H. As the metric DRIM only takes into account the luminance information of the input images, we first convert them into luminance maps: we apply the transformation (Eq. 5.42) on the LDR image L_0 and extract the luminance channel Y_0, and use the function *pfs_read_luminance*[Mantiuk and Heidrich, 2009] to extract the luminance information of H.

2 We perform the preprocessing described in Section 5.3.1 on the image Y_0 with the following parameters: the variance σ of the Gaussian smoothing kernel is set to 0.62, and the constants α, β are respectively 0.7 and 0.5. These values provide good results and have been fixed for all the experiments in this chapter.

3 We apply the gradient descent algorithm (Eq. 5.39) where the initial condition is the output of the preprocessing. We test different values for k and different domains for the summation in the expression of the discrete gradient operator (Eq. 5.40). Indeed, because the variational problem we propose is non local, the gradient operator is an integral operator, meaning that its computation might be very time-consuming. In order to decrease the execution time of the algorithm, we adopt the following two

approaches: we consider 50×50 and 200×200 neighborhoods, and make use of computers equipped with multiple cores to compute the gradient operator over the whole image domain. The algorithm stops when the energy does not decrease anymore, i.e. when we reach a local or global minimum.

The pseudocode of the gradient descent algorithm is given below.

Pseudocode 2 Gradient descent for optimizing tone mapped images

while $E_{new} < E_{old}$ **do**
  for each pixel p **do**
  $L_n^+ = L_n + \epsilon_1(p)\delta_p$
  $L_n^- = L_n - \epsilon_2(p)\delta_p$, ▷ ϵ_1, ϵ_2 from Eq. 5.28 and Eq. 5.29
  $(LVC^{+/-}, AIC^{+/-}, INV^{+/-}) = DRIM(L_n^{+/-}, H)$,
  $Diff_{type} = (type^+ - type^-)/d(p),\ type \in Type = \{LVC, AIC, INV\}$ ▷ d from Eq. 5.30
  $gradient(p) = \sum_X (\sum_{Type} type^{k-1} \times Diff_{type}) / (\sum_{Type} type^k)^{1-1/k}$
  ▷ X is either 50×50 or full image domain
  end for
$L_{n+1} = L_n - \alpha_n \times gradient$, ▷ α_n determined in Section 5.3.1
$(LVC, AIC, INV) = DRIM(L_{n+1}, H)$
$E_{old} = E_{new}$
$E_{new} = \sum_X (LVC^k + AIC^k + INV^k)^{1/k}$
end while
$L_* = L_n$ ▷ L_* steady-state of the gradient descent
Return $P^D L_*$ ▷ P^D from Eq. 5.45

4 The final output LDR color image is then obtained by combining the X_0 and Z_0 components of the output of the preprocessing with the output $P^D L_*$ of the gradient descent algorithm.

5.3.2 Experiments

We test our algorithm dedicated to optimize tone mapped images on different HDR images taken from the MPI [MPI, 2004] and Fairchild [Fairchild, 2013] databases, and tone mapped images produced by the TMOs [Ferradans et al.,

2011], [Drago et al., 2003b], [Reinhard and Devlin, 2005], and [Mantiuk et al., 2008].

The evaluation of our algorithm is two-fold: global and pixel-wise. As a pixel-wise measure, we consider the distortion maps given by the metric DRIM and compare the initial condition and final output. As a global measure, we compare their averaged perceptual distance with H, making use of the following energy

$$E(L,H) = \frac{1}{|\Omega|} \sum_{(a,b)\in\Omega} \|DRIM(L,H)(a,b)\|_2 \tag{5.48}$$

The first experiment we propose consists in evaluating the preprocessing. In Figure 5.1, we show the input tone mapped image, output color image of the preprocessing, as well as their distortion maps. We have applied the Eq. 5.47 to the luminance channel of tone mapped image obtained with the method of [Ferradans et al., 2011]. The HDR source image is taken from the MPI database [MPI, 2004]. We observe that the LVC distortion has been reduced, whereas the INV distortion has increased a bit. As we can see in Eq. 5.47, the preprocessing is only devoted to reduce the LVC and AIC distortions, and does not take into account the INV distortion. Hence, some choices of the parameters α, β might yield an increase of the INV distortion.

On Table 5.1, we present results of the preprocessing (amongst other results) tested on images of the Fairchild database [Fairchild, 2013] for the TMOs aforementioned. Note that the images have been rescaled to 200×200 pixels in order to speed up the gradient descent algorithm. Average results have been computed over 10 images of the dataset. The results confirm that the preprocessing reduces the perceptual distance with respect to the HDR source image.

In the second experiment, we evaluate the final output of our method described above for different perceptual distances (Eq. 5.32) parametrized by a strictly positive number k. Table 5.1 shows the distance (Eq. 5.48) of the initial tone mapped images and output images with a given HDR image for the following values of the parameter k : $0.8, 1, 1.2, 2, 5$ and 50. The summation for the computation of the gradient operator (Eq. 5.40) has been done on 50×50 neighborhoods. The results show that the minimum distance is always reached for k close to 1 ($k = 0.8$ in most of the cases), and the distance increases when k gets away from

Figure 5.1: Evaluation of preprocessing stage for [Ferradans et al., 2011] TMO. Top row: input tone mapped image, output of preprocessing stage. bottom row: distortion map of the input and preprocessed image.

this optimal k value. Note that the same behavior occurs when k is lower than the optimal value but we do not show the result in this chapter. We would like to point out that these results were not expected.

In Figure 5.2, we show a region of the output color images obtained with $k = 0.8, 1$ and 5 along with their distortion maps. The HDR source is the image

Figure 5.2: Comparison of output with different k values. Top left: final output with $k = 0.8$. Top middle: final output with $k = 1.2$. Top right: final output with $k = 5$. Bottom row: corresponding distortion maps

"snow" from the MPI database [MPI, 2004]. The input tone mapped image is provided by the TMO [Mantiuk et al., 2008]. Taking a close look, we can see in the distortion maps that there is an increase in AIC (blue patches) and INV (red patches) with the increase in k value, which is coherent with the fact that the distance increases with the value of k. However, visual differences are almost impossible to perceive for such close values of k.

In the third experiment, we compare the output of the preprocessing stage with the final output of our method using 50×50 neighborhoods and parameter $k = 0.8$, as well as their distortions maps (see Figure 5.3). The HDR source is the image "PeckLake" from the Fairchild database, and the input tone mapped image is provided by the TMO [Drago et al., 2003b]. Its corresponding distortion map reveals a great loss of contrast (green patches). We observe that the preprocessing stage reduces such a distortion, and the gradient descent algorithm applied to the output of the preprocessing reduces it to even greater extent. These improvements are confirmed when computing the perceptual distances (Eq. 5.48) at each stage with the HDR source image: initial (0.64), preprocessing (0.51), and final (0.39). There is a 23% reduction in distance in the final image compared to the preprocessed image. From this result and the ones shown in Table 5.1 we can claim that applying the gradient descent algorithm to the output of the preprocessing provides much better results than applying only the preprocessing.

Figure 5.4 shows some regions of images on Figure 5.3 and "BarHarborPresunrise" from the Fairchild database. We observe that the preprocessing and the

Table 5.1: Distance with different k values.

Image \ TMO		Drago	Reinhard	Mantiuk	Ferradans
AmikBeav DamPM1	Initial	0.73	0.74	0.76	0.75
	Preprocess	0.57	0.61	0.66	0.71
	Final (k = 0.8)	0.46	0.52	0.59	0.55
	Final (k = 1)	0.47	0.52	0.59	0.55
	Final (k = 1.2)	0.47	0.52	0.59	0.55
	Final (k = 2)	0.49	0.53	0.6	0.56
	Final (k = 5)	0.54	0.57	0.62	0.56
	Final (k = 50)	0.56	0.59	0.64	0.60
Barharbor Presun	Initial	0.66	0.67	0.53	0.56
	Preprocess	0.53	0.55	0.49	0.51
	Final (k = 0.8)	0.42	0.45	0.45	0.46
	Final (k = 1)	0.42	0.45	0.44	0.46
	Final (k = 1.2)	0.42	0.45	0.44	0.45
	Final (k = 2)	0.43	0.47	0.45	0.46
	Final (k = 5)	0.47	0.51	0.47	0.49
	Final (k = 50)	0.50	0.53	0.48	0.50
Average (10 images)	Initial	0.65	0.66	0.57	0.59
	Preprocess	0.53	0.58	0.52	0.54
	Final (k = 0.8)	0.42	0.46	0.46	0.46
	Final (k = 1)	0.42	0.46	0.46	0.46
	Final (k = 1.2)	0.43	0.46	0.46	0.46
	Final (k = 2)	0.44	0.48	0.47	0.47
	Final (k = 5)	0.48	0.51	0.48	0.49
	Final (k = 50)	0.49	0.53	0.50	0.51

final output have higher contrast than the original tone mapped images which means that the LVC distortion has been reduced. Note that the results on top-row are coherent with the distortion maps in Figure 5.3.

At last, we analyze the impact of modifying the neighborhood size used for the summation in the expression of the gradient operator (Eq. 5.40). We test our algorithm for 50×50 as well as 200×200 neighborhoods (the whole image domain). The parameter k has been set to 0.8, which is the value giving the best results in the case of 50×50 neighborhoods according to the second experiment. In Figure 5.5, we show the output color images obtained with the two neighborhood sizes along with their distortion maps. The HDR source images

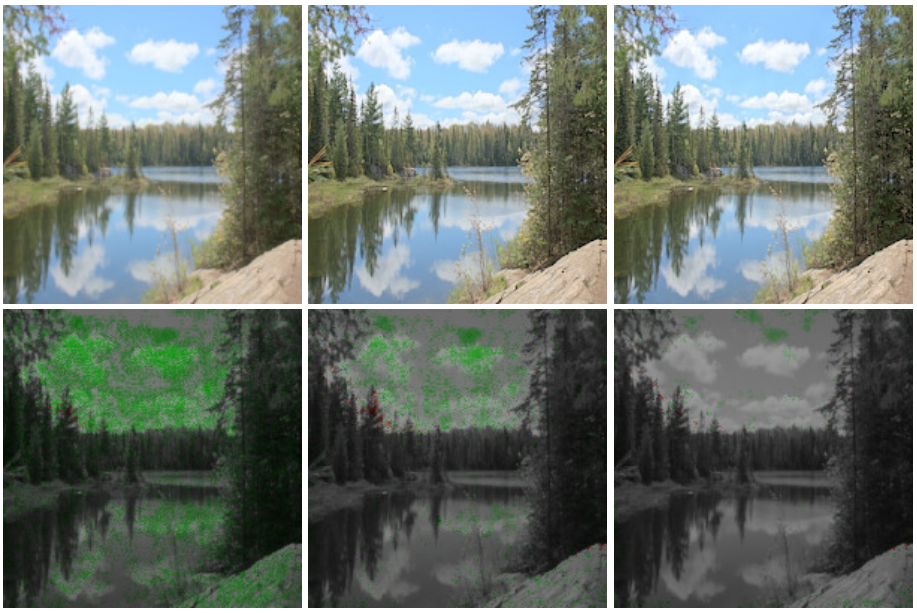


Figure 5.3: Comparison between the output of preprocessing stage and the final stage. Top left: input tone mapped image of [Drago et al., 2003b]. Top middle: output of preprocessing stage. Top right: final output with $k = 0.8$. Bottom row: corresponding distortion maps.

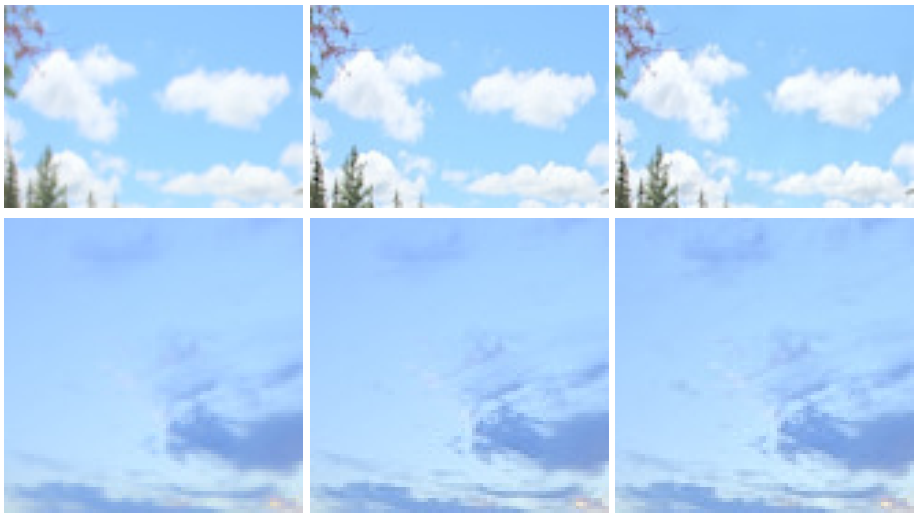


Figure 5.4: Visual comparisons between the output images of the different stages. Top left: "PeckLake" tone mapped by [Drago et al., 2003b]. Bottom left: "BarHarborPresun" tone mapped by [Drago et al., 2003b]. Middle: output of preprocessing stage. Right: final output with $k = 0.8$.

are "BarHarborPresunrise" and "PeckLake" from the Fairchild database. The input tone mapped image is provided by the TMO of [Reinhard and Devlin, 2005] and [Drago et al., 2003b]. The distortion maps show that using the whole image domain substantially reduces the LVC distortion (less green patches appear in the distortion map), and reduces (in less proportion) the INV distortion (less red patches). This improvement is due to the fact that the metric DRIM uses the information of the whole image to estimate the distortions at each pixel. Then, restricting the domain where we compute the gradient operator yields some information loss.

Table 5.2 compares the distance between the final output images (using the aforementioned domains) with the HDR source image. We can see that the results obtained by using a 200×200 neighborhood have smaller error than the results achieved by using a 50×50 neighborhood, and this numerical behavior is consistent with the visual distortion maps presented in Figure 5.5.

In Figure 5.6, we show some results of our algorithm with a 200×200 neighbor-

Table 5.2: Distance at the final stage of our method with 50×50 and 200×200 neighborhoods. Percentile improvement at each stage with respect to the original tone mapping result is given in brackets.

Image \ TMO		Drago	Reinhard	Mantiuk	Ferradans
AmikBeav DamPM1	Initial	0.73	0.74	0.76	0.75
	Preprocess	0.57 (21%)	0.61 (13%)	0.66 (14%)	0.70 (6%)
	Final (50×50)	0.46 (36%)	0.52 (31%)	0.59 (23%)	0.55 (27%)
	Final (200×200)	0.26 (64%)	0.36 (52%)	0.47 (38%)	0.39 (48%)
Barharbor Presun	Initial	0.66	0.67	0.53	0.56
	Preprocess	0.53 (20%)	0.55 (17%)	0.49 (7%)	0.51 (9%)
	Final (50×50)	0.42 (37%)	0.45 (32%)	0.45 (16%)	0.46 (18%)
	Final (200×200)	0.21 (69%)	0.25 (63%)	0.35 (34%)	0.34 (39%)

hood tested on the different TMOs mentioned above applied to the same HDR image. By a close observation of the output images, we can notice an enhancement of details of the initial tone mapped images which is confirmed by comparing the corresponding distortion maps (reduction of LVC distortion).

5.4 Conclusion

Based on perceptual metrics that measure distortions between images of different dynamic range, we propose in this chapter a method to improve tone mapped images based on a non local variational problem. We tested this approach with the metric DRIM [Aydin et al., 2008] for different tone mapped images provided by several TMOs. The experiments show that our approach improves the tone mapped images tested in the sense that it reduces their perceptual distance to their HDR source. Our method provides an average reduction of the distance by more than 25%.

Further work will be devoted to apply the proposed framework to contexts where minimization of a perceptual distance could also be useful. One such application may be to optimize gamut mapping methods by considering a perceptual metric that measures color distortions between images. We are also investigating im-

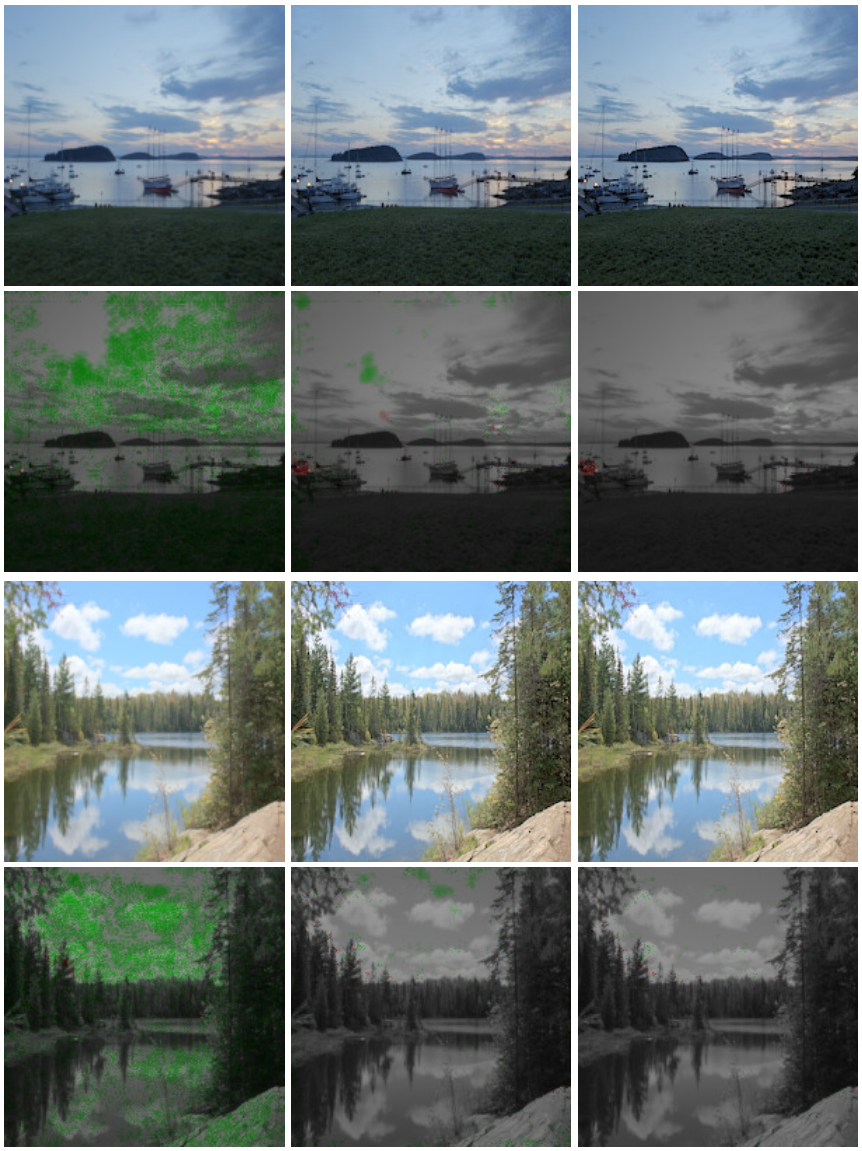


Figure 5.5: Comparison between the final output with 50×50 and 200×200 neighborhoods. First row left: input tone mapped image of [Reinhard and Devlin, 2005], third row left: input tone mapped image of [Drago et al., 2003b]. Middle: final output with 50×50 neighborhood and their distortions maps. Right: final output with 200×200 neighborhood and their distortions maps.

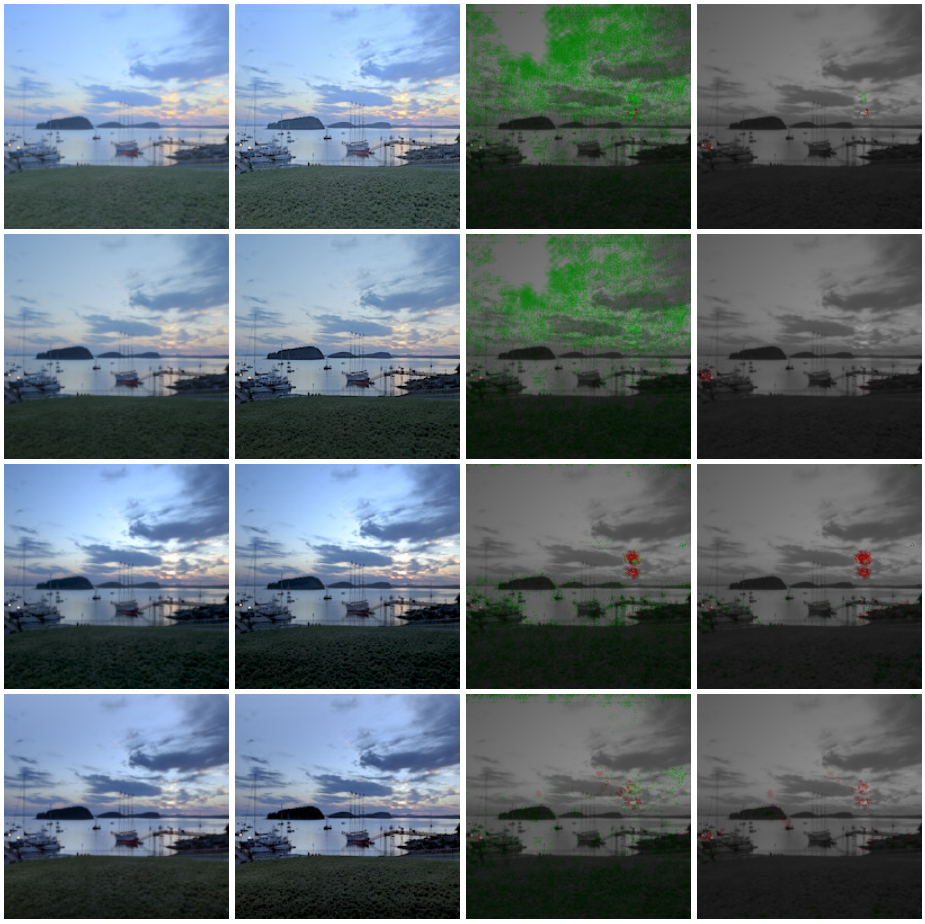


Figure 5.6: The final output. First column: input tone mapped images (TMOs from top to bottom: [Drago et al., 2003b], [Reinhard and Devlin, 2005], [Mantiuk et al., 2008], [Ferradans et al., 2011]). Second column: final output images with a 200×200 neighborhood. Third column: distortion maps of input tone mapped images. Fourth column: distortion maps of final outputs. See Table. 5.2 (image “BarharborPresun”) for the corresponding distances.

provements of our mathematical model. Indeed, we were using in this chapter a gradient descent approach in order to reduce perceptual distances between images. However, this approach suffers from two main issues when dealing with tone mapping: first, the gradient descent algorithm might stop at local minima that are not global and that are close to the initial condition; second, the representation of LDR and HDR images as smooth functions is not very realistic since we are not taking into account the specificity of their dynamic ranges: the use of other functional spaces might be more appropriate and lead to better algorithms.

Chapter 6

TONE MAPPING BASED ON VISUAL PERCEPTION MODELS

In this chapter we propose two TMOs that are based on models of visual perception. Both TMOs have a global first stage followed by a local contrast enhancement stage. Figure 6.1 shows the simplified block diagram of our proposed methods.

The chapter is organized as follows: we introduce our tone mapping Algorithm 1 in Section 6.1. Then in Section 6.2 we present the details of our tone mapping Algorithm 2. Each section includes implementation details, experiments and results. The content of this chapter is derived from the work published in [Cyriac et al., 2015a] [Cyriac et al., 2015b] [Cyriac et al., 2016b].

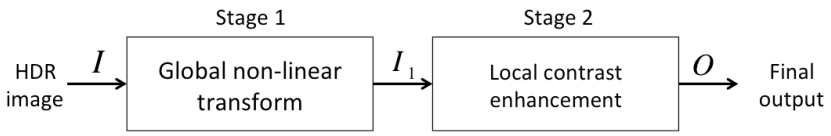


Figure 6.1: Block diagram of the proposed TMOs.

6.1 Tone Mapping: Algorithm 1

6.1.1 Overview

The first stage applies an optimal gamma non-linearity that best equalizes the lightness histogram [Kane and Bertalmío, 2016]. The second stage adopts the neural model [Bertalmío, 2014], which is an extension of the contrast and color enhancement method [Bertalmío et al., 2007] used in [Ferradans et al., 2011], with larger capabilities in terms of redundancy reduction and the ability to reproduce assimilation phenomena. Both [Bertalmío, 2014] and [Bertalmío et al., 2007] are closely related to the Retinex theory [Land and McCann, 1971] of color vision and to the perceptually inspired color correction approach of [Rizzi et al., 2003]. We validate the method with two image quality metrics DRIM [Aydin et al., 2008] and TMQI [Yeganeh and Wang, 2013] that incorporate a model of human vision.

6.1.2 Stage 1: Constrained Histogram Equalization based on Psychophysical Data

Natural scenes tend to have a low-key luminance distribution. This means that low luminance values occur much more frequently than high luminance values. This is especially true for HDR images such as a picture taken directly into sunlight. In Figure 6.2, we show the relationship between the median luminance and the dynamic range of each of the images from the Fairchild HDR dataset [Fairchild, 2007]. The median luminance is calculated after the luminance range has been normalized between 0 and 1. As the dynamic range increases the median luminance decreases. This result indicates that, when HDR images are presented linearly on low dynamic range media, they are dominated by dark, low contrast regions (see Figure 6.3, top). This problem can be mitigated by a process called histogram equalization that flattens the luminance distribution of an image. This technique is well established and is effective at increasing the contrast, and in turn, the detail visible in an image. Complete histogram equalization is achieved by computing the cumulative histogram of an image and applying this as a pointwise non-linearity as follows, where H is the normalized cumulative histogram, I the original normalized image and x a pixel location:

$$I_{eq}(x) = H(I(x)). \tag{6.1}$$

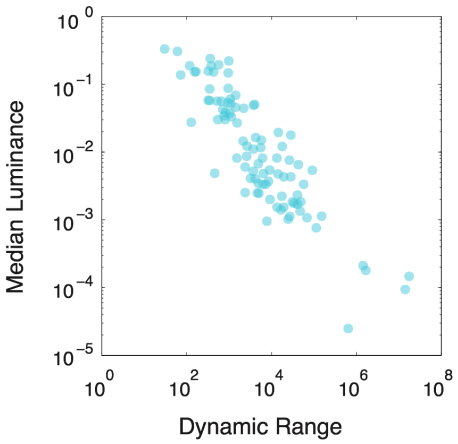


Figure 6.2: Relation ship between the median luminance and the dynamic range of each of the images from the Fairchild HDR dataset. Figure from [Kane and Bertalmío, 2016].

Although complete histogram equalization is highly effective at increasing image contrast, it can lead to very sharp changes in contrast and frequently produces unnatural looking images (see Figure 6.3, bottom left). Thus some form of constrained histogram equalization is necessary (see Figure 6.3, bottom right). One approach is to apply a smooth function that approximates the cumulative histogram.

A recent psychophysical study [Kane and Bertalmío, 2016] investigated image quality scores for images presented with different system gammas, where system gamma is the end-to-end pixel-wise exponent that describes the relationship between the relative luminance values in the original scene and the displayed image (i.e. the product of the decoding gamma of the monitor and a variable encoding gamma set by the experimenter). The stimuli were images from the Fairchild HDR dataset [Fairchild, 2007] and spanned a broad range of dynamic ranges from two to seven orders of magnitude. The images were displayed with a system gamma of between 1/16 to 4 using a logarithmic sampling. Subjects were asked to rate the perceived quality of each image using a sliding scale. Figure 6.4 shows the major finding, that image quality scores could be predicted by the degree of flatness in the perceived lightness distribution, where lightness was

Figure 6.3: Comparison of complete histogram equalization and constrained histogram equalization. Top row: linearly scaled HDR image. Bottom row: result of complete histogram equalization of the HDR image and result of constrained histogram equalization (first stage of Algorithm 1) applied on the HDR image.

modeled as gamma function of on-screen luminance. Accordingly, the optimal system gamma is the one which produces the flattest lightness histogram.

$$\gamma_{opt} = \arg\min_{\gamma} F(\gamma) \tag{6.2}$$

where

$$F(\gamma) = 1 - \frac{1}{N}\sqrt{\sum_{i=1}^{N} \left(H\left((L^{\gamma})^{\gamma_{psy}}\right)_i - i\right)^2} \tag{6.3}$$

F is the flatness of the cumulative histogram H, N is the number of bins for the histogram (we fix that to be 2^{16}) and L is the on-screen luminance. We assume that the perceived lightness of the displayed luminance can be modeled as a gamma function γ_{psy} and we fix it as 0.4 in our method. Figure 6.5 shows the schematic of the model.

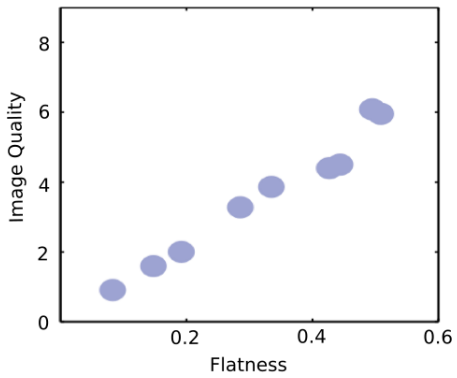


Figure 6.4: Relationship between average perceived image quality vs the degree of flatness of histogram. Figure from [Kane and Bertalmío, 2015].

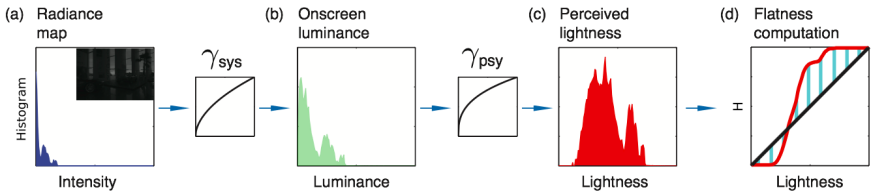


Figure 6.5: (a) The luminance histogram of the original HDR image after normalization between the range 0 and 1. (b) The onscreen luminance histogram after being passed through the system gamma γ_{sys} (γ_{opt} in optimal case) of the imaging pipeline. (c) The estimated perceived lightness distribution after being passed through γ_{psy}. (d) We compute the flatness of a distribution as the root mean square difference between the cumulative lightness distribution and the identity line (Eq. 6.3). Figure from [Kane and Bertalmío, 2015].

We model the first stage of our TMO based on the above psychophysical study as follows:

$$I_1 = (I_{hdr})^{\gamma_{enc}} \tag{6.4}$$

where $\gamma_{enc} = \gamma_{opt}/\gamma_{dec}$, $\gamma_{dec} = 2.2$ and I_{hdr} is the input HDR image normalized to [0,1] range by linear scaling and clipping (implementation details in Section 6.1.4)

6.1.3 Stage 2: Contrast Enhancement based on a Neural Model

The first stage of our TMO applies an 'optimal' gamma non-linearity. However, this approach alone is not always sufficient to produce a high quality image. In particular, those images with a high DR tend to have a flat, low contrast appearance. Also a global approach is not able to model the spatially variant operation of the human visual system.

In [Ferradans et al., 2011], they showed that by using the color enhancement model of [Bertalmío et al., 2007] as a second stage, the local contrast and color constancy properties of the human visual system can be approximated. The following energy functional proposed by [Bertalmío et al., 2007] is an improvement from the energy functional proposed by [Sapiro and Caselles, 1997] (that performs histogram equalization when it is minimized) by incorporating basic visual perception principles, such as locality, color contrast and white patch:

$$\begin{aligned} E(I) = \frac{\alpha}{2}\sum_x (I(x) - \frac{1}{2})^2 dx - \gamma \sum_x \sum_y w(x,y)|I(x) - I(y)|dxdy \\ + \frac{\beta}{2}\sum_x (I(x) - I_0(x))^2 dx \end{aligned} \tag{6.5}$$

where $\alpha, \beta, \gamma > 0$, I is a color channel (R, G, or B) of an image which is in the range [0, 1] and x, y are pixel positions. The first term measures the average difference of the image pixels with the mid-value of $1/2$. The average value of the original image is used in [Ferradans et al., 2011] instead of $1/2$. The second term calculates the local contrast, where w is a Gaussian kernel with standard deviation σ and $I(x)$ and $I(y)$ are intensity values at pixel positions x and y respectively. The last term measures the average departure of the new image from the original image (I_0). Now, by minimizing $E(I)$ one could maximize the

contrast without departing too much from the original image and mid-value, and this can be achieved by a gradient descent approach, with

$$I_t(x) = -\alpha(I(x) - \frac{1}{2}) + \gamma \sum_y w(x,y) sign(I(x) - I(y)) dy - \beta(I(x) - I_0(x)). \tag{6.6}$$

The Wilson-Cowan equations [Wilson and Cowan, 1972] [Wilson and Cowan, 1973] that describe the temporal evolution of the neural activity in the $V1$ region of the visual cortex, could be a gradient descent of certain energy and Eq. 6.6 is closely related to it [Bertalmío et al., 2007] [Bertalmío and Cowan, 2009]. Also [Bertalmío, 2014] showed that [Bertalmío et al., 2007] always produces local contrast enhancement but not assimilation, and proposed a modification that incorporates all the features of [Bertalmío et al., 2007] along with lightness induction. The modified gradient descent function is

$$\begin{aligned} I_t(x) = -\alpha(I(x) - \mu(x)) + \gamma(1 + (\sigma(x))^c) \sum_y w(x,y) sign(I(x) - I(y)) dy \\ -\beta(I(x) - I_0(x)), \end{aligned} \tag{6.7}$$

where the mid-value of the first term is no longer global but the local mean of the original image computed with a Gaussian kernel: $\mu(x) = (G * I_0)(x)$ and a constant weight for the second term is replaced by a spatially and temporally varying one, based on the local standard deviation σ, while c is a constant.

The image illustrated in the top row left of Figure 6.6 has equally spaced gray bars either superimposed on a dark or a light background. The observer will perceive the bars as darker on a dark background and vice versa. However, in the original formulation [Bertalmío et al., 2007], lightness contrast is predicted where the bars are perceived as lighter on a dark background and vice versa (second row). The adaptation of [Bertalmío, 2014] correctly predicts lightness assimilation (third row).

If we use a linear combination of two Gaussian kernels to compute the local mean, we can drop the term $(1 + \sigma^c(x))$ from Eq. 6.7 but still produce assimilation, as Figure 6.6 (bottom row) shows. Accordingly the modified gradient

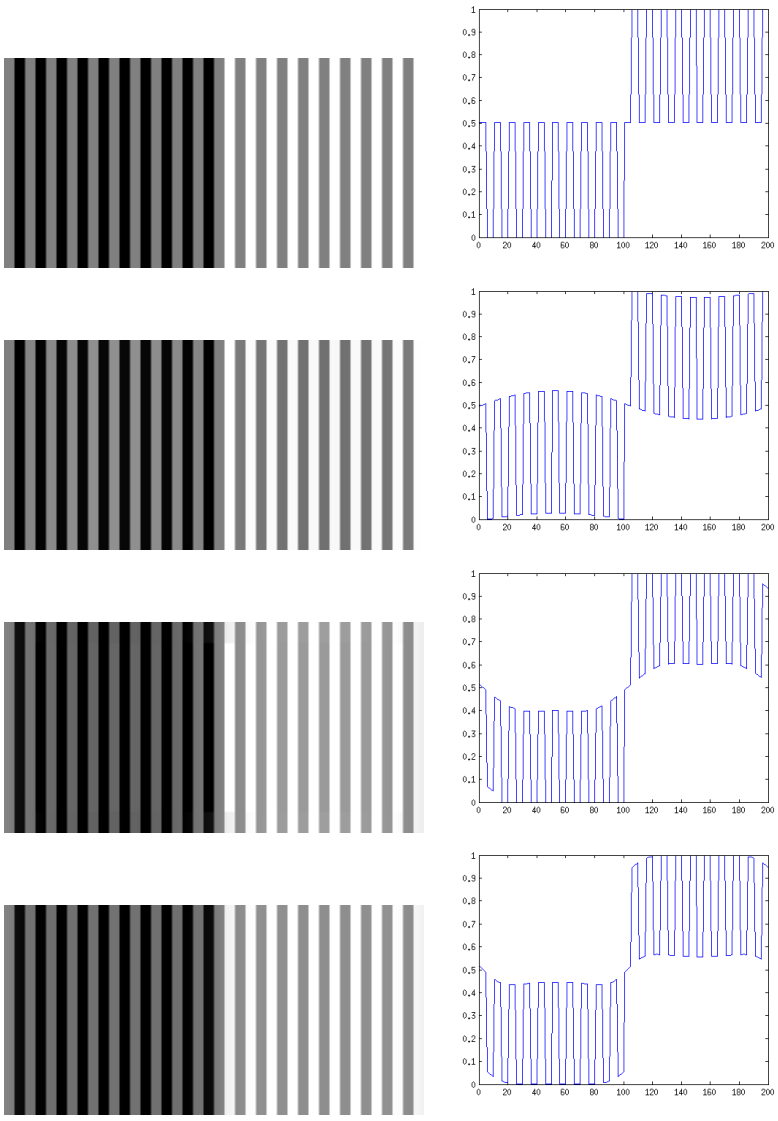


Figure 6.6: Illustration of lightness assimilation and contrast. First column , from top to bottom: input image, result of applying [Bertalmío et al., 2007] to input image, result of applying [Bertalmío, 2014], result of gradient descent by Eq. 6.8. Second column, shows the profile of line from the corresponding images.

descent function is:

$$I_t(x) = -\alpha(I(x)-\mu(x))+\gamma \sum_y w(x,y)sign(I(x)-I(y))dy-\beta(I(x)-I_0(x)). \tag{6.8}$$

where I_0 is one of the color channels of the output image from first stage, $\alpha, \beta, \gamma = 1$, w is a normalized 2D Gaussian kernel of standard deviation σ_w. We fix σ_w as 200. We compute the local mean $\mu(x)$ by convolving the image with a kernel K: $K = n1 \times G1 + n2 \times G2$, where $n1 = 1, n2 = 0.5$, the standard deviation of Gaussian kernel $G1$ is 10 and that of $G2$ is 250.

We develop the second stage based on the above neural model as a gradient descent method by iterating:

$$I^{n+1}(x) = I^n(x) + \Delta t(I_t^n(x)) \tag{6.9}$$

where at the start (when $n = 0$) I^n is the output of the first stage, $\Delta t = 0.15$, and the iteration stops when the difference between the current and the previous result is less that 0.005.

6.1.4 Implementation

In this section we present the implementation details. Our method consists of two stages, one global operation (Eq. 6.4) and a local operation (iterating Eq. 6.9), and are applied separately to each of the red (R), green (G) and blue (B) color channels of the image. Initially, the input image is linearly scaled to [0,1] range and the luminance component L is computed using the formula $L = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B$.

Then, we clip L, R, G and B values that are above 99 percentile of the L channel. In Figure 6.7, we show the importance of clipping to get an image with more contrast. The left image is obtained by applying our algorithm without clipping. This image looks darker and is of low contrast. On the other hand the right image obtained by first applying clipping has more contrast and more details of the scene are visible.

Now, we estimate the optimal system gamma (γ_{opt}) from the L channel and compute the encoding gamma γ_{enc}. Then the R, G and B values are non-linearly

Figure 6.7: Illustration of the importance of clipping. Left image is the final output of our TMO without clipping, and right image is the final output of our TMO with clipping.

transformed by Eq. 6.4. The pseudocode of the complete first stage is given below:

Pseudocode 3 First stage

Input: R, G, B channels of an HDR image
Result: R^1, G^1, B^1 LDR image
$L = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B$
$\gamma_{sys} = 0.4;\ \gamma_{psy} = 0.4\ ; F_{old} = 0;\ \gamma_{dec} = 2.2;$

While $diff > 0$ and $\gamma_{sys} < 2$
$\gamma_{sys} = \gamma_{sys} + 0.1$
$L^* = (L)^{\gamma_{psy} \times \gamma_{sys}}$
$F = 1 - \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(H(L^*)_i - i\right)^2}$ (Eq. 6.3)
$diff = F - F_{old}$
$F_{old} = F$

$\gamma_{enc} = \frac{\gamma_{sys}}{\gamma_{dec}}$
$R^1 = R^{\gamma_{enc}};\ \ G^1 = G^{\gamma_{enc}};\ \ B^1 = B^{\gamma_{enc}}$

The second stage takes as input the result of the first stage and apply the gradient descent by iterating Eq. 6.9.

6.1.5 Experiments and Results

In order to evaluate the performance of our TMO, we use two quantitative metrics: DRIM [Aydin et al., 2008] and TMQI [Yeganeh and Wang, 2013]. Both metrics are explicitly designed to compare the original HDR image to the tone-mapped LDR counterpart.

The metric DRIM estimates three types of distortions for each pixel: loss of visible contrast (LVC), amplification of invisible contrast (AIC), and reversal of visible contrast (INV). The output of the metric is visualized using a color coded distortion map. Green represents LVC, blue indicates AIC, red indicates INV and the saturation of each color indicates the magnitude of distortion. Detailed explanation of DRIM is given in Section 4.3. The metric DRIM does not provide a single global quality score. To generate a single output we average across space the estimated loss, amplification and reversal of contrast:

$$GS = \frac{1}{N} \sum \sqrt{LVC^2 + AIC^2 + INV^2}. \tag{6.10}$$

A low GS indicates less contrast distortions. We note that this formulation of GS does not differentiate between the distortion types and their spatial distributions. These factors can affect the perceived quality of the image. Hence GS has to be considered along with the distortion map to get a clear indication of the image quality.

The TMQI metric has three terms: structural fidelity (S), Naturalness (N) and overall quality (Q) obtained by a weighted average of S and N (details in Section 4.3).

We estimate the performance of the first and second stages of the model using the metric DRIM and illustrate this in Figure 6.8. The top left image is the output of the first stage and the top right image is the output of the second stage of our algorithm. We can see considerable improvement of contrast in the output of the second stage. The bottom left image shows the distortion map of the first stage and the bottom right image shows the distortion map of the second stage. The LVC (green color) is considerably reduced in the distortion map of the second stage. For a numerical comparison we use GS. The reading for the first stage is GS = 0.589 and for the second stage is GS = 0.554. The results confirm that the

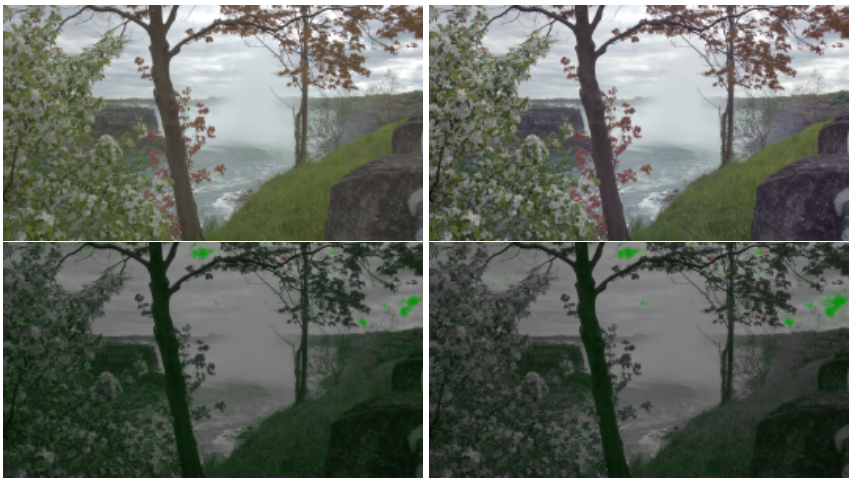


Figure 6.8: Comparison between the output of first and second stage. Top row, output of the first stage and output of the second stage. Bottom row, corresponding distortion map.

second stage produces an output with less error than the first stage.

To estimate how our model compares to contemporary approaches we apply the DRIM and TMQI metrics to the following TMOs: [Mantiuk et al., 2008], [Drago et al., 2003b], [Reinhard and Devlin, 2005] and [Ferradans et al., 2011]. We use the tone mapping functions provided by $pfstools$ [Mantiuk and Heidrich, 2009] to generate tone mapped LDR images from HDR images, except for [Ferradans et al., 2011] for which the author has supplied the code. We also use the default parameters for all TMOs. In Figure 6.9 and 6.10, the HDR images are 'BloomingGorse2' and 'CemeteryTree1' from the Fairchild dataset [Fairchild, 2007]. According to DRIM, for both examples our algorithm produces less contrast distortions than the other methods, except for [Mantiuk et al., 2008] to which we are equal in terms of the GS (see Table 6.1). In the first example, our method has less LVC (green color) in the bright regions but the result of [Mantiuk et al., 2008] has less LVC in the dark regions. And in the second example, our algorithm results in low INV (red color) but has higher LVC (green color) in the dark regions. We refer to Table 6.1 for the numerical global error for these two examples. The TMQI gives different result with that of DRIM. For instance in the

Table 6.1: Quantitative evaluation

Images	TMO	DRIM (GS)	TMQI		
			Q	S	N
BloomingGorse2	Algorithm 1	0.55	0.92	0.94	0.60
	[Mantiuk et al., 2008]	0.55	0.90	0.93	0.48
	[Ferradans et al., 2011]	0.58	0.91	0.91	0.56
	[Drago et al., 2003b]	0.61	0.92	0.94	0.60
	[Reinhard and Devlin, 2005]	0.58	0.95	0.91	0.80
CemeteryTree1	Algorithm 1	0.64	0.95	0.95	0.73
	[Mantiuk et al., 2008]	0.64	0.92	0.95	0.59
	[Ferradans et al., 2011]	0.70	0.93	0.93	0.64
	[Drago et al., 2003b]	0.72	0.83	0.93	0.13
	[Reinhard and Devlin, 2005]	0.76	0.92	0.93	0.61

Table 6.2: Quantitative evaluation of images in Fairchild dataset [Fairchild, 2007]

Images	TMO	DRIM (GS)	TMQI		
			Q	S	N
Average of 41 images for DRIM 105 images for TMQI	Algorithm 1	0.48	0.89	0.92	0.50
	[Mantiuk et al., 2008]	0.46	0.9	0.92	0.51
	[Ferradans et al., 2011]	0.5	0.89	0.89	0.50
	[Drago et al., 2003b]	0.54	0.87	0.87	0.42
	[Reinhard and Devlin, 2005]	0.51	0.87	0.88	0.44

first example, TMQI rates our method equal to [Drago et al., 2003b] in terms of S but ranks bests the method of [Reinhard and Devlin, 2005] in terms of Q due to its high score of N. In Figure 6.11 we show more results of our tone mapping approach.

In Table 6.2, we show the global error with DRIM and TMQI for the above considered TMOs over 41 images (in the case of DRIM) and 105 images (in the case of TMQI) from the Fairchild dataset [Fairchild, 2007]. According to the GS of DRIM our algorithm outperforms all the other compared approaches except [Mantiuk et al., 2008]. The TMQI rates our algorithm lower than [Mantiuk et al., 2008] and equal to [Ferradans et al., 2011] in terms of overall quality (Q). Whereas rates our method equal to [Mantiuk et al., 2008] in terms of S.

Figure 6.9: Comparison of TMOs. TM results (first column) and distortion maps (second column) for these TMOs from top to bottom: our method (Algorithm 1), [Mantiuk et al., 2008], [Drago et al., 2003b], [Reinhard and Devlin, 2005] and [Ferradans et al., 2011].

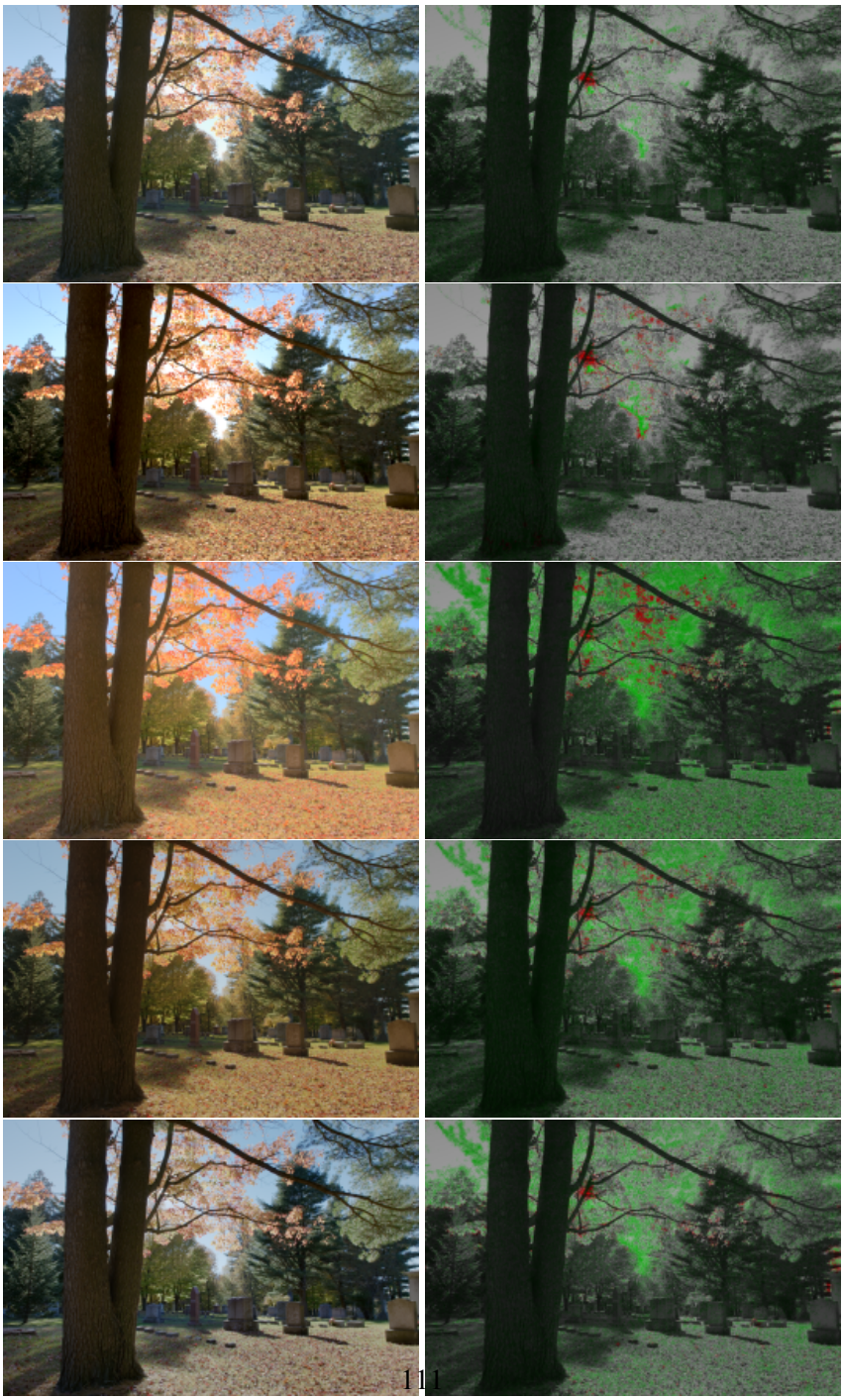


Figure 6.10: Another comparison of TMOs. TM results (first column) and distortion maps (second column) for these TMOs from top to bottom: our method (Algorithm 1), [Mantiuk et al., 2008], [Drago et al., 2003b], [Reinhard and Devlin, 2005] and [Ferradans et al., 2011].

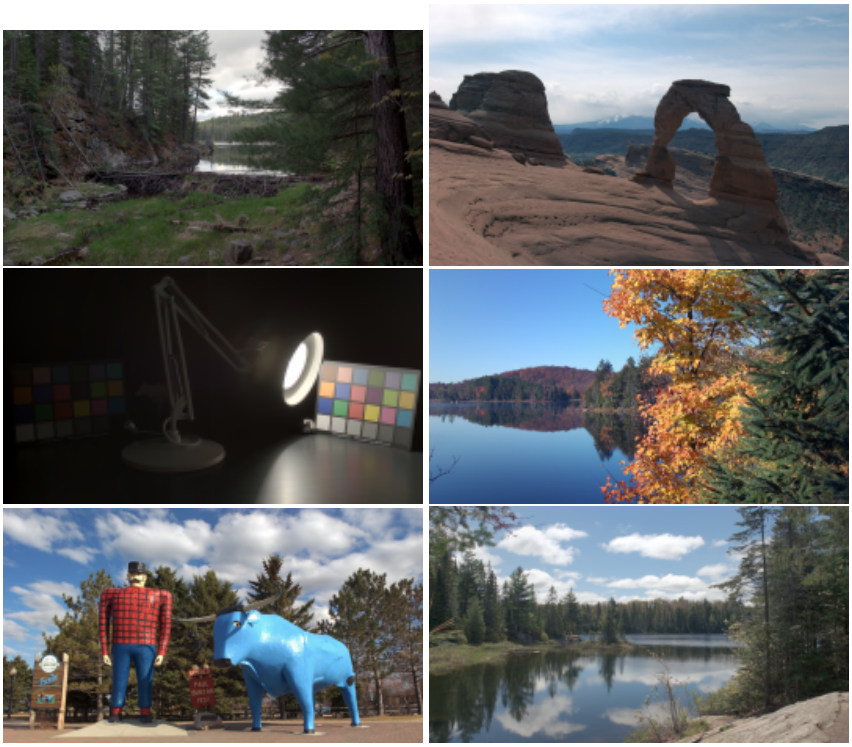


Figure 6.11: Results of our method for several HDR images obtained from the Fairchild database [Fairchild, 2007].

6.2 Tone Mapping: Algorithm 2

The above approach (Algorithm 1) is constrained to use a simple gamma non-linearity in the first stage. This non-linearity is not enough in many cases, especially with images of higher dynamic range (see Figure 6.8 and 6.13, left). The result looks less contrasty with washed out blacks. Also the gradient descent operation in the second stage is computationally expensive and the execution time is around 18 seconds for a full HD image. Hence the Algorithm 1 is not suitable for real time operation.

To overcome the limitation caused by using a simple gamma non-linearity, we modeled the non-linearity as a function γ estimated using some key features of the cumulative histogram of the input image based on natural image statistics. The second stage performs a contrast normalization operation that replicates efficient coding behavior that occurs both in the retina and cortical areas of the human visual system. The method is automatic (no need for user-selected parameters) and has low computational complexity. We validate the method through psychophysical tests that confirm that it outperforms other state of the art algorithms in terms of users' preference. Algorithm 2 can be used in the in-camera image processing pipeline for the non-linear transformations, and also can be used as an off-line tone mapping method for converting HDR images into LDR ones, with applications to cinema shoots (on-set use of LDR monitors with an HDR camera), cinema post-production (color grading), television broadcast (making HDR signals compatible with LDR equipment), and rendering in computer graphics (for video games, 3D animation, the integration of CGI onto real footage, etc.)

6.2.1 Stage 1: Non-linear Adaptation based on Natural Image Statistics

In the vision science community the prevailing view is that the visual system transforms the input image to ensure an efficient representation (see [Olshausen and Field, 2000] and references therein). The human visual system has evolved so as to adapt best to the statistics of natural images. Several works on natural image statistics (e.g. [Ruderman, 1994, Huang and Mumford, 1999]) report that the average shape of the luminance histogram for a natural image is triangular in log-log coordinates: it increases linearly up to a peak, obtained for an image

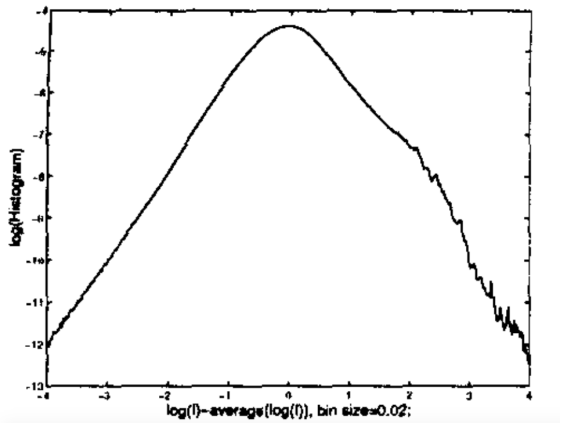


Figure 6.12: Average histogram of natural scenes, in log-log coordinates. Figure from [Huang and Mumford, 1999].

intensity value of M (related to the average of the intensity), and then decreases linearly with a slightly different slope, see Figure 6.12. This implies that the cumulative histogram, being simply the integral of the histogram, will also be a piece-wise linear function in log-log coordinates, increasing linearly with some slope γ_L until the intensity value M, then increasing linearly with a different slope γ_H. In our method we use this insight from the above-mentioned results on natural image statistics to estimate, for the input image, the particular values of M, γ_L and γ_H that best fit the specific histogram of the image. That is, instead of using fixed values of M, γ_L and γ_H that may adequately represent the average statistics of natural images, we tailor these values to the particular image at hand, obtaining an image-dependent and smooth sigmoid curve $\gamma(I)$ such that $\gamma(I) \simeq \gamma_L$ for small intensities, $\gamma(I) \simeq \gamma_H$ for large intensities, and M_{lin} is the exponential of M:

$$\gamma(I) = \gamma_H - (\gamma_L - \gamma_H)(1 - \frac{I^n}{I^n + M_{lin}^n}). \tag{6.11}$$

Then, the cumulative histogram of the image can be approximated in log-log coordinates by

$$H(I) = \gamma(I)(I), \tag{6.12}$$

Figure 6.13: Comparison of the result of the first stage of the Algorithm 1 (left) with the first stage of Algorithm 2 (right).

and in linear-linear coordinates the cumulative histogram has the form

$$H(I) = (I)^{\gamma(I)}. \tag{6.13}$$

From this we introduce our first stage that performs constrained histogram equalization:

$$I_1 = (I)^{\gamma(I)} \tag{6.14}$$

Figure 6.13, compares the result of the first stage of the Algorithm 1 with that of the current method. We can see that the non-linear function based on natural image statistics provides more reliable result in terms of contrast and detail visibility.

6.2.2 Stage 2: Contrast Normalization

In the neuroscience literature there is abundant neurophysiological evidence (see [Brenner et al., 2000, Carandini and Heeger, 2012] and references therein) that the visual system is performing an operation called contrast normalization, in which the contrast (the difference between light intensity and its mean value) is divided by a factor depending on the standard deviation of the light intensity (see Figure 6.14). This re-scaling already occurs at the retina and optimizes information transmission and coding efficiency [Brenner et al., 2000, Kastner and Baccus, 2014]. Given that contrast normalization is a key element of the human visual system we have incorporated it to our method with the following second and final stage:

$$O(x) = \mu(x) + (I_1(x) - \mu(x)) \times k/\sigma, \tag{6.15}$$

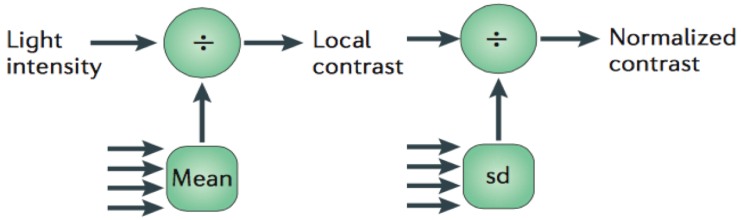


Figure 6.14: A schematic of the contrast normalization. Figure from [Carandini and Heeger, 2012].

where x is a pixel position, I_1 is the output of the first stage of our Algorithm 2, $\mu(x)$ is the local mean of I_1, k is a constant, σ is the standard deviation of I_1, and O is the final output. Figure 6.15 illustrates the various components and the operation of Eq. 6.15. A simple stimulus I_1 (top left), is smoothed by applying a Gaussian kernal to get μ (top right). The difference of the above two signals provides the contrast (bottom left) which is further added to μ to yield the contrast enhanced signal (bottom right).

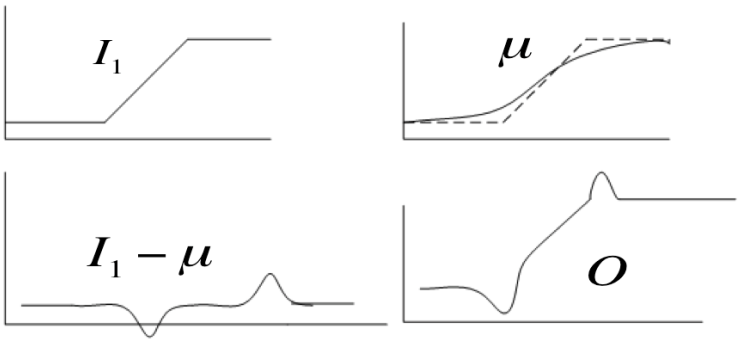


Figure 6.15: Illustration of various component and the operation of the proposed contrast normalization step. A simple stimulus I_1 (top left), smoothed version μ (top right), contrast (bottom left) and final contrast enhanced signal (bottom right).

6.2.3 Implementation

Our TMO (Algorithm 2) consists of two stages, one global operation (Eq. 6.14) followed by a local operation (Eq. 6.15), and are applied separately to each of the red (R), green (G) and blue (B) color channels of the image. Initially, the luminance component L is computed using the formula $L = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B$.

In the first stage, we initially compute the non-linear function $\gamma(I)$ in Eq. 6.11 using the parameter values of γ_H, γ_L and M_{lin}. The parameter values are estimated from the cumulative histogram (H) in log-log coordinates (see Figure 6.16) to correlate well with the subject-chosen parameters in the psychophysical experiment in Section 6.2.5. We compute M_{lin} as the exponential of M, where M is the average of the intermediate values L_m and L_M on the horizontal axis (log luminance), which respectively correspond to the values of 1 and 90% in the vertical axis (log cumulative histogram).

To compute γ_H and γ_L we distinguish 3 cases depending on the shape of the histogram.

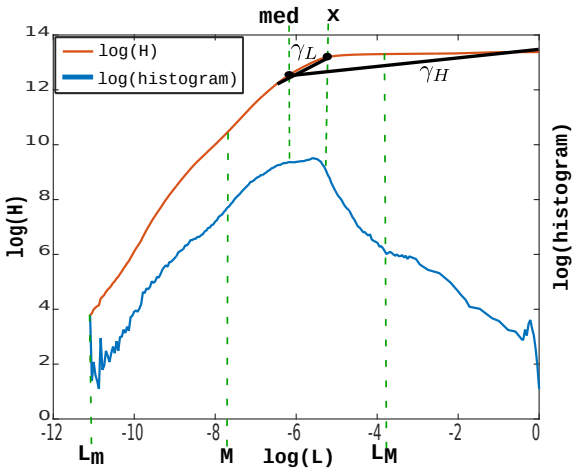


Figure 6.16: Example of a cumulative histogram and histogram for a single natural image (in log-log axes) and our estimated parameters γ_L, γ_H and M.

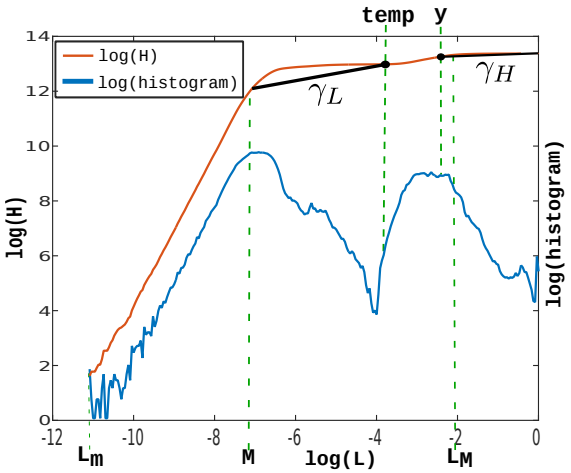


Figure 6.17: Example of a cumulative histogram and histogram for a single natural image (in log-log axes) with a bi-modal distribution and our estimated parameters γ_L, γ_H and M.

Case 1: The histogram of the input image has roughly a triangular shape in log-log coordinates(see Figure 6.16). This is the general case and the majority of images fall in this category. The values of γ_H and γ_L are estimated with respect to $med = log(median(L))$ and $x = log(\sqrt{median(L) \times trm(L)})$ respectively, where $trm(L)$ is the $mean(L)$ after removing 1% of the extreme values.

Case 2: The histogram of the input image has a bi-modal distribution in log-log coordinates (see Figure 6.17). Our estimate of γ_H and γ_L are with respect to the log of $median$ of the upper half of L (denoted by y) and $temp = log(mean(L))$ respectively.

Case 3: The bin of the histogram near to the $median$ is over-populated, resulting in a spike in the histogram (see Figure 6.18). We estimate γ_H and γ_L with respect to v, the intensity value just lower than those high frequent intensities.
More detailed explanations and an algorithm for the parameter estimation are given below. In all the cases, we set the slope value n equal to γ_L.

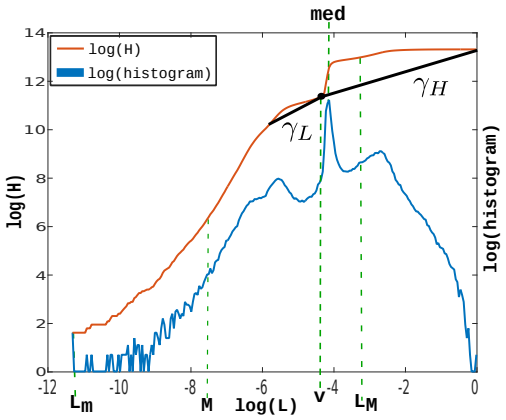


Figure 6.18: Example of a cumulative histogram and histogram with a spike for a single natural image (in log-log axes) and our estimated parameters γ_L, γ_H and M.

Pseudocode 4 Parameter estimation

Input: Image I

Result: γ_L and γ_H

L: Luminance channel of I normalized to [0,1].

$H(L)$: Cumulative histogram of L in log-log domain.

$\mathbf{S}(a, b)$: Slope of line joining point on H corresponding to a and b; where a, b $\in log(L)$

$trm(L)$: Mean of L after removing 1% of extreme values.

1. $\gamma_H = \mathbf{S}\Big(med, 0\Big)$; $\quad med = log(median(L))$

$\gamma_L = \mathbf{S}\Big(x, H^{-1}\big(H(x) - 1\big)\Big)$; $\quad x = log(\sqrt{median(L) \times trm(L)})$

2. If $\gamma_H < \mathbf{S}\Big(temp, 0\Big)$ **then** $\quad temp = log(mean(L))$

2.1. $\gamma_H = \mathbf{S}\Big(y, 0\Big)$; $y = log\Big(median\big(L > median(L)\big)\Big)$

2.2. $\gamma_L = \mathbf{S}\Big(temp, H^{-1}\big(H(temp) - 1\big)\Big)$

3. If $\frac{dH}{d(log(L))} > Te$, for some $v \in log(L)$ and within unit distance from med, **then**

3.1. $\gamma_H = \mathbf{S}\Big(v - \delta, 0\Big)$

3.2. $\gamma_L = \mathbf{S}\Big(v - \delta, H^{-1}\big(H(v - \delta) - 1\big)\Big)$

Initially, γ_H and γ_L are computed by step 1 of the Pseudocode 4, as illustrated in Figure 6.16. γ_H is the slope of the line joining the point in H that corresponds to $med = log(median(L))$ and the top end of H and γ_L is the slope of the line joining the point in H corresponding to $x = log((sqrt(median(L) \times trm(L)))$ and the point in H that is 1 unit lower than the above point in the vertical axis.

If the condition in step 2 of the Pseudocode 4 is satisfied, i.e., γ_H is less than the slope of the line joining the point in H corresponding to $temp = log(mean(L))$ and the top end of H, then γ_H and γ_L are computed as in step 2.1 and 2.2 respectively. This condition is an indication of the bi-modality of the distribution of the luminance histogram. Then a new $median$ is computed by excluding all the luminance value below the original $median$, and γ_H is recomputed. γ_L is recomputed as the slope of the line joining the point in H corresponding to $temp$ and the point in H that is 1 unit lower than the above point in the vertical axis.

But if the condition in step 3 of the Pseudocode 4 is satisfied, i.e., the slope of H is greater than a threshold Te in some region around the $median(L)$ due to a high concentration of some luminance values, then γ_H and γ_L are computed as in step 3.1 and 3.2 respectively and explained in what follows. Select a value $v \in log(L)$ that satisfies the condition in step 3 and is within a unit distance from med, then γ_H is the slope of the line joining the point in H corresponding to $v - \delta$ and the top end of H and γ_L is the slope of the line joining the same point and the point in H that is 1 unit lower than the above point on the vertical axis. We set threshold $Te = 4$ and $\delta = 0.1$.

Adaptive clipping We added an adaptive clipping term C to the first stage to preserve the global contrast of the image. The new first stage is given by:

$$I_1 = (I(x))^{\gamma(I(x))} C(I(x)) \tag{6.16}$$

where C is also a smooth curve with $C \simeq C_L$ for small intensities, $C \simeq C_H$ for large intensities, and with a smooth transition at M_{lin} with a slope m as follows:

$$C(I) = C_L + (C_H - C_L)\Big(\frac{I^m}{I^m + M_{lin}^m}\Big), \tag{6.17}$$

where $C_L$and C_H are also computed from the cumulative histogram of the input

Figure 6.19: Comparison of the result of our TMO without clipping (left) and with clipping (right).

intensity image as follows:

$$\begin{aligned} C_L &= e^{a-\gamma_L Low_x}; \qquad a = log\Big(\frac{1}{255}\Big) \\ C_H &= e^{b-\gamma_H High_x}; \qquad b = log\Big(1-\frac{1}{255}\Big) \end{aligned} \tag{6.18}$$

where Low_x and $High_x$ correspond to $(1/255) \times 100$ and $(1 - 1/255) \times 100$ percentile of L in the log domain.

Figure 6.19 compares the result of stage one with and without adaptive clipping. It is clear that the clipping improves the global contrast of the image.
The output of the first stage is passed to the second stage. For the second stage (Eq. 6.15), the local mean $\mu(x)$ is computed by the convolution of a kernel W with image I_1, where W is generated by the linear combination of two Gaussian kernels with standard deviations $\sigma = 5$ and $\sigma = 25$ and the kernels are weighted by 0.9 and 0.1 respectively. The constant k determines the contrast level of the image, larger k values give images with higher contrast. We set the value of $k = 0.33$ that produces final images that have a natural appearance and good contrast.

6.2.4 Extension to Videos

The proposed method if applied separately to each frame of a video then may produce flickering artifacts due to sudden changes in the parameter values. Figure 6.20 shows the fluctuation of the parameter values of the first stage and also the average luminance of the resulting frames. We overcome this problem by in-

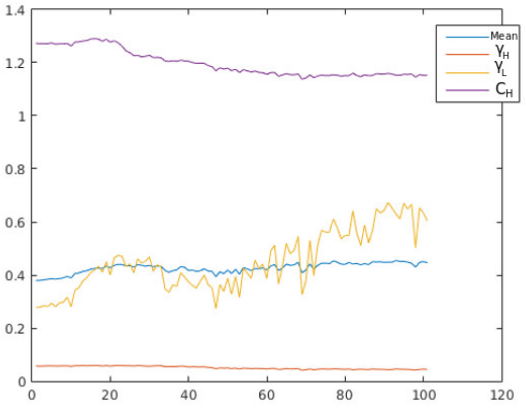


Figure 6.20: Parameter values and average luminance of the result, when the first stage of our TMO is applied to each frame of a video sequence independently without any temporal filtering.

corporating temporal coherence between the parameter values of the consecutive frames as in [Kiser et al., 2012]:

$$P_i = \left(\hat{P}_i + 15 \times P_{i-1}\right)/16 \tag{6.19}$$

where $\hat{P}_i$ is the initial parameter value of the i^{th} frame and P_i is the temporal coherent parameter value. Figure 6.21 shows the smooth change of parameter values and the average frame luminance across the frames of the video sequence after incorporating temporal coherence. The resultant video sequence is with no flickering artifacts. Video results can be downloaded from the following link: `http://ip4ec.upf.edu/tonemapping`

6.2.5 Experiments and Results

In this section we first show the potential of the Algorithm 2 to be used as a method for in-camera non-linear processing and also to be used as a TMO. Then we validate our approach through psychophysical tests in 2 experiments.

In Figure 6.22 we illustrate the advantage of our method over the conventional

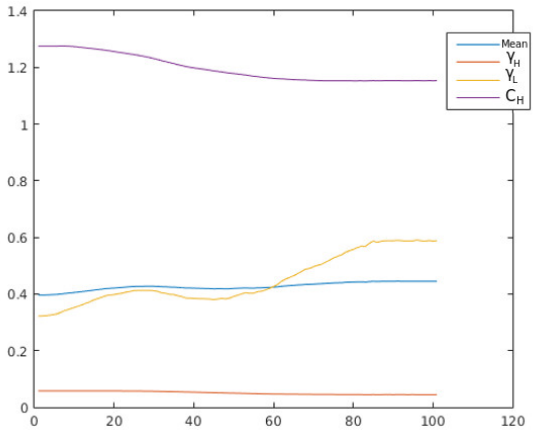


Figure 6.21: Parameter values and average luminance of the result, when the first stage of our TMO is applied to each frame of a video sequence with temporal filtering.

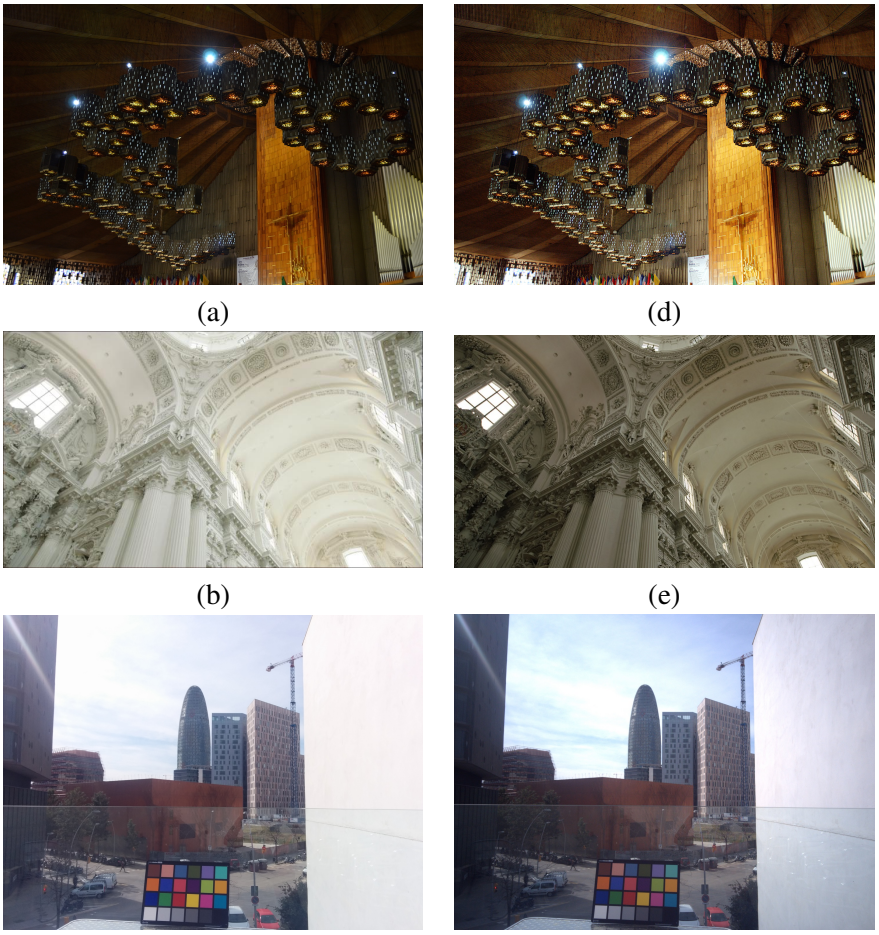
non-linearity applied in a camera imaging pipeline. Three sample images are shown each from a consumer camera, smart phone and cinema camera, along with the results of applying the Algorithm 2 on the corresponding RAW sensor values. Our results look natural in appearance, with enhanced overall contrast and without any visual artifacts.

In Figure 6.23 we show some results of applying the Algorithm 2 to HDR images. Our method when applied to HDR video sequences from the ARRI dataset [Froehlich et al., 2014] produces results that are natural looking with no visible flicker and without any sort of spatio-temporal artifacts. Video results: `http://ip4ec.upf.edu/tonemapping`

Psychophysical experiment

Now we discuss the setup of two psychophysical experiments that we performed to validate our method.

Subjects: Seven subjects completed both experiments. All had corrected to normal vision. Ethics was approved by the Comité Etico de Investigacion Clin-



(a) (d)

(b) (e)

(c) (f)

Figure 6.22: First column: original JPEG images as recorded by the camera, with the exception of image (b) which is generated by applying a S-shaped curve to a LogC image. Second column: results of applying our method to the corresponding RAW images. Camera models: top, Nikon D3100 consumer photography camera; middle, ARRI Alexa digital cinema camera [Andriani et al., 2013]; bottom, Nexus 5 smartphone camera.

Figure 6.23: Results of our method applied to HDR images from the Fairchild dataset (top row) [Fairchild, 2007] and to video frames from the ARRI dataset (bottom row) [Froehlich et al., 2014].

ica, Parc de Salut MAR, Barcelona, Spain, and all procedures complied with the declaration of Helsinki.

Apparatus: Both experiments were conducted on an ASUS VS197D LCD monitor set to 'sRGB' mode with a luminance range from $0.1cdm^{-2}$ to $106cdm^{-2}$, with spatial and temporal resolutions of 1366 by 768 pixels and 50 $\sim$ 75 Hz. The display was viewed at a distance of approximately 70 cm so that 40 pixels subtended 1 degree of visual angle. The full display subtended 33 by 18 degrees. The decoding nonlinearity of the monitor was recorded using a Konica Minolta LS 100 photometer and was found to be closely approximated by a gamma function with an exponent of 2.2. Stimuli were generated under Ubuntu 12.04 LTS running MATLAB (MathWorks) with functions from the Psychtoolbox[Brainard, 1997, Pelli, 1997].

The experiment was conducted in an office environment and the ambient luminance levels were recorded with a Sinometer LX1010B which could record the incoming light from 180 degree angle. The results indicated the average ambient illumination was 147 lux. The surround luminance of the display as measured by the photometer was 65 cd/m^2.

Stimuli:
Experiment one: 20 base images were taken from the Fairchild HDR dataset [Fairchild, 2007] including indoor and outdoor scenes, night time images and landscapes. Images were resized to a quarter of the original areas using Matlabs `imresize` and the setting `nearest` which performs a simple subsampling of pixel value. Each image then covered approximately 80% of the viewing area. The remaining area was presented with a surround luminance of 65 cd/m^2 corresponding to the average luminance of the surrounding area.

Experiment two: 30 base images were taken from the Fairchild HDR dataset [Fairchild, 2007], excluding the 20 used in experiment one. Tone mapped versions of the original HDR images were produced by the Algorithm 1 (Section 6.1), Algorithm 2 (Section 6.2) and the TMO of [Mantiuk et al., 2008] created using `pfstools` [Mantiuk and Heidrich, 2009] using the parameters of 'lcd_office' display type except for ambient illumination which is set to 147 lux. The images were viewed side by side and each image was presented over 14 by 10

Figure 6.24: Psychopysical experiment one setup.

degrees.

In both the experiments the images were presented without correcting for the decoding non-linearity of the monitor.

Procedure:
Experiment one: A screenshot of the experiment procedure is shown in Figure 6.24. Subjects manipulated the parameter γ_H and γ_L of Eq. 6.11 via two scrollbars. The subject interacted with the scrollbar via a mouse and a press of the space key initiated the next image. The range of values corresponding to the scrollbar location was fixed at 0 to 1 for γ_H and 0 to 3 for γ_L. The initial position of the scrollbar was the midway point and the scaling was linear. The parameter values, dynamically and in real time, updated the displayed image by applying Eq. 6.16 to the original HDR image. The subject was asked to manipulate the scrollbars until the most pleasing image was achieved. No further instruction was provided. Subjects had unlimited time to choose the appropriate values.

Experiment two: Subjects were asked to select which of the two simultaneously presented images they preferred. Given the 30 base images and 3 tone-mapped versions for each image, the total number of comparisons was $30 \times 3 = 90$. Sub-

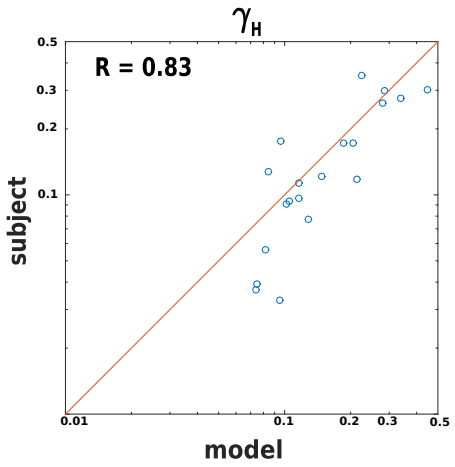


Figure 6.25: Average subject selected values from 7 observers plotted against model estimated parameter values for γ_H.

jects had unlimited time to make the comparison.

Figure 6.25 and 6.26 shows that for both parameters γ_H and γ_L there is a strong correlation, correlation coefficient $R = 0.83$ and $R = 0.88$, between the automatically estimated and user chosen parameters. This result indicates that the parameters estimated by the proposed Algorithm 2 match well with the users' choice.

In Figure 6.27 we show the result of experiment two, in which the subjects evaluated three TMOs: [Mantiuk et al., 2008], Algorithm 1 and Algorithm 2. We choose these TMOs as they were the ones performing best according to DRIM. We compute the accuracy score from the experimental data using the method in [Morovic, 1998] which is based on Thurstone's law of comparative judgments [Thurstone, 1927]. The experiment shows that subjects have a preference for the results of the Algorithm 2 over the other operators. This order is different according to the prediction of the tone mapping metrics DRIM and TMQI (see Table 6.3). According to metrics, the Algorithm 2 increases the contrast errors in the reproduced image, relative to those of Algorithm 1 and [Mantiuk et al., 2008]. Both the DRIM and TMQI metrics estimate threshold differences in per-

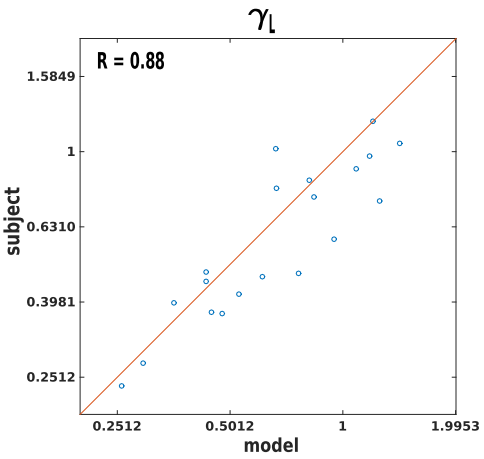


Figure 6.26: Average subject selected values from 7 observers plotted against model estimated parameter values for γ_L.

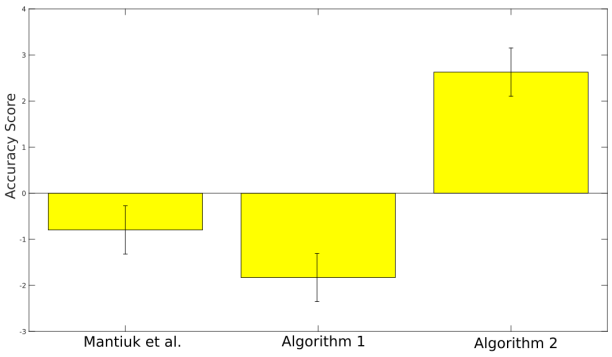


Figure 6.27: Result of experiment two. Pairwise comparison of 3 TMOs: [Mantiuk et al., 2008], Algorithm 1 and Algorithm 2.

TMO	DRIM	TMQI		
		Q	S	N
Algorithm 2	0.51	0.89	0.9	0.54
Algorithm 1	0.48	0.89	0.92	0.50
[Mantiuk et al., 2008]	0.46	0.90	0.92	0.51

Table 6.3: Quantitative evaluation using the Fairchild dataset [Fairchild, 2007].

ceived contrast and thus, if accurate, provide a metric that can determine whether image details are lost or in the case of DRIM whether contrast artifacts have been introduced. However, it is not clear to what extent image preference judgment are related to detail reproduction. Other factors such as global contrast, color vibrancy, etc may play a key role in determining the perceived quality of an image. Also, the reproduction of suprathreshold contrast levels may be important but DRIM and TMQI only consider threshold contrast.

6.3 Conclusion

In this chapter, we have proposed two image dependent TMOs with a global first stage and a local second stage.

The first TMO (Algorithm 1) is based on psychophysical and neural models of visual perception. The first stage performs range compression by a gamma transform, where the gamma curve is the one that best equalizes the lightness histogram. The second stage performs local contrast enhancement using neural activity models for the visual cortex. We compared our method to other contemporary TMOs using two computation metrics that estimate either the perceived contrast differences between high and low DR images (DRIM) or produce estimates of the perceived image quality (TMQI). The metrics indicate that our method compares well with the state of the art. The advantage of using a computational assessment is that many TMOs and base images can be tested in a short space of time, however we acknowledge that this analysis relies heavily upon the reliability of the metrics. Indeed, if the metrics are to be considered as ground truth, then the aims and motivations of any TMO should ultimately be to optimize the metric outputs (e.g. method discussed in Chapter 5). In the opinion of the author, our approach produces realistic images with no noticeable color

artifacts for all images tested.

The novelty of the results stems from the use of a perceptually derived model for estimating the preferred system gamma. This is an area of ongoing research and there are two research directions that could improve the model. The first is to develop a model of lightness perception so that it adapts precisely to the end viewing conditions and to the image in question. The second is to try different non-linearities. The gamma function is chosen because of its simplicity, but it is not optimized for the statistics of natural scenes.

The second TMO (Algorithm 2) operates in real time and overcomes the limitation of the previous TMO by adapting the non-linearity of the first stage based on natural image statistics. The second stage performs a simple contrast enhancement operation that mimics the contrast normalization process in the visual system. The results look natural, without halos. We optimized the automated parameter estimation method in an office viewing environment based on the users' parameter choice . The method works for video sequences and produces no flicker or spatio-temporal artifacts of any kind. We conducted a pairwise comparison experiment to compare our method with other state of the art TMOs. Results confirms that the proposed TMO (Algorithm 2) outperforms the state of the art in terms of user's preference.

Chapter 7

TONE MAPPING DEPENDENT ON VIEWING CONDITION

It is well known that viewing conditions and the display's capabilities in terms of contrast and brightness can significantly affect the perceived image quality. This happens mainly because the brightness perception of the HVS varies based on the viewing condition. The parameters of the first stage of the Algorithm 2 in the Chapter 6 are optimized based on a psychophysical experiment conducted on a LCD display in an office environment. Therefore, the result is ideal to be viewed in that condition and may be sub-optimal in other viewing conditions. This chapter proposes a method to adapt the result of our TMO depending on viewing condition. The content of this chapter is derived from [Cyriac et al., 2016a].

7.1 Brightness Perception and Viewing Condition Dependency

The effect of background and surround on the perception of brightness has been studied extensively for many decades.

In a classical work in imaging science, [Bartleson and Breneman, 1967] con-

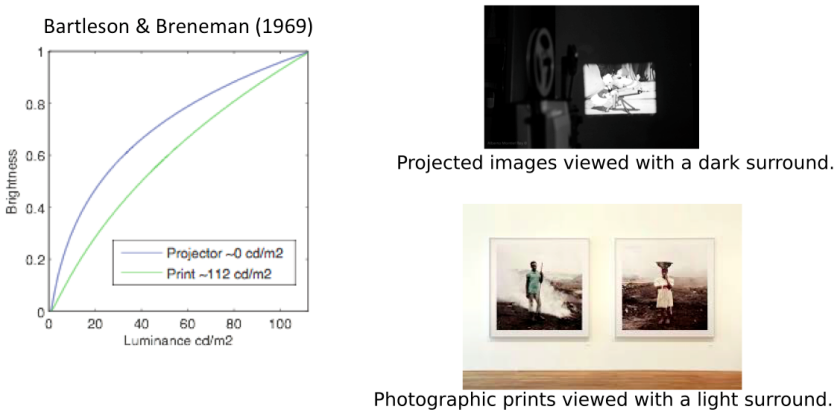


Figure 7.1: Brightness perception according to [Bartleson and Breneman, 1967] for different viewing conditions.

ducted experiments on brightness scaling and matching, with various elements of pictures viewed with different levels of luminance and surround lighting. They found that the surround luminance and the overall luminance level significantly affect the perceived brightness and proposed a simplified formula to predict the relative brightness [Bartleson, 1975]. The formula is similar to the CIELAB L^* with exponent 0.33 for dark, 0.41 for dim and 0.5 for light surrounds respectively (see Figure 7.1). And the ratio of the exponent is 1.5 : 1.25 : 1.

Liu and Fairchild [Liu and Fairchild, 2007] later conducted a pair comparison experiment to find the optimal gamma value that compensates for the surround effect. The experiment considered four tone mapped HDR images, six different surround conditions from total black to 125% of image white and eleven different gamma values. They proposed a simple linear formula to predict the experimental data:

$$\gamma = -0.1356 L_s + 1.137 \tag{7.1}$$

where L_s is the luminance of the surround proportional to the image white. By this equation, the ratio of the exponent is 1.14 : 1.1 : 1. This ratio is considerably less when compared to that of [Bartleson, 1975]. They also point out that the image content plays an important role in determining the optimal gamma.

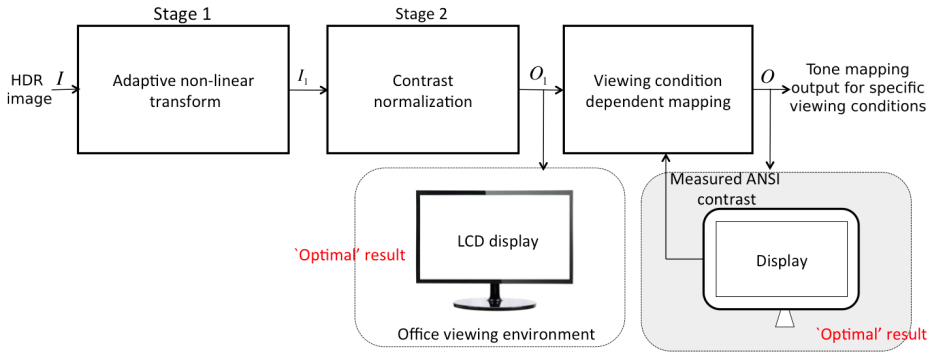


Figure 7.2: Block diagram of the complete model.

Recently, [Pindoria and Thompson, 2016] conducted a pair comparison experiment to determine the optimal system gamma for displays with different peak luminance (L_{peak}). They tested it for the peak luminance levels of 500, 1000, 2000 and 4000 cd/m^2 on a SIM2 HDR display. They proposed a formula to predict the experiment data:

$$\gamma = 1.2 + 0.42 log_{10}\left(\frac{L_w}{1000}\right) \tag{7.2}$$

The experiment considered only the effect of change in display brightness and did not take into account the change in surround and background luminance.

All the above studies are limited in terms of the different display types they used and/or in terms of the background and surround luminance condition they considered. Also, the advancement in the display technologies in the recent years has made available many new varieties of displays in the market. This all makes very relevant the further study of optimal gamma adjustment to compensate for the surround effect.

7.2 Proposed Approach

Figure 7.2 shows the block diagram of the complete proposed model. The first two stages are already explained in the previous chapter. The last stage performs the viewing condition dependent mapping that aims to adapt our tone mapped

result to each particular display and viewing environment in terms of luminance. We follow the approach of the related works in the previous section and propose the last stage as a simple gamma non-linear transform. In order to develop a general model that determines the necessary non-linear adjustment value γ_{adj} needed for our tone mapping results to look optimal under some given conditions, we perform the following steps:

1 Conduct psychophysical experiments to determine the optimal non-linear adjustment the subjects prefer for different surround conditions and displays.

2 Record several physical measurements including minimum and maximum luminance of the display, surround luminance, and ambient illuminance.

3 With the above data in hand we develop a general formula to predict the subject's choice of non-linearity value for a specific surround environment and display type.

7.3 Psychophysical Experiment

In this section we discuss the setup of the psychophysical experiment.

Seven subjects completed the experiment. All had corrected to normal vision. Ethics was approved by the Comité Etico de Investigacion Clinica, Parc de Salut MAR, Barcelona, Spain and all procedures complied with the declaration of Helsinki.

The viewing conditions that we consider are
Two surround environments:

- Office room: ambient illuminant of 47 nits and average near surround luminance of 65 nits
- Dark room: ambient illuminant of 0.3 nits and average near surround luminance ≈ 0 nits

and three display types:

- LCD: ASUS VS197D LCD monitor set to sRGB mode

Figure 7.3: Psychophysical experiment setup.

- OLED: Sony Trimaster PVM
- HDR: SIM2 HDR47ES4MB monitor set to HDR mode

The stimuli for the experiment are 20 tone mapped versions of the HDR images from the Fairchild HDR dataset [Fairchild, 2007]. The images were chosen to cover a variety of scenarios: night images, indoor scenes, bright outdoor scenes, landscapes, etc. The HDR images were tone mapped by the tone mapping Algorithm 2 discussed in Chapter 6.

Experiment procedure: A schematic of the experiment setup is shown in Figure 7.3. Subjects were asked to adjust the gamma non-linearity via a scrollbar such that the image achieves an optimal appearance. Table 7.3 shows the average subject choice for γ_{adj}, for seven observers (two observers in the case of HDR display), 20 images, 3 display types and two surround environments.

7.4 Contrast Measurement

For each combination of display type and surround condition we measured both the sequential and ANSI contrasts. A Konica Minolta LS 100 photometer was used to measure the luminance and the reading was taken at the distance at which observers view the display (approximately 3 times the display height). Results are shown in Tables 7.1 and 7.2.

7.5 Model to Predict the Experiment Result

Analyzing the results of the psychophysical experiment and the contrast measurement, we found that the subject-preferred non-linear adjustment value γ_{adj} can be predicted from the ANSI contrast and the maximum brightness of the display by the following formula:

$$\gamma_{adj} = (1 + 0.2|C|)^{sign(C)} \tag{7.3}$$

where

$$C = log_{10}\left(\frac{L_{vc}^{peak}}{L_{grad}^{peak}}\right) + log_{10}\left(ANSI_{vc}\right) - log_{10}\left(ANSI_{grad}\right), \tag{7.4}$$

L_{vc}^{peak} and $ANSI_{vc}$ are the peak brightness and the ANSI contrast of the intended display in which the image is to be viewed. L_{grad}^{peak} and $ANSI_{grad}$ are the peak brightness and the ANSI contrast of the grading display: an LCD display in an office environment in our experiments, since the parameter values for the first stage of our model were selected to optimize image appearance in that scenario. Table 7.3 shows that the model predicts well the result of the psychophysical experiment.

7.6 Results and Discussion

In Tables 7.1 and 7.2, we show the luminance and contrast measurements taken under the different display-surround setups designed above (unfortunately, for the HDR display we could only collect data from the office room condition). In general, the ANSI contrast is substantially smaller than the sequential contrast in the dark surround condition. This is mainly because some of the light emitted by the display is reflected back onto the screen by objects in the surround, resulting in raising the effective minimum value of the display. For example, the minimum luminance measured on the OLED is 0.001 cd/m^2 in a dark surround when the full screen is black, whereas the measured luminance is 0.3 cd/m^2 when a checkerboard pattern is displayed. This result is consistent with the results of [Schuck and Lude, 2015], [Tydtgat et al., 2015], where they measured both sequential contrast and the effective contrast by varying the image white content on cinema projectors (see Figure 7.4). We may interpret from these

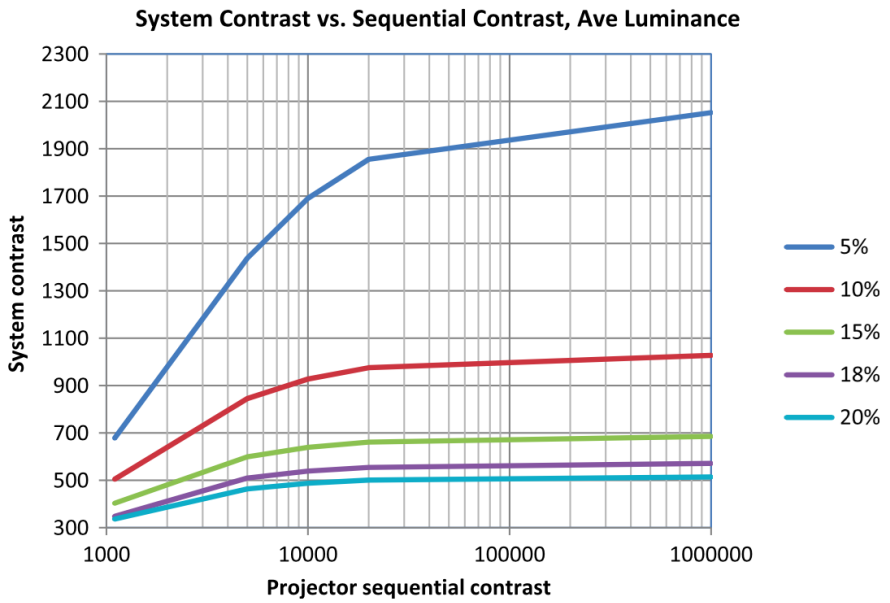


Figure 7.4: Plot of projector sequential contrast vs system contrast with different level of white content. Figure from [Schuck and Lude, 2015].

measurements that the effective contrast produced by a display depends not only on its maximum contrast ratio but also on the surround and the reproduced image content.

In Figure 7.5 we show the result produced by our method for three different viewing conditions. As the effective contrast of the display (ANSI contrast) increases, our method tries to compensate it by decreasing the contrast of the lower mid-intensities of the input image. Note that these images will look optimal only under the intended viewing conditions and display type.

7.7 Conclusion

Viewing conditions and the display's capabilities can significantly affect the perceived image quality. We recorded luminance measurements of several display types and surround conditions and showed that the effective contrast (ANSI con-

Table 7.1: Sequential contrast measurement.

	Dark room			Office room		
	Min luminance	Max luminance	Sequential contrast	Min luminance	Max luminance	Sequential contrast
LCD	0.35	170	486	2.3	170	74
OLED	0.001	97	97000	1.2	97	80
HDR	-	-	-	1.7	2700	1588

Table 7.2: ANSI contrast measurement.

	Dark room			Office room		
	Min luminance	Max luminance	ANSI contrast	Min luminance	Max luminance	ANSI contrast
LCD	0.7	170	242	2.6	170	65
OLED	0.3	97	323	1.3	97	74
HDR	-	-	-	2	2700	1350

Table 7.3: Comparison between user-chosen and model-predicted gamma adjustment.

	Dark		office	
	subject	model	subject	model
LCD	1.11	1.12	1	1
OLED	1.1	1.1	0.99	0.98
HDR	-	-	1.5	1.51

LCD office (65) OLED dark room (323) SIM2 office (1350)

Figure 7.5: Results of our method applied to an HDR video frame from the ARRI dataset [Froehlich et al., 2014] for three different viewing conditions. ANSI contrast given in parenthesis.

trast) produced by a display depends not only on its maximum contrast capability but also on the surround and the reproduced image content. We have conducted psychophysical experiments and developed a mathematical model to predict the users' chosen non-linear adjustment for the results produced by our TMO to look best in each viewing condition.

Chapter 8

APPLICATIONS

In this chapter we present potential applications of the tone mapping model (Algorithm 2) presented in Chapter 6.

8.1 Substitute for In-camera Gamma Correction

The non-linearity applied in most digital cameras is well approximated by a simple power law, and while this type of method is very fast and may perform well on average, in general it is suboptimal because it is not tailored to the specificity of each image. Existing apps for smartphones do increase the visible contrast, but at the expense of introducing noticeable artifacts when there is motion in the scene, or needing a considerable amount of user interaction. By implementing our model in-camera as a substitute for gamma correction, we can capture better, more realistic and natural-looking pictures and videos on our smartphones and cameras. See Figure 8.1 for a comparison between the result of our method and the default camera output and the result of highly rated mobile app proHDR [EyeApps, 2013].

Figure 8.1: Left column from top to bottom: output of Nikon D3100 camera, output of Samsung S6 mobile phone, output of Nexus 5 mobile phone, output of highly rated mobile app proHDR. Right column from top to bottom: results of applying our method to RAW sensor data, with better contrast, natural appearance and no artifacts.

8.2 Use at Several Stages of the Production Chain in the Media Industry

In photography and cinema the contrast transforms are applied off-line, in post-production, but not automatically but manually by a skilled technician: still, the artistic work most often starts only after the original image has been manually corrected so as to emulate the visual perception of the contrast of the scene at the shooting location, therefore in the context of professional post-production there is also a need for an automated, good-quality, perceptually-consistent contrast transform like our method provides. Figure 8.2 shows that the result of our method (bottom right) is similar to that of the manual output by a skilled colorist (bottom left), whereas conventional non-linear transform results in loss of visible details (top).

For the media industries, our model can also be used at several stages of the production chain, allowing for better and faster work during shoots and a lesser

Figure 8.2: Top: conventional non-linear transform of a frame from a video sequence captured by an ARRI- Alexa professional digital cinema camera. Bottom left: manual output by a skilled colorist. Bottom right: result of our method. Notice visibility of details in background. Image courtesy of ARRI.

need for post-production time:

- on the set, for cinema and TV shoots that use state-of-the-art extended dynamic range cameras, allowing correct monitorization of video signals on standard displays;
- during post-production (both for photography and cinema), automatically emulating the results produced manually by skilled technicians;
- for distribution, exhibition and broadcast, providing real-time conversion of a high contrast video signal (of the original content or coming from the TV control room) to the standard dynamic range of regular displays and projection systems.

8.3 Cascaded Linear and Nonlinear Neural Models

In the article "Derivatives and Inverse of Cascaded Linear and Nonlinear Neural Models" [Martinez et al., 2017] we present a specific four-layer vision model that accounts for brightness perception, contrast computation, local masking, and orientation/scale masking. The first layer is based on the model (Algorithm 2) presented in Chapter 6, the second and third layers were previously introduced in [Malo and Simoncelli, 2015], and the fourth layer performs a multi-scale/orientation image decomposition and divisive normalization. Each of the four layers consists of a linear and a non-linear part. We optimize the parameters of each of these layers. The second and third layers are optimized through the Maximum Differentiation (MAD) competition methodology [Wang and Simoncelli, 2008] as in [Malo and Simoncelli, 2015]. But this methodology was not suitable for the first layer because of the large inter-subject variability in choosing the parameter settings. Therefore, we optimize the parameters through a brute force approach by searching in the parameter space for values that maximize the correlation between the model response and the average subject rating of the image distortions in the TID2008 dataset [Ponomarenko et al., 2009] (details in Appendix A). A pictorial illustration of the complete model is given in Figure 8.3 and is explained below.

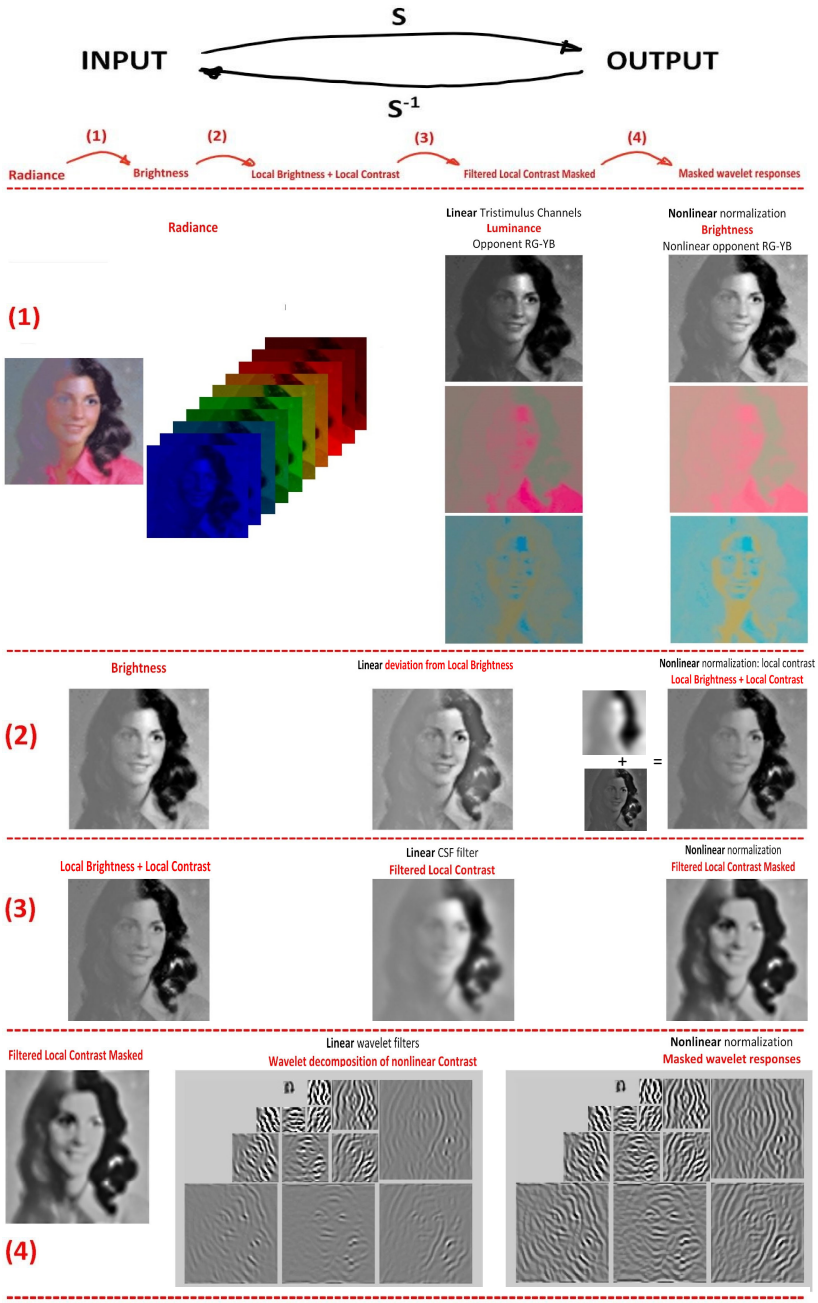


Figure 8.3: Cascaded Linear and Nonlinear Neural Models. Figure from [Martinez et al., 2017].

First layer: luminance to brightness mapping

The linear part of the first stage accounts for the spectral integration at the photoreceptors and the separate processing of achromatic and opponent chromatic information. It applies a linear spectral integration of the input signal (x_0) in each spatial location and then a transform to an opponent color space:

$$y_1 = A.T^{\lambda}.x_0, \tag{8.1}$$

where T^{λ} contains the spectral sensitivities (or color matching functions) for each spatial location, (e.g. functions tuned to long, medium and short wavelengths, LMS), and the matrix A performs the LMS-to-opponent color space transform.

Then, the non-linear part transforms the luminance to brightness using the method developed in Section 6.2.1 (the current implementation considers only the luminance channel):

$$x_1 = (L_1)^{\gamma_1(L_1)}, \tag{8.2}$$

where L_1 is the luminance linearly scaled to [0,1], and

$$\gamma_1(L_1) = \gamma_H - (\gamma_L - \gamma_H)(1 - \frac{L_1^m}{L_1^m + \mu^m}). \tag{8.3}$$

The exponent γ_1 has value γ_L for low luminance and γ_H for high luminance, with a smooth transition at $L_1 = \mu$. The slope of the transition is controlled by the exponent m. The optimal parameter values are: $\gamma_L = 0.2$, $\gamma_H = 0.3$, $\mu = 0.018$ and $m = 1.12$. The details of the parameter optimization are given in Appendix A.

Second layer: local brightness and local contrast

The second layer accounts for local brightness and local contrast estimation. The linear part computes the local brightness by convolving the brightness image with a Gaussian kernel g_A:

$$y_2 = g_A \odot x_1. \tag{8.4}$$

In the non-linear part the local contrast is estimated through divisive normalization [Carandini and Heeger, 2012], where the brightness is subtracted and

normalized by the local average brightness:

$$\mathcal{N}(x_1) = \frac{x_1 - g_B \odot x_1}{b_2 + g_C \odot x_1}, \tag{8.5}$$

where g_B and g_C are Gaussian kernels and b_2 a constant. The output of the second layer is a weighted average of the local brightness and local contrast:

$$x_2 = \alpha_1 y_2 + \alpha_2 \mathcal{N}(x_1). \tag{8.6}$$

The parameters of this stage are optimized as in [Malo and Simoncelli, 2015] and are as follows: the spatial width (in degrees) of the Gaussian kernels g_A, g_B and g_C is 0.066, $\alpha_1 = 1$, $\alpha_2 = 180$ and $b_2 = 30$.

Third layer: frequency sensitivity and local masking

The third layer accounts for the frequency-dependent sensitivity (through the Contrast Sensitivity Function, CSF [Mullen, 1985] [Campbell and Robson, 1968]) and the local masking. The linear part applies a CSF filter [Watson and Malo, 2002], L_{CSF} to the output of second layer:

$$y_3 = L_{CSF} \odot x_2. \tag{8.7}$$

The non-linear part performs a divisive normalization to the result of the linear part:

$$x_3 = sign(y_3) \frac{|y_3|^{\gamma_3}}{b_3 + g_3 \odot |y_3|^{\gamma_3}}, \tag{8.8}$$

where $\gamma_3 = 1.5$, $b_3 = 0.04$ and the width (in degrees) of the Gaussian kernel g_3 is 0.02.

Fourth layer: masking in wavelet domain

The fourth layer accounts for the band-pass local orientation filters in V1 [Marĉelja, 1980] [Watson, 1987] and their mutual inhibitory interactions responsible for frequency dependent masking [Watson and Solomon, 1997]. The linear part applies linear multi-scale, multi-orientation image decomposition [Simoncelli and Freeman, 1995] of the output of the third stage, to which a divisive normalization is applied to get the final output of our vision model

$$x_4 = sign(y_4) \frac{|y_4|^{\gamma_4}}{b_4 + g_4 \odot |y_4|^{\gamma_4}}, \tag{8.9}$$

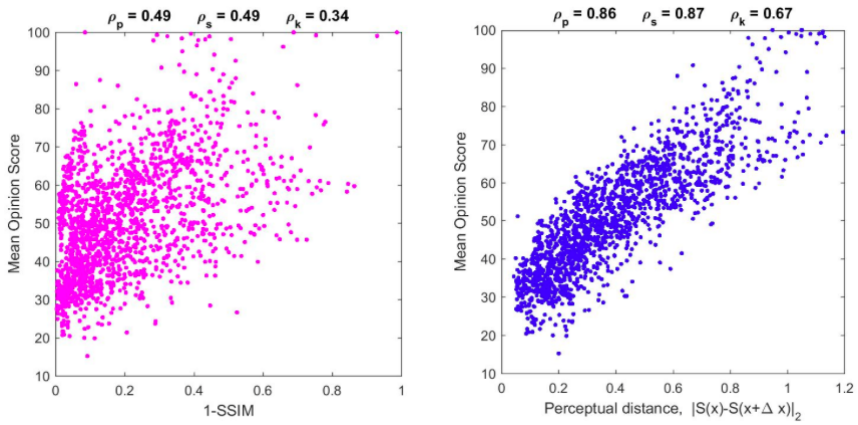


Figure 8.4: The subjective image distortion problem consists in predicting the opinion of human observers on the perceived difference between pairs of images, x_A, and $x_B = x_A + \Delta x$. This may be thought as assessing the visibility of the distortion Δx when put on top of x_A. The TID database [Ponomarenko et al., 2009] includes the experimental visibility of 1700 deviations Δx of different nature and energy put on top of a set of images of different content. These plots show the behavior of two models (abscesses) in predicting the perception (ordinates): a widely used model in engineering, SSIM [Wang et al., 2004] (at the left); and the model we propose in [Martinez et al., 2017] (at the right). Pearson, Spearman and Kendall correlations between predictions and perception are given on top.

where y_4 is the output of the linear stage, $b_4 = 0.6$, $\gamma_4 = 0.6$ and the width (in degrees) of the Gaussian kernel g_4 is 0.24.

In order to see how well the proposed four-layer framework models the human visual system response, we compare the model predicted result with that of the psychophysical data of rating the perceived distortions between pairs of images. We found that the proposed model correlates well with the psychophysical data on subjective distance [Ponomarenko et al., 2009] and outperforms the prediction of SSIM (see Figure 8.4).

8.4 Dehazing

The presence of light scattering substances such as fog, haze or smoke can greatly reduce the visibility of the objects in a scene and results in producing an image with low contrast and faded colors (see Figure 8.5, top middle). The dehazing problem tries to recover the details and color of a hazy image (see Figure 8.5, top left), and has gained much attention in the recent years. The main limitation of the current dehazing approaches is that they tend to produce images with a considerable amount of artifacts (see Figure 8.6).

In [Vazquez-Corral et al., 2016], we show that by adapting the tone mapping Algorithm 2 explained in Chapter 6, we can perform dehazing with a very limited amount of image artifacts (see Figure 8.6). We operate in the HSV color space based on the observation that when an image of a haze-free scene is compared with an image of the same scene with haze, the hue remains almost constant, the saturation is always higher in the haze-free image and the value is always greater in the hazy image. We illustrate the above observation in Figure 8.5. To have a better understanding of the difference between the corresponding HSV components of the two images we also show the histogram of their differences in the third column. Note that as the hue component is circular, and normalized between 0 and 1, having a distance of 1 is equivalent to having the same hue. The saturation difference is always positive while the value difference is always negative. Another point to note is that the histogram of the V component of a hazy image is more skewed towards 1. Distributing these values uniformly would help in increasing the contrast and visibility.

8.4.1 Model in Operation

The input hazy image I is initially converted into the HSV color space. We apply our two-stage model only to the value (V) component while the saturation (S) component is scaled based on the modified V. The channel V is linearized by applying the inverse of the encoding gamma (assumed to be gamma = 2.2) and is passed through the two-stage model. The first global stage (Section 6.2.1) performs a constrained histogram equalization resulting in a uniform distribution of the intensity values and hence produces better contrast. The second stage performs a local contrast enhancement (Section 6.2.2) resulting in recovering local image details.

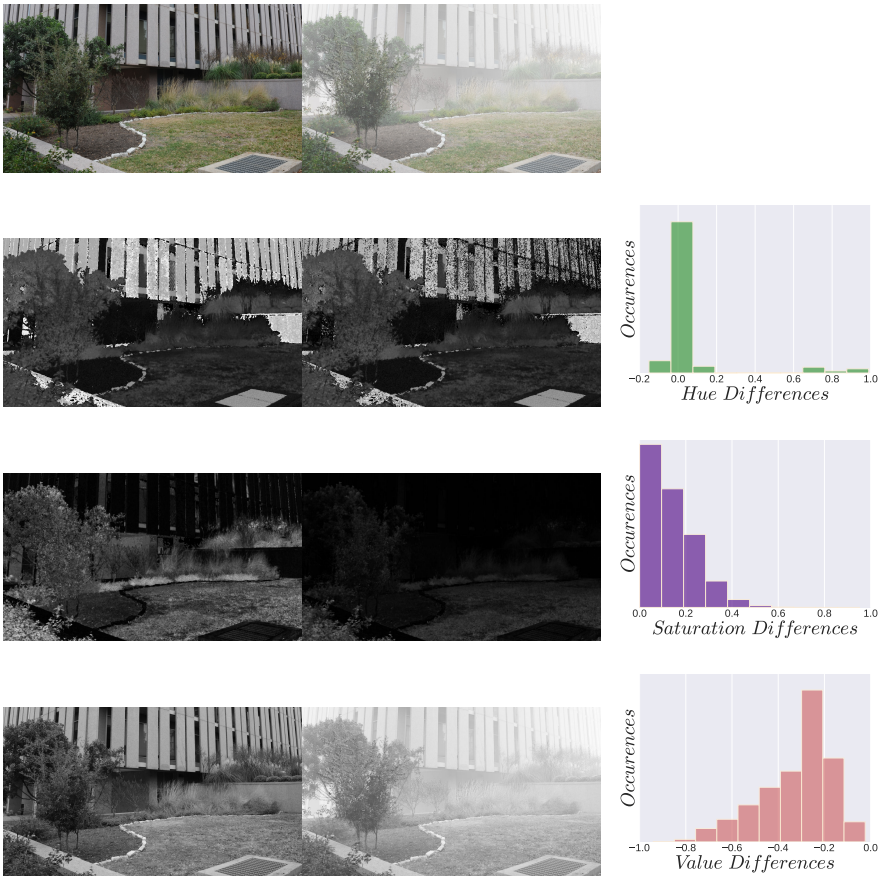


Figure 8.5: Comparison between a fog and a non-fog image in the HSV space. First two columns, from top to bottom: Original image, hue, saturation and value. Last column: Histogram of differences between the fog and the non-fog image. We can see that hue stays the same, while saturation is always higher for the non-fog image, and value is always higher for the fog image.

Figure 8.6: An example of artifacts appearing in current image dehazing methods. From left to right: Original image, result of [He et al., 2011], result of [Meng et al., 2013], result of the method we propose in [Vazquez-Corral et al., 2016].

The S component is modified based on the observation that subjects tend to perceive the same level of colorfulness in a scene independent of the quantity of haze present [Mizokami et al., 2016] [Mizokami et al., 2015]. The measure of colorfulness can be related to the HSV-Chroma, i.e., the product of S and V. Now, to keep the chroma of the hazy and dehazed images similar we modify the saturation of the hazy image S_{in}, taking into account the modification that we made to the value of the hazy image V_{in} as follows:

$$S_{out} = \left(\frac{V_{in}}{V_{out}} S_{in} \right)^{\alpha} \tag{8.10}$$

where α controls the amount of saturation (set to 0.8 in our experiment) and V_{out} is the output value component.

Figure 8.7 shows the comparison of our result with some state of the art dehazing methods. We can see that for standard hazy images our results are comparable with the results of the state of the art.

Figure 8.8 shows the result of popular dehazing algorithms applied to challenging images. Other methods tend to produce considerable amount of artifacts while enhancing the contrast, whereas the proposed method produces reliable results while keeping the artifacts to a minimum.

Figure 8.7: Dehazing applied to standard images. From top to bottom: Original, [Choi et al., 2015], [Meng et al., 2013], [Tarel and Hautiere, 2009], [He et al., 2011] and result of the method we propose in [Vazquez-Corral et al., 2016].

Figure 8.8: Challenging images. From top to bottom: Original, [Choi et al., 2015], [He et al., 2011], [Meng et al., 2013], [Tarel and Hautiere, 2009] and result of the method we propose in [Vazquez-Corral et al., 2016].

8.5 Conclusion

We have presented several applications of the proposed tone mapping model in this chapter: for the media industry, as the first layer of a state-of-the art vision model, and for dehazing. For the media industry, our model can be used for on-set monitoring on regular displays, as a substitute for gamma-correction, and as a way of providing the colorist with content that is both natural looking and has a crisp and clear appearance. We showed that the response of the four layer vision model with the first layer based on the proposed tone mapping method (Algorithm 2) correlates well with the psychophysical data of rating the perceived distortions between pairs of images. We also showed that our tone mapping model can be used for a dehazing application and the result competes well with the state of the art for standard images, and clearly outperforms them when dealing with challenging images.

Chapter 9

CONCLUSION

The main goal of this thesis was to propose tone mapping algorithms that ensure that the displayed content appears realistic and appealing to the viewers in terms of detail and contrast visibility. To this end, we looked into neural and psychophysical models of the visual system, natural image statistics, and also the effect of display capabilities and the viewing environment on perception.

State of the art TMOs are still unable to produce LDR images that perfectly match the perceived contrast of the HDR images. In Chapter 5, we presented a general framework for improving any tone mapped image by reducing the distance with the corresponding HDR image in terms of a non-local perceptual metric. The distance is minimized by means of a gradient descent algorithm. Dealing with the dynamic range independent metric DRIM, we tested the framework for different tone mapped images provided by several TMOs. Experiments confirmed that our method substantially reduces the perceptual distance with the HDR source image, providing an average improvement of more than 25% for several state of the art TMOs. The proposed gradient descent method might stop in local minima. Therefore, when provided with an input tone mapped image that is closer, in terms of the quality metric, with the corresponding HDR image our method might provide a better result.

In Chapter 6, we proposed two TMOs based on the visual perception models. The efficient coding hypothesis and neuroscience experiments indicate that the visual system, when presented with natural scenes, transforms input signals to

ensure a uniform distribution of the output levels. In addition, a psychophysical study showed that subjects tend to prefer images with a flat lightness histogram, however, complete histogram equalization may produce results with unnatural appearance due to sharp changes in contrast. The proposed TMOs consist of two stages, a global adaptive non-linear transform followed by a local contrast enhancement step.

In the first TMO (Section 6.1), the non-linear transform is achieved by means of a simple gamma curve that best flattens the lightness histogram. Contrast enhancement is performed by a variational method that models the neural activity of the visual cortex. We compared the proposed method with other TMOs using two image quality metrics that incorporate a model of human vision. The results showed that our method compares well with the state of the art.

In the second TMO (Section 6.2), we improved the first stage by replacing the simple gamma transform by a function γ estimated using some key features of the cumulative histogram of the input image based on the natural image statistics. We replaced the second stage with a simple contrast enhancement step that is based on the neuroscience studies showing that the visual system is performing an operation called contrast normalization that normalizes the local contrast by a factor depending on the standard deviation of the light intensity. These modifications allow executing the proposed tone mapping in real time. The method also ensures that the results look natural and are without any spatio-temporal artifacts. The parameters of the first stage were optimized through psychophysical experiments in an office viewing environment. We validated our method through psychophysical tests that confirm that proposed approach outperforms other state of the art algorithms in terms of users' preference.

The viewing condition in which the display is viewed greatly affects the perceived image quality. This happens mainly because of the difference in the human brightness perception in different surround condition and also because the effective display contrast (ANSI contrast) varies depending on the surround environment. As a result, the image produced by our tone mapping method, which is optimized for office condition, may not be ideal for viewing in other surround conditions. In Chapter 7, we conducted a psychophysical experiment to determine the adequate gamma non-linearity needed for the result produced by our

TMO to look best in each viewing condition. We also developed a mathematical model to predict the experimental data from the ANSI contrast and peak luminance of the display.

The applications of the proposed approach include: providing a fast and automatic graded content to the colorist, replacing the gamma correction stage in the camera image processing pipeline for reliable results, use as a TMO, on-set monitoring of HDR cameras, real time grading of HDR content for television broadcast, and dehazing.

Future work

The tone mapping method developed in this thesis is based on the natural image statistics derived from photographs and not from HDR images. Further research into image statistics of HDR images (real world luminance) would allow us to improve our model.

We also recognize that the proposed method does not take into account the artistic intent when tone mapping HDR contents. We would like to investigate into the manual color grading process and try to adapt it into our model.

The mathematical model developed in Chapter 7 performs viewing condition dependent mapping based on the psychophysical experiment conducted with limited number of display types, viewing conditions and subjects. We would like to extend the psychophysical experiment to see if we can improve the model.

Bibliography

[ITU, 2016] (2016). Report ITU-R BT.2390-0, high dynamic range television for production and international programme exchange.

[Adams, 1983] Adams, A. (1983). *The print. The Ansel Adams Photography Series*. Little, Brown and Company.

[Adams et al., 2012] Adams, J., Deever, A. T., Morales, E. O., and Pillman, B. H. (2012). Perceptually based image processing algorithm design. *Perceptual Digital Imaging: Methods and Applications*, 6:123.

[Aggarwal and Ahuja, 2004] Aggarwal, M. and Ahuja, N. (2004). Split aperture imaging for high dynamic range. *International Journal of Computer Vision*, 58(1):7–17.

[Ahn et al., 2013] Ahn, H., Keum, B., Kim, D., and Lee, H. S. (2013). Adaptive local tone mapping based on retinex for high dynamic range images. In *2013 IEEE International Conference on Consumer Electronics (ICCE)*, pages 153–156. IEEE.

[Andriani et al., 2013] Andriani, S., Brendel, H., Seybold, T., and Goldstone, J. (2013). Beyond the kodak image set: A new reference set of color image sequences. In *2013 IEEE International Conference on Image Processing*, pages 2289–2293. IEEE.

[Ashikhmin, 2002] Ashikhmin, M. (2002). A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics workshop on Rendering*, pages 145–156. Eurographics Association.

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

[Aydin et al., 2008] Aydin, T. O., Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2008). Dynamic range independent image quality assessment. In *ACM SIGGRAPH 2008 Papers*, SIGGRAPH '08, pages 69:1–69:10, New York, NY, USA. ACM.

[Aydin et al., 2014] Aydin, T. O., Stefanoski, N., Croci, S., Gross, M., and Smolic, A. (2014). Temporally coherent local tone mapping of hdr video. *ACM Transactions on Graphics (TOG)*, 33(6):196.

[Baker, 2010] Baker, S. (2010). Led backlighting, tft central. `http://www.tftcentral.co.uk/articles/led_backlighting.htm`.

[Banterle et al., 2011] Banterle, F., Artusi, A., Debattista, K., and Chalmers, A. (2011). *Advanced high dynamic range imaging: theory and practice*. CRC Press.

[Banterle et al., 2012] Banterle, F., Artusi, A., Sikudova, E., Bashford-Rogers, T., Ledda, P., Bloj, M., and Chalmers, A. (2012). Dynamic range compression by differential zone mapping based on psychophysical experiments. In *Proceedings of the ACM Symposium on Applied Perception*, pages 39–46. ACM.

[Barten, 1999] Barten, P. G. (1999). *Contrast sensitivity of the human eye and its effects on image quality*, volume 72. SPIE press.

[Barten, 2003] Barten, P. G. (2003). Formula for the contrast sensitivity of the human eye. In *Electronic Imaging 2004*, pages 231–238. International Society for Optics and Photonics.

[Bartleson, 1975] Bartleson, C. (1975). Optimum image tone reproduction. *Journal of the SMPTE*, 84(8):613–618.

[Bartleson and Breneman, 1967] Bartleson, C. and Breneman, E. (1967). Brightness perception in complex fields. *JOSA*, 57(7):953–956.

[Belle, 2016] Belle, K. V. (2016). Laser-phosphor illumination in projectors (barco white paper).

[Bertalmío, 2014] Bertalmío, M. (2014). From image processing to computational neuroscience: a neural model based on histogram equalization. *Frontiers in computational neuroscience*, 8.

[Bertalmío et al., 2007] Bertalmío, M., Caselles, V., Provenzi, E., and Rizzi, A. (2007). Perceptual color correction through variational techniques. *IEEE Transactions on Image Processing*, 16(4):1058–1072.

[Bertalmío and Cowan, 2009] Bertalmío, M. and Cowan, J. D. (2009). Implementing the retinex algorithm with wilson–cowan equations. *Journal of Physiology-Paris*, 103(1):69–72.

[Boitard et al., 2012] Boitard, R., Bouatouch, K., Cozot, R., Thoreau, D., and Gruson, A. (2012). Temporal coherency for video tone mapping. In *SPIE Optical Engineering+ Applications*, pages 84990D–84990D. International Society for Optics and Photonics.

[Boitard et al., 2014] Boitard, R., Cozot, R., Thoreau, D., and Bouatouch, K. (2014). Zonal brightness coherency for video tone mapping. *Signal Processing: Image Communication*, 29(2):229–246.

[Borer and Cotton, 2016] Borer, T. and Cotton, A. (2016). A display-independent high dynamic range television system. *SMPTE Motion Imaging Journal*, 125(4):50–56.

[Brainard, 1997] Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10:433–436.

[Brajovic et al., 1998] Brajovic, V. M., Miyagawa, R., and Kanade, T. (1998). Temporal photoreception for adaptive dynamic range image sensing and encoding. *Neural Networks*, 11(7):1149–1158.

[Brenner et al., 2000] Brenner, N., Bialek, W., and Van Steveninck, R. d. R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702.

[Buchsbaum, 1980] Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26.

[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.

[Campbell and Robson, 1968] Campbell, F. W. and Robson, J. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of physiology*, 197(3):551.

[Carandini and Heeger, 2012] Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62.

[Cardei and Funt, 1999] Cardei, V. C. and Funt, B. (1999). Committee-based color constancy. In *IS&T/SID?s Color Imaging Conference*, pages 311–313.

[Cerdá-Company et al., 2016] Cerdá-Company, X., Párraga, C. A., and Otazu, X. (2016). Which tone-mapping operator is the best? a comparative study of perceptual quality. *arXiv preprint arXiv:1601.04450*.

[Choi et al., 2015] Choi, L. K., You, J., and Bovik, A. C. (2015). Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901.

[Choudhury and Tumblin, 2005] Choudhury, P. and Tumblin, J. (2005). The trilateral filter for high contrast images and meshes. In *ACM SIGGRAPH 2005 Courses*, page 5. ACM.

[Croci et al., 2016] Croci, S., Aydın, T. O., Stefanoski, N., Gross, M., and Smolic, A. (2016). Real-time temporally coherent local hdr tone mapping. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 889–893. IEEE.

[Cyriac et al., 2013] Cyriac, P., Batard, T., and Bertalmío, M. (2013). A variational method for the optimization of tone mapping operators. In *Pacific-Rim Symposium on Image and Video Technology*, pages 505–516. Springer.

[Cyriac et al., 2014] Cyriac, P., Batard, T., and Bertalmío, M. (2014). A non-local variational formulation for the improvement of tone mapped images. *SIAM Journal on Imaging Sciences*, 7(4):2340–2363.

[Cyriac et al., 2015a] Cyriac, P., Bertalmio, M., Kane, D., and Vazquez-Corral, J. (2015a). A tone mapping operator based on neural and psychophysical models of visual perception. In *SPIE/IS&T 2015 Human vision and electronic imaging XX*. International Society for Optics and Photonics.

[Cyriac et al., 2015b] Cyriac, P., Kane, D., and Bertalmío, M. (2015b). Perceptual dynamic range for in-camera image processing. In *British Machine Vision Conference (BMVC); 2015 Sept 7-10; p. 19.1-11*.

[Cyriac et al., 2016a] Cyriac, P., Kane, D., and Bertalmío, M. (2016a). Automatic, viewing-condition dependent contrast grading based on perceptual models. In *Annual Technical Conference and Exhibition, SMPTE*, pages 1–11.

[Cyriac et al., 2016b] Cyriac, P., Kane, D., and Bertalmío, M. (2016b). Optimized tone curve for in-camera image processing. *Electronic Imaging*, 2016(13):1–7.

[Daly, 1992] Daly, S. J. (1992). Visible differences predictor: an algorithm for the assessment of image fidelity. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pages 2–15. International Society for Optics and Photonics.

[Damera-Venkata and Chang, 2009] Damera-Venkata, N. and Chang, N. L. (2009). Display supersampling. *ACM Transactions on Graphics (TOG)*, 28(1):9.

[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.

[Delahunt et al., 2005] Delahunt, P. B., Zhang, X., and Brainard, D. H. (2005). Perceptual image quality: Effects of tone characteristics. *Journal of electronic imaging*, 14(2):023003–023003.

[Dierickx et al., 1996] Dierickx, B., Scheffer, D., Meynants, G., Ogiers, W., and Vlummens, J. (1996). Random addressable active pixel image sensors. In *Advanced Imaging and Network Technologies*, pages 2–7. International Society for Optics and Photonics.

[Drago et al., 2003a] Drago, F., Myszkowski, K., Annen, T., and Chiba, N. (2003a). Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426.

[Drago et al., 2003b] Drago, F., Myszkowski, K., Annen, T., and Chiba, N. (2003b). Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426.

[Duan et al., 2010] Duan, J., Bressan, M., Dance, C., and Qiu, G. (2010). Tone-mapping high dynamic range images by novel histogram adjustment. *Pattern Recognition*, 43(5):1847–1862.

[Dudley et al., 2003] Dudley, D., Duncan, W. M., and Slaughter, J. (2003). Emerging digital micromirror device (dmd) applications. In *Micromachining and Microfabrication*, pages 14–25. International Society for Optics and Photonics.

[Durand and Dorsey, 2002] Durand, F. and Dorsey, J. (2002). Fast bilateral filtering for the display of high-dynamic-range images. In *ACM transactions on graphics (TOG)*, volume 21, pages 257–266. ACM.

[Ebner and Fairchild, 1998] Ebner, F. and Fairchild, M. D. (1998). Development and testing of a color space (ipt) with improved hue uniformity. In *Color and Imaging Conference*, volume 1998, pages 8–13. Society for Imaging Science and Technology.

[Eilertsen et al., 2015] Eilertsen, G., Mantiuk, R. K., and Unger, J. (2015). Real-time noise-aware tone mapping. *ACM Trans. Graph*, 34(6):198.

[Eilertsen et al., 2013] Eilertsen, G., Wanat, R., Mantiuk, R. K., and Unger, J. (2013). Evaluation of tone mapping operators for hdr-video. In *Computer Graphics Forum*, volume 32, pages 275–284.

[EyeApps, 2013] EyeApps (2013). `https://play.google.com/store/apps/details?id=com.eyeappsllc.prohdr\&hl=en`.

[Fairchild, 2007] Fairchild, M. D. (2007). The hdr photographic survey. In *Color and Imaging Conference*, volume 2007, pages 233–238. Society for Imaging Science and Technology.

[Fairchild, 2013] Fairchild, M. D. (2013). *Color appearance models*. John Wiley & Sons.

[Fattal et al., 2002] Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 249–256. ACM.

[Ferradans et al., 2011] Ferradans, S., Bertalmio, M., Provenzi, E., and Caselles, V. (2011). An analysis of visual adaptation and contrast perception for tone mapping. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):2002–2012.

[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*, pages 249–258. ACM.

[Froehlich et al., 2014] Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A., and Brendel, H. (2014). Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays.

[Gil Rodríguez and Bertalmío, 2016] Gil Rodríguez, R. and Bertalmío, M. (2016). High quality video in high dynamic range scenes from interlaced dual-iso footage. *Electronic Imaging*, 2016(18):1–7.

[Gil Rodríguez et al., 2015] Gil Rodríguez, R., Vazquez-Corral, J., and Bertalmío, M. (2015). The intrinsic error of exposure fusion for hdr imaging, and a way to reduce it. In *British Machine Vision Conference (BMVC); 2015*.

[Ginosar and Gnusin, 1997] Ginosar, R. and Gnusin, A. (1997). A wide dynamic range cmos image sensor. In *IEEE Workshop on CCD and Advanced Image Sensors*, volume 32.

[Granados et al., 2013] Granados, M., Kim, K. I., Tompkin, J., and Theobalt, C. (2013). Automatic noise modeling for ghost-free hdr reconstruction. *ACM Transactions on Graphics (TOG)*, 32(6):201.

[Gu et al., 2010] Gu, J., Hitomi, Y., Mitsunaga, T., and Nayar, S. (2010). Coded rolling shutter photography: Flexible space-time sampling. In *Computational*

Photography (ICCP), 2010 IEEE International Conference on, pages 1–8. IEEE.

[Guild, 1932] Guild, J. (1932). The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 230(681-693):149–187.

[Hajsharif et al., 2014] Hajsharif, S., Kronander, J., and Unger, J. (2014). Hdr reconstruction for alternating gain (iso) sensor readout. In *Eurographics, Strasbourg, France, April 7-11, 2014*.

[He et al., 2011] He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353.

[Hecht, 1934] Hecht, S. (1934). Vision: Ii. the nature of the photoreceptor process.

[Huang and Mumford, 1999] Huang, J. and Mumford, D. (1999). Statistics of natural images and models. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE.

[Hung, 2005] Hung, P. (2005). *Image Sensors and Signal Processing for Digital Still Cameras, chapter Color theory and its application to digital still cameras*. CRC.

[Hunt, 1995] Hunt, R. (1995). The reproduction of colour in photography, printing and television, 5th edition.

[IMS-CHIPS, 2002] IMS-CHIPS (2002). `http://www.ims-chips.de/content/pdftext/HDRC_Imager_Camera_Feature3.pdf`.

[Kane and Bertalmio, 2015] Kane, D. and Bertalmio, M. (2015). Dynamic range, light scatter in the eye and perceived image quality. In *PERCEPTION*, volume 44, pages 123–124.

[Kane and Bertalmío, 2015] Kane, D. and Bertalmío, M. (2015). Is there a preference for linearity when viewing natural images? In *SPIE/IS&T 2015 Image Quality and System Performance XX*. International Society for Optics and Photonics.

[Kane and Bertalmío, 2016] Kane, D. and Bertalmío, M. (2016). System gamma as a function of image-and monitor-dynamic range. *Journal of vision*, 16(6):4–4.

[Kastner and Baccus, 2014] Kastner, D. B. and Baccus, S. A. (2014). Insights from the retina into the diverse and general computations of adaptation, detection, and prediction. *Current opinion in neurobiology*, 25:63–69.

[Kiser et al., 2012] Kiser, C., Reinhard, E., Tocci, M., and Tocci, N. (2012). Real time automated tone mapping system for hdr video. In *IEEE International Conference on Image Processing*, pages 2749–2752.

[König and Dieterich, 1892] König, A. and Dieterich, C. (1892). *Die Grundempfindungen in normalen und anomalen Farbensystemen und ihre Intensitätsverteilung im Spektrum*. L. Voss.

[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):461–468.

[Kuang et al., 2007] Kuang, J., Johnson, G. M., and Fairchild, M. D. (2007). icam06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414.

[Kunkel and Reinhard, 2010] Kunkel, T. and Reinhard, E. (2010). A reassessment of the simultaneous dynamic range of the human visual system. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 17–24. ACM.

[Lam and Fung, 2008] Lam, E. and Fung, G. (2008). *Single-sensor imaging: Methods and applications for digital cameras, chapter Automatic White Balancing in Digital Photography*. CRC press.

[Land and McCann, 1971] Land, E. H. and McCann, J. J. (1971). Lightness and retinex theory. *JOSA*, 61(1):1–11.

[Larson et al., 1997] Larson, G. W., Rushmeier, H., and Piatko, C. (1997). A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306.

[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 Transactions on Graphics (TOG)*, volume 24, pages 640–648. ACM.

[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.

[Liu and Fairchild, 2007] Liu, C. and Fairchild, M. D. (2007). Re-measuring and modeling perceived image contrast under different levels of surround illumination. In *Color and Imaging Conference*, volume 2007, pages 66–70. Society for Imaging Science and Technology.

[Ma et al., 2015] Ma, K., Yeganeh, H., Zeng, K., and Wang, Z. (2015). High dynamic range image compression by optimizing tone mapped image quality index. *IEEE Transactions on Image Processing*, 24(10):3086–3097.

[MagicLantern, 2014] MagicLantern (2014). `http://www.magiclantern.fm/index.html`.

[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*, 20(6):1558–1571.

[Malo et al., 2006] Malo, J., Epifanio, I., Navarro, R., and Simoncelli, E. P. (2006). Nonlinear image representation for efficient perceptual coding. *IEEE Transactions on Image Processing*, 15(1):68–80.

[Malo and Simoncelli, 2015] Malo, J. and Simoncelli, E. P. (2015). Geometrical and statistical properties of vision models obtained via maximum differentiation. In *SPIE/IS&T Electronic Imaging*, pages 93940L–93940L. International Society for Optics and Photonics.

[Mandle, 2010] Mandle, G. (2010). Oled; what is it and how does it work? In *SMPTE Tech Conference & Expo, 2010 Annual*, pages 1–17. SMPTE.

[Mann and Picard, 1995] Mann, S. and Picard, R. W. (1995). On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. *Proceedings of IS&T*, pages 442–448.

[Mantiuk et al., 2008] Mantiuk, R., Daly, S., and Kerofsky, L. (2008). Display adaptive tone mapping. *ACM Trans. Graph.*, 27(3):68:1–68:10.

[Mantiuk et al., 2005] Mantiuk, R., Daly, S. J., Myszkowski, K., and Seidel, H.-P. (2005). Predicting visible differences in high dynamic range images: model and its calibration. In *Electronic Imaging 2005*, pages 204–214. International Society for Optics and Photonics.

[Mantiuk and Heidrich, 2009] Mantiuk, R. and Heidrich, W. (2009). Visualizing high dynamic range images in a web browser. *Journal of Graphics, GPU, and Game Tools*, 14(1):43–53.

[Mantiuk et al., 2011] Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W. (2011). Hdr-vdp-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on Graphics (TOG)*, volume 30, page 40. ACM.

[Mantiuk et al., 2004] Mantiuk, R., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2004). Perception-motivated high dynamic range video encoding. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 733–741. ACM.

[Mantiuk et al., 2009] Mantiuk, R., Tomaszewska, A., and Heidrich, W. (2009). Color correction for tone mapping. In *Computer Graphics Forum*, volume 28, pages 193–202.

[Marĉelja, 1980] Marĉelja, S. (1980). Mathematical description of the responses of simple cortical cells. *JOSA*, 70(11):1297–1300.

[Martinez et al., 2014] Martinez, L. M., Molano-Mazón, M., Wang, X., Sommer, F. T., and Hirsch, J. A. (2014). Statistical wiring of thalamic receptive fields optimizes spatial sampling of the retinal image. *Neuron*, 81(4):943–956.

[Martinez et al., 2017] Martinez, M., Cyriac, P., Batard, T., Bertalmio, M., and malo, J. (2017). Derivatives and inverse of cascaded linear + nonlinear neural models (in preparation).

[Massof et al., 1988] Massof, R. W., Marcus, S., Dagnelie, G., Choy, D., Sunness, J., and Albert, M. (1988). Theoretical interpretation and derivation of flash-on-flash threshold parameters in visual system diseases. *Applied optics*, 27(6):1014–1024.

[Meng et al., 2013] Meng, G., Wang, Y., Duan, J., Xiang, S., and Pan, C. (2013). Efficient image dehazing with boundary constraint and contextual regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 617–624.

[Menon and Calvagno, 2011] Menon, D. and Calvagno, G. (2011). Color image demosaicking: An overview. *Signal Processing: Image Communication*, 26(8):518–533.

[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.

[Mizokami et al., 2015] Mizokami, Y., Takahashi, Y., and Yaguchi, H. (2015). Colorfulness perception of natural images adjusting to haze. In *The 23rd Symposium of the International Colour Vision Society*.

[Mizokami et al., 2016] Mizokami, Y., Takahashi, Y., and Yaguchi, H. (2016). Stable colorfulness perception of scene through haze. In *Vision Sciences Society (VSS)*.

[Morovic, 1998] Morovic, J. (1998). To develop a universal gamut mapping algorithm (condensed format edition).

[MPI, 2004] MPI (2004). `http://www.mpi-inf.mpg.de/resources/hdr/gallery.html`.

[Mullen, 1985] Mullen, K. T. (1985). The csf of human colour vision to red-green and yellow-blue chromatic gratings.

[Myers, 2003] Myers, R. L. (2003). *Display interfaces: fundamentals and standards*. John Wiley & Sons.

[Nakamura, 2005] Nakamura, J. (2005). *Image sensors and signal processing for digital still cameras*. CRC press.

[Normann and Perlman, 1979] Normann, R. A. and Perlman, I. (1979). The effects of background illumination on the photoresponses of red and green cones. *The Journal of physiology*, 286:491.

[Olshausen and Field, 2000] Olshausen, B. A. and Field, D. J. (2000). Vision and the coding of natural images. *American Scientist*, 88(3):238–245.

[Olshausen and Field, 2005] Olshausen, B. A. and Field, D. J. (2005). How close are we to understanding v1? *Neural computation*, 17(8):1665–1699.

[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*, pages 47–54. ACM Press/Addison-Wesley Publishing Co.

[Pelli, 1997] Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4):437–442.

[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(7):1020–1030.

[Pharr and Humphreys, 2004] Pharr, M. and Humphreys, G. (2004). *Physically based rendering: From theory to implementation*. Morgan Kaufmann.

[Pindoria and Thompson, 2016] Pindoria, M. and Thompson, S. (2016). Image adaptation requirements for high dynamic range video under reference and non-reference viewing conditions.

[Ponomarenko et al., 2009] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., and Battisti, F. (2009). Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45.

[Poynton, 2012] Poynton, C. (2012). *Digital video and HD: Algorithms and Interfaces*. Elsevier.

[Preiss et al., 2014a] Preiss, J., Fairchild, M. D., Ferwerda, J. A., and Urban, P. (2014a). Gamut mapping in a high-dynamic-range color space. In *IS&T/SPIE Electronic Imaging*, pages 90150A–90150A. International Society for Optics and Photonics.

[Preiss et al., 2014b] Preiss, J., Fernandes, F., and Urban, P. (2014b). Color-image quality assessment: From prediction to optimization. *IEEE Transactions on Image Processing*, 23(3):1366–1378.

[Purves and Lotto, 2003] Purves, D. and Lotto, R. B. (2003). *Why we see what we do: An empirical theory of vision.* Sinauer Associates.

[Ramanath et al., 2005] Ramanath, R., Snyder, W. E., Yoo, Y., and Drew, M. S. (2005). Color image processing pipeline. *IEEE Signal Processing Magazine*, 22(1):34–43.

[Ramsey et al., 2004] Ramsey, S. D., Johnson III, J. T., and Hansen, C. (2004). Adaptive temporal tone mapping. In *Proceedings of the 7th IASTED International Conference on Computer Graphics and Imaging*, pages 124–128. Citeseer.

[Reinhard and Devlin, 2005] Reinhard, E. and Devlin, K. (2005). Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):13–24.

[Reinhard et al., 2010] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K. (2010). *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.

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

[Rizzi et al., 2003] Rizzi, A., Gatta, C., and Marini, D. (2003). A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, 24(11):1663–1677.

[Ruderman, 1994] Ruderman, D. L. (1994). The statistics of natural images. *Network: computation in neural systems*, 5(4):517–548.

[Rust and Movshon, 2005] Rust, N. C. and Movshon, J. A. (2005). In praise of artifice. *Nature neuroscience*, 8(12):1647–1650.

[Sapiro and Caselles, 1997] Sapiro, G. and Caselles, V. (1997). Histogram modification via differential equations. *Journal of Differential Equations*, 135(2):238–268.

[Schlick, 1995] Schlick, C. (1995). Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*, pages 7–20. Springer.

[Schuck and Lude, 2015] Schuck, M. and Lude, P. (2015). An analysis of system contrast in digital cinema auditoriums. In *SMPTE 2015 Annual Technical Conference and Exhibition*, pages 1–19. SMPTE.

[Seetzen et al., 2004] Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., and Vorozcovs, A. (2004). High dynamic range display systems. *ACM Transactions on Graphics (TOG)*, 23(3):760–768.

[Simoncelli and Freeman, 1995] Simoncelli, E. P. and Freeman, W. T. (1995). The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *ICIP (3)*, pages 444–447.

[Singnoo and Finlayson, 2010] Singnoo, J. and Finlayson, G. D. (2010). Understanding the gamma adjustment of images. In *Color and Imaging Conference*, volume 2010, pages 134–139. Society for Imaging Science and Technology.

[Smith et al., 2006] Smith, K., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2006). Beyond tone mapping: Enhanced depiction of tone mapped hdr images. In *Computer Graphics Forum*, volume 25, pages 427–438.

[SMPTE, 2014] SMPTE, S. (Aug, 2014). 2084: 2014. *High dynamic range electro-optical transfer function of mastering reference displays*.

[Soneira, 2005] Soneira, R. M. (2005). Display technology shoot-out: Comparing crt, lcd, plasma and dlp displays. `http://www.displaymate.com/ShootOut_Part_1.htm`.

[Soneira, 2014] Soneira, R. M. (2014). 2014 innovative displays and display technologies smartphones, tablets, tvs, and wearable displays. `http://www.displaymate.com/Display_Technology_2014.htm#Quantum_Dots`.

[Spencer et al., 1995] Spencer, G., Shirley, P., Zimmerman, K., and Greenberg, D. P. (1995). Physically-based glare effects for digital images. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 325–334. ACM.

[Stevens and Stevens, 1963] Stevens, J. and Stevens, S. S. (1963). Brightness function: Effects of adaptation. *JOSA*, 53(3):375–385.

[Stockman and Sharpe, 2000] Stockman, A. and Sharpe, L. T. (2000). The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737.

[Tarel and Hautiere, 2009] Tarel, J.-P. and Hautiere, N. (2009). Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2201–2208. IEEE.

[Thurstone, 1927] Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

[Tomasi and Manduchi, 1998] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE.

[Tomaszewska and Mantiuk, 2007] Tomaszewska, A. and Mantiuk, R. (2007). Image registration for multi-exposure high dynamic range image acquisition.

[Tumblin and Rushmeier, 1993] Tumblin, J. and Rushmeier, H. (1993). Tone reproduction for realistic images. *IEEE Computer graphics and Applications*, 13(6):42–48.

[Tumblin and Turk, 1999] Tumblin, J. and Turk, G. (1999). Lcis: A boundary hierarchy for detail-preserving contrast reduction. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 83–90.

[Tydtgat et al., 2015] Tydtgat, C., Maes, D., Stojmenovik, G., and Grillet, A. (2015). Modelling of achievable contrast and its impact on hdr projection in commercial cinema environments. In *SMPTE 2015 Annual Technical Conference and Exhibition*, pages 1–14. SMPTE.

[Vazquez-Corral et al., 2016] Vazquez-Corral, J., Galdran, A., Cyriac, P., and Bertalmio, M. (2016). A fast image dehazing method that does not introduce color artifacts (submitted).

[Wang et al., 2010] Wang, J., Xu, D., Lang, C., and Li, B. (2010). An adaptive tone mapping method for displaying high dynamic range images. *J. Inf. Sci. Eng.*, 26(3):977–990.

[Wang and Bovik, 2002] Wang, Z. and Bovik, A. C. (2002). A universal image quality index. *IEEE signal processing letters*, 9(3):81–84.

[Wang et al., 2004] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

[Wang and Simoncelli, 2008] Wang, Z. and Simoncelli, E. P. (2008). Maximum differentiation (mad) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12):8–8.

[Watson, 1987] Watson, A. B. (1987). The cortex transform: rapid computation of simulated neural images. *Computer vision, graphics, and image processing*, 39(3):311–327.

[Watson and Malo, 2002] Watson, A. B. and Malo, J. (2002). Video quality measures based on the standard spatial observer. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages III–41. IEEE.

[Watson and Solomon, 1997] Watson, A. B. and Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *JOSA A*, 14(9):2379–2391.

[Wilson and Cowan, 1972] Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1.

[Wilson and Cowan, 1973] Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80.

[Wright, 1929] Wright, W. D. (1929). A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141.

[Wyszecki and Stiles, 1982] Wyszecki, G. and Stiles, W. S. (1982). *Color science*, volume 8. Wiley New York.

[Yeganeh and Wang, 2013] Yeganeh, H. and Wang, Z. (2013). Objective quality assessment of tone-mapped images. *Image Processing, IEEE Transactions on*, 22(2):657–667.

[Yoshida et al., 2005] Yoshida, A., Blanz, V., Myszkowski, K., and Seidel, H.-P. (2005). Perceptual evaluation of tone mapping operators with real-world scenes. In *Electronic Imaging 2005*, pages 192–203. International Society for Optics and Photonics.

[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. In *Computer Graphics Forum*, volume 25, pages 415–426.

[Zimmer et al., 2011] Zimmer, H., Bruhn, A., and Weickert, J. (2011). Freehand hdr imaging of moving scenes with simultaneous resolution enhancement. In *Computer Graphics Forum*, volume 30, pages 405–414.

Appendix A

PARAMETER OPTIMIZATION

The parameters of the second and third layers of the neural models in Chapter 8 are optimized through the Maximum Differentiation (MAD) competition methodology [Wang and Simoncelli, 2008] as in [Malo and Simoncelli, 2015]. In the following sections, we explain the MAD competition methodology for parameter optimization and look into its applicability for optimizing the parameters of the first layer. Then, we fit the parameters of the first layer to maximize the correlation with the psychophysical data on subjective distance [Ponomarenko et al., 2009].

A.1 Parameter Optimization: MAD Competition

This is a methodology for optimizing parameters of a perceptual model using synthesized images that can optimally distinguish the model behavior with different parameter settings. For a parameter setting of the model, an image pair is synthesized from an image I_{org} in such a way that these images maximize/minimize the model response while keeping the mean square error (MSE) with I_{org} fixed. The image pair is synthesized based on the second-order approximation of the model response distance [Malo et al., 2006]. The image regions in I_{org}, which upon distorting maximize/minimize the model response, are given by the subspace corresponding to high/low eigenvalues of the transformed Riemannian metric [Malo and Simoncelli, 2015]. Therefore, we synthesize the image pair

by adding a fixed amount of noise to the corresponding image regions. This pair of images is the most perceptually different according to the model with the given parameter setting. By varying the parameter values, different image pairs are synthesized. The best parameter setting should be the one that produces the most perceptually distinguishable pair. Through a psychophysical experiment, subjects can select the most distinct pair indicating the best parameter setting.

We follow the approach of [Malo and Simoncelli, 2015] and optimize the parameters of the first layer by keeping the parameters of other layers fixed. Our model operates only on the luminance channel, so we directly give the display luminance (L_{disp}) as input to the nonlinear part of the first layer. Due to the unavailability of an HDR display we optimize the parameters based on LDR images displayed on an LDR display.

We optimize the parameters γ_L and γ_H of Eq. 8.3 while keeping the semi-saturation μ and slope m fixed per image:

$$\mu = exp\left(mean\left(log(L_{disp})\right)\right) \tag{A.1}$$

$$m = -4.5/log(\mu). \tag{A.2}$$

This formulation to compute m makes sure that the γ function (Eq. 8.3) has the value γ_L and γ_H for lowest and highest luminance values respectively. The parameter space considered for γ_L and γ_H is: $\gamma_L \in \{0.3, 0.405, 0.548, 0.74, 1\}$ and $\gamma_H \in \{0.1, 0.1565, 0.245, 0.3834, 0.6\}$.

In the MAD competition (psychophysical experiment), five image pairs are shown in each trial, where those images are generated by keeping one parameter fixed and varying the other. Subjects are asked to choose the most distinct pair. The order in which the parameters are optimized is randomized. Therefore, in the second trial, if the parameter being optimized is different from the first trial, then five new image pairs are generated by keeping the selected parameter value from the previous trial and varying the other, else the same image pairs as in the first trial are displayed. This process is repeated for 10 trials. The order of displaying the image pairs in each trial is also randomized. The final parameter values of γ_H and γ_L after 10 trials are recorded. A screenshot of the image pairs displayed in a trial is shown in Figure A.1.

Figure A.1: A screenshot of a MAD optimization trial.

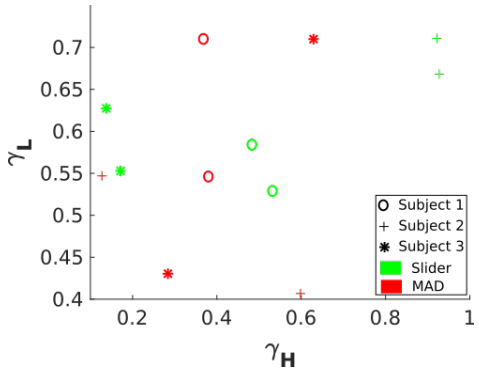


Figure A.2: Comparison between MAD and slider approaches of parameter optimization.

A.2 Slider Optimization

Alternatively, we consider the slider experiment, similar to the approach used in the Experiment one of Section 6.2.5, to optimize the parameters of the first layer. The output of the first layer is given as input to the display. Subjects manipulate the parameters γ_H and γ_L via two scrollbars until the most pleasing image is achieved. The range of values for γ_H is from 0 to 1 and for γ_L is from 0 to 3. μ and m are computed as in the MAD optimization case.

A.3 Comparison between MAD and Slider Optimization

Figure A.2 shows the scatter plot between the subject selected optimal value for γ_L and γ_H for an image. The three different symbols correspond to three subjects and the green color for the slider optimization and the red color for the MAD optimization. Each subject performs both experiments twice. We can see in the figure that, for the MAD optimization, the subject-selected optimal parameter values vary considerably between different subjects and also with the same subject in different trails. The slider optimization also shows a similar trend but the variability within the subjects is lower.

Now, we look into the psychophysical data from the tone mapping parameter optimization experiment (Experiment one of Section 6.2.5) to see the relationship between the dynamic range and the inter-subject variability. In Figure A.3,

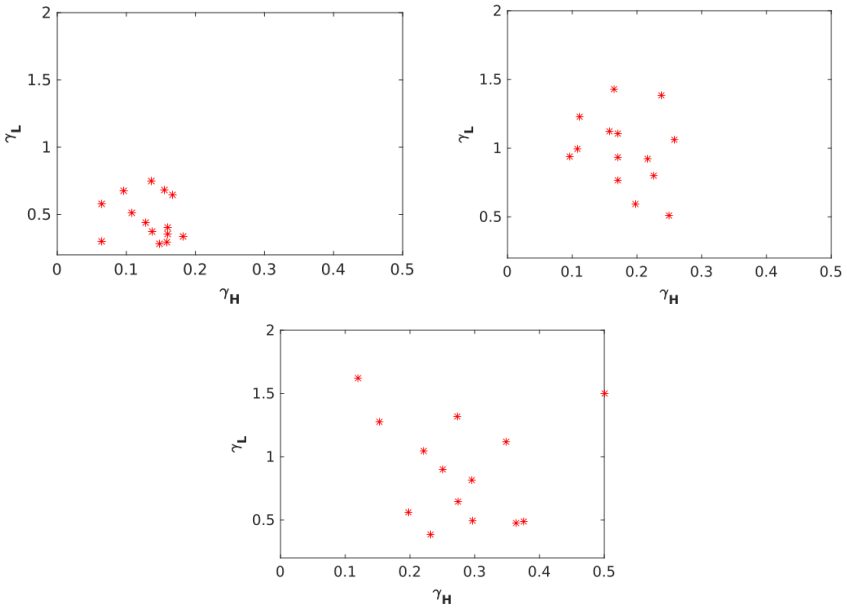


Figure A.3: Scatter plot between the user-selected value of γ_L and γ_H from the slider experiment for a high, medium and low dynamic range image in clockwise order.

we show the scatter plots between the user-selected value for γ_L and γ_H for a high, medium and low dynamic range image in clockwise order. We can see that as the dynamic range of the image decreases the inter-subject variability increases. The large inter-subject variability in choosing the parameter settings for LDR images makes the optimization process difficult, and the MAD or slider optimization approaches may not provide an efficient solution.

A.4 Optimization based on TID2008 Dataset

The TID2008 dataset [Ponomarenko et al., 2009] consists of 25 original LDR images along with their distorted images. Seven different types of distortions in four different levels were used to generate 28 distorted images for each original image. Subjects' Mean Opinion Score (MOS) consisting of the average subjects rating (0 -100 scale) of the distorted images with respect to the original images

is also provided along with the dataset.

We extend our model into a quality metric to take advantage of the MOS and the TID2008 dataset to optimize the parameters of the first layer. The metric takes both the original and the distorted images as inputs and processes them separately through the four-layer model and then computes the mean square error. If the four-layer model is a good approximation of the HVS response, a simple mean square error between the two outputs of the model should give a good approximation of the subject ratings of the distortions.

In order to optimize the parameters of the first layer (γ_H, γ_L and μ), we follow a brute force approach and search in the parameter space to find the value that maximizes the correlation between the output of the developed quality metric and the MOS. The best correlation is obtained for $\gamma_H = 0.3$, $\gamma_L = 0.2$ and $\mu = 0.018$.