UNIVERSIDADE DA BEIRA INTERIOR
Engenharia

Quality Perception and Chromatic Changes in Digital Images

Marco André Vieira Andrade Bernardo

Tese para obtenção do Grau de Doutor em
Engenharia Informática
(3º Ciclo de Estudos)

Orientador: Prof. Doutora Maria Manuela Areias da Costa Pereira de Sousa
Co-orientador: Prof. Doutor Paulo Torrão Fiadeiro

Covilhã, Junho 2017

Dedicatory

To my parents,
for all the love...

Acknowledgments

This thesis is more than a “physical” result of my latest years of work. It is also the product of a journey of learning, knowledge sharing, friendship and lots of intense emotions!

I would therefore like to thank first to the person who gave me the opportunity to start this journey, Prof. Manuela Pereira, my supervisor, for accepting this challenge from the beginning, for the guidance and for being the keystone in this research group.

I would like also to express my gratitude and friendship to my co-supervisor, Prof. Paulo Fiadeiro, for encouraging my research and allowing to grow as a researcher. Your advice, during all of this time, were priceless.

Thanks also to Prof. António Pinheiro, my co-supervisor (not officially), for all of your assertiveness, commitment capacity, and for bringing to the research group all of the salutary discussions. Without you the end of this thesis would be unachievable.

Part of the work described in this thesis would have not also been possible without the collaborations with many brilliant people in EPFL and CTU Prague, with whom I had the opportunity to cooperate. Therefore, I would like to thank specially to Touradj Ebrahimi, Karel Fliegel, Philippe Hanhart, Pavel Korshunov and Lukas Krasula.

To all of my academic lecturers, at the Department of Computer Science, that from the beginning of my academic path tried to transmit me the best of their knowledge. Thereby, helping me to achieve the outcome of this thesis. For this reason, thanks to you all.

I am also grateful to all of my colleges of the Optics Center, namely Toni, Jorge, Daniel, Rui, Anita, Pedro, Miguel, Rafael, and particularly to Gil from the SOCIA Lab. Thank you all for the support and friendship.

To the AC people, namely Sérgio, Francisco, Joaquim, João, Ricardo, Jorge, Nuno, and in general to all friends in Leomil. Thanks to you all. Without you guys I would have lost completely my sanity.

A very special thanks to my family. Words cannot express how grateful I am to my parents António and Maria, my sister Sílvia, my brother-in-law Rafael and to my nephews Dinis and Pedro, for supporting and encouraging me during these long years, as always!

Finally to Monica. Thanks for being who you are in my life. It is great to know that your patience and support has no limits.

Thanks to all again and I hope to continue and to count on your support for the next stages of my future life!

Tudo vence uma vontade obstinada, todos os obstáculos abate o homem que integrou na sua vida o fim a atingir e que está disposto a todos os sacrifícios para cumprir a missão que a si próprio se impôs.

Agostinho da Silva

Resumo

O trabalho apresentado nesta tese pretende fornecer contribuições significativas para a investigação e desenvolvimento na avaliação da qualidade multimédia, particularmente centrada sobre a avaliação da percepção e modelos de qualidade de imagem e também no estudo da sensibilidade do consumidor para alterações cromáticas induzidas em imagens digitais.
A motivação principal é a avaliação da percepção da qualidade com o objetivo de fornecer meios para melhorar a capacidade de interação do consumidor com novas tendências em tecnologia de imagem, uma percepção mais realista dos conteúdos, e consequentemente, alcançando novos níveis de qualidade de experiência. O desenvolvimento de novos modelos de representação de informação visual está a ser disponibilizado como uma fonte de fidelidade visual, que por sua vez é mais próxima dos limites de percepção humana. Estes modelos de representação, tais como tecnologias de sistemas hiperespectrais e de multi-exposição, podem resultar em novos formatos de imagem que recentemente têm originado grande interesse das comunidades científicas, profissionais e artísticas.
O primeiro tema desenvolvido nesta tese descreve uma investigação sobre os efeitos de alterações cromáticas induzidas em imagens e a subsequente influência na qualidade da experiência percebida pelos consumidores. Os resultados demonstraram que os observadores revelam maior sensibilidade com a perda da naturalidade da cor do que com a própria magnitude do erro induzido. Além disso, os dados de fixação visual revelaram uma alta correlação com a qualidade de imagem percebida pelos observadores.
O segundo tema de investigação desta tese é dedicado à avaliação do desempenho do JPEG XT na codificação de imagens HDR.
Este estudo, feito em colaboração com outros grupos de investigação pertencentes ao Qualinet, beneficiou da experiência e conhecimento adquiridos no primeiro tema de pesquisa. Além disso, permitiu a aplicação de conhecimento adquirido sobre qualidade em atividades de normalização. Os perfis JPEG XT foram verificados e a aplicação de métricas objetivas para avaliação da qualidade foram analisadas. Foi concluído que as métricas desenvolvidas especificamente para o domínio HDR, demonstram ser indicadores confiáveis de qualidade. Além disso, se as imagens HDR forem convertidas para domínios perceptuais as métricas objetivas desenvolvidos para o domínio LDR também revelam um bom desempenho.
Esta tese é concluída com um comentário final sobre os resultados alcançados e as possíveis direções de investigação para desenvolvimentos futuros.

Palavras-chave

Informação Cromática, Qualidade de Experiência, Avaliação da Qualidade de Imagem, Métricas de Qualidade de Imagem, Percepção de Qualidade, Mapas de Fixação, Representação de Imagem, Codificação de Imagem, HDR, JPEG XT.

Resumo alargado

Nos últimos anos, com o crescimento da oferta de serviços multimédia resultou também num aumento da demanda de qualidade. Os consumidores de multimédia possuem diariamente um conjunto de novos serviços que estimulam por sua vez a competição entre os diferentes prestadores de serviços e da evolução das próprias tecnologias multimédia. Medidas relacionadas com o desempenho objetivo de um sistema, tais como a Qualidade de Serviço, tornam-se altamente insatisfatórias para uma definição real de um sistema ou mesmo da qualidade da aplicação. Isto acontece porque esta medida possui lacunas na avaliação da percepção da qualidade do utilizador dos serviços, uma vez que este tipo de avaliação não é considerado. Por este motivo, uma nova tendência aponta para a aplicação do conceito de Qualidade de Experiência, que tende a avaliar a experiência perceptual do consumidor, os sentimentos e sensações produzidas, sendo também altamente dependente do contexto e expectativas do próprio consumidor. Assim, a Qualidade de Experiência é um tipo de avaliação que envolve fatores subjetivos do utilizador final, e por esta razão é mais apropriada para a avaliação do desempenho dos sistemas reais. Embora os métodos subjetivos de avaliação da qualidade sejam considerados os que resultam em medidas de qualidade mais fiáveis, tais avaliações são dispendiosas, geralmente demoradas e complexas de serem implementadas. Assim, a avaliação com métodos objetivos da qualidade são também de extrema importância.
Esta tese visa fornecer contribuições científicas sobre a avaliação da qualidade de multimédia, com foco na percepção de qualidade e alterações cromáticas em imagens digitais, com o objetivo de melhorar a capacidade de interação do utilizador com as novas tendências em tecnologia de imagem, proporcionando uma percepção mais realista do conteúdo, e, consequentemente, atingindo novos níveis de Qualidade de Experiência.
Sabendo que a forma como a cor é exibida influência diretamente como ela é percebida, consequentemente a sua representação incorreta pode induzir o julgamento errôneo das cenas observadas. Assim, o primeiro tema de pesquisa desta tese é investigar de que forma são percebidos os efeitos de variações cromáticas induzidas em imagens e como influenciam a Qualidade de Experiência.
Com as recentes tendências de crescimento de novas tecnologias de imagem, alguns novos sistemas estão a tornar-se disponíveis para o público em geral. Um destes novos sistemas é o HDR, que como sistemas de aquisição estão presentes numa grande variedade de equipamento fotográfico e mesmo em alguns dispositivos móveis. Estes têm como objetivo a representação de toda a gama de cores dinâmica percebida pelo sistema visual humano. Por esta razão, e também porque é essencial a avaliação dos sistemas HDR em termos de Qualidade de Experiência, o segundo tema de pesquisa desta tese é investigar como o novo padrão para codificação de conteúdo HDR (JPEG XT), influencia a qualidade da imagem no domínio HDR.

Esta tese está organizada como descrito a seguir. O Capítulo 1 fornece uma breve introdução sobre a tese, incluindo a motivação e o âmbito, bem como a sua organização e contribuições. O restante conteúdo desta tese está dividido em três partes.

A Parte I aborda o estado da arte, onde particularmente no Capítulo 2 é apresentada uma detalhada revisão de conceitos e princípios da visão das cores e sistemas fundamentais de representação de cor, dando maior ênfase a aspetos que são fundamentalmente relevantes no âmbito desta tese. Começa por descrever toda a avaliação da percepção visual da cor, seguida por uma

descrição dos sistemas de representação de cores. No Capítulo 3 é fornecida uma introdução aos conceitos básicos de avaliação de qualidade de imagem. Inclui uma descrição geral dos principais passos a serem considerados quando se projeta uma avaliação da qualidade subjetiva e objetiva, bem como metodologias de validação. Estes conceitos serão amplamente utilizados em todos os estudos relatados nesta tese.

A Parte II descreve um dos principais temas desta tese, a percepção e modelação de alterações cromáticas induzidas em imagens coloridas. A influência de alterações cromáticas na qualidade da imagem é estudada no Capítulo 4. Este capítulo começa com uma descrição detalhada do método proposto para induzir alterações cromáticas em imagens coloridas, seguido da descrição de uma avaliação subjetiva realizada para quantificar a sensibilidade dos observadores a estas alterações cromáticas. Além disso, é também descrita uma avaliação objetiva da qualidade da imagem usando um extenso conjunto de métricas de qualidade existentes, seguido pela descrição de uma nova métrica de qualidade proposta para a avaliação da qualidade de imagem na presença de alterações cromáticas. A partir deste estudo resultaram, uma avaliação subjetiva de qualidade [1], uma avaliação objetiva de qualidade através de métricas com referência [2], uma apresentação num seminário [3], e uma avaliação objetiva de qualidade comparando resultados provenientes de métricas com referência e sem referência [4]. Um estudo mais genérico neste domínio foi publicado em revista [5]. No Capítulo 5 é detalhado o estudo na qualidade da percepção em imagens com alterações cromáticas com erros específicos, permitindo dar continuidade ao estudo apresentado no capítulo anterior. Começa com uma descrição detalhada do método proposto para induzir erros cromáticos específicos em imagens coloridas, seguida por duas experiências de avaliação subjetiva, incluindo também o registo dos movimentos dos olhos dos observadores. Além disso, é descrita também uma análise comparativa entre os resultados subjetivos e os dados resultantes do rastreamento ocular. A partir deste estudo resultaram as seguintes três publicações: a primeira inclui uma avaliação subjetiva de qualidade [6], e uma avaliação subjetiva da qualidade em conjunto com uma avaliação do comportamento do movimento ocular [7]. Todas estas publicações em conjunto e adicionando um estudo da importância da medição do tempo de fixação relativo sob erros cromáticos específicos, originaram um artigo de revista submetido [8].

A Parte III aborda a qualidade proporcionada pela extensão da norma JPEG, conhecida como JPEG XT especificamente a parte para codificação de imagens HDR. Estes estudos foram feitos em colaboração com parceiros da Qualinet, especificamente a EPFL (Suíça) e a CTU em Praga (República Checa), e permitiu utilizar o conhecimento adquirido na Parte II num estudo colaborativo e aplicado a atividades de normalização.
No Capítulo 6 são reportados dois estudos com objetivo de analisar a qualidade objetiva da performance do standard JPEG XT para a codificação HDR. Inicialmente é apresentado um primeiro estudo, resultante de um trabalho colaborativo com dois laboratórios de investigação, no EPFL (Suíça) e no CTU (República Checa), e está focada no desempenho dos perfis do JPEG XT usando um conjunto de imagens de teste HDR. Além disso, é também descrita uma análise comparativa dos resultados. A partir deste estudo resultou na seguinte publicação [9]. Na secção seguinte, é reportado um segundo estudo que incide sobre a influência da codificação JPEG XT na geração de TMO's. Tal como no estudo anterior, é avaliada a performance dos perfis JPEG XT usando um conjunto de imagens de teste HDR. Além disso, um método de avaliação da qualidade objetiva, e uma análise comparativa dos resultados é também descrito, resultando na seguinte publicação [10]. É reportado e analisado no Capítulo 7 um estudo que resultou de um trabalho

colaborativo com um laboratório de investigação no EPFL (Suíça), tendo como objetivo a avaliação comparativa de métricas objetivas de qualidade para avaliação da qualidade de imagem HDR. Inicia com uma pequena descrição sobre uma avaliação subjetiva de qualidade, seguido pela avaliação objetiva da qualidade de imagem. De seguida são descritos os resultados dos índices de desempenho com o objetivo de avaliar a precisão, monotonicidade, e consistência das métricas de qualidade objetiva selecionadas, seguido por uma análise estatística dos índices de desempenho tendo em conta a discriminação entre essas métricas objetivas. Este capítulo é constituído por um estudo anterior sobre compressão HDR [11], e um estudo extensivo sobre a avaliação comparativa de métricas objetivas de qualidade para avaliação da qualidade de imagem HDR, que em conjunto resultou num artigo de revista [12].

Por fim, esta tese é concluída no Capítulo 8, resumindo os resultados obtidos e também fornecendo indicações para futuras contribuições sobre a avaliação da qualidade de imagem.

Abstract

The work presented in this thesis intends to provide significant contributions to the research and development on multimedia quality assessment. It is particularly focused on the evaluation of image quality perception and models, and also on the study of consumer sensitivity to induced chromatic changes in digital images.
The main motivation is to evaluate the perception of quality, providing the means to improve the consumer immersion capability of new trends in imaging technology, a more realistic perception of the contents, and consequently to reach new levels of quality of experience. New representation models of visual information are being developed to offer visual fidelity closer to the human perceptual limits. Such representation models, like hyperspectral and multi-exposure technologies, lead to new image formats that recently had acquired large attraction from scientific, professional, and artistic communities.
The first research topic in this thesis reports a study on the induced chromatic changes effects on perceived colors in images and subsequent quality of experience influence on consumers. The results had demonstrated that subjects reveal larger perception to a loss of color naturalness than to the error magnitude itself. Moreover, the eye tracking data revealed a high correlation with the perceived image quality.
The second research topic is devoted to the quality evaluation of the JPEG XT performance on high dynamic range images coding. This study, made in cooperation with other research groups inside Qualinet, benefit from the experience and knowledge acquired on the first research topic. Furthermore, allowed to apply the acquired knowledge on quality in standardization activities. The JPEG XT profiles were verified and the application of quality metrics for quality evaluation was studied. It was concluded that metrics specifically developed for high dynamic range images revealed to be reliable predictors. Moreover, good performance can also be archived with objective metrics developed for low dynamic range domain, when used on perceptual domains. This thesis is concluded with a final comment on the achieved results and possible future research and development directions.

Keywords

Chromatic Information, Quality of Experience, Image Quality Assessment, Quality Perception, Image Quality Metrics, Fixation Maps, Image Representation, Image Coding, High Dynamic Range, JPEG XT.

Contents

List of Figures

List of Tables

Acronyms

CCT Correlated Color Temperature

CI Confidence Interval

CIE Commission Internationale de l'Eclairage

CMY Cyan-Magenta-Yellow

CMYK Cyan-Magenta-Yellow-blacK

CRT Cathode Ray Tube

DFT Discrete Fourier Transform

DMOS Difference Mean Opinion Score

DSCQS Double Stimulus Continuous Quality Scale

DSIS Double Stimulus Impairment Scale

FR Full-Reference

HD High Definition

HDR High Dynamic Range

HFR High Frame Rate

HVS Human Visual System

IQM Image Quality Metric

ITU International Telecommunication Union

IVT Image and Video Technology

JND Just Noticeable Difference

JPEG XT JPEG eXTension

JPEG Joint Photographic Experts Group

LDR Low Dynamic Range

LED Light-Emitting Diode

MOS Mean Opinion Score

NR No-Reference

NTSC National Television System Committee

OR Outlier Ratio

PCC Pearson Correlation Coefficient

PQ Perceptual Quantizer

PU Perceptually Uniform

QoE Quality of Experience

QoS Quality of Service

RFT Relative Fixation Time

RGB Red-Green-Blue

RMSE Root Mean Square Error

RR Reduced-Reference

SROCC Spearman Rank-Order Correlation Coefficient

SSCQE Single Stimulus Continuous Quality Evaluation

TFT Thin-Film Transistor

TMO Tone Mapping Operator

UCS Uniform Color Space

UHD Ultra High Definition

VQEG Video Quality Experts Group

WCG Wide Color Gamut

Chapter 1

Introduction

1.1 Motivation and Scope

In recent years, the growth in multimedia services has resulted in increased demand for quality. Multimedia consumers have daily a set of new services, that stimulates an amazing competition between different service providers and multimedia technologies. Typical technical measures related to the objective performance of a system, like Quality of Service (QoS), become highly unsatisfactory for a real definition of a system or application quality. The QoS lacks sufficient evidence regarding the user's quality perception, since the consumer of a visual system output, the human observer, is not considered. Therefore, a new trend points out to the application of the concept of Quality of Experience (QoE), which tends to evaluate the perceptual experience of a consumer, the produced sensations and feelings, and it is also highly dependent on the context and expectations.

QoE is an evaluation that involves the subjective factors of the end user, and for that reason it is more appropriate for the evaluation of the real systems performance. Although the subjective quality assessment methods are considered to give the most reliable results of quality measuring since subjects are the ultimate consumers of multimedia content, they are costly, complex and a very time-consuming. Thus, objective quality assessment is also of utmost importance.

This thesis intends to provide significant contributions to the research on multimedia quality assessment, focusing on the quality perception and chromatic changes in digital images. It aims to improve user immersion capability of the new trends in imaging technology, providing a more realistic perception of content, and consequently reaching new levels of QoE.

New representation models for the visual information are under development that provides a source of visual fidelity, closer to the human perceptual limits. New trends like hyperspectral and multi-exposure representations lead to new image formats like High Dynamic Range (HDR), that recently had acquired large attraction from scientific, professional, and artistic communities.

The color displaying method influences how the visual information is perceived. Consequently, incorrect representation of color might induce erroneous judgment of the displayed scenes. Thereby, the first research topic in this thesis is to investigate how the perceived effects of induced chromatic variations in color images influence the QoE. To provide an accurate color representation hyperspectral images are used.

Furthermore, HDR systems pursue the representation of the entire dynamic range and color gamut perceived by the Human Visual System (HVS). HDR acquisition systems are already available in a large variety of photographic equipment, and even on some mobile devices. As it is essential, the evaluation of HDR systems in terms of QoE, was the second research topic in this thesis. The need for the quality evaluation of the JPEG eXTension (JPEG XT) part on HDR coding, lead to a collaborative study that is also reported in this thesis. The knowledge provided by the research on the chromatic quality was of utmost importance to the success of the JPEG XT

quality analysis in cooperation with Qualinet[1].

1.2 Thesis Organization and Contributions

This thesis is organized as outlined in the following. This chapter gives a brief introduction to the thesis, including the motivation and scope, and also the thesis organization and contributions. The remainder of this thesis is divided into three parts.

Part I addresses the state of the art. Particularly the Chapter 2 reviews several fundamental concepts and principles of color vision and imaging systems, focusing mainly on aspects that are relevant for the scope of this thesis. It starts with a review of the visual color perception, followed by a description of the color representation systems. Chapter 3 provides an introduction to the basic concepts of image quality assessment. It includes an overview of the main steps for subjective tests design, the models for objective quality assessment, and validation methodologies. These concepts will be widely used across all reported studies in the thesis.

Part II focuses on the perception and modeling of chromatic impairments induced in color images. The influence of chromatic impairments on the perceived image quality is studied in Chapter 4. A method to induce chromatic impairments on color images is described. Then, a subjective assessment experiment held to quantify the subjects' sensitivity to the applied chromatic changes is reported and analyzed. Finally, an objective evaluation of the image quality using an extensive benchmark of existing quality metrics complements this quality study. Moreover, new objective models for quality evaluation in this context are reported.
Hence, the developed research leads to the following, 1) a subjective quality assessment [1], 2) a objective quality assessment with Full-Reference (FR) metrics [2], 3) a seminar presentation [3], and 4) a objective quality assessment with FR *versus* No-Reference (NR) metrics [4]. Furthermore, a generic study on this domain was published in a journal article [5].
In Chapter 5 a study on the quality perception of specific chromatic impairments is described. This chapter extends the work presented in Chapter 4 analyzing specific situations that revealed to be worthy of further analysis. The specific chromatic impairments generation is described initially. Then two subjective assessment experiments, including the subjects' eye movements recording is described. Moreover, a comparative correlation analysis between the subjective scores and the eye tracking data is described. This study comprises the subjective quality assessment [6], and a subjective quality assessment plus an eye gaze behavior experiment [7]. Furthermore, a generic study on this domain, including the Relative Fixation Time (RFT) analysis have been submitted to a new journal article [8].

Part III addresses a collaborative study on JPEG XT quality for HDR coding. In Chapter 6 are reported two studies to analyze the objective quality of JPEG XT performance for HDR coding. Initially is presented a first study that resulted from a joint work performed with two Qualinet research laboratories, namely EPFL (Switzerland) and CTU (Czech Republic). The main objective was to verify the conformance of JPEG XT profiles using a test set of HDR images. A comparative analysis of the results is also described [9]. These results were also used for presentation [13] about Evaluation of Current and Future Image Compression Technologies. The following section is focused on the influence of the JPEG XT encoding on Tone Mapping Operator (TMO) generation.

[1] http://www.qualinet.eu/

As in the previous study, is evaluated the performance of the JPEG XT profiles using a test set of HDR images. Furthermore, an objective quality assessment methodology and a comparative analysis of the results [10] is reported.
Finally, a collaborative work performed with a research group of EPFL (Switzerland), on an extensive benchmarking of objective quality metrics for HDR image quality assessment is reported and analyzed in Chapter 7. A short report of the subjective quality assessment, followed by a description of the used objective quality assessment. Furthermore, the accuracy, monotonicity, and consistency of the objective quality metrics are also reported [11, 12].
The thesis is concluded with a summary of the achieved results in Chapter 8. Moreover drawing directions for future research in this domain are also discussed briefly.

Part I

State of the Art

Chapter 2

Color Vision and Imaging Systems

The way color is displayed influences, broadly, how it is perceived. Hence, color incorrectly represented might induce erroneous judgment of the observed displayed scenes. This may differ from the form, or use, as the subject is viewing the considered scene or image. For instance, if the reddish of the forest, which should be resplendent and vibrant becomes weak, an apathetic and depressing image is transmitted rather than an exciting one.
Jacques Rivière[1] wrote "*The true propose of painting is to represent objects as they really are, that it is to say, differently from the way we see them. It tends always to give us their sensible essence, their presence, this is why the image it forms does not resemble their appearance because the appearance changes from moment to moment*". Therefore, regardless the influence of time or environment might cause, a reliable judgment of a scene is the main goal of the HVS [14].

2.1 Visual Color Perception

The visual color perception of a scenario is due to the interaction between the light produced by a light source, the object surface in the scene and the HVS, as represented in Fig. 2.1. The human eye perceives the color of an object as a set of multiple wavelengths within the visible electromagnetic spectrum (as represented in Fig. 2.2) that can be reflected, absorbed or transmitted by the object surface [15].

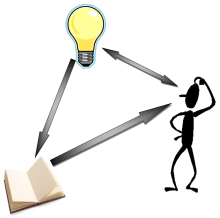


Figure 2.1: Human vision of a reflective object.

For instance, assuming that you are in a picnic on a sunny day, about to catch a red apple, the sunlight shines on the apple and the red wavelengths of the illuminating light are reflected from the apple to the eye. Most of the wavelengths of blue and green are absorbed by the apple surface. Thus, the sensors in the eye structure react to the reflected light that enters the eye by sending information interpreted by the brain as being a red color.

2.1.1 Human Visual System

The visual perception is strongly influenced by the anatomical structure of the human eye. Due to its structure, the human eye can be compared to a photographic camera, an optical microscope or a telescope: a lens system (cornea and crystalline), one variable aperture system (iris

[1] Jacques Rivière, 1886-1925, art critic

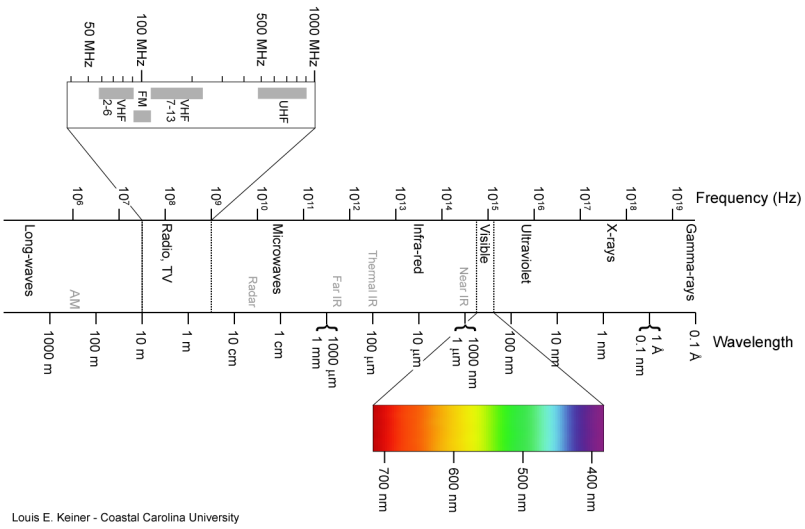


Figure 2.2: Electromagnetic spectrum *(Louis E. Keiner - Coastal Carolina University)*.

and pupil) and a photographic film (retina). In Fig. 2.3, it is represented the schematic cross-section of the human eye, showing the main structural elements.

While the pupil controls the amount of light admitted through the eye, the cornea and crystalline function as a focusing system of a camera. These structures, when focusing on a real scene or object provide jointly an adaptation to the viewing distance. The retina, at the posterior part of the eye, provides the first processing layer; i.e., it retains the formed real images and transmits that information to the brain by means of electrical pulses sent through the optic nerve. These and other structures have a significant impact on the color perception [16, 17].

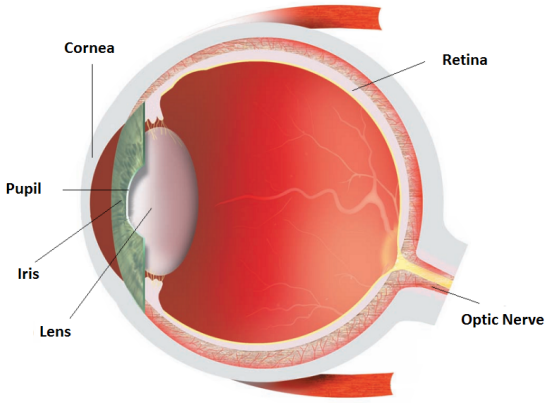


Figure 2.3: Schematic diagram of the human eye with some key structures labeled (Adapted from [18]).

The cornea is a transparent refractive surface, which is located in front of the eye which the light passes through and is a fundamental forming-image element.

The iris is a contractile structure that controls the size of the pupil, which adjust the amount of light entering the eye.

The crystalline lens acts as a camera objective lens, participating in the refractive process of the light, being able to increase the degree of focusing of images (accommodation). This

structure absorbs and scatters energy of shorter wavelength (blue-violet and ultra-violet), and according to crystalline lens age, the level of energy absorption and dispersion is increased [19].

The optic nerve is the structure of the retina, constituted by the axons and the ganglion cells. This cells, have the function of transmitting electric pulses to the visual cortex in the brain, that corresponds to the last level of the retina process [20].

The retina is the inner layer of the eyeball, where can be found the visual system's photosensitive cells or photoreceptors [20]. As can been seen in Fig. 2.4, the retina it's composed of several layer cells (pigmented epithelium, photoreceptors, horizontal, bipolar, amacrine and ganglion), that are responsible for different tasks in image processing. The first cell layer contains two types of photoreceptor sensitive to light being grouped in order to filter different types of wavelength.

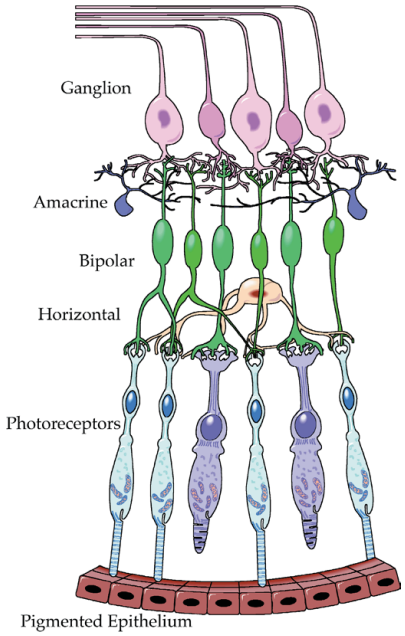


Figure 2.4: Schematic cross section of the retina (Adapted from [19]).

There are two types of photoreceptors in the human retina, the rods and the cones. The rods, about 120 million per retina, are light achromatic (black or white) sensitive cells responsible for night vision (scotopic), capturing only the lightness of the colors and is distributed mostly on the periphery of the retina [21].
The cones, approximately 8 million cells per retina, are color light-sensitive cells, and are responsible for daylight vision (photopic). They are distributed with larger density in a central region of the retina (macular area). The cone cells operate as filters for different wavelengths and are subdivided into three types:

- **S-Short:** sensitive to short wavelengths with a sensitivity peak of *440 nm* (violet color, but often referred erroneously by blue color);
- **M-Middle:** sensitive to medium wavelengths with a sensitivity peak of *540 nm* (referred to the green color, but perceived as yellowish-green);
- **L-Large:** sensitive to large wavelengths with a sensitivity peak of *570 nm* (referred to the red color, but perceived as orange) [22, 23].

The Fig. 2.5 depicts the normalized spectral sensitivity of the human cone cells.
The nervous system gets the color information, comparing the responses of these three types of cones, and through these three values, describes the sensation of color. So, any method of association of these three values (called tristimulus) and a given color, is called a color space.

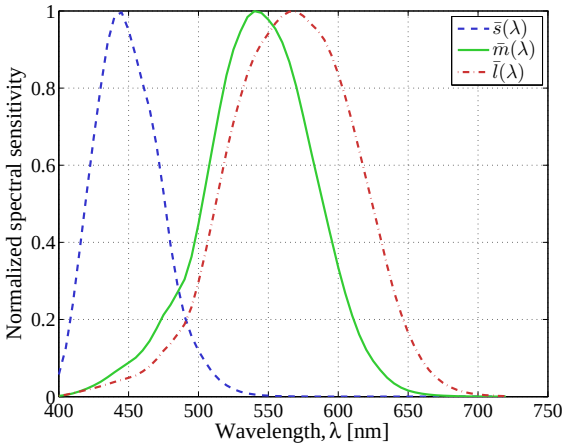


Figure 2.5: Normalized spectral sensitivity curves $\bar{s}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{l}(\lambda)$, of the three different types of cones (*S*, *M*, *L*) responsible for the photopic vision (Adapted from [24]).

2.1.2 Colorimetry

Colorimetry refers to the science and technology used to quantify and describe (with the help of mathematical models) the human color perception. This is a branch of color science, centered on the quantitative specification of a visual color stimulus, physically defined such that [16]: 1) when viewed by an observer with normal color vision and for the same observation conditions, stimuli with the same specification look the same; 2) stimuli that look the same, have the same specification; 3) values comprising the color specification, are continuous functions of the physical parameters defining the spectral distribution of the radiant power of the stimulus.

Trichromatic Theory

In the early XIX century, Thomas Young[2] proposed the trichromatic theory, according to which the human eye has receptors for three primary colors (red, green, blue). The perception of any color would result from the combination of these three primaries in a process that occurs in the brain. The loss of ability to perceive a primary color through the retina, produce the inability to perceive a color that has any of these components [25].
The trichromatic theory proposed by Young was later developed by Hermann von Helmholtz[3], and according to his study, it is the synthesis of the three primary colors that allows the human brain to carry out all the shades of colors of the visible spectrum [26].
Later, Maxwell[4] showed that any color can be produced by the three primary colors and provided an explanation for the inability of some people to perceive colors, confirming the ideas of Young and Helmholtz. This theory is today a basis for color vision understanding, although this theory does not have any regard to certain aspects of vision, namely on the operation of the chromatic

[2]Thomas Young, 1773-1829, polymath
[3]Hermann Ludwig Ferdinand von Helmholtz, 1821-1894, physicist
[4]James Clerk Maxwell, 1831-1879, physicist and mathematician

cells, or on color information processing in the human brain. According to this theory, it can be considered that the existing three types of cones present in the retina are respectively sensitive to red, green and blue, which are designated by the initials R, G, and B. For this reason, they are called primary colors [27].

Grassmann's Laws

The Grassmann's[5] laws of additive color mixture, represent one of the most important foundations of the colorimetry [28, 29, 16].

Grassmann's first law - Three independent variables are necessary and sufficient to psychophysically characterize a color [30]. This law states that the color space is three-dimensional. Thus, any color stimulus can be produced from three fixed primary stimuli, whose radiant power can be adjusted individually by an observer. The only restriction for these primaries choice is the requirement of colorimetrically independence, that is, none of the primaries can be color-matched by mixing the other two. The three most common set of primaries is composed by the red (R), green (G) and blue (B). This law is the support base of the color matching experiments. Generally, as can be seen in Equation 2.1, the color stimulus is represented by C, which have a visual equivalence ($\equiv$) to the sum of $\mathcal{R}$ units of the primary R, $\mathcal{G}$ units of the primary G, and $\mathcal{B}$ units of the primary B. The terms R, G, B denotes the primary stimuli required to match the color. The terms $\mathcal{R}, \mathcal{G}, \mathcal{B}$ indicates the adjustment factors (amount) for each primary.

$$C \equiv \mathcal{R}R + \mathcal{G}G + \mathcal{B}B \quad (2.1)$$

Grassmann's second law - The result of an additive mixture of colored light depends only on the psychophysical characterization and not on the spectral composition of the colors. This means that the human eye can not differentiate components, but only the resulting color [30]. For example, an observer can visualize a yellow color, but, he can not identify the components that gave rise to it.
The term additive mixture means a color stimulus for which the radiant power in any wavelength interval is equal to the sum of the powers in the same interval of the mixture elements, and it is expressed by Equation 2.2 using the previous notation.

$$\begin{aligned} \forall C_1 &\equiv \mathcal{R}_1R + \mathcal{G}_1G + \mathcal{B}_1B \\ \forall C_2 &\equiv \mathcal{R}_2R + \mathcal{G}_2G + \mathcal{B}_2B \\ C_1 + C_2 &\equiv (\mathcal{R}_1 + \mathcal{R}_2)R + (\mathcal{G}_1 + \mathcal{G}_2)G + (\mathcal{B}_1 + \mathcal{B}_2)B \end{aligned} \quad (2.2)$$

Grassmann's third law - If the components of a color stimuli mixture are amplified by a given factor, the psychophysical color result is weighted by the same factor [30]. Thus, if k is a constant, this law is characterized by Equation 2.3.

$$\begin{aligned} C &\equiv \mathcal{R}R + \mathcal{G}G + \mathcal{B}B \\ kC &\equiv k\mathcal{R}R + k\mathcal{G}G + k\mathcal{B}B \end{aligned} \quad (2.3)$$

[5] Hermann Günter Grassmann, 1809-1877, polymath

2.1.3 Perceptual Attributes of Color

The subjective color characterization of an object surface by a normal color vision observer is usually described by three parameters, which show the function of their three-dimensionality: hue, saturation, and brightness. It can be said that the color is a sensation produced in the brain in response to incidence of light on the retina. This sensation is caused by different types of light (composed of a spectrum with different energies at different wavelengths), emitted by a light source or reflected by the observed scene. Only in the eye, and later in the brain, the specific spectrum is translated and that will give the feeling of a certain color, although different spectra composition might be perceived as the same color. This type of effect is called *metamerism* [31]. Despite the difficulty in defining "color" as a generic term, there are properties that can be set. Thus, the terms hue, saturation, and brightness are very important for modeling the appearance of colors [19].

Hue

Hue is the attribute of a visual sensation that is mostly used to describe a color. It refers to the quantity of redness, greenness, blueness, etc., that a color has. For a color physical description, this attribute is related to the dominant wavelength of the incident light into the eye, and is possible to sort the perception of color into two categories: 1) **Achromatic perception:** Devoid of tone, and is associated with the perception of white, black and gray mid-tones; 2) **Chromatic perception:** Has tonality, and is associated with the perception of a predefined color.

Saturation

Saturation defines the attribute of visual perception that determines the proportion of pure color in full color perception, and is related with the proportion of white that is present in a color. This attribute is associated with the behavior of the dominant wavelength maximum value [19]. In common usage, changes in saturation are given the terms pale, weak, strong or vivid.

Lightness or Brightness

The terms lightness or brightness refers to the amount of light emitted or reflected by the color. While lightness is related to the amount of reflected, the brightness is used in the context of light sources, and is associated with the amount of emitted light [19].

2.1.4 Light Sources and Illuminants

One of the three components of Fig.2.1 is the light source as it provides the electromagnetic energy required to initiate the visual responses. The specification of the color properties of these light sources can be done in two ways, by measurements or standardization [19]. It can then be said that a light source is a physical phenomenon of a kind of radiation, such as a candle, a tungsten lamp or the daylight.
On the other hand, an illuminant is a tabulated set of specifications for a potential light source. All light sources can be specified as an illuminant, but not all illuminants may be represented by a light source. The illuminants are specified typically in terms of relative energy tabulated for each wavelength [32]. Therefore, it can be said that it is the radiation with a relative spectral

flux distribution, defined over the range of wavelengths, which influences the perception of color [33, 34].
If the function $S(\lambda)$ represents the spectral power radiance distribution of a light source and the spectral reflectance at a given point of the surface of a scene is expressed by $r_e(\lambda)$, then the distribution of the spectral radiance of the reflected light in such point (color signal) is represented by $R_e(\lambda)$ and it can be expressed by the Equation 2.4,

$$R_e(\lambda) = S(\lambda) r_e(\lambda) \quad (2.4)$$

The mathematical representation of this color signal does not takes into account the geometrical effects of the surfaces. Hence, as an example, the spectral radiance from the surface of an object may depend on the incidence angles of the light and on the observation angles.

CIE Standard Illuminants

The Commission Internationale de l'Eclairage (CIE) established various standard types of spectral power distributions to describe light sources. Among them, the *A*, the *D65* and the *F2* are very relevant. These illuminants are the standard representations of incandescent light, daylight, and fluorescent light sources, respectively [16].

Incandescent light - The typical illuminants A, B and C (described in Fig.2.6) where recommended in 1931 by the CIE, and intended to represent an incandescent light, sunlight and an average daylight, respectively.
The illuminant A aims to represent a typical tungsten lamp, and its relative radiance represents a Planckian radiator[6]. This illuminant has a Correlated Color Temperature (CCT) of *2856K* and it should be used in all applications that are related to the colorimetry, which involves the use of incandescent light [16].
The illuminants B and C were developed to represent the sunlight and the average daylight, with a CCT approximately of *4900K* and *6800K*, respectively [19].

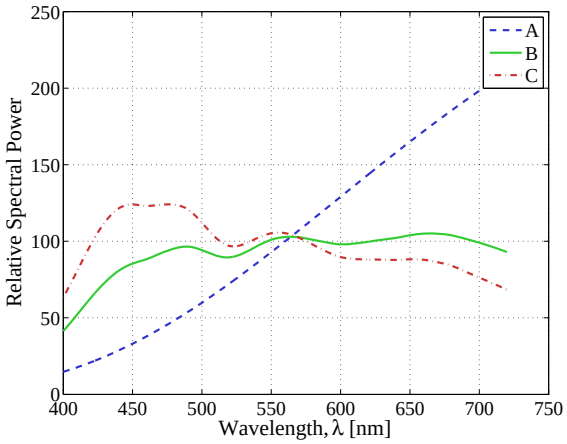


Figure 2.6: Relative spectral power distributions of the typical CIE illuminants *A*, *B* and *C* (Adapted from [16]).

[6]Planckian radiator - blackbody emitter that changes color over a continuous spectrum when heated to high temperatures, as originally determined Max Planck.

Daylight - Posteriorly, in 1963 the CIE defined a series of illuminants (described in Fig.2.7) aiming to simulate the standard daylight, based on the combination of different measurements that vary both with regard to geographic location as compared to the measurement time, weather and climatic conditions. The D series illuminants (D50, D55, D65, D75) were defined from chromaticity equations and spectral power distributions, with a CCT approximately of *5000K*, *5500K*, *6504K*, and *7500K*, respectively [16]. Particularly, the CIE recommended the D65 as the standard illuminant to use in order to represent the average daylight in the visible spectrum.

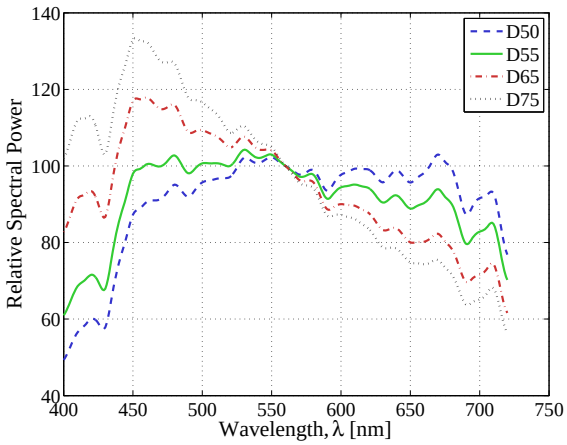


Figure 2.7: Relative spectral power distributions of CIE illuminants *D50*, *D55*, *D65*, e *D75* (Adapted from [16]).

Fluorescent light - The CIE also defined the F series illuminants (12 in total) to represent the spectrum of various types of fluorescent sources. Considering the three most representative (F2, F7 and F11), which are shown in Fig.2.8, can be reported that the F2 illuminant tends to represent a source of white and cold fluorescent light with a CCT of *4230K*, the F7 illuminant represent a source of fluorescent light broadband with a CCT of *6500K* that simulates the D65 illuminant, and the F11 illuminant that represent a triband fluorescent source with a CCT of *4000K* [19, 32].

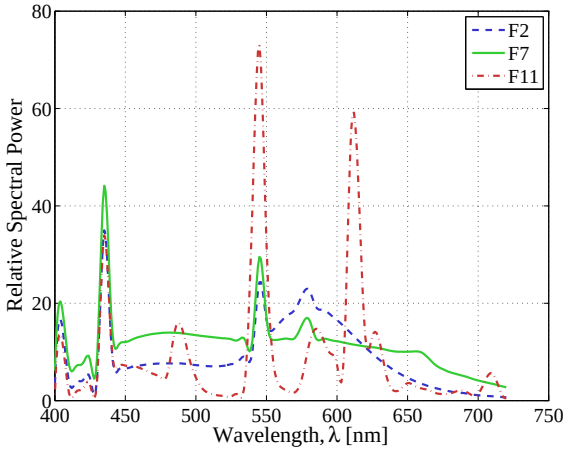


Figure 2.8: Relative spectral power distributions of CIE illuminants *F2*, *F7* e *F11* (Adapted from [16]).

2.2 Color Representation Systems

2.2.1 CIE Standard Colorimetric System

The standard colorimetric system developed in 1931 by CIE, is currently the reference system in chromatic stimuli specification. Deduced from psychophysical experiments, based on experimental results obtained by Wright [35, 36] and Guild [37], this system provides standardized methods to describe a given color stimuli under controlled lighting conditions and observation, based on an average of the HVS responses. Thus, the CIE defines a standard observer, based on two different, but equivalent, sets of color matching functions. The first set is based on the values of the color matching functions for a bipartite field of 2° obtained from three monochromatic primary stimuli of wavelengths $\lambda_r = 700$ nm, $\lambda_g = 546$ nm, and $\lambda_b = 436$ nm, and for a reference equienergetic white [16]. Since the spectral power of any color stimulus corresponds to the result of an additive mixture of monochromatic stimuli, the tristimulus values can be determined integrating over the visible spectrum (Grassmann's second law) the product of the color matching functions, $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$, and the amount of stimuli energy in each wavelength (Grassmann's third law), as represented in Fig. 2.9. This defines the CIE 1931 Standard RGB Colorimetric System.

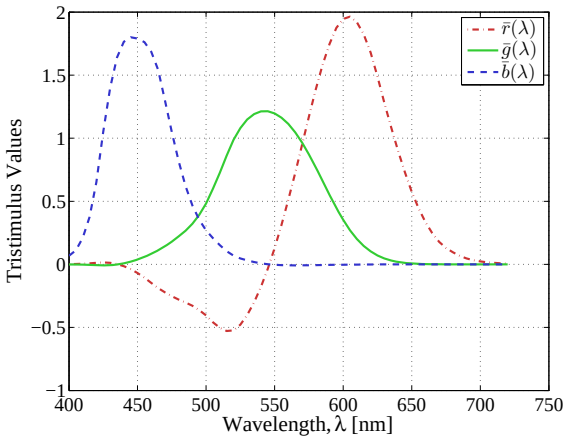


Figure 2.9: CIE 1931 RGB $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ color matching functions.

Thus, a given color stimulus with spectral radiance distribution $R_e(\lambda)$ can be represented as three RGB tristimulus values and computed by the Equations 2.5, 2.6, and 2.7.

$$R = \int_{\lambda} R_e(\lambda)\bar{r}(\lambda)d\lambda \quad \textbf{(2.5)}$$

$$G = \int_{\lambda} R_e(\lambda)\bar{g}(\lambda)d\lambda \quad \textbf{(2.6)}$$

$$B = \int_{\lambda} R_e(\lambda)\bar{b}(\lambda)d\lambda \quad \textbf{(2.7)}$$

A second set of color matching functions can be defined from the RGB primary stimuli, using another set of primaries XYZ, as a new color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, as presented in Fig.2.10. This new standard, referred as CIE 1931 Standard XYZ Colorimetric System, it was recommended by the CIE with the main objective to eliminate the negative values of the

tristimulus at all wavelengths of the CIE RGB color matching functions [16]. Thus, this new set of color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ can be calculated (Equation 2.8), using a linear transformation of the previous set of color matching functions [38].

$$\begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} = \begin{bmatrix} 0.49 & 0.31 & 0.2 \\ 0.17697 & 0.8124 & 0.01063 \\ 0.0 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{bmatrix} \quad (2.8)$$

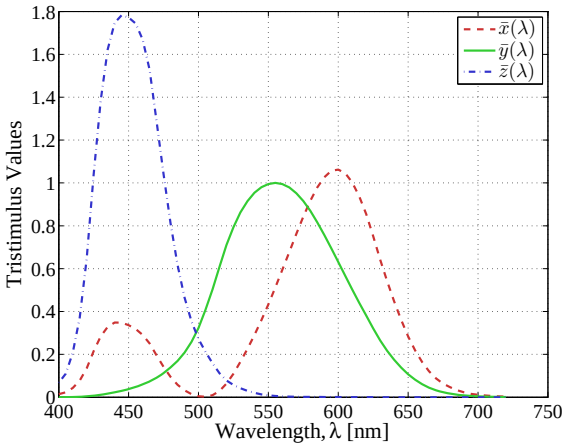


Figure 2.10: CIE 1931 XYZ $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ color matching functions.

Then, a given color stimulus with spectral radiance distribution $R_e(\lambda)$ can be represented as three XYZ tristimulus values and computed by the Equations 2.9, 2.10, and 2.11. Note that, κ (defined by Equation 2.12) is a normalization constant and λ_{Max} and λ_{Min} represent the upper and the lower limits of the wavelength range in the visible region. $S(\lambda)$ represents the spectral power radiance distribution of a selected illuminant.

$$X = \kappa \int_{\lambda_{Min}}^{\lambda_{Max}} R_e(\lambda)\bar{x}(\lambda)d\lambda \quad (2.9)$$

$$Y = \kappa \int_{\lambda_{Min}}^{\lambda_{Max}} R_e(\lambda)\bar{y}(\lambda)d\lambda \quad (2.10)$$

$$Z = \kappa \int_{\lambda_{Min}}^{\lambda_{Max}} R_e(\lambda)\bar{z}(\lambda)d\lambda \quad (2.11)$$

$$\kappa = \frac{100}{\int_{\lambda} S(\lambda)\bar{y}(\lambda)d\lambda} \quad (2.12)$$

Since, a color stimulus is specified by a triplet of values, the projection of the tristimulus values in a unitary plane $(X + Y + Z = 1)$ can represent the color stimulus based on the tristimulus values of only two [39]. The Y value represents the stimuli luminance and can also be denoted by L. The transformation of tristimulus values in a unitary plan is followed by a normalization of

values that suppresses information of the tristimulus "intensity", preserving information about the direction [40].
Thus, one color with XYZ tristimulus values can be specified by its chromaticity coordinates (x, y, z), and defined by the Equations 2.13, 2.14, 2.15. As there are only two dimensions of the chromaticity coordinates, a third coordinate it can always be obtained by the other two. So, z can be also calculated from x and y, using the Equation 2.16.

$$x = \frac{X}{X+Y+Z} \quad (2.13)$$

$$y = \frac{Y}{X+Y+Z} \quad (2.14)$$

$$z = \frac{Z}{X+Y+Z} \quad (2.15)$$

$$z = 1.0 - x - y \quad (2.16)$$

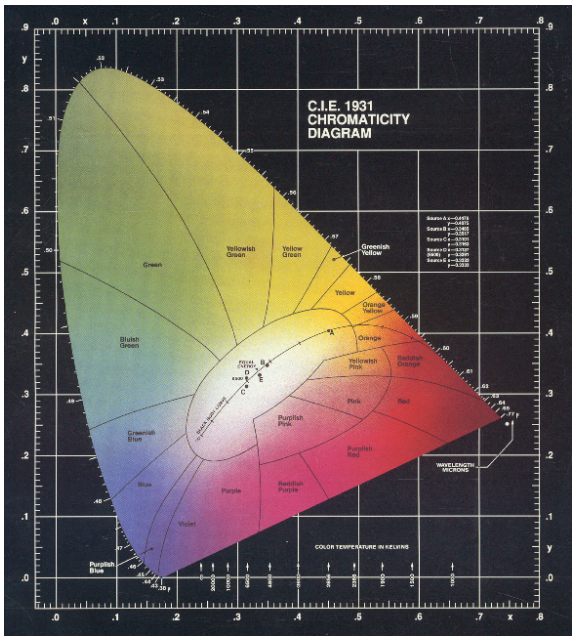


Figure 2.11: CIE 1931 (x, y) chromaticity diagram for the standard observer in the CIE 1931 Standard XYZ Colorimetric System (Photo Research Product Bulletin Nº 66, 100-4/89-7.50)

In the Fig. 2.11 is represented graphically the projection of the CIE 1931 XYZ in the chromaticity diagram CIE 1931 (x, y). However, this chromaticity diagram only shows the proportions of tristimulus values for a given color, since they only represent colors with the same level of luminance. Thus, the chromaticity diagrams are maps related only to color stimuli. For this reason, and for uniformity issues, there are other CIE color models, which describe a color in terms of hue, brightness, and saturation. Among them are included the CIELUV and the CIELAB color spaces.

2.2.2 CIE Uniform Color Spaces

The major limitation of the CIE 1931 XYZ color model, is the lack of uniformity, as the same variations in the values of x, y or Y do not correspond to the same perceived visual difference, designated as Just Noticeable Difference (JND). In non-uniform color spaces, the representation of JND is a line segment, which varies in length according to the area of the corresponding chromaticity diagram [33]. Furthermore, in this chromaticity diagram all colors have the same luminance L. In an attempt to obtain an expression for the chromatic differences wherein the value obtained in relation to a pair of samples was proportional to the observed visual difference, the CIE in 1976 established the Uniform Color Space (UCS), which resulted in the CIE 1976 (u', v') or CIELUV, and later the CIE 1976 (L^*, a^*, b^*) or CIELAB [41]. The main difference between these two color spaces is that straight lines in CIE 1931 (x, y) color space remain straight lines in CIELUV color space, but not in the CIELAB color space. This feature was considered important when colored lights are mixed additively, for example in color television [19].
In Fig.2.12 is represented the chromaticity diagram for the CIELUV color space, and the chromaticity coordinates are obtained by transforming the CIE 1931 XYZ tristimulus values or chromaticity coordinates with the Equations 2.17 and 2.18.

$$u' = \frac{4X}{X + 15Y + 3Z} = \frac{4x}{-2x + 12y + 3} \quad (2.17)$$

$$v' = \frac{9Y}{X + 15Y + 3Z} = \frac{9y}{-2x + 12y + 3} \quad (2.18)$$

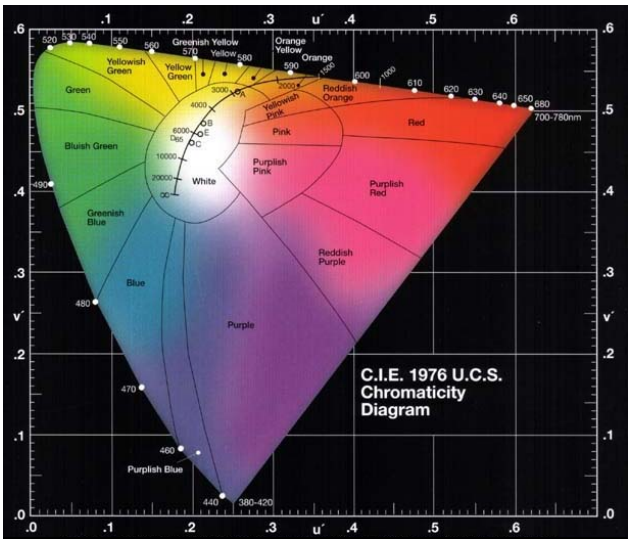


Figure 2.12: CIE 1976 (u', v') chromaticity diagram for the standard observer (Photo Research Product Bulletin Nº 66, 100-4/89-7.50).

CIELAB Color Space

The CIELAB is a numerical color space, in which the perpendicular axes a^* and b^* link around a pair of the hue of colors. The green-red for the a^* axis and the blue-yellow for the b^* axis. With these variables are defined the hue and the saturation. On the other hand, the variable

L^* defines the lightness. By agreement, the typical numerical ranges are $[-60, 60]$ for a^* and b^* and $[0, 100]$ for the L^* component. In the latter range, 0 represents the ideal black and 100 the white reference [32, 19]. Furthermore, the L^*, a^*, and b^* are obtained through the relationships between the CIE 1931 XYZ tristimulus values for a given color and their tristimulus values X_w, Y_w, and Z_w for a reference white, as can be seen in Equations 2.19, 2.20, 2.21, and through the Fig.2.13.

$$L^* = 116\left[f\left(\frac{Y}{Y_w}\right) - \left(\frac{16}{116}\right)\right] \quad (2.19)$$

$$a^* = 500\left[\left(\frac{X}{X_w}\right) - \left(\frac{Y}{Y_w}\right)\right] \quad (2.20)$$

$$b^* = 200\left[\left(\frac{Y}{Y_w}\right) - \left(\frac{Z}{Z_w}\right)\right] \quad (2.21)$$

where,

$$f\left(\frac{Y}{Y_w}\right) = \left(\frac{Y}{Y_w}\right)^{\frac{1}{3}}, \text{ para } \frac{Y}{Y_w} > 0.008856$$

$$f\left(\frac{Y}{Y_w}\right) = 7.787\left(\frac{Y}{Y_w}\right) + \frac{16}{116}, \text{ para } \frac{Y}{Y_w} \leq 0.008856$$

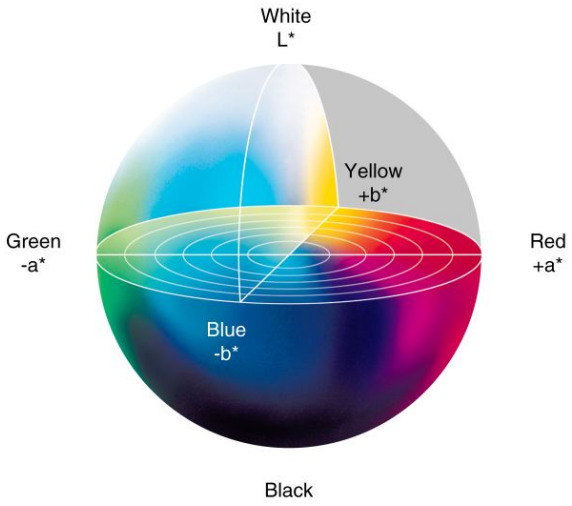


Figure 2.13: CIE LAB color space (Adapted from [42]).

An alternative representation of the CIELAB color space can be achieved by the use of cylindrical coordinates C^* and H_{ab}. The chroma (C^*), is determined by the Equation 2.22,

$$C^*_{ab} = \sqrt{(a^*)^2 + (b^*)^2} \quad (2.22)$$

The hue angle (H_{ab}) is determined by the Equation 2.23,

$$H_{ab} = \arctan\left(\frac{b^*}{a^*}\right) \quad (2.23)$$

CIELAB Color Differences

Based on the CIELAB color space and considering that is sufficiently uniform for applying comparative metrics, it is possible to calculate differences between two stimuli. Since these differences consider the response of the HVS under controlled conditions of illumination and observation, they can provide a good indication as to the quality of display and reproduction of a given stimulus. In 1976, the CIE defined the color difference as the Euclidean distance between two colors (two points) that represent the two stimuli in the CIELAB color space [43], expressed by the Equation 2.24,

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (2.24)$$

where, ΔL^* denotes the lightness differences of the stimuli, Δa^* and Δb^* denotes the chromaticity differences of the stimuli.
In order to improve the uniformity of color differences measures, especially when considering small differences in color between stimuli, several proposals for changes to the ΔE^*_{ab} method were being presented by the scientific community. In 1994, the CIE then sets a new color difference model, namely the ΔE^*_{94} or CIEDE94 [44], and is determined by the Equation 2.25,

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{K_L S_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{K_C S_C}\right)^2 + \left(\frac{\Delta H^*_{ab}}{K_H S_H}\right)^2} \quad (2.25)$$

where, S_L, S_c, and S_H are the lightness, chroma and hue weighting factors, respectively. The K_L, K_c, and K_H are correction factors related to observation environment.
The latest model recommended by the CIE is the CIEDE2000 color difference [45], also represented by ΔE^*_{00} or DE00, includes not only weighting factors for lightness, chroma, and hue, but also factors to handle the relationship between chroma and hue. This CIE color difference was developed because the ΔE^*_{ab} and the ΔE^*_{94} magnitude in different regions of the color space did not appear correlated with the perceptual color differences. The CIEDE2000 is computed using the Equation 2.26, and detailed in Appendix A,

$$\Delta E^*_{00} = \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{K_C S_C}\right)^2 \left(\frac{\Delta H'}{K_H S_H}\right)^2} \quad (2.26)$$

where $\Delta L'$ is the lightness difference, $\Delta H'$ is the hue difference and $\Delta C'$ is the chroma difference. K_L, K_C, and K_H are correction factors related to observation environment. The lightness, chroma and hue weighting factors are represented by S_L, S_C, and S_H respectively. The rotation factor R_T is used to correct the deflection between chroma and hue differences.

2.2.3 Additive and Subtractive Systems

According to the assumptions that gave rise to the trichromatic theory made by Young, Helmholtz and later by Maxwell, three colorimetric independent lights can generate a wide range of colors through the additive mixture.

Thus, to the simultaneous projection process of monochromatic lights (Red, Green, and Blue) of equal intensity on a white target (when illuminated by white light), is named by additive mixture [46]. As can be seen in the Fig. 2.14, when overlapping the lights: 1) **Red and Blue:** generates the magenta color; 2) **Blue and Green:** generates the cyan color; 3) **Red and Green:** generates the yellow color.

The resulting colors from the overlap of the primary colors (Red, Green, and Blue), two by two, are called the secondary colors.

On the other hand, two colors are said to be complementary when the overlap results in the white color, for example, blue and yellow [46]. The spectra filtering process of reflected beams is called subtractive mixture. As an example, when mixing a yellow pigment and a cyan pigment, is obtained the green color, as can be seen in Fig. 2.15. This result is due to the subtraction of the white light to the complementary colors of yellow and cyan, which are respectively blue and red.

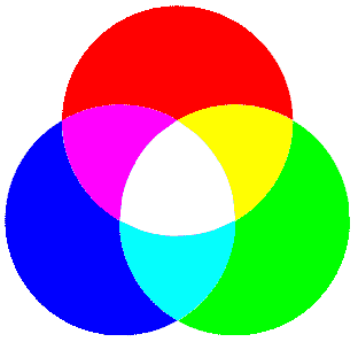


Figure 2.14: Additive mixture of the primary colors

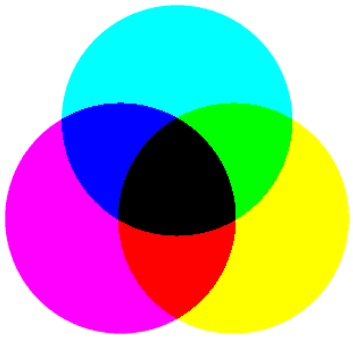


Figure 2.15: Subtractive mixture of the complementary colors

In additive or subtractive mixture, the set of primary and secondary colors are the basis of color reproduction systems. Examples are respectively the Red-Green-Blue (RGB) system (used in displays) and Cyan-Magenta-Yellow (CMY) (used in printers).

The RGB color system is used in most electronic technologies involving the transmission of light, such as Cathode Ray Tube (CRT), Thin-Film Transistor (TFT) or Light-Emitting Diode (LED) display systems. In this system, the color is defined by the numerical value assigned to each primary color (red, green, blue), and a three-dimensional volume is described by the non-negative component values as cartesian coordinates in a euclidean space. If all color values are normalized between $[0, 1]$, this is a unit color cube and is represented in Fig. 2.16.

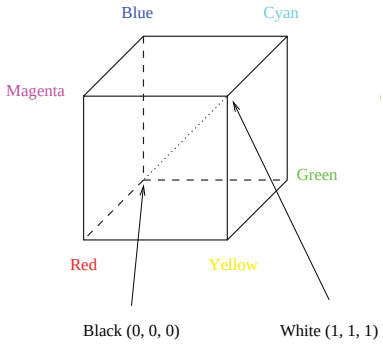


Figure 2.16: RGB color space

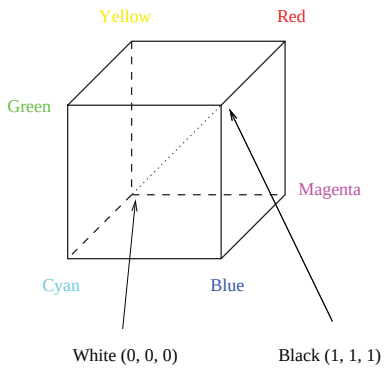


Figure 2.17: CMY color space

Thus, this color system should be considered as a three-dimensional space with the origin at zero $(0,0,0)$ representing the black color, and the white color at the corner farthest from the origin $(1,1,1)$. The red, green and blue maximum values are represented by $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$, respectively. In the opposite corners are presents the cyan, magenta, and yellow with maximum values equal to $(0,1,1)$, $(1,0,1)$, and $(1,1,0)$, respectively. If the values of the three primary colors are equal the color gray get more or less dark depending on the distance of these values to the maximum. This system is device dependent either on the acquisition or the representation, this is due to the individual characteristics of the materials.
As mentioned in 2.2.3 and in contrast with the RGB system, there is a color space based on the subtractive mixture named by CMY, represented in Fig. 2.17. This color system is built from the secondary colors and is used in technologies involving the reflected light, such as printers or photocopiers [46]. Because of this correlation, it is possible to convert the RGB system to the CMY system (see Equation 2.27).

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.27)$$

For technical reasons relating to the physical properties of printing inks, the overlapping of three colors (cyan, magenta, and yellow) with maximum saturation may not be sufficient to absorb the entire spectrum, thereby making difficult the reproduction of the black color. For this reason, the Cyan-Magenta-Yellow-blacK (CMYK) system was implemented using a fourth color, black (represented by K).

2.2.4 Video Color Systems

In order to overcome huge performance requirements to process and handling tasks for video compression, and transmission some color systems are used. Like in HVS, the luminance have a large perceptual range relative to the chrominance [47], which is why it is useful to have color spaces using Y as one of its three parameters. In RGB color space, the three colors are of equal importance and are usually stored with the same resolution. For this reason, video color systems make the separation of the luminance and chrominance components.
Developed by the National Television System Committee (NTSC) in 1953, the YIQ is used in commercial color TV broadcasting. The component Y is used to represent the luminance information (with more representative bits), allowing backwards-compatibility with monochrome TV standards. The color information (with less representative bits) is represented by I (orange-blue) and Q (purple-green) components. The RGB to YIQ conversion is defined in the Equation 2.28.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & -0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.28)$$

The YCbCr color system is similar to the previous system but more suitable for imaging and digital video compression and was developed as part of the ITU-R BT.601 recommendation [48]. The Y component (usually with Y' notation) describes the luminance, the Cb (blue-yellow) and Cr (red-green) represents the chrominance components [49]. The RGB to YCbCr conversion is defined in the Equation 2.29 relative to the R'G'B' (RGB with γ correction), where Y' can take values ranging from 16 to 235; Cb and Cr can take values in the range 16 to 240, with 128

representing the origin of chromatic system.

$$\begin{bmatrix} Y' \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 & 16 \\ -0.148 & -0.291 & 0.439 & 128 \\ 0.439 & -0.368 & -0.071 & 128 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \\ 1 \end{bmatrix} \quad (2.29)$$

The luminance produced by a CRT is related to the signal voltage that it receives in input, and this relation is nonlinear. To represent the nonlinearity in luminance reproduction typically is applied a gamma transformation, given by $L = \nu^{\gamma}$, where L is the luminance on the monitor screen and ν is the input value to the monitor. The parameter (γ) is usually around 2.2 for most color displays. Moreover, it is also necessary to join a nonlinearity present in the HVS, that occurs in inverse form. Thus, for the same objective change of luminance the variation that the HVS perceive it is high in low luminance levels and low in high. To express this luminance nonperceptual linearity is then used a power function with exponent $(1/\gamma)$. Thus, typically the acquisition devices use a gamma correction to modify the input values in accordance with the nonlinear mappings of luminance intensity in human perception.
To compensate the nonlinear characteristics the displays systems corrects the (R, G, B) signals to (R', G', B'), such that $R' = R^{1/\gamma}$, $G' = G^{1/\gamma}$, and $B' = B^{1/\gamma}$. When these signals are applied to the display, the images values are proportional to (R, G, B).

2.2.5 Multi and Hyperspectral Imaging

The trichromatic theory based on the nature of the HVS, as mentioned before, not only forms the foundations of colorimetry but also all the technology involved in color imaging systems. However, in an attempt to simulate the tridimensional color of the HVS, most of the technologies used in acquisition and color reproduction comprising only three channels (red, green, blue) have several limitations [50]. These limitations can be summarized by the difference between the spectral sensitivities and color matching functions (defined by the CIE), and by the metameric nature of these systems. Being the spectral reflectance an intrinsic characteristic of a given material, this knowledge allows minimizing the limitations imposed by the trichromatic systems, since the reproduction of a given color stimulus is based on the spectral power distribution function and not in a metameric match between the original and registered stimulus.
The usage of spectral images aims to fill the gaps of these limitations, due to the ability of such images to get a higher volume of spectral information of the object or scene in study. The use of the information contained in the spectral images allow the possibility of segmenting the digital image signals into two parts; one referring to the illuminant and the other related to the object or scene [51]. Such information also enables to detect and distinguish object colors. This case is characterized by the failure of metameric pictures due to the inability of the spectral separation of the illuminant from the object.
The continuous evolution of equipment associated with the spectral images acquisition and the subsequently increased number of spectral bands susceptible of being recorded, led to the recent emergence of the terms “multispectral” and “hyperspectral”. According to some authors, the “multispectral” term corresponds to the results of spectral scanning images in 4 to 20 different bands [52], and the “hyperspectral” term it is usually assigned when the number of registered bands is high, between 16 and 40 [53].
In the context of this thesis, an hyperspectral image is obtained using a set of pictures of a scene, captured at specified wavelengths intervals. In this study was adopted the usual situation of

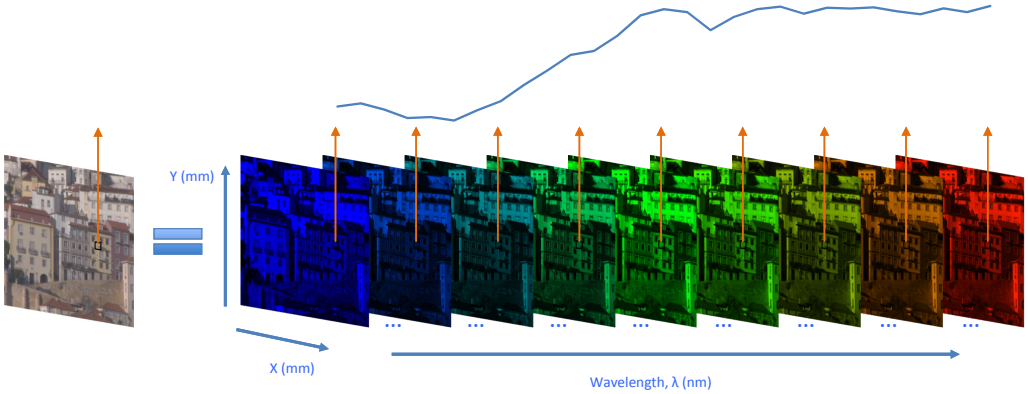


Figure 2.18: Example of a typical hyperspectral image, fused from 33 bands and mapped to the RGB color space for display.

capturing 33 different wavelengths between 400 nm and 720 nm, in the visible electromagnetic spectrum, spaced in steps of 10 nm. Fig. 2.18 summarizes an example of a typical hyperspectral image, fused from 33 bands and mapped to the RGB color space for display.

Each pixel of an hyperspectral image depicts the spectral reflectance information in a specific region of the space. The radiance information is by definition dependent of the spectral reflectance properties of the objects on the scene and also by the spectral power distribution of the illumination source that illuminates the scene. In general, a white reference surface with a well-known spectral reflectance property (perfect diffuser) is placed in the scene. Hence, the spectral power distribution of the illuminant can be estimated by measuring the spectral radiance distribution on the surface of the white reference. Then, the corresponding spectral reflectance of the objects on the scene can be computed independently of the illumination source, for each pixel. This set of pictures represents the reflectance spectra (one spectrum per pixel) of the final hyperspectral image.

2.2.6 Multi-Exposure Imaging

Recently the world of multimedia has been observing a growth in new imaging technology aiming at improving user immersion capability, providing a more realistic perception of content, and consequently, reaching new levels of QoE. This trend has begun with introduction of 3D capable devices in the consumer market, providing depth perception, followed by Ultra High Definition (UHD), focused on higher pixel resolutions beyond High Definition (HD), High Frame Rate (HFR), to provide more fluid motion, and, more recently, HDR, intended to capture and represent a wider range of luminance values.

The HDR imaging systems pursue the acquisition of images where all the brightness information of the visible range of a scene is represented, even in extreme lighting conditions. Hence, it pursues the representation of the entire dynamic range and color gamut perceived by the HVS. HDR images cope with high and low illuminated regions better compared to conventional Low Dynamic Range (LDR) images, which makes HDR images to be more suitable for capturing richer information from the scenes. The problem with the correction of the white balance is more pronounced in a LDR image than with a HDR image. This is a perfect example of the limitation of the overexposed pixels in LDR images.

Many applications can greatly benefit from the adoption of HDR imaging. For example, HDR imaging can be exploited to improve the quality of experience in multimedia applications [54]

and to enhance intelligibility in security applications where lighting conditions cannot be controlled [55].

Nowadays, HDR systems are becoming available for the general public as acquisition systems are present in a large variety of photographic equipment and even on some mobile devices. Typically, merging of multiple LDR images taken at different exposure settings is the method used to generate HDR images [56], as represented in Fig. 2.19.

There are two forms of visualization in HDR images. The first and the best solution is to use a specific HDR display that has the ability to represent a wider luminance range and color gamut. The second solution is to use a TMO that maps the HDR content into the luminance range and color gamut for display with current display technology LDR [59].

The TMO's used to display HDR content on legacy monitors are mostly divided into two groups, global and local operators. The global operators apply the same dynamic range compression function on every pixel in the image, while with local operators, this function varies depending on a neighborhood of the pixels under consideration. An overview and an extensive comparison of existing TMO's was presented by Čadík et al. [60].

In the context of this thesis were selected the following TMO's based on their performance. The *gamma* TMO, first clips a certain percentage of luminance values from both ends of the dynamic range, then linearly scales the remaining values into the interval between 0 and 1, and finally, applies a gamma correction with an exponent of 2.2. This is a very simple global tone mapping operator that implements a naïve auto-exposure strategy. The Reinhard is a global version of the photographic operator [59], which is a popular choice for many applications. The Drago is a global logarithmic TMO [57], which was found to give good compression performance [61]. The Mantiuk is a local TMO for contrast enhancement [62]. The iCAM06 incorporates the spatial

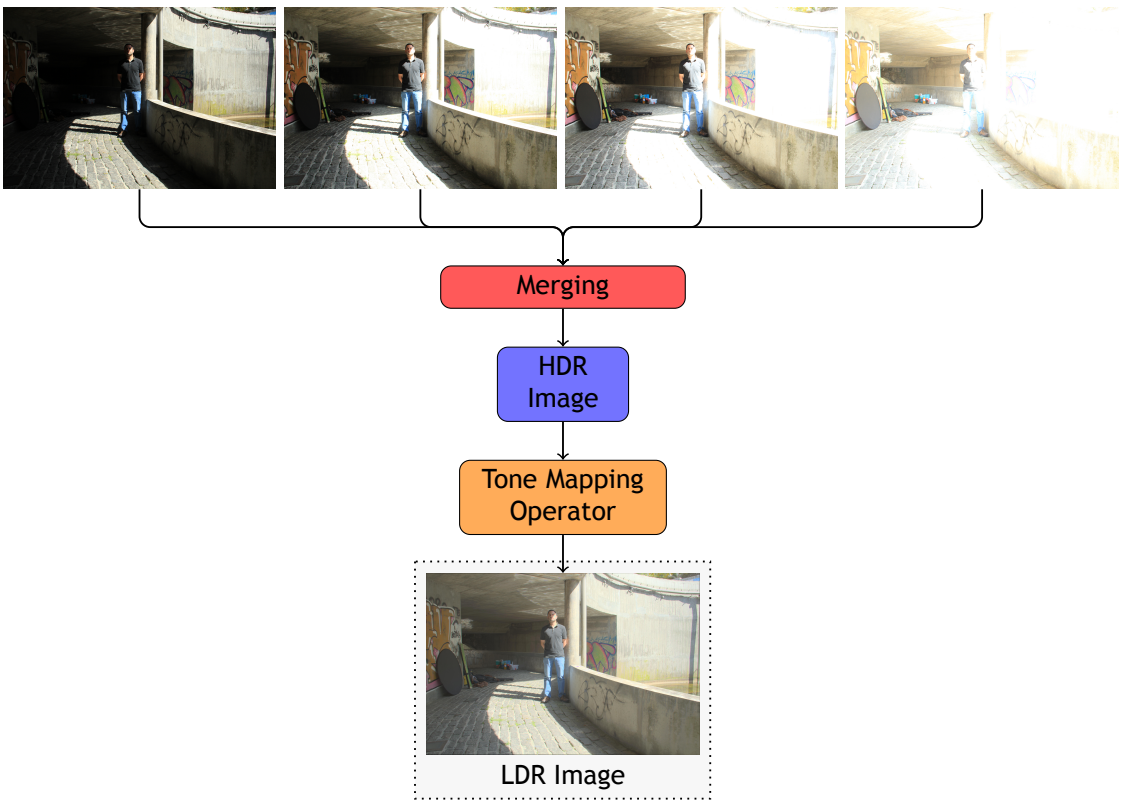


Figure 2.19: Example of a typical process to create an HDR image, fused from four exposures tone mapped with *Drago* TMO [57] (Adapted from [58]).

(a) *gamma* TMO (b) *Reinhard* TMO

(c) *Drago* TMO (d) *Mantiuk* TMO

Figure 2.20: Example of the resultant LDR images represented with *gamma*, *Reinhard*, *Drago* and *Mantiuk* TMO's (Adapted from [58]).

processing models in the HVS for contrast enhancement [63]. The Mai is a TMO optimized for the best encoding performance in a backward-compatible scheme [64]. And, finally, the Reinhard & Devlin, a local TMO based on the simulation of the visual system adaptation [56]. The Fig. 2.20 present an example of the differences of four TMO's.

Chapter 3

Image Quality Assessment

Nowadays, the recent growth in the offer of multimedia services has resulted in increased demand for quality. Multimedia consumers have daily a set of new services that stimulates an amazing competition between different service providers and multimedia technologies. Typical concepts like QoS become highly unsatisfactory for a real definition of a system or application quality [65, 66]. A new trend points out to the application of the concept of QoE, which tends to evaluate the perceptual experience of a consumer, the produced sensations, and feelings, and it is also highly dependent on the context and expectations [65, 67]. The QoS is a technical measure related to the objective performance of a system. The QoS lacks sufficient evidence regarding the user's quality perception, because the consumer of a visual system output, the human observer, is not considered [68]. In opposition, the QoE is an evaluation that involves the subjective factors of the user, and for that reason, it is more appropriate for the evaluation of the real systems performance. This evaluation is carried out to provide the Mean Opinion Score (MOS) or the Difference Mean Opinion Score (DMOS) computed by averaging all the gathered subjective measurements. However, while subjective assessments are usually expensive, very time consuming and the results are strongly dependent on the multimedia content itself, objective metrics are of utmost importance.

3.1 Subjective Quality Assessment

The subjective evaluation of the visual quality according to standardized methods has been the subject of study in recent years. Thus, subjective assessment of audio and visual quality is considered to be the most accurate method reflecting the human perception [69, 70]. To make these quality assessments the International Telecommunication Union (ITU), which is also been adhered by the Video Quality Experts Group (VQEG) [71], has issued several recommendations including the widely used ITU-R Rec. BT.500 [72] for subjective assessment of the quality of television pictures. The ITU-T Rec. P.910 [73] for subjective video quality assessment methods for multimedia applications and the ITU-R Rec. BT.2022 [74] for subjective quality evaluation taking into account the emerging multimedia data, such as Higher Definition TV formats, HDR, or 3DTV.

3.1.1 Standardized Test Methods

Both of the ITU-T Rec. P.910 [73] and ITU-R Rec. BT.2022 [74] recommendations reports similar methods to the ITU-R Rec. BT.500 [72], and the most commonly used are briefly described below.

Double Stimulus Continuous Quality Scale (DSCQS) In this method are shown alternatively to the subjects multiple images or sequences pairs, one of which is the reference. These stimuli are shown to the subjects in a random order and they are not informed which is the reference

one. Additionally, the subjects rate each of the two stimuli separately on a continuous quality scale ranging from “bad” to “excellent”. The DMOS is calculated based on the difference in rating for each pair, ranging in the interval $[0, 100]$. As this method is more sensitive to small differences, it is also preferred when the quality of the reference is quite similar to the image or sequence under test.

Double Stimulus Impairment Scale (DSIS) This is an opposite approach to the DSCQS method. The reference is always shown to the subjects followed by the impaired one. Subjects rate the quality of the impaired version on a discrete five-level scale ranging from “very annoying”, “annoying”, “slightly annoying”, “perceptible but not annoying”, “imperceptible”, and then the MOS is calculated. The DSIS method is intended for quality evaluation of stimuli which cover a wide range of impairments.

Single Stimulus Continuous Quality Evaluation (SSCQE) In this method, the stimuli are shown in separate to the subjects without a source reference. Using a slider, the subjects continuously rate the instantaneously perceived quality on a continuous scale ranging the interval $[0, 100]$, and then the MOS could be calculated. This method is more suitable for a real evaluation of digital services because in real systems users do not have access to the reference, like in home conditions. Thus, this method overcomes most of the conventional double stimulus methodologies issues to assess image quality in digital systems.

3.1.2 General Experiment Requirements

In subjective assessment methods there are two essential elements for conducting the evaluation sessions properly: the test conditions and the environmental viewing conditions. The main test conditions are related to the maximum test duration per session, the maximum number of subjects per session, and the viewing distance.
For the environmental viewing conditions, three factors must be considered: the lighting, the ambiance noise and the quality and calibration of the display. The display and laboratory characteristics used in the subjective tests are related to the height of the picture shown on the screen, the properly viewing distance, the background room illumination, the peak luminance of the display, the luminance of inactive screen, the luminance of background behind the display, the ratio of luminance of inactive screen to peak luminance, and the ratio of luminance of background behind the display to peak of luminance.
The detailed description of these general experiment requirements is reported in the ITU-R Rec. BT.500 [72] for further analysis.

3.1.3 QoE Assessment Using Crowdsourcing

The term crowdsourcing is a popular approach for employing a huge crowd of anonymous *workers* to perform short and simple online *tasks*, and can be considered as a further development of the outsourcing principle [75, 76]. The multiple commercial platforms available of crowdsourcing provide online *workers* with varying cultural and social backgrounds from around the world. Since the payment for a crowdsourcing job is very low, this popular approach can be a powerful and a cost effective tool for performing a great variety of works. Subjective quality assessment or QoE assessment of multimedia is another easily implemented task suitable for the crowdsourcing approach as it reduces the costs effectiveness and also considers more practical

environments [77]. Such flexibility enables a faster achievement compared to traditional forms of assessment, as more potential participants are available, and the assessment is done directly on the subjects devices. Moreover, this approach also allows the possibility to choose, using the ITU-R Rec. BT.500 [72], which common subjective test methodology for quality assessment should be used. Therefore, crowdsourcing is increasingly used in various research areas, including rebuffering in streaming video [78], aesthetics of images [79], the emotional reaction caused by image content [80], and privacy issues in HDR images [55, 58].
Besides the great number of advantages of using this quality assessment tool, there are also some disadvantages or limitations. Technical issues like the bandwidth constraints and support of the *workers* devices to display a required stimulus are the major disadvantages of the use of this quality assessment method. For example, subjective tests with HDTV or UHDTV formats demands large bandwidth requirements and the *workers* may not have a large enough connection to visualize correctly the stimulus with this type of formats. Also, the display disability to represent every kind of image or video formats, as the 3D video or HDR. Furthermore, the major shortcoming of the crowdsourcing-based subjective evaluation is the inability to supervise the *workers* behavior and to control their test conditions. To identify and exclude a *worker* as 'non-trustworthy', there are some approaches like the inclusion of the tests obvious easy-to-answer questions, and measure the mean time spent on each question [81].

3.1.4 Subjective Data Analysis

Once the subjective experiment has performed, following the methodologies described above, a large amount of subjective data is collected. These data, provided by the subjects, must be condensed to be statistically analyzed. In accordance with the guidelines established in Section 2 of Annex 2 following the ITU-R Rec. BT.500 recommendation [72], to perform this analysis the outlier detection, the MOS, as well as the Confidence Interval (CI) must be calculated to summarize the performance of the systems under test.

Outlier Detection

This procedure detects and discards the subjects whose ratings exhibits a strong shift compared to the average behavior. In order to perform the outlier detection, it must be checked if the data is statistically normally distributed or not, by calculating the *kurtosis* coefficient. If the obtained result is between 2 and 4, then the data follows a normal distribution, otherwise, the data is considered to follow a non-normal distribution. The methodology increment the number of times the subject fall outside the mean value plus/minus the associated standard deviation times 2 (if normal) or times $\sqrt{20}$ (if non-normal). If the first ratio is higher than 5% and the relative absolute difference of occurrences is lower than 30% (below and above the specified interval), then the subject is considered as an outlier.

Mean Opinion Score (MOS)

After screening of possible outlier subjects, the MOS is computed to describe the score distribution across the selected subjects for each test as,

$$MOS(j) = \frac{\sum_{i=1}^{N} s_{ij}}{N}, \quad (3.1)$$

where N is the number of subjects and s_{ij} is the score given by subject i to test image j.

Confidence Interval (CI)

The CI for each stimulus over all subjects are also computed with a degree of significance of 95%, which is given by:

$$[MOS(j) - \delta_j, MOS(j) + \delta_j] \quad (3.2)$$

where,

$$\delta_j = 1.96\frac{\sigma_j}{\sqrt{N}}, \quad (3.3)$$

$$\sigma_j = \sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}(s_{ij} - MOS(j))^2}, \quad (3.4)$$

N is the number of subjects and σ_j is the standard deviation of the scores for each test image j. It should also be added that the increase in the standard deviation is an indicative factor of the decrease in quality, even if the MOS remains constant. According to the statistical formalism, the standard deviation is a valid measure when the samples follow a normal distribution, which limits somewhat its use [82].

3.2 Objective Quality Assessment

Subjective quality assessment methods are considered to give the most reliable results of measuring quality since human subjects are the ultimate consumers of multimedia content. However, such subjective evaluations are costly, complex and a very time-consuming process because complex equipment must be used, the environment must be controlled, and the rating is based on the subjects feedback of the output quality. Hence, there has been an increasing interest from the research community to overcome some of the constraints of subjective image and video quality assessment methods [83]. The requirements for an objective perceptual multimedia quality model were also been recommended by the ITU-T Rec. J.148 [84]. Objective Image Quality Metric (IQM)s are developed to give quality predictions, which should be in accordance with subjective assessment results.
Furthermore, these predictions could be correlated with the subjective assessment results for evaluation of the objective IQMs, and consequently develop the ability to provide an enhanced performance for the representation of the QoE, following the ITU-T Rec. P.1401 [85].

3.2.1 Objective Quality Metrics

The objective IQM's can be classified into three categories based on the availability of reference information: FR, Reduced-Reference (RR), and NR [86]. The FR image quality assessment requires a reference image prior to any distortion. For the RR, a reference image is not available, but it is represented by a set of extracted features representative of the image quality. Finally, the perceived quality is computed in the absence of the original image for the NR approaches. This thesis focuses only on a set of state of the art FR and NR metrics, which are studied and the representation of is perceptual quality is analyzed. The FR metrics are generally more reliable

as they provide a comparison to a reference image, while the NR metrics might be very useful to estimate the image quality in the absence of an original image.

Full-Reference Metrics

In this thesis, the studied FR metrics can be divided into different categories: difference measures and statistical-oriented metrics, structural similarity measures, visual information measures, information weighted metrics, HVS inspired metrics, and color difference measures (studied in vision science). The considered FR metrics are described bellow.

Difference measures and statistical-oriented metrics These metrics are based on pixel values differences and provide measures between the reference and the distorted image. The following metrics of this category were considered: the Mean Squared Error (MSE), the Peak Signal-to-Noise Ratio (PSNR), the Signal to Noise Ratio (SNR) [87], and the Gradient Magnitude Similarity Deviation (GMSD) [88].

Structural similarity measures These metrics model the quality based on pixel statistics to model the luminance (using the mean), the contrast (variance), and the structure (cross correlation) [89]. The metrics considered in this category are the following: the Universal Quality Index (UQI) [90], the Mean Structural Similarity Index (MSSIM) [86] and the downsampling version (MSSIMD), the Multiscale SSIM index (MS-SSIM) [91], the Measure - Singular Value Decomposition (M-SVD) [92], the Quality Index on Local Variance (QILV) [93], and the Internal Generative Mechanism (IGM) [94].

Visual information measures These metrics aim at measuring the image information by modeling the psycho-visual features of the HVS or by measuring the information fidelity. Then, the models are applied to the reference and distorted image, resulting in a measure of the difference between them. The following metrics on this category were considered: the Visual Signal to Noise Ratio (VSNR) [95], the Image Fidelity Criterion (IFC) [96], the Visual Information Fidelity (VIF) [97], the Visual Information Fidelity pixel-based (VIFp) [97], the Feature SIMilarity Index (FSIM) [98] and the extension (FSIM$_c$), and the RIESZ Feature SIMilarity Index (RFSIM) [99].

Information weighted metrics The metrics in this category are based on the modeling of the relative local importance of the image information. As not all regions of the image have the same importance in the perception of distortion, the image differences computed by any metrics have allocated local weights resulting in a more perceptual measure of quality. The following metrics were computed: the Information Content Weighting MSE (IW-MSE) [100], the Information Content Weighting PSNR (IW-PSNR) [100], and the Information Content Weighted SSIM (IW-SSIM) [100].

HVS inspired metrics These metrics try to model empirically the human perception of images from real scenes. The following metrics were considered: the Just Noticeable Distortion (JND) [101], the Weighted Signal-to-Noise Ratio (WSNR) [102, 103], the Divisive Normalization (DN) [89], the PSNR - Human Visual System (PSNR-HVS) [104], the PSNR - Human Visual System Masking (PSNR-HVS-M) [105], the Most Apparent Distortion (MAD) [106], the Additive impairment and Detail loss Measure (ADM) [107], and metrics developed for HDR images, like the High

Dynamic Range Visual Detection Predictor (HDR-VDP2) [108, 109], and the High Dynamic Range Video Quality Metric (HDR-VQM) [110].

Color difference measures The color difference metrics were developed because the CIE1976 color difference magnitude [32] in different regions of the color space did not appear correlated with perceived colors. These metrics were designed to compensate the non-linearities of the HVS present on the CIE1976 model. The following metrics were computed: the Mean CIE1976 (MDEab) [32], the Mean CIE94 (MDE94) [44], the Mean CMC (MDECMC) [111], the Mean CIEDE2000 (MDE00) [45], the SCIELAB [112], and the Spatial Hue Angle MEtric (SHAMEI and SHAMEII) [113].

No-Reference Metrics

The studied NR assessments were based on the analysis of a set of well-known sharpness measures. The following NR metrics were considered: the Variance (VAR) [114], the Laplacian (LAP) [115], the Gradient (GRAD) [115], the Frequency Threshold Metric (FTM) [116], the HP Metric (HPM) [117], the Marziliano blurring metric (Marziliano) [118], the Kurtosis based metric (KurtosisZhang) [119], the Kurtosis of wavelet coefficients (KurtosisWavelet) [120], the Auto correlation (AutoCorr) [115], the Riemannian Tensor based metric (RTBM) [121], the Just Noticeable Blur (JNB) [122], the Cumulative Probability Blur Detection (CPBD) [123] and the Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [124].

Other state of the art NR metrics, like BLind Image Integrity Notator using DCT Statistics - II (BLIINDS-II) [125], Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [126] or Naturalness Image Quality Evaluator (NIQE) [127] are not reported in this thesis, as they depend of machine learning models for the quality estimation. Hence, if used in this thesis context they will need appropriate training data, which should be defined with subjective results for this kind of induced distortion. Nevertheless the DIIVINE metric that is also into this state of the art group, is reported as an example.

3.2.2 Evaluation of Objective Models

To evaluate the performance of the IQM's described above, the metrics are compared with the subjective assessment results and the ability to provide a representation of the QoE. The statistical relation between the MOS and the IQM's will be reported bellow.

The subjective assessment results should be normalized into MOS_n on the range $[0, 1]$, shown in Equation 3.5,

$$\text{MOS}_n(j) = \frac{\text{MOS}(j) - \text{MOS}_{min}}{\text{MOS}_{max} - \text{MOS}_{min}} \quad (3.5)$$

where j is the j^{th} tested stimulus, MOS_{max} and MOS_{min} represent respectively the maximum and minimum MOS value obtained for all the stimuli. Hence, the different content and distortion levels relations are preserved and also a continuous relationship between the original MOS and the impairment. In this range, "one" represents no impairment and "zero" represents maximum impairment.

Then, the non-linear regression, are fitted to the IQM's and the mapped MOS_n values and restricted to be monotonic over its range. The data $[\text{MOS}_p, \text{MOS}_n]$ are fitted using Equation 3.6 for

the regression,

$$\text{MOS}_p(j) = b1 + \frac{b2}{1 + e^{(-b3 \times (\text{MR}(j) - b4))}} \quad (3.6)$$

where MOS_p represents the MOS prediction, j is the j^{th} tested stimulus, MR is the corresponding metric result, and $b1$, $b2$, $b3$ and $b4$ denote the regression parameters.
To evaluate the metrics performance four measures were chosen as suggested in [71]:

1. The Pearson Correlation Coefficient (PCC) r_P measures the model prediction accuracy [128],

$$r_P = \frac{\sum_{j=1}^{N}(S_j - \bar{S})(O_j - \bar{O})}{\sqrt{\sum_{j=1}^{N}(S_j - \bar{S})^2}\sqrt{\sum_{j=1}^{N}(O_j - \bar{O})^2}} \quad (3.7)$$

   In this case S_j denotes the subjective score MOS_n, and O_j the objective MOS_p. $\bar{S}$ and $\bar{O}$ are the means of the respective data sets, and N represents the total number of image samples considered in the analysis.

2. The Spearman Rank-Order Correlation Coefficient (SROCC) r_S evaluates the model prediction monotonicity [129],

$$r_S = 1 - \left(\frac{6 \sum_{j=1}^{N} d_j^2}{N(N^2 - 1)} \right) \quad (3.8)$$

   where d is the difference between the ranked MOS_n and the ranked MOS_p, and N represents the total number of image samples considered in the analysis. In practice is computed as the Pearson correlation, but instead of the variables are used the ranked variable.

3. The Outlier Ratio (OR) r_O is a measure of the model prediction consistency. A data pair (S_j, O_j) may be declared as an outlier when the absolute difference between S_j and O_j is greater than twice as much the MOS_n standard deviation σ_{S_j},

$$|S_j - O_j| > 2\sigma_{S_j} \quad (3.9)$$

   Therefore the r_O can be calculated by,

$$r_O = R_O / N \quad (3.10)$$

   where R_O denotes the total number of outliers, and N represents the total number of image samples considered in the analysis.

4. The Root Mean Square Error (RMSE) also measures the quality prediction error that is maximized when the RMSE was minimized,

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}(S_j - O_j)^2}{N}} \quad (3.11)$$

   where N represents the total number of image samples considered in the analysis.

All of these four measures will be used in this thesis to investigate the performance of the proposed objective metrics.

Part II

Perception and Modeling of Chromatic Impairments

Chapter 4

Image Quality under Chromatic Impairments

The influence of chromatic impairments on the perceived image quality is studied in this chapter. Under the $D65$ standard illuminant a set of hyperspectral images were represented into the CIELAB color space and the corresponding chromatic coordinates were subdivided into clusters with the K-Means algorithm. Each color cluster was shifted by a predefined chromatic impairment ΔE^*_{ab} with a random direction in a^*b^* chromatic coordinates only. Applying impairments of 3, 6, 9, 12 and 15 units of ΔE^*_{ab} to the a^*b^* coordinates of five hyperspectral images a set of modified images was generated. Those images were shown to subjects that were asked to rank their quality based on their naturalness. The MOS of the subjective evaluations was computed to quantify the sensitivity to the chromatic changes. This study is also complemented with an objective evaluation of the quality using several state of the art metrics and using the CIEDE2000 color difference among others.

4.1 Introduction

The subjective evaluation of visual quality, according to standardized methods, has been subject of study in recent years. Subjective assessment of audio and visual quality is considered to be the most accurate method for quality evaluation reflecting the human perception [69, 70]. To make these quality assessments ITU, also joined by VQEG [71], has issued several recommendations, as presented in Section 3.1. Furthermore, ITU has also released requirements for an objective perceptual multimedia quality models, as presented in Section 3.2.

Research on this domain had led to a considerable progress on the development of objective models based on the HVS for both audio and video [70]. Perceptual evaluation of video quality, standardized in ITU-T J.247 [130], defines the chrominance indicator for spatial distortion analysis, together with the luminance indicator (based on luminance difference of edge images between the reference and distorted videos), and two temporal variability indicators for omitted and introduced components.

More recently, the CIE has proposed some compulsory and also some recommended test images [131]. Previous studies of color gamut have used computer generated images of colored test charts, which include highly saturated colors and more realistic pictorial images, such as photographic prints [132, 133]. However, the representation of real scenes in these images is limited and may not adequately represent the spectral variety or sampling density of natural colors. The availability of hyperspectral images of real scenes with high chromatic resolution has made possible the extension of the gamut-mapping analysis to real stimuli. The color gamut of real scenes is known to be constrained to the more central region of the chromaticity diagrams [134, 135, 136, 137, 138, 139, 140], which enables the induction of chromatic impairments that might affect the image appearance and its quality. In this context, any chromatic distortion may affect the image appearance, especially when images of real scenes are acquired by hyperspectral imaging systems to be used as stimuli.

In this study, the perceived effects of induced chromatic changes in color images are investi-

gated, and it also extends the previous works on color discernibility introduced in [141, 138, 142]. The study is based on a subjective evaluation of the chromatic impairments, induced in a set of five images. It is impossible to model the large number of possible chromatic impairments created by the typical displaying systems, caused by gamut limitations, graphic cards, and processing software. However, the chromatic changes method introduced in this study will provide a simulation of generic chromatic impairments that can be comparable to those produced by the real displaying systems. The testing system was calibrated to represent the colors of the original real scene images and the corresponding colors of the modified images with the chromatic changes. Moreover, a study using state of the art objective metrics to represent the sensitivity to chromatic impairments will also be reported.

Five hyperspectral images were selected to cover different types of content, namely rural and urban scenes. As the selected images are going to be changed by different chromatic impairments, becomes difficult to use a larger dataset and produce a reliable subjective assessment. After converting them to the CIE 1976 $(L^*a^*b^*)$ color space a change on their initial chromatic components a^*b^* was applied. For that, colors were clustered into similar colors with the application of the K-Means algorithm [143]. Then, a giving chromatic impairment is applied to each of this color cluster in the a^*b^* chromatic components. After that, images are displayed in a calibrated color displaying system, as reported in Section 4.2. Pixels with colors out of the display gamut are clipped to the closest color on the display gamut boundary.

Typically, the MOS is the most used measure to quantify QoE. In this study, the MOS was measured on a scale from 1 (low quality/artificial) to 10 (high quality/naturalness). In this context, naturalness means realistic or not artificial images. Testing should be conducted providing the sensation of a continuous scale that is discretized for the final measure. Although this measure is very reductive of the QoE concept itself, it is still the most conceptual and widely used.

When performing subjective tests using a set of subjects is important to notice that the MOS is going to be influenced by the content itself. This is an important remark that needs always to be considered in any resulting analysis. Using MOS, the sensitivity to the chromatic impairments is studied and analyzed in Section 4.3. The perceptual testing was performed using a group of 102 subjects of different social and educational backgrounds, different age, and satisfying gender equity.

This study also analyses the representation provided by the state of the art objective metrics for the quality representation of chromatic impairments. A good objective metric should take the psychophysical process of the human vision and perception system into account. The main characteristics of the HVS include modeling of contrast and orientation sensitivity, frequency selection, spatial and temporal pattern masking and color perception [144]. In the present study, the most common FR and NR objective metrics were compared with the subjective results in Section 4.4. Moreover, a study on the use of the perceptual CIEDE2000 [45] color difference for objective quality evaluation is also introduced. Their performance was evaluated using three prediction attributes: accuracy, monotonicity, and consistency. In particular, the metrics that reveal the best representation of the subjective results will be analyzed in further detail.

Hence, the following summarizes the main contributions: 1) A method to introduce chromatic impairments that do not create spatial artifacts produced by color discontinuities; 2) A subjective study of the sensitivity to chromatic impairments; 3) An extensive benchmark of existing quality metrics performance when used in the presence of chromatic impairments. These metrics usually only become reliable when the average value of the three color channels is used; 4) A method to use the CIEDE2000 color difference for quality evaluation in the presence of chromatic impairments.

4.2 Induced Chromatic Impairments Method

The hyperspectral images selected for this study (represented in Fig. 4.1) were manipulated on their spectral reflectance data to generate images with different chromatic impairments. These images were converted into spectral radiances using a $D65$ illuminant in order to obtain the corresponding conversion to the CIELAB color space that is device independent, perceptually uniform and based on the HVS [32, 38, 145, 16]. This color space was selected for our experiments as it reveals to be the most appropriate, mainly due to its uniform characteristic.

(a) *Image 1* (b) *Image 2* (c) *Image 3* (d) *Image 4* (e) *Image 5*

Figure 4.1: Original test images (urban and rural scenes).

In a previous study introduced by Aldaba *et al* [141] the CIELAB color space was segmented into cubes. Each cube suffers the application of a ΔE^*_{ab} error and was concluded the existence of the error perception for ΔE^*_{ab} larger than 5 units. However, the error method produced spatial artifacts, that might have a stronger influence on the subject's assessment than the chromatic impairments. So, it is impossible to quantify if the subjective quality evaluation obtained depends exclusively on the chromatic impairment, which was the aim of this study.
To avoid the effect of those spatial artifacts, the colors of the images were subdivided into clusters applying the K-Means algorithm [143] instead of cubes. To identify the value of K that would lead to better results, several numbers of visual inspections were conducted, with K ranging from 2 to 12 clusters. The final results allowed to conclude that $K = 4$ formed groups of similar colors, reducing the influence of the spatial artifacts to unperceptive levels.
Any color pixel i of an image, represented by $(L^*_i a^*_i b^*_i)$, that belong to a specific cluster were chromatically transformed into a new color pixel $(L^*_{ei} a^*_{ei} b^*_{ei})$ by adding a ΔE^*_{ab} error (Equation 4.1) with a predefined magnitude and random direction,

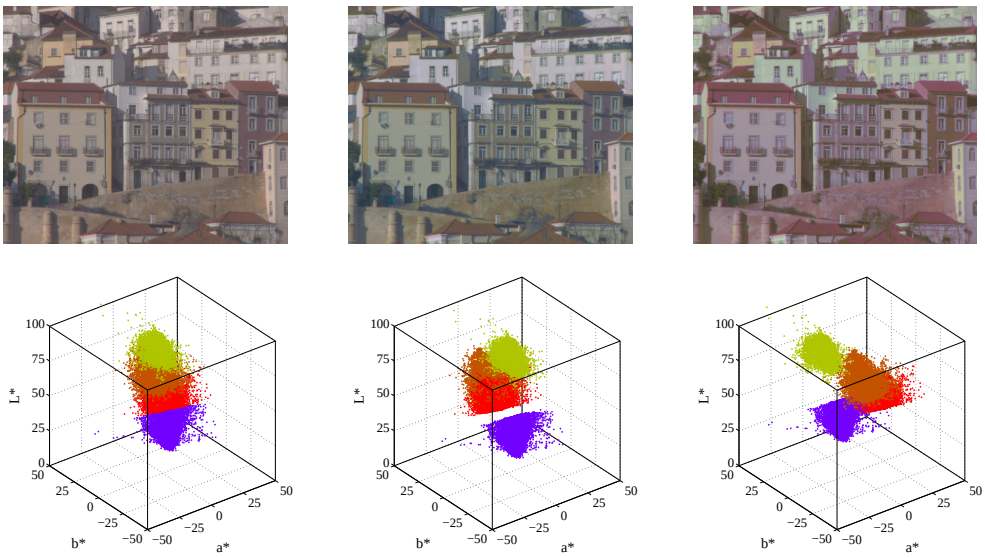
$$\Delta E^*_{ab} = \sqrt{(L^*_i - L^*_{ei})^2 + (a^*_i - a^*_{ei})^2 + (b^*_i - b^*_{ei})^2} \quad (4.1)$$

Considering that the HVS is more sensitive to luminance changes and due to display color gamut limitations [146], was decided to keep the lightness L^* of the resulting image always remained unchanged ($L^*_{ei} = L^*_i$) in Equation 4.1 for all the image pixels. The new color coordinates were computed accordingly to the Equation 4.2,

$$\begin{aligned} L^*_{ei} &= L^*_i + (\Delta E^*_{ab} \cos(\theta)) \\ a^*_{ei} &= a^*_i + (\Delta E^*_{ab} \cos(\alpha) \sin(\theta)) \\ b^*_{ei} &= b^*_i + (\Delta E^*_{ab} \sin(\alpha) \sin(\theta)) \end{aligned} \quad (4.2)$$

where α is the angle direction of the applied chromatic impairment in the range 0^o to 360^o, and the angle $\theta = 90^o$ was kept constant to maintain the lightness L^* unchanged.
This procedure was applied to all color clusters using a set of fixed chromatic impairments. A

different random direction α was applied to each cluster, moving all pixels in a cluster for a predefined ΔE^*_{ab} in the same directions. Hence, it was guaranteed that groups of similar colors were changed in the same direction (see examples shown in Fig. 4.2).



(a) $\Delta E^*_{ab} = 0$ (b) $\Delta E^*_{ab} = 9$ (c) $\Delta E^*_{ab} = 15$

Figure 4.2: Image 3 and corresponding clusters represented in CIELAB color space, when a specified chromatic impairment ΔE^*_{ab} was applied to each chromatic cluster in different directions.

Also, the evaluation provided will depend exclusively on the chromatic impairments, which is the aim of the study, because only images with unperceived spatial artifacts were selected for the test. The magnitudes of ΔE^*_{ab} ranged from 3 to 15 units in steps of 3 units. A set of 5 different images was generated for each predefined magnitude error to cover a larger number of different directions in the CIELAB color space. It is also important to recall that the CIELAB color space is considered to be perceptually uniform [32, 145] and all the chromatic changes were numerically equal.
At the end, 150 images were obtained, including the five samples of each original image ($\Delta E^*_{ab} = 0$).

4.3 Subjective Quality Assessment

4.3.1 Laboratory Apparatus and Calibration

The experiments for subjective quality evaluation were conducted at the Image and Video Technology (IVT) laboratory of the Optics Center at "Universidade da Beira Interior" (UBI), which is compliant with the recommendations for subjective evaluation of visual data issued by ITU-R BT.500 [72]. In this study, the color representation of images was made using a 20-inch color display (Sony, GDM-F520), with a screen resolution of 800×600 pixels, and a refresh rate of 100Hz. The brightness and the contrast were set to the factory defaults settings, with a white point of 6500K CCT. This display device was driven by a visual stimulus generator card (ViSaGe, Cambridge Research Systems), installed in a PC, enabling a 8 to 14 bits resolution in each of R, G and B channels. The voltage-luminance function (gamma function) of the display device was determined by interpolating the measured luminances (L_R, L_G, L_B), expressed in $\mathrm{cd/m^2}$, for each of the three channels using input voltage levels, spaced over sixteen equally steps within the range

$[0, 1]$ volt, in full screen. The luminance measurements were performed using a photometer (OptiCAL Photometer, Cambridge Research Systems), and the corresponding measurements for the chromaticity components (x, y, z) of the RGB primaries, $[(x_r, y_r, z_r), (x_g, y_g, z_g), (x_b, y_b, z_b)]$, were performed using a spectroradiometer (SpectraColorimeter, PR-650 SpectraScan, Photo Research Inc.) [147]. Through the Equation 4.3 the chromatic coordinates distribution of the display device in the CIE 1931 XYZ color space can be calculated in order to compute its color gamut assuming constancy on the chromaticity of the RGB primaries for all range of luminances (0 to L_{max}) [148].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x_r}{y_r} & \frac{x_g}{y_g} & \frac{x_b}{y_b} \\ 1 & 1 & 1 \\ \frac{z_r}{y_r} & \frac{z_g}{y_g} & \frac{z_b}{y_b} \end{bmatrix} \begin{bmatrix} L_R \\ L_G \\ L_B \end{bmatrix} \quad (4.3)$$

The color gamut of the display, calculated in the CIE 1931 XYZ, was then transformed in CIE 1931 xyL, and in the CIELAB color space.

4.3.2 Dataset

The hyperspectral test images used in this study were obtained from two sources: 3 real scenes (rural and urban) from the University of Manchester hyperspectral image database [138, 149], and 2 real urban scenes from the "Universidade da Beira Interior" and the "Universidade de Coimbra" hyperspectral image database. The 3 hyperspectral images from the University of Manchester were collected by D. H. Foster [138] during the summers of 2002 and 2003, almost always under a clear sky in the Minho region of Portugal. The two hyperspectral images from "Universidade da Beira Interior" and "Universidade de Coimbra" were collected during the summer of 2011 under a clear sky in the Costa de Prata region of Portugal, in particular in the city of Coimbra. The radiance spectrum (color signal) at each pixel for all the selected scenes was estimated from calibration measurements based on a neutral reference object located in the scene whose radiance spectrum was known for a specific illuminant (in this particular case a $D65$ illuminant was chosen). Fig. 4.1 represents pictures of these 5 images.
Nevertheless, due to the color gamut of the used display device some colors pixels could not be represented (see an example in Fig. 4.3). So, these colors were displayed by clipping them to the closest color of the display gamut surface in the CIELAB color space, as mentioned and studied by N. Katoh [132]. The number of clipped pixels was, in general, less than 2%.
All the generated test images were displayed on the described displaying setup with a spatial resolution of 600×500 pixels. The area around the displayed images was filled with the average

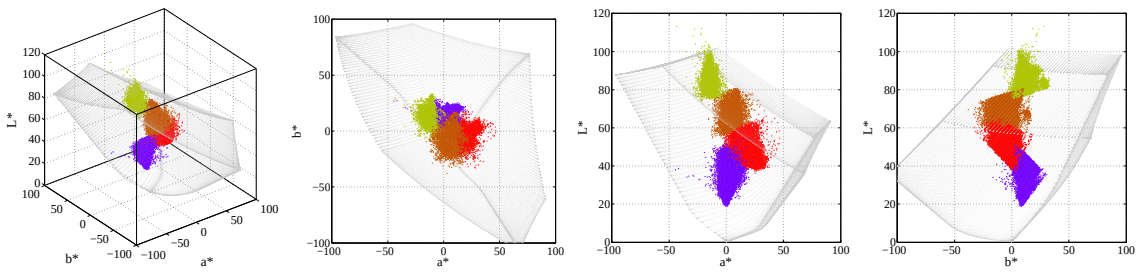


Figure 4.3: Clusters of Image 3 and the display gamut volume, in 3D and different projections, represented in CIELAB color space before clipping, when a chromatic impairment $\Delta E^*_{ab} = 15$ was applied in different directions.

color (gray) of all the pixels that comprise the test images with a mean luminance of $15\ \mathrm{cd/m^2}$. The complete dataset[1] for the tests is composed of the 5 original RGB images plus the images that resulted from using 5 different error magnitudes in 5 different random directions (125 RGB images with generated chromatic impairments).

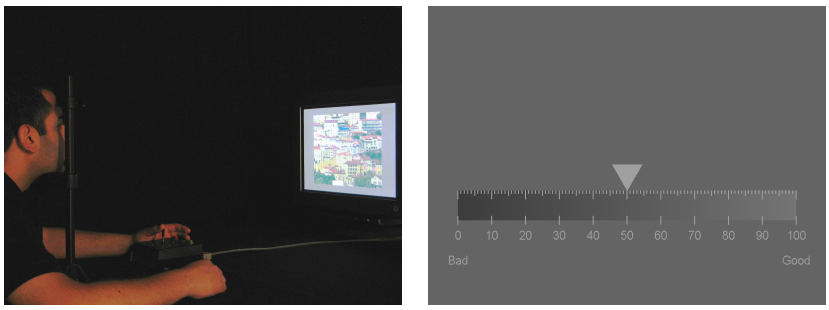
4.3.3 Experimental Procedure

The SSCQE standard test methodology [72] was chosen for the experiments. This methodology is more suitable for a real evaluation of the chromatic sensitivity because in real systems users do not have access to a reference image. Moreover, humans can compare colors to a certain degree, but it is also known that they might not be sensitive to such chromatic change. Prior to the beginning of the test, an experiment protocol (presented in Appendix B), was shown to the subjects. The test was composed of two parts: a first one called "training period" (using a different image set), where the subjects were allowed to be familiarized with the procedure, followed by the "test session". Only the answers given by the subjects during the "test session" were considered for the results of this study. Each subject visualized a randomly selected group of 50 images, taken from all the images that comprise the complete test dataset, in a random sequence. The resulting test had duration time of approximately 20 minutes. Longer test sessions are not advisable because the effects of the subjects tiredness would influence the final results. During the tests, the subject visualizes an image during a 10 second period, followed by the respective evaluation. For the image evaluation, subjects used a joystick that allowed the selection of a grade between 0 (low quality/artificial) and 100 (high quality/naturalness) in a simulated continuous scale. Fig. 4.4 shows an example of a test session and the used voting screen.
The images were observed by the subjects seated in line with the display center, inside a dark room with dim illumination. According to the display size (40×30 cm), the image viewing distance was set to 235 cm in order to guarantee that the image of the stimulus falls in the parafovea region in the retina [150], which corresponds to an angular viewing of 9.69°.

4.3.4 Panel of Subjects

The panel of volunteers comprised 102 subjects, all within the range of 17-65 years old, with an average age of 25 ± 9 years old. This panel contained 54 female and 48 male subjects, keeping the gender balance. All of them were naïve as to the aim of the experiment and all had normal

[1] Available at http://ubicid.di.ubi.pt



(a) Experimental Setup (b) MOS Scale

Figure 4.4: Example of a test session (a) and the used voting scale (b).

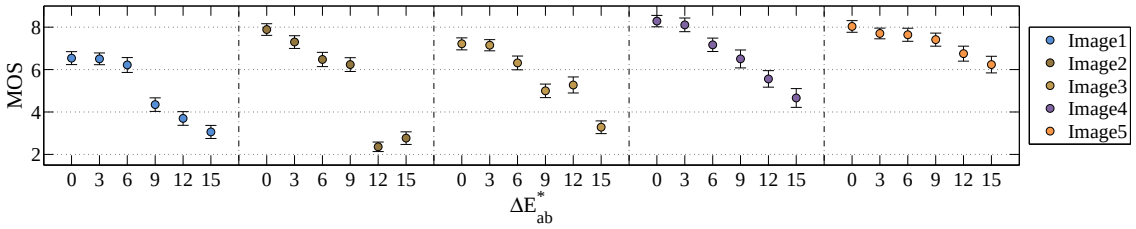


Figure 4.5: MOS values and the 95% CI as function of ΔE^*_{ab}.

color vision, which was assessed by the Cambridge Color Test [151], also performed. Informed consent was obtained from all participants and the research was conducted according to the Declaration of Helsinki [152].

4.3.5 Results Analysis

According to Section 3.1.4, the processing of the subjective data taking into account the outlier detection, the MOS scores and the respective CI were properly calculated. The subjects were screened following the guidelines described for the SSCQE method in Section 2.3 of Annex 2 of the ITU-R Rec. BT.500 [72]. This procedure allows the detection and discards subjects with ratings exhibiting a strong shift compared to the average behavior. However, in this study, no subject was detected as an outlier. Following the same methodology, the MOS scores and the CI were also computed for each test image.

The quality scores obtained for each chromatic impairment ΔE^*_{ab} and the corresponding CI for the 5 test images are presented in Fig. 4.5. As expected, the MOS value decrease with the increase of the chromatic impairment ΔE^*_{ab} for all images. The MOS quality scores stay relatively high for $\Delta E^*_{ab} \leq 6$ units. A degradation on quality scores was observed for $\Delta E^*_{ab} \geq 9$ units. For $\Delta E^*_{ab} = 12$ and 15 units the quality scores are normally poor and depict an erratic behavior.

Furthermore, it was also observed that the quality scores depend not only on the ΔE^*_{ab} error but also on the image content [1]. Namely, Image 5 reveals interesting results since the subjects were very tolerant to the chromatic impairments, although the differences are very visible with a direct comparison between original and impaired images. The quality scores stay stable and high even for large chromatic impairments produced with $\Delta E^*_{ab} = 12$ and 15 units. Moreover, Image 2 and Image 3 exhibits an unexpected behavior when the chromatic impairment increases from $\Delta E^*_{ab} = 12$ to 15 and from $\Delta E^*_{ab} = 9$ to 12, respectively. This behavior is due to the random nature of the chromatic impairment direction, that leads to different types of chromatic changes and hence, creates different perceptual impact, as studied in [6]. Moreover, and considering that Shapiro-Wilk test [153] rejects a Gaussian distribution, a Kruskal-Wallis one way analysis followed by a multiple comparison test [154] reveal that these two cases are statistically similar with $p = 0.14$ and $p = 0.39$, respectively.

4.4 Objective Quality Assessment

A set of FR and NR metrics (mentioned in Section 3.2.1) were studied, and their representation of the perceptual quality was analyzed. Moreover, they were compared with the MOS obtained in the subjective quality assessment experiment. The FR metrics were generally more reliable

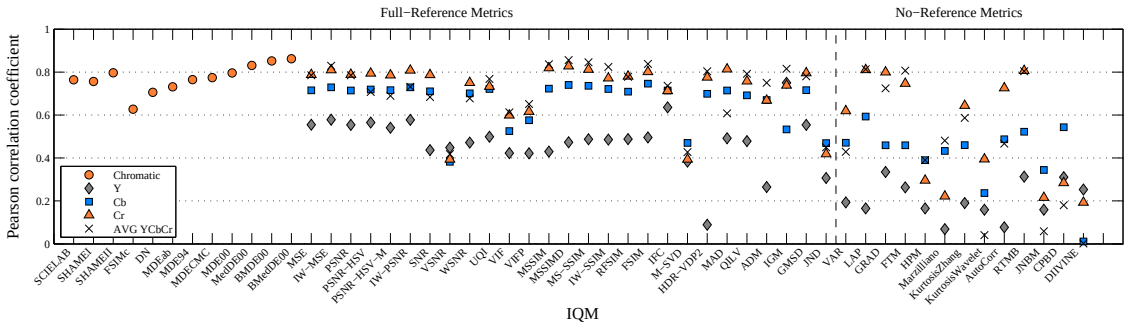


Figure 4.6: Absolute values of PCC for the studied IQM's.

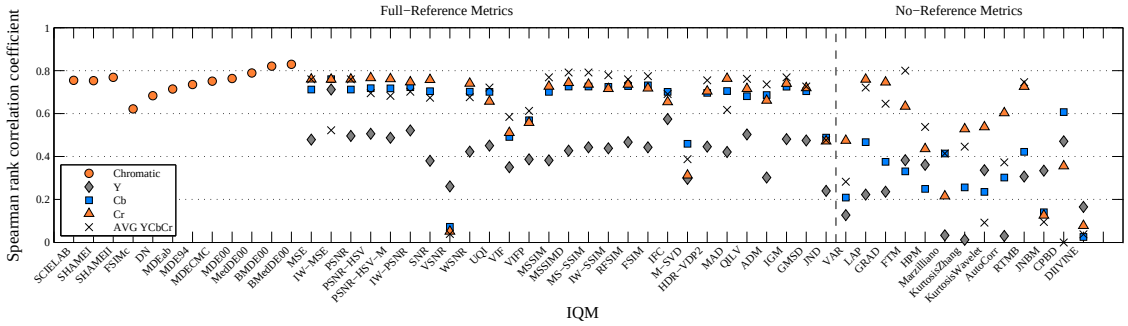


Figure 4.7: Absolute values of SROCC for the studied IQM's.

as they provide a comparison to the reference image, while the NR metrics might be very useful to estimate the image quality in the absence of the original image.

These FR metrics are suitable to be used in this context if applied to the chromatic components as they suffer a distortion caused by the generated impairments. It is possible to argue if FR metrics are suitable to represent a SSCQE subjective study, as they have at most a hidden reference. However, studies like [155, 156, 157] reveal that there is no statistical difference between single and double stimulus studies. Hence, FR metrics can be used for both types of subjective data, as it is a common situation in several studies.

There are a large number of FR metrics, although it was tried to provide an exhaustive study of a significant number of them in this thesis. Other metrics have been tested, but because of the lack of representation and the relative minor popularity, they are not reported.

All these metrics were typically applied to the luminance channel. In this study, all the original color images were converted into the YC_bC_r color space [48]. For the objective quality assessment some metrics were applied to the original RGB images (in the case of objective color difference measures), while others were applied to the components Y, C_b, C_r separately. Furthermore, the average of the three components metrics was also considered.

Moreover, 3 metrics based only on the CIEDE2000 color difference are proposed. They will be explained on Section 4.4.2, and they are: the Median CIEDE2000 (MedDE00), the Balanced Mean of CIEDE2000 (BMDE00), and the Balanced Median of CIEDE2000 (BMedDE00).

In order to evaluate the objective models, all of these IQM's were compared with the available MOS for the studied images, according to the procedure mentioned in Section 3.2.2. Thus, the ability to provide a representation of the QoE was also assessed.

The absolute values of the Pearson and Spearman correlations for all studied IQMs are represented in Fig. 4.6 and 4.7, respectively. In these figures can be observed that in most cases

the objective evaluation that results from averaging the metrics computed for the three color components Y, C_b, C_r resulted in a better evaluation than the individual ones. However, considering the Spearman correlation coefficient, some NR metrics computed for the C_r chromatic component reveal a higher monotonicity than the metrics average. Moreover, in opposition to the typical quality estimation, the metrics computed using the Y component do not correlate properly with the subjective quality estimation in this case. This was somehow expected because in this experiment the luminance was kept constant, while only the chromatic components were changed.

4.4.1 Selected Objective Metrics

Among the analyzed metrics, 10 were selected for further analysis because they revealed the best MOS correlation on the considered category. These metrics are the SHAMEII, the CIEDE2000, 3 metrics based on CIEDE2000, described in the following section, the MSSIMD, the MS-SSIM, the FSIM, the LAP, and the FTM.

Based on the concepts of the CIELAB color difference [32], the SHAME color difference metric initially transforms the reference and the transformed images to the opponent colorspace [112], followed by the application of a spatial filter. The SHAME I and the SHAME II differ on the selected filter, which are based on the spatial filters defined in [112] or [158], respectively. Then, the metric computes the hue angle histogram and divides it into 4 quartiles, which is highly discriminative on the perceptual domain. The different quartiles are weighted $1/4$, $1/2$, 1 and $9/4$ in ascending order, and the hue difference for each pixel is weighted by the respective quartile weight. The final metric value is given by the weighted mean value.

The CIEDE2000 color difference [45], typically represented by ΔE^*_{00} or DE00, includes not only weighting factors for lightness, chroma, and hue, but also factors to handle the relationship between chroma and hue. This CIE metric was developed because the ΔE^*_{ab} magnitude in different regions of the color space did not appear to be correlated with the perceptual color differences. However, in this case it can be used with reliability because the tested images were represented in the CIELAB color space that allows a precise computation. The detailed description of this color difference metric is explained on Section 2.2.2.

Between the FR metrics, the best correlation with the MOS is provided by the SSIM derived metrics. The MSSIMD is an extension of the SSIM that reveals the higher Pearson correlation between all the studied metrics, closely followed by the MS-SSIM. The SSIM is a perceptual metric based on the content features extraction and abstraction. This quality metric considers that the HVS uses the structural information from a scene [86]. The structure of the objects in the scene can be represented by its attributes, which are independent of both contrast and average luminance. Hence, the changes in the structural information from the reference and distorted images can be perceived as a measure of the image distortion. The MSSIMD index is the mean value of the SSIM when a downsampling is applied to the reference and test images, reducing the smaller dimension in an integer proportion to be as close as possible to 256 pixels wide.

The MS-SSIM index is also based on the SSIM. This algorithm calculates multiple SSIM values at multiple image scales. By running the algorithm at different scales, the quality of the image is evaluated for different viewing distances. The MS-SSIM also puts less emphasis on the luminance component compared to the contrast and the structure components [91]. The MS-SSIM formula

is described by Equation 4.4.

$$\text{MS-SSIM} = [l_M]^{\alpha_M} \prod_{j=1}^{M} [c_j]^{\beta_j} [s_j]^{\gamma_j} \quad (4.4)$$

where M is the finest scale obtained after $M-1$ scaling iterations and l_M, c_j and s_j are the luminance, contrast and structure components at their different scales. α_M, β_j and γ_j are defined according the scale to match the contrast sensitivity function of the HVS.

The FSIM is a perceptual metric that results from SSIM and it is also based on the structural information of the image. However, the FSIM adds the comparison of the low-level feature sets between the reference image and the distorted image [98]. Hence, the FSIM analyses the high Phase Congruency (PC) extracting highly informative features and the Gradient Magnitude (GM), to encode contrast information. This analysis is complementary and reflects different aspects of the HVS in assessing the local quality of the image. The FSIM index is calculated through the Equation 4.5,

$$\text{FSIM} = \frac{\sum_{x \in \Omega} S_L(x) PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \quad (4.5)$$

where Ω is the whole image space domain, S_L is the combined similarity between two signals and PC_m is the maximum value of the phase congruency between the same two signals.

The LAP is a simple metric that uses the mean square value of the Laplacian of the image as a metric [115]. It is selected because is the NR metric that leads the highest Pearson correlation between the NR metrics.

Finally, despite the fact of being a NR metric, the FTM results in the highest SROCC, and also results in high PCC. This method is based on the spectral analysis, and on the initial works from Firestone *et al* [159], which describes a method using spectral analysis for optical microscopy, also applicable to electron microscopy. This spectral analysis was applied to the evaluation of the images focus. When the image is focused, sharpen edges and fine details become more visible since they correspond to high spatial frequencies. Some differences to the method were introduced by Murthy *et al* [116] defining the FTM as the summation of the absolute values of the image Discrete Fourier Transform (DFT) at low frequencies components (u, v), without considering the DC component ($F(0,0)$). After computing the DFT of the whole image, the frequencies belonging to the first quarter are selected. This selection are the frequencies lower or equal to $\frac{N'}{4} - 1$ and $\frac{M'}{4} - 1$ when considering an image with $N' \times M'$ dimension. Thus, the FTM can be calculated using the Equation 4.6,

$$\text{FTM} = -|F(0,0)| + \sum_{u=0}^{\frac{N'}{4}-1} \sum_{v=0}^{\frac{M'}{4}-1} |F(u,v)| \quad (4.6)$$

where $F(u, v)$ is the DFT of the image.

4.4.2 Objective Assessment Based on CIEDE2000

The CIEDE2000 defines a perceptual difference between two colors. However, when applied to images this difference does not explore the interspatial characteristics of the HVS, and the CIEDE2000 mean value reveals a reduced correlation with the subjective results. In this case study, the colors have been changed in the same amount in the CIELAB color space, with dif-

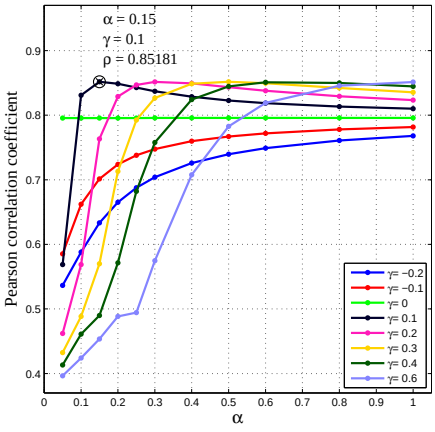


Figure 4.8: Proposed BMDE00 objective metric results for Equation 4.7.

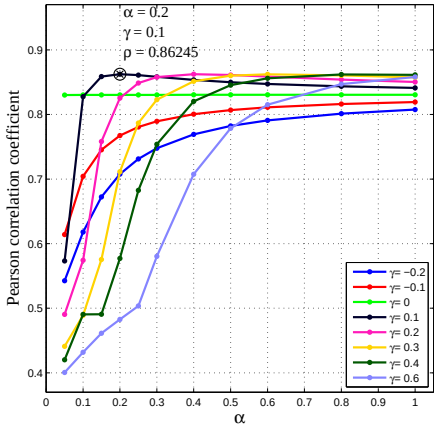


Figure 4.9: Proposed BMedDE00 objective metric results for Equation 4.8.

ferent directions, depending on the color cluster where the pixel is included. Nevertheless, the CIEDE2000 mean value does not lead to the highest correlation with the subjective scores. This is caused by the lack of interspatial relations that are used in most of the metrics analyzed in this work. To reduce these interspatial relations, isolated errors need to have a reduced effect on the final mean CIEDE2000 computation. Three new formulations based on the CIEDE2000 are reported, the median of CIEDE2000, the Balanced Mean CIEDE2000 given by the Equation 4.7 and the Balanced Median CIEDE2000 given by the Equation 4.8,

$$\text{BMDE00} = \frac{\mu^{\alpha}}{(1+\sigma^{\gamma})} \quad (4.7)$$

$$\text{BMedDE00} = \frac{\tilde{\mu}^{\alpha}}{(1+\sigma^{\gamma})} \quad (4.8)$$

In these equations, μ is the CIEDE2000 mean, σ is the CIEDE2000 standard deviation and $\tilde{\mu}$ is the CIEDE2000 median. The fitting parameters α and γ were selected using the best result of the Pearson correlation coefficient ρ shown in Figs. 4.8 and 4.9. Moreover, in the plots of these figures, it is possible to understand that the proposed metrics are stable for small variations of the parameters α and γ. The best Pearson correlation results with $\alpha = 0.15$ and $\gamma = 0.1$ for the Balance Mean CIEDE2000 (Equation 4.7), and with $\alpha = 0.2$ and $\gamma = 0.1$ for the Balance Median CIEDE2000 (Equation 4.8).

4.4.3 Performance Analysis

Fig. 4.10 illustrates the analysis of fitting the MOS *vs* IQM's for all images, using the average of the 3 components Y, C_b, C_r metrics. The obtained regression parameters of logistic fitting for the selected IQM's are shown in Table 4.1.

The Table 4.2 represents the performance measures for the different IQM's studied in this work. The best correlation was obtained for the Balanced Median CIEDE2000. Between the established FR metrics, the MSSIMD has a slightly better PCC for all the tested images. The OR is very small for all IQM's. Finally, the MedDE00 provides the best RMSE. Moreover, some NR metrics also provide a very acceptable representation of the quality. In particular, the LAP has a

Table 4.1: Regression parameters of logistic fitting for the selected IQM's.

IQM	$b1$	$b2$	$b3$	$b4$
MSSIMD	0.072	0.900	8.024	0.822
MS-SSIM	0.052	1.067	5.721	0.852
FSIM	0.009	0.876	6.501	0.665
LAP	0.089	0.764	24.723	0.886
FTM	0.022	0.975	7.592	0.750
SHAMEII	0.150	1.000	4.697	0.883
MDE00	0.056	0.771	6.183	0.465
MedDE00	0.014	0.816	5.983	0.444
BMDE00	−0.024	0.833	14.766	0.145
BMedDE00	−0.087	0.896	10.465	0.179

Table 4.2: Performance measures for the selected IQM's.

IQM	PCC	SROCC	OR	RMSE
MSSIMD	0.856	0.791	0.007	0.295
MS-SSIM	0.846	0.791	0.007	0.285
FSIM	0.838	0.774	0.007	0.242
LAP	0.813	0.721	0.007	0.363
FTM	0.807	0.800	0.0	0.265
SHAMEII	0.797	0.769	0.0	0.266
MDE00	0.796	0.763	0.0	0.178
MedDE00	0.831	0.789	0.0	**0.167**
BMDE00	0.852	0.821	0.0	0.438
BMedDE00	**0.862**	**0.828**	0.0	0.394

close performance to the best FR metrics, while the FTM reveals the highest SROCC between the established metrics, being only over passed by the proposed Balanced metrics based on CIEDE2000.

4.5 Conclusion

In this work, a study on the perception of chromatic changes was carried out using color images with chromatic impairments. The MOS was used to quantify the perceived quality. It is well known that human vision can identify chromatic impairments for $\Delta E^*_{ab} \geq 2.2$ units [141]. However, it was verified that the sensitivity of subjects to chromatic impairments below $\Delta E^*_{ab} = 6$ units was well tolerated, independently of the content. Moreover, chromatic impairments with high amplitude error are in some cases very well tolerated.

The representation of quality obtained by the 39 FR metrics and the 13 NR metrics were analyzed using the subjective results provided by the MOS. From these metrics, the set that provides a better representation of the overall quality, considering the Pearson and Spearman correlations, were selected for further study. Particularly, the metrics SHAMEII, and the metrics based on the color difference CIEDE2000, the average of MSSIMD and MS-SSIM, the FSIM, the LAP, and the FTM were computed over the considered color images providing a MOS representation with acceptable reliability. The best correlation is given by the proposed BMedDE00. Moreover, considering only the established metrics, the Pearson correlation of the FR metric MSSIMD reveals the highest accuracy, while the NR metric FTM reveals the highest monotonicity considering the Spearman correlation. This is particularly interesting due to the NR nature of the FTM metric.

Also, an important conclusion of this work results from the fact that under chromatic impair-

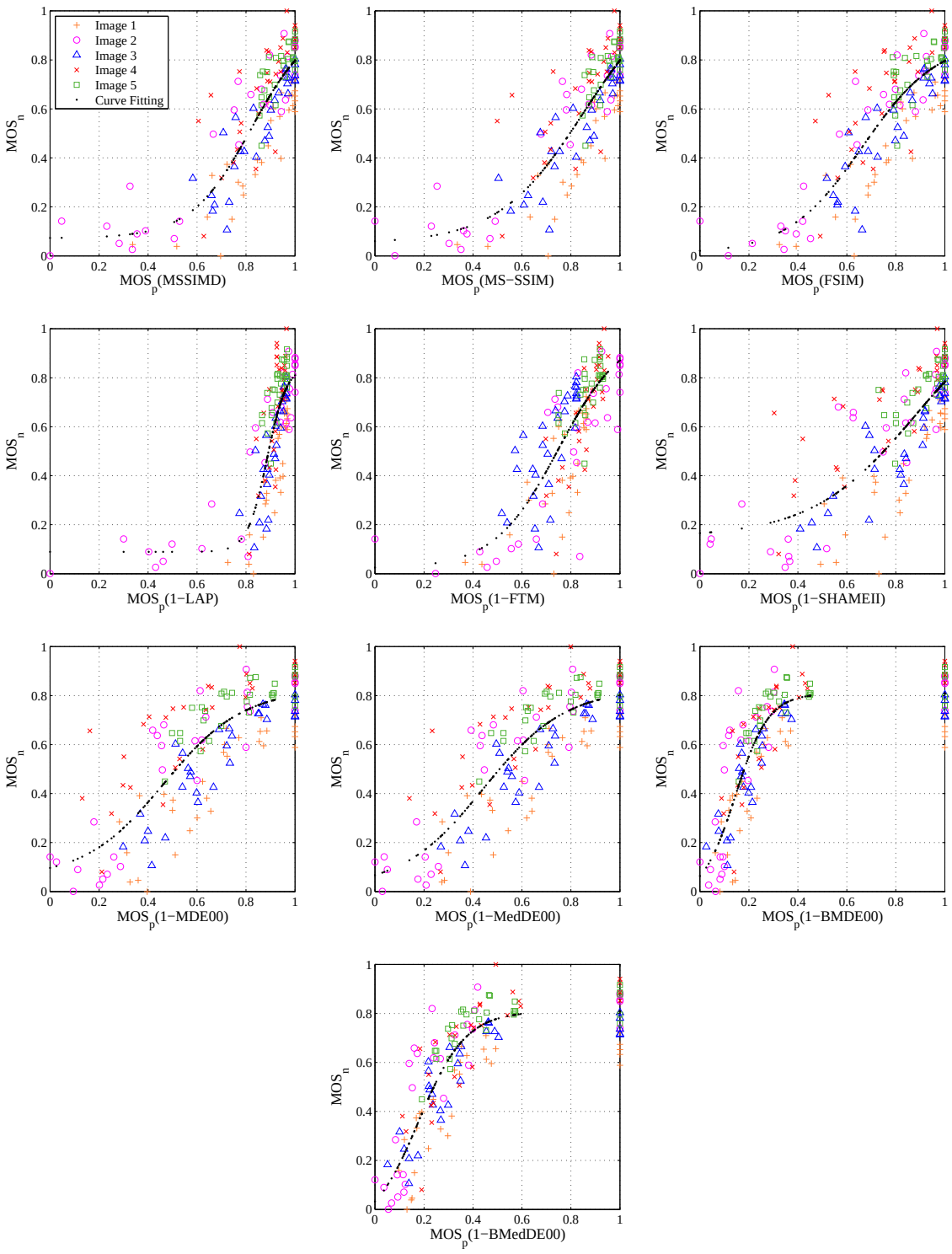


Figure 4.10: Logistic fitting analysis for MOS vs IQM's.

ments, the established metrics like the SSIM based, and also the NR metrics only become more reliable when the average of the metrics computed on the 3 color components is used. This is opposed to the typical metric computation on the luminance channel only.

It can also be concluded that subjects reveal a low sensitivity to chromatic quality. This is, in fact, a very common knowledge, which is particularly considered in video technology, where the chromatic components are usually subsampled. With this research was intended to add a

contribution on a domain where not many reliable studies have been conducted, by modeling, quantifying, and estimating the sensitivity of the HVS to chromatic impairments.

Chapter 5

Quality Perception of Specific Chromatic Impairments

A study of the quality perception caused by the influence of specific chromatic errors is presented in this chapter. Initially, the colors of two hyperspectral images were represented in the CIELAB color space under the $D65$ standard illuminant. Then, the (a^*, b^*) chromatic components of a set of selected pixels clustered by color similarity were changed with a ΔE^*_{ab} chromatic error. This error was applied by moving the color cluster centroid inside the same chromatic quadrant, or in the direction of an adjacent quadrant, resulting in a set of new images with induced chromatic errors. The influence of these chromatic errors on the images naturalness was assessed using a SSCQE, and the MOS was computed. Also, the eye movements were registered using an eye tracking device during the image quality evaluations, and the RFT was computed. Moreover, a comparative correlation analysis between the MOS and the RFT's was studied.

5.1 Introduction

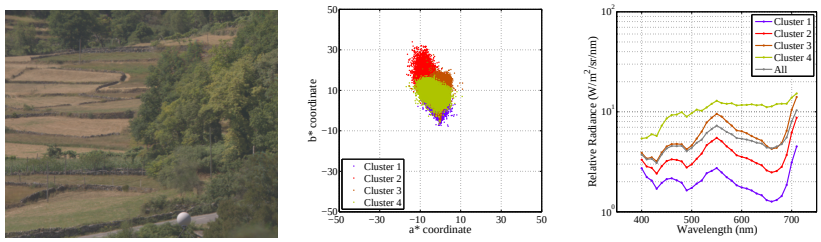
As mentioned in the previous chapter, the QoE tends to evaluate the perceptual quality of experience of a subject, the produced sensations, and feelings, and it is also highly dependent on the context and expectations [65, 67]. Thus, it is well known that the subjective assessment of audio and visual quality is considered to be the most accurate method that considers the human perception [69, 70]. Also, the research on QoE lead many researchers to use eye tracking devices to study the eye gaze changes when subjects visualize images or videos with some kind of chromatic error [160, 161]. Despite the difficulties to model this type of data [162, 163], the eye gaze changes with the impairments might lead to new subjective quality models and might provide a good indicator of the perceived quality. It is also difficult to correlate eye gaze with the measured QoE, typically evaluated with the MOS in a subjective testing [164, 165, 166]. An attempt to create an eye gaze database to help on the definition of new objective quality metrics is reported by Engelke *et al* [167]. A quality model for image inpainting based on the usage of an eye tracking experiment is defined and reported by Venkatesh *et al* [168], and it correlates eye gaze densities with the subjective quality. Another important attempt to correlate quality assessment with eye tracking data was made by Liu *et al* [169], where the called natural scene salience obtained with the eye tracking data was integrated on 4 different metrics resulting in an improved quality estimation. Recently, several works also tried to model quality based on eye gaze for 3D content [170, 171, 172].

In this chapter is described an experiment that was carried out to study the quality perception of specific chromatic errors in images, and aims to extend the work presented in Chapter 4 on user perception to global chromatic errors [1, 5]. This experiment was divided into two parts. The first one corresponds to a subjective score evaluation test [6]. The second part, beyond the subjective score evaluation test, the register of the eye movements (tracking data) of the subjects, was also performed [7]. These subjective score evaluations allow the comparison of specific chromatic changes with global chromatic errors induced in an image. To accomplish

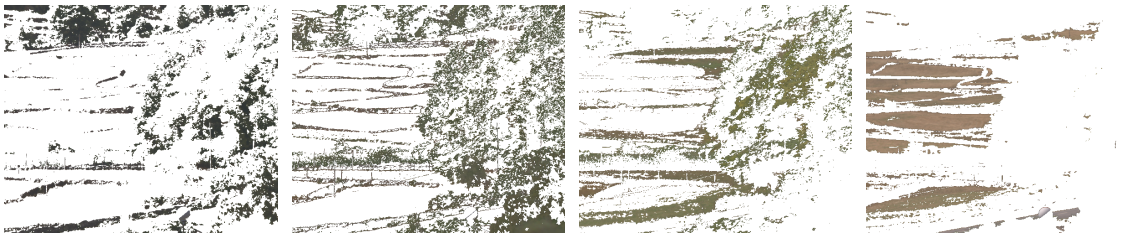
this experiment 2 hyperspectral images (already used in previous work [5]) were selected. The availability of these real image scenes with high a chromatic resolution made possible the extension of gamut mapping analysis to real natural stimuli. After converting these test images into the CIELAB color space under the $D65$ illuminant, the (a^*, b^*) chromatic components of a selected set of pixels clustered by color similarity were changed with a ΔE^*_{ab} chromatic error, according to a predefined procedure detailed in Section 5.2. A dataset of images with chromatic errors was then created. This dataset of images were displayed for visualization in a calibrated CRT display in true color, and a subjective quality assessment experiment was performed using a group of subjects, as detailed in Section 5.3. Using the results of the subjective score evaluation, the MOS was computed and analyzed, as reported in Section 5.4. The eye movements were registered to compute the RFT, and their correlation with the MOS was described and analyzed, as reported in Section 5.5. Conclusions are reported in Section 5.6.
The main contributions in the present work are: 1) The understanding of the perceptual sensitivity to specific chromatic errors using subjective scores and eye tracking data evaluation; 2) A study of the relationship between the MOS and the subjects RFT's.

5.2 Induced Specific Chromatic Impairments Method

The human sensitivity to chromatic errors was tested and quantified using a subjective assessment [5], which can be seen as an extension of the previous works on color discernibility [141, 142]. In Section 4.2 was described a methodology on how to induce chromatic impairments.
In the present work, and in order to study the quality perception of specific chromatic errors, two hyperspectral images were selected from the database used in [5] and specific chromatic errors were induced. These images are a real rural scene representing the country side, and a typical real urban scene, represented in Figs. 5.1 (a) and 5.2 (a), respectively.
These images were divided into 4 clusters and sorted by L^* of each cluster centroid, following the methodology presented in Section 4.2. The colors representation for each cluster in CIELAB color space are represented in Figs. 5.1 (b) and 5.2 (b). These figures show that the color distribution in the rural scene is less saturated (concentrated around the central coordinates (a^*, b^*)), while the color distribution in the urban scene is more disperse.
The average of the relative radiance spectra for each cluster and full image are presented in Figs. 5.1 (c) and 5.2 (c). In these figures can be noticed that clusters 1, 2 and 3 of the rural scene depicts a greenish color signal, while cluster 4 depicts a yellowish one. Moreover, the complete image average radiance spectra signal keeps the greenish appearance. However, the urban scene average radiance spectra depict a grayish color signal in all clusters and also in the complete image.
The corresponding color cluster masks with the corresponding percentage of pixels are presented by the colored pixels in Figs. 5.1 (d) to (g) and 5.2 (d) to (g), respectively.
In this study and in contrast with the methodology proposed in the previous chapter, the chromatic errors were induced in two controlled directions. The first (referred as "in direction"), moved the chromatic components (a^*, b^*) of a cluster by a specific error magnitude with a random direction confined to keep the cluster centroid in the same quadrant. On the other hand, the second induced direction (referred as "out direction"), moved the chromatic components (a^*, b^*) of the cluster centroid towards an adjacent quadrant. The magnitude of the movement corresponds to the induced chromatic error ΔE^*_{ab}.

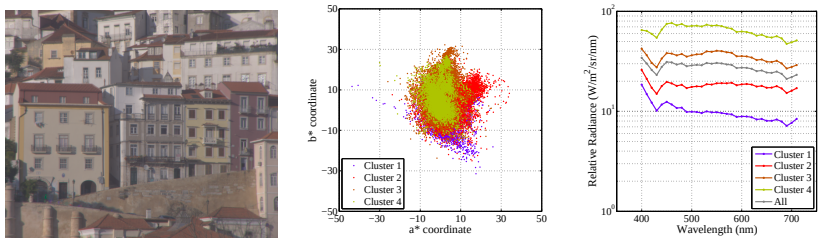


(a) Original (b) CIELAB (c) Relative radiance



(d) Cluster 1 (22.27%) (e) Cluster 2 (33.96%) (f) Cluster 3 (25.04%) (g) Cluster 4 (18.73%)

Figure 5.1: Image 1 (rural scene), representation in CIELAB color space, average of the relative radiance spectra for each cluster and full image, and respective cluster masks (with pixels percentage) (d) to (g).



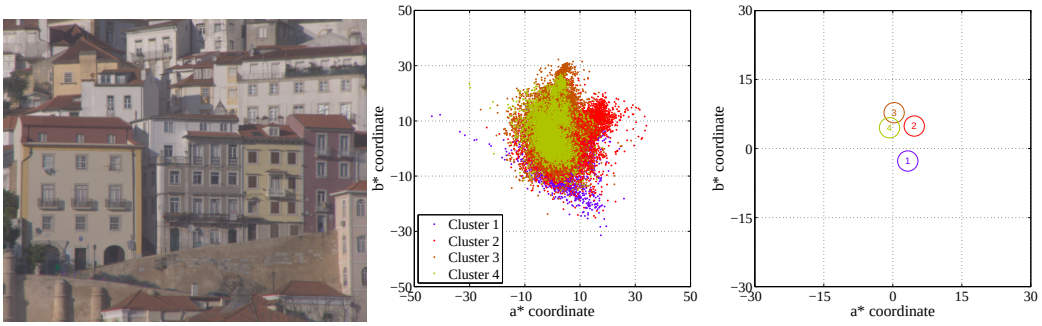
(a) Original (b) CIELAB (c) Relative radiance

(d) Cluster 1 (26.92%) (e) Cluster 2 (30.58%) (f) Cluster 3 (27.33%) (g) Cluster 4 (15.17%)

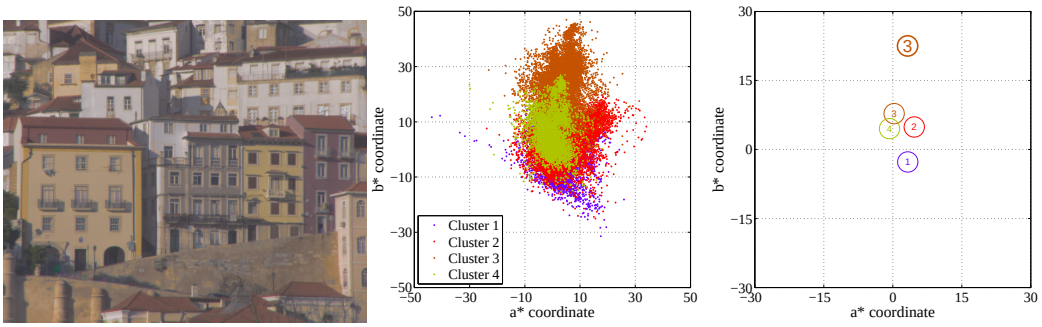
Figure 5.2: Image 2 (urban scene) (a), representation in CIELAB color space (b), the average of the relative radiance spectra for each cluster and full image (c), and respective cluster masks (with pixels percentage) (d) to (g).

When a color cluster and its centroid remains in the same chromatic quadrant the hue remains close from the initial, and the saturation changes accordingly with the chromatic error magnitude. However, a movement to a different quadrant produces a hue change while the saturation varies accordingly with the chromatic error magnitude. The Fig. 5.3 shows an example of such procedure.

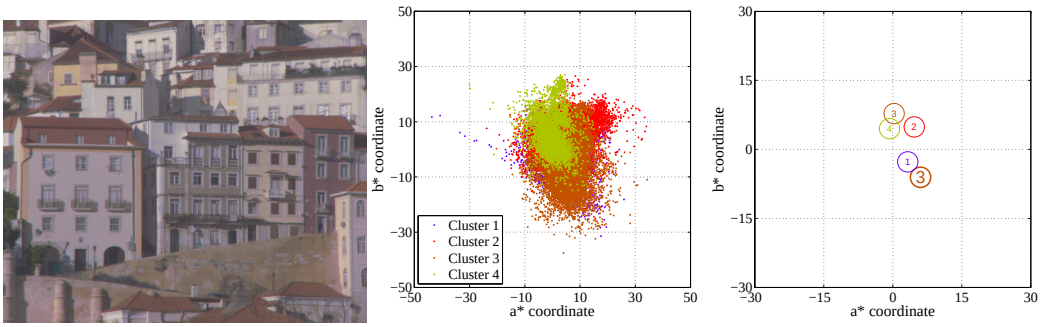
As a result, a new testing image dataset was created adding this predefined chromatic errors to all color clusters, and also to each color cluster individually. Moreover, and considering the different characteristics of the average relative radiance spectra of each cluster, for the two original hyperspectral images, different visual effects were produced with this methodology (see example in Fig. 5.3).



(a) Original image with the initial clusters representation.



(b) Cluster 3 changing, “in direction”.



(c) Cluster 3 changing, “out direction”.

Figure 5.3: Example of a test image and the corresponding clusters representation in CIELAB, when a chromatic error of $\Delta E^*_{ab} = 15$ was applied to cluster 3.

5.3 Subjective Quality Assessment

5.3.1 Laboratory Apparatus and Calibration

The experiment for subjective quality evaluation was conducted with the same laboratory apparatus and display calibration characteristics as described in Section 4.3.
Also, a High-Speed Video Eye Tracker (High-Speed VET, Cambridge Research Systems), properly calibrated for each subject, was used to register the eye gaze (tracking data) during each stimulus presentation.

5.3.2 Dataset

The dataset of testing images was created by adding chromatic errors with predefined directions (“in” and “out”) to all color clusters, and to each color cluster individually, as described in Section 5.2. The original hyperspectral images used in this dataset were obtained from two

Table 5.1: Corresponding ΔE^*_{00} average errors between the original image and the generated dataset with chromatic errors expressed in ΔE^*_{ab}.

Image	Induced Chromatic Changes	6		9		12		15		ΔE^*_{ab}
		In	Out	In	Out	In	Out	In	Out	**Direction**
1	All	5.06	6.43	7.15	9.53	9.01	12.36	10.64	14.87	**Average** ΔE^*_{00}
	Cluster 1	1.33	1.71	1.87	2.53	2.34	3.26	2.75	3.87	
	Cluster 2	1.33	1.78	1.92	2.67	2.45	3.51	2.95	4.30	
	Cluster 3	1.33	1.60	1.87	2.39	2.32	3.10	2.75	3.74	
	Cluster 4	1.06	1.34	1.50	1.96	1.88	2.49	2.21	2.98	
2	All	5.43	6.36	7.72	9.42	9.75	12.27	11.54	14.85	
	Cluster 1	1.58	1.92	2.23	2.83	2.80	3.67	3.29	4.41	
	Cluster 2	1.87	2.11	2.65	3.10	3.33	4.02	3.93	4.85	
	Cluster 3	1.17	1.48	1.69	2.24	2.16	2.95	2.59	3.61	
	Cluster 4	0.81	0.85	1.16	1.26	1.46	1.63	1.73	1.97	

sources: 1) the real rural scene (see Fig. 5.1 (a)) from the University of Manchester (UK) hyperspectral image database [138, 149]; 2) the real urban scene environment (see Fig. 5.2 (a)) from a hyperspectral image database provided by a joint research project between Universidade de Coimbra and Universidade da Beira Interior (PT).

The magnitudes of the chromatic error ΔE^*_{ab} ranged from 6 to 15 units in steps of 3 units. The chromatic errors with the magnitude equal to 3 are not used, since in the previous work [5] their MOS results were very close to the MOS of the original images. Moreover, as some color pixels could not be reproduced by the CRT display, they were clipped to the closest color of the display gamut surface represented in the CIELAB color space, as mentioned and studied by N. Katoh [132]. Typically, an average of 1% pixels were clipped. All the test images were displayed on the described visualization setup with a spatial resolution of 600×500 pixels. The area around the displayed image was filled with the average color (gray) of all the pixels that comprise the test images with a mean luminance of 15 cd/m^2. Hence, 80 transformed images were obtained, plus one sample of each original image ($\Delta E^*_{ab} = 0$), resulting in an 82 images dataset.

Table 5.1 reports the ΔE^*_{00} [45] corresponding average values that resulted with the application of the predefined chromatic errors, expressed in ΔE^*_{ab}, in each image. As expected, chromatic changes in all clusters produce larger values of ΔE^*_{00} than the produced by individual transformations in each cluster. These differences occur due to the different number of modified pixels in each case. However, it is also important to note that errors in the "out direction" always produce a higher ΔE^*_{00} average values although the number of pixels that suffer the chromatic error is exactly the same and suffered a ΔE^*_{ab} exactly with the same value. Only the direction is different.

5.3.3 Experimental Procedure

The standard test methodology SSCQE [72] was chosen for the experiment, as in the previous chapter. The experiment described in the following sections was divided into two parts and was carried out by a panel of volunteer subjects.

Experiment - Part 1

A subjective score evaluation test using 66 images [6] was achieved in this experiment. These images were selected from the created image dataset, where 2 are the original ones, and 64

correspond to different impairments, in two directions, with ΔE^*_{ab} errors of 6, 9, 12 and 15 units in each of the four clusters. This part of the experiment took about 22 minutes per subject.

Experiment - Part 2

In this part of the experiment, beyond the subjective score evaluation test, it was performed the register of the eye movements (tracking data) using 26 selected images [7]. Similarly, these images were selected from the created image dataset, where 2 are the original ones, and 24 have induced impairments. The induced impairments had ΔE^*_{ab} errors with 9 and 15 units in two directions, which were applied to all clusters and to the selected clusters of the original test images. The duration of this part of the experiment was around 12 minutes per subject.

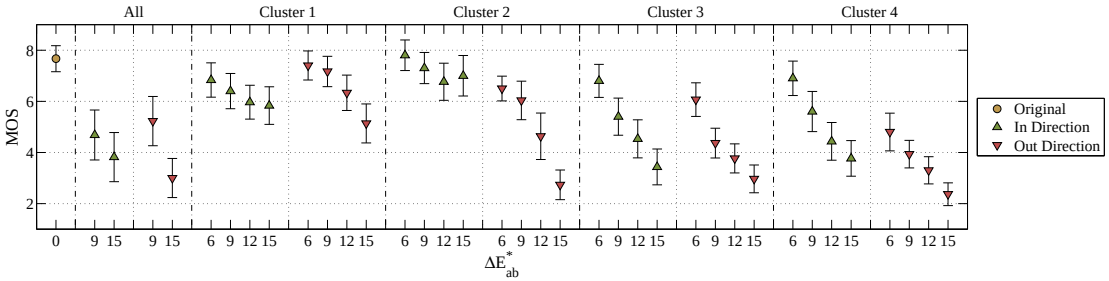
Prior to the beginning of the tests, an experimental protocol (similar to Appendix B), was shown to the panel of subjects. Each test starts with a "training period" where the subject was allowed to be familiarized with the score evaluation procedure using a different image set. After this period the "test session" itself takes place. Only the scores given by the subject during the "test session" were considered for the study. During the tests, the subject visualizes an image during 10 seconds, followed by the score evaluation using a joystick for selection of a grade between 1 (low quality/artificial) and 100 (high quality/naturalness) on a simulated continuous scale. In this context, naturalness means realistic image, which is opposed to an artificial one.
The images were observed by the subjects according to the same procedure described in Section 4.3.3.

5.3.4 Panel of Subjects

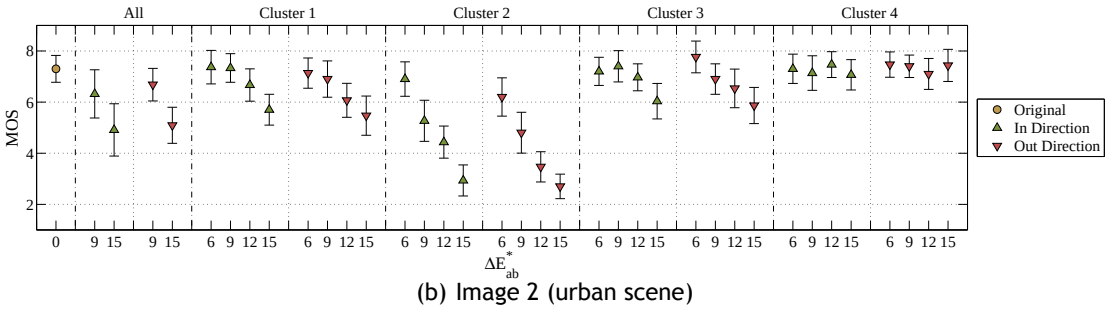
The panel of subjects volunteers for the first part of the experiment comprised 30 subjects, all within the range of 18-61 years old, with an average age of 28 ± 10 years old. This panel contained 15 female and 15 male subjects, keeping the gender balance.
For the second part of the experiment, a panel of subjects volunteers comprised 22 subjects, all within the range of 19-38 years old, with an average age of $22 \pm 5,87$ years old. This panel contained 10 female and 12 male subjects, also keeping the gender balance.
All subjects were naïve as to the aim of the experiment and all had normal color vision, which was assessed by the Cambridge Color Test [151] and the Ishihara's plates [173] (CSO Vision Chart, Costruzione Strumenti Oftalmici). The informed consent was obtained from all the subjects and the research was conducted according to the Declaration of Helsinki [152].

5.4 Subjective Score Data Analysis

According to Section 3.1.4, the processing of the subjective data taking into account the outlier detection, the MOS scores and the respective CI were properly calculated. The subjects were screened following the guidelines described for the SSCQE method in Section 2.3 of Annex 2 of the ITU-R Rec. BT.500 [72]. This procedure allows the detection and discards of subjects with ratings exhibiting a strong shift compared to the average behavior. However, in this study, no subject was detected as an outlier. Following the same methodology, the MOS scores and the CI were also computed for each test image.
The MOS values obtained for each chromatic error ΔE^*_{ab} and the respective CI for the experiment described in Section 5.3.3, are presented in Fig. 5.4. The results for the original images



(a) Image 1 (rural scene)



(b) Image 2 (urban scene)

Figure 5.4: MOS values and the 95% CI for the original images, and for images with chromatic changes in all clusters simultaneously, and in each cluster individually.

and images with chromatic errors in each cluster individually were obtained in part 1 of the subjective experiment, whereas the results for the images with chromatic changes in all clusters simultaneously were obtained in part 2 of the experiment.

As expected, the MOS results for the original images are higher than the obtained results for images with chromatic errors.

When the chromatic error is applied in all clusters simultaneously, independently of the magnitude and direction, the MOS for Image 1 is lower than the MOS for the Image 2. This difference can be explained by the fact that the Image 1 represents a rural scene with natural colors. Any chromatic changes on the original image, whose average spectra is shown in Fig. 5.1 (c), results in spectral shift inducing a non-natural appearance of the image. The image with chromatic impairments tend to loose the typical colors from nature, and subjects can perceive that loss of "naturalness". As a consequence, the obtained MOS was lower and decreases with the increase of the chromatic error.

In contrast, Image 2 represents an urban scene with a majority of colors that are typical from urban scenes and not from nature. The average spectral distributions of the colors of the original image, shown in Fig. 5.2 (c), reveals a neutral straight format. Any chromatic change on this image results in spectral shifts that keep the urban appearance of the image. As a result, the obtained MOS was higher. Nevertheless, it still decreases with the increase of the chromatic error.

It is important to observe that the MOS results for chromatic errors applied in some clusters individually are very similar to the MOS results obtained for chromatic errors applied to the whole image. Thus, these results show that there are also some color clusters that produce higher perceptual sensitivity to chromatic errors than others. In particular, the cluster 4 of Image 1 and cluster 2 of Image 2 lead to this higher perception. This difference in perception seems to be related to the covered areas, which correspond to the largest uniformity distribution of the natural colors presented on the ground fields of Image 1 (see Fig. 5.1 (g)), and on the stone wall of Image 2 (see Fig. 5.2 (e)). In this case, it is also important to note, that the

stone colors are also nature colors, and hence subjects become aware of this change when the color loses their naturalness. In opposition, it can also be noted that cluster 1 of Image 1 and cluster 4 of Image 2 produce less sensitivity to chromatic errors. This lower sensitivity seems to be related to the peripheral areas present in each of these two clusters. Moreover, any chromatic changes on cluster 4 of Image 2 always result in a MOS value similar to the original image, independently of the direction and the chromatic error magnitude of the induced error. This is due to the neutral chromatic characteristics of the unsaturated colors on this cluster (whites). A change on these colors represents the painting of the building with a different color and that is highly tolerated by subjects. These clusters (1 and 4 on image 1, and 2 and 4 on image 2) were then selected to be used in part 2 of the subjective experiment because they produce extreme perceptual behaviors.
It can also be noticed that the MOS sharply decreases when an increase of the chromatic error is induced in the "out direction". If the chromatic error changes are made to keep the cluster centroid in the same quadrant ("in direction") the MOS depicts acceptable values, even for a chromatic error of $\Delta E^*_{ab} = 15$. From these results can be concluded that some clusters can change up to $\Delta E^*_{ab} = 15$ keeping an acceptable quality, while others, with just $\Delta E^*_{ab} = 6$, are already in the limit of an acceptable quality.

5.5 Eye Tracking Data Evaluation

To better understand the visual behavior of the subjects during the stimulus visualization an eye tracker device was used to measure the eye movements. These measurements allow the subjects eye fixations computation for each visualized image (stimulus), and the generation of the overall fixation maps, allowing the visual RFT attainment.
The eye tracking data evaluation was carried out for the images dataset with chromatic changes in all clusters and on the 2 selected representative color clusters only, avoiding the subjects long time exposition. The chosen clusters were selected based on the particular characteristics of the MOS results obtained in part 1 of the subjective experiment, as mentioned in Section 5.4. Moreover, only 2 chromatic errors were considered, since smaller errors are well tolerated (lower sensitivity) by the subjects, and a chromatic step of 6 units was shown to be representative. Therefore, chromatic errors ΔE^*_{ab} equal to 9 and 15 were selected.

5.5.1 Eye Fixations and Relative Fixation Time

It is widely known that there is a relation between the subjects eye fixations and their information processing or cognitive activities [174]. An eye fixation occur when a subject focuses on a certain area during image visualization for longer than 100 ms. Therefore, most of the cognitive processes occurs during this period [175].
These eye fixations occurs in the foveola region of the retina that corresponds to an angular viewing of 0.8° [150]. This corresponds to a circular area with 65 pixels in diameter, considering the used display resolution and the viewing distance. Thus, it is expected that the data acquired during our experiment, and processed as eye fixations, will provide a useful understanding of the subject cognitive process to perform the quality evaluation. In this study, only the data of subjects with more than 90% of successfully acquired tracking was considered. In order to determine the ideal fixation time, a non-linear mapping correlation study [176] between the average RFT values in each cluster and the corresponding subjective MOS results was performed.

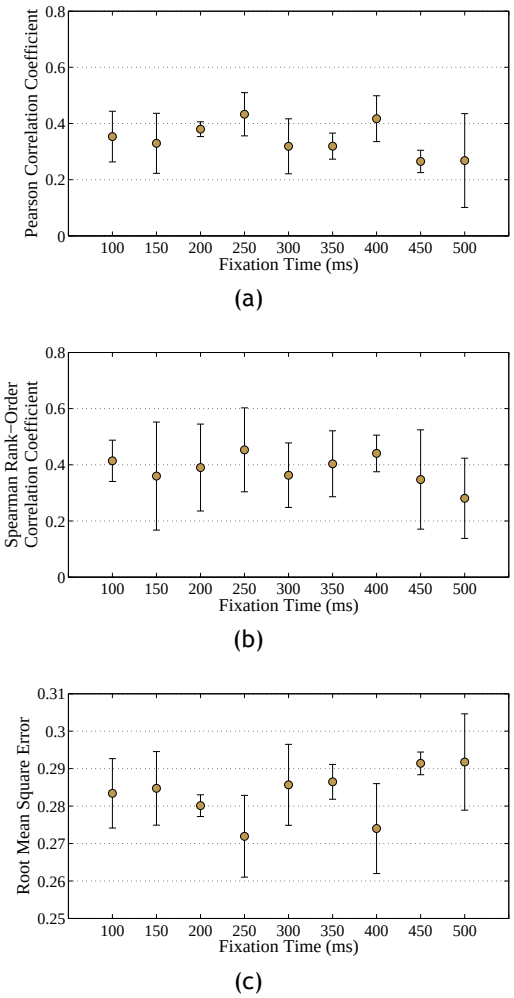


Figure 5.5: Quantitative methods (PCC, SROCC, RMSE) evaluating the MOS relation with fixation time.

For this, it was considered as fixation time the elapsed time value in the interval between 100 ms and 500 ms with a step of 50 ms. The Fig. 5.5 shows the results of 3 quantitative methods applied to evaluate the fixation time performance, the PCC, the SROCC, and the RMSE, detailed in Chapter 3. From the observation of these results, a fixation time of 250 ms was selected for analysis as it provides the best correlation with the MOS. Then, the number of fixations was determined for each image visualization. Since the dataset images were decomposed in 4 clusters, the number of subjects fixations in each cluster was weighted according to the percentage of pixels on each. Finally, the subjects overall fixation maps and the cluster fixation maps for each image was defined, and the corresponding RFT values were computed.

5.5.2 Results Analysis

The fixations maps and the RFT values for each chromatic error ΔE^*_{ab} are represented in Tables 5.2 and 5.3, for both test images. These tables show in each column the induced chromatic change, the visualized image, the global fixation map, the cluster fixation maps, and the corresponding RFT's boxplots, where the symbol $(\times)$ denotes the average RFT values for each cluster. The global fixation map is simply the overlap of the 4 individual cluster fixation maps. The rows in these tables depict the results for the original image, the images with chromatic changes ("In", "Out") in all clusters, and the images with chromatic changes in the 2 selected clusters

(C1 or C4).
A comparative analysis will be explained in the following sections for the 2 test images individually in order to study the correlation between the average RFT values and the different color errors that were induced, and also between the average RFT values and the corresponding MOS. For a statistical significance analysis of the results, a Shapiro-Wilk test [153] with a significance level of 0.05 was performed, and it was concluded that these results reject the hypothesis of a Gaussian distribution. Then, a Kruskal-Wallis one-way analysis followed by a multiple comparison test [154] was performed and the respective p-values in each case are presented in Tables 5.4 and 5.5 for each test image.

Image 1 analysis

In this analysis are considered the following groups of images: 1) the original image; 2) the group of images with chromatic errors in all clusters; 3) the group of images with chromatic errors induced only on cluster 1; 4) the group of images with chromatic errors induced only on cluster 4. Considering the fixation maps and the RFT's results for each cluster (see Table 5.2) the following can be observed.
Image 1 reveals certain areas that tend to attract the subjects fixation. However, the average RFT's are identical in all the clusters, as shown in Table 5.4, where no statistically significant differences were found $(p > 0.05)$.
The fixations maps of the second group of images, with changes in all clusters, reveal noticeable differences. The RFT's comparison between the original image and for clusters 3 and 4, revealed an increase with the chromatic error. Some of these differences exhibit a statistical significance $(p < 0.05)$, as shown in Table 5.4. Furthermore, the subjects become aware of color differences in some directions, resulting in fixation maps that are either more disperse or focused in some areas with colors that become unnatural.
For the third group of images, with chromatic changes induced on cluster 1, the average RFT's are similar when compared with the original image, which means that the chromatic changes are somehow meaningless. The cause for this behavior might be explained with the low luminosity on these cluster pixels and the location of the cluster regions that are mostly in the image periphery. Nevertheless, it is noticeable a slightly grew of the average RFT's on regions of cluster 4 that are very close to the limits of cluster 1. This increase is somehow balanced with the decrease of the fixations average on cluster 1, mainly due to local contrast. However, some of these differences present a statistical significance $(p < 0.05)$, as shown in Table 5.4.
For the fourth group of images with chromatic errors induced on cluster 4, subjects reveal high sensitivity to the chromatic error. Most of the average RFT's show a major increase on cluster 4 fixations when compared with the original image results. Moreover, some of these differences have a statistical significance $(p < 0.05)$, as also shown in Table 5.4.
Another relevant observation results when chromatic errors are induced to change the cluster centroid to a different quadrant ("out" direction) and a magnitude of $\Delta E^*_{ab} = 15$ units. In these cases, where the MOS revealed the worst values, the average RFT's also increased in cluster 3 and 4 for the second group, in cluster 3 for the third group, and in cluster 4 for the fourth group of images.

Table 5.2: Chromatic changes, visualized image, fixation map, and RFT boxplot for Image 1.

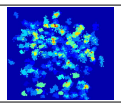
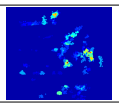
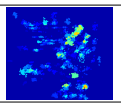
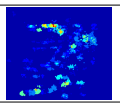
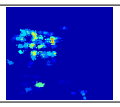
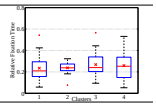
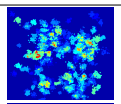
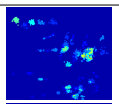
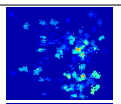
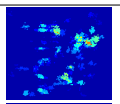
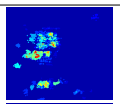
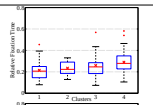
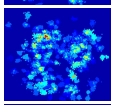
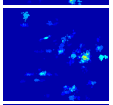
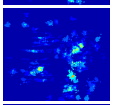
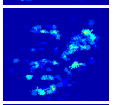
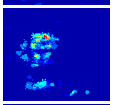
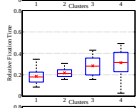
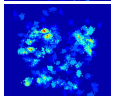
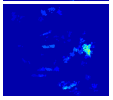
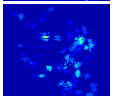
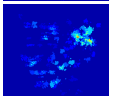
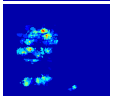
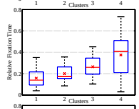
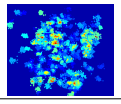
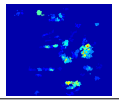
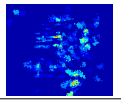
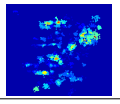
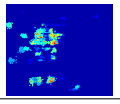
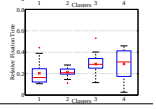
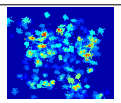
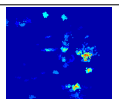
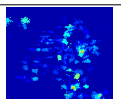
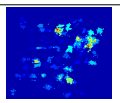
Chromatic changes	Image	Fixation Map					RFT
		Global	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Original							
$(All, 9, In)$							
$(All, 15, In)$							
$(All, 9, Out)$							
$(All, 15, Out)$							
$(C1, 9, In)$							
$(C1, 15, In)$							
$(C1, 9, Out)$							
$(C1, 15, Out)$							
$(C4, 9, In)$							
$(C4, 15, In)$							
$(C4, 9, Out)$							
$(C4, 15, Out)$							

Image 2 analysis

As in the previous section, the following group of images are defined: 1) the original image; 2) the group of images with chromatic errors in all clusters; 3) the group of images with chromatic errors induced only on cluster 2; 4) the group of images with chromatic errors induced only on cluster 4.
Considering the obtain results (in each cluster) for the fixation maps and the RFT's (see Table 5.3) it can be noticed as follows.
It is noticeable the influence of the stone wall where most of the pixels belong to cluster 2. The stone has nature colors, in opposition to most of the remaining areas that have human made colors. Moreover, and unlike Image 1, the average RFT's are not identical in all clusters, and statistically significant differences were found $(p < 0.05)$, as shown in Table 5.5. A larger fixation on cluster 2 and 3 is observed. This can be explained by the extension of the distribution of contiguous areas (stone wall and painted building frontages).
For the second group with images obtained with chromatic errors applied to all clusters, there are a few noticeable differences between fixations maps when compared with the original image. In general, these results revealed that there are some significant statistic differences between clusters fixation maps, as shown in Table 5.5. When chromatic errors are induced to move the cluster centroid to a different quadrant ("out" direction), independently of the error magnitude (9 or 15 ΔE^*_{ab} units), a larger dispersion on cluster 2 fixation maps is observed, when compared with the dispersion caused by chromatic errors induced to keep the cluster centroid in the same quadrant ("in" direction).
For the third group of images with chromatic errors induced only on cluster 2, there are few noticeable differences between fixations maps. Most of the average RFT's reveal similar fixation maps when compared with the results of the original image. In general, these results revealed that there are few significant statistic differences between cluster 2 and the others, as shown in Table 5.5. Subjects revealed high sensitivity to the chromatic errors on this cluster because it contains a large area of nature colors (stone wall), revealing a slightly grew on this cluster average RFT's. For chromatic errors of $\Delta E^*_{ab} = 15$ units, independently of the direction, the average RFT differences between clusters revealed no statistical significance, and also present the worst MOS values for this test image.
For the fourth group of images with chromatic errors induced only on cluster 4, there are few noticeable differences between each of the fixation maps. Most average RFT's reveal similar fixation maps when compared with the original image results. Moreover, and in general, there are few significant statistic differences between them, as shown in Table 5.5. Furthermore, this cluster is composed mainly of peripheral areas of the image (typical colors are very close to white), and it has the smallest number of pixels. The induced chromatic error tend to be undetected by subjects and are perceived as common building frontages painting. The subjects tend to change the attention behavior mainly when the chromatic errors are induced to change the cluster centroid for a different quadrant ("out" direction). However, none of these changes are reflected in the MOS results.

Table 5.3: Chromatic changes, visualized image, fixation map, and RFT boxplot for Image 2.

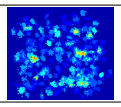
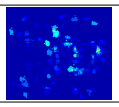
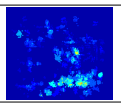
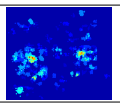
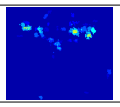
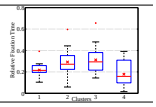
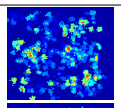
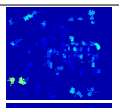
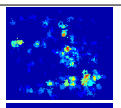
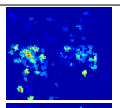
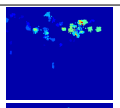
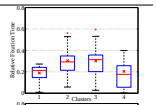
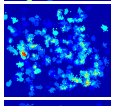
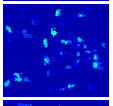
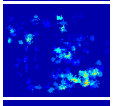
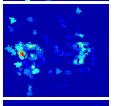
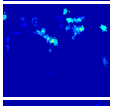
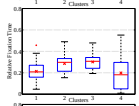
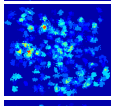
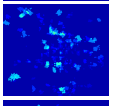
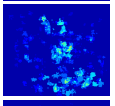
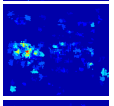
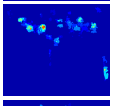
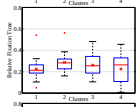
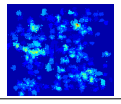
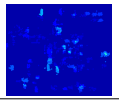
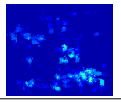
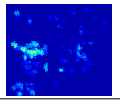
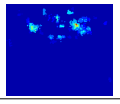
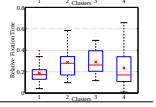
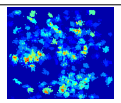
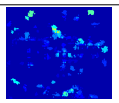
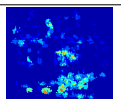
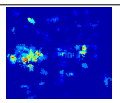
Chromatic changes	Image	Fixation Map					RFT
		Global	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Original							
$(All, 9, In)$							
$(All, 15, In)$							
$(All, 9, Out)$							
$(All, 15, Out)$							
$(C2, 9, In)$							
$(C2, 15, In)$							
$(C2, 9, Out)$							
$(C2, 15, Out)$							
$(C4, 9, In)$							
$(C4, 15, In)$							
$(C4, 9, Out)$							
$(C4, 15, Out)$							

Table 5.4: p-values for the multiple comparison of the relative fixation time between clusters in Image 1.

Chromatic changes	Clusters	1	2	3
Original	2	**0.867**		
	3	**0.611**	**0.970**	
	4	**0.769**	**0.998**	**0.994**
$(All, 9, In)$	2	**0.687**		
	3	**0.442**	**0.980**	
	4	**0.067**	**0.530**	**0.771**
$(All, 15, In)$	2	**0.428**		
	3	0.002	**0.144**	
	4	0.000	0.034	**0.937**
$(All, 9, Out)$	2	**0.428**		
	3	0.006	**0.288**	
	4	0.000	0.006	**0.456**
$(All, 15, Out)$	2	**0.933**		
	3	0.010	**0.060**	
	4	0.020	**0.098**	**0.996**
$(C1, 9, In)$	2	**0.829**		
	3	**0.237**	**0.735**	
	4	0.000	0.003	**0.066**
$(C1, 15, In)$	2	**0.864**		
	3	0.041	**0.243**	
	4	0.000	0.002	**0.343**
$(C1, 9, Out)$	2	**0.942**		
	3	**0.980**	**0.998**	
	4	**0.968**	**0.734**	**0.831**
$(C1, 15, Out)$	2	**0.689**		
	3	0.000	0.018	
	4	0.004	**0.103**	**0.913**
$(C4, 9, In)$	2	**0.545**		
	3	0.034	**0.511**	
	4	0.000	0.000	0.008
$(C4, 15, In)$	2	**0.998**		
	3	**0.078**	**0.118**	
	4	0.004	0.006	**0.734**
$(C4, 9, Out)$	2	**0.513**		
	3	0.003	**0.159**	
	4	0.003	**0.164**	**1.000**
$(C4, 15, Out)$	2	**0.967**		
	3	**0.634**	**0.886**	
	4	0.000	0.003	0.029

Table 5.5: p-values for the multiple comparison of the relative fixation time between clusters in Image 2.

Chromatic changes	Clusters	1	2	3
Original	2	**0.151**		
	3	**0.083**	**0.993**	
	4	**0.643**	0.006	0.002
$(All, 9, In)$	2	**0.112**		
	3	**0.300**	**0.960**	
	4	**0.999**	**0.085**	**0.243**
$(All, 15, In)$	2	**0.108**		
	3	0.021	**0.924**	
	4	**0.995**	**0.059**	0.009
$(All, 9, Out)$	2	**0.134**		
	3	**0.565**	**0.822**	
	4	**0.924**	**0.415**	**0.908**
$(All, 15, Out)$	2	**0.108**		
	3	0.021	**0.924**	
	4	**0.995**	**0.059**	0.009
$(C2, 9, In)$	2	**0.123**		
	3	**0.229**	**0.990**	
	4	**0.998**	**0.174**	**0.306**
$(C2, 15, In)$	2	0.039		
	3	**0.205**	**0.894**	
	4	**0.569**	**0.517**	**0.911**
$(C2, 9, Out)$	2	0.000		
	3	**0.503**	0.030	
	4	**0.981**	0.001	**0.743**
$(C2, 15, Out)$	2	**0.912**		
	3	**1.000**	**0.899**	
	4	**0.961**	**0.659**	**0.969**
$(C4, 9, In)$	2	**0.213**		
	3	0.010	**0.638**	
	4	**0.393**	0.003	0.000
$(C4, 15, In)$	2	0.023		
	3	**0.847**	**0.174**	
	4	**0.998**	0.014	**0.761**
$(C4, 9, Out)$	2	**0.155**		
	3	0.012	**0.774**	
	4	**0.965**	**0.366**	0.049
$(C4, 15, Out)$	2	**0.673**		
	3	**0.482**	**0.991**	
	4	**1.000**	**0.698**	**0.508**

Comparative correlation analysis

In order to provide quantitative measures on the RFT's performance, a non-linear mapping [176] was performed between the average RFT in each cluster and the subjective MOS results for the 2 test images (see Table 5.6 and 5.7 respectively).

The SROCC results (model prediction monotonicity) revealed that a MOS variation with a determined trend does not imply a similar variation of the RFT's values. However, the high PCC results (model prediction accuracy) revealed a strong correlation between the MOS and the average RFT's. This means that the RFT values are representative of the MOS behavior. However, the lack of monotonicity does not allow using the RFT's as a measure of the MOS.

Table 5.6: Quantitative methods (PCC, SROCC, RMSE) applied to evaluate the average RFT performance in each cluster for the Image 1.

Chromatic changes	Cluster	PCC	SROCC	RMSE
All	1	0.893	0.425	0.154
	2	0.894	0.400	0.154
	3	0.774	0.600	0.218
	4	0.879	0.300	0.163
Cluster 1	1	0.880	0.800	0.170
	2	0.897	0.700	0.160
	3	0.721	0.400	0.248
	4	0.897	0.675	0.160
Cluster 4	1	0.826	0.000	0.194
	2	0.826	0.425	0.194
	3	0.633	0.500	0.267
	4	0.826	0.400	0.194

Table 5.7: Quantitative methods (PCC, SROCC, RMSE) applied to evaluate the average RFT performance in each cluster for the Image 2.

Chromatic changes	Cluster	PCC	SROCC	RMSE
All	1	0.824	0.875	0.219
	2	0.941	0.825	0.130
	3	0.335	0.300	0.364
	4	0.672	0.300	0.286
Cluster 2	1	0.565	0.300	0.303
	2	0.883	0.700	0.172
	3	0.802	0.500	0.219
	4	0.846	0.300	0.195
Cluster 4	1	0.942	0.875	0.133
	2	0.898	0.900	0.174
	3	0.692	0.300	0.285
	4	0.577	0.375	0.322

5.6 Conclusion

This chapter describes a study on the quality perception of specific chromatic errors that were induced in 2 stimulus test images. With these test images, 2 classes are represented, a typical countryside scene (image 1), and a typical urban scene (image 2).

A subjective score, expressed by the MOS, and the visual behavior of the subjects during the stimulus visualization were measured. It should be noticed that the clusters size of the test images were sufficiently balanced, and there is no relation between that size and the sensitivity to chromatic errors. The RFT's were determined using a fixation time equal to 250 ms, as this value revealed to be the most representative choice for the objectives of this study.

The MOS results for the original images are higher than the MOS results obtained for the images with induced chromatic errors. Moreover, the average RFT's for Image 1 are identical in all the clusters, unlike the average RFT's for Image 2. When chromatic errors are applied in all clusters simultaneously, independently of the magnitude and direction, the MOS for Image 1 is lower than the MOS for Image 2. In terms of RFT's, there are differences in the resulting cluster fixation maps for the impaired images, when compared with the resulting ones for the original images. It can be observed that subjects revealed a larger perception to a loss of color naturalness than to the error magnitude itself. It was also verified that in presence of images with nature colors (rural scenes for instance) subjects become more sensitive to chromatic errors than when these kinds of colors are reduced, as it is typical in urban scenes.

There are color clusters closely related to the global perception of quality. This means that a chromatic error in a specific color cluster results in a subjective quality that is very close to the subjective quality produced by a global change (the cases where all color clusters were changed simultaneously). Considering Image 1, when the induced chromatic errors result in unexpected colors that should not be present in such real scenes, a large reduction of the MOS was also observed [6]. For Image 2, when the induced chromatic errors result in unexpected colors in some regions with nature colors (stone wall), a large decrease of the MOS was also observed [6]. In general, this region revealed an average RFT increase, when compared with the other image regions.

Smaller MOS values were obtained when the cluster color centroid tends to change to a different chromatic quadrant. By contrast, a higher chromatic error tolerance was found when the cluster

color centroid is kept inside the same chromatic quadrant.
Moreover, the RFT values revealed a high correlation with the MOS, although there is no proportional relationship between the measured RFT's and the MOS.

Part III

HDR Image Quality with JPEG XT

Introduction to JPEG XT for HDR Coding

Nowadays, HDR systems are becoming available for the general consumer as acquisition systems that are available in a large variety of photographic equipment and even in some mobile devices. In addition to the several HDR acquisition and display technologies, the Joint Photographic Experts Group (JPEG) has been standardizing new codecs for HDR content. HDR images are able to represent a wider range of luminance values, closer to the luminance range of the HVS, as described in Section 2.2.6. For this reason, HDR representation of image or video content is gaining increased interest as a mean to improve QoE, in particular, in imaging applications such as photography, TV, and Cinema. Typically, HDR images are displayed on legacy monitors using TMO's, which map HDR in a wider range of contrasts and colors to the range available in the displays.

The JPEG has recently developed a new standard, JPEG XT, with a set of extensions to the widely used JPEG codec, keeping the backward compatibility with any JPEG decoder. One of these extensions provides the means for HDR content encoding [177]. This standard keeps the backward compatibility with the conventional JPEG format by encoding a LDR version of the initial HDR image in the base layer using a conventional JPEG encoder. An enhancement layer with the encoded HDR residual information is added into the JPEG metadata field. The novelty of this scheme is that any conventional JPEG decoder can read the LDR image encoded in the base layer, while the HDR image is available for any JPEG XT decoder. Hence, JPEG XT keeps the JPEG legacy allowing, for example, a direct display on a conventional LDR monitor. The scheme shown in Fig. 5.6 represents a generic JPEG XT encoder model. Furthermore, a JPEG XT compliant decoder can use the residual layer to reconstruct a lossy or even lossless version of the HDR image.

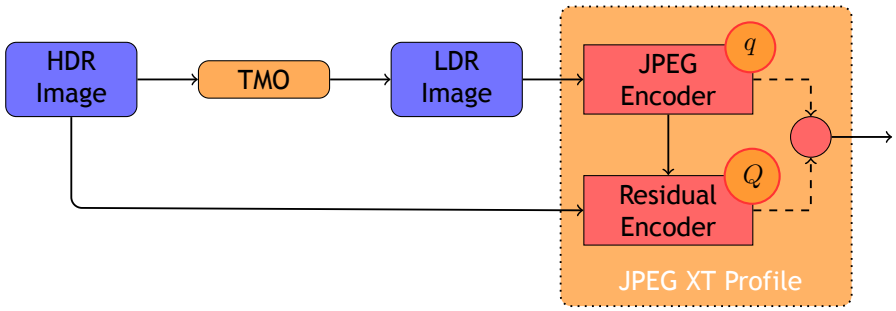


Figure 5.6: JPEG XT encoder model.

The legacy JPEG format is associated with a compressed image limited to a precision of 8 bits per component, while the JPEG XT encoder enlarges its precision up to 16 bits per component. Currently, JPEG XT defines four profiles (A, B, C, and D) for HDR image compression, of which profile D is a very simple entry-level decoder that roughly uses the 12 bit mode of JPEG. Profiles A, B, and C take into account the non-linearity of the HVS, and they essentially differ on the strategy used for creating the residual information and on the pre- and post-processing techniques [178, 179].

In profile A, the residual is represented as a ratio of the luminance of the HDR image and the tone-mapped image after inverse gamma correction. The residual is encoded in the logarithmic space and compressed as an 8-bit greyscale image [180].

In profile B, the image is split into "overexposed" areas, and LDR areas. The extension image

is represented as a ratio of the HDR image and the tone-mapped image, after inverse gamma correction. Profile B uses a difference of logarithms to approximate the ratio.
Finally, profile C works almost like profile A as it computes the residual image as a ratio of the HDR image and the inverse tone-mapped image. However, unlike the other profiles, the inverse TMO is not a simple inverse gamma, but rather a global approximation of the inverse of the TMO that was used to generate the base-layer image. Similarly to profile B, the ratio is implemented as a difference of logarithms. However, instead of using the exact mathematical log operation, profile C uses piecewise-linear approximation functions that have exact inverses, enabling near lossless coding [181, 182].
Moreover, the JPEG XT encoding standard depends mainly on two parameters: 1) The parameter q that controls the base layer coding quality index, which is exactly the same parameter that is used in the conventional JPEG implementations; 2) The parameter Q that controls the quality of the residual HDR information.

Chapter 6

Objective Quality Analysis of JPEG XT Performance

In this chapter is described the objective quality analysis of JPEG XT performance for HDR coding.
The first part that results from a joint work performed with two Qualinet research laboratories, namely EPFL (Switzerland) and CTU (Czech Republic), focus on the performance of the three profiles of JPEG XT, referred to as profiles A, B and C, using a test set of six HDR images. Four TMO techniques were used for the base layer image generation to assess the influence of the TMO's on the performance of JPEG XT profiles. Then, the HDR images were coded with different quality levels for the base layer and for the residual layer, and the performance of each profile was evaluated using the SNR and FSIM objective metrics.
The second part focus on the influence of JPEG XT codec modes on LDR generation using TMO's. JPEG XT encodes HDR images into a two layer scheme, encoding a LDR version of the image in a base layer, and the residual HDR information in an enhancement layer. The question addressed here is to understand if this model allows to extract a new LDR representation using a different TMO, independently of the TMO used to generate the LDR image encoded in the base layer. In particular, the use of *Drago* [57], *Mai* [64], *Mantiuk* [62], *Reinhard* [59], and *Reinhard & Devlin* [56] TMO's will be considered together with the three main profiles defined on the JPEG XT standard.
To study this effect of JPEG XT on HDR image coding, images generated with a TMO applied to the decoded HDR image are compared with the TMO's generated from the original HDR image, using the popular FSIM metric. Moreover, in cases with similar perceptual quality, the bit rate increase, compared with the TMO encoding with JPEG, is also studied.

6.1 Introduction

As reported in the previous section, the upcoming JPEG XT was recently adopted for image compression. A very important feature of this new standard is the backward compatible coding of HDR images. In that case, the standard defines a codec where a LDR version of the HDR image generated by a TMO is encoded using the conventional JPEG coding as a base layer and the extra HDR information is encoded in a residual layer. However, what are the advantages of using it, as the residual information represents a grew on bit rate? Currently, there are two possibilities for HDR images visualization, 1) In a HDR capable display, which requires the ability to display very high contrasts. These displays are very expensive and for that reason, they are not easily available nowadays; 2) To display a LDR version of the image, which is the typical situation. In this case JPEG XT already made that version available, but why should some one receive the HDR residual information of the enhancement layer if it is useless in this scenario?
Even if a HDR display is not available they might be in the near future. Then, it might be interesting to save images in a HDR format anyway. Moreover, and in practice, even if no HDR display is available, the HDR information encoded on the JPEG XT stream can be used to generate new LDR versions of the HDR image using different techniques. In fact, this is a useful application,

as typically the LDR generation in the encoder side results from an automatic procedure using a predefined TMO. It is well known that TMO algorithms tend to produce unnatural colors. Hence, if the encoded LDR version of the HDR image in the JPEG XT stream is not visually satisfactory, the encoded HDR image can be used to generate a new LDR version, using a different TMO algorithm or even a new adjustment on the initial TMO method parameterization.

The first study of JPEG XT performance on HDR domain is described in Section 6.2. This study objectively investigates the performance of the three main profiles, of the JPEG XT standard for JPEG backward-compatible HDR image compression, using a set of HDR images. Moreover, TMO techniques were used for the base layer image generation to assess the influence of the TMO's on the performance of JPEG XT profiles. The performance of each profile was evaluated using two objective metrics, and the results were studied and compared.

Furthermore, in Section 6.3, an analysis of the influence of the JPEG XT encoding on the generation of new LDR representations extracted from the HDR image is studied. This study also analyses the performance of the three main profiles of JPEG XT. LDR images generated using a specific TMO applied to the original HDR image are compared with the TMO generated using the decoded HDR image using perceptual metrics. The novelty is that the study addresses the LDR generation quality, instead of the HDR quality offered by JPEG XT [9, 11, 12].

6.2 JPEG XT Performance on HDR Domain

This section is focused on the objective evaluation of JPEG XT performance on HDR domain. The evaluation is a follow-up from verification tests performed jointly by JPEG HDR Ad Hoc Group and the European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003 - Qualinet). To accomplish this analysis performance, six HDR images were selected and coded using three profiles of JPEG XT, referred to as profiles A, B and C in accordance with the way they are identified in JPEG XT specifications. Moreover, four TMO's were used for the base layer, and different combinations of compression qualities were used for the base layer and for the HDR residual layer coding. The following are the four main objectives of this evaluation. 1) To verify the profiles of JPEG XT by checking if they produce an intended HDR JPEG backward-compatible stream when given an original HDR image and a tone-mapped LDR version; 2) To better understand the influence of TMO's on JPEG XT performance; 3) To better understand the degree of content dependency in JPEG XT performance; 4) To study the influence of parameters q and Q that control the balance of quality and bit rate between LDR and HDR portions of a JPEG XT coded image.

6.2.1 Evaluation Methodology

In this study, six HDR images have been selected by researchers of the Multimedia Signal Processing Group laboratory from EPFL, Switzerland, based on their visual content and quality. These test images were obtained from two sources: one scene, named *cadik-desk01*, from Čadík dataset[1], and five scenes, named *Knossos8*, *LowerLewisFalls*, *RevelStoke*, *SwissSunset*, and *Zurich2*, from the EMPA dataset[2]. The dynamic range of the content was also among the selection criteria.

[1] http://dcgi.felk.cvut.cz/home/cadikm/tmo/

[2] http://empamedia.ethz.ch/hdrdatabase/index.php

(a) *cadik_desk01* (b) *Knossos8* (c) *LowerLewisFalls*

(d) *RevelStoke* (e) *SwissSunset* (f) *Zurich2*

Figure 6.1: Tone mapped versions of the selected HDR images.

Table 6.1: Selected HDR images resolution information.

Scene	Resolution (pixels)
cadik_desk01	1920 × 1080
Knossos8	4916 × 3273
LowerLewisFalls	3800 × 2516
RevelStoke	3846 × 2560
SwissSunset	4916 × 3273
Zurich2	4916 × 3273

Natural HDR images are often acquired following a process that requires a set of pictures with different exposures from the same scene, which are then fused in a single one. If this fusion is carried out without a special care, HDR images generated by this mechanism tend to have poor edges or exhibit various motion blur distortions in regions where objects move between exposures.

In this image selection, any potential edge and blur distortions were carefully examined by observing images on a SIM2 HDR display to make sure that no such distortions were present. Tone mapped versions of the selected HDR images are shown in Fig. 6.1. Moreover, the dynamic range was optimized considering a more robust maximum and minimum estimation procedure described in [183]. The resolutions of the selected HDR images are provided in Table 6.1.

Moreover, three global and one local TMO's based on their performance on the selected images were chosen, namely, a *Gamma* TMO, *Reinhard* global TMO [59], *Drago* TMO [57], and *iCAM06* TMO [63]. Further details about these TMO's were described in the Section 2.2.6. The *Gamma*, *Reinhard* and *Drago* TMO's implementations available in Banterle's HDR Toolbox [183] were used. In the case of *iCAM06*, the MATLAB code provided by the authors was employed.

The parameters of particular TMO's were set in order to provide visually appealing results. For the *Drago* and *iCAM06* TMO's, the default setting parameters were sufficient, whereas for the other operators, the parameters were optimized according to the method proposed by Krasula et al. [184]. In some cases, the final parameters were additionally manually adjusted to create more natural images. The setting parameters for TMO's can be found in Table 6.2.

Table 6.2: Setting parameters of the selected TMO's for each selected scenes.

Scene	*Gamma* TMO		*Reinhard* TMO		*Drago* TMO		*iCAM06* TMO		
	clipped %	γ	p_α	p_{white}	Ld_{max}	bias	L_{max}	p	γ
cadik_desk01	3.39	3.6162	0.6470	111.2977	100	0.85	20000	0.7000	1
Knossos8	0.00	3.2710	0.7423	18.0095	100	0.85	20000	0.7000	1
LowerLewisFalls	2.50	2.2000	0.3809	28.6045	100	0.85	20000	0.7000	1
RevelStoke	0.19	2.9150	0.8705	24.7873	100	0.85	20000	0.7000	1
SwissSunset	0.00	4.5830	0.4948	52.7505	100	0.85	20000	0.4617	1
Zurich2	2.15	2.4479	1.0039	5.9081	100	0.85	20000	0.7000	1

Moreover, a large set of the two quality encoding parameters (q and Q) was used for the tests. The base layer parameter q varied from 40 to 95 in steps of 5, plus the value of 99. Smaller values than $q = 40$ were not used, since the quality of the base layer for lower q values was considered to be too low for any practical application. As for the residual layer quality parameter Q, it varied from 20 to 95 in steps of 5, plus the 99 value. For the residual encoding, it makes sense to use smaller quality values, since Q essentially controls the compression of the dynamic range of an image. Furthermore, $q = 99$ and $Q = 99$ were added to represent a near lossless example in all JPEG XT profiles under study. However, because of the inherent characteristics of profiles A and B, near lossless representations were not achieved even with these parameters. Only profile C, which combines two layers information in an additive manner, could reach near lossless performance.

The procedure scheme for the JPEG XT performance evaluation on HDR domain is represented in Fig. 6.2. In this study, results of two practical use case scenarios are considered. One is a typical JPEG encoding case, when the value of the parameter q is equal to 75, providing tone mapped images with common JPEG compression quality. And another is an optimal case with optimal rate distortion achieved by each profile when the distortion is estimated by SNR and without any a priori constraint on the quality of LDR image. The latter was obtained by varying both parameters q and Q, and selecting the combinations that provided the best performance after a full search.

The performance of each test case was evaluated with two FR objective metrics, refereed in Section 3.2.1. The SNR and the FSIM [98] were computed for different bit rates and images resulting from the use of different quality parameters. SNR is given by,

$$SNR_{dB} = 10 \log_{10} \frac{P_{image}}{P_{noise}} \quad (6.1)$$

where P_{image} is the power of the reference image (computed over the three color components) and P_{noise} is the power of the distortion due to compression.

The FSIM [98] is a perceptual metric based on the SSIM [86] concepts. FSIM adds the comparison of the low-level feature sets between the reference image and the distorted image [98], and is detailed in Section 4.4.1. The disadvantage of this metric is that it saturates very quickly for the operating points selected in these experiments. However, this saturation may also be related to the fact that the images perceived quality also reaches saturation for higher bit rates when observed by human subjects.

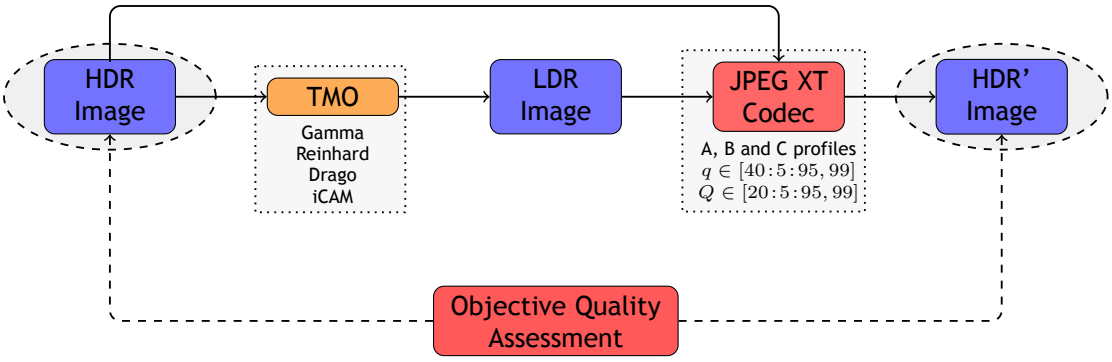


Figure 6.2: JPEG XT performance evaluation flowchart.

6.2.2 Results Analysis

The rate-distortion (SNR versus overall bit rate) curves were calculated for the three profiles (A, B, and C), four TMO's (*Gamma*, *Reinhard*, *Drago*, and *iCAM06*), and six images (*cadik-desk01*, *Knossos8*, *LowerLewisFalls*, *RevelStoke*, *SwissSunset*, and *Zurich2*). A significant number of operating points were produced by varying the base layer parameter q and the residual layer parameter Q as specified in Section 6.2.1.

To provide performance assessments under realistic situations, in a first use case, the LDR base layer quality was constrained to $q = 75$, since it is a common quality parameter used in JPEG compression. A second use case considers the optimal combinations of q and Q that guarantee the best relation between the SNR and bit rate. Moreover, no constraints were imposed on the quality of the base layer of the JPEG XT coded image.

This study is focused on the following performance comparative scenarios: 1) four TMO's are used for the same image (*RevelStoke*); 2) all six images are encoded while the same TMO is used (*Reinhard*).

The Fig. 6.3 presents both scenarios for SNR metric when the base layer quality parameter is set to $q = 75$, while Fig. 6.4 presents both scenarios for SNR metric when the optimal combinations of two quality parameters were used. All plots in both figures have a logarithmic scale on their horizontal axis.

From Fig. 6.3 the difference in the used profiles behavior can be observed. Profiles A and B behave similarly. That can be explained by the intrinsic nature of their encoding mechanisms. Profile C outperforms profiles A and B with higher SNR for the higher bit rates, as the SNR values grow with the increasing bit rate. Implementations of profiles A and B tend to saturate the SNR, while profile C implementation exhibits an increasing SNR as the overall bit rate grows. This may be due to the additive nature of the profile C, since its residual layer is added to the base layer, providing a continuous mechanism to improve the proximity to the base layer data. However, this approach also leads to worse SNR values for lower bit rates. Also, from Fig. 6.3, profile A outperforms other profiles for the lower bit rates. However, it reaches the saturation quicker than others in the higher bit rates.

The influence of the TMO in the rate-distortion characteristics of the different profiles can be observed in Fig. 6.3 as well. Different profiles present different rate-distortion curves for different TMO's. Some TMO's can lead to the best performance of one profile and to the worst performance of the other profile. For instance, *Reinhard* TMO performs better in profile B and worse in the two other profiles. The Fig. 6.3 also suggests that profile A is the least affected by the influence of different TMO's. Profile B, for the *Gamma* TMO and for a fixed q keeps the SNR stable for the increasing Q. This was noticed for all images, except for *LowerLewisFalls*, which

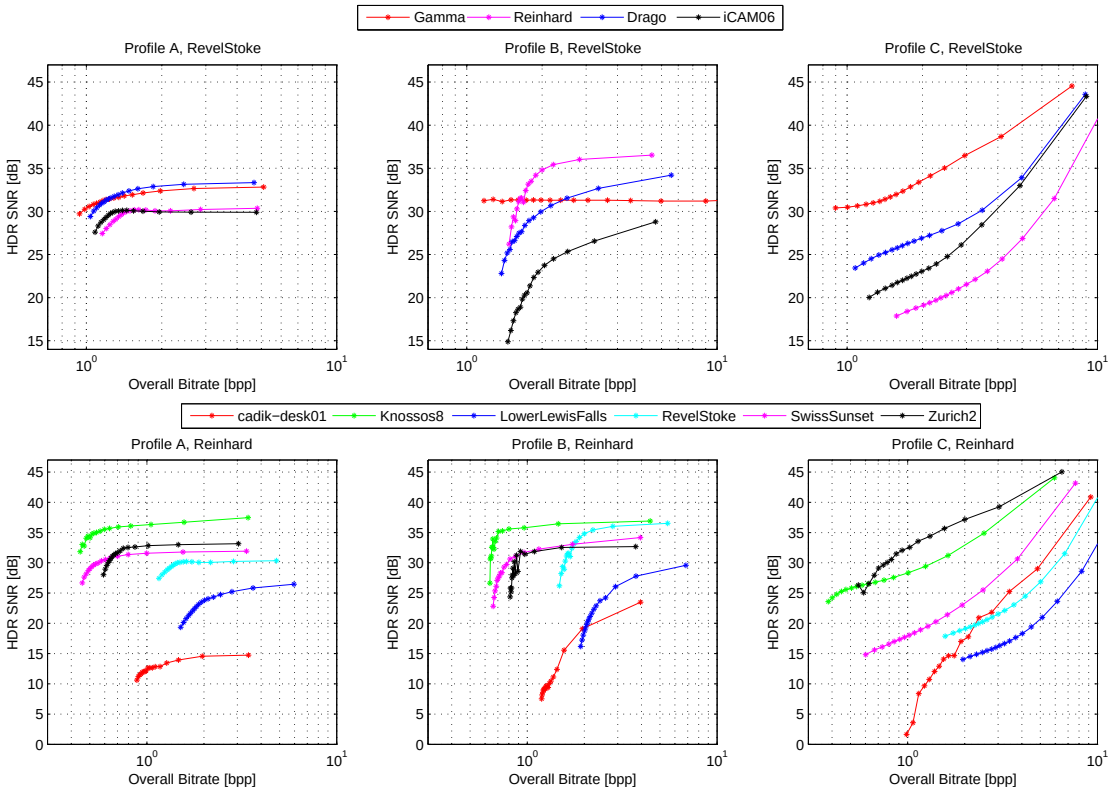


Figure 6.3: SNR metric values (varying Q values for $q = 75$) for the three profiles of JPEG XT.

passed through an automatic gamma correction, while the others had a fixed gamma correction.

The Fig. 6.3 also indicates that image content has an important influence on the SNR in all profiles. This is expected since all profiles compress the TMO versions in the base layer, keeping the backward compatibility with the conventional JPEG decoders. All profiles showed a similar behavior considering the content dependency. This means that the most demanding images, in terms of the relation SNR vs. bit rates, are the same for all profiles. The images *LowerLewisFalls* and *cadik-desk01* exhibit the lowest SNR values. The similarity between profiles A and B is evident also in this domain as they clearly separate the images into two different groups of performance. Hence, *LowerLewisFalls* and *cadik-desk01* are in the lower performance group, while the other images are in the higher performance group. For the image *cadik-desk01* all profiles show a poor SNR for the lower bit rates. Moreover, profile A always shows the lowest performance for this image. This image has the highest dynamic range as can be seen in Table 6.1. Finally, it is also important to observe that the SNR versus overall bit rate curve of *Zurich* image in Fig. 6.3 for profile B presents an irregular behavior for the lowest bit rates considered in this study.

The Fig. 6.4 reveal the results obtained by using the optimal combination of the two quality parameters, q, and Q. It can be observed that some curves in the figure exhibit an abrupt change in their tangential behavior. This is due to the discrete nature of the selected operating points that have a fixed step and also a limited range of values. Hence, this figure represents the optimal combinations for the set of computed operating points. This also confirms most of the observations pointed out in Fig. 6.3. The SNR metric saturates for profiles A and B. Moreover, it saturates faster for profile B than for profile A. For the tested values and images, profile C does not present this saturation effect.

Furthermore, from Fig. 6.4, it can be observed that profile A is the least dependent on the TMO.

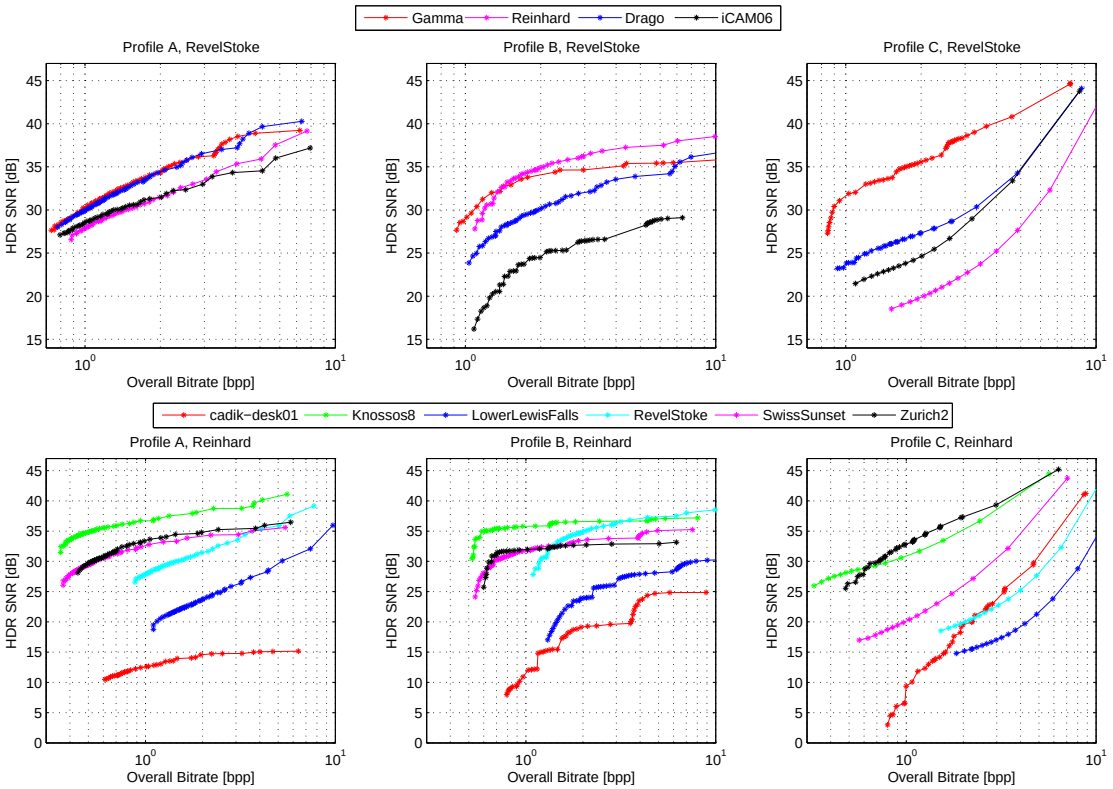


Figure 6.4: SNR metric values (combination of Q and q maximizing SNR) for the three profiles of JPEG XT.

In opposition, profile C seems to be the most dependent on the TMO. These plots also show the content dependency of the profiles performance. Profile B revealed the lowest dependency while profile C the highest. For profile A, the SNR saturates very fast for the image *cadik-desk01*. Furthermore, the profile A reveals to be the most sensitive to the base layer parameter q, while the residual layer Q parameter has a higher influence on profile C. Moreover, profile B is sensitive to both parameters, in the sense that a change in any is always reflected in the relation between bit rate versus SNR.

The FSIM versus bit rate curves are also shown for the three profiles (A, B, and C) in Fig. 6.5. The figure presents the performance comparison for the six images using the *Reinhard* TMO with varying Q parameter and fixed $q = 75$. As already discussed in Section 6.2.1, FSIM metric reaches the saturation very quickly. Profile C is the profile with the quickest saturation, which is the opposite behavior compared to the SNR metric performance. The lowest performance denoted in SNR evaluation for *cadik-desk01* image (see Fig. 6.3 and Fig. 6.4) is not observed in the case of FSIM. Otherwise and generally, this metric corroborates the SNR results. The *Knossos8* image leads to the highest FSIM, which is inline with the SNR case observations. Furthermore, the FSIM also confirms the close behavior between profiles A and B observed with the SNR analysis. The FSIM is a perceptual metric and this lack of discrimination could be due to the fact that high bit rates lead to a significantly higher perceived quality for all profiles.

6.2.3 Comments

The evaluation results verify the performance of all profiles to be appropriate for compression of HDR images. The results for the SNR metric demonstrate that rate-distortion functions of profiles A and B are similar, whereas rate-distortion behaves differently for profile C. It is also

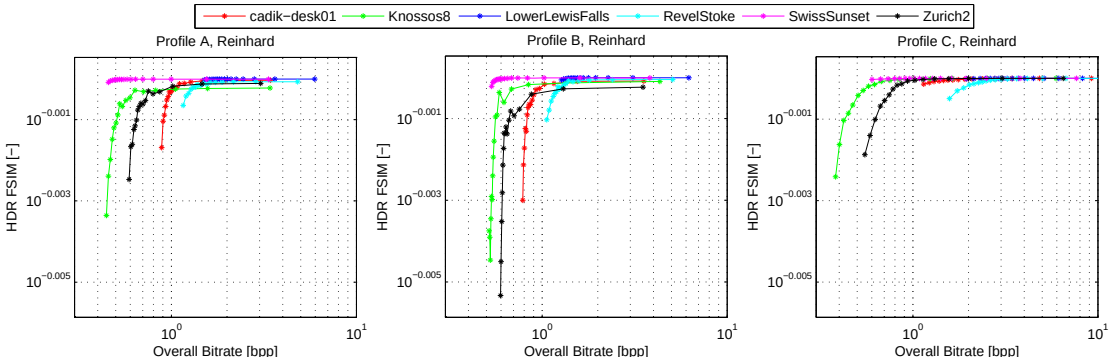


Figure 6.5: FSIM metric values (varying Q values for $q = 75$) for three profiles of JPEG XT.

evident that profile A exhibits less dependency on the TMO's used for base layer; profile B shows faster saturation for higher bit rates; profile C, while demonstrating no saturation and being able to encode images with high bit rates, performs worse on low bit rates when compared to profiles A and B. The results using the FSIM metric show faster saturation for all profiles when compared with the SNR. Nevertheless, the profiles show similarly consistent behavior for the FSIM metric and for the SNR metric.

6.3 Influence of the JPEG XT Encoding on TMO Generation

As analyzed before, JPEG XT decoded HDR images might be used for the generation of new TMO's. This application might be useful when the consumer of the JPEG XT codec has no visualization mechanism for HDR content. In this case, the consumer might not be happy with the LDR version existing on the JPEG XT stream. This is a very common situation because TMO's often generate images that become very unnatural, lacking from visual appeal.
This section deals with the quality of LDR images generated from HDR images encoded with JPEG XT. Similarly to what was done in the previous section, to accomplish this analysis performance, a set of variables needed to be considered. Firstly the testing content needed to be defined, and then a set of TMO algorithms were chosen. This performance analysis will also depend on the JPEG XT encoding parameters and profile. In the following the options assumed in this study are detailed.

6.3.1 Evaluation Methodology

Seven HDR images were selected for this experiment based on its perceptual quality and on their dynamic range characteristic. LDR test images generated with a TMO applied to these HDR images are represented in Fig. 6.6. These test images were obtained from three sources: 1) four scenes (a), (b), (c) and (g) from the Fairchild dataset[3]; 2) two scenes (d) and (e) from the EPFL dataset[4]; 3) one scene (f) from the EMPA dataset[5]. Moreover, all of the test images have resolution equal to 1920×1080 pixels.
At the beginning of this study, several TMO's have been considered. Finally a set of TMO's available in the pfstmo library[6] were selected: 1) *Drago* TMO [57] that is a global logarithmic

[3]http://http://rit-mcsl.org/fairchild/HDR.html
[4]http://http://mmspg.epfl.ch/jpegxt-hdr
[5]http://empamedia.ethz.ch/hdrdatabase/index.php
[6]http://pfstools.sourceforge.net/

(a) *507*

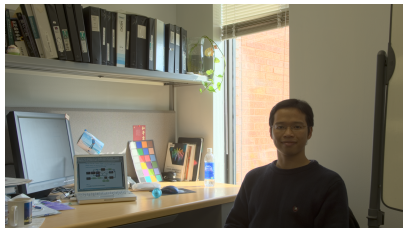
(b) *HancockKitchen*

(c) *LasVegasStore*

(d) *oldfactorywindow*

(e) *set33*

(f) *SwissSunset*



(g) *WillyDesk*

Figure 6.6: Tone mapped versions of the selected HDR images.

operator; 2) *Mantiuk* TMO [185] that is a local operator for contrast enhancement; 3) *Reinhard* TMO [59] that is a very popular global photographic TMO; 4) *Reinhard* & Devlin TMO [56] that is a local operator based on the simulation of the visual system adaptation; and 5) *Mai* TMO [64] that was developed for encoding optimization.

An example of the images generated with this five different TMO's using the pfstmo library is presented in Fig.6.7. As it is well known different TMO's generate different images with different visual attributes. Even the same TMO might lead to different images, depending on the parameterization. Hence, the generation of new TMO images from the JPEG XT encoded HDR images can make sense in the case of a consumer that has no HDR capable display and is not satisfied with the initial LDR version present on the JPEG XT encoding.

In the study referred in the previous section, a large set of the two quality encoding parameters (q and Q) was used for the tests. In this study, the option of fixing $q = 75$ was assumed as it is the typical value used for JPEG compression near the perceptual limit of accepting quality. Moreover, the encoding parameter Q was varied from 40 to 95 in steps of 10, plus the value of 99.

(a) *Drago* TMO (b) *Mantiuk* TMO

(c) *Reinhard* TMO (d) *Reinhard & Devlin* TMO

(e) *Mai* TMO

Figure 6.7: Tone mapped images generated with the five selected TMO's of the HDR image *LasVegasStore*.

To analyze the LDR image generated from the decoded JPEG XT HDR images, they will be compared with the LDR image generated before any encoding with the same TMO. Fig. 6.8 represents the whole comparison model.

As an example, an HDR image is encoded using as base layer an LDR image generated with the

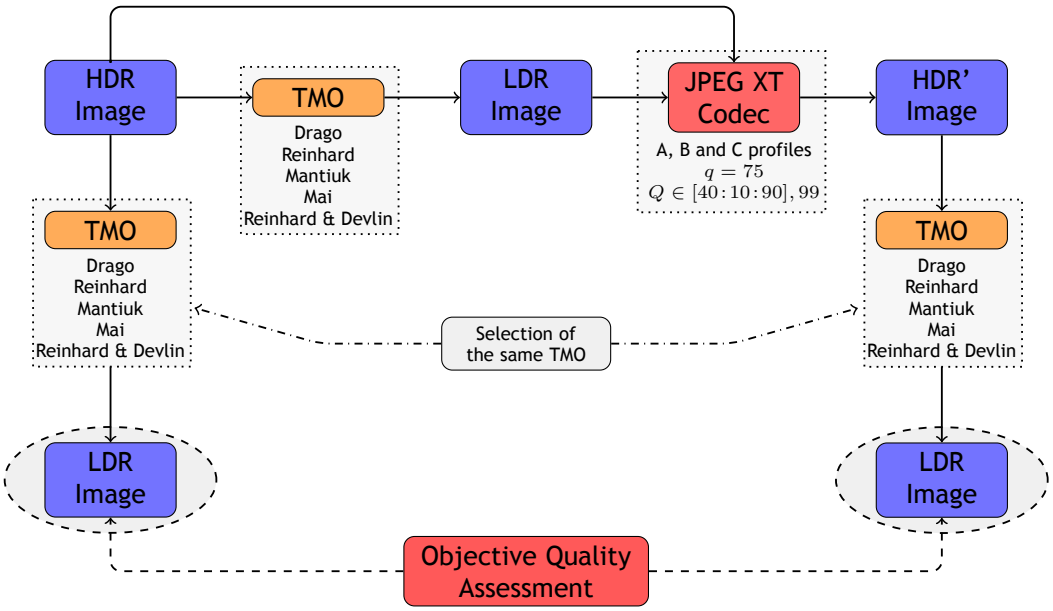
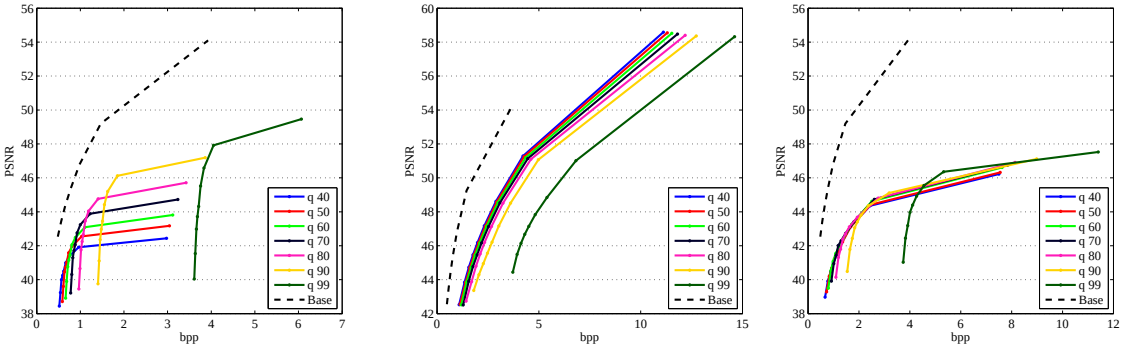
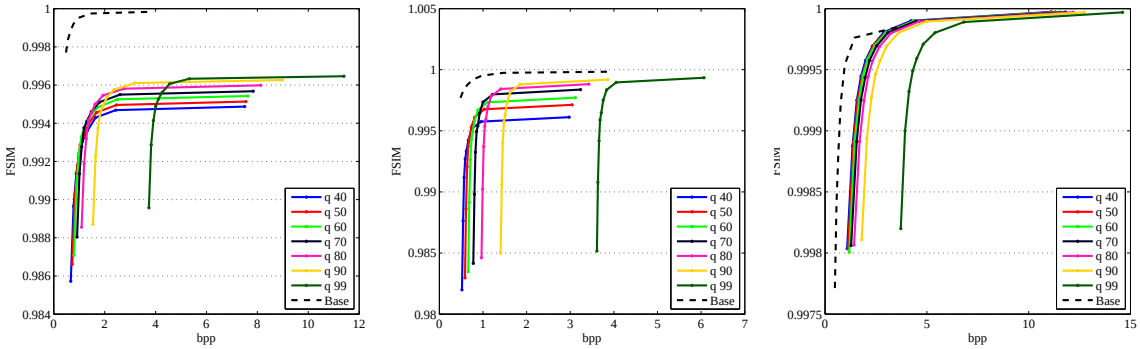


Figure 6.8: JPEG XT encoding on TMO generation flowchart.



(a) Profile A (b) Profile B (c) Profile C

Figure 6.9: PSNR values for the three profiles, obtained for the *Mantiuk* TMO. The *Reinhard* TMO was used for LDR encoding.



(a) Profile A (b) Profile B (c) Profile C

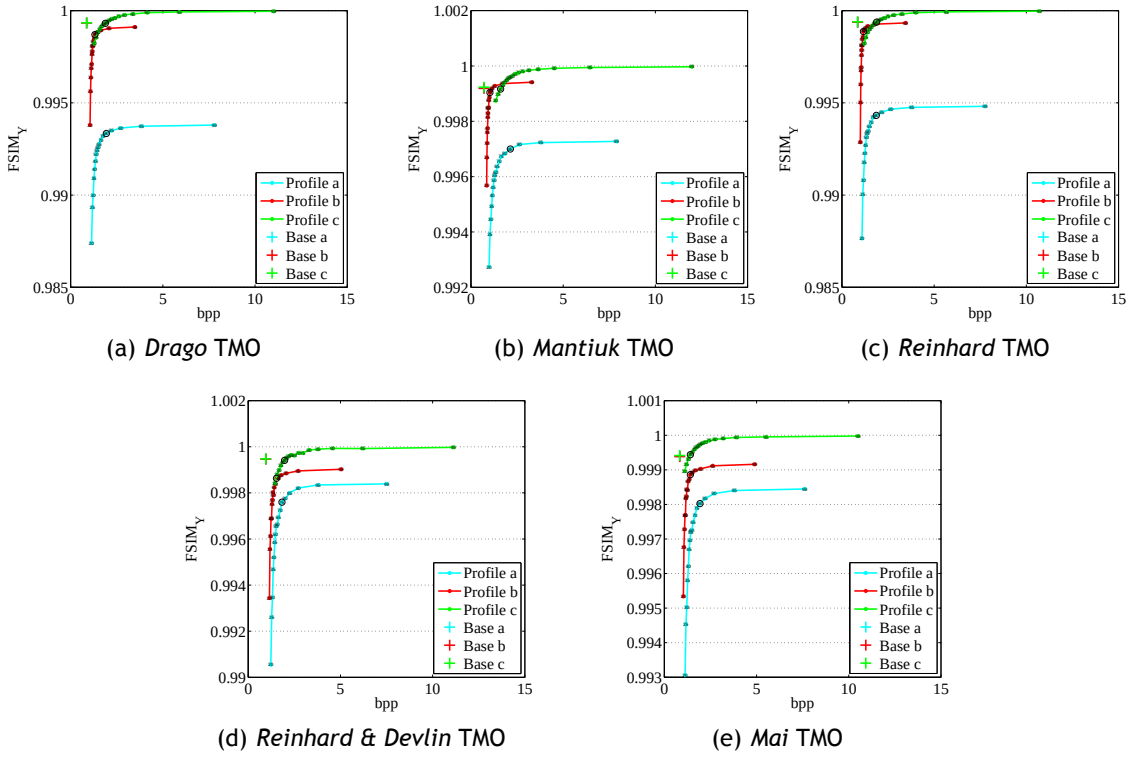
Figure 6.10: FSIM values for the three profiles, obtained for the *Mantiuk* TMO. The *Reinhard* TMO was used for LDR encoding.

Reinhard TMO, can lead to an LDR image generated with *Mantiuk* TMO after decoding with a JPEG XT decoder. This image is compared with the image generated from the original HDR image with the same *Mantiuk* TMO. As in this study were used 5 different TMO's, this leads to 20 comparisons (considering that the same TMO used in the encoder side is also tested in the decoder side). This study also involves the 3 profiles A, B and C.

To compare the images generated by a TMO before and after coding, initially the PSNR was used, but the results analysis shown that this was not the best approach (see Figs. 6.9 and 6.10).

Hence, perceptual metrics, like the SSIM [86] or FSIM [98] were tested. Both provided a better method for the comparison and FSIM was chosen, as it seems to be more stable in this context. As can be seen in Figs. 6.9 and 6.10 there is a value of Q for each q case where the perceived quality seems to reach a certain level that almost cannot be improved. These plots are representative of typical behavior for the different study cases (different content and used TMO's). For reference, the metrics results computed for the LDR base layer are also shown with a dash line. Moreover, this saturation point, provided very similar results between the LDR image generated from the decoded HDR image and from the original HDR image.

Furthermore, it is important to observe that in case of profile C, the quality can be grown in the enhancement layer until a near lossless situation. However, that is not the case with profiles A and B.



(a) *Drago* TMO (b) *Mantiuk* TMO (c) *Reinhard* TMO

(d) *Reinhard & Devlin* TMO (e) *Mai* TMO

Figure 6.11: Decoder FSIM result versus bit rate example for $q = 75$ using *Reinhard* TMO in the base layer in *LasVegasStore* image.

6.3.2 Results Analysis

For the case where the JPEG XT encoder uses an LDR image generated, for instance, with the *Reinhard* TMO, would be interesting to understand what is the increase on bit rate considering the JPEG encoding of the LDR image only. In practice, the use of JPEG XT result on the increase of bit rate but with the advantage of being able to generate new LDR versions using different TMO's in the decoder side instead of using the predetermined TMO used on the encoder side.

To analyze this increase of bit rate, the point that is closer to the base layer (JPEG encoding) was computed. The base layer point is represented with $'+'$ in the Fig. 6.11. The process to compute the base layer closest point is described in the following.

After normalizing the bit rate and the FSIM metric values was used the Manhattan distance between the different working points generated for each Q and the base layer (the Euclidean distance gives very similar results, but in this case did not seem to be so stable). The resulting points are signalized with a circle for each profile in the examples of Fig. 6.11 and the complete set of plots for all images dataset is presented in Appendix D on Figs. D.1 and D.2. Using this criterion, was possible to compute the closer point in bit rate and FSIM metric for each case.

Furthermore, the bars plot of Figs. D.3 and D.4 show the results for the three studied profiles using all images and the *Reinhard* and *Mai* TMO's on the encoder, respectively. Considering all the tested results, the median of all Q values selected by the distance criterion was computed for each profile. An interesting fact is that this median is $Q = 75$ for the profiles A and B while results in $Q = 45$ for profile C. The final results for the median Q, mean bit rate, mean bit rate for the base layer, mean bit rate increase and bit rate increase standard variation for each profile are shown in Table 6.3. Moreover, the mean bit rate increase for each TMO used as a base layer is also shown. These values are shown considering each TMO (in the encoder side)

Table 6.3: Results for the median Q, mean bit rate, mean bit rate for the base layer, mean bit rate increase and bit rate increase standard variation for each profile.

	Profile		
	A	B	C
Median Q	75	75	45
Mean bit rate (bpp)	1.64	1.08	1.96
Mean bit rate base layer (bpp)	0.64	0.64	0.64
Mean bit rate increase (Δ bpp [%])	100.7	44.5	131.9
Δ bpp [%] (*Drago* TMO)	104.06	41.54	126.33
Δ bpp [%] (*Mantiuk* TMO)	103.88	33.23	159.30
Δ bpp [%] (*Reinhard* TMO)	99.74	32.90	130.97
Δ bpp [%] (*Reinhard & Devlin* TMO)	99.31	56.95	127.45
Δ bpp [%] (*Mai* TMO)	96.56	57.78	115.30

and for all cases (global).
From Table 6.3, profile B reveals the best case, with a global bit rate increase of about 44.5%, much smaller than the other two profiles, with a mean grow of 100.7% and 131.9% respectively for profiles A and C. Hence, can be concluded that for this kind of application the profile B requires the smallest bit rate increase.
In Figs. 6.12 to 6.17 are shown some examples of TMO's generation from the original HDR images and decoded images when *Reinhard* and *Mai* where used in JPEG XT base layer. The Figs. 6.12, 6.13 and 6.14 represent an example using respectively profiles A, B and C. The Figs. 6.15, 6.16 and 6.17 represent other examples using profile B as it produces lower mean bit rates for the selected Q value. In all cases, the JPEG XT encoder used an LDR version generated with the *Reinhard* and *Mai* TMO's for better comparison. As can be observed the differences between the images are hardly seen.

6.3.3 Comments

It is important to have the freedom to generate different LDR versions of the same HDR content, providing the best viewing conditions to the consumers of HDR content in the absence of HDR capable displays. However, this leads to an increase of bit rate that in the case of profile B would be around 44.5%. However, in some cases, profile B can not reach the level of quality that can be found in the other two profiles. In particular, profile C can provide a near lossless encoding while profile A in some cases leads to higher levels for the used objective metrics.
This study also helped to define a working point for the parameter Q that in most of the cases results in an acceptable quality.

6.4 Conclusion

In this chapter were described two studies to analyze the objective quality of JPEG XT performance for HDR coding.
A study of JPEG XT performance on HDR domain described in Section 6.2, and it tries to investigate the performance of the three main profiles, of the upcoming JPEG XT standard using a set of HDR images. Moreover, TMO techniques were used for the base layer image generation to assess the influence of the TMO's on the performance of JPEG XT profiles. The performance of each profile was evaluated using two objective metrics, and the results were studied and

(a) (b) (c)

Figure 6.12: Example of *Mantiuk* TMO generation for *LasVegasStore* image, from (a) original HDR, (b) decoded HDR (*Reinhard* encoded as JPEG XT base layer), and (c) decoded HDR (*Mai* encoded as JPEG XT base layer), with profile A.

(a) (b) (c)

Figure 6.13: Example of *Mantiuk* TMO generation for *LasVegasStore* image, from (a) original HDR, (b) decoded HDR (*Reinhard* encoded as JPEG XT base layer), and (c) decoded HDR (*Mai* encoded as JPEG XT base layer), with profile B.

(a) (b) (c)

Figure 6.14: Example of *Mantiuk* TMO generation for *LasVegasStore* image, from (a) original HDR, (b) decoded HDR (*Reinhard* encoded as JPEG XT base layer), and (c) decoded HDR (*Mai* encoded as JPEG XT base layer), with profile C.

compared. The evaluation results verify the performance of all profiles to be appropriate for compression of HDR images. The evaluation results demonstrate that profiles A and B lead to similar saturation effect for quality on higher bit rates, while profile C reveals a near lossless behavior for high bit rates. Profiles B and C reveal to be more dependent of the used TMO's on the base layer when compared with profile A. The profiles show similarly consistent behavior for the FSIM and SNR objective metrics.

Furthermore, in Section 6.3, a study on the influence of the JPEG XT encoding on TMO generation was described. This study also investigates the performance of the three main profiles of JPEG XT. LDR images generated using a specific TMO applied to the original HDR image are compared with the TMO generated using the decoded HDR image using perceptual metrics. This work is also important because nowadays HDR capable displays are not available for common consumers. HDR images are displayed using TMO's that convert the HDR images to LDR images making possible to be represented on conventional displays. However, different TMO's generate different LDR images, that are often visually quite different. Hence, it is important to have the freedom to generate different LDR versions of the same HDR content, providing the best viewing conditions to the consumers of HDR content in the absence of HDR capable displays.

(a) (b) (c)

Figure 6.15: Example of *Reinhard* TMO generation for *LasVegasStore* image, from (a) original HDR, (b) decoded HDR (*Reinhard* encoded as JPEG XT base layer), and (c) decoded HDR (*Mai* encoded as JPEG XT base layer), with profile B.

(a) (b) (c)

Figure 6.16: Example of *Mantiuk* TMO generation for *HancockKitchen* image, from (a) original HDR, (b) decoded HDR (*Reinhard* encoded as JPEG XT base layer), and (c) decoded HDR (*Mai* encoded as JPEG XT base layer), with profile B.

(a) (b) (c)

Figure 6.17: Example of *Drago* TMO generation for *set33* image, from (a) original HDR, (b) decoded HDR (*Reinhard* encoded as JPEG XT base layer), and (c) decoded HDR (*Mai* encoded as JPEG XT base layer), with profile B.

Moreover, was observed that HDR images encoded with JPEG XT can be efficiently used for LDR generation using different TMO's. Results reveal to be very similar to the LDR generated with the TMO applied to the original HDR image. Moreover, was also computed the increase of bit rate that represents the encoding with JPEG XT over JPEG, to provide reliable TMO generation on the decoder side. It is also interesting to realize that typical encoding artifacts did not have a strong influence on the TMO results.

Chapter 7

Benchmarking of Objective Quality Metrics for HDR Image Quality Assessment

As any technology, the evaluation of HDR systems in terms of quality of experience is essential. Subjective evaluations are time consuming and expensive, and thus objective quality assessment tools are needed as well. In this chapter, are reported and analyzed the results of an extensive benchmarking of objective quality metrics for HDR image quality assessment. In total, 35 objective metrics were benchmarked on a database of 20 HDR images with different content, encoded with three JPEG XT profiles at four bit rates, leading to a total of 240 compressed HDR images, using subjective quality scores as ground truth. Performance indexes were computed to assess the accuracy, monotonicity, and consistency of the subjective scores metrics estimation. Statistical analysis was also performed on the performance indexes to discriminate small differences between metrics.

7.1 Introduction

Few objective metrics have been developed so far for HDR content. The most relevant work in this domain is the HDR visual detection predictor (HDR-VDP) metric proposed by Mantiuk *et al* [186], which is an extension of Daly's VDP [187] for the HDR domain. The second version of this metric, HDR-VDP-2 [108, 109], is considered as the state-of-the-art in HDR image quality assessment. Recently, the high dynamic range video quality metric (HDR-VQM) was proposed by Narwaria *et al* [110]. The metric was designed for quality assessment of HDR video content, but can also be used for HDR still images.

To overcome the lack of HDR objective metrics, LDR metrics, e.g., PSNR, were also used to evaluate HDR quality, especially in early HDR studies. However, LDR metrics are designed for gamma encoded images, typically having luminance values in the range $[0.1, 100]\ \mathrm{cd/m^2}$, while HDR images have linear values and are meant to capture a much wider range of luminance. Originally, gamma encoding was developed to compensate for the characteristics of CRT displays, but it also takes advantage of the non-linearity in HVS to optimize quantization when encoding an image [188]. Under common illumination conditions, the HVS is more sensitive to relative differences between darker than brighter tones. According to Weber's law, the HVS sensitivity approximately follows a logarithm function for light luminance values [189]. Therefore, in several studies, LDR metrics have been computed in the log domain to predict HDR quality. However, at the darkest levels, the HVS sensitivity is closer to a square-root behavior, according to Rose-DeVries law [190, 191]. To extend the range of LDR metrics and to consider the sensitivity of the HVS, Aydin *et al* [192] have proposed the Perceptually Uniform (PU) encoding. For LDR content, extensive studies have shown that not all metrics can be considered as reliable predictors of perceived quality [193, 194], while only a few recent studies have benchmarked objective metrics for HDR quality assessment [195, 196, 11, 110]. Thus, the main limitation of these studies lies in the small number of images or video sequences used in their experiments, which was limited to five or six contents. Also, a proper adaptation of the contents to the HDR

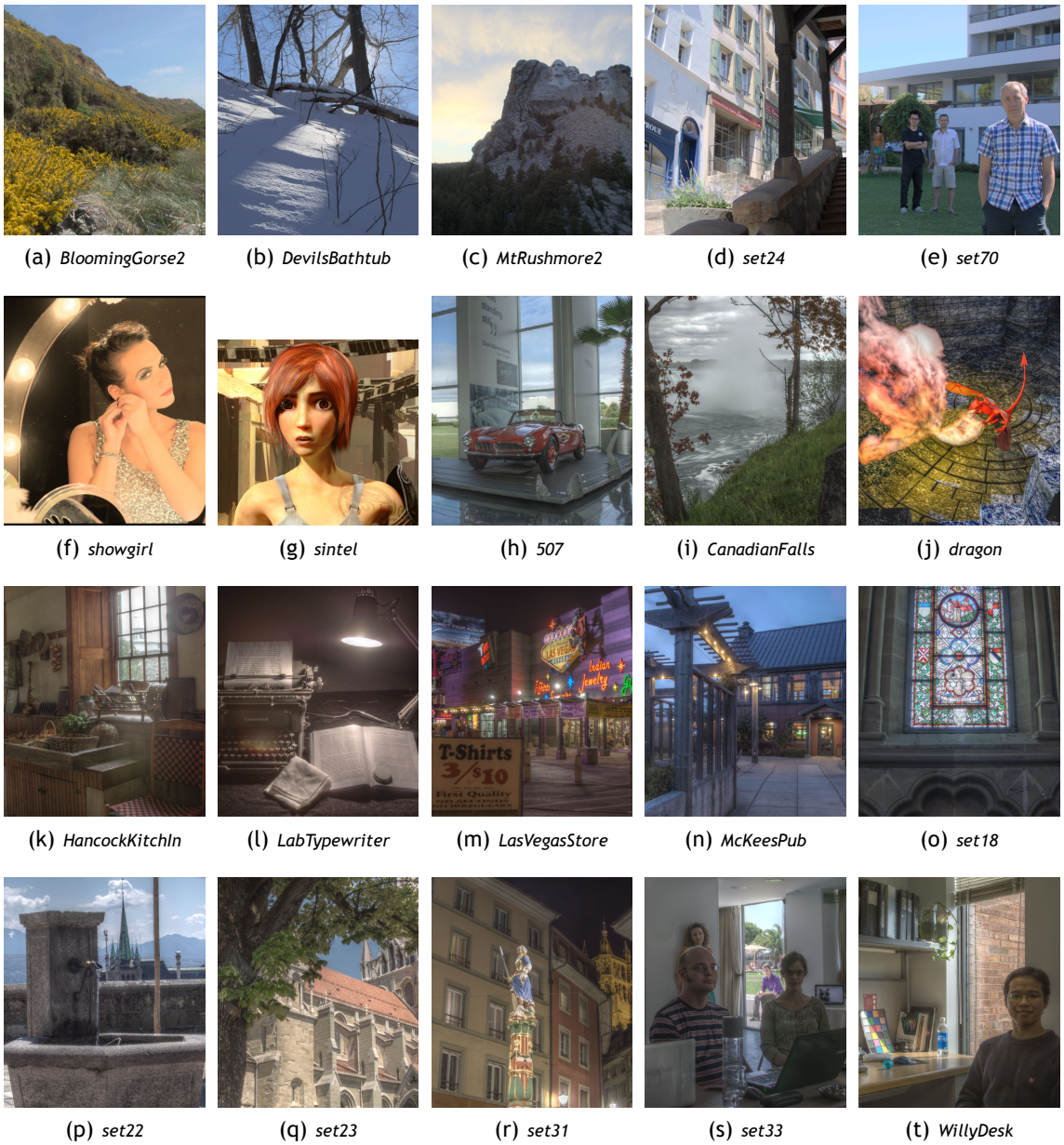
display and correction of the metrics for non-uniformity were not always considered.
Therefore, in this study, are reported and analyzed the results of an extensive benchmarking of IQMs for HDR image quality assessment. In total, 35 objective metrics were benchmarked using subjective scores as ground truth. The IQM's were computed in the linear, logarithmic, PU [192], and Dolby Perceptual Quantizer (PQ) [197] domains. Additionally, the metrics were computed both on the luminance channel alone and as the average quality score of the Y, C_b, and C_r channels. For each metric, objective scores were fitted to subjective scores using a logistic fitting. Performance indexes were computed to assess the accuracy, monotonicity, and consistency of the subjective scores metrics estimation.

7.2 Subjective Quality Assessment

The results of subjective tests can be used as ground truth to evaluate the level of perceived quality estimation using an IQM. In this chapter, were used the publicly available dataset provided by Korshunov *et al* [198, 199] to benchmark 35 objective metrics.
The test dataset is composed of 20 HDR images with a resolution of 944×1080 pixels and with different contents. Contains scenes of architecture, landscapes, and portraits and is composed of HDR images fused from multiple exposure pictures, frames extracted from HDR video, and computer generated images. Since publicly available HDR images are usually not graded, the images are adjusted for a SIM2 HDR monitor using a display-adaptive TMO [185] to map the relative radiance representation of the images to an absolute radiance and color space of the HDR display. These display-adapted images are then considered as original images and compressed with JPEG XT using profiles A, B, and C. For each content and profile, four different bit rates were selected, leading to a total of 240 compressed HDR images. Figure 7.1 shows tone mapped versions of the images in the dataset.
The DSIS methodology [72] was used for the evaluation and to show the stimuli to the subjects was used a full HD 47" SIM2 HDR display. The MOS was computed for each test stimulus as the mean score across the 22 valid subjects, as well as the associated 95% CI, assuming a Student's *t*-distribution of the scores. More details about the dataset and subjective evaluations can be found in [198].

7.3 Objective Quality Assessment

A set of 35 FR and NR metrics, were studied and their representation of the perceptual quality was analyzed. Moreover, they were compared with the MOS obtained in the subjective quality assessment experiment. The FR metrics are generally more reliable as they provide a comparison to the reference image, while the NR metrics might be very useful to estimate the image quality in the absence of an original image.
In this study the FR metrics can be divided into different categories (as mentioned in Section 3.2.1). The complete list of these metrics is provided in the following. In the category of difference measures and statistical-oriented metrics, were selected the MSE, the PSNR, and the SNR [87]. For the structural similarity measures category, were selected the UQI [90], the SSIM [86], the MS-SSIM [91], the M-SVD [92], and the QILV [93]. In the category of visual information measures, were selected the IFC [96], the VIF [97], the VIFp [97], and the FSIM [98]. For the information weighted metrics category, were chosen the IW-MSE [100], the



(a) *BloomingGorse2* (b) *DevilsBathtub* (c) *MtRushmore2* (d) *set24* (e) *set70*

(f) *showgirl* (g) *sintel* (h) *507* (i) *CanadianFalls* (j) *dragon*

(k) *HancockKitchIn* (l) *LabTypewriter* (m) *LasVegasStore* (n) *McKeesPub* (o) *set18*

(p) *set22* (q) *set23* (r) *set31* (s) *set33* (t) *WillyDesk*

Figure 7.1: Display-adapted images of the dataset [198, 199].

W-PSNR [100], and the IW-SSIM [100]. In the category of HVS inspired metrics, were selected the JND_st [101], the WSNR [102, 103], the DN [89], and metrics developed for HDR content, like the HDR-VDP2 [108, 109], and the HDR-VQM [110]. For the color difference measures category, were selected the CIE1976[32], the CIE94 [44], the CMC [111], and the CIEDE2000 [45].

Furthermore, the studied NR metrics are based on the analysis of a set of well-known sharpness measures. The following were considered, the JND [101], the VAR [114], the LAP [115], the GRAD [115], the FTM [116], the HPM [117], the Marziliano [118], the KurtosisZhang [119], the KurtosisWavelet [120], the AutoCorr [115], and the RTBM [121].

Moreover, the selected IQMs were computed in the linear, logarithmic, PU [192], and Dolby PQ [197] domains. Additionally, the metrics were computed both on the luminance channel alone and as the average quality score of the Y, C_b, and C_r channels.

7.3.1 Metrics Computation and Transform Domains

LDR metrics are designed for gamma encoded images, typically having luminance values in the range $[0.1, 100]\ \mathrm{cd/m^2}$, while HDR images have linear values and are meant to capture a much wider range of luminance. Therefore, in this study, metrics were computed not only in the linear space but also in transformed spaces that provide a more perceptual uniformity. This space transformation was not applied to HDR-VDP-2 and HDR-VQM, which are calibrated metrics and require absolute luminance values as input. The color difference metrics, i.e., CIE1976, CIE94, CMC, and CIEDE2000, were also not computed in transformed spaces. These color difference measures require a conversion from the RGB representation to the CIELAB color space, considering a D65 $100\ \mathrm{cd/m^2}$ reflective white point as reference white point.

Before any metric was computed, images were clipped to the range $[0.001, 4000]\ \mathrm{cd/m^2}$ that is the theoretical range of luminance values that the HDR monitor used in the subjective tests can render. Therefore, were mimicked the physical clipping performed by the HDR display. To compute the metrics in the linear domain, these luminance values were normalized to the interval $[0, 1]$. For the reasons explained before, this normalization was not applied to HDR metrics and to color difference metrics.

In addition to the linear domain, the remaining metrics were computed in three transform domains: the log domain, the PU domain [192], and the PQ domain [197]. The PU transform is derived using the threshold-integration method [200]. The transform is constrained such that luminance values are transformed to the range $[0.1, 80]\ \mathrm{cd/m^2}$, as produced by a typical CRT display, and mapped to the range $[0, 255]$ to mimic the RGB non-linearity. The PQ transform is derived from the Barten contrast sensitivity function [201]. The PQ curve has a square-root and log behavior at the darkest and highest light levels, respectively, while it exhibits a slope similar to the gamma non-linearity between those extreme luminance regions.

These transformations were applied before any normalization and only after their application the resulting color components were normalized to the interval $[0, 1]$. After the normalizations, the values considered to be in the RGB color space were transformed to the YC_bC_r color space [202]. The exception is the DN metric, which uses directly these RGB components. The metrics were computed on each of these components separately and two final metrics were considered: the quality score computed on the luminance channel alone (*_Y* suffix) and the average quality score of the Y, C_b, and C_r channels (*_M* suffix).

7.3.2 Evaluation of the Objective Models

In order to evaluate the objective models, all of these IQMs were compared with the MOS available for the studied test images, according to the procedure mentioned in Section 3.2.2. Thus, the ability to provide a representation of the QoE was assessed, and performance indexes to assess the accuracy of objective metrics were computed following the procedure explained in Section 3.2.2.

Furthermore, to determine whether the difference between two performance index values corresponding to two different metrics is statistically significant, two-sample statistical tests were performed on all four performance indexes, according to the guidelines of recommendation ITU-T P.1401 [85]. In particular, for the PCC and SROCC, a *Z*-test was performed using Fisher z-transformation. For the RMSE, a *F*-test was performed, whereas a *Z*-test for the equality of two proportions was performed for the OR.

7.4 Results Analysis

The accuracy, monotonicity, and consistency indexes for the IQMs computed in the different domains, as defined in the previous section, are reported in Appendix C Tables C.1-C.4. The metrics are sorted from best (top) to least (bottom) performing, based on the different performance indexes (higher PCC/SROCC and lower RMSE/OR values indicate better performance). The statistical analysis results are reported in the same tables. This analysis was performed on the performance indexes computed from 240 data points to discriminate small differences between two metrics. Metrics whose performance indexes are connected by a line are considered statistically not significantly different. For example, in the linear domain, according to PCC, there is no statistical evidence to show performance differences between IFC and FSIM computed on the luminance channel, but they are statistically different from HDR-VDP-2 (see Appendix C Table C.1).

7.4.1 Best Performing Metrics

As expected, HDR-VDP-2 and HDR-VQM, which are the only true HDR quality metrics considered in this study, computed on absolute luminance values, are the best performing metrics when compared to all other metrics and domains. Both metrics have a correlation above 0.95 and a particularly low RMSE (around 0.35) and low OR, whereas all other metrics have an OR above 0.48. HDR-VDP-2 (OR $= 0.35$) has a slightly lower OR than HDR-VQM (OR $= 0.4083$), but there is no statistical evidence to show a significant difference.
The results for HDR-VDP-2 are in line with the finding of [11], slightly better than that of Valenzise *et al* [195], but in contradiction with Mantel *et al* [196], who reported a much lower correlation. However, Mantel *et al* used unusual combinations of parameters for the base and extension layers, especially for content *BloomingGorse*. Narwaria *et al* [110] found that HDR-VQM was performing significantly better than HDR-VDP-2 for both, video and still image content. However, our results show that both metrics have similar performance, while it was reported in [203] that HDR-VQM performs lower than HDR-VDP-2 for HDR video compression. The divergence between these findings might be due to the contents and types of artifacts considered in the different studies.
In contrast to the HDR metrics, the NR metrics show the worst performance with PCC and SROCC values below 0.5 and RMSE and OR values above 1 and 0.8, respectively, independently of the domain in which the metric was computed. These results show that NR metrics do not reach satisfactory prediction accuracy considering a perceptual domain and that specific NR metrics should be designed for HDR image quality assessment.

7.4.2 Difference Measures and Statistical-Oriented Metrics

Results show that MSE-based metrics, i.e., MSE, SNR, and PSNR, are not very reliable predictors of perceived quality when computed in the linear domain, with a correlation between 0.65 and 0.75. Higher PCC values were reported in [11] for MSE and SNR (PCC $= 0.88$), but the study was performed considering only five contents. These metrics are known to be very content dependent [204], which might explain the drop in performance when considering 20 images. The correlation of MSE-based metrics computed on the luminance channel alone can be improved by about 0.1 considering a more perceptual domain than the linear domain, which does not take into account the contrast sensitivity response of the HVS. In the log and PU domains, the correlation

is about 0.83 and 0.84, respectively, which is in line with the results from [195]. Nevertheless, the performance of the MSE-based metrics computed as the average quality score of the Y, C_b, and C_r channels did not improve when considering perceptual domains. These observations indicate that the log, PU, and PQ domains provide a better representation of the HVS luminance sensitivity than the linear domain, but they might not be optimal for chrominance sensitivity.

7.4.3 Color Difference Measures

In the linear domain, the color difference metrics, with the exception of the original CIE1976 color difference metric, are the best performing pixel-based metrics. They outperform the MSE-based metrics, but there is no statistical evidence to show a significant improvement over SNR computed on the luminance alone. Nevertheless, their correlation with perceived visual quality is only about 80%, with an OR above 69%, which cannot be considered as a reliable prediction. Since the release of the CIE1976 color difference metric, two extensions have been developed in 1994 and 2000 to better address perceptual non-uniformities of the HVS. But, according to the benchmarking results, further improvements might be necessary for HDR images to handle non-uniformities in low and high luminance ranges, outside of the typical range of LDR displays. The color difference metrics are computed in the CIELAB color space, which considers relative luminance values with respect to a reference white point, typically a reflective D65 white point with about 100-$120\ \mathrm{cd/m^2}$. This reference white point is similar to the targeted peak luminance that is typically considered when calibrating LDR reference monitors. Therefore, for HDR images, one would be tempted to set the luminance of the reference white point considered in the color conversion equal to the peak luminance of the HDR monitor. However, this leads to a lower performance of the color difference metrics and the reflective white point should also be used for HDR content instead.

7.4.4 Structural Similarity and Visual Information Measures

The performance of SSIM and its multi-scale extension, MS-SSIM, is improved by considering logarithm instead of linear values, and are even further improved by considering the PU or PQ transform. In particular, on the luminance channel, the correlation of SSIM increased about 0.15 from linear to logarithm, while MS-SSIM improved only about 0.03. From log to PU/PQ, improvements are relatively low for SSIM, whereas MS-SSIM exhibits a gain of about 0.04. Results show that MS-SSIM (luminance only) performs the best in PU and PQ spaces according to the PCC, SROCC, and RMSE indexes. The correlation obtained for SSIM in the log and PU domains is similar to the results of Valenzise *et al* [195]. On the other hand, UQI, which corresponds to the special case of SSIM when the constants C_1 and C_2 are set to 0, does not perform better in the log, PU, or PQ space than in the linear domain. Similar correlation results for SSIM and MS-SSIM are reported in [11] as in this study (for the linear domain). However, it is reported that the relative change between the worst and best qualities for SSIM and MS-SSIM was less than 0.003% and 0.0003%, respectively. In this study, the average relative change computed over all domains is 16.5% and 11.5% for SSIM and MS-SSIM, respectively. One major difference between the two works is the use of absolute luminance values in [11], whereas luminance values were linearly mapped from the theoretical display range to the range $[0, 1]$ in this study. For LDR content, SSIM uses different values for C_1 and C_2 depending on whether the images are in the range $[0, 1]$ or $[0, 255]$. For HDR content, our findings suggest that the value of these constants should be adjusted according to the luminance range and depending on whether scaling of the values is

Table 7.1: Statistical analysis comparing the HDR metrics and best performing metric of each domain.

PCC		SROCC		RMSE		OR	
HDRVDP2	0.9604	HDRVDP2	0.9564	HDRVDP2	0.3498	HDRVDP2	0.3500
HDRVQM	0.9602	HDRVQM	0.9564	HDRVQM	0.3506	HDRVQM	0.4083
PU2MSSSIM_Y	0.9447	PU2MSSSIM_Y	0.9501	PU2MSSSIM_Y	0.4132	LOG_VIFP_Y	0.4833
PQ2MSSSIM_Y	0.9380	PQ2MSSSIM_Y	0.9435	PQ2MSSSIM_Y	0.4366	PU2MSSSIM_Y	0.5417
LOG_VIFP_Y	0.9230	LOG_VIFP_Y	0.9200	LOG_VIFP_Y	0.4832	PQ2MSSSIM_Y	0.5542
LOG_PSNR_Y	0.8348	LOG_PSNR_Y	0.8399	LOG_PSNR_Y	0.6911	LOG_PSNR_Y	0.6208

performed or not.

Metrics that quantify the loss of image information, i.e., VIF, its pixel-based version, VIFP, and its predecessor, IFC, also show good performance. In particular, IFC (luminance only) is the second best performing metric in the linear domain. While the domain in which the metric is computed does not influence IFC performance, the performance of VIFP is significantly improved when considering a more perceptual domain than the linear space. In the log domain, results show that VIF computed on the luminance alone is the best performing metric. Note that the correlation reported for VIFP in this study is significantly better than the one reported in [11]. Similarly to MS-SSIM, the difference might be due to the scaling procedure. Among the other HVS-based metrics, FSIM also shows good performance, especially in the PU and PQ space (RMSE below 0.5).

7.4.5 Statistical Analyses

To determine how the best metrics of each domain compare to each others, a direct benchmarking of the two HDR metrics, which are the best performing metrics in the linear space, and the best performing metric of the log, PU, and PQ spaces was performed. The PSNR metric computed on the luminance channel in the log space was added to this comparison, as this metric is widely used in HDR compression studies. Table 7.1 reports the results of the statistical analysis of the six metrics. To identify metrics computed in the log, PU, and PQ spaces, the *LOG_*, *PU2*, and *PQ2* prefixes are used, respectively. According to PCC and SROCC, there is no statistical evidence to show performance differences between HDR-VDP-2, HDR-VQM, and MS-SSIM computed on the luminance channel in the PU space. However, HDR-VDP-2 and HDR-VQM have a significantly lower RMSE than all other metrics. Figure 7.2 depicts the scatter plots of subjective versus objective results for these metrics. As it can be observed, the data points are well concentrated near the fitting curve for HDR-VDP-2, as well as for HDR-VQM, while they are more scattered for the other metrics, especially in the case of LOG_PSNR_Y, which shows higher content dependency. These findings indicate that HDR-VDP-2 and HDR-VQM have a very high consistency when compared to the other metrics. Nevertheless, HDR-VDP-2 is complex and requires heavy computational resources, which limits its use in many applications. HDR-VQM and MS-SSIM computed in the PU space are lower complexity alternatives to HDR-VDP-2.

The statistical analysis was also used to understand whether there is a statistically significant difference between the performance of each metric when computed on the luminance component alone and when computed on all components. Only results from the analysis performed on the 28 metrics that were computed both on the Y channel alone and as the average quality score of the Y, C_b, and C_r channels were considered. Table 7.2 reports the number of metrics for which one approach was significantly better than the other one, as well as when no significant

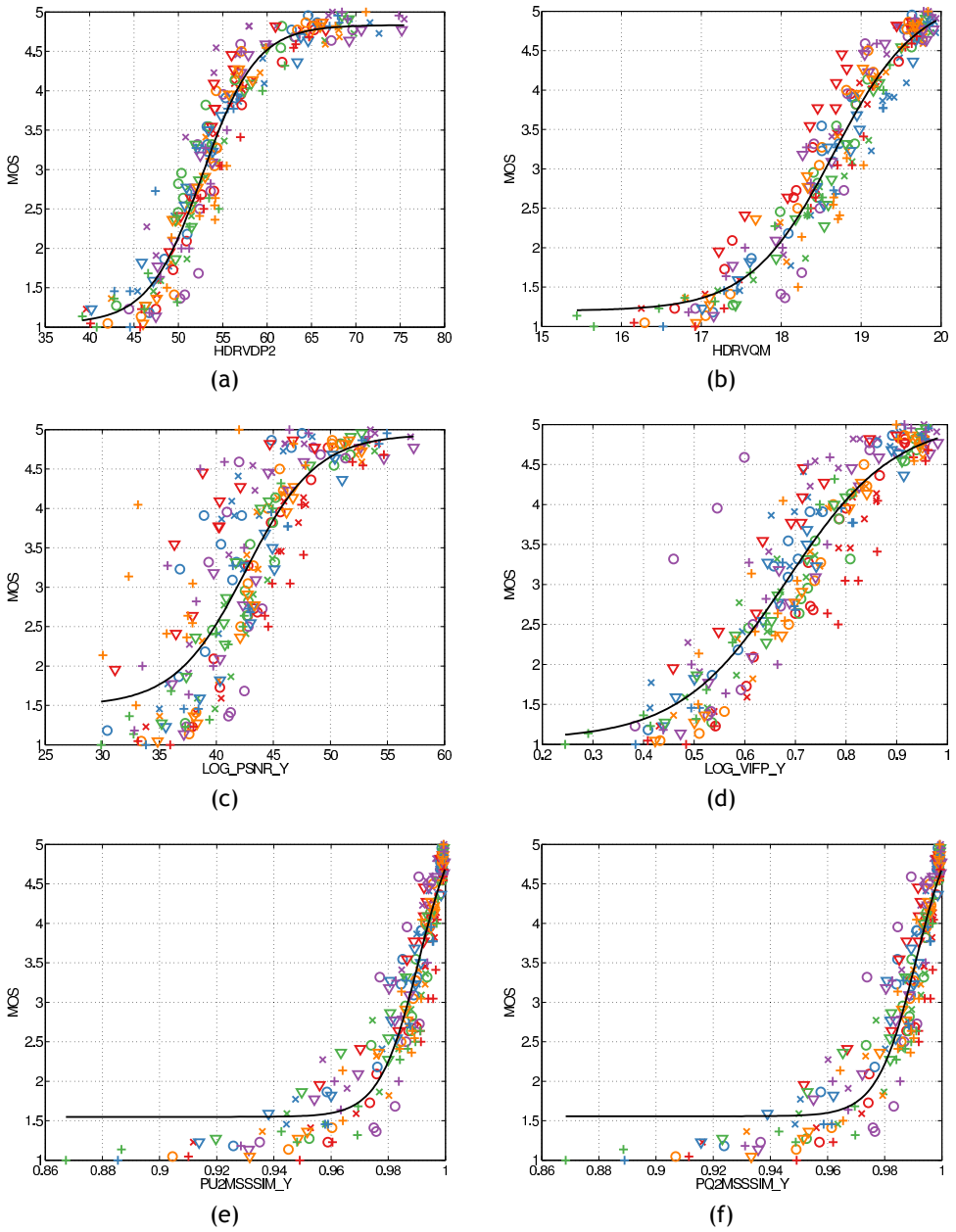


Figure 7.2: Subjective versus objective evaluations results for the HDR metrics and best performing metric of each domain. Each symbol, i.e., combination of marker and color, corresponds to a specific content.

difference between the two approaches was observed. The analysis was performed individually for each performance index and domain. In the linear domain, there is no statistical evidence to show performance differences between the two approaches for about 80% of the metrics. However, in the log, PU, and PQ space, roughly half of the metrics perform significantly better when computed on the luminance channel alone. According to PCC, the JND metric, FR version, computed in the log domain, is the only case for which better performance is achieved when considering all channels. As HDR is often considered in combination with Wide Color Gamut (WCG), it is expected that the color reproduction fidelity will play a more important role in the context of HDR when compared to LDR. It is possible that improvements can be achieved by considering different domains for computing the metrics on the chrominance channels and by using better

Table 7.2: Comparison of the 28 objective quality metrics computed on the Y and YC_bC_r channels. Comparison of the metrics computed as the average quality score of the Y channel alone and as the average quality score of the YC_bC_r channels.

	lin				log				PU				PQ			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Y is better	6	7	8	3	16	14	16	8	16	14	15	14	14	14	15	14
similar	22	21	20	25	11	14	12	20	12	14	13	14	14	14	13	14
YC_bC_r is better	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Table 7.3: Comparison of the 57 objective quality metrics computed on all domains. Results represent the number of times a metric computed in the domain i performs significantly better than when computed in the domain j, where i and j are the row and column of the table.

	PCC				SROCC				RMSE				OR			
	lin	log	PU	PQ	lin	log	PU	PQ	lin	log	PU	PQ	lin	log	PU	PQ
lin	0	9	5	6	0	9	6	8	0	8	5	5	0	2	1	1
log	10	0	2	2	10	0	4	4	8	0	0	0	8	0	3	3
PU	14	17	0	11	13	15	0	11	11	16	0	10	9	7	0	7
PQ	13	17	8	0	11	15	8	0	11	16	6	0	9	7	3	0

pooling strategies.

Similarly, the statistical analysis was also used to understand whether there is a statistically significant difference between the performance of a particular metric computed in one domain and another domain. Only results from the analysis performed on the 57 metrics that were computed in all domains were considered. Table 7.3 reports the number of times a metric computed in the domain i performs significantly better than when computed in the domain j, where i and j are the row and column of the table. Results show that most metrics perform the best in the PU and PQ spaces when compared to the linear and log spaces, which is in line with our previous observations. Note that results based on PCC, SROCC, and RMSE are in agreement, while the OR metric shows fewer cases where statistically significant difference are observed. Additionally, there are also metrics for which computations performed in the linear and logarithm domains perform better than in the PU and PQ space. Overall, there is no optimal domain that performs the best for all metrics. Instead, different metrics should use different domains to maximize the correlation with perceived quality.

7.5 Conclusion

In this study, 35 objective metrics were benchmarked on a database of 240 compressed HDR images using subjective quality scores as ground truth. Additionally to the linear space, metrics were computed in the logarithm, PU, and PQ domains to mimic non-linearities of the HVS. Results showed that the performance of most FR metrics could be improved by considering perceptual transforms when compared to linear values. On the other hand, findings suggested that a lot of work remains to be done for NR quality assessment of HDR content. This study benchmark demonstrated that HDR-VDP-2 and HDR-VQM are ultimately the most reliable predictors of perceived quality. HDR-VQM is over three times faster, which makes it a suitable alternative to HDR-VDP-2. Alternatively, MS-SSIM computed in the PU space is another lower complexity substitute, as there is no statistical evidence to show performance differences between these metrics in terms of PLCC and SROCC. Even though the numbers of contents and compressed images considered in the experiments are quite large, different performance might be observed

for other contents and types of artifacts.

Chapter 8

Conclusions

8.1 Final Comments

This thesis aims to provide contributions to the research and development on multimedia quality assessment, focusing on the quality perception of chromatic changes in digital images. It aims at improving user immersion capability of the new trends in imaging technology, providing more realistic perception of content, and consequently reaching new levels of QoE.

The first topic of the thesis deals with the perception and modeling of chromatic impairments induced on color images. A new method that induces chromatic changes on images without creating visible artifacts was described. The images transformed using this method were evaluated on a subjective test, and it was possible to assess the HVS perception to this type of distortions. Moreover, the previous study is complemented with the application of IQM methods to evaluate the provided representation when dealing with the chromatic distortions. Furthermore, were proposed quality estimators based on the CIEDE2000, which provided the best representation. A complementary study is also reported aiming to extend it by defining specific chromatic impairments is also reported. It was observed the influence of the content and type of the impairments on the perception of quality.
In both studies could be observed that subjects revealed larger perception to a loss of a color naturalness than to the error magnitude itself. It was also verified that in presence of images with nature colors (rural scenes for instance) subjects become more sensitive to chromatic errors than when these colors are non-natural (as it is typical in urban scenes). Moreover, it can be noted that the usage of the eye movements data might lead to new subjective quality models and might provide a good indicator of the perceived quality, since the RFT variations revealed some correlation with the subjective tests.

The second topic addresses HDR image quality encoded with JPEG XT. Two studies on objective quality analysis of JPEG XT performance on HDR were also reported. The first study initially analyzes the quality provided by the HDR encoding, relative to the parameterization and profile. Then, it is followed by a quality analysis of TMO versions extracted from JPEG XT decoded HDR images. This study allowed a verification of the JPEG XT standard. The three profiles revealed to be appropriate for HDR encoding.
Specifically, profile A exhibits less dependency on the TMO's used for base layer; Profile B shown faster saturation for higher bit rates. Profile C, while demonstrating no saturation, is able to encode images with high bit rates, whereas performs worse on low bit rates when compared to profiles A and B. Moreover, it was verified that HDR images encoded with JPEG XT can be efficiently used for LDR generation using different TMO's.
The second study analyses a benchmarking of objective quality metrics applied to HDR image encoding using JPEG XT.
Both HDR-VDP-2 and HDR-VQM objective metrics revealed to be the most reliable predictors of perceived quality on HDR domain. Moreover, the performance of most FR objective metrics

developed for LDR domain, also revealed a good performance, when perceptual transforms are used. Hence, these metrics are a reliable alternative to measure the decoded HDR images quality.

8.2 Future Work

This thesis focus on a small part of the broad domain of quality assessment. Therefore this thesis allows multiple research directions.
The domains that might be considered as added value for future research directions may include,

- A larger study on the development of the CIEDE2000 color difference applied to the standard databases; As the PSNR, the CIEDE2000 color difference is usually applied as the mean of pixel based differences. Hence, it does not consider the complexity of inter-spatial relations that exists on the perceptual domain. A study on the use of this measure, modeling the spatial factors present in the HVS perceptual domain, should be considered for objective quality evaluation.
- The validation of the results presented in Section 6.3 on TMO generation and bit rate increase with an appropriate subjective quality assessment. This is a difficult issue, because of the small differences in image quality, and the multiple parameters and profiles involved. Eventually, crowdsourcing platforms and flickering tests could be considered.
- To develop a conversion model from hyperspectral images to the HDR domain, providing an accurate color information representation that is contained on this type of images.
- To explore the high correlation between eye fixation data and subjective quality with an extensive perceptual study, eventually leading to new quality models.

List of Publications

Articles included in the thesis resulting from these doctoral research program

- António M. G. Pinheiro, Marco V. Bernardo, and Manuela Pereira: **LDR images generation with JPEG-XT decoded HDR images**. – *in 2016 IEEE International Conference on Image Processing (ICIP)*, pages 1399–1403, Phoenix, Arizona, USA, Sept 2016.
- Marco V. Bernardo, António M. G. Pinheiro, Paulo T. Fiadeiro, and Manuela Pereira: **Quality Perception of Specific Chromatic Errors**. (manuscript submitted for consideration) 2016
- Marco V. Bernardo, António M. G. Pinheiro, Paulo T. Fiadeiro, and Manuela Pereira: **Image Quality Under Chromatic Impairments**. – *ACM Transactions on Applied Perception*, 14(1):6:1–6:20, August 2016
- Philippe Hanhart, Marco V. Bernardo, Manuela Pereira, António M. G. Pinheiro, and Touradj Ebrahimi: **Benchmarking of objective quality metrics for HDR image quality assessment**. – *EURASIP Journal on Image and Video Processing*, 2015(1):1–18, December 2015
- M. V. Bernardo, A. M. G. Pinheiro, P. T. Fiadeiro, and M. Pereira: **Eye gaze behavior under chromatic impairments and quality assessment**. – *in Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, pages 1–6, May 2015.
- Philippe Hanhart, Marco V. Bernardo, Pavel Korshunov, Manuela Pereira, Antonio M. G. Pinheiro, and Touradj Ebrahimi: **HDR image compression: a new challenge for objective quality metrics**. – *in Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, September 2014.
- A. Pinheiro, K. Fliegel, P. Korshunov, L. Krasula, Marco V. Bernardo, M. Pereira, and T. Ebrahimi: **Performance evaluation of the emerging JPEG XT image compression standard**. – *in 16th International Workshop on Multimedia Signal Processing (MMSP)*, Jakarta, Indonesia, September 2014.
- Marco V. Bernardo, António M. G. Pinheiro, Paulo Torrão Fiadeiro, and Manuela Pereira: **Quality assessment of chromatic variations: A study of full-reference and no-reference metrics**. – *in European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September 2014.
- Marco V. Bernardo, A.M.G. Pinheiro, M. Pereira, and P.T. Fiadeiro: **Specific chromatic errors: A quality assessment**. – *in Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pages 1–5, July 2013.
- Marco V. Bernardo, António M. G. Pinheiro, Manuela Pereira, and Paulo Torrão Fiadeiro: **Objective evaluation of chromatic quality assessment**. – *in Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6, July 2013.
- A. M. G. Pinheiro, M. V. Bernardo, M. Pereira, and P. T. Fiadeiro: **Influence of chromatic information on QoE**. – *in Colloquium on Quality of Experience in Multimedia Systems and Services - COST Action IC1003 QUALINET*, Alpen-Adria-Universität Klagenfurt, Austria, November 2012.

- M. V. Bernardo, A. M. G. Pinheiro, M. Pereira, and P. T. Fiadeiro: **A study on the user perception to color variations**. – *in Proceedings of the 20th ACM international conference on Multimedia*, Nara, Japan, 2012

Other publications resulting from this doctoral program not included in the thesis

- Rafael Rodrigues, Peter Pocta, Hugh Melvin, Marco V. Bernardo, Manuela Pereira and António M. G. Pinheiro: **Audiovisual Quality of Live Music Streaming over Mobile Networks using MPEG-DASH**. (manuscript submitted for consideration) 2016

- Pedro Fernandes, Marco V. Bernardo, António M. G. Pinheiro, Paulo T. Fiadeiro, and Manuela Pereira: **Quality comparison of the HEVC and VP9 encoders performance**. – *Multimedia Tools and Applications*, pages 1–17, 2016.

- M. Fidalgo-Fernandes, M. V. Bernardo, and A. M. G. Pinheiro: **A bag of words description scheme based on SSIM for image quality assessment**. – *in 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, June 2016.

- M. Bernardo, T. Bruylants, T. Ebrahimi, K. Fliegel, P. Hanhart, L. Krasula, A. Pinheiro, M. Rerabek, P. Schelkens, and H. Xu: **Objective and subjective evaluations of some recent image compression algorithm**. – *in Evaluation of Current and Future Image Compression Technologies I - 31st Picture Coding Symposium (PCS)- COST Action IC1003 QUALINET*, Cairns, Australia, June 2015.

- Pavel Korshunov, Marco V. Bernardo, Antonio M.G. Pinheiro, and Touradj Ebrahimi: **Impact of tone-mapping algorithms on subjective and objective face recognition in HDR images**. – *in Proceedings of the Fourth International Workshop on Crowdsourcing for Multimedia, CrowdMM '15*, pages 39–44, New York, NY, USA, 2015. ACM.

- Gil Santos, Emanuel Grancho, Marco V. Bernardo, and Paulo T. Fiadeiro: **Fusing iris and periocular information for cross-sensor recognition**. – *Pattern Recognition Letters*, 57:52 – 59, 2015.

- Daniel Piedade, Marco V. Bernardo, António M. G. Pinheiro, Paulo Torrão Fiadeiro, and Manuela Pereira: **Chromatic variations on 3D video and QoE**. – *in European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September 2014.

- Gil Santos, Marco V. Bernardo, Hugo Proença, and Paulo T. Fiadeiro: **Iris recognition: Preliminary assessment about the discriminating capacity of visible wavelength data**. – *in Multimedia (ISM), 2010 IEEE International Symposium on*, pages 324–329, December 2010.

Bibliography

[1] M. V. Bernardo, A. M. G. Pinheiro, M. Pereira, and P. T. Fiadeiro, "A study on the user perception to color variations," in *Proceedings of the 20th ACM international conference on Multimedia*, Nara, Japan, 2012, pp. 1009–1012. xii, 2, 43, 51

[2] M. V. Bernardo, A. M. G. Pinheiro, P. T. Fiadeiro, and M. Pereira, "Objective evaluation of chromatic quality assessment," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, July 2013, pp. 1–6. xii, 2

[3] A. M. G. Pinheiro, M. V. Bernardo, M. Pereira, and P. T. Fiadeiro, "Influence of chromatic information on QoE," Colloquium on Quality of Experience in Multimedia Systems and Services - COST Action IC1003 QUALINET, Alpen-Adria-Universität Klagenfurt, Austria, November 2012. xii, 2

[4] M. V. Bernardo, A. M. G. Pinheiro, P. T. Fiadeiro, and M. Pereira, "Quality assessment of chromatic variations: A study of full-reference and no-reference metrics," in *European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September 2014. xii, 2

[5] ——, "Image quality under chromatic impairments," *ACM Transactions on Applied Perception*, vol. 14, no. 1, pp. 6:1–6:20, Aug. 2016. xii, 2, 51, 52, 55

[6] ——, "Specific chromatic errors: A quality assessment," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, July 2013, pp. 1–5. xii, 2, 43, 51, 55, 65

[7] ——, "Eye gaze behavior under chromatic impairments and quality assessment," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, May 2015, pp. 1–6. xii, 2, 51, 56

[8] ——, "Quality perception of specific chromatic errors," 2016, (manuscript submitted for publication). xii, 2

[9] A. Pinheiro, K. Fliegel, P. Korshunov, L. Krasula, M. V. Bernardo, M. Pereira, and T. Ebrahimi, "Performance evaluation of the emerging JPEG XT image compression standard," in *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*. IEEE, 2014, pp. 1–6. xii, 2, 72

[10] A. M. G. Pinheiro, M. V. Bernardo, and M. Pereira, "LDR images generation with JPEG XT decoded HDR images," in *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, Arizona, USA, Sept 2016, pp. 1399–1403. xii, 3

[11] P. Hanhart, M. Bernardo, P. Korshunov, M. Pereira, A. Pinheiro, and T. Ebrahimi, "HDR image compression: A new challenge for objective quality metrics," in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*, Sept 2014. xiii, 3, 72, 87, 91, 92, 93

[12] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for hdr image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–18, 2015. xiii, 3, 72

[13] M. Bernardo, T. Bruylants, T. Ebrahimi, K. Fliegel, P. Hanhart, L. Krasula, A. Pinheiro, M. Rerabek, P. Schelkens, and H. Xu, "Objective and subjective evaluations of some recent image compression algorithm," Evaluation of Current and Future Image Compression Technologies I - 31st Picture Coding Symposium (PCS)- COST Action IC1003 QUALINET, Cairns, Australia, June 2015. 2

[14] K. Nassau, *The Physics and Chemistry of Color: The Fifteen Causes of Color*, 2nd ed. New York, NY: Jonh Wiley & Sons, Inc., 2001. 7

[15] W. R. McCluney, *Introduction to Radiometry and Photometry*. Boston, London: Artech House, Inc., 1994. 7

[16] G. Wyszecki and W. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., ser. Wiley Series in Pure and Applied Optics. John Wiley & Sons, 2000. 8, 10, 11, 13, 14, 15, 16, 39

[17] J. Morovic, *Color Gamut Mapping*. Jonh Wiley & Sons, Inc., 2008, ch. Basics of Color Science. 8

[18] L. R. Barone, *Anatomy and Physiology of the Human Body*. Buenos Aires,Argentina: CLASA, 2004. 8

[19] M. D. Fairchild, *Color Appearance Models*, 2nd ed. Jonh Wiley & Sons, Inc., 2005. 9, 12, 13, 14, 18, 19

[20] E. N. Marieb, *Human Anatomy & Physiology*, 7th ed. Benjamin-Cummings Publishing Company, 2007, ch. 15. 9

[21] N. R. Carlson, *Physiology of Behavior*, 9th ed. Boston, Massachusetts: Allyn and Bacon, 2006. 9

[22] R. Sekuler and R. Blake, *Perception*, 2nd ed. New York, NY: McGraw-Hill Publishing Company, 1990. 9

[23] B. A. Wandell, *Foundations of Vision*, 1st ed. Sunderland, MA: Sinauer Associates, Inc., 1995. 9

[24] A. Stockman and L. T. Sharpe, "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype," *Vision Research*, vol. 40, no. 13, pp. 1711 – 1737, 2000. 10

[25] T. Young, "The bakerian lecture: On the theory of light and colours," *Philosophical Transactions of the Royal Society of London*, vol. 92, pp. 12–49, 1802. 10

[26] H. L. Helmholtz, "Über die zusammensetzung von spectralfarben," *Annalen der Physik*, vol. 94, pp. 1–28, 1855. 10

[27] J. C. Maxwell, "Experiments on colour as perceived by the eye, with remarks on colour blindness," *Transactions Of The Royal Society Of Edinburgh*, vol. 21, pp. 275–298, 1855. 11

[28] R. Sève, *Physique de la couleur: de l'apparence colorée à la technique colorimétrique*, 1st ed. Masson, 1996. 11

[29] P. Kowaliski, *Vision et mesure de la couleur*, 2nd ed. Masson et CIE, 1990. 11

[30] H. G. Grassmann, "Zur theorie der farbenmischung," *Annalen der Physik (Leipzig)*, vol. 89, pp. 60–84, 1853. 11

[31] I. Newton, *Opticks: Or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*. New York, N.Y.: Dover Publications, Inc., 1952. 12

[32] CIE, "Colorimetry official recommendation of the international commission on illumination," CIE Central Bureau, CIE Publication 15.2, 1986. 12, 14, 19, 32, 39, 40, 45, 89

[33] R. W. G. Hunt, *Measuring Color*, 2nd ed. New York: Ellis Horwood, 1991. 13, 18

[34] K. Nassau, "The causes of color," *Scientific American*, vol. 243, pp. 124–154, 1995. 13

[35] W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colours," *Transactions of the Optical Society*, vol. 30, no. 4, p. 141, 1929. 15

[36] ——, "A re-determination of the mixture curves of the spectrum," *Transactions of the Optical Society*, vol. 31, no. 4, p. 201, 1930. 15

[37] J. Guild, "The colorimetric properties of the spectrum," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 230, pp. 149–187, 1932. 15

[38] T. Smith and J. Guild, "The C.I.E. colorimetric standards and their use," *Transactions of the Optical Society*, vol. 33, no. 3, pp. 73–134, 1931. 16, 39

[39] H. S. Fairman, M. H. Brill, and H. Hemmendinger, "How the cie 1931 color-matching functions were derived from wright-guild data," *Color Research & Application*, vol. 22, no. 1, pp. 11–23, 1997. 16

[40] R. S. Berns, *Billmeyer and Saltzman's Principles of Color Technology*, 3rd ed. New York, United States: Jonh Wiley & Sons, Inc., 2000. 17

[41] I. C. on Illumination, *Recommendations on Uniform Color Spaces, Color-difference Equations, Psychometric Color Terms*, ser. CIE publication. CIE, 1978. 18

[42] M. Bicchierini, A. Davalli, R. Sacchetti, and S. Paganelli, "Colorimetric analysis of silicone cosmetic prostheses for upper-limb amputees," *Journal of Rehabilitation Research & Development*, vol. 42, no. 5, pp. 655–664, 2005. 19

[43] M. K., "Development of CIE 1976 (Lab) uniform color space and color-difference formula," *Journal of the Society of Dyers and Colourists*, vol. 92, no. 9, pp. 338–341, 1976. 20

[44] CIE, "Industrial colour-difference evaluation," CIE Central Bureau, CIE Publication 116, 1995. 20, 32, 89

[45] M. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Research & Application*, vol. 26, no. 5, pp. 340–350, 2001. 20, 32, 38, 45, 55, 89

[46] E. J. Giorgianni and T. E. Madden, *Digital Color Management: Encoding Solutions*, 2nd ed. Jonh Wiley & Sons, Inc., 2008. 21, 22

[47] A. Poynton, *A technical introduction to digital video*. New York, NY, USA: John Wiley & Sons, Inc, 2006. 22

[48] ITU-R Recommendation BT 601-7, "Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios," International Telecommunication Union, Tech. Rep., March 2011. 22, 44

[49] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996. 22

[50] J. Hardeberg, *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches*. Universal Publishers, 2001. 23

[51] M. S. Drew and G. D. Finlayson, "Analytic solution for separating spectra into illumination and surface reflectance components," *J. Opt. Soc. Am. A*, vol. 24, no. 2, pp. 294–303, Feb 2007. 23

[52] A. Ribés and F. Schmitt, "A fully automatic method for the reconstruction of spectral reflectance curves by using mixture density networks," *Pattern Recognition Letters*, vol. 24, no. 11, pp. 1691 – 1701, 2003, colour Image Processing and Analysis. First European Conference on Colour in Graphics, Imaging, and Vision (CGIV 2002). 23

[53] S. Westland and C. Ripamonti, *Computational Colour Science using MATLAB*. Wiley, 2004. 23

[54] P. Hanhart, P. Korshunov, and T. Ebrahimi, "Subjective evaluation of higher dynamic range video," in *SPIE Applications of Digital Image Processing XXXVII*, Aug. 2014. 24

[55] P. Korshunov, H. Nemoto, A. Skodras, and T. Ebrahimi, "Crowdsourcing-based evaluation of privacy in HDR images," in *SPIE Photonics Europe 2014, Optics, Photonics and Digital Technologies for Multimedia Applications*, Brussels, Belgium, Apr. 2014. 25, 29

[56] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 13–24, Jan. 2005. 25, 26, 71, 79

[57] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," in *Proc. of EUROGRAPHICS 2003*, ser. Computer Graphics Forum, P. Brunet and D. W. Fellner, Eds., vol. 22, no. 3. Granada, Spain: Blackwell, 2003, pp. 419–426. 25, 71, 73, 78

[58] P. Korshunov, M. V. Bernardo, A. M. Pinheiro, and T. Ebrahimi, "Impact of tone-mapping algorithms on subjective and objective face recognition in HDR images," in *Proceedings of the Fourth International Workshop on Crowdsourcing for Multimedia*, ser. CrowdMM '15. New York, NY, USA: ACM, 2015, pp. 39–44. 25, 26, 29

[59] E. Reinhard, M. Stark, P. Shirley, and J. Fewerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267–276, 2002. 25, 71, 73, 79

[60] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Comput. Graph.*, vol. 32, no. 3, pp. 330–349, 2008. 25

[61] R. Mantiuk and H.-P. Seidel, "Modeling a generic tone-mapping operator," *Computer Graphics Forum*, vol. 27, no. 2, pp. 699–708, 2008. 25

[62] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward compatible high dynamic range MPEG video compression," in *ACM SIGGRAPH 2006 Papers*, ser. SIGGRAPH'06. New York, NY, USA: ACM, Aug. 2006, pp. 713–723. 25, 71

[63] J. Kuang, G. M. Johnson, and M. D. Fairchild, "iCAM06: A refined image appearance model for HDR image rendering," *Journal of Visual Communication and Image Representation*, vol. 18, pp. 406–414, 2007. 26, 73

[64] Z. Mai, C. Doutre, P. Nasiopoulos, and R. Ward, "Subjective evaluation of tone-mapping methods on 3D images," in *17th International Conference on Digital Signal Processing (DSP)*. Piscataway, NJ, USA: IEEE, Jul. 2011, pp. 1–6. 26, 71, 79

[65] K. U. R. Laghari and K. Connelly, "Toward total quality of experience: A qoe model in a communication ecosystem," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 58–65, April 2012. 27, 51

[66] P. L. Callet, S. Möller, and A. Perkis, *Qualinet White Paper on Definitions of Quality of Experience*. EU COST Action 1003 QUALINET, June 2012. 27

[67] U. Reiter, K. Brunnström, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank, *Quality of Experience: Advanced Concepts, Applications and Methods*. Springer International Publishing, 2014, ch. Factors Influencing Quality of Experience, pp. 55–72. 27, 51

[68] P. Reichl, J. Fabini, M. Happenhofer, and C. Egger, "From QoS to QoX: A Charging Perspective." in *Proc. 18th ITC Specialist Seminar on Quality of Experience, Blekinge Institute of Technology, Karlskrona*, May 2008. 27

[69] ITU-T Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia application," International Telecommunication Union, Tech. Rep., December 1998. 27, 37, 51

[70] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio-visual services: A survey." *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482–501, 2010. 27, 37, 51

[71] VQEG, "Final report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," Phase II VQEG, March 2000. [Online]. Available: http://www.vqeg.org/ 27, 33, 37

[72] ITU-R Recommendation BT.500-12, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Tech. Rep., September 2009. 27, 28, 29, 40, 42, 43, 55, 56, 88

[73] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Tech. Rep., April 2008. 27

[74] ITU-R Recommendation BT.2022, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays." International Telecommunication Union, Tech. Rep., August 2012. 27

[75] J. Howe, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*, 1st ed. New York, NY, USA: Crown Publishing Group, 2008. 28

[76] E. Schenk and C. Guittard, "Crowdsourcing: What can be outsourced to the crowd, and why?" Tech. Rep., Dec. 2009. 28

[77] T. Hoßfeld, M. Hirth, P. Korshunov, P. Hanhart, B. Gardlo, C. Keimel, and C. Timmerer, "Survey of web-based crowdsourcing frameworks for subjective quality assessment," in *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*, Sept 2014, pp. 1–6. 29

[78] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of YouTube QoE via crowdsourcing," in *Symposium on Multimedia*, Dana Point, USA, Dec. 2011. 29

[79] J. Redi, T. Hoßfeld, P. Korshunov, F. Mazza, I. Povoa, and C. Keimel, "Crowdsourcing-based multimedia subjective evaluations: a case study on image recognizability and aesthetic appeal," in *ACM CrowdMM 2013*, Barcelona, Spain, Oct. 2013. 29

[80] I. Hupont, P. Lebreton, T. Mäki, E. Skodras, and M. Hirth, "Is it possible to crowdsource emotions?" in *International Conference on Communications and Electronics*, Da Nang, VN, jul 2014. 29

[81] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541–558, Feb 2014. 29

[82] T. Hoßfeld, R. Schatz, and S. Egger, "Sos: The mos is not enough!" in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, Sept 2011, pp. 131–136. 30

[83] L. J. Karam, T. Ebrahimi, S. S. Hemami, T. N. Pappas, R. J. Safranek, Z. Wang, and A. B. Watson, "Introduction to the issue on visual media quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 189–192, April 2009. 30

[84] ITU-T Recommendation J.148, "Requirements for an objective perceptual multimedia quality model," International Telecommunication Union, Tech. Rep., May 2004. 30

[85] ITU-T Recommendation P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Tech. Rep., jul 2012. 30, 90

[86] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 –612, April 2004. 30, 31, 45, 74, 81, 88

[87] D. Beixing, Y. Bo, L. Ling, and L. Xing, "Steganography based on baseline sequential jpeg compression," *Wuhan University Journal of Natural Sciences*, vol. 11, no. 2, pp. 370–372, 2006. 31, 88

[88] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *Image Processing, IEEE Transactions on*, vol. 23, no. 2, pp. 684–695, February 2014. 31

[89] V. Laparra, J. M. noz Marí, and J. Malo, "Divisive normalization image quality metric revisited," *J. Opt. Soc. Am. A*, vol. 27, no. 4, pp. 852–864, Apr 2010. 31, 89

[90] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81 –84, March 2002. 31, 88

[91] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, November 2003, pp. 1398–1402. 31, 45, 88

[92] A. Shnayderman, A. Gusev, and A. Eskicioglu, "An SVD - based grayscale image quality measure for local and global assessment," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 422–429, Feb 2006. 31, 88

[93] S. Aja-Fernandéz, R. Estepar, C. Alberola-Lopez, and C. F. Westin, "Image quality assessment based on local variance," in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 2006, pp. 4815–4818. 31, 88

[94] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 43–54, Jan 2013. 31

[95] D. Chandler and S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on*, vol. 16, no. 9, pp. 2284–2298, 2007. 31

[96] H. Sheikh, A. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2117–2128, 2005. 31, 88

[97] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430 –444, February 2006. 31, 88

[98] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2378–2386, 2011. 31, 46, 74, 81, 88

[99] L. Zhang, D. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using riesz transforms," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 321–324. 31

[100] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 5, pp. 1185–1198, 2011. 31, 88, 89

[101] X. Yang, W. Ling, Z. Lu, E. Ong, and S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Processing: Image Communication*, vol. 20, no. 7, pp. 662 – 680, 2005. 31, 89

[102] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *Information Theory, IEEE Transactions on*, vol. 20, no. 4, pp. 525–536, 1974. 31, 89

[103] T. Mitsa and K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 5, 1993, pp. 301–304. 31, 89

[104] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, January 2006. 31

[105] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2007. 31

[106] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy." *Journal of Electronic Imaging*, vol. 19(1), January 2010. 31

[107] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *Multimedia, IEEE Transactions on*, vol. 13, no. 5, pp. 935–949, Oct 2011. 31

[108] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 40:1–40:14, Jul. 2011. 32, 87, 89

[109] M. Narwaria, R. K. Mantiuk, M. Perreira Da Silva, and P. Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, p. 010501, 2015. 32, 87, 89

[110] M. Narwaria, M. Perreira Da Silva, and P. Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015. 32, 87, 89, 91

[111] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC79 colour-difference formula," *Journal of the Society of Dyers and Colourists*, vol. 100, no. 4, pp. 128–132, 1984. 32, 89

[112] X. Zhang, B. A. Wandell, and A. W. Brian, "A spatial extension of CIELAB for digital color image reproduction," *Journal of the Society for Information Display*, vol. 5, no. 1, pp. 61–63, 1997. 32, 45

[113] M. Pedersen and J. Y. Hardeberg, "SHAME: A new spatial hue angle metric for perceptual image difference," *Journal of Vision*, vol. 9, no. 8, p. 343, 2009. 32

[114] S. J. Erasmus and K. C. A. Smith, "An automatic focusing and astigmatism correction system for the SEM and CTEM," *Journal of Microscopy*, vol. 127, no. 2, pp. 185–199, 1982. 32, 89

[115] C. F. Batten, "Autofocusing and astigmatism correction in the scanning electron microscope," Master's thesis, University of Cambridge, U.K., 2000. 32, 46, 89

[116] A. Murthy and L. Karam, "A MATLAB-based framework for image and video quality evaluation," in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, June 2010, pp. 242–247. 32, 46, 89

[117] D. Shaked and I. Tastl, "Sharpness measure: towards automatic image enhancement," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 1, 2005, pp. I–937–40. 32, 89

[118] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual Blur and Ringing Metrics: Application to JPEG2000," *Signal Processing : Image Communication*, vol. 19, no. 2, pp. 163–172, 2004. 32, 89

[119] N. Zhang, A. Vladar, M. Postek, and B. Larrabee, "A kurtosis-based statistitcal measure for two-dimensional processes and its application to image sharpness," in *Proceedings Section of Physical and Engineering Sciences of American Statistical Society*, 2003, pp. 4730–4736. 32, 89

[120] R. Ferzli, L. J. Karam, and J. Caviedes, "A robust image sharpness metric based on kurtosis measurement of wavelet coefficients," in *Proceedings of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005. 32, 89

[121] R. Ferzli and L. J. Karam, "A no-reference objective sharpness metric using riemannian tensor," in *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, Arizona, January 2007. 32, 89

[122] ——, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *Image Processing, IEEE Transactions on*, vol. 18, no. 4, pp. 717–728, 2009. 32

[123] N. Narvekar and L. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *Image Processing, IEEE Transactions on*, vol. 20, no. 9, pp. 2678–2683, 2011. 32

[124] A. Moorthy and A. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3350–3364, Dec 2011. 32

[125] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, Aug 2012. 32

[126] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012. 32

[127] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, March 2013. 32

[128] J. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference, Fourth Edition: Revised and Expanded*, ser. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 2003. 33

[129] C. Spearman, "Correlation calculated from faulty data," *British Journal of Psychology, 1904-1920*, vol. 3, no. 3, pp. 271–295, 1910. 33

[130] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," International Telecommunication Union, Tech. Rep., August 2008. 37

[131] CIE, "Guidelines for the evaluation of gamut mapping algorithms," CIE Central Bureau, CIE Publication 156 8-03, 2004. 37

[132] N. Katoh and M. Ito, "Gamut mapping for computer generated images (ii)," *Color and Imaging Conference*, vol. 1996, no. 1, pp. 126–128, 1996. 37, 41, 55

[133] W. Wallace and M. Stone, "Gamut mapping computer generated imagery. image handling and reproduction systems integration." in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, vol. 1460, 1991, pp. 20–28. 37

[134] P. Fiadeiro, S. Nascimento, D. Foster, K. Amano, and V. Almeida, "Reproduction of colours of natural scenes by CRT and TFT display devices," in *Proceedings of the 10th Congress of the AIC*, Granada, 2005, pp. 1191–1194. 37

[135] C. Parraga, G. Brelstaff, T. Troscianko, and I. Moorehead, "Color and luminance information in natural scenes," *Measurement*, vol. 15, no. c, pp. 1–19, 1998. 37

[136] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, pp. 2036–2045, Aug 1998. 37

[137] S. Nascimento, F. Ferreira, and D. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *Journal of the Optical Society of America A*, vol. 19, no. 8, pp. 1484–1490, Aug 2002. 37

[138] D. Foster, S. Nascimento, and K. Amano, "Information limits on neural identification of colored surfaces in natural scenes." *Visual Neuroscience*, vol. 21, no. 3, pp. 331–336, 2004. 37, 38, 41, 55

[139] M. Pointer, "The gamut of real surface colours," *Color Research Application*, vol. 5, no. 3, pp. 145–155, 1980. 37

[140] F. Long and D. Purves, "Natural scene statistics as the universal basis of color context effects." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 25, pp. 15 190–15 193, Dec. 2003. 37

[141] M. Aldaba, J. Linhares, P. Pinto, S. Nascimento, K. Amano, and D. Foster, "Visual sensitivity to color errors in images of natural scenes," *Visual Neuroscience*, vol. 23, no. 3-4, pp. 555–559, 2006. 38, 39, 48, 52

[142] J. M. Linhares, P. D. Pinto, and S. M. Nascimento, "The number of discernible colors in natural scenes," *J. Opt. Soc. Am. A*, vol. 25, no. 12, pp. 2918–2924, 2008. 38, 52

[143] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall PTR, 2011. 38, 39

[144] S. Winkler, *Digital Video Quality: Vision Models and Metrics*, 1st ed. John Wiley & Sons, March 2005. 38

[145] R. Hunt, *The Reproduction of Colour*, 3rd ed. New York: John Wiley & Sons, March 1975. 39, 40

[146] W. Wintringham, "Color television and colorimetry," *Proc. of the IRE*, vol. 39, no. 10, pp. 1135 –1172, October 1951. 39

[147] D. Brainard, "Calibration of a computer controlled color monitor," *Color Research & Application*, vol. 14, no. 1, pp. 23–34, 1989. 41

[148] W. Cowan, *The Perception of Colour*. CRC Press, 1991, ch. Colour Television, pp. 294–307. 41

[149] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *J. Opt. Soc. Am. A*, vol. 23, no. 10, pp. 2359–2372, Oct 2006. 41, 55

[150] R. Krueger, R. Applegate, and S. MacRae, *Wavefront Customized Visual Corrections: The Quest for Super Vision II*. SLACK, 2004. 42, 58

[151] J. Mollon and J. Reffin, "A computer-controlled colour vision test that combines the principles of chibret and stilling," *Journal of Physiology*, vol. 414, no. 5P, 1989. 43, 56

[152] WMA, "World Medical Association Declaration of Helsinki - Ethical principles for medical research involving human subjects," 2009. 43, 56

[153] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, Dec. 1965. 43, 60

[154] W. Kruskal and W. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, pp. 583–621, 1952. 43, 60

[155] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," *Computer Graphics Forum*, vol. 31, no. 8, pp. 2478–2491, 2012. 44

[156] M. Pinson and S. Wolf, "An objective method for combining multiple subjective data sets," in *Proceedings of the SPIE Video Communication and Image Processing Conference*, Lugano, Switzerland, July 2003. 44

[157] S. Winkler and R. Campos, "Video quality evaluation for internet streaming applications," in *SPIE Conference on Human Vision and Electronic Imaging VII*, Jan 2003, pp. 104–115. 44

[158] G. M. Johnson and M. D. Fairchild, "Darwinism of color image difference models," in *Proc. of IS&T/SID 9 th Color Imaging Conference*, 2001, pp. 108–112. 45

[159] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991. 46

[160] J. S. Babcock, J. B. Pelz, and M. D. Fairchild, "Eye tracking observers during color image evaluation tasks," in *Proceedings on Human Vision and Electronic Imaging VIII (SPIE)*, vol. 5007, June 2003, pp. 218–230. 51

[161] K. Fliegel, "Eyetracking based approach to objective image quality assessment," in *2008 42nd Annual IEEE International Carnahan Conference on Security Technology*, Oct 2008, pp. 371–376. 51

[162] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky, "Eye tracking in web search tasks: Design implications," in *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ser. ETRA '02. New York, NY, USA: ACM, 2002, pp. 51–58. 51

[163] R. J. K. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," *Mind*, vol. 2, no. 3, p. 4, 2003. 51

[164] H. Alers, L. Bos, and I. Heynderickx, "How the task of evaluating image quality influences viewing behavior," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, Sept 2011, pp. 167–172. 51

[165] O. L. Meur, A. Ninassi, P. L. Callet, and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 547 – 558, 2010, special Issue on Image and Video Quality Assessment. 51

[166] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in *2007 IEEE International Conference on Image Processing*, vol. 2, Sept 2007, pp. II – 169–II – 172. 51

[167] U. Engelke, A. Maeder, and H. Zepernick, "Visual attention modelling for subjective image quality databases," in *Multimedia Signal Processing, 2009. MMSP '09. IEEE International Workshop on*, Oct 2009, pp. 1–6. 51

[168] M. Vijay Venkatesh and S. Cheung, "Eye tracking based perceptual image inpainting quality analysis," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 1109–1112. 51

[169] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 7, pp. 971–982, July 2011. 51

[170] L. Jansen, S. Onat, and P. König, "Influence of disparity on fixation and saccades in free viewing of natural scenes," *Journal of Vision*, vol. 9, no. 1, 2009. 51

[171] D. A. Wismeijer, C. J. Erkelens, R. v. Ee, and M. Wexler, "Depth cue combination in spontaneous eye movements," *Journal of Vision*, vol. 10, no. 6, pp. 1–15, Jun 2010. 51

[172] Q. Huynh-Thu and L. Schiatti, "Examination of 3D visual attention in stereoscopic video content," in *Proceedings on Human Vision and Electronic Imaging XVI (SPIE)*, vol. 7865, Jan 2011, p. 15. 51

[173] S. Ishihara, *Ishihara's Tests for Colour Blindness: 24 Plate Edition*, ser. Abridged 24 Plate Edition. Taylor & Francis, 1998. 56

[174] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes," *Cognitive Psychology*, vol. 8, no. 4, pp. 441 – 480, 1976. 58

[175] A. Poole and L. J. Ball, "Eye tracking in human-computer interaction and usability research: Current status and future," in *Prospects", Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc*, 2005. 58

[176] ITU-T Tutorial, "Objective perceptual assessment of video quality: Full reference television," International Telecommunication Union, Tech. Rep., 2004. 58, 64

[177] T. Richter, "Backwards compatible coding of high dynamic range images with JPEG," in *Proceedings of the 2013 Data Compression Conference*, ser. DCC'13. Washington, DC, USA: IEEE Computer Society, Mar. 2013, pp. 153–160. 69

[178] A. Artusi, R. K. Mantiuk, T. Richter, P. Hanhart, P. Korshunov, M. Agostinelli, A. Ten, and T. Ebrahimi, "Overview and evaluation of the JPEG XT HDR image compression standard," *Journal of Real-Time Image Processing*, pp. 1–16, 2015. 69

[179] A. Artusi, R. K. Mantiuk, T. Richter, P. Korshunov, P. Hanhart, T. Ebrahimi, and M. Agostinelli, "JPEG XT: A compression standard for hdr and wcg images [standards in a nutshell]," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 118–124, March 2016. 69

[180] G. Ward and M. Simmons, "JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG," in *ACM SIGGRAPH 2006 Courses*, ser. SIGGRAPH '06. New York, NY, USA: ACM, 2006. 69

[181] T. Richter, "On the integer coding profile of JPEG XT," in *SPIE Applications Of Digital Image Processing XXXVII*, vol. 9217, 2014. 70

[182] T. Richter, A. Artusi, and T. Ebrahimi, "JPEG XT: A New Family of JPEG Backward-Compatible Standards," *IEEE MultiMedia*, vol. 23, no. 3, pp. 80–88, July 2016. 70

[183] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*. Natick, MA, USA: AK Peters (CRC Press), 2011. 73

[184] L. Krasula, M. Narwaria, and P. Le Callet, "An automated approach for tone mapping operator parameter adjustment in security applications," in *SPIE Photonics Europe*. International Society for Optics and Photonics, 2014, pp. 913 803–913 803. 73

[185] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 68:1–68:10, 2008. 79, 88

[186] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images: model and its calibration," in *SPIE Human Vision and Electronic Imaging X*, vol. 5666, 2005. 87

[187] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *SPIE Human Vision, Visual Processing, and Digital Display III*, vol. 1666, 1992. 87

[188] C. Poynton, *Digital video and HD: Algorithms and Interfaces*. Burlington, Vermont, USA: Elsevier/Morgan Kaufmann, 2012. 87

[189] S. K. Shevell, *The science of color*. Boston, Massachusetts, USA: Elsevier, 2003. 87

[190] A. Rose, "The sensitivity performance of the human eye on an absolute scale," *Journal of the Optical Society of America A*, vol. 38, no. 2, pp. 196–208, 1948. 87

[191] H. De Vries, "The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye," *Physica*, vol. 10, no. 7, pp. 553–564, 1943. 87

[192] T. O. Aydın, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Extending quality metrics to full luminance range images," in *SPIE Human Vision and Electronic Imaging XIII*, vol. 6806, Feb 2008. 87, 88, 89, 90

[193] H. Sheikh, M. Sabir, and A. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, November 2006. 87

[194] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of Subjective and Objective Quality Assessment of Video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, June 2010. 87

[195] G. Valenzise, F. De Simone, P. Lauga, and F. Dufaux, "Performance evaluation of objective quality metrics for HDR image compression," in *SPIE Applications of Digital Image Processing XXXVII*, vol. 9217, Aug 2014. 87, 91, 92

[196] C. Mantel, S. Ferchiu, and S. Forchhammer, "Comparing subjective and objective quality assessment of HDR images compressed with JPEG-XT," in *16th International Workshop on Multimedia Signal Processing*, Sept 2014. 87, 91

[197] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual Signal Coding for More Efficient Usage of Bit Codes," *SMPTE Motion Imaging Journal*, vol. 122, no. 4, pp. 52–59, May 2013. 88, 89, 90

[198] P. Korshunov, P. Hanhart, T. Ricther, A. Artusi, R. Mantiuk, and T. Ebrahimi, "Subjective quality assessment database of HDR images compressed with JPEG XT," in *7th International Workshop on Quality of Multimedia Experience*, May 2015. 88, 89

[199] Subjective quality assessment database of HDR images compressed with JPEG XT. [Online]. Available: http://mmspg.epfl.ch/jpegxt-hdr 88, 89

[200] H. Wilson, "A transducer function for threshold and suprathreshold human vision," *Biological Cybernetics*, vol. 38, no. 3, pp. 171–178, 1980. 90

[201] P. G. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. Bellingham, Washington, USA: SPIE Optical Engineering Press, 1999. 90

[202] ITU-R Recommendation BT.709, "Parameter values for the HDTV standards for production and international programme exchange," International Telecommunication Union, Tech. Rep., Apr 2002. 90

[203] P. Hanhart, M. Rerabek, and T. Ebrahimi, "Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies," in *Proc. SPIE, Applications of Digital Image Processing XXXVIII*, Aug. 2015. 91

[204] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, June 2008. 91

Appendix A

CIEDE2000 Color Difference Formula

$$\Delta E_{00}^{*} = \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{K_C S_C}\right)^2 \left(\frac{\Delta H'}{K_H S_H}\right)^2} \quad \text{(A.1)}$$

where,

$$\begin{aligned} L' &= L^* \\ a' &= (1+G)a^* \\ b' &= b^* \\ C' &= \sqrt{(a')^2 + (b')^2} \\ h' &= \tan^{-1}(b'/a') \end{aligned}$$

$$G = 0.5\left(1 - \sqrt{\frac{\overline{\Delta E_{ab}^{*}}^{7}}{\overline{\Delta E_{ab}^{*}}^{7} + 25^7}}\right)$$

$$\begin{aligned} \Delta L' &= L'_b - L'_s \\ \Delta C' &= C'_b - C'_s \\ \Delta H' &= 2\sqrt{C'_b C'_s}\sin\left(\frac{\Delta h'}{2}\right) \\ \Delta h' &= h'_b - h'_s \end{aligned}$$

$$S_L = 1 + \frac{0.015(\overline{L'} - 50)^2}{\sqrt{20 + (\overline{L'} - 50)^2}}$$

$$\begin{aligned} S_C &= 1 + 0.045\overline{C'} \\ S_H &= 1 + 0.015\overline{C'}T \\ R_T &= -\sin(2\Delta\theta)R_C \\ \Delta\theta &= 30\exp\{-[(\overline{h'} - 275^\circ)/25]^2\} \end{aligned}$$

$$R_C = 2\sqrt{\frac{\overline{h'}^7}{\overline{h'}^7 + 25^7}}$$

and where,

$$T = 1 - 0.17\cos(\overline{h'} - 30^\circ) + 0.24\cos(2\overline{h'}) + 0.32\cos(3\overline{h'} + 6^\circ) - 0.20\cos(4\overline{h'} - 63^\circ)$$

Appendix B

Evaluation of Perceptual Sensitivity to Chromatic Changes

Protocol

Objective

In this psychophysical experience, observers shall classify in terms of naturality several images that are displayed on the monitor.

Methodology

The observer will sequentially visualize a set of digital images that correspond to natural scenes from both rural and urban areas. The observation of each image is performed under a gray background and has a duration of 10 seconds. After observing the images a rating scale with a triangular cursor is shown. Follows a classification period with a duration of 10 seconds also. The image classification implies assigning a degree of quality/naturalness that extends from good quality images/natural (value 100) to poor quality images/unnatural (value 0). The classification is assigned with values between 0 and 100, moving the triangular cursor in the horizontal directions using the joystick. After the selection is done press any button to validate the selected rating.

In order to become familiar with the automatic data collection, the experience is preceded by a short period of training and customizing to the system. The beginning of the experiment is properly marked and will last approximately 20 minutes. During the experiment, a “no response” is not a problem and the observer should not speak or look away from the monitor.

On behalf of the research team, thank you for your participation and collaboration.

Appendix C

Benchmarking of Quality Metrics Performance Indexes

Table C.1: Accuracy, consistency, and monotonicity indexes for each objective metric computed in the linear space.

PCC		SROCC		RMSE		OR	
HDRVDP2	0.9604	HDRVDP2	0.9564	HDRVDP2	0.3498	HDRVDP2	0.3500
HDRVQM	0.9602	HDRVQM	0.9564	HDRVQM	0.3506	HDRVQM	0.4083
IFC_Y	0.9140	IFC_Y	0.9205	IFC_Y	0.5109	IFC_Y	0.5458
FSIM_Y	0.8938	FSIM_Y	0.9160	FSIM_Y	0.5643	UQI_Y	0.5667
UQI_Y	0.8873	IFC_M	0.8952	UQI_Y	0.5792	IWPSNR_Y	0.6167
IFC_M	0.8855	DN	0.8926	IFC_M	0.5845	IWSSIM_Y	0.6167
DN	0.8814	WSNR_Y	0.8791	DN	0.5936	IWPSNR_M	0.6417
WSNR_Y	0.8786	MSSSIM_Y	0.8776	WSNR_Y	0.5995	IWSSIM_M	0.6417
FSIM_M	0.8571	FSIM_M	0.8768	FSIM_M	0.6476	IFC_M	0.6458
MSSSIM_Y	0.8545	UQI_Y	0.8737	VIF_Y	0.6521	WSNR_Y	0.6500
VIF_Y	0.8545	VIF_Y	0.8617	MSSSIM_Y	0.6527	VIF_Y	0.6583
IWPSNR_Y	0.8352	IWMSE_Y	0.8374	IWPSNR_Y	0.6907	WSNR_M	0.6667
IWSSIM_Y	0.8352	IWPSNR_Y	0.8374	IWSSIM_Y	0.6907	VIFP_Y	0.6667
VIFP_Y	0.8275	IWSSIM_Y	0.8374	VIFP_Y	0.7050	DN	0.6667
WSNR_M	0.8178	VIFP_Y	0.8344	WSNR_M	0.7225	FSIM_M	0.6750
UQI_M	0.8082	IWMSE_M	0.8298	UQI_M	0.7396	FSIM_Y	0.6833
IWPSNR_M	0.8052	WSNR_M	0.8273	IWPSNR_M	0.7446	UQI_M	0.6917
IWSSIM_M	0.8052	IWPSNR_M	0.8145	IWSSIM_M	0.7446	CMC	0.6917
CMC	0.8045	IWSSIM_M	0.8145	CMC	0.7458	VIF_M	0.7000
CIE94	0.7987	UQI_M	0.8112	CIE94	0.7558	IWMSE_M	0.7042
CIEDE00	0.7951	CIE94	0.8031	CIEDE00	0.7615	MSSSIM_M	0.7042
IWMSE_Y	0.7951	CMC	0.8019	IWMSE_Y	0.7622	CIEDE00	0.7083
IWMSE_M	0.7907	VIF_M	0.7946	IWMSE_M	0.7689	CIE94	0.7083
VIF_M	0.7813	CIEDE00	0.7908	VIF_M	0.7836	SNR_M	0.7208
MSVD_Y	0.7629	MSSSIM_M	0.7877	MSVD_Y	0.8120	SNR_Y	0.7333
MSSSIM_M	0.7610	MSVD_Y	0.7758	MSSSIM_M	0.8146	PSNR_M	0.7375
SNR_Y	0.7535	VIFP_M	0.7657	SNR_Y	0.8253	CIE1976	0.7458
VIFP_M	0.7520	MSVD_M	0.7600	VIFP_M	0.8275	VIFP_M	0.7500
MSVD_M	0.7466	SNR_Y	0.7574	MSVD_M	0.8355	IWMSE_Y	0.7542
SSIM_Y	0.7374	SSIM_Y	0.7559	SSIM_Y	0.8482	MSSSIM_Y	0.7583
CIE1976	0.7254	MSE_Y	0.7262	CIE1976	0.8642	MSVD_M	0.7667
PSNR_Y	0.7152	PSNR_Y	0.7262	PSNR_Y	0.8774	SSIM_M	0.7708
MSE_Y	0.7050	CIE1976	0.7256	MSE_Y	0.8924	PSNR_Y	0.7792
SNR_M	0.6872	MSE_M	0.6978	SNR_M	0.9120	QILV_M	0.7792
MSE_M	0.6777	SNR_M	0.6969	MSE_M	0.9232	MSVD_Y	0.7833
PSNR_M	0.6525	PSNR_M	0.6628	PSNR_M	0.9513	MSE_M	0.8000
QILV_Y	0.6183	QILV_Y	0.6615	QILV_Y	0.9867	QILV_Y	0.8042
QILV_M	0.6164	QILV_M	0.6535	QILV_M	0.9886	SSIM_Y	0.8125
SSIM_M	0.5904	SSIM_M	0.5948	SSIM_M	1.0132	MSE_Y	0.8167
Marziliano_Y	0.4551	Marziliano_Y	0.4160	Marziliano_Y	1.1178	Marziliano_M	0.8292
Marziliano_M	0.3669	HPM_M	0.3341	Marziliano_M	1.1678	Marziliano_Y	0.8333
HPM_Y	0.3625	HPM_Y	0.3211	HPM_Y	1.1700	JND_St_M	0.8375
HPM_M	0.3503	Marziliano_M	0.3093	HPM_M	1.1758	RTBM_Y	0.8375
JND_St_Y	0.2913	JND_St_Y	0.2393	JND_St_Y	1.2009	RTBM_M	0.8375
JND_St_M	0.2509	JND_St_M	0.1599	JND_St_M	1.2152	JND_St_Y	0.8500
GRAD_M	0.1060	LAP_M	0.1197	GRAD_M	1.2483	VAR_Y	0.8500
GRAD_Y	0.1012	LAP_Y	0.1030	GRAD_Y	1.2489	VAR_M	0.8500
KurtZhang_M	0.0829	GRAD_M	0.0784	KurtZhang_M	1.2513	LAP_M	0.8500
KurtWav_M	0.0709	GRAD_Y	0.0742	KurtWav_M	1.2522	GRAD_M	0.8500
RTBM_Y	0.0709	KurtWav_M	0.0607	RTBM_Y	1.2522	FTM_Y	0.8500
AutoCorr_Y	0.0675	KurtZhang_Y	0.0518	AutoCorr_Y	1.2525	HPM_Y	0.8500
AutoCorr_M	0.0664	KurtZhang_M	0.0366	AutoCorr_M	1.2526	KurtZhang_Y	0.8500
RTBM_M	0.0623	AutoCorr_Y	0.0301	RTBM_M	1.2529	KurtWav_Y	0.8500
LAP_M	0.0537	AutoCorr_M	0.0297	LAP_M	1.2535	KurtWav_M	0.8500
LAP_Y	0.0458	KurtWav_Y	0.0265	LAP_Y	1.2540	JND_Y	0.8542
FTM_Y	0.0324	JND_M	0.0138	FTM_Y	1.2547	JND_M	0.8542
KurtWav_Y	0.0290	JND_Y	0.0110	KurtWav_Y	1.2548	LAP_Y	0.8542
JND_Y	0.0200	RTBM_Y	−0.0095	JND_Y	1.2551	FTM_M	0.8542
JND_M	0.0198	VAR_Y	−0.0221	JND_M	1.2551	KurtZhang_M	0.8542
KurtZhang_Y	0.0194	VAR_M	−0.0354	KurtZhang_Y	1.2551	AutoCorr_Y	0.8542
FTM_M	0.0082	FTM_Y	−0.0396	FTM_M	1.2553	AutoCorr_M	0.8542
VAR_M	0.0068	RTBM_M	−0.0421	VAR_M	1.2553	GRAD_Y	0.8625
VAR_Y	0.0067	FTM_M	−0.0748	VAR_Y	1.2553	HPM_M	0.8625

Table C.2: Accuracy, consistency, and monotonicity indexes for each objective metric computed in the logarithm space.

PCC		SROCC		RMSE		OR	
VIFP_Y	0.9230	VIFP_Y	0.9200	VIFP_Y	0.4832	VIFP_Y	0.4833
VIF_Y	0.9185	VIF_Y	0.9174	VIF_Y	0.4974	IFC_Y	0.5500
IFC_Y	0.9051	IFC_Y	0.9112	IFC_Y	0.5355	VIF_Y	0.5583
MSSSIM_Y	0.8971	MSSSIM_Y	0.9091	MSSSIM_Y	0.5560	UQI_Y	0.5917
IFC_M	0.8928	IFC_M	0.9037	IFC_M	0.5672	SSIM_Y	0.6125
SSIM_Y	0.8900	SSIM_Y	0.8952	SSIM_Y	0.5727	PSNR_Y	0.6208
UQI_Y	0.8780	FSIM_Y	0.8817	UQI_Y	0.6009	IFC_M	0.6250
FSIM_Y	0.8553	UQI_Y	0.8603	FSIM_Y	0.6516	MSSSIM_Y	0.6292
WSNR_Y	0.8404	UQI_M	0.8441	WSNR_Y	0.6803	MSE_Y	0.6375
UQI_M	0.8373	WSNR_Y	0.8416	UQI_M	0.6870	MSVD_Y	0.6500
PSNR_Y	0.8348	MSE_Y	0.8399	PSNR_Y	0.6911	IWPSNR_Y	0.6542
MSVD_Y	0.8316	PSNR_Y	0.8399	MSVD_Y	0.6975	SNR_Y	0.6542
MSE_Y	0.8272	MSVD_Y	0.8370	MSE_Y	0.7055	IWSSIM_Y	0.6542
SNR_Y	0.8269	SNR_Y	0.8333	SNR_Y	0.7060	UQI_M	0.6583
IWPSNR_Y	0.8160	IWPSNR_Y	0.8165	IWPSNR_Y	0.7256	WSNR_Y	0.6625
IWSSIM_Y	0.8160	IWSSIM_Y	0.8165	IWSSIM_Y	0.7256	DN	0.6625
VIF_M	0.8079	IWMSE_Y	0.8165	VIF_M	0.7401	FSIM_Y	0.6708
VIFP_M	0.7986	VIF_M	0.8152	VIFP_M	0.7556	MSE_M	0.6958
IWMSE_Y	0.7912	VIFP_M	0.8082	IWMSE_Y	0.7680	VIF_M	0.7000
DN	0.7877	DN	0.7993	DN	0.7737	IWPSNR_M	0.7083
MSSSIM_M	0.7482	FSIM_M	0.7603	MSSSIM_M	0.8330	IWSSIM_M	0.7083
FSIM_M	0.7363	MSSSIM_M	0.7584	FSIM_M	0.8498	VIFP_M	0.7125
WSNR_M	0.7252	WSNR_M	0.7324	WSNR_M	0.8644	MSSSIM_M	0.7125
QILV_Y	0.6918	QILV_Y	0.6913	QILV_Y	0.9088	WSNR_M	0.7208
SSIM_M	0.6855	SSIM_M	0.6847	SSIM_M	0.9139	PSNR_M	0.7250
MSE_M	0.6785	MSVD_M	0.6772	MSE_M	0.9222	SSIM_M	0.7250
MSVD_M	0.6779	IWPSNR_M	0.6580	MSVD_M	0.9229	IWMSE_Y	0.7333
IWPSNR_M	0.6646	IWSSIM_M	0.6580	IWPSNR_M	0.9380	SNR_M	0.7417
IWSSIM_M	0.6646	PSNR_M	0.6409	IWSSIM_M	0.9380	IWMSE_M	0.7458
SNR_M	0.6415	SNR_M	0.6394	SNR_M	0.9630	MSVD_M	0.7583
PSNR_M	0.6412	MSE_M	0.6360	PSNR_M	0.9633	QILV_Y	0.7583
IWMSE_M	0.6162	IWMSE_M	0.5931	IWMSE_M	0.9890	FSIM_M	0.8125
HPM_Y	0.4900	QILV_M	0.5265	HPM_Y	1.0944	KurtWav_M	0.8333
Marziliano_Y	0.4855	HPM_Y	0.4874	Marziliano_Y	1.0975	HPM_Y	0.8417
Marziliano_M	0.4059	Marziliano_Y	0.4303	Marziliano_M	1.1473	AutoCorr_Y	0.8458
GRAD_Y	0.3736	HPM_M	0.3518	GRAD_Y	1.1645	JND_St_Y	0.8542
GRAD_M	0.2844	Marziliano_M	0.3056	GRAD_M	1.2035	JND_St_M	0.8542
JND_St_M	0.2591	GRAD_Y	0.2570	JND_St_M	1.2125	JND_Y	0.8542
LAP_Y	0.2153	GRAD_M	0.1915	LAP_Y	1.2270	JND_M	0.8542
VAR_M	0.1654	LAP_M	0.1736	VAR_M	1.2381	VAR_Y	0.8542
QILV_M	0.1427	LAP_Y	0.1642	KurtWav_M	1.2425	LAP_M	0.8542
KurtWav_M	0.1425	VAR_M	0.1548	VAR_Y	1.2426	FTM_Y	0.8542
VAR_Y	0.1423	FTM_M	0.1314	QILV_M	1.2427	KurtZhang_M	0.8542
KurtZhang_Y	0.1053	KurtZhang_Y	0.1268	KurtZhang_Y	1.2484	KurtWav_Y	0.8542
FTM_M	0.0736	VAR_Y	0.1226	FTM_M	1.2520	AutoCorr_M	0.8542
HPM_M	0.0599	KurtWav_M	0.0961	HPM_M	1.2531	RTBM_Y	0.8542
AutoCorr_M	0.0560	AutoCorr_Y	0.0754	AutoCorr_M	1.2534	RTBM_M	0.8542
KurtZhang_M	0.0429	JND_St_M	0.0752	KurtZhang_M	1.2542	QILV_M	0.8583
JND_M	0.0402	KurtZhang_M	0.0726	JND_M	1.2543	FTM_M	0.8583
JND_Y	0.0401	KurtWav_Y	0.0543	JND_Y	1.2543	HPM_M	0.8583
AutoCorr_Y	0.0379	JND_M	0.0516	AutoCorr_Y	1.2545	Marziliano_Y	0.8583
RTBM_M	0.0315	JND_Y	0.0498	RTBM_M	1.2547	KurtZhang_Y	0.8583
KurtWav_Y	0.0312	RTBM_M	0.0429	KurtWav_Y	1.2547	LAP_Y	0.8625
LAP_M	0.0306	FTM_Y	0.0230	LAP_M	1.2548	GRAD_Y	0.8625
FTM_Y	0.0227	AutoCorr_M	−0.0019	FTM_Y	1.2551	VAR_M	0.8667
RTBM_Y	0.0173	JND_St_Y	−0.0482	RTBM_Y	1.2553	Marziliano_M	0.8750
JND_St_Y	0.0038	RTBM_Y	−0.0494	JND_St_Y	1.2553	GRAD_M	0.8792

Table C.3: Accuracy, consistency, and monotonicity indexes for each objective metric computed in the PU space.

PCC		SROCC		RMSE		OR	
MSSSIM_Y	0.9447	MSSSIM_Y	0.9501	MSSSIM_Y	0.4132	VIF_Y	0.4833
FSIM_Y	0.9376	FSIM_Y	0.9470	FSIM_Y	0.4377	IWPSNR_Y	0.5167
VIF_Y	0.9291	VIF_Y	0.9276	VIF_Y	0.4649	IWSSIM_Y	0.5167
VIFP_Y	0.9288	VIFP_Y	0.9228	VIFP_Y	0.4656	VIFP_Y	0.5208
IWPSNR_Y	0.9130	IFC_Y	0.9170	IWPSNR_Y	0.5121	IFC_Y	0.5375
IWSSIM_Y	0.9130	IWMSE_Y	0.9109	IWSSIM_Y	0.5121	MSSSIM_Y	0.5417
IFC_Y	0.9110	IWPSNR_Y	0.9109	IFC_Y	0.5196	WSNR_Y	0.5583
DN	0.9078	IWSSIM_Y	0.9109	DN	0.5275	FSIM_Y	0.5625
SSIM_Y	0.9060	DN	0.9090	SSIM_Y	0.5316	SSIM_Y	0.5792
WSNR_Y	0.8959	SSIM_Y	0.9072	WSNR_Y	0.5577	UQI_Y	0.5833
IFC_M	0.8928	IFC_M	0.9043	IFC_M	0.5670	DN	0.5875
IWMSE_Y	0.8841	WSNR_Y	0.8950	IWMSE_Y	0.5878	PSNR_Y	0.5917
UQI_Y	0.8777	MSVD_Y	0.8638	UQI_Y	0.6016	SNR_Y	0.5958
MSVD_Y	0.8612	UQI_Y	0.8610	MSVD_Y	0.6392	IWMSE_Y	0.6250
PSNR_Y	0.8526	MSE_Y	0.8564	PSNR_Y	0.6562	IFC_M	0.6375
SNR_Y	0.8472	PSNR_Y	0.8564	SNR_Y	0.6669	MSVD_Y	0.6375
MSE_Y	0.8352	SNR_Y	0.8556	MSE_Y	0.6915	VIF_M	0.6625
FSIM_M	0.8310	FSIM_M	0.8484	FSIM_M	0.6991	VIFP_M	0.6667
UQI_M	0.8278	MSSSIM_M	0.8442	UQI_M	0.7049	UQI_M	0.6833
VIF_M	0.8275	VIF_M	0.8373	VIF_M	0.7053	IWPSNR_M	0.6917
MSSSIM_M	0.8273	UQI_M	0.8335	MSSSIM_M	0.7059	WSNR_M	0.6917
VIFP_M	0.8242	VIFP_M	0.8327	VIFP_M	0.7109	IWSSIM_M	0.6917
WSNR_M	0.8163	WSNR_M	0.8233	WSNR_M	0.7254	MSE_Y	0.7000
IWPSNR_M	0.7848	QILV_Y	0.8047	IWPSNR_M	0.7781	QILV_M	0.7042
IWSSIM_M	0.7848	IWPSNR_M	0.7937	IWSSIM_M	0.7781	SSIM_M	0.7083
MSVD_M	0.7837	IWSSIM_M	0.7937	MSVD_M	0.7804	MSSSIM_M	0.7083
QILV_Y	0.7779	MSVD_M	0.7862	QILV_Y	0.7922	IWMSE_M	0.7458
IWMSE_M	0.7414	IWMSE_M	0.7700	IWMSE_M	0.8426	SNR_M	0.7458
MSE_M	0.7280	MSE_M	0.7444	MSE_M	0.8651	FSIM_M	0.7458
SSIM_M	0.7194	SSIM_M	0.7324	SSIM_M	0.8719	PSNR_M	0.7542
SNR_M	0.7085	SNR_M	0.7147	SNR_M	0.8859	MSVD_M	0.7542
PSNR_M	0.7033	PSNR_M	0.7088	PSNR_M	0.8925	QILV_Y	0.7542
QILV_M	0.6789	QILV_M	0.6739	QILV_M	0.9218	MSE_M	0.7708
Marziliano_Y	0.5114	HPM_Y	0.4442	Marziliano_Y	1.0788	HPM_Y	0.8375
HPM_Y	0.4548	Marziliano_Y	0.4179	HPM_Y	1.1181	Marziliano_M	0.8375
Marziliano_M	0.4217	HPM_M	0.3679	Marziliano_M	1.1383	JND_St_Y	0.8417
HPM_M	0.4004	Marziliano_M	0.3378	HPM_M	1.1503	LAP_Y	0.8417
JND_St_Y	0.2975	GRAD_Y	0.2040	JND_St_Y	1.1985	AutoCorr_M	0.8458
LAP_Y	0.1824	GRAD_M	0.1869	LAP_Y	1.2343	RTBM_M	0.8458
VAR_Y	0.1736	VAR_M	0.1387	VAR_Y	1.2363	VAR_Y	0.8500
GRAD_M	0.1618	VAR_Y	0.1258	GRAD_M	1.2389	GRAD_M	0.8500
GRAD_Y	0.1599	RTBM_Y	0.1223	GRAD_Y	1.2397	FTM_Y	0.8500
VAR_M	0.1031	KurtZhang_Y	0.1044	VAR_M	1.2487	FTM_M	0.8500
LAP_M	0.0948	LAP_M	0.0858	LAP_M	1.2497	AutoCorr_Y	0.8500
RTBM_Y	0.0946	RTBM_M	0.0744	RTBM_Y	1.2498	JND_St_M	0.8542
AutoCorr_Y	0.0860	LAP_Y	0.0713	AutoCorr_Y	1.2507	JND_Y	0.8542
KurtZhang_Y	0.0803	KurtWav_M	0.0634	KurtZhang_Y	1.2513	JND_M	0.8542
AutoCorr_M	0.0609	KurtWav_Y	0.0596	AutoCorr_M	1.2530	GRAD_Y	0.8542
RTBM_M	0.0577	FTM_Y	0.0578	RTBM_M	1.2533	HPM_M	0.8542
FTM_Y	0.0560	AutoCorr_Y	0.0518	FTM_Y	1.2534	KurtZhang_M	0.8542
JND_St_M	0.0545	JND_M	0.0515	JND_St_M	1.2538	KurtWav_Y	0.8542
KurtWav_M	0.0422	JND_Y	0.0499	JND_M	1.2545	KurtWav_M	0.8542
JND_M	0.0361	KurtZhang_M	0.0357	JND_Y	1.2545	RTBM_Y	0.8542
JND_Y	0.0360	AutoCorr_M	0.0356	KurtWav_Y	1.2552	LAP_M	0.8625
KurtWav_Y	0.0143	JND_St_M	0.0321	KurtZhang_M	1.2553	KurtZhang_Y	0.8625
KurtZhang_M	0.0093	JND_St_Y	0.0313	FTM_M	1.2553	VAR_M	0.8667
FTM_M	0.0090	FTM_M	−0.0193	KurtWav_M	1.2553	Marziliano_Y	0.8708

Table C.4: Accuracy, consistency, and monotonicity indexes for each objective metric computed in the PQ space.

PCC		SROCC		RMSE		OR	
MSSSIM_Y	0.9380	MSSSIM_Y	0.9435	MSSSIM_Y	0.4366	VIF_Y	0.4917
VIFP_Y	0.9301	FSIM_Y	0.9361	VIFP_Y	0.4613	VIFP_Y	0.4958
VIF_Y	0.9292	VIF_Y	0.9272	VIF_Y	0.4646	IFC_Y	0.5333
FSIM_Y	0.9240	VIFP_Y	0.9242	FSIM_Y	0.4812	IWPSNR_Y	0.5458
SSIM_Y	0.9107	IFC_Y	0.9151	SSIM_Y	0.5188	IWSSIM_Y	0.5458
IFC_Y	0.9093	SSIM_Y	0.9117	IFC_Y	0.5243	MSSSIM_Y	0.5542
IWPSNR_Y	0.9025	IFC_M	0.9039	IWPSNR_Y	0.5407	WSNR_Y	0.5750
IWSSIM_Y	0.9025	IWPSNR_Y	0.9024	IWSSIM_Y	0.5407	SNR_Y	0.5792
IFC_M	0.8930	IWSSIM_Y	0.9024	IFC_M	0.5667	SSIM_Y	0.5792
WSNR_Y	0.8893	IWMSE_Y	0.9022	WSNR_Y	0.5743	UQI_Y	0.5875
DN	0.8887	DN	0.8917	DN	0.5768	PSNR_Y	0.5917
UQI_Y	0.8767	WSNR_Y	0.8890	UQI_Y	0.6039	FSIM_Y	0.6042
IWMSE_Y	0.8730	MSE_Y	0.8656	IWMSE_Y	0.6132	IWMSE_Y	0.6250
MSVD_Y	0.8604	PSNR_Y	0.8656	PSNR_Y	0.6400	DN	0.6375
PSNR_Y	0.8603	MSVD_Y	0.8646	MSVD_Y	0.6409	MSVD_Y	0.6417
SNR_Y	0.8511	UQI_Y	0.8603	SNR_Y	0.6592	IFC_M	0.6500
MSE_Y	0.8451	SNR_Y	0.8589	MSE_Y	0.6721	VIF_M	0.6542
UQI_M	0.8285	MSSSIM_M	0.8359	UQI_M	0.7037	VIFP_M	0.6583
VIF_M	0.8282	VIF_M	0.8358	VIF_M	0.7039	WSNR_M	0.6833
VIFP_M	0.8224	UQI_M	0.8344	VIFP_M	0.7143	UQI_M	0.6875
MSSSIM_M	0.8181	FSIM_M	0.8315	MSSSIM_M	0.7224	QILV_Y	0.6917
FSIM_M	0.8129	VIFP_M	0.8302	FSIM_M	0.7318	QILV_M	0.6917
WSNR_M	0.8047	WSNR_M	0.8127	WSNR_M	0.7455	MSE_M	0.7083
QILV_Y	0.7744	QILV_Y	0.7964	QILV_Y	0.7946	IWPSNR_M	0.7083
MSVD_M	0.7671	IWPSNR_M	0.7734	MSVD_M	0.8058	SSIM_M	0.7083
IWPSNR_M	0.7653	IWSSIM_M	0.7734	IWPSNR_M	0.8081	IWSSIM_M	0.7083
IWSSIM_M	0.7653	MSVD_M	0.7690	IWSSIM_M	0.8081	MSE_Y	0.7167
SSIM_M	0.7243	IWMSE_M	0.7368	SSIM_M	0.8655	MSSSIM_M	0.7292
MSE_M	0.7219	MSE_M	0.7362	MSE_M	0.8688	PSNR_M	0.7417
IWMSE_M	0.7125	SSIM_M	0.7359	IWMSE_M	0.8810	SNR_M	0.7417
SNR_M	0.7041	SNR_M	0.7117	SNR_M	0.8914	IWMSE_M	0.7458
PSNR_M	0.7007	PSNR_M	0.7088	PSNR_M	0.8956	MSVD_M	0.7500
QILV_M	0.6601	QILV_M	0.6411	QILV_M	0.9430	FSIM_M	0.7708
Marziliano_Y	0.5065	HPM_Y	0.4685	Marziliano_Y	1.0824	LAP_Y	0.8375
HPM_Y	0.4717	Marziliano_Y	0.4199	HPM_Y	1.1069	HPM_Y	0.8375
Marziliano_M	0.4213	HPM_M	0.3486	Marziliano_M	1.1385	Marziliano_M	0.8417
HPM_M	0.4108	Marziliano_M	0.3267	HPM_M	1.1445	AutoCorr_M	0.8458
LAP_Y	0.1929	GRAD_Y	0.2241	LAP_Y	1.2318	RTBM_M	0.8458
VAR_M	0.1797	GRAD_M	0.1917	VAR_M	1.2349	JND_St_Y	0.8500
GRAD_Y	0.1733	VAR_M	0.1483	GRAD_Y	1.2368	FTM_Y	0.8500
JND_St_Y	0.1603	VAR_Y	0.1273	GRAD_M	1.2403	HPM_M	0.8500
GRAD_M	0.1556	RTBM_Y	0.1177	JND_St_Y	1.2411	JND_St_M	0.8542
VAR_Y	0.1048	KurtZhang_Y	0.1062	VAR_Y	1.2485	JND_Y	0.8542
LAP_M	0.0978	LAP_M	0.1004	LAP_M	1.2493	JND_M	0.8542
KurtZhang_Y	0.0867	LAP_Y	0.0846	KurtZhang_Y	1.2506	GRAD_Y	0.8542
AutoCorr_Y	0.0724	KurtWav_M	0.0780	AutoCorr_Y	1.2521	GRAD_M	0.8542
RTBM_Y	0.0651	FTM_Y	0.0642	RTBM_Y	1.2527	FTM_M	0.8542
KurtWav_M	0.0587	RTBM_M	0.0639	KurtWav_M	1.2532	KurtZhang_M	0.8542
JND_Y	0.0373	AutoCorr_Y	0.0578	JND_Y	1.2545	KurtWav_Y	0.8542
FTM_Y	0.0372	KurtWav_Y	0.0546	FTM_Y	1.2545	KurtWav_M	0.8542
JND_M	0.0362	JND_M	0.0511	JND_M	1.2545	AutoCorr_Y	0.8542
AutoCorr_M	0.0333	JND_Y	0.0506	AutoCorr_M	1.2547	RTBM_Y	0.8542
RTBM_M	0.0324	KurtZhang_M	0.0358	RTBM_M	1.2547	VAR_M	0.8583
KurtWav_Y	0.0166	JND_St_Y	0.0310	KurtWav_Y	1.2552	Marziliano_Y	0.8583
JND_St_M	0.0128	AutoCorr_M	0.0264	JND_St_M	1.2552	KurtZhang_Y	0.8625
KurtZhang_M	0.0119	JND_St_M	−0.0238	KurtZhang_M	1.2553	LAP_M	0.8667
FTM_M	−0.0086	FTM_M	−0.0521	FTM_M	1.2553	VAR_Y	0.8708

Appendix D

Q Values Selection

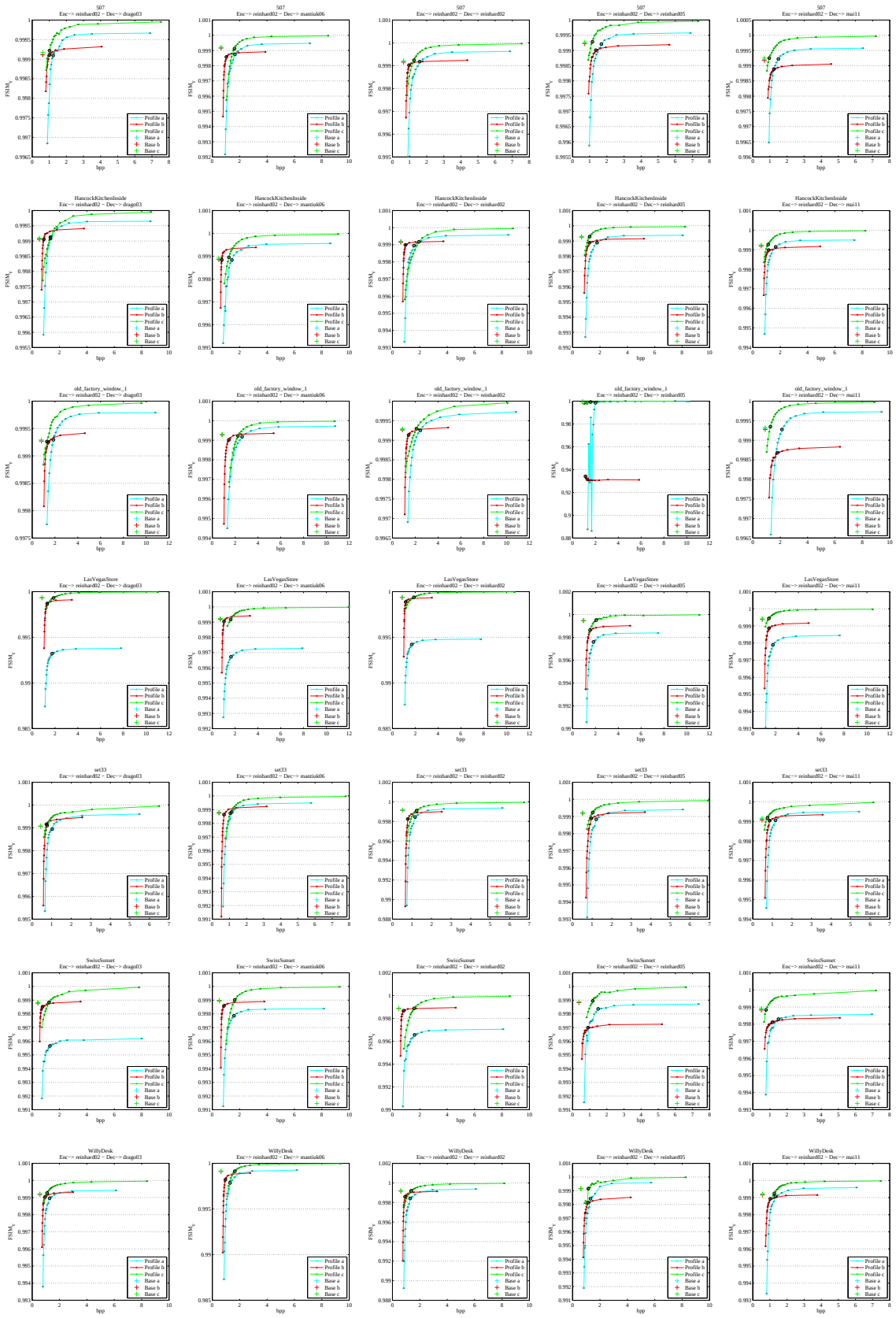


Figure D.1: Decoder FSIM result versus bit rate for $q = 75$ using *Reinhard* TMO in the base layer in all images.

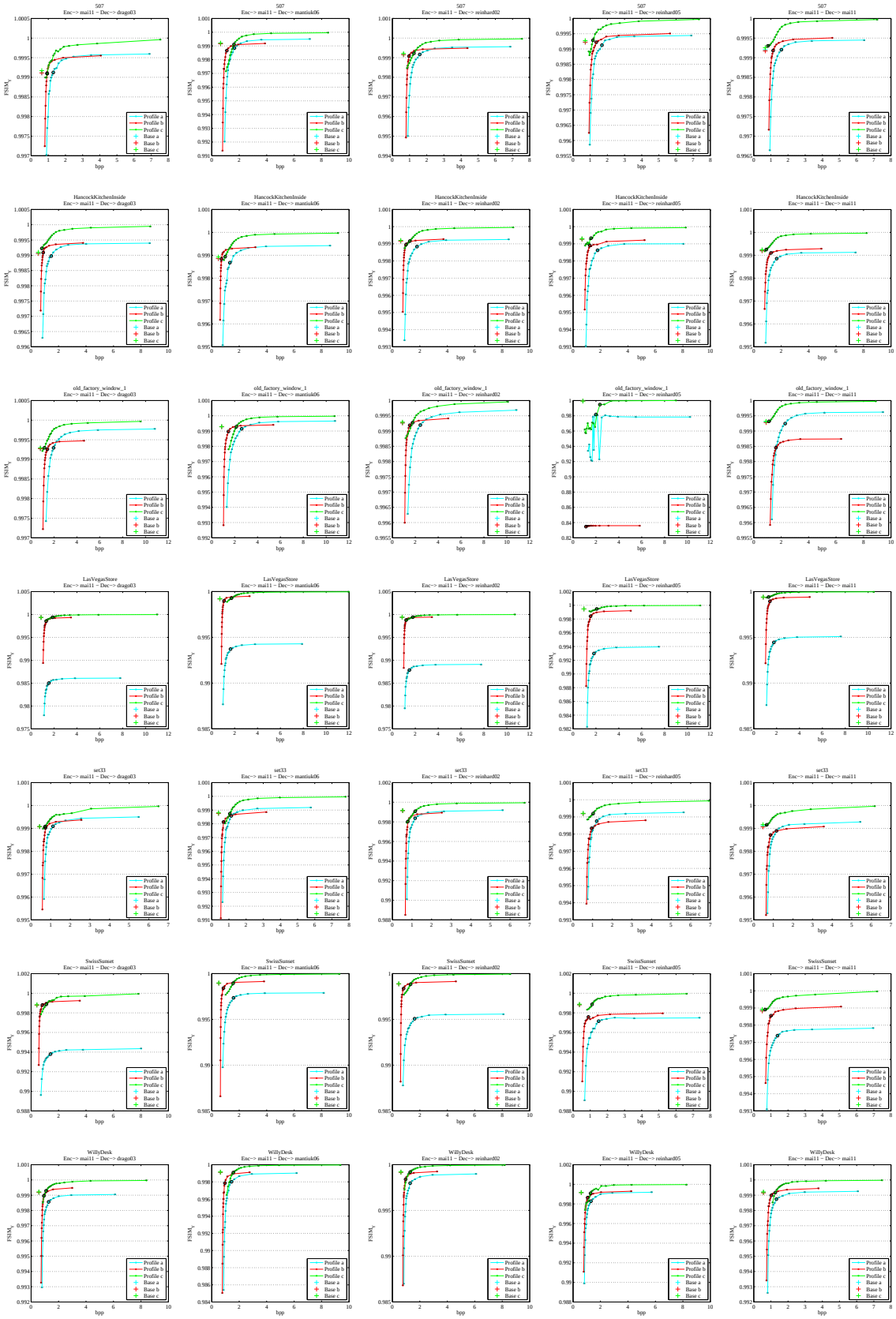


Figure D.2: Decoder FSIM result versus bit rate for $q = 75$ using *Mai* TMO in the base layer in all images.

Figure D.3: Q values for the closest quality to the base layer when *Reinhard* TMO is on the encoder. Each bar has the bit rate increment for each case.

Figure D.4: Q values for the closest quality to the base layer when *Mai* TMO is on the encoder. Each bar has the bit rate increment for each case.