HAL
open science

# Visuo-perceptual validation methods for physically based image synthesis

Victor Medina

## To cite this version:

Victor Medina. Visuo-perceptual validation methods for physically based image synthesis. Graphics [cs.GR]. Université Paris sciences et lettres, 2016. English. NNT : 2016PSLEM029 . tel-01585882

**HAL Id: tel-01585882**

**https://pastel.hal.science/tel-01585882**

Submitted on 12 Sep 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

**Préparée à MINES ParisTech**

# Visuo-perceptual validation methods for physically-based image synthesis.

Méthodes de validation visuo-perceptive en synthèse d'image physico-réaliste.

**Ecole doctorale n°432**

SCIENCES DES METIERS DE L'INGENIEUR

**Spécialité** Informatique temps-réel, robotique et automatique.

**COMPOSITION DU JURY :**

**Mme. Dominique LAFON-PHAM**
École des Mines d'Alès, Examinateur

**M. Philippe FUCHS**
MINES ParisTech, Examinateur

**M. Alexis PALJIC**
MINES ParisTech, Examinateur

**M. Marius PEDERSEN**
Norwegian University of Science and Technology, Rapporteur

**M. Ludovic MACAIRE**
LAGIS - FRE CNRS 3303, Université Lille 1, Rapporteur

**M. Alain TREMEAU**
Université Jean-Monnet (Saint-Etienne), President

**M. Christian PERROT-MINNOT**
PSA Peugeot-Citroën, Examinateur

**Soutenue par** Victor **MEDINA**
**le 23 mai 2016**

Dirigée par : **Philippe FUCHS**

# THÈSE DE DOCTORAT

Université de recherche Paris Sciences et Lettres
PSL Research University

préparée à
MINES ParisTech

Spécialité "Informatique temps-réel, robotique et automatique"

par **Victor MEDINA**

# VISUO-PERCEPTUAL VALIDATION METHODS FOR PHYSICALLY-BASED IMAGE SYNTHESIS

Méthodes de validation visuo-perceptive en synthèse d'image physico-réaliste

*Thèse CIFRE avec le groupe PSA Peugeot Citroën*

Directeurs de thèse:
**Dominique LAFON-PHAM (École des Mines d'Alès)**
**Philippe FUCHS (MINES ParisTech)**

Tuteur de thèse: **Alexis PALJIC**
Tuteur Industriel: **Philippe PORRAL**

version: August 23, 2016

PHD THESIS

Université de recherche Paris Sciences et Lettres
PSL Research University

prepared at
MINES ParisTech

Specialization: "Real-time informatics, robotics, and automatics"

by **Victor MEDINA**

# VISUO-PERCEPTUAL VALIDATION METHODS FOR PHYSICALLY-BASED IMAGE SYNTHESIS

Méthodes de validation visuo-perceptive en synthèse d'image physico-réaliste

*CIFRE PhD Thesis with the PSA Peugeot Citroën group*

Thesis directors:
**Dominique LAFON-PHAM (École des Mines d'Alès)**
**Philippe FUCHS (MINES ParisTech)**

Academic supervisor: **Alexis PALJIC**
Industrial supervisor: **Philippe PORRAL**

version: August 23, 2016

# Contents


**1 Introduction** **1**
1.1 Research framework . . . 1
1.2 Perceptually realistic rendering . . . 2
1.2.1 Definition of realistic image rendering . . . 4
1.2.2 Generation of color stimuli in real conditions . . . 6
1.2.3 Generation of color stimuli with a computer . . . 8
1.3 Thesis organization . . . 8

**2 Human visual perception of materials** **13**
2.1 Transforming electromagnetic waves into a visual response . . . 14
2.1.1 Electromagnetic radiation and visible light . . . 14
2.1.2 Mechanisms of the human eye for sensing and processing light . . . 15
2.2 Mathematical representation of color stimuli . . . 22
2.3 Visual adaptation to the environment . . . 25
2.3.1 Luminance adaptation . . . 25
2.3.2 Chromatic adaptation . . . 25
2.3.3 Modeling visual adaptation mathematically . . . 27
2.4 Measuring human visual acuity . . . 27
2.4.1 Spatial frequency and image resolution . . . 27
2.4.2 Contrast sensitivity . . . 29
2.4.3 Acuity of color vision . . . 29
2.5 Binocularity and depth perception in humans . . . 30
2.6 Vision as a temporal combination of partial views of a scene . . . 31
2.7 Conclusion . . . 31

**3 Ensuring color signal consistency in the visualization chain** **33**
3.1 Choosing the right display device . . . 34
3.1.1 Display Technologies . . . 34
3.1.2 Depth perception in displays . . . 35
3.1.3 Our choice . . . 37
3.2 Color signal transformations introduced by the display . . . 37
3.2.1 Previous considerations . . . 39
3.2.2 Modeling the response of the display . . . 48
3.3 Effects of stereoscopic visualization on the color signal . . . 55
3.4 Conclusion . . . 56

**4 Visual Psychophysics: Perception-guided display of sparkling materials** **59**
4.1 Description of the problem . . . 60
4.2 Perceptual characteristics of sparkling materials . . . 61
4.3 Modeling the human visual response with a digital camera . . . 62
4.3.1 Effects of camera optics on sparkle perception in photographs . . . 63
4.3.2 Obtaining radiometric information from the scene using photographs . . . 64
4.4 Perceptually-aware acquisition of photographs of sparkling materials . . . 68
4.4.1 Representing the real observation conditions in a photograph . . . 68
4.4.2 Adding binocular information to the photographs . . . 71
4.4.3 Image calibration: Adapting color information to the visualization environment 71
4.4.4 Effects of image calibration in the perception of sparkle in photographs . . . 73
4.5 Perceptual reproduction of luminance dynamics in photographs of sparkling materials. 75

4.5.1 Perceptual effects of image exposure in photographs of sparkling materials . . 75
4.5.2 Perceptual effects of dynamic range compression in photographs of sparkling materials . . . . . 80

**5 Objective validations of the rendering model against photographs of real scenes 87**
5.1 Introduction . . . . . 88
5.2 Setup and characterization of a controlled test scene . . . . . 89
5.3 Data acquisition from test scene photographs . . . . . 91
5.4 Simulation of the scene in the target rendering engine . . . . . 91
5.5 Objective comparison of photographs with simulations . . . . . 92
5.6 Conclusion . . . . . 99

**6 Discussion and future work 103**

**A Appendix 1: Glossary 107**

**B Appendix 2: Scientific production 115**
B.1 International conference articles . . . . . 115
B.2 Oral presentations . . . . . 116
B.3 Poster presentations . . . . . 116

**C Appendix 3: Additional data 117**
C.1 Color set used for the display characterization. . . . . 117

**Index 119**

**Bibliography 123**

# List of Figures


1.1 Several types of physical samples for automobile paints. . . . . 2
1.2 LIMA Project: general task organization . . . . . 3
1.3 Functional realism in technical illustrations. . . . . 5
1.4 Optical light-matter interaction phenomena. . . . . 7
1.5 Stages of the simulation process . . . . . 9
1.6 Visuo-perceptual validation methodology comparing photographs with simulations. . 11

2.1 Electromagnetic light wave model. . . . . 14
2.2 Scalar simplification of the electromagnetic wave model. . . . . 15
2.3 Electromagnetic spectrum. . . . . 15
2.4 Diagram of a human eye. . . . . 16
2.5 Diagram of the human optic nerves. . . . . 17
2.6 Diagram of the retinal cell structure. . . . . 18
2.7 Representation of the On-center/Off-center ganglion cells mechanisms. . . . . 19
2.8 Density distribution of cones and rods in the retina. . . . . 20
2.9 Normalized human photoreceptor absorbances for different wavelengths of light. . . . 20
2.10 CIE spectral luminous efficiency functions. . . . . 21
2.11 Boynton's color vision model diagram. . . . . 22
2.12 CIE 1931 RGB and XYZ color matching functions. . . . . 25
2.13 Dark adaptation curve . . . . . 26
2.14 Illuminant color shift. . . . . 26
2.15 Snellen chart for visual acuity . . . . . 28
2.16 Conversion between Snellen's notation and spatial frequency . . . . . 29
2.17 Simultaneous contrast . . . . . 30
2.18 Spatial frequency response of opponent color channels . . . . . 30

3.1 Color gamut comparison of stereoscopic vs monoscopic monitors. . . . . 38
3.2 Monitor characterization environment. . . . . 39
3.3 Gamut comparison of the Asus VG248QE preset settings. . . . . 40
3.4 Gamut comparison of the Asus VG248QE with the stereoscopic mode On and Off. . 41
3.5 Luminance comparison of the Asus VG248QE with stereoscopic mode On and Off. . 42
3.6 Monitor characterization setup. . . . . 43
3.7 Asus VG248QE screen luminance uniformity measurements. . . . . 43
3.8 Display stabilization graphs. . . . . 44
3.9 Asus VG248QE channel constancy graphs. . . . . 45
3.10 Asus VG248QE channel additivity test graph. . . . . 46
3.11 Luminance distributions for the red, green, blue, and gray input series. . . . . 47
3.12 CIE 1931 chromaticity coordinates of all the characterization color samples. . . . . 50
3.13 Spectral radiance curves of the Asus VG248QE RGB primary colors. . . . . 52
3.14 Zoom of Figure 3.13. . . . . 53
3.15 Average $R^2$ in XYZ-DAC estimation with models GOG, GOGO, and GGO. . . . . 54
3.16 Average $R^2$ of DAC-XYZ estimation with models GOG, GOGO, and GGO. . . . . 54
3.17 Asus VG248QE RGB correlation between the real and estimated RGB input values with the GOG model. . . . . 55
3.18 Normalized spectral transmittance of the Nvidia 3D shutter glasses. . . . . 56
3.19 Stereoscopic characterization with a spectrometer and the Nvidia 3D shutter glasses. 56

4.1 Material structure diagram and microscopic closeup view of a paint layer. . . . . 62

4.2 Zoomed detail of a paint sample with a x700 factor. . . . . 64
4.3 Intensity profile of an image area around a sparkle pixel as a function of the total exposure. . . . . 65
4.4 Acquisition setup for sample photographs. . . . . 69
4.5 Custom support for stereoscopic sample acquisition. . . . . 70
4.6 Methods for stereo-pair creation: Off-axis frustum vs Toe-in. . . . . 71
4.7 Illustration of positive parallax. . . . . 72
4.8 Illustration of zero parallax. . . . . 72
4.9 Raw image processing workflow from sensor data to display input. . . . . 73
4.10 Effects of a display's low-dynamic range on calibrated image contrast (illustration). . 74
4.11 Effects of a display's low-dynamic range on calibrated image contrast (photograph). 74
4.12 Visualization setup for the comparison of virtual images with real references. . . . . 76
4.13 Paint samples used in the experiments. . . . . 77
4.14 Average image similarity scores with respect to a real reference for several exposure times. . . . . 79
4.15 Paint sample similarity scores with the real reference. . . . . 79
4.16 Difficulty to discern each sample image within the same series. . . . . 80
4.17 Three suggested scenarios to compare virtual and real images. . . . . 81
4.18 Logarithmic Bradley-Terry scores for different tone-mapped samples. . . . . 84

5.1 Visuo-perceptual validation methodology comparing photographs with simulations (Figure 1.6-Bis). . . . . 88
5.2 Spectral power distribution of the Solux lights. . . . . 90
5.3 Acquisition light cabin. . . . . 90
5.4 Wireframe views of the validation scene for comparing photographs with simulations. 92
5.5 Render preview of the scene used for the validations. . . . . 93
5.6 Spectral sensitivity of the Nikon D800. . . . . 93
5.7 Photograph and simulation of a Macbeth color chart with a shutter speed of 1/8 and 1/15 seconds. . . . . 94
5.8 Luminance of the Macbeth color chart for the camera characterization process. . . . 95
5.9 Luminance comparison between a photograph (at a 1/8-second shutter speed) and its simulation, after scaling by a luminance factor. . . . . 96
5.10 Luminance comparison between a photograph (at a 1/15-second shutter speed) and its simulation, after scaling by a luminance factor. . . . . 97
5.11 Graphical representation of the CIEDE2000 color differences between the photograph and the simulated images in Figure 5.7. . . . . 98
5.12 Photograph and simulation of a sample green paint plate. . . . . 98

C.1 Characterization color samples. . . . . 117

# List of Tables


3.1 Technical specifications of the Asus VG248QE monitor. . . . . 39
3.2 Asus VG248QE settings used for its characterization and the observation tests. . . . 40
3.3 Individual and average $R^2$ in XYZ-DAC estimation with models GOG, GOGO, and GGO. . . . . 51
3.4 Individual and average $R^2$ of DAC-XYZ estimation with models GOG, GOGO, and GGO. . . . . 51
3.5 Asus VG248QE characterization results with the forward GOG model. . . . . 53
3.6 Asus VG248QE characterization results with the inverse GOG model. . . . . 55

4.1 Fixed parameters for the characterization of the Nikon D800. . . . . 68
4.2 Exposure sets with variable luminance for the characterization of the Nikon D800. . 68
4.3 Characteristics of the image series used in the tone-reproduction experiment. . . . . 82
4.4 Bradley-Terry scores and Log scores for different tone-mapped samples. . . . . 84

5.1 Fixed Nikon D800 parameters for the comparison of photographs and simulations. . 91
5.2 xyY tristimulus values of the Macbeth color samples and chromaticity distances. . . 100
5.3 CIEDE2000 color differences between the photograph and the simulated images in Figure 5.7. . . . . 101
5.4 xyY tristimulus values and chromaticity distances of the photograph and the simulated images in Figure 5.12. . . . . 101
5.5 CIEDE2000 color differences between the photograph and the simulated images in Figure 5.12. . . . . 101

C.1 RGB tristimulus values of the 120 characterization color samples. . . . . 118

Chapter 1

# Introduction

## Contents


**1.1 Research framework** . . . . . **1**
**1.2 Perceptually realistic rendering** . . . . . **2**
1.2.1 Definition of realistic image rendering . . . . . 4
1.2.2 Generation of color stimuli in real conditions . . . . . 6
1.2.2.1 Additive color synthesis . . . . . 6
1.2.2.2 Subtractive color synthesis . . . . . 6
1.2.3 Generation of color stimuli with a computer . . . . . 8
**1.3 Thesis organization** . . . . . **8**


## 1.1 Research framework

This thesis is a part of a french national research project[1], named LIMA (Light, Interaction, Matter, and Aspects) project, which is a collaboration between 3 research laboratories of École Nationale Supérieure des Mines de Paris, 3 SMEs[2], and an industrial partner, the PSA Peugeot Citröen group. The LIMA project ran from January 2011 to January 2016; this thesis started in March 1st, 2013, financed within the framework of a CIFRE (*Conventions Industrielles de Formation par la Recherche*) convention. The goal of the LIMA project is to create perceptually realistic simulations of paint materials with specific known properties. Interactive head-tracking and stereoscopic visualization are suggested as ways to improve perceptual realism. We will say that the simulations are perceptually-realistic when they trigger the same visual response on a human observer as the physical instances, under the same observation conditions. Our research takes place in the context of automobile paints, with a special focus on complex sparkling coating materials.

The aspect of paint materials is a very important factor in industrial product design, more so after the proliferation of complex finishes such as those used in pearlescent, sparkling, glossy, or goniochromatic materials. The use of simulation applications for product design introduces important cost and time reductions, in addition to offering greater control and flexibility to designers in the creative process; however, replacing the visual evaluation of physical samples or prototypes (Figure 1.1) by virtual representations of the product also carries important challenges. Industrial design applications rely strongly on the aspect of the product, so it is paramount to ensure that the product simulation is accurate with respect to the physical product.[1]

The production of industrial materials requires the participation of three main agents: manufacturer, designer and final user. In digital simulations, the manufacturing process defines the physical composition of the materials, which often entails the modelization at several levels (chemical, mechanical, mathematical, optical, etc) of the material's microstructure in terms of spatio-chromatic distribution. This model is then used by a rendering engine to produce a color image representation

[1] ANR (Agence Nationale de la Recherche) grant #2011 RMNP 01401
[2] Small and medium-sized enterprises

Figure 1.1: Several types of physical samples for automobile paints.

of the material, under some predefined environment, which simulates the optical properties of the material's microstructure in an accurate, or realistic, manner. Realism is defined based on perceptual criteria that relate perceptual image attributes to spatio-chromatic properties which, in turn, are related to properties of the material composition.

The LIMA project proposes a double approach, from material composition to material aspect, and vice versa. Consequently, visual simulations of the materials could be generated from their composition and, conversely, their composition be predicted from a desired visual aspect. We can distinguish three main lines of work: material measurement and modeling, simulation rendering, and perceptual validation of the results (Figure 1.2). First, optical measurements are taken from each individual physical component of the paint coating materials, and used to describe and model their morphological microstructure [2]. At the same time, the same physical base components are combined to produce real reference samples [3], which are later used for physical validation. A rendering engine integrates these models to generate simulations of the materials, using virtual reality techniques like head tracking and stereoscopy to provide dynamic and interactive visualization [4]. This allows the user to move around the objects and view them from different angles, reproducing more accurately the real observation conditions. In this context, the work presented in this thesis deals with the perceptual validation of the simulations: we use feedback from visual evaluations by human observers, under some controlled conditions (illumination, observation, display, etc.), to ensure the perceptual realism of the simulations with respect to the real samples.

## 1.2 Perceptually realistic rendering

When we observe an object illuminated by one or several light sources, its surface is hit by some portion of the electromagnetic waves emitted by these light sources. A series of optical interactions then take place between the light and the object, resulting —in the case of non-fully absorbent objects— in a series of secondary signals re-emitted by the object, some of which will reach the eye of the observer. When this light arrives in the eye, it is captured by the photoreceptors in the retina, and transmitted to deeper processing stages. These further stages process the perceived signal in terms of higher-level criteria, using information about the environment and our personal prior knowledge. Image synthesis, or image rendering, is the process by which a computer simulates these processes, to generate an image that reproduces, to some extent, the information present in the real scene.

The first part of this section discusses different interpretations of realistic image rendering, using objective criteria to evaluate the realism of an image. In this context, we define a perceptually realistic object simulation as one that produces the same visual response in the observer as the physical object. The following section discusses different ways in which the light can interact with the objects in a real environment to produce visual stimuli, and how that affects our visual response

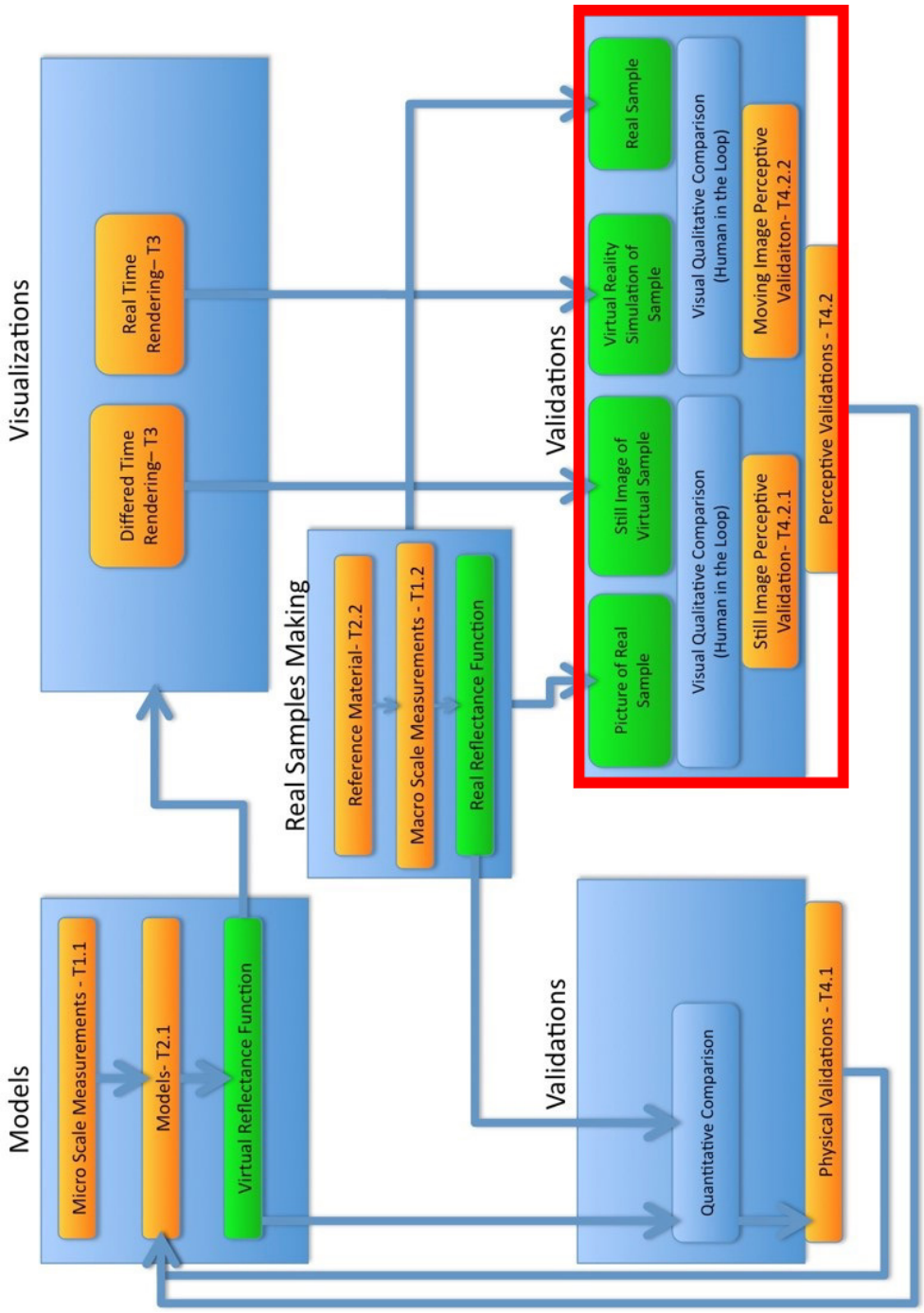



Figure 1.2: LIMA Project: general task organization. The bottom-right block, marked in red, concerns the perceptual validations.

when we observe the environment. We will finish by describing the process to generate and display visual stimuli with a computer.

### 1.2.1 Definition of realistic image rendering

Image synthesis, or image rendering, is the process by which computers can generate a digital representation of a given environment. These images represent selected properties of a real scene with varying degrees of realism. However, although extensive research work has been done in the area of realistic image rendering, one of the most important issues remains the lack of common criteria to define and measure the level of realism [5]. We can find in the literature many different interpretations of realism, attending to various criteria. Some interpretations require the existence of a reference, whereas others simply judge the realism of a scene or object in itself. Authors Shirley et al. [6] and Ferwerda [5] present two overlapping, but different, classifications of special interest for our work.

Shirley et al. [6] distinguish the following three rendering approaches attending to the accuracy of the physics in the scene:

- **Predictive rendering:** the results must **match the real scene** (design and simulation applications).
- **Plausible rendering: plausibly real** look, but the physics can be wrong (entertainment and visualization applications).
- **Visually-rich rendering:** visual richness of reality, but can be **highly stylized**.

In this work we are concerned by the first and second types of rendering. We do not care for the aesthetic aspect of a scene so much as for how it is constructed according to a set of input models and physics laws. This rendering paradigm is known as physically-based rendering (PBR). In the past, PBR required a big amount of computation time and resources, but the latest technological advances have reduced these computation demands drastically. However, as the complexity of a scene increases, the amount of computation needed to simulate a high-fidelity scene is such that even modern GPUs are incapable of doing it in reasonable time. This problem becomes specially important in interactive simulation applications such as ours. The key to reducing computation times, as discussed by Chalmers et al. [7], is to keep in mind that the scenes are rendered for a human observer, so that the amount of required information is limited by the capabilities of the human visual system (HVS).

This takes us to the next classification, provided by Ferwerda [5], based on the type of visual information depicted in the scene:

- **Physical realism:** the image provides the same **visual stimulation** as the real scene.
- **Photo-realism:** the image produces the same **visual response** as the real scene.
- **Functional realism:** the image provides the same **visual information** as the real scene.

In this second classification we can identify physical realism with predictive rendering, or more specifically, PBR; however, there is no correspondence between the remaining categories in the two classifications —which is why we provide both classifications here. We find an additional variety of realism in this classification: *functional realism*. Functional realism refers to a type of realism where, even if the simulation does not resemble the real scene as much (stylized, sketched, etc), it still provides the same (or even more) information needed to perform a given task. This is illustrated in Figure 1.3. As explained by Ferwerda, although the photographs in this illustration are sufficiently clear, one could argue if they are preferred in terms of providing information because

"the drawings offer a number of benefits over the photos. First, the drawings can eliminate irrelevant details produced (in this case) by shading, shadows, and surface texture. Second, the drawings can facilitate visual segmentation and grouping by color (e.g. hand white, tools gray, wood tan). Third, the drawings make it possible to show viewpoints that would be difficult or impossible in a photograph, such as the edge shot shown in the first panel or the point-of-view shot shown in the second. Finally, the rendering of the saw blade shows that the drawings can make use of 'special effects' like artificial transparency to depict important features that would be hidden in photographs" [5].

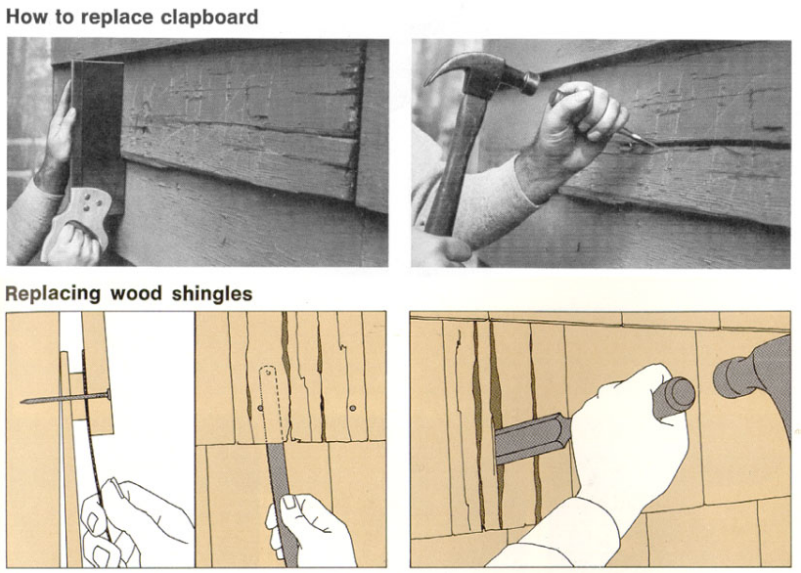



Figure 1.3: Functional realism in technical illustrations. *Adapted by Ferwerda [5] from the Reader's Digest Complete Do-It Yourself Manual [8].*

As we mentioned above, physically-realistic rendering has several disadvantages: the highly demanding computation requirements of PBR; the low capabilities of current conventional displays in terms of tone reproduction; and the fact that producing physically-realistic images for human observers is excessive, since it ignores the limitations of the HVS. All of these factors, reduce the suitability of a strictly physics-based approach for interactive simulation applications like ours. For this reason, we find the latter classification —proposed by Ferwerda— more appropriate for our purposes.

When a human observer looks at the simulations produced by our rendering model, the perceived information should be the same as if they were looking at the actual physical objects, under the same environment. In other words, the response produced by the observer's visual system should be the same when looking at the simulation and the real scene under the same conditions. This definition of perceptual realism can be identified with Ferwerda's definition of photo-realism, but also with that of functional realism. Indeed, human visual perception is known to be selective and, often, biased and dependent on factors like the context or the perceiver's motivation [9, 10]. For this reason, we use immersive and stereoscopic visualization to introduce functional information in the simulations with the aim of improving their perceptual realism. With this in mind, we will use the term "perceptual realism" in this manuscript to refer to realistic material simulations in the sense of material simulations that produce the same visual response in the observer as the real material, under predefined environmental and observation conditions.

### 1.2.2 Generation of color stimuli in real conditions

Color stimuli can be produced directly from the radiant energy emitted by a light source, or indirectly, as the result of the interaction of that radiant energy with an object. When a color stimulus is produced by a light source, we speak of light-source colors, and the resulting visual response is produced by additive synthesis of the contribution from each individual source. When the color stimulus results from the interaction of some illuminating light with an object, we speak of object colors, and the visual response is the result of a subtractive synthesis.

#### 1.2.2.1 Additive color synthesis

Light radiation is a combination of waves at multiple wavelengths, known as the light spectrum, as we will explain in Section 2.1.1. Because each wavelength corresponds to one color, the perceived color of the emitted radiation is the result of adding the contribution of each individual wavelength, which can be seen as mixing multiple monochromatic lights (light with only one wavelength, or a narrow band of wavelengths). A color resulting from the addition of multiple lights is said to be produced by additive color synthesis.

Color stimuli produced by additive synthesis can be measured in two different ways, depending on the use: directly or by reflection. The direct (or incident) measurement method is done with the meter facing the source, whereas the reflection method measures the light reflected by a near-perfect white diffuser (or barium white)[3] positioned in front of the light source. Multiple devices may be used to perform this measurement, but the most common ones are[4]:

- Spectroradiometer: the most complete device, it measures the spectral power distribution (SPD)[5] of the radiance emitted by surface.
- Spectrocolorimeter: measures the spectral reflectance, transmittance, or relative irradiance of a color sample [11], and calculates its tristimulus values.
- Tristimulus colorimeter: measures the chromaticity coordinates of a color surface.

#### 1.2.2.2 Subtractive color synthesis

When light hits a (non-fully absorbent) solid object, a color is produced as the result of its interaction with the object's material. The color resulting from said interaction depends on the object's properties, but also on the illumination and observation angles. A light incident on an object's surface can be subject to many different physical phenomena, but the most relevant in this context are those shown in Figure 1.4.

Some wavelengths of the incident light is reflected by the object, bouncing off the surface, in two forms: specular and diffuse reflections. In perfectly planar surfaces, the specular component is reflected with the same angle in a direction opposite to the incident light; the diffuse component, however, is reflected at multiple angles and magnitudes (luminance). Under most observations conditions —except when the observer is at the angle of maximum specularity— the diffuse component is what creates the primary color sensation, since it provides most of the light rays arriving to the retina of the eyes. Some wavelengths of the incident light may also be absorbed by the object, disappearing. Finally the remaining light is refracted through the object at different angles —due to differences between the refraction indices of the object's material (n2) and the surrounding medium

[3] A near-perfect white diffuser is a surface —made of barium sulfate— whose reflectance approximates a lambertian reflectance, a theoretical surface that reflects back all the light that it receives.

[4] See Appendix A for a more complete definition of these devices.

[5] Radiant flux ($W/sr/m^2$) or luminous flux ($lm/m^2$) per unit wavelength (See a more complete definition in Section A.12 of the glossary)

(n1), as indicated by the Snell-Descartes law (Section A.16 of the glossary)— and transmitted out of the object.

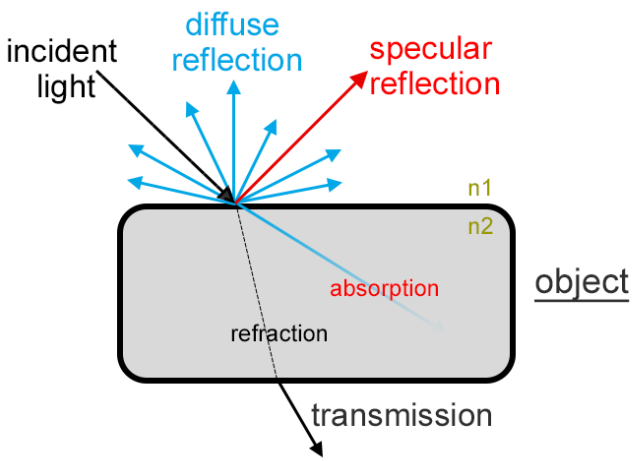



Figure 1.4: Optical light-matter interaction phenomena that may take place when light reaches the surface of an object (with refraction index n2), from another medium (with refraction index n1).

Since the colors produced in this way are the result of subtracting components from the illuminating light[6], via one or several of the phenomena mentioned above, this process is called subtractive color synthesis. A fully-absorbent object, or an object which is not illuminated, will not reflect any light, and it will be seen as black. Conversely, an object that reflects every wavelength of the illuminating light in its entirety (lambertian reflectance) will be seen as having the same color of the illuminant. When an object is painted a certain color, the properties of its surface are modified by pigments in the paint that act as a filter, absorbing certain wavelength components and reflecting others. Therefore, subtractive color synthesis requires the existence of light to produce color stimuli —object's colors cannot be seen in the absence of light. In order to measure subtractive color stimuli, we must always characterize two components: the optical behavior of the object's surface and the illumination. The optical behavior of a surface is always given as the percentage of the illumination that is affected by each phenomenon (reflection, absorption,...) per unit wavelength; the illumination is specified by its SPD.

In the case of perfectly planar surfaces, the characterization of the optical properties of an object is simple; but the computation becomes more difficult for irregular surfaces. In that case, more complex functions such as the bidirectional reflectance distribution function (BRDF) are used. The BRDF (which has units inverse steradians, $sr^{-1}$) describes the amount (proportion) of light reflected when light makes contact with a material, as a function of the illumination and observation angles, relative to the surface normal and tangent, and the wavelength of the light [12]. This is indicated by Expression 1.1, where $\omega_i$ is a vector representing the incoming direction of the light, $\omega_r$ represents the outgoing or reflection direction of the light, L is radiance ($W \cdot sr^{-1} \cdot m^{-2}$), and E is irradiance ($W \cdot m^{-2}$).

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{dE_i(\omega_i)} \tag{1.1}$$

[6]Since it is out of the scope of this research, we will not consider here the special case of luminescent objects.

### 1.2.3 Generation of color stimuli with a computer

The process of image rendering generates a digital representation of the information in a real scene. Since the rendered images are meant to be viewed on a display, the computed information has to be adapted to the capabilities of the reproduction medium. Depending on the display technology used to visualize the images, different methods can be used to display the color information present in an image (see Chapter 3). Traditionally, displays —such as cathodic raytube (CRT) displays— have used a set of three colored-light sources to create the color for each pixel, and modulate the intensity of each one to obtain the necessary proportion that produces the desired color. Practically any color can be produced by combining a set of three primary lights; displays typically chose red, green, and blue, as primary lights, to emulate the human visual system as described by the trichromatic theory (explained in Section 2.1.2.4). As display technologies have evolved, this display model has become more complex. For example, as we will see in Section 3.1.1, modern liquid-crystal displays (LCD) use more advanced techniques like backlight illumination with multi-color filters, or color dithering (integration of the color response over time) to display colors in the screen.

The presence of different electronic and optical components —like filters, light diffusers, etc— inside the display device obscures the computation of the synthesized color stimuli. However, in practice, a display can be seen as a "black box" where colors are essentially produced by means of additive synthesis using different color lights. When a display is observed in a perfectly dark environment, our visual response will depend exclusively on the stimuli generated by the display. In the presence of ambient light, however, our visual response will depend on many other factors, such as the displayed color stimuli, the ambient illumination, the contribution from the environment, and the properties of the display's surface (absorption, reflection, etc). This is the reason why displays must be calibrated for each observation environment.

## 1.3 Thesis organization

Physically-based rendering, or PBR, also known as predictive rendering, aims at creating images that produce the exact same radiometric information as the real scene that they simulate. Since the simulations contains the same information as in reality, one might think that the simulation is realistic. This is partly true and, in fact, this type of precision is necessary in some cases (remote sensing, multi-spectral imaging, etc); however, when the images are made to be viewed by humans, there are additional constraints that must be taken into account to produce realistic simulations. The human visual system (HVS) has a limited acuity that depends on many factors such as contrast, frequency, shape, or illuminant [13, 14, 15, 16, 17]. As a result, the original radiometric signals that hit our eyes are compressed and transformed at different stages of our visual system, before a visual response is created in our brain. This means that a lot of the visual information present in the real scene cannot be perceived by a human observer and, therefore, it is unnecessary (and, at times, even wrong, as we will see in Section 4.1) to compute it. Furthermore, since images are visualized on a display, additional constraints are introduced by the capabilities and limitations of the display and visualization devices participating in the process (Figure 1.5). Taking all these constraints into account in the rendering process allows us to find and resolve aspects that are not properly dealt with by the rendering engine, as well as to identify perceptually superfluous information that may be removed from the computation. Consequently, we can reduce the amount of image information and computation times, without affecting the visual response produced by the scene for a human observer. We have defined this type of rendering as perceptually realistic rendering. By defining accuracy in terms of perceptual realism, we can optimize the rendering model to produce images much closer than PBR to what we see in reality while reducing computation time. However, given the complexity of human visual perception, and the large amount of factors that take part in the process, the addition of perceptual constraints to a physically-based rendering model might lead to incorrect results. For this reason, a perceptual validation process is needed to verify that the

simulations remain perceptually correct after introducing these simplifications.

One big issue in the context of perceptual realism is the problem of *tone reproduction* (or *tone mapping*). This problem consists in finding how to represent the light richness of the physical reality on a display device that is much more limited in terms of reproducible light ranges (dynamic range, or DR). Ample research exists in this area [18, 19], and there are multiple models to reproduce the appearance of a scene with a more limited dynamic range; however, the use of generic reproduction models is not desired in our case, given the distinct characteristics of the materials that we will study. In order to control the realism of the results, we must first be able to understand the relationship between light and perceived material aspect in our specific context, and how it affects our perception of the materials.

Chapter 2 describes human visual perception as the result of a series of complex processes distributed along different stages of the visual system. As we will see, the light emitted by the objects in the scene that we are observing arrives into our eyes, and is captured and converted into sensory information by a series of selective photosensitive cells distributed on its surface. This information is transmitted via several channels to deeper stages of the visual system. It is then interpreted by different physiological mechanisms, responsible for producing visual attributes like contrast, color constancy, depth, etc. Each of these mechanisms produce a series of neurological responses that are interpreted and combined in the brain to construct a global visual response.

This work is part of a research project that integrates material models into a physico-realistic rendering model, as described by Da Graça et al [20, 4]. As we will see in Chapter 2, physical representation and perceptual (sensory) appearance are connected by means of the colorimetric space used to express the color information in the image. Representing image information using perceptual spaces such as those directly related to cone-response values (e.g. LMS) or device-independent spaces like XYZ —closely related to radiometric values entering the retina— gives us important information about the relationship between analyzed image parameters and visual attributes. This information may be used in determining the amount of contribution of said attributes to the appearance of the final images, identifying relevant information to maximize perceptual similarity with the original scene.

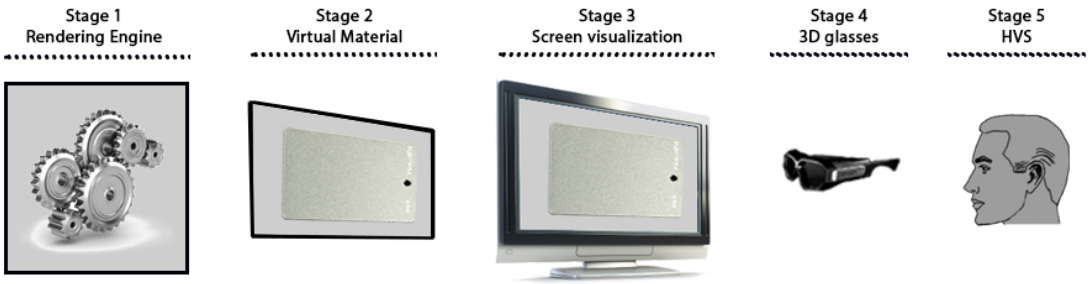



Figure 1.5: Stages of the simulation process

The rendering process produces images which are viewed on a display. Therefore, our visual system is not the only element that affects the perception of the images after the simulation. In effect, choosing the right display device, as well as analyzing and modeling its characteristics in the right way, is also crucial to control the visualization process. Chapter 3 presents an overview of the most common display technologies, the effects that each technology has on visualization, and the most important criteria to select a display for our purposes. The same chapter also describes the process to fully characterize a computer display, showing characterization results from our specific model. A proper display characterization will allow us to understand the transformations that take place inside the display, and to control its colorimetric response, favoring the conversion between

representation spaces without altering the original perceptual information.

Because we use a two-dimensional (2D) medium (a display) to reproduce a dynamic[7] three-dimensional (3D) scene, part of the depth information provided by binocular visual cues, such as stereopsis (depth by triangulation) or eye convergence, is lost during the simulation —both at rendering and visualization time. This becomes specially true in the case of sparkling coating materials. The intensity of the sparkles generated by the metal flakes —and therefore the overall material appearance— in sparkling coatings depends on factors such as the depth and size of the metal flakes, and the illumination and observation angles. Given the separation between our eyes, we may perceive a different image with each eye when we look at one point in space, because the observation angle is slightly different for each one. If we ignore this information, we will be neglecting a very important aspect of the material, and therefore the simulation will lose accuracy.

To overcome some of these disadvantages and improve perceptual realism, we use stereoscopic visualization to provide additional depth information from binocular disparity. As we will show later on, our tests show that in all of our cases observers find that stereoscopic visualization produces a visual response that is much closer to the reference than monoscopic visualization. In this sense, Chapter 3 provides an analysis of the most important technologies to produce depth perception in displays, to serve as the basis to justify our choice of stereoscopy. Similarly, the characterization methodology is extended to incorporate criteria specific to stereoscopic displays.

Chapter 4 presents an analysis of the particularities of human visual perception when observing sparkling materials. We will use digital photographs to reproduce the integrating function of the human eye in terms of trichromatic photoreceptor responses, showing how the sparkling effect in metallic flaked paints is strongly linked to the incident light. This effect is the macrometric optical manifestation of a physical phenomenon, which results from the interaction at micrometric level between light and matter (i.e. micrometric flakes). Since the digital camera emulates the mechanisms of the human visual system to produce visual responses (respectively, images), it must be possible to control the imaging process to maximize perceptual similarity between the visual response of an observer looking at the real scene and the resulting photograph of the scene.

The sparkling effect in metallic flaked paints is the macrometric optical manifestation of a physical phenomenon, which results from the interaction at micrometric level between light and matter (i.e. micrometric flakes). When the material is illuminated, the light bouncing off the metal flakes is much stronger due to their high specular reflectivity. When these specular reflections reach the camera lens and pass through its optical components, they are spread, producing the optical effect associated with sparkling. Therefore, due to this light spreading, the sparkles may be perceived as being much larger than the actual metal flakes at their origin. Consequently, we can then see that the sparkling effect is strongly linked to the amount and intensity of the incident light, which in turn determines the intensity of the light reflected by the metal flakes, or sparkle luminance, as well as the global luminance (brightness) of the image.

The camera sensor is not capable of capturing the vast luminance dynamics (values) of a real scene into an image, so we must introduce simplifications in the image, to produce a similar visual response with less information. We can maximize the amount of perceptual information present in the images, by selecting those camera settings that best reproduce human visual perception under the experimentation conditions. By fixing all imaging parameters but the shutter speed, we can establish a direct relationship between image luminance and exposure time, under those same conditions. We have conducted an experimental psychophysics study, where human observers were presented a series of stereoscopic images taken with different exposures, and asked to evaluate their similarity to a physical reference. The results from this experiment show that image luminance has indeed an effect on its brightness and contrast, and that this affects the perceived size of the

[7] As we will see in Section 2.6 real scenes, unlike synthetic scenes, are dynamic in the sense that our perception of the scene changes constantly due to the temporal integration properties of our visual system, which is periodically examining the scene for changes.

sparkling, texture depth, and sparkling-background contrast. Furthermore, they suggest optimal global luminance values to maximize perceptual similarity with the physical references.

A second psychophysics study has been performed, to analyze the perceptual effects of modifying the range of luminance values in the images, with respect to the found optimal global luminance, as well as the dynamic range (number of luminance values). The results suggest that indeed the global luminance found in the first experiment seems to maximize perceptual similarity, whereas no improvements result from using different luminance or dynamic ranges. Furthermore, under the selected luminance conditions, there seems to be a inverse relation between sparkling contrast and global image brightness.

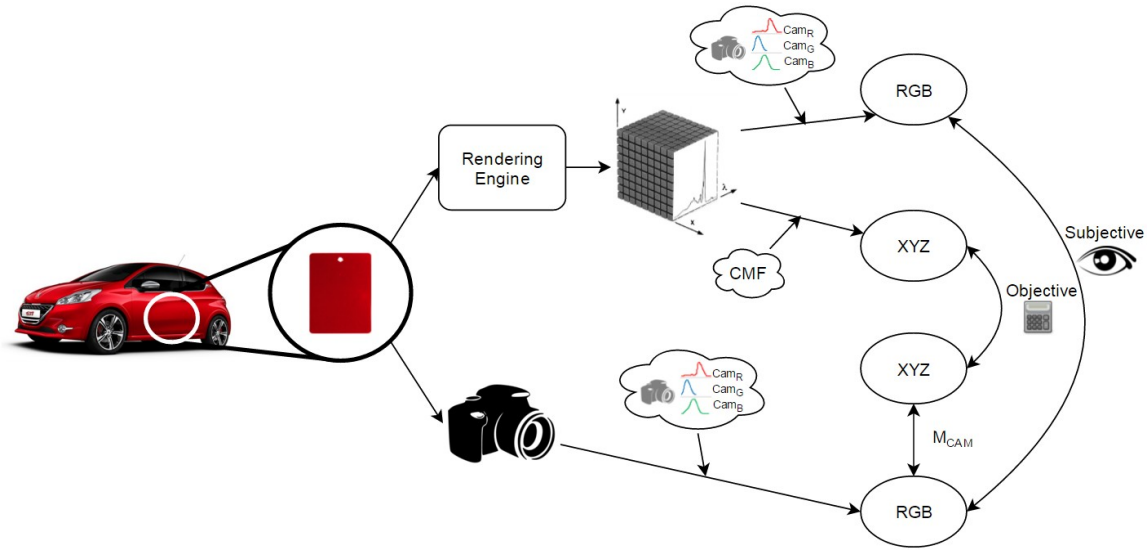



Figure 1.6: Visuo-perceptual validation methodology comparing photographs with simulations.

Chapter 5 presents a methodology to validate objectively the images produced by a rendering engine against photographs of a real scene (Figure 1.6). We describe the process to ensure identical conditions in the virtual and real scenes, which results in two types of images, photographs and simulations, sufficiently similar to permit a meaningful objective comparison between them. We will show how, with the proposed methodology, we are capable of simulating a scene with a very high level of perceptual colorimetric and radiometric realism. This is a fundamental step towards the validation of any physico-realistic rendering engine, because it ensures the reliability of the comparisons, and paves the way for the integration of the observations from the psychophysics experiments into the rendering model.

We will finish this PhD thesis manuscript with a discussion of the research work performed during this thesis, and by suggesting possible future lines of work in the context of this research.

Chapter 2

# Human visual perception of materials

## Contents


**2.1 Transforming electromagnetic waves into a visual response** . . . . . **14**
2.1.1 Electromagnetic radiation and visible light . . . . . 14
2.1.2 Mechanisms of the human eye for sensing and processing light . . . . . 15
2.1.2.1 General physiology of the human eye . . . . . 15
2.1.2.2 The human photoreceptors . . . . . 17
2.1.2.3 Human sensitivity to light . . . . . 19
2.1.2.4 Traditional approaches to color vision . . . . . 19
Trichromatic theory . . . . . 19
Opponent-colors theory . . . . . 21
Stage theory . . . . . 21
**2.2 Mathematical representation of color stimuli** . . . . . **22**
**2.3 Visual adaptation to the environment** . . . . . **25**
2.3.1 Luminance adaptation . . . . . 25
2.3.2 Chromatic adaptation . . . . . 25
2.3.3 Modeling visual adaptation mathematically . . . . . 27
**2.4 Measuring human visual acuity** . . . . . **27**
2.4.1 Spatial frequency and image resolution . . . . . 27
2.4.2 Contrast sensitivity . . . . . 29
2.4.3 Acuity of color vision . . . . . 29
**2.5 Binocularity and depth perception in humans** . . . . . **30**
**2.6 Vision as a temporal combination of partial views of a scene** . . . . . **31**
**2.7 Conclusion** . . . . . **31**


The computer simulations that we aim at producing are meant to be viewed by humans, who interact with the application and, ultimately, will be judging the aspect of the images. Therefore we must know how the human visual system (HVS) works when viewing these images, and understand its characteristics and limitations, in order to understand the complex processes behind human visual perception.

Humans are capable of seeing because our visual system can translate the light present in a scene into visual stimuli that can be interpreted by our brain. When the light interacts with the objects in a scene, it is transformed in different ways, depending on their physical properties and chemical composition, resulting in what we commonly know as object colors (see Section 1.2.2). Therefore, we cannot say that objects have a certain color; colors are the psychophysical sensations produced in our visual system by the energy emitted by those objects in a specific environment. That energy hits our eyes, which are covered by multiple photoreceptors that are sensitive to some of this energy, known as light, producing a neuron signal that is sent through the visual system and on to the brain, where it is processed.

Colors are indeed intimately linked to light. As we have seen, they are the result of the interaction of the light with the objects in a scene, and ultimately with the visual system of the observer [21],

which means that colors cannot exist in the absence of light. We could then say that we do not see objects, but *connected sets*, grouped by color criteria, that form shapes.

Visual perception processes are not the result of the action of individual areas in the HVS, but rather, of a collaborative interaction between them. For this reason, perceptual processes are difficult to describe completely, although many key features can be modeled [22]. In this chapter we will describe the structure of the HVS and present some important notions that are required to understand the mechanisms behind human visual perception, how the human eye captures the energy in a scene and transmits it to our brain to be translated into a visual response.

## 2.1 Transforming electromagnetic waves into a visual response

### 2.1.1 Electromagnetic radiation and visible light

Objects are constantly emitting electromagnetic radiation. A specific portion of this radiation can be seen by the human eye, and it is therefore known as *visible light*, or just light[1]. The nature of light can be typically explained by two fundamental theories, according to which light can behave both as a wave and a particle. This is known as the wave-particle duality of light. These two theories are complementary in the sense that, while many observations can be explained by both models, some can only be explained by one or the other.

According to the wave theory — first proposed by Christian Huygens in 1678 — light waves are made up of two mutually coupled vector waves, the electric and magnetic fields (electromagnetic wave), as shown in Figure 2.1. However, this model is often simplified by describing light as a single scalar wave function (Figure 2.2). This model is typically used to explain phenomena such as polarization, diffraction, absorption or imaging theory.

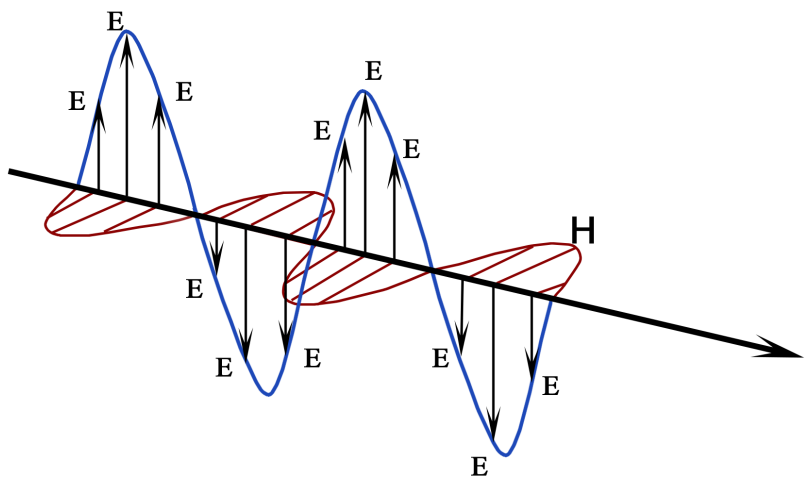



Figure 2.1: Electromagnetic waves are made up of electric (E) and magnetic (H) fields, perpendicular to the travel direction of the wave, and perpendicular to each other.

The particle theory (or quantum theory) — developed initially by Isaac Newton and re-emerged in the 20th century — treats light as a group of particles, known as photons, with energy and momentum, and without mass. This model is typically used to explain phenomena that cannot be explained with classical optics, such as light-matter interactions or lasers.

[1]Some authors use the term *light* more widely to refer to any electromagnetic radiation and the term *visible light* for the radiation that can be seen, but here we will use both as synonyms for *visible light*.

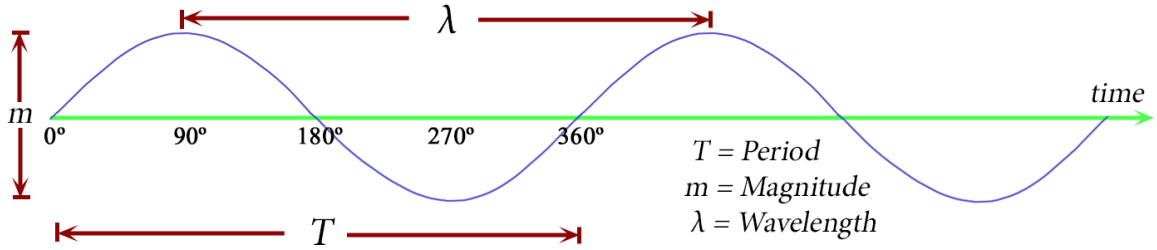



Figure 2.2: The electromagnetic wave model can be simplified in many cases as a scalar wave function.

From the point of view of the wave theory, the electromagnetic spectrum can be classified into different types of radiation attending to its wavelength. According to this classification, the term *visible light* refers to all the radiation with wavelengths roughly between 400 and 700 nanometers. This part of the electromagnetic spectrum is known as the *visible spectrum*, and corresponds to the wavelength interval that the photoreceptors in our eyes are sensitive to. The visible spectrum goes from violet light to red light, but often we also speak of ultraviolet light for the radiation with wavelengths immediately below the visible spectrum (100 nm - 400 nm), and infrared light for the radiation with wavelengths immediately above it (700 nm - 1 mm), even though we cannot see this radiation and therefore it should not be considered as light. Figure 2.3 summarizes this classification.

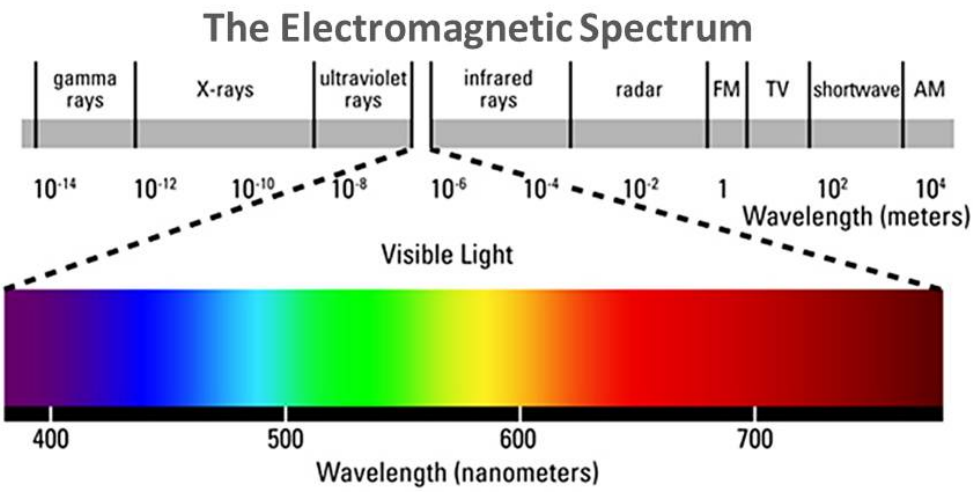



Figure 2.3: The electromagnetic spectrum. The visible light is that with wavelengths roughly between 400 and 700 nanometers. Ultraviolet and infrared light is that with wavelengths immediately below and above the visible spectrum, respectively. Image obtained from *The Photovoltaic Lighting Group* [23].

### 2.1.2 Mechanisms of the human eye for sensing and processing light

#### 2.1.2.1 General physiology of the human eye

The processing of information in the HVS takes place in three main stages: the eyes, the visual pathway that connects them to the brain, and the brain cortex. The eyes are two pseudo spheres of about 24 mm in diameter, with their centers separated by a distance of about 6.3 cm in average

for an adult — although it varies with the person. The walls of the eyes are formed by three layers (Figure 2.4): schlera, choroid, and retina, respectively from the innermost to the outermost layer. Towards the anterior pole, the schlera transforms into the cornea, a transparent tissue that is curved outwards about 1 mm with respect to the ocular globe. [24, 25]

At the intersection with the cornea, the choroid continues into the ciliary muscle and the suspensory ligament which, in turn, holds the lens —a biconvex structure that focuses the light on the retina by changing its shape to adapt to different focal distances, a process known as accommodation. The iris and the ciliary muscle are responsible for controlling the pupil size and the amount of light that reaches the retina. The space between the cornea and the lens, divided by the iris into the anterior and posterior chambers, is filled with a saline solution called the aqueous humour which controls the intraocular pressure and the refractive index, amongst other important functions.

The retina, which is an extension of the optic nerve, contains a series of photoreceptor cells and different layers of visual neurons. The most sensitive area in the retina is the fovea and, more specifically, its central region —an area of about $400\mu m$, corresponding to a visual field of 1.3°— called the foveola, which contains the highest concentration of photoreceptors. The optic nerve, which contains no photoreceptors, connects the information gathered in the retina with further visual processing stages that take place in the brain.

The images of what we see are formed at the fovea. Therefore, if we consider the eye as an optical system, the visual axis —which passes by the fixation point at the center of the pupil and the fovea— is deviated about five degrees from the horizontal optic axis, as shown in Figure 2.4.

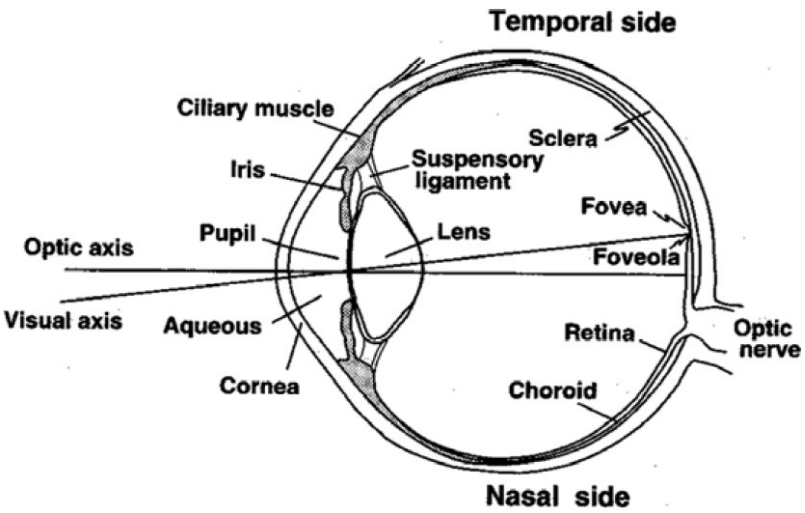



Figure 2.4: Top view of the right human eye [26].

The neural signals from the retina are transmitted from the eye into the visual pathway until they reach the visual cortex (Figure 2.5). The retinal neurons travel to the optic chiasm via the optic nerve, where the information perceived by the right and left-hand sides of the eyes is separated, transmitting the axons from right-field neurons to the left brain hemisphere, and the axons from left-field neurons to the right brain hemisphere. From there, the axons follow the optic tracts on to their respective hemisphere's LGN (lateral geniculate nucleus), where the information is processed and transformed. The new axons then follow the optic radiations to the primary visual cortex (the V1 area) and successive visual areas, where the information is finally processed and interpreted. [24, 26]

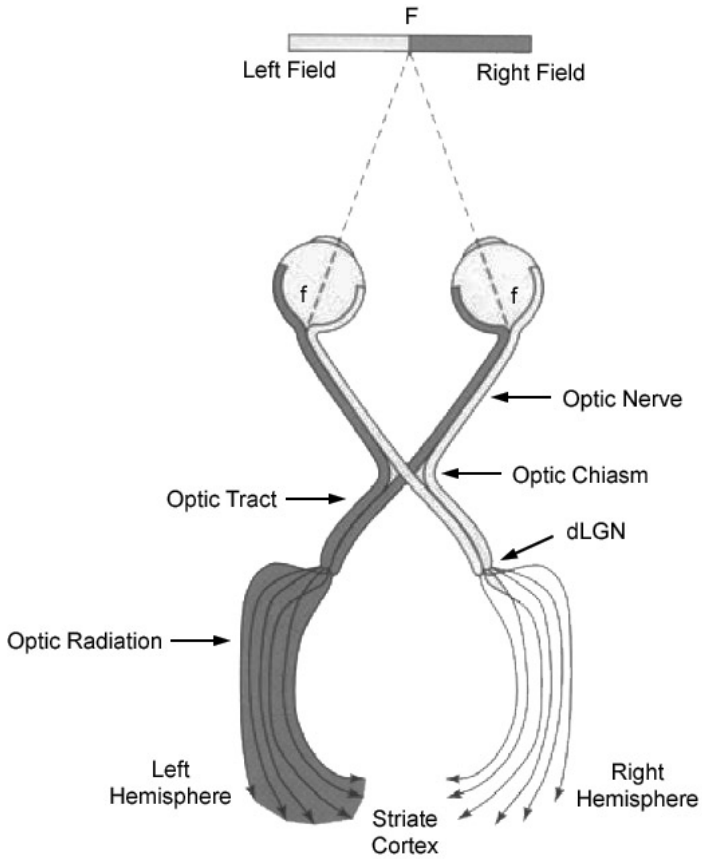



Figure 2.5: Diagram of the human optic nerves [26].

#### 2.1.2.2 The human photoreceptors

The surface of the retina is covered by different photosensitive and neuron cells of five types: photoreceptor, bipolar, ganglion, horizontal, and amacrine cells. All these cells are arranged in a double radial and transversal structure which maintains the spatial properties of the image while introducing lateral inhibition effects that increase contrast and sharpness in the visual response (Figure 2.6). [25]

As we can see in Figure 2.6, the photoreceptors are very elongated cells arranged perpendicular to the wall of the choroid which can discriminate light of different wavelengths. They are sensitive to the photons hitting the eye and can generate signals of varying amplitudes depending on the intensity of the light.

Bipolar cells can synapse with the photoreceptors —directly or through horizontal cells— to transmit signals to the ganglion cells —again, directly or through amacrine cells. Signals are sent using graded potentials via ON and OFF bipolar cells: in the presence of light, ON cells are depolarized (excited) and OFF cells are hyperpolarized (inhibited); inversely, in the dark, ON cells are inhibited and OFF cells are excited. [27]

Horizontal cells are responsible for the external synapses, transversal connections between the photoreceptors and bipolar cells. They allow the eye to adjust to changing light conditions and also introduce the lateral inhibition that participates in the center-surround inhibition in the ganglion cells.

Ganglion cells provide a mechanism for perceiving contrast and detecting edges. They have a receptive field connected with underlying levels in a way to cover isotropic areas of the visual

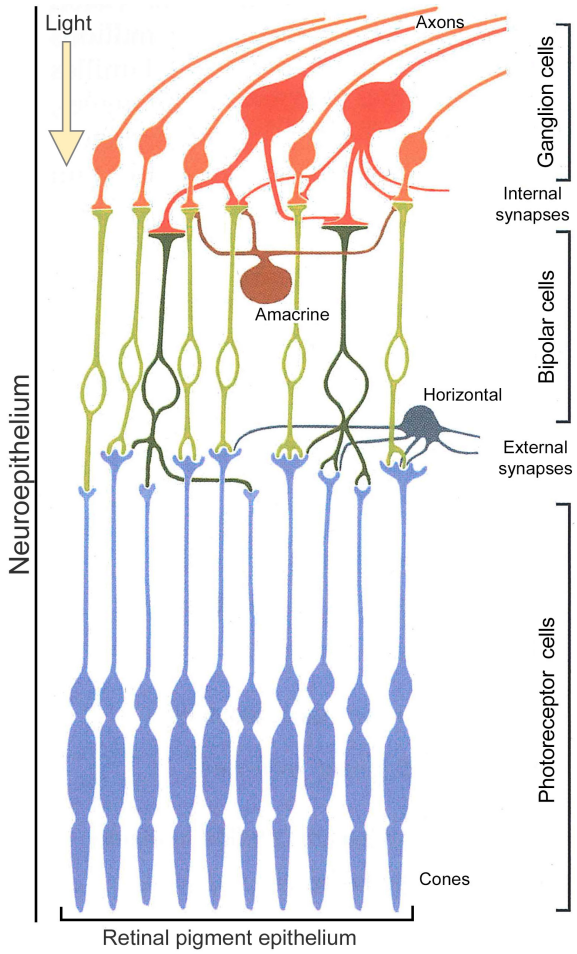



Figure 2.6: Diagram of the retinal cell structure. From bottom to top, cones are represented in blue, different types of bipolar cells in shades of green, and different types of ganglion cells in shades of red. At the very top, the axons from the ganglion cells continue towards the optic nerve. Diagram from Robert Sève's *Science de la couleur* [25].

field. The ganglion cell receives information about the contrast (or lack thereof) between a central area and a surrounding area that produce opposing responses. According to the response of their receptive field, ganglion cells can be of two types: "on-center" and "off-center". On-center cells are excited when the center of their receptive field is exposed to light, and inhibited when the surround area is exposed to light; off-center cells have the exact opposite reaction (Figure 2.7).

The amacrime cells create the internal synapses, transversal connections between the bipolar and ganglion cells. They work laterally, like bipolar cells, but are more specialized and supplement their action. Their function, although complex and not entirely known [25], is related to directional motion detection and modulation of light detection, amongst others.

We find two distinct types of photoreceptor cells in the retina: rods and cones. Rods are the most sensitive photoreceptors, only active at medium light levels —mesopic vision, at luminance levels of 0.001 to 3 $cd/m^2$— and low light levels —scotopic vision, at luminance levels of $10^{-3.5}$ to $10^{-6}$ $cd/m^2$. Cones are less sensitive to light than rods but provide a faster response time to stimuli; they are responsible for the vision at high luminance levels —photopic vision, at luminance levels of 10 to $10^8$ $cd/m^2$— and color vision. Due to their low sensitivity to light, cones can be divided in three types depending on the light wavelengths that they are sensitive to. [25]

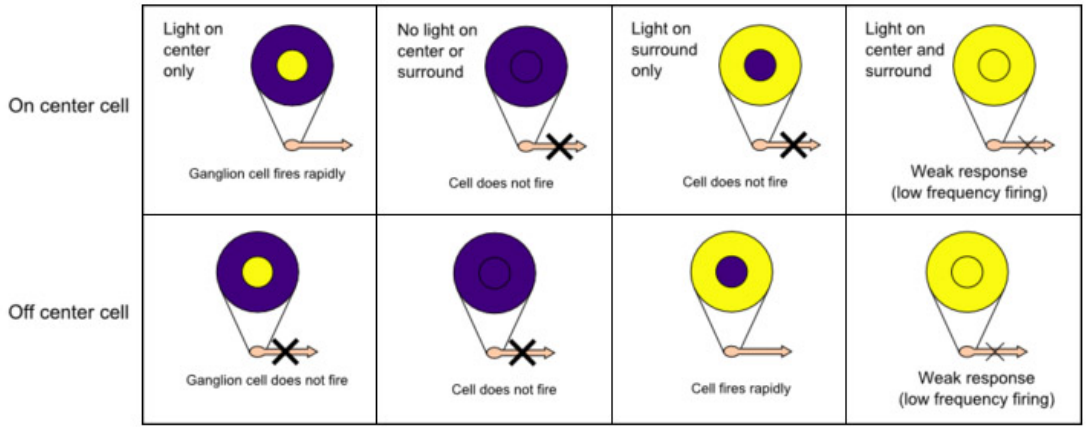



Figure 2.7: Representation of the On-center/Off-center ganglion cells mechanisms. Image taken from *Wikipedia* [28].

There are around 140 millions of rods and 5 to 7 millions of cones in the human retina. Cones are mainly located in and around the fovea, and highly concentrated in its central region, the foveola, where all three types are mixed (Figure 2.8). Furthermore, the central region of the foveola —an angular diameter of about 0.2°— only contains cones and no rods. In the 20° around this area, the number of cones reduces drastically, whereas the amount of rods increases rapidly to achieve its peak at about 15° of the visual axis, then reducing slowly until the peripheral area of the retina [25]. There is a small part —7.5° high and 5.5° wide— of the visual field, located roughly at 12°–15° horizontally towards the nasal retina and 1.5° below the visual axis, where nothing is visible. This area, which is known as the "blind spot", corresponds to the area where the optic nerve passes through the retina, and therefore there are no photoreceptors.

#### 2.1.2.3 Human sensitivity to light

We saw earlier that the photoreceptors in the HVS are classified according to the interval of the light wavelengths that they are sensitive to. Rods have their peak sensitivity at around 498nm; cones on the other hand are separated into three different categories: L cones, sensitive to long-wavelength light; M cones, sensitive to medium-wavelength light; and S cones, sensitive to short-wavelength light. The peak sensitivities of L, M, and S cones, are around 564nm, 534nm, and 420nm, respectively (Figure 2.9). Sometimes LMS cones are also referred to as RGB cones, hinting at the color of the light that they are sensitive to; however, as we can see in Figure 2.9, this denomination is inaccurate and LMS is preferred instead.

The spectral luminous efficiency functions shown in Figure 2.10 represent the sensitivity of the photoreceptors in scotopic and photopic vision. With scotopic vision only the rods are active, so the luminous efficiency function V'($\lambda$) corresponds exactly with that of the spectral sensitivity shown in Figure 2.9; the function V($\lambda$), however, is a combination of the three spectral sensitivity functions of L, M, and S cones, since all three are active under photopic conditions. Note that with scotopic vision we are more sensitive to shorter wavelengths which means, for example, that a red object that appears the same lightness as a blue one in daylight will appear much darker than the blue one under very low luminance levels —this is known as the Purkinje effect or Purkinje shift. [30]

#### 2.1.2.4 Traditional approaches to color vision

**Trichromatic theory**

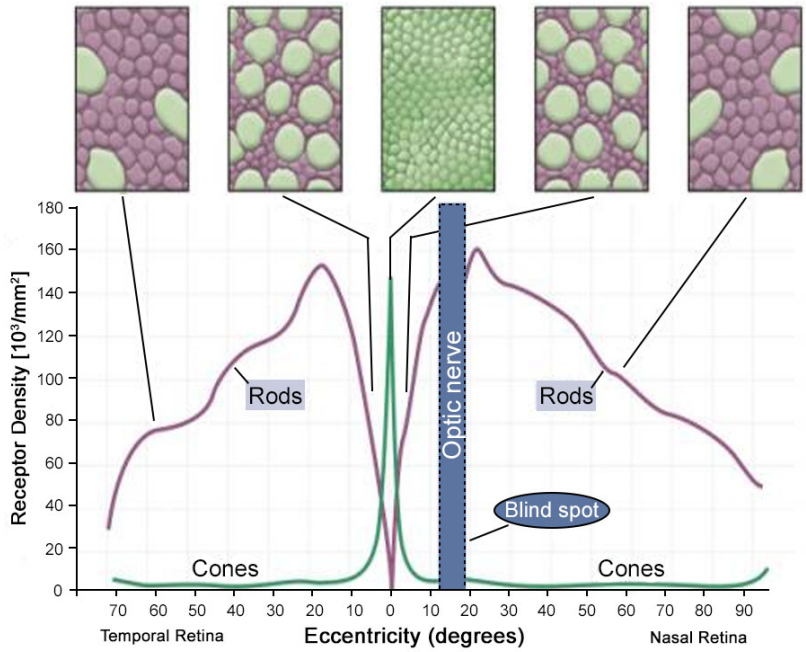



Figure 2.8: Density distribution of cones and rods in the retina. The area where the optic nerve passes through the retina is called the blind spot because it does not have any photoreceptors, which means that images cannot be formed in that area. We can see that there is a much higher concentration of cones towards the central region (the fovea), whereas rods are located in the surrounding areas of the retina. [26]

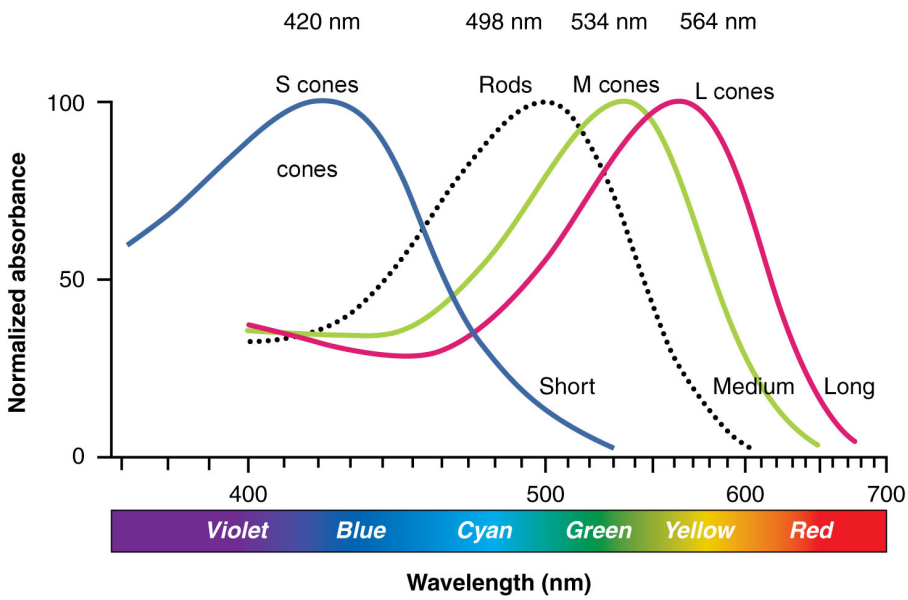



Figure 2.9: Normalized human photoreceptor absorbances for different wavelengths of light. Cones sensitive to short, medium, and long wavelengths are typically referred to as Blue, Green, and Red cones, respectively. Image taken from *Wikipedia* [29].

Traditionally, each type of cone has been associated with the perception of one color, which is why L, M and S cones are also known as Red , Green, and Blue cones, respectively. This is known

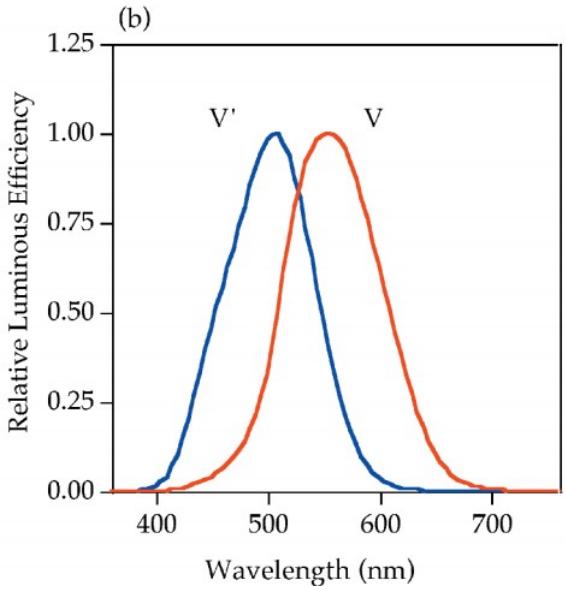



Figure 2.10: CIE spectral luminous efficiency functions for scotopic, V'($\lambda$), and photopic, V($\lambda$), vision. [30]

as the trichromatic color vision theory.

Proposed originally by Thomas Young (1773-1829) in 1802, it was first considered as a hypothesis, and then as a theory when the experiments of James Clerk Maxwell (1831-1879) proved its validity by means of color light matching experiments. This theory was extended later on by Hermann von Helmholtz (1821-1894) with results from quantitative experiments. He proposed three spectral sensitivity curves for the photoreceptors, and stated that the perceived color was in fact a combination of the relative responses from each type of photoreceptor. This theory explained the additive properties of color mixtures, and several other aspects of color generation and reproduction; however, it failed to explain other phenomena of color perception. [26]

### Opponent-colors theory

An alternative theory, was proposed in 1878 by Ewald Hering (1834-1948), who observed that there never seemed to be color combinations such as reddish-green or yellowish-blue. According to this theory, called the opponent theory of color vision, there would be three types of photoreceptors with an opponent response to red and green, blue and yellow, and black and white. It was then based on six basic colors: red, green, blue, yellow, black and white, and stated that color perception would be the result of the combined responses of these three types of color-pair receptors, which cannot produce signals for both colors in the pair at the same time. This theory explains phenomena such as negative afterimages —where a color appears when a presented stimulus of its opponent color is removed— or color blindness —caused by the lack of a particular chemical in the eye.

However, although both the trichromatic and opponent color vision theories do explain many color perception phenomena, none of them are capable of explaining all the aspects of color vision, so we cannot say that either one is right or wrong. In fact, most recent color vision models assume a combination of both. [26]

### Stage theory

The opponent colors theory regained popularity in the mid-20th century after a series of independent studies such as those by Jameson and Hurvich (1955), Svaetichin (1956), DeValois et al.

(1958), as well as many other researchers, provided data that reinforced Hering's theory [30]. This resulted in a modern version of the opponent-colors vision theory, also called the *stage theory*. This theory shows that color perception is indeed the result of complex processes performed in the brain cortex using opponent signals received from the retina.

According to this theory, color processing would take place at least in two stages: the first stage, corresponding to the photoreceptors, is trichromatic as described by Young and Helmholtz; the second stage follows the opponent theory, and takes place partly in the retina and partly in the visual pathway. Given the complexity of the processes that occur in the second stage, they are usually simplified using color vision models which vary from one author to another, such as those proposed by Boynton (1986) or DeValois (1993).

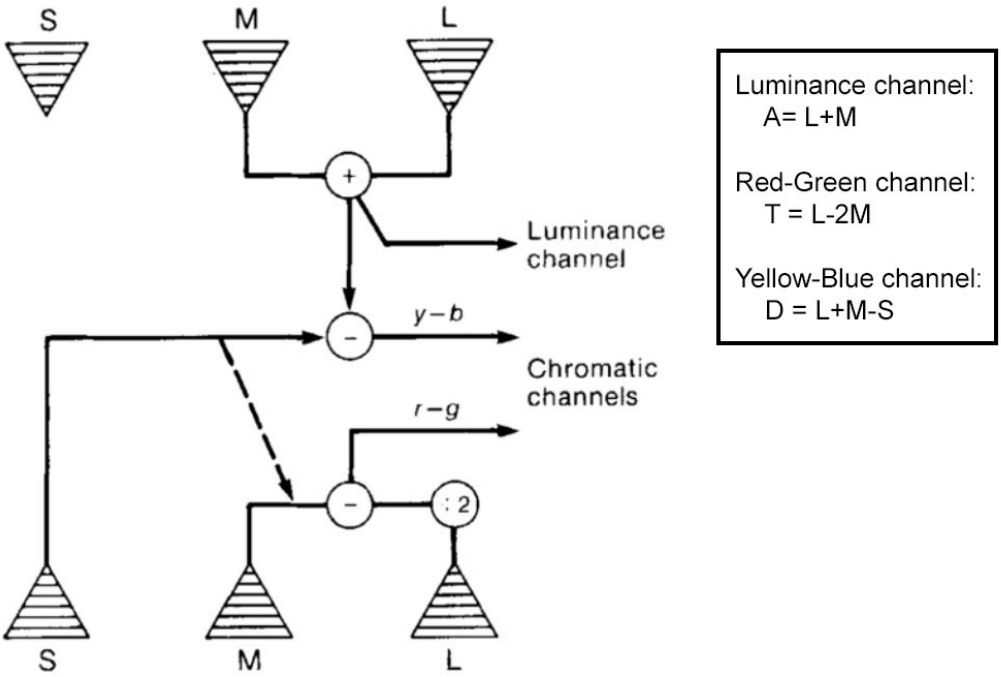



Figure 2.11: Left: Boynton's color vision model diagram. [31]; Right: Input combinations that produce each output. [26]

Boynton's is one of the simplest models. It represents the output of the processed opponent signals as three channels A, T, and D, which are a combination of the input signals from the photoreceptors (Figure 2.11). Channel A is the luminance channel, and corresponds to the addition of signals L and M; channel T represents the red-green opponent signal and is obtained as the result of subtracting L minus 2M; finally, channel D represents the blue-yellow opponent signal, which is obtained as L+M-S. This is illustrated in matrix form by Expression 2.1.

$$\begin{pmatrix} A \\ T \\ D \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -2 & 0 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix} \tag{2.1}$$

## 2.2 Mathematical representation of color stimuli

Color perception is the physiological sensation generated by our visual system to a specific interaction between light and matter. Indeed, all there is in reality is just light and surfaces that interact

with it. This light is measured in terms of its spectral power distribution (SPD)[2], however most of this information is redundant because the HVS samples color signals according to the sensitivity of the photoreceptors in the retina [32]. Therefore, when we work with color stimuli we must take into account this compression of information, and modify the perceived signal's SPD accordingly.

Most color vision approaches assume the existence of an initial trichromatic retinal stage corresponding to the three types of photoreceptors, L, M, and S. Regardless of whether or not, depending on the theory, further processing may occur in deeper areas of the visual pathway, the processing of the original signals received by the photoreceptors in the HVS is always assumed to be trichromatic. As shown by Maxwell, any given color light can be perceptually matched by mixing different amounts of a set of three known color lights, known as primaries [25]. The amount of each primary light required to match a stimulus are known as tristimulus values (see Appendix A).

The results of the color matching experiments proving the principles of the trichromatic theory were summarized by Hermann Grassmann (1809-1877) into the two following laws, known as Grassmann Laws [33]:

1. *To specify a color, three elements are necessary and sufficient: the hue, the luminance, and the luminance of the intermixed white, which defines the saturation.*
2. *For every color, there is a complementary color, which, when mixed, becomes a colorless gray.*
3. *Two lights of different color, with the same hue and saturation, when mixed, produce another color with identical hue and saturation independently of their power spectra.*
4. *The total luminance of any mixture of light is the sum of each light's luminance.*

All mathematical operations in colorimetry are based on these laws. If $[C_1]$, $[C_2]$, $[C_3]$, and $[C_4]$ are two pairs of matching color stimuli, we can then write that $[C_1] = [C_2]$ and $[C_3] = [C_4]$, where $=$ represents a perceptual match. Using this notation, Grassmann Laws can be summarized in mathematical terms as follows [34]:

1. Proportionality: Color matching holds when the intensity of all the lights are multiplied by a constant.
$$\alpha[C_1] = \alpha[C_2] \quad and \quad \alpha[C_3] = \alpha[C_4] \tag{2.2}$$
2. Additivity: Color matching holds for new stimuli obtained by adding color-matched stimuli.
$$[C_1] + [C_3] = [C_2] + [C_4] \quad and \quad [C_1] + [C_4] = [C_2] + [C_3] \tag{2.3}$$

The trichromatic theory then states than any color stimulus [F] can be matched mixing different amounts of reference stimuli [R], [G], and [B]; we can then write the color matching equation as:

$$[F] = R[R] + G[G] + B[B] \tag{2.4}$$

where R, G, B are the amounts of stimuli [R], [G], and [B], respectively, required to match stimulus [F]. Stimuli [R], [G], and [B], are the primary stimuli and R, G, and B, are called the tristimulus values. We can generalize for the spectral case as shown in Expression 2.5, in which case the tristimulus values become the color matching functions (CMF)[3].

$$[F_\lambda] = \bar{r}_\lambda[R] + \bar{g}_\lambda[G] + \bar{b}_\lambda[B] \tag{2.5}$$

[2]Radiant flux ($W/sr/m^2$) or luminous flux ($lm/m^2$) per unit wavelength (See a more complete definition in Section A.12 of the glossary)

[3]See a more complete definition of color matching functions further in this chapter and in Section A.33

The CMF indicate the tristimulus values of the spectral primaries for each wavelength. In general, for any given set of primaries, we can write the color equation as:

$$[F] = T1[C_1] + T2[C_2] + ... + Tn[C_n] \tag{2.6}$$

where $n$ is the number of primaries used to match the color stimulus, $Tk$ are the tristimulus values, and $[C_k]$ are the primary colors. A light with spectral power distribution (SPD)(see Appendix A) $P(\lambda)$ can be seen as made of a group of lights with a narrow band of wavelengths $\Delta\lambda$ [34], so using Expression 2.6 we can obtain any tristimulus value T from the SPD of the light, $P(\lambda)$, and its CMF, f, as shown in Expression 2.7. Ohta and Robertson [34] provide a detailed description of the steps to obtain this expression.

$$T = \sum_{i=1}^{n} P(\lambda) f_{\lambda i} \Delta\lambda \tag{2.7}$$

In order to represent color information more simply, omitting spectral information redundant to our visual system, color stimuli can be represented using tuples of numbers (the tristimulus values) and a set of primaries. Choosing a specific tuple representation and a set of primaries results in a color model; defining the conditions to interpret the model's components —such as the primaries' chromaticity or the set of CMF— results in a color space, that is, the space of all the possible colors that can be produced using that model under those conditions. Color stimuli are then specified as coordinates within a reference system; the specific color corresponding to those coordinates depends on the color space used.

Color spaces are always defined for an observer. Given the physiological differences between observers in terms of cone distribution, the tristimulus values may vary from one to another depending on their visual field. In order to produce standard color conversion functions, the CIE (see Appendix A) defined an average chromatic response function within a 2° arc inside the fovea —the area with the highest concentration of cones— known as the *CIE 1931 Standard Observer* or, simply, the standard observer. Alternatively, the *CIE 1964* 10° *standard observer* is also provided for observations covering a field larger than 4°.

There are any different color spaces, each one appropriate for different applications. The first attempt to connect spectral color signals with physiological responses in humans was the CIE 1931 RGB color space (or CIE RGB) based on three monochromatic additive color primaries: Red, Green, and Blue. It was created based on the trichromatic color theory and the results from color matching experiments, and it is mainly used for color representation in electronic systems such as computers and displays. The results from the observation experiments showed that some colors perceived by humans could not be matched using the three primaries, because they were too saturated. Matching could only be achieved by subtracting some amount of one of the primaries from the reference color, which can be observed in the negative values present in the CMF shown in Figure 2.12(a).

In order to correct the existence of these negative values, the CIE derived a new space from the RGB color space based on experimental results from Wright and Guild [36, 37], named the CIE 1931 XYZ (CIE XYZ) color space. This space is a linear transformation of the CIE RGB color space, as shown in Expression 2.8. In fact, the spectral chromatic response of the CIE 1931 Standard Observer was defined from the CMF of the CIE XYZ color space (Figure 2.12(b)).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.8}$$

The spectral sensitivities of the three types of cones in the HVS (Figure 2.9) originate another important color space, the LMS color space. In this space, which can be approximated to CIE XYZ

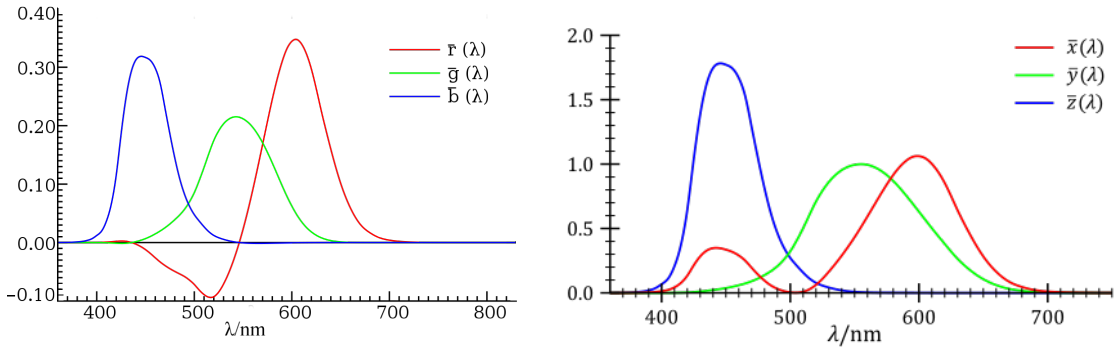



Figure 2.12: Left (a): CIE 1931 RGB color matching functions; Right (b): CIE 1931 XYZ and CIE 2° standard observer color matching functions. [35]

by a linear transformation, each tristimulus value represents the response of one type of cone, so it is useful when we need to model human visual perception, because it is easier to simulate the mechanisms performed by the human eye.

## 2.3 Visual adaptation to the environment

Human perception works in a way such that the aspect of objects remain relatively constant even when the observation conditions change. Two important adaptation phenomena take place without us realizing it when the luminance levels or the spectral distribution of the illuminant change, in order to adapt to the new conditions while preserving our perception of the environment. These two phenomena are, respectively, luminance adaptation and chromatic adaptation, or color constancy.

### 2.3.1 Luminance adaptation

The HVS can work at varying levels of illumination, ranging from about 100000 lx to about 0.0003 lx. To adapt to such different levels, the human eye has two mechanisms: on the one hand, the pupil can adjust its diameter, similar to the diaphragm of a camera, to control the amount of light reaching the retina; on the other hand, rods and cones alternate their function —rods in dark (scotopic) conditions, cones in bright (photopic) conditions, and both at the same time in intermediate brightness (mesopic) conditions— to change the responsivity of the retina. The adaptation from scotopic to photopic conditions is much quicker (about 1 min) than the adaptation from photopic to scotopic (around 30 min), as illustrated by the experimental dark adaptation curve in Figure 2.13.

### 2.3.2 Chromatic adaptation

As explained by Ohta and Robertson [34], chromatic adaptation occurs in two steps: first, an illuminant (or colorimetric) shift takes place, during which our perception of the colors will change according to the spectral distribution of the new illuminant; then, an adaptive color shift takes place, where our eyes adapt to the new illuminant and the color appearance changes to a corresponding color that matches that under the old illuminant. In theory these two steps should perform a color shift of similar length in opposite directions of the color space, resulting in perfect color constancy; however, in practice, color constancy is rarely perfect, and there is always a resultant color shift that is typically much smaller than either one of the two shifts separately. This process is illustrated in Figure 2.14, where illuminants D and A indicate daylight and an incandescence lamp, respectively;

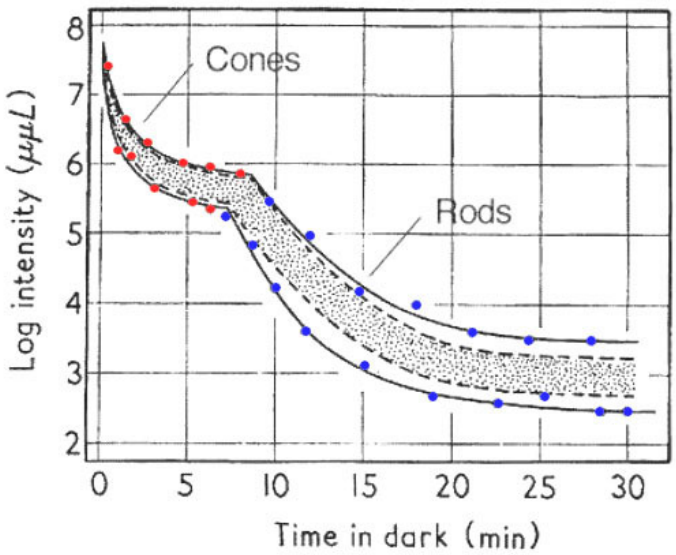



Figure 2.13: Dark adaptation curve. The shaded area represents 80% of the group of subjects. Hecht and Mandelbaum's data from From Pirenne M. H., Dark Adaptation and Night Vision. Chapter 5. In: Davson, H. (ed), The Eye, vol 2. London, Academic Press, 1962. [38]

$S_D$ and $S_A$ are the chromaticity points of an object under illuminant D and A, respectively; and $S_D{}'$ is the corresponding color for $S_D$ under illuminant A. Colorimetric changes in the illuminant become a problem in cases such as when we look at photographs or displayed images; in these cases, the radiometric information reaching our eyes is not that of the scene depicted in the images, but that of the environment under which we observe such images, which means that color constancy cannot take place. For this reason, we must introduce in the images a color correction to the illuminant color shift between the acquisition and the observation environments, known as white balance.

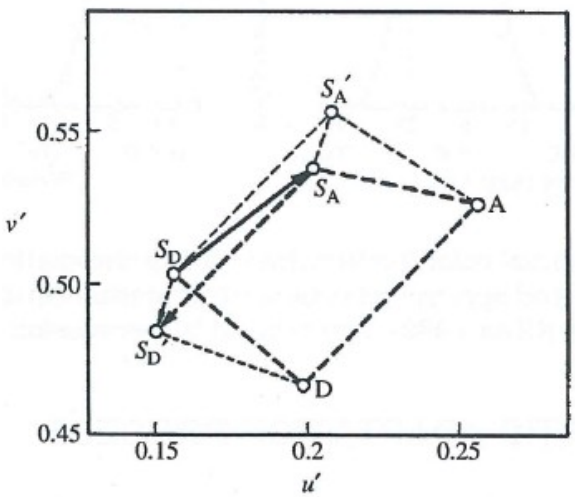



Figure 2.14: Illuminant color shift (from $S_D$ to $S_A$), adaptive color shift (from $S_A$ to $S_D{}'$), resultant color shift (from $S_D$ to $S_D{}'$), and corresponding color $S_D{}'$ for $S_D$, when illuminant changes from D to A. [34]

### 2.3.3 Modeling visual adaptation mathematically

The complex nature of human perception requires that new elements be introduced to fully describe a color, not only from a mathematical point of view, but also in terms of their perceptual aspect. Color appearance models (CAM) attempt to quantify the human sensory response to color stimuli, describing relationships between changes in viewing conditions and changes in appearance, by introducing appearance parameters such as hue, lightness, or chroma, which represent a measure of subjective sensations. Since the LMS color space is related to human visual perception, chromatic adaptation operations are typically performed in that space, so a transformation between the CIE XYZ and the LMS space must be defined. However, given the diffuse boundaries between the sensitivity bandwidths of the three types of cones, no objective transformation exists between these two spaces, and instead several authors provide various chromatic adaptation transforms (CAT) as part of different CAMs.

One example of a CAT is given by the von Kries transformation matrix. This matrix is based on the von Kries model, which proposes an independent color adaptation between two viewing conditions for each type of cone, by means of a linear transformation, as described in Expression 2.9, where $L$, $M$, and $S$, are the LMS tristimulus in the initial viewing conditions; $L_a$, $M_a$, and $S_a$, are the LMS tristimulus in the new viewing conditions after the adaptation; $L_{max}$, $M_{max}$, and $S_{max}$, are the maximum value for L, M, and S, respectively; and $L_{white}$, $M_{white}$, and $S_{white}$, are the LMS tristimulus of a white object.

$$\begin{array}{lll} L_a = k_L L & (k_L = 1/L_{max} \quad or & k_L = 1/L_{white}) \\ M_a = k_M M & (k_M = 1/M_{max} \quad or & k_M = 1/M_{white}) \\ S_a = k_S S & (k_S = 1/S_{max} \quad or & k_S = 1/S_{white}) \end{array} \tag{2.9}$$

For convenience, the same transforms are presented in matrix form in Expression 2.10, as a transformation between spaces under different illuminants.

$$\begin{vmatrix} L_a \\ M_a \\ S_a \end{vmatrix} = \begin{vmatrix} 1/L_{max} & 0.0 & 0.0 \\ 0.0 & 1/M_{max} & 0.0 \\ 0.0 & 0.0 & 1/S_{max} \end{vmatrix} \begin{vmatrix} L \\ M \\ S \end{vmatrix} \tag{2.10}$$

## 2.4 Measuring human visual acuity

### 2.4.1 Spatial frequency and image resolution

Visual acuity (or sharpness) is the ability of the HVS to resolve details in an image, and it is inversely proportional to the observation distance. The sensitivity to changes in brightness increases with the spatial frequency of the changes. The acuity is related to the type of photoreceptors being used: under scotopic vision rods are active, so our eyes are more sensitive to small changes in luminance, but acuity for detail and color is poor; conversely, under photopic vision, cones are active, so color vision and acuity are sharp, but sensitivity to changes in luminance is low. [22]

Visual acuity is typically measured using the Snellen eye chart (Figure 2.15) or any of its more modern variants such as the LogMAR chart. According to these charts, a normal person has a visual acuity of 20/20 (6/6 in the metric system) which means that this person is capable of seeing clearly at 20 feet (respectively, 6 meters) an object that should be seen clearly at that distance. In general, the numerator of this fraction, known as the Snellen fraction, indicates the distance at which a person can see clearly an object that should normally be seen clearly at the distance indicated in the denominator – e.g. 20/10 (6/3), 20/25 (6/7.5), or 20/100 (6/30) [39].

Each letter on the 8th line (20/20 or 6/6) of the Snellen chart should subtend exactly 5 minutes of arc (and the width of each black line should subtend 1 minute of arc) from a distance of 6 meters.

Therefore, the chart must be sized so that the letters on the 8th line are 8.86 mm tall. This means that from a 6-meter distance, a person with normal visual acuity can resolve clearly any visual element above 8.86 mm in size —which corresponds to a visual resolution of 1.48 mm at a 1 meter distance.

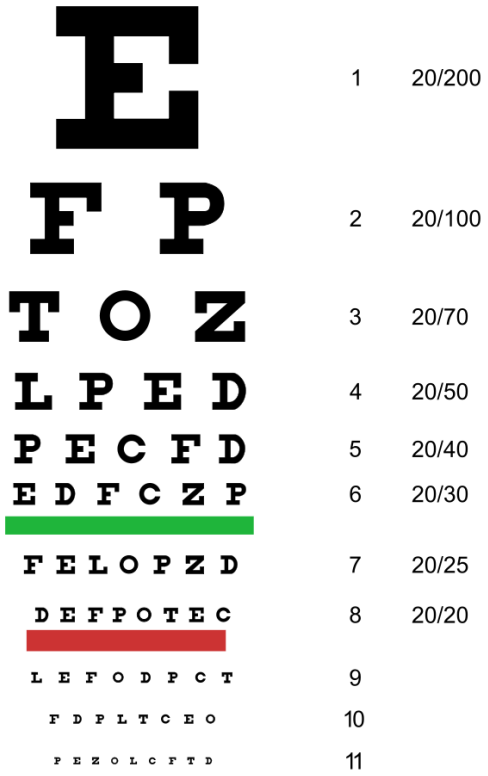



Figure 2.15: The Snellen chart is used for measuring visual acuity by showing it to the subject at a distance of 20 feet (or 6 meters). It was developed in 1862 by Dutch ophthalmologist Herman Snellen.

Another way to measure visual acuity is through the use of sinusoidal gratings, which measure acuity in terms of spatial frequency. Snellen's fractions can be converted to spatial frequency (Figure 2.16), which is more convenient when working with displays given the grating-like arrangement of pixels in an image. In a grating, a cycle is the distance between two consecutive black lines (i.e. one black line followed by a white line). A person with normal vision is known to have a maximum resolution of 60 cycles per degree of vision, which means that an image must contain at least 60 pixels per degree of vision to avoid its pixels from being seen.[40]

The viewing angle or field of view (FOV) of an observer located at a given distance can be calculated using trigonometry, as shown in Expression 2.11.

$$FOV = 2 \times \arctan\left(\frac{Sample\ size}{2 \times Observation\ distance}\right)_{deg} \ (degrees) \tag{2.11}$$

Since we know that a person with normal vision has a maximum resolution of 60 cycles per degree of vision, we can derive from Expression 2.11 the minimum resolution (MR) of a displayed sample photograph to avoid the observer from seeing the pixel grid of the image (pixelation) using Expression 2.12

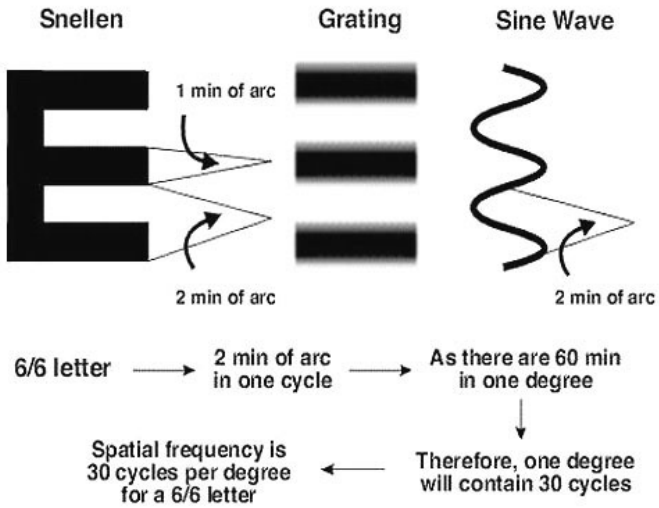



Figure 2.16: Conversion between Snellen's notation and spatial frequency.[40]

$$MR = FOV \times 60 \ (cycles) \tag{2.12}$$

### 2.4.2 Contrast sensitivity

We cannot see patterns with frequencies over the maximal resolution of the eye. The sensitivity of the HVS to the spatial frequency of visual stimuli can be measured using the Contrast Sensitivity Function (CSF), which represents the contrast sensitivity of a person as a function of the spatial frequency of a stimulus. [22]

Contrast perception follows the Weber–Fechner law, according to which the human perception of a given stimulus follows a logarithmic function, so that the just-noticeable difference (JND) between two adjacent stimuli is proportional to their magnitudes. Depending on the situation, contrast can be defined in several ways, with slight differences, the most common being the ratio between the luminance difference between two adjacent stimuli and the average luminance. The most relevant in our context is the *Weber contrast*, defined as shown in Expression 2.13.

$$\frac{I_f - I_b}{Ib} \tag{2.13}$$

where $I_f$ is the luminance of an image feature, and $I_b$ is the luminance of the background. This formula is used in cases with small features on a large uniform background, that is, with small differences between the feature and background luminance [41]. This is illustrated by a phenomenon called simultaneous contrast, where a stimulus of constant luminance is perceived with a different contrast as the luminance of the background changes (Figure 2.17)

### 2.4.3 Acuity of color vision

The spatial acuity of the HVS varies for different parts of the light spectrum or locations in a color space. Some contributing factors are the chromatic aberration of the eye, the photoreceptors density on the retina, or the spatial acuity of opponent channels [42]. Due to the refraction index of the human eye, each wavelength arriving into the eye is focused to a different distance from the retina that is directly proportional to the wavelength of the light, which limits our ability

Figure 2.17: Simultaneous contrast: the internal squares all have the same luminance, but the changes in luminance in the surrounding areas change the perceived luminance of the internal squares.[22]

to resolve each wavelength. This is called axial chromatic aberration. Another limiting factor is the density of each type of photoreceptor on the retina: it is known that all three types of cones are not equally distributed over the retina, with approximate double amount of L cones than M cones, which samples the original spectral stimulus down to only three components; as shown by Meyer and Liu [42], there are approximately 62 L cones, 32 M, and 6 S per degree of vision, which determines the visual acuity for each wavelength. Furthermore, Meyer and Liu also cite results by Middleton [43] and Schade [44] as evidence that color spatial acuity is less than monochrome, supported by results from Mullen [45] showing a cutoff of 34 cycles/degree for the luminance channel, and 11 cycles/degree for the red/green and yellow/blue opponent channels (see Section 2.1.2.4), as illustrated by Figure 2.18.

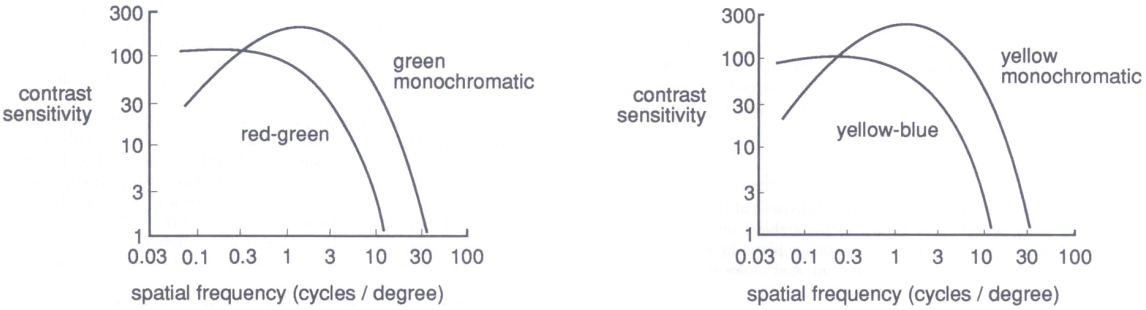



Figure 2.18: Spatial frequency response of opponent color channels. Left: red-green versus monochromatic green; Right: yellow-blue versus monochromatic blue.[42, 45]

## 2.5 Binocularity and depth perception in humans

A very important characteristic of the HVS is its binocularity, or the possession of two eyes. Indeed, having binocular vision confers the human being many advantages, the most remarkable being a wider field of view and a more precise depth perception. The monocular field of view of an average person is approximately 60° nasally, 60° superiorly, 70°–75° inferiorly, and 100°–110° temporally [46]; however, with binocular vision, the field of view increases up to approximately 200° horizontally and 135° vertically, with a common binocular overlap that is 120° wide [47]. The position of the eyes in humans —located on the front of the head and separated in average by an interpupillary distance of 64.7 mm in men and 62.3 mm in women [48] — offers a narrower visual field than other animals —some birds may extend up to 360°— in favor of better stereopsis. Stereopsis is the exploitation of the small disparity in the position of the objects between the images observed by both eyes —also known as binocular disparity or parallax— which allows for a very accurate depth perception and distance estimation by means of triangulation. Binocular vision helps separate regions of interest from background, providing figure-ground separation and

spatial localization; it also improves the assessment of surface curvature and the perception of surface materials using luster perception. Combining the information seen by both eyes, known as binocular summation, permits to average the visual input from both eyes, correcting for monocular inaccuracies, and improving absolute light detection at the perception threshold and the CSF. [49]

The HVS infers depth perception from information received by two types of depth cues: monocular and binocular. Some monocular cues are: perspective, texture gradients, occlusion, relative and known object sizes, haze, or defocus [50]. Binocular cues are provided by four main physiological mechanisms: binocular disparity, accommodation, eye convergence, and motion parallax. From these four mechanisms, we have already described binocular disparity and accommodation (see Section 2.1.2); eye convergence is the capability of the human eye to rotate inwards, increasing or reducing the amount of parallax; motion parallax refers to the additional parallax introduced by movement. Without these four physiological cues, it is not possible to perceive a true three-dimensional view [51].

## 2.6 Vision as a temporal combination of partial views of a scene

The human eye is constantly sampling information from images projected onto the retina. This information is integrated in time at several stages along the visual system, notably the retina and the visual cortex. Information is integrated at the retina to stabilize the position of objects and produce the impression of smooth movement. This temporal integration has a limited resolution given the finite amount of time required to process the information. Intermittent stimuli are perceived smoothly when presented to eye above a certain rate; below that rate, they are perceived as separate, producing the sensation called flicker [52, 53]. These mechanisms are exploited, for example, by television or cinema, to produce the desired effect from a succession of static images. A second integration level is located at the visual cortex, where the brain creates a representation of a visual scene from different spatial eye fixations. Processing times take place between different saccades, while the eye is static, so rapid saccades (microsaccades) do not provide visual information. On the other hand, smooth pursuit movements allow us to follow and fixate moving objects to stabilize gaze.

At an upper integration level, the environment can be reconstructed as a homogeneous whole where each element appears perfectly clear. However, in reality, this reconstruction is put together from a number of different elements, captured at successive times, which only contain some clear elements and many blurred ones. Such a complex puzzle is reassembled using immediate memory to provide it with continuity and coherence.

These cognitive mechanisms are used, for example, to provide color constancy under different illuminations. [53]

## 2.7 Conclusion

Human visual perception, as we mentioned concisely in Chapter 1, is produced as the combination of a series of bottom-up and top-down processes. In the bottom-up process, the sensory responses produced by the light that arrives in our eyes constitutes a stimulus representation which is accepted by the observer passively. This representation includes low-level information such as color discrimination and light information [10]. Human perception continues as the information is transmitted via different channels to deeper layers, where top-down influences enrich our perception of the environment. The signals are processed at several layers of the LGN to provide information about depth, or fine details about brightness and color contrast. The information from the LGN is then

sent to the visual cortex. In early cortex stages, spatial information is provided, such as orientation and contours, being responsible for the occurrence of some basic visual illusions related to figure contour and occlusion [54]. Deeper areas of the visual cortex provide more complex information, influenced by factors such as the environment, the perception context, the perceiver's cognitive and psychological state, or the motivation. At this point, perception processing specializes into two different paths: the "where" (or "how") stream, involved in spatial attention, affecting eye and hand movements; and the "what" stream, involved in recognition, identification, and categorization of visual stimuli.

When we display a simulated virtual stimulus to a human observer, the information corresponding to top-down influences is already provided by the observers themselves. Indeed, when observers look at a displayed simulation, they already contribute with high-level information processing beyond what is provided by the simulation. For this reason, we only offer here an introduction to the mechanisms behind passive bottom-up processing. To maximize perceptual realism in the simulations, we must minimize the amount of perceptual information from bottom-up influences that gets lost in the simulation process, that is, chromaticity (color) and luminance information.

The visual studies that we perform in this thesis research also need the consideration of the higher-level mechanisms that participate in the visual perception, in order to understand what the observers see when they look at the materials. As we will see in the following sections, color and luminance contrast, in turn determined by how the color and luminance information in the scene in represented in the images, play an important role in the perception of material aspect in images. Since they affect the spatial relationships within sub elements in a texture, this also affects the ability of an observer to identify and recognize visual image attributes such as texture density, or randomness.

Chapter 3

# Ensuring color signal consistency in the visualization chain

## Contents


**3.1 Choosing the right display device** . . . . . **34**
3.1.1 Display Technologies . . . . . 34
3.1.1.1 Cathodic Raytubes (CRT) vs Liquid-crystal displays (LCD) . . . . 34
3.1.1.2 Common LCD display technology . . . . . 34
Twisted Nematic Liquid Crystal Display (TN-LCD) . . . . . 34
In Plane Switching Liquid Crystal Display (IPS-LCD) . . . . . 35
3.1.1.3 Color management in LCD displays . . . . . 35
3.1.2 Depth perception in displays . . . . . 35
3.1.2.1 Stereoscopic displays . . . . . 36
3.1.2.2 Holographic displays . . . . . 37
3.1.3 Our choice . . . . . 37
**3.2 Color signal transformations introduced by the display** . . . . . **37**
3.2.1 Previous considerations . . . . . 39
3.2.1.1 Pixel independence . . . . . 41
3.2.1.2 Warm-up . . . . . 42
3.2.1.3 Angular dependence . . . . . 42
3.2.1.4 Channel constancy . . . . . 45
3.2.1.5 Channel independence . . . . . 46
3.2.1.6 Gamma correction . . . . . 46
3.2.2 Modeling the response of the display . . . . . 48
**3.3 Effects of stereoscopic visualization on the color signal** . . . . . **55**
**3.4 Conclusion** . . . . . **56**


We have seen in previous chapters that visualizing images on a display implies the representation of a radiometric reality with an unlimited dynamic range, in a medium with a much more restricted dynamic range. Although the human eye is not capable of perceiving all the luminance variations present in a real scene, it can perceive a much wider range of luminance levels than what can be shown by a display.

The perceptual aspect of displayed images is directly related to three main factors: the display technology, the transfer function, or colorimetric response, of the visualization device (inherent to its electronic components) and the tone mapping operator used to display the image.

Section 3.1 starts by introducing liquid-crystal display (LCD) technologies —presently the most commonly used in the general consumer market— in comparison with their predecessor, cathodic ray-tube (CRT) displays. The most prevalent LCD technologies are presented, as well as the mechanisms that they use to produce color. Next, we describe the two most popular technologies for depth visualization, stereoscopy and holography, naming some advantages and disadvantages of

each one. We finish the section by explaining the display technology chosen for our tests, justifying this choice in terms of the suitability of the technology for the experimentations. Section 3.2 describes the most relevant characteristics of LCD displays that may affect the visualization under our specific test conditions, as well as a methodology to perform the characterization of the device. Each aspect of this process is illustrated with specific results obtained for our chosen display. Finally, Section 3.3 will describe some particularities of the characterization methodology required by the use of stereoscopic visualization.

## 3.1 Choosing the right display device

The technology used in the visualization device has a very strong influence on the quality of the displayed images. Some technologies may be inherently limited by the electronics, while others may skyrocket the price of the solution. Choosing the right display technology is therefore a very important step in obtaining a solution that provides good results within the requirements of the project.

LCD technology has replaced CRT displays as the main technology for desktop computer monitors. We present in this section an overview of the main technologies used in LCD displays —focusing on those most relevant to our test display— providing in each case some advantages and disadvantages relevant to psychophysical experimentation.

### 3.1.1 Display Technologies

#### 3.1.1.1 Cathodic Raytubes (CRT) vs Liquid-crystal displays (LCD)

CRT displays produce light by means of three primary phosphors —red, green, and blue— for each pixel. Each phosphor is excited independently by an electron beam of varying intensity to form a pixel. This results in a large gamut of colors and brightness levels [55]. Conversely, LCD displays are formed by an array of liquid crystal molecules which can change their polarization state from "on" to "off" by applying a varying electric voltage. In their off state the molecules block the light, whereas in their on state they let the light pass.

Although there are many advantages to LCDs over CRTs in terms of consumer applications, such as smaller space requirements or lower energy consumption, their use in vision science applications introduces several complications [56, 57, 58]. For example, Gaurav Sharma reported that while earlier LCD models already provided a larger color gamut than CRTs, changing the viewing angle resulted in colorimetric variations larger that what was estimated by the characterization model. Indeed, the large variations of luminance as a function of both the viewing angle and the distance from the surface's central position, is a well known disadvantage of LCD displays. Elze and Tanner [58] provide a very good analysis of the positive and negative aspects of the main technologies used in LCD displays in comparison with CRT displays.

#### 3.1.1.2 Common LCD display technology

Depending on the type of liquid crystal molecules, their arrangement, and their orientation, we can distinguish different types of LCD displays. Next we describe the most common technologies: Twisted Nematics LCD and In Plane Switching LCD.

**Twisted Nematic Liquid Crystal Display (TN-LCD)** TN-LCD is the most common technology in the desktop display market [59], used also for watches and calculators. It consists of a layer of nematic liquid crystals, between two layers of glass. As a voltage is applied to the nematic

crystals they stand up, changing their orientation and polarization, and blocking light with the selected polarization. Varying voltage can produce almost any gray level [60]. They present however certain drawbacks, most notably their narrow viewing angles (especially vertically) and low color depth [61, 59, 62]

**In Plane Switching Liquid Crystal Display (IPS-LCD)** This technology was designed to resolve the limitations of TN-LCDs such as brightness change an color shift depending on the viewing angle. In IPS-LCD, liquid crystal molecules never change their orientation with respect to the display panel, and instead they rotate in parallel with the panel. For this reason, the variations related to the viewing angle are much smaller than in TN-LCDs [63]. Besides preserving accurate color display from all viewing angles, they eliminate other issues of TN-LCDs like the appearance of tailing when touched [64], which makes them more convenient for touch-screen devices, such as smartphones and tablets.

#### 3.1.1.3 Color management in LCD displays

The human eye can discriminate around 10 million different color shades. Modern graphic cards are capable of displaying RGB images with up to 8 bits per color channel, or 24 bits in total, which can represent 16.7 million color shades.

Color in LCDs is created by a backlight —which is usually made of LEDs— that can be separated into different color components via optical elements, which are then let pass by the liquid crystal molecules depending on their state. These molecules can typically present only two states (on/off) which translates into a color depth of one bit per pixel. Some newer models use variations with molecules that can present a few different states, with up to 6 bits per bit, so they cannot produce as many colors as graphic cards. It is therefore necessary to use additional techniques to increase the precision of pixel colors on LCD displays.

The most common technique is frame rate control (FRC), which takes advantage of the color persistence properties of the human retina [65, 66] (see Section 2.3.1) to increase in the number of perceivable gray scale values. The number of color shades may therefore be increased by combining different colors in time at a fast rate (30 Hz), through the use of dithering algorithms —a pixel density variation method that introduces intentional noise to reduce quantization patterns. In theory some problems —such as transverse striping or perceived flickering [55]— may arise due to the use of FRC, specially in first-generation LCD models and the most inexpensive modern ones, but in practice they are hardly noticeable to the human eye [67].

### 3.1.2 Depth perception in displays

Recent years have seen the emergence of numerous visualization technologies that try to reproduce the perception of depth in the human visual system (HVS). Section 2.5 described how the humans can obtain depth information from monocular —perspective, texture gradients, occlusion, etc— and binocular cues —binocular disparity, accommodation, convergence, and motion parallax. These binocular cues can be reproduced by some display technologies to produce depth perception in the viewer. The most important of such technologies are stereoscopy and holography. Stereoscopic displays are developing at a very fast pace, and are already well established as a mass-market product; however, they have some limitations that make holographic displays seem like a more promising technology in the future.

#### 3.1.2.1 Stereoscopic displays

This is the most common type of displays that provide depth perception. They are often referred to as 3D displays, however this name is not correct because they do not provide all four binocular depth cues required to see true 3D images. Indeed, although the images shown by this type of displays are perceived by the observer as being at different depths —both with respect to each other and to the observer— the elements depicted are all at the same depth plane, namely, the image plane corresponding to the screen. Therefore, when looking at these images our eyes must accommodate to the screen's distance and not to the perceived distance, which provides conflicting accommodation cues. Moreover, since the displayed images have been taken from a fixed point of view, there is no true binocular disparity between the images seen by each eye.

Nonetheless, although this type of displays does not provide true three-dimensional images, they can mimic stereoscopic visualization and simulate the perception of depth by adding the missing binocular depth cues artificially. This can be done in several ways, as we will describe next.

1. **Stereo-Pair displays**

   Stereo-Pair stereoscopic displays create artificial parallax by distributing a separate image of the scene for each eye, so they normally require an additional viewing device —generally a pair of special glasses— in order to combine two separate images sent by the display into one single view. There are two types, distinguished by the type of viewing device used:

   - Active (shutter) stereoscopy:

     The image-eye synchronization —known as alternate-frame sequencing— electronics are contained within the glasses themselves; normally, two independent LCD filters (one per eye) switch on and off, alternatively for the right and left eye, in phase with the displaying device, at rates from 60 to 120 frames per second or above. This technology provides a larger color spectrum and higher image resolution, however, the filters in the glasses alter the original image color and reduce the luminance, which may also modify the contrast in LCD displays as a result of the reduction of the blacklight intensity. This technology has proven so far to be the best solution for 3D home entertainment such as video games and home cinema.

   - Passive stereoscopy:

     Images are sent to the viewing device in a special way so that the images for each eye can be separated on the other end. The most common solutions rely on using a different polarization (usually linear or circular) or complementary colors (known as anaglyph stereoscopy) for each eye's image. The glasses contain special filters to be able to only see the right image. This technology has the advantage that glasses are inexpensive, but the displays are much more expensive and, since the images have to share the screen, the resolution is reduced. This technology is more common in applications for large audiences such as cinema theaters.

2. **Autostereoscopic displays**

   Unlike stereo-pair displays, autostereoscopic displays —also known as glasses-free 3D or glasses-less 3D— do not require an additional external device in order to provide the perception of depth [68]. They have the ability to multiplex several views at a time, so a series of different perspective views can be delivered to the observers depending on their viewpoint, a process called *direction-multiplex* [69]. This type of stereoscopy reproduces two of the four eye mechanisms required to create a 3D image, namely, motion parallax and binocular dis-

parity. Common examples of autostereoscopic displays are parallax barrier and re-imaging displays [70].

#### 3.1.2.2 Holographic displays

Holographic displays work by encoding at a spatial light modulator (SLM) the light diffracted from a 3D object, which essentially corresponds to its Fourier transform. Color is typically produced using either monochrome (most often) or light-emitting diodes (LED) [71, 72]. Although some early working implementations exist [73, 74, 75], they are much less developed than stereoscopic displays; however, the limitations inherent to the latter, make holographic technologies a more promising bet for 3D displays. Unlike stereoscopy, holography does provide all four binocular depth cues necessary to reconstruct natural looking three-dimensional scenes with unlimited depth, however, it still has some issues such as insufficient display resolution, inadequate data volume, and large computational requirements [76].

### 3.1.3 Our choice

We have presented the most relevant types of display technology, as well as the main techniques to reproduce depth perception in displays. The nature of our work imposes certain requirements that reduce our number of choices in terms of display technology, so the advantages and disadvantages of each type had to be taken into account when choosing a display for our experiments. Indeed, initial material analysis performed by Couka et al. [77, 78] reveal that the effect particles contained in some of the paint coating materials used in our research are distributed at different depths and orientations. Consequently, as we will explain in later chapters, depth plays a very important role in the visual perception of the material aspect, and neglecting this information would reduce the perceptual realism of the displayed images.

Although true depth would be preferable, the current development state and low availability of holographic displays effectively restricts our choice of displays to stereoscopic displays. We tested several types of stereoscopic technology, such as a stereoscopic projector, an autostereoscopic display, and several types of active-viewing stereoscopic displays, and found that active-viewing stereoscopy gave the best results in terms of colorimetry, resolution and quality/price ratio. In order to avoid an elevated cost for the system, we decided to choose general-purpose technology, so the final choice was an *Asus VG248QE* stereoscopic display with a set of Nvidia3D vision shutter glasses as the viewing device. We purposely selected this display's specifications (see Table 3.1) to obtain the best results for our requirements, keeping in mind the numerous disadvantages of using an LCD display for visual experiments [58, 57]. This monitor uses TN-LCD technology but, in spite of the drawbacks mentioned earlier associated with this technology, this specific model ranks very well in terms of color reproduction in existing benchmarks [79]. Furthermore, the color depth is increased greatly through the use of FRC, which we could corroborate from the color gamuts resulting from the measured primaries in each of the tested monitors (Figure 3.1).

The only important problem left to workaround is the reduced viewing angles which, given the size of our samples, we can minimize by restricting the position of the displayed samples during the experiments to avoid using the outer edges of the display, and also limiting the observer's position to the center of the screen to minimize the angle with respect to the image position.

## 3.2 Color signal transformations introduced by the display

Traditionally, CRT displays have always been preferred over LCDs for visual experiments because their behavior has been widely studied. However, due to the advances in flat-screen display tech-

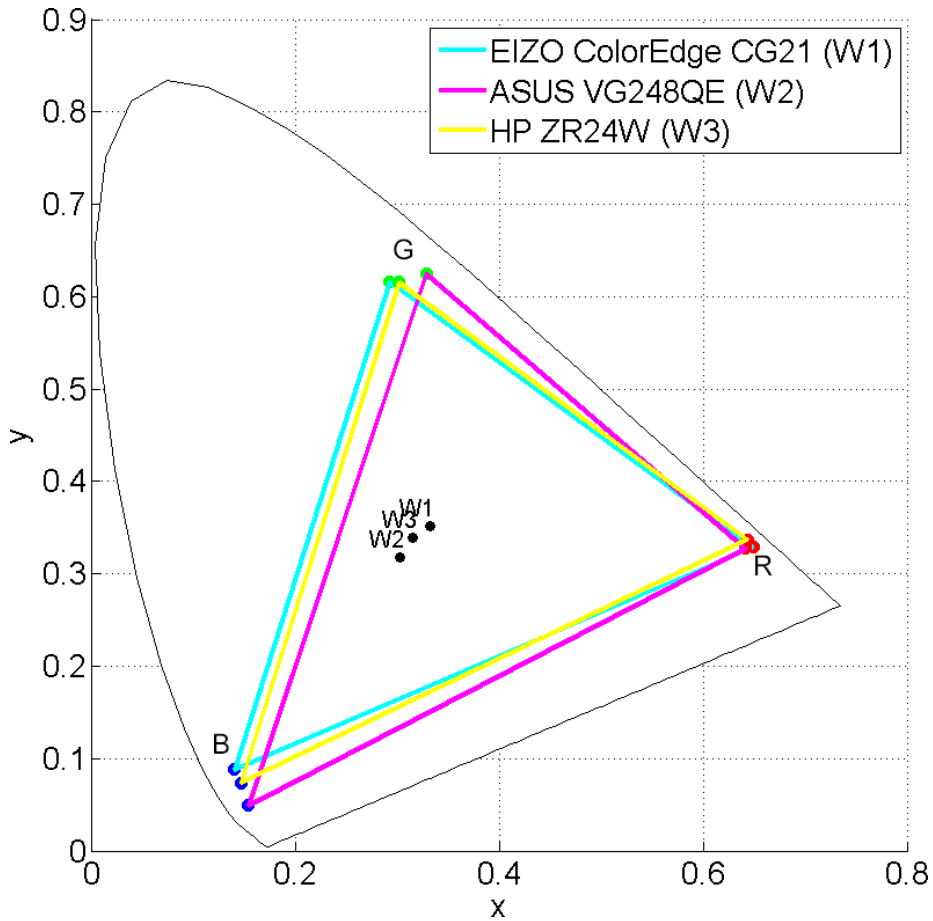



Figure 3.1: Color gamuts obtained by measuring the color coordinates of the three primaries for the same configuration in several monitors: Asus VG248QE (stereoscopic), Eizo ColorEdge CG21 (monoscopic), and HP ZR24w (monoscopic). As we can see, the gamut size is very similar in all three cases, and it is even slightly larger in the stereoscopic (Asus) case.

nology, CRT display technology has been replaced by LCD as the standard for display technology. Although many properties of LCDs, such as their photometric and temporal response, still remain unknown [80, 58], ample research is being done to characterize and improve their performance [81, 82, 83]. The characterization models traditionally used for CRTs may also be used for LCDs, although it is necessary to perform extensive measurements to ensure their suitability, since violations of some model assumptions might invalidate the results. In effect, an accurate characterization is required to identify the effects of possible imprecisions in the results of the experiments, and verify whether or not a specific display is suitable for the tests.

As we have seen earlier, human visual perception is mainly the result of the interaction between an electromagnetic signal hitting the retina and the HVS. In order to ensure the perceptual realism of the displayed stimuli, we must characterize and control the transformations sustained by the visual signal during visualization. The color characterization of a display allows us to find a transfer function that predicts the display's colorimetric response from a given input (forward transformation) and, conversely, to find the input needed to produce a given output (inverse transformation) [84, 85]. Therefore, the transfer function permits us to move between device-dependent and device-independent representations.

It is difficult to obtain a perfect characterization model that is valid in all possible situations; the recommendation is to find a simple model that is acceptable in most cases and then extend it

| | |
|---|---|
| Panel size: | 24.0” |
| True resolution: | 1920x1080 |
| Pixel pitch: | 0.2768 mm |
| Max. brightness: | 350 cd/$m^2$ |
| Viewing angle(CR+/-10): | 170°(H)/160°(V) |
| Gray-to-gray response: | 1 ms |
| Color depth: | 6 bits + FRC |

Table 3.1: Technical specifications of the Asus VG248QE monitor.

to deal with exceptions, rather than to find a very specific model that may not allow to be inverted to cover more general cases. Following this recommendation, we base our characterization model on the standard CRT model, as described by Brainard et al. [57], adapting the model for certain particularities of stereoscopic LCD displays. Indeed, most of the assumptions made in the standard CRT model also apply to LCDs, and this model can be easily adapted to cover the particular characteristics of the latter.

### 3.2.1 Previous considerations

The CRT characterization model considers that the light emitted by any given pixel in the screen is the result of a linear combination of the contribution from each of the color filters in the pixel plus the reflected light from the ambient illuminant.

$$\boldsymbol{C(\lambda) = rR(\lambda) + gG(\lambda) + bB(\lambda) + A(\lambda)} \tag{3.1}$$

Equation 3.1 represents this model, where $C(\lambda)$ is the spectrum measured at a pixel, $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$ are the spectral power distributions (SPD) measured when each of the color filters are fully saturated (primary colors), r, g and b are real numbers between 0 and 1 that indicate the weights for each filter, and $A(\lambda)$ represents the amount of ambient illumination (or "flare") emitted or reflected by the display when it shows a black signal. It is recommended to work with no ambient light in order to simplify the computations; to adapt the model for the presence of ambient illumination, we must simply replace $C(\lambda)$ with $C'(\lambda) = C(\lambda) - A(\lambda)$ and characterize supposing no ambient. We performed the characterization process in a dark room —where all external light sources were blocked, and bright surfaces were covered with black tissue to minimize reflections — with only the monitor light on, so that no ambient illumination was present (Figure 3.2).

Figure 3.2: Monitor characterization environment.

In order to select the right settings for the display before the characterization process, we tested several preset configurations and found that they mostly affect images in terms of brightness, contrast, and off-axis[1] quality, whereas, as shown in Figure 3.3, the color gamut barely changes. For this reason, we preferred to use the monitor settings values recommended by more exhaustive benchmarks [79], as shown in Table 3.2.

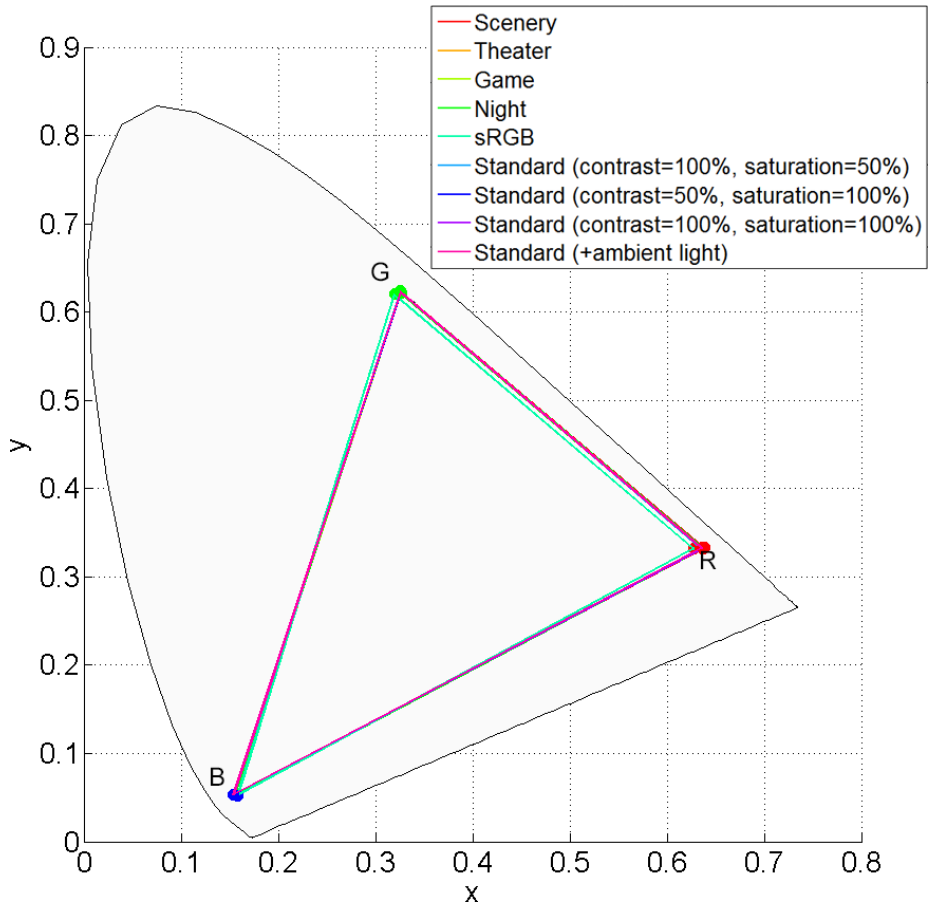



Figure 3.3: Gamut comparison of the Asus VG248QE preset settings. As we can see, there is not much difference in the resulting gamuts.

| | |
|---|---|
| Splendid mode: | Standard |
| Contrast (%): | 80 |
| Brightness (%): | 35 |
| RGB (%): | Red 96 / Green 94 / Blue 88 |

Table 3.2: Asus VG248QE settings used for its characterization and during all the observation tests.

Since we want to display images in stereoscopic mode, the entire characterization process was done using stereoscopic color samples, so that the monitor's stereoscopic mode is kept enabled throughout the process, and placing a pair of shutter glasses in front of the characterization device. This is important because the stereoscopic mode has an effect on both the colorimetry and the luminance of the output, as shown in Figures 3.4 and 3.5. We will discuss more in detail the effects of stereoscopy in the characterization process in Section 3.3. From now on, all the results will be shown under these conditions, unless otherwise stated.

[1] The off-axis angle is the horizontal offset of the viewing angle with respect to the display's normal.

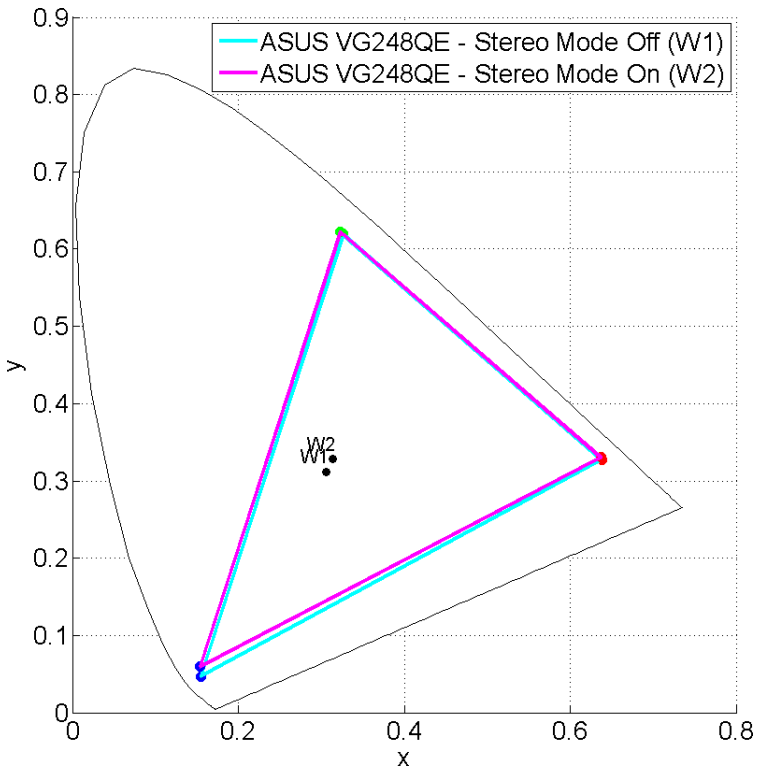



Figure 3.4: Gamut comparison of the Asus VG248QE with the stereoscopic mode On and Off. In stereoscopic mode, a pair of shutter glasses was also placed in front of the characterization device. Although the gamut size does not change much, the white point moves towards the green in stereo mode as a result of the filters in the shutter glasses.

We created a series of large RGB uniform color patches (see Table C.1 and Figure C.1) by setting the intensities of all R, G, and B channels to zero and increasing each channel's intensity separately from 0 to 255, with increments of 10 units. This resulted in three ramp sets for the red, green, and blue. Additional shorter series were created for Magenta, Yellow and Cyan, with intensity increments of 50 units, plus an *achromatic* series of grays. This results in seven color series and a total of 120 samples, used to estimate the parametric model that characterizes the response of the display. Additionally, an image of the 24-path Macbeth color chart was used to validate the estimated models, and black and white samples were measured throughout the characterization process to ensure the stability of the white and black point signals. Figure 3.6 shows the characterization setup and the measuring distances. Both the described color samples and the setup were used throughout the entire characterization process in this chapter.

#### 3.2.1.1 Pixel independence

One of the assumptions of the most common characterization models is that of pixel independence. Pixel independence means that the intensity of each primary color at each pixel location depends only on the input signal for that pixel, independent of the neighboring pixels [86]. In old CRT displays and the first LCD models this assumption of pixel independence could fail due to inadequate video bandwidth, DC restoration, and high-voltage regulation, caused by inaccuracies of the digital-to-analog converters (DACs) [87]. However, newer LCD displays – such as ours – are provided with a digital video input (DVI) which eliminates the conversion stage, thus improving pixel independence.

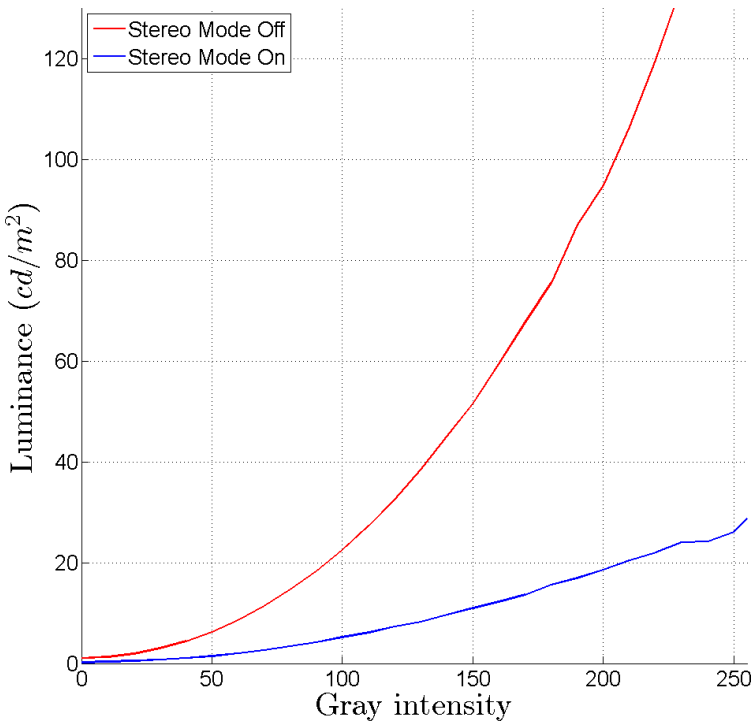



Figure 3.5: Luminance comparison of the Asus VG248QE with stereoscopic mode On and Off, for the characterization gray series with intensity from 0 to 255. We can see that the stereoscopic mode reduces drastically the luminance.

#### 3.2.1.2 Warm-up

LCD panels can take a long time to warm up due to the backlight used to illuminate the liquid crystal panel. Although our model uses a white LED (WLED) backlight —which is known to be more efficient— we performed stabilization measurements to ensure that the peak luminance was reached before proceeding with the characterization. We displayed a medium-tone gray image —code named *gray120*, as shown in Table C.1, at approximately a 50% intensity level, with RGB tristimulus [120 120 120]— on the monitor. Using a spectroradiometer[2], we took regular measurements (at about 10-minute intervals) of the SPD of the signal emitted by the display on three points evenly distributed along the central horizontal line of the screen (points P1, P2 and P3 in Figure 3.7). Once the graphs were stable during several measurements (Figure 3.8), we considered that the display had already reached its peak luminance. Unlike in older LCDs, we can see that our monitor has a very low warm-up time, which is almost immediate after being switched on.

#### 3.2.1.3 Angular dependence

We have seen that one of the biggest disadvantages of using LCD displays for visual experiments is the strong dependency of the screen's luminance on the viewing angle. We positioned the spectroradiometer facing the center of the screen, parallel to its normal; we then took several measurements on the screen by rotating the spectroradiometer's head so as to reproduce off-axis visualization. This way, we can represent more faithfully what an observer would perceive when positioned in its place. Figure 3.7 shows the different luminance levels —relative to the maximum luminance obtained for

[2] As described in Section 1.2.2.1 and A.19, a spectroradiometer is a device that measures the SPD of a light source (units $W \cdot sr^{-1} \cdot m^{-2}$ per unit wavelength). We used a Konica Minolta CS-2000, which is a non-contact spectroradiometer that measures contrasts of 100,000:1 and low luminance levels down to 0.003 $cd/m^2$.

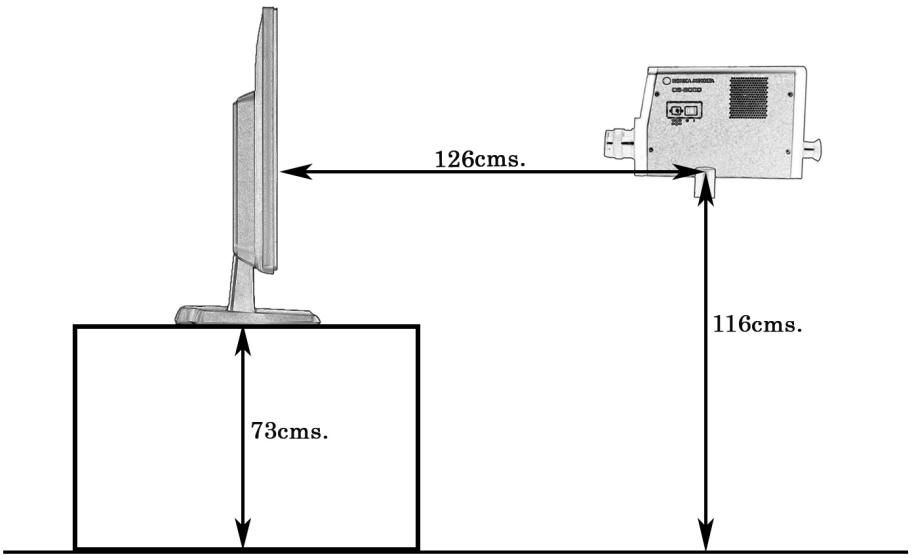
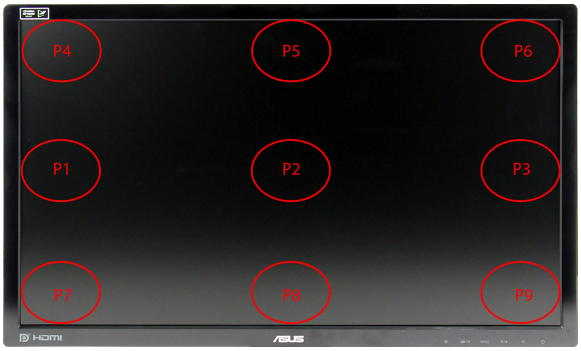



Figure 3.6: Monitor characterization setup.




Luminance values ($cd/m^2$):

| | | |
|---|---|---|
| 0.0697 | 0.1047 | 0.0955 |
| 0.1088 | 0.1587 | 0.1434 |
| 0.1340 | 0.1774 | 0.1608 |

Luminance deviations (%):

| | | |
|---|---|---|
| 8.90% | 5.40% | 6.32% |
| 4.99% | 0% | 1.53% |
| 2.47% | -1.87% | -0.21% |

Figure 3.7: Left: approximate locations of the points measured with the spectroradiometer on the Asus VG248QE monitor; Right: luminance uniformity values ($cd/m^2$) (top) and corresponding percentage deviations from the central position (bottom) obtained by spectroradiometer rotation for the gray120 color patch, with RGB values [120 120 120]. The measurements were taken on the nine points indicated on the left image.

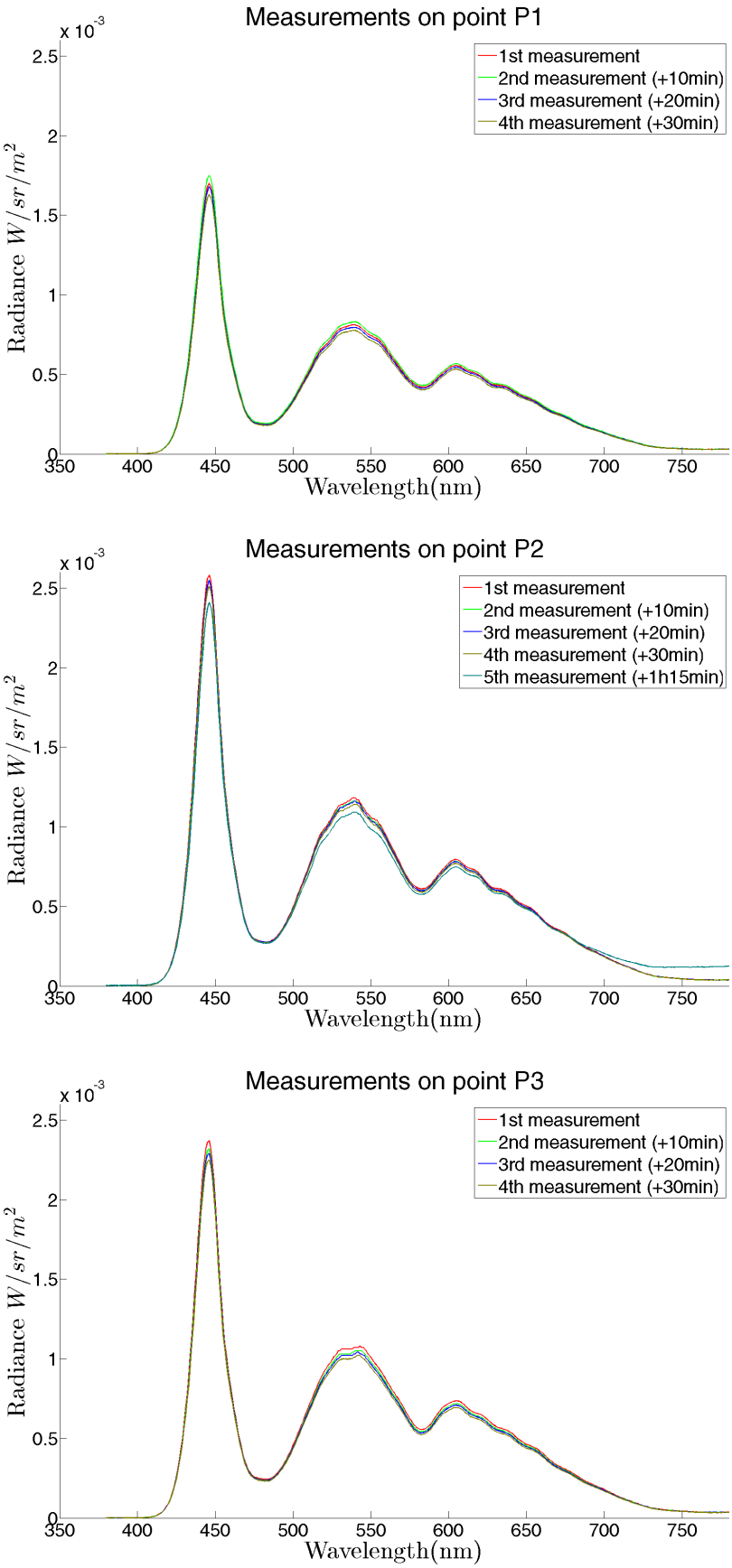



Figure 3.8: Display stabilization graphs measured at 10-minute intervals on points P1, P2, and P3 (respectively from top to bottom) for a 30-minute period (4 measurements). For reference purposes, we also included a fifth measurement taken on point P2 (central point), 1 hour and 15 minutes later.

all the color patches— measured on the display's screen for the same color patch, gray120, used in section 3.2.1.2. In effect, we can see that the luminance varies greatly for off-axis observation angles; however, as mentioned earlier, we believe that this effect can be minimized by restricting the position of the displayed objects to the center of the display, thus reducing the amount of off-axis observations. Decreasing the observation distance would also help mitigate this effect, given that shorter distances require larger deviation angles from the center of the screen for the observer to reach the edges of the display.

#### 3.2.1.4 Channel constancy

The model described by equation 3.1 assumes the presence of channel constancy in the display. Channel constancy implies that, when only one channel is active, varying the input intensity for that channel only changes the luminance, whereas the chromaticity remains constant. In terms of the SPD of the signal emitted by the display, channel constancy means that the responses of a given channel for two different inputs are proportional, so that $P^c i_\lambda = P^c j_\lambda \times k$, where $P^c i_\lambda$ and $P^c j_\lambda$ are the SPDs emitted by channel C as output for two input intensity values, i and j, respectively, and k is a multiplicative constant. In terms of chromatic properties, if $(x_i, y_i)$ and $(x_j, y_j)$ are the chromaticity coordinates of inputs i and j, respectively, then $x_i = x_j$ and $y_i = y_j$, and their respective luminance, $Y_i$ and $Y_j$, are such that $Y_i = Y_j \times k$.

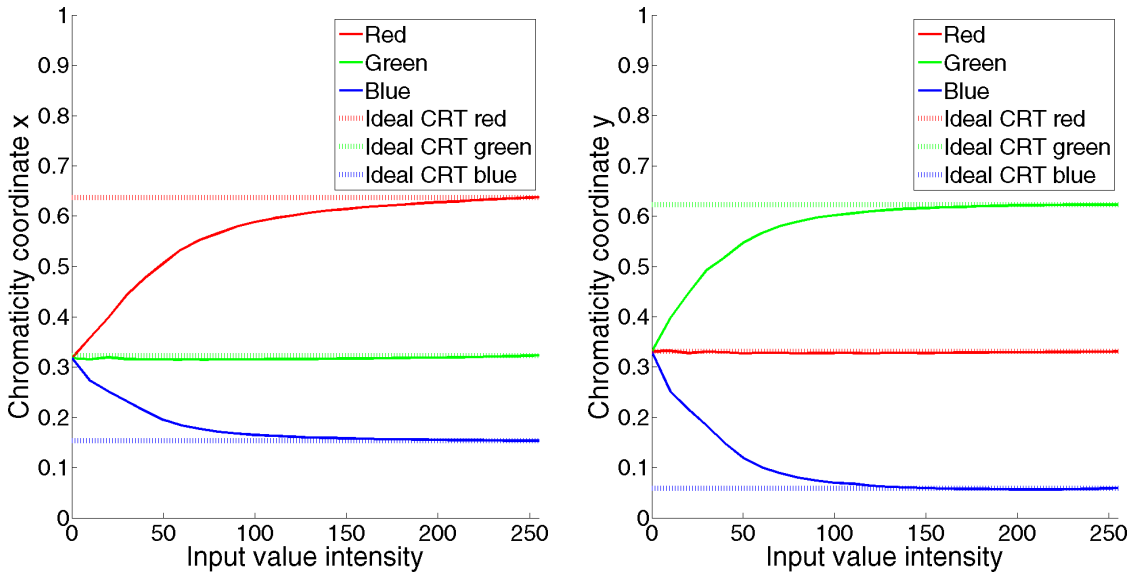



Figure 3.9: Asus VG248QE channel constancy graphs for the x (left) and y (right) chromaticity coordinates. The dashed lines represent the chromatic response of ideally constant red, green, and blue phosphors in a CRT display.

Figure 3.9 shows the channel constancy results for the Asus VG248QE. The graph displays the x and y chromaticity coordinates for all the red, green, and blue color series, in comparison with a constant signal with the same coordinates obtained by each channel at their maximum intensity. In the case of the x coordinate, the green channel remains almost constant, whereas channels red and blue have a very poor constancy for low intensity inputs and improve as the intensity increases; similarly, in the case of the y coordinate, channel red remains constant, but channels green and blue deviate from constancy at low intensities and behave almost linearly for higher values. Our studies suggest that the reason for this low color constancy is that, as we already explained LCD displays, unlike CRT displays —for which this calibration model was designed— do not produce light by means of independent monochromatic light sources, but through a common backlight and

several color filters. In the case of the Asus VG248QE, a white backlight is used, so the method used to produce this white light influences the response of each color channel. As we will see later, we can use more complex characterization models to take into account this lack of constancy.

#### 3.2.1.5 Channel independence

Another assumption of the CRT characterization model is that the light signals emitted by each color channel are independent from each other, so that no interferences exist between color channels. Given the current advances in the electronic components used for display devices, the assumption of channel independence is true in most modern displays. However, since a poor channel dependence may introduce additional color changes in the input data, it is important to verify that independence does indeed exist. Channel independence can be verified with an additivity test, consisting in comparing the intensity emitted for a white input with that of the combined primaries, to check whether changes on any of the channels has an influence on the others.

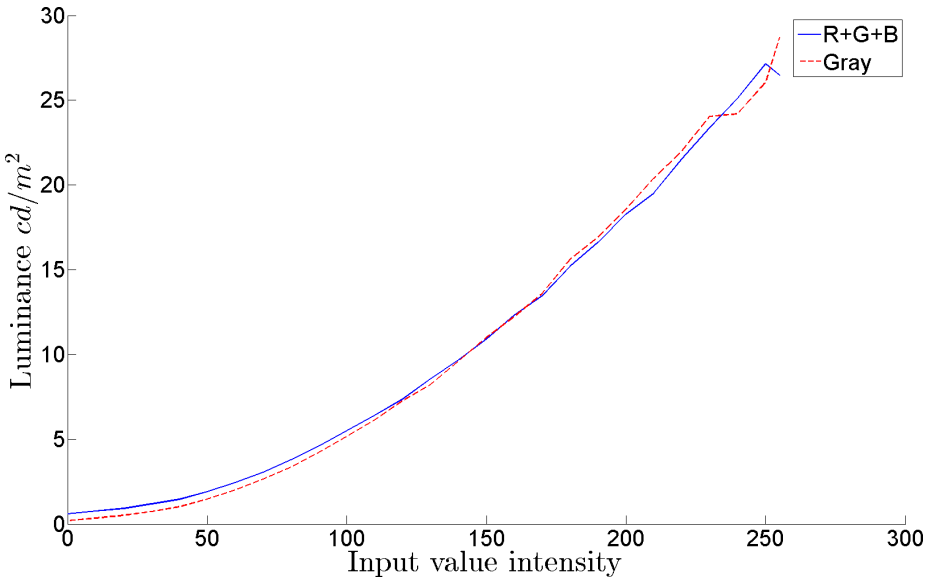



Figure 3.10: Asus VG248QE channel additivity test graph comparing the luminance of a gray series input of increasing intensity (dashed red line) with the summed luminance of the red, green and blue inputs of the same intensity (solid blue line).

Figure 3.10 shows the additivity of the luminance emitted for each a series of increasing-intensity inputs by the Asus VG248QE. We can see that there is a very small offset between the graphs, except at the limit of the display's gamut, where the measurements are less reliable. We attribute this offset to the the fact that, again due to their backlight, LCD displays do not produce a full-zero signal for black inputs. This can be observed both in Figure 3.9 and 3.10, which show that neither the luminance nor the chromaticity of the black input is zero. The signal emitted by the display for a black input is known as the display's black point and, if needed, can also be corrected in the characterization process by subtracting the black point color signal from every color output.

#### 3.2.1.6 Gamma correction

The pixel luminance response of a display is typically nonlinear, and is characterized by a function in the form of a power-law for CRT displays or a sigmoid, or S-curve, for LCDs. By generalization, this function is usually known as *gamma function*, and represents the relationship between the input color values and the luminance output for each pixel. By measuring the color samples in

Table C.1, we obtained the corresponding luminance curves for each color channel. As shown in Figure 3.11, these measurements can be used to find the parameters of the gamma function that best approximates the resulting shape for each color channel.

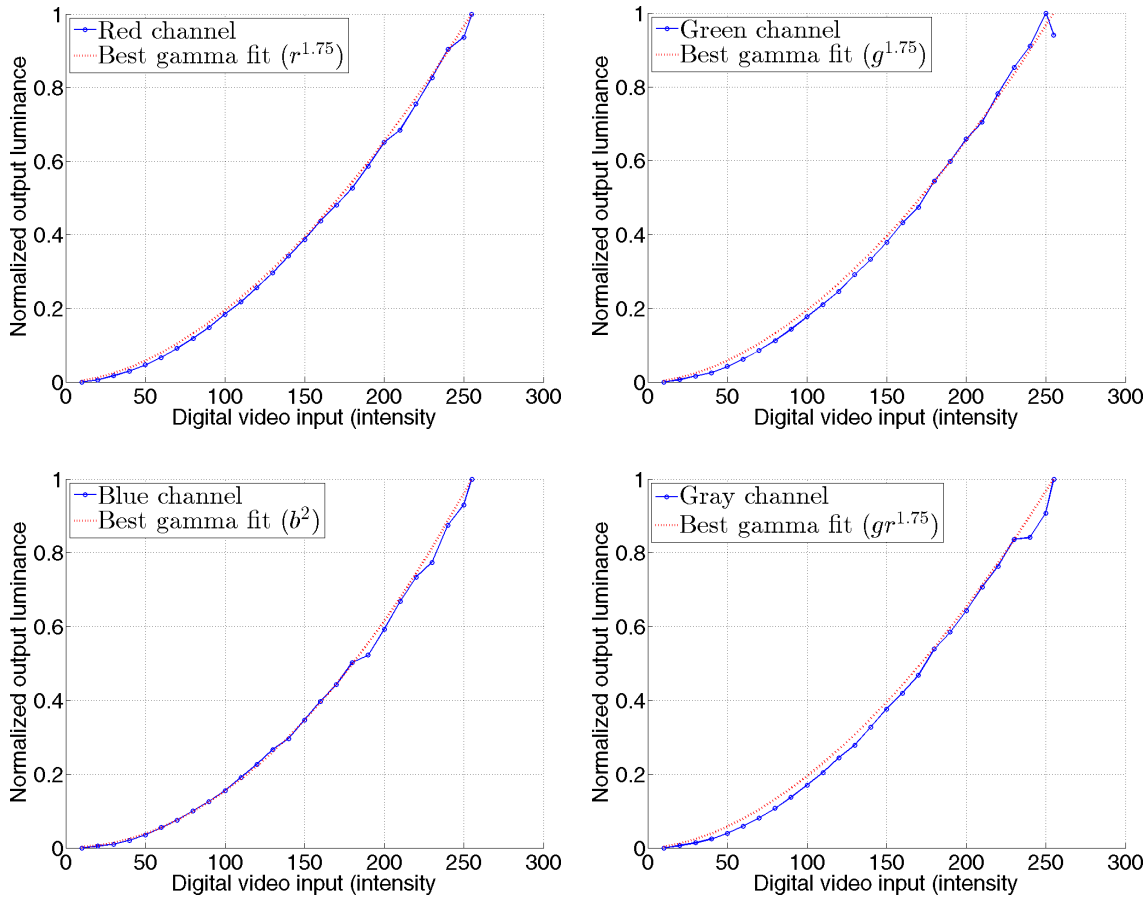



Figure 3.11: From left to right and top to bottom: normalize luminance distributions for the red, green, blue, and gray input series. The blue line shows the measured luminance values for a series of color samples with intensities from 10 to 255 in equal increments of 10 units; the dotted red line shows the best-fit gamma function approximation for the same input values, with $\gamma = 1.75$, $\gamma = 1.75$, $\gamma = 2$, and $\gamma = 1.75$, respectively.

Indeed, Figure 3.11 shows that the curves are very closely approximated by a gamma function, specially for intensity values below 200. We can then say that each color channel's luminance can be approximated by the following expression:

$$
\begin{aligned}
r = gamma(R) &= \left[\frac{R}{max(RGB)}\right]^{\gamma}, \; for\ R > 0 \;\; / \;\; r = 0 \;\; otherwise \\
g = gamma(G) &= \left[\frac{G}{max(RGB)}\right]^{\gamma}, \; for\ G > 0 \;\; / \;\; g = 0 \;\; otherwise \\
b = gamma(B) &= \left[\frac{B}{max(RGB)}\right]^{\gamma}, \; for\ B > 0 \;\; / \;\; b = 0 \;\; otherwise
\end{aligned}
\tag{3.2}
$$

where r, g, and b are the output luminance for each color channel, R, G, and B are the values in each pixel for the red, green and blue channels, max(RGB) is the maximum intensity for each

RGB color channel – which should be 255 or 1 if RGB values are given in a [0-255] or [0-1] scale, respectively – and $\gamma$ is the parameter that determines the shape of the gamma function. In our case, as illustrated in Figure 3.11, initial analyses via best-fit have estimated the gamma parameter as 1.75, 1.75, and 2, for the red, green and blue channels, respectively.

Similarly, we can use Expression 3.3 to estimate the input color value needed to obtain a given output.

$$\begin{aligned} R &= gamma(r)^{-1} = (max(RGB))r^{1/\gamma} \\ G &= gamma(g)^{-1} = (max(RGB))g^{1/\gamma} \\ B &= gamma(b)^{-1} = (max(RGB))b^{1/\gamma} \end{aligned} \tag{3.3}$$

However, although this approximation may be sufficient for simple characterizations, in cases like ours where colorimetry is important we must use more complex models, as we will see in Section 3.2.2.

### 3.2.2 Modeling the response of the display

The transfer function that characterizes the response of a display device is generally estimated from a set of known color samples. The premise is that, due to the electronics in the display, an input RGB value is displayed with a given output chromaticity and luminance which cannot necessarily be predicted from the input color intensities; the chromaticity response depends on the response of the display's color pixels, and the luminance response depends on the display's gamma function explained in Section 3.2.1.6.

Several groups of models can be used to perform this estimation: physical models, numerical models, and models based in 3D lookup tables (LUT). Physical models assume a channel independence and that the chromaticity of the primaries is constant; the colorimetric transform is done by weighting a combination of the chromaticity of the primaries by a standard luminance function (such as a power law, more appropriate for CRT displays, or a sigmoid function in the case of LCDs, as we saw in Section 3.2.1.6). Numerical models use a training color set to estimate the parameters of a polynomial function via a fitting process. Models based in 3D LUTs construct the LUT from a series of known color patches, and use interpolation methods to estimate unknown values; the quality of the results depends on the interpolation method and the amount of color measurements, although a large data set is often required to obtain precise results [88].

Physical models are the most commonly used for display characterization because, although it has been shown that they do not fit well for some LCD cases, they are fast, it is easy to estimate the response curve with a small amount of data [88], and allow for both forward and inverse transformations [89]. The general model used in CRTs assumes the presence of a brightness gain in the screen and two types of chromaticity offsets, which can also be applied to modern LCDs in many cases [90]. Fixing the values of some of these terms gives several variants of the model:

- Gain-Offset-Gamma-Offset (GOGO) $(a, b, c \neq 0)$

$$y_h = (a_h \cdot x_h + b_h)^{\gamma_h} + c \tag{3.4}$$

- Gain-Offset-Gamma (GOG) $(c = 0)$

$$y_h = (a_h \cdot x_h + b_h)^{\gamma_h} \tag{3.5}$$

- Gain-Gamma-Offset (GGO) $(b = 0)$

$$y_h = (a_h \cdot x_h)^{\gamma_h} + c \tag{3.6}$$

- Simple Gamma model ($a = 1, b = c = 0$)

$$y_h = x_h^{\gamma_h} \tag{3.7}$$

where $y_h$ is the gray level of channel H for a given digital input $x_h$; $a_h$, $b_h$, and $\gamma_h$, are the gain, offset, and gamma for the channel, respectively; and $c$ is a general offset that represents the environmental flare [88].

We measured the color samples in Table C.1 (Appendix C) in full-screen mode —so that the contribution due to the point-spread function from neighboring pixels can be neglected— and measured their spectra in the central position of the monitor using a Konica Minolta CS-2000 spectroradiometer[3]. Initial results showed that the simple gamma model could not estimate a good approximation to our problem. We computed a solution using each of the models above for a given set of measured color data, and obtained the coefficient of determination, $R^2$, for each solution. The simple model yielded very low average $R^2$ values, between 0.13 and 0.6, which were very far off from those obtained with the other three models. For that reason we decided to test only the first three models (GOGO, GOG and GGO) .

Figure 3.12 shows the chromaticity coordinates of all the measured color samples with regularly-increasing intensities for the Asus VG248QE. We can see that there is a strong nonlinearity (especially when for samples near the maximum channel's intensity), which justifies our choice to choose the more complex characterization models over the simple gamma model. Figure 3.13 shows the SPD of the three primaries —i.e. the red, green and blue color samples at full intensity— measured with the spectroradiometer under the same illumination conditions used in our experiments. If we look at the zoomed-in graph in Figure 3.14, we can see that the response of the red and green channels overlap with the blue channel (see the small blue peak around 450 nm); this overlap affects especially the red channel, whose response is much more powerful in this area of the spectrum. This is part of the reason why the chromaticity response shown in Figure 3.9 was not linear and, indeed, the response of the blue channel always affects that of the red (x chromaticity) or the green (y chromaticity) channels. This signal overlap must be taken into account, since it complicates things if we want to characterize the display with a simple model.

Let $[R_{in}\ G_{in}\ B_{in}]$ denote the device-dependent RGB tristimulus values of the input color stimulus displayed at each time on the monitor, and let coordinates $(r_{in}\ g_{in}\ b_{in})$ be the normalized RGB color coordinates corresponding to that same input stimulus, such that:

$$(r_{in} = R_{in}/255;\ g_{in} = G_{in}/255;\ b_{in} = B_{in}/255) \tag{3.8}$$

Similarly, let $[X_{out}\ Y_{out}\ Z_{out}]$ denote the device-independent XYZ tristimulus values, corresponding to the radiance measured by the spectroradiometer on the surface of the display, as output for the corresponding input stimulus. These values can be converted from the XYZ color space to RGB by multiplying by the corresponding transformation matrix, $M_{xyz2rgb}$, such that:

$$RGB_{out} = XYZ_{out} * M_{xyz2rgb} \tag{3.9}$$

where $[R_{out}\ G_{out}\ B_{out}]$ are the RGB tristimulus values estimated from the output stimulus $XYZ_{out}$, and similarly $(r_{out}\ g_{out}\ b_{out})$ refer to the normalized RGB color coordinates of the estimated stimulus $RGB_{out}$.

A simple method to estimate the characterization models is to estimate the parameters directly from the set of input RGB and output XYZ data via a Least Squares (LS) regression analysis which tries to fit the set of RGB-XYZ data correspondences to the chosen model —as given by

[3]The spectroradiometer is a Konica Minolta CS-2000. It is a non-contact spectroradiometer that measures contrasts of 100,000:1 and low luminance levels down to 0.003 $cd/m^2$.

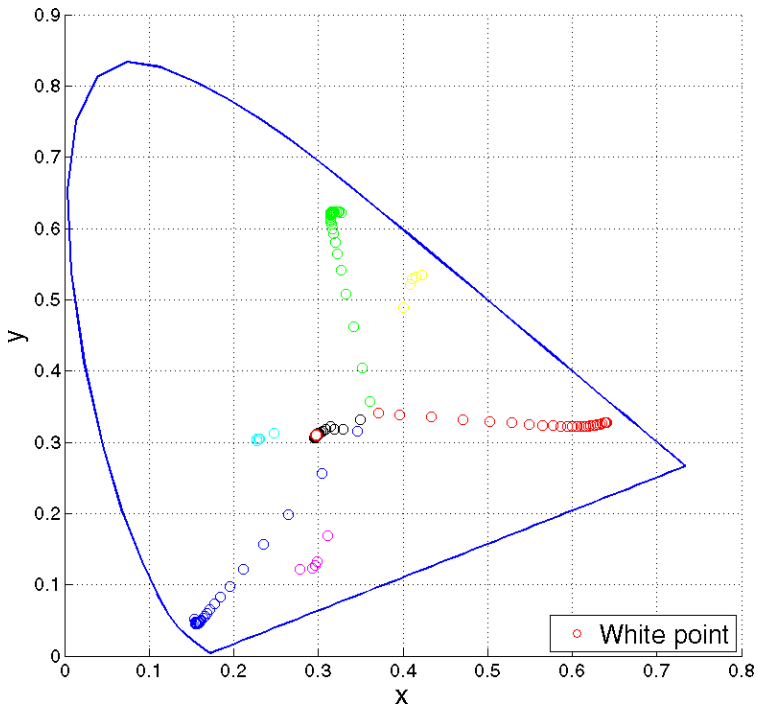



Figure 3.12: CIE diagram 1931 showing the chromaticity coordinates of all the characterization color samples measured at the Asus VG248QE, and the location of the white point.

the corresponding equation from Expressions 3.4 to 3.7— replacing $x$ by $RGB_{in}$ and $y$ by $XYZ_{out}$ in the model equation. Since the three chosen characterization models (GOGO, GOG and GGO) are defined by a nonlinear function, we perform a nonlinear least squares (NLLS) analysis[4], where there can be up to four parameters, and a much larger number of observations (in this case, 120). Tables 3.3 and 3.4 and Figures 3.15 and 3.16 show the individual and average goodness —averaged from the individual determination coefficients obtained for X, Y, and Z— of the fitted functions estimated via NLLS regression analysis for each of the three models for the direct DAC-XYZ[5] transform and the inverse XYZ-DAC transformation functions for each color channel. As we mentioned earlier, the non-linear response of the red channel complicates the estimation of XYZ from DAC, and viceversa, as demonstrated by the low determination coefficient ($R^2$) with respect to the other channels, notably in the case of the estimation of/from Z values, typically associated with the blue channel, which shows once again the dependency between both channels that we discussed earlier.

As we can see, $R^2$ is similar in most cases for all three models, although the best results in our case are obtained by the GOG and GGO models. The results for these two models are very similar, and they both use the same number of parameters; however, we decided to choose the GOG model because it does not include the flare offset term and therefore it is more appropriate for a characterization in the dark. However, although it can be used as a first approach to obtain the most adequate model, the estimation method shown above presents several problems. First, it performs the estimation between two different color spaces, XYZ and RGB. Furthermore, each channel is estimated separately without accounting for the effect of other channels, which is not correct because the tristimulus values in both RGB and XYZ are not independent from each other.

To improve the characterization results, we propose a different methodology, where data fitting

[4]Using Matlab R2012b's *fit* function [91].

[5]DAC is the acronym for Digital-to-Analog Converter; although our display has a digital input and therefore this term is not strictly correct, the term is widely used in the context of CRT characterization, so we continue to use it here for consistency.

| GOG (XYZ-DAC) | | | | |
|---|---|---|---|---|
| | ${R^2}_X$ | ${R^2}_Y$ | ${R^2}_Z$ | $AVG\ \ {R^2}_{XYZ}$ |
| Blue channel | 0.999 | 0.999 | 0.999 | 0.999 |
| Green channel | 0.998 | 0.998 | 0.996 | 0.997 |
| Red channel | 0.999 | 0.999 | 0.869 | 0.956 |
| **GOGO (XYZ-DAC)** | | | | |
| | ${R^2}_X$ | ${R^2}_Y$ | ${R^2}_Z$ | $AVG\ \ {R^2}_{XYZ}$ |
| Blue channel | 0.998 | 0.964 | 0.999 | 0.987 |
| Green channel | 0.998 | 0.998 | 0.975 | 0.990 |
| Red channel | 0.999 | 0.999 | 0.867 | 0.955 |
| **GGO (XYZ-DAC)** | | | | |
| | ${R^2}_X$ | ${R^2}_Y$ | ${R^2}_Z$ | $AVG\ \ {R^2}_{XYZ}$ |
| Blue channel | 0.999 | 0.999 | 0.999 | 0.999 |
| Green channel | 0.998 | 0.998 | 0.996 | 0.997 |
| Red channel | 0.999 | 0.999 | 0.869 | 0.956 |

Table 3.3: Individual and average $R^2$ for the estimation of output XYZ values, from input DAC values, with each of the proposed models: GOG, GOGO, and GGO.

| GOG (DAC-XYZ) | | | | |
|---|---|---|---|---|
| | ${R^2}_X$ | ${R^2}_Y$ | ${R^2}_Z$ | $AVG\ \ {R^2}_{XYZ}$ |
| Blue channel | 0.999 | 0.998 | 0.999 | 0.998 |
| Green channel | 0.999 | 0.999 | 0.997 | 0.998 |
| Red channel | 0.999 | 0.998 | 0.870 | 0.955 |
| **GOGO (DAC-XYZ)** | | | | |
| | ${R^2}_X$ | ${R^2}_Y$ | ${R^2}_Z$ | $AVG\ \ {R^2}_{XYZ}$ |
| Blue channel | 0.992 | 0.998 | 0.999 | 0.996 |
| Green channel | 0.997 | 0.999 | 0.997 | 0.998 |
| Red channel | 0.999 | 0.989 | 0.869 | 0.952 |
| **GGO (DAC-XYZ)** | | | | |
| | ${R^2}_X$ | ${R^2}_Y$ | ${R^2}_Z$ | $AVG\ \ {R^2}_{XYZ}$ |
| Blue channel | 0.999 | 0.999 | 0.999 | 0.999 |
| Green channel | 0.999 | 0.999 | 0.997 | 0.998 |
| Red channel | 0.999 | 0.998 | 0.867 | 0.955 |

Table 3.4: Individual and average $R^2$ for the inverse estimation of input DAC values, from output XYZ values, with each of the proposed models: GOG, GOGO, and GGO.

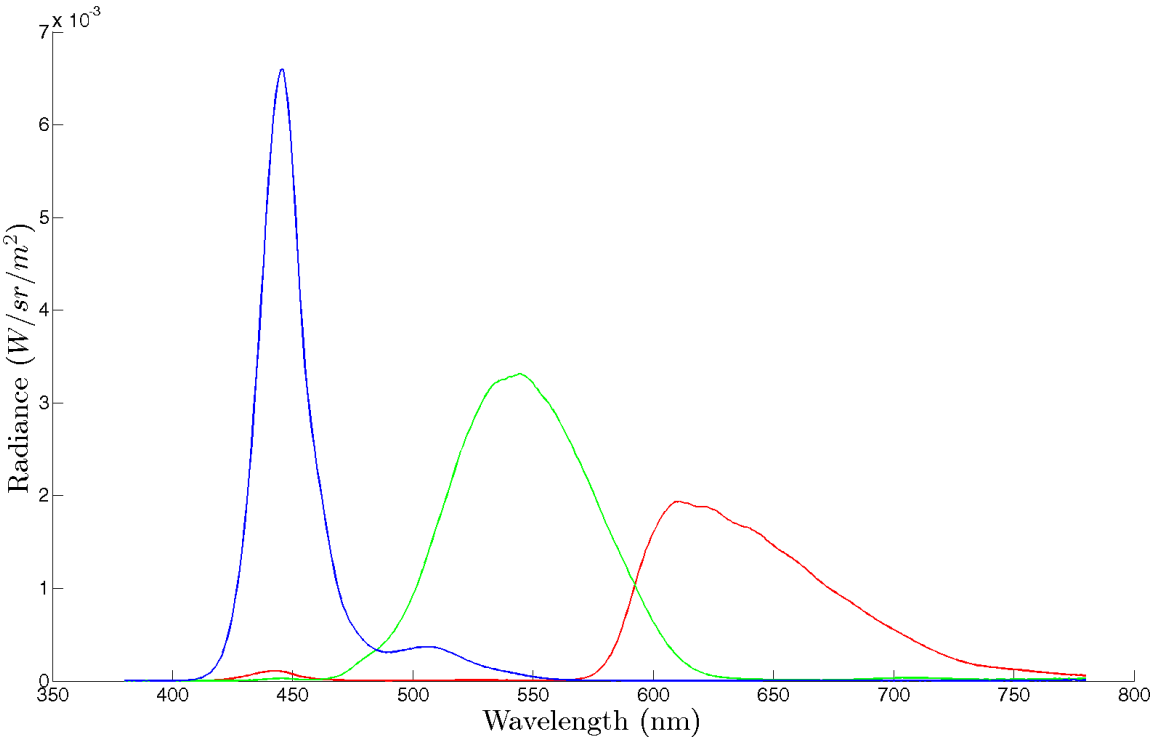



Figure 3.13: Spectral radiance curves for each of the three primary colors of the Asus VG248QE.

is not performed directly between output XYZ values and input RGB values. Instead, we perform a previous color space conversion from XYZ to RGB, and then obtain the model parameters by NLLS analysis within the same color space. The general structure of the algorithm that finds the parameters of the GOGO characterization model (Expression 3.4), or any of its simplified variants (Expressions 3.5- 3.7), is the following:

1. Compute the transformation matrix from the XYZ (device-independent) color space to the RGB (device-dependent) color space.
2. Use the matrix obtained previously to estimate output RGB ($RGB_{out}$) tristimuli from displayed XYZ output tristimuli ($XYZ_{out}$) —i.e. the cone responses to the measured radiances.
3. Perform data fitting between $RGB_{in}$ and $RGB_{out}$, and estimate the model's parameters for each channel's data.

As we said earlier, the characterization model that we are using assumes color constancy and channel independence. According to these assumptions, since we know the XYZ tristimulus values of the display's RGB primaries, we can estimate the XYZ tristimulus of any RGB input as a linear combination of the primaries. We can then estimate the inverse transformation matrix $M_{rgb2xyz}$, and the corresponding direct transformation $M_{xyz2rgb}$, from the tristimulus values of the primaries, as shown in Expressions 3.10 and 3.11 respectively.

$$M_{rgb2xyz} = \begin{bmatrix} X_{R255} & X_{G255} & X_{B255} \\ Y_{R255} & Y_{G255} & Y_{B255} \\ Z_{R255} & Z_{G255} & Z_{B255} \end{bmatrix} \tag{3.10}$$

$$M_{xyz2rgb} = M_{rgb2xyz}^{-1} \tag{3.11}$$

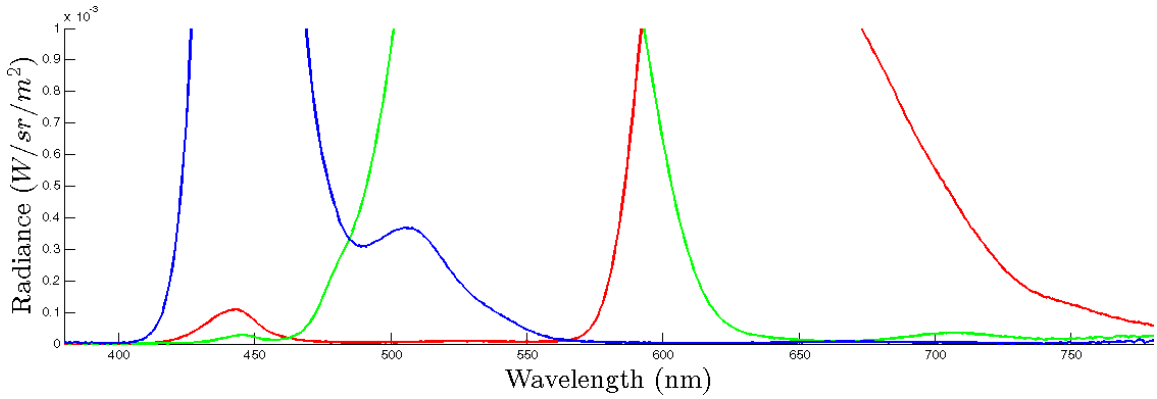



Figure 3.14: Zoom of the bottom section of Figure 3.13. The response of the red and green channels overlap with the blue channel for wavelengths around 450 nanometers.

$$M_{rgb2xyz} = \begin{bmatrix} 10.83 & 9.67 & 5.68 \\ 5.62 & 18.65 & 2.20 \\ 0.54 & 1.63 & 29.18 \end{bmatrix} \tag{3.12}$$

$$M_{xyz2rgb} = \begin{bmatrix} 0.12645 & -0.06383 & -0.01980 \\ -0.03808 & 0.07320 & 0.00189 \\ -0.00021 & -0.00291 & 0.03453 \end{bmatrix} \tag{3.13}$$

where $[X_{R255}, Y_{R255}, Z_{R255}]$, $[X_{G255}, Y_{G255}, Z_{G255}]$, and $[X_{B255}, Y_{B255}, Z_{B255}]$ are the XYZ tristimulus values of the maximum-intensity red, green, and blue samples, respectively. Matrix $M_{xyz2rgb}$ can be used to convert any XYZ output value to RGB output, using Expression 3.9. The transformation matrices obtained for the Asus VG248QE are shown in Expressions 3.12 and 3.13.

For the set of characterization color samples shown in Figure C.1, we obtained the parameters shown in Table 3.5 for the forward transformation using the GOG model on the Asus VG248QE monitor.

| Parameter | Red channel | Green channel | Blue channel |
|---|---|---|---|
| Gain (a) | 267.6079 | 258.6558 | 257.9680 |
| Offset (b) | 0.0458 | 0.0572 | 0.0202 |
| Gamma ($\gamma$) | 0.5256 | 0.5055 | 0.5534 |
| $R^2$ | 0.9995 | 0.9984 | 0.9986 |

Table 3.5: Asus VG248QE characterization results with the forward GOG model.

We are more interested, however, in the inverse transformation, because it allows to estimate the input that is required to obtain the desired output. In order to obtain the inverse transformation, we performed the same NLLS regression analysis, but using the inverse GOG function, that is:

$$y_h = a_h \cdot (x_h^{\gamma_h} - b_h) \tag{3.14}$$

where $y_h$ is the required input gray level of channel H to obtain a given digital output $x_h$, and $a_h$, $b_h$, and $\gamma_h$, are the inverse gain, offset, and gamma, for the channel, respectively. The results obtained for the inverse model are shown in Table 3.6; notice how the values obtained for the gamma parameter are very similar to those obtained experimentally in Section 3.2.1.6 (Figure 3.11).

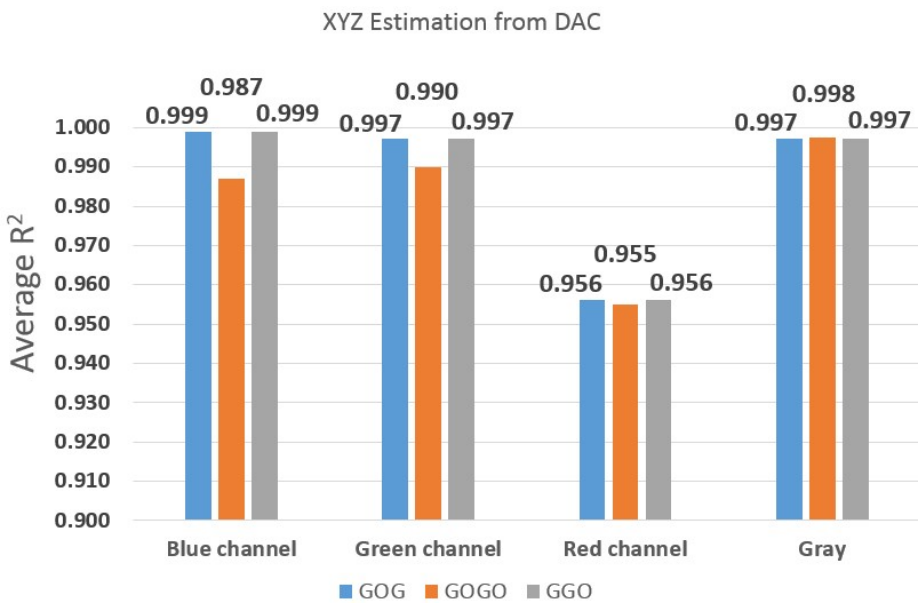



Figure 3.15: Average $R^2$ for the estimation of output XYZ values, from input DAC values, with each of the proposed models: GOG, GOGO, and GGO (values in Table 3.3).

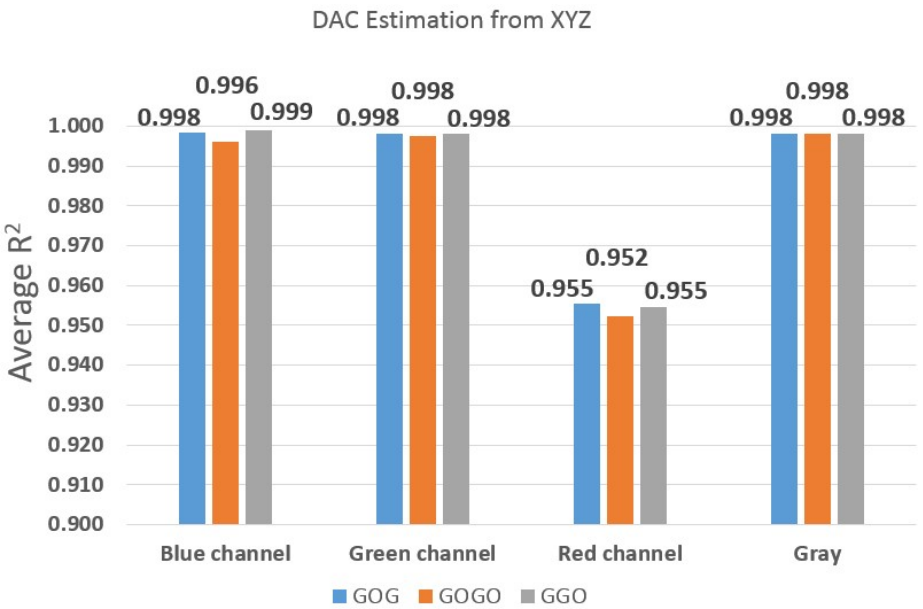



Figure 3.16: Average $R^2$ for the inverse estimation of input DAC values, from output XYZ values, with each of the proposed models: GOG, GOGO, and GGO (values in Table 3.4).

| Parameter | Red channel | Green channel | Blue channel |
|---|---|---|---|
| Gain (a) | 0.0038 | 0.0041 | 0.0039 |
| Offset (b) | 0.0218 | 0.0000 | 0.0000 |
| Gamma ($\gamma$) | 1.8132 | 1.7649 | 1.7156 |
| $R^2$ | 0.9995 | 0.9976 | 0.9991 |

Table 3.6: Asus VG248QE characterization results with the inverse GOG model.

Figure 3.17 shows the correlation between the real RGB input values and those estimated from the measured output in our tests; as we can see, there is an almost linear correlation, which is consistent with the high values obtained for the coefficient of determination $R^2$. At the same time, this method produces much better results than the one shown earlier.

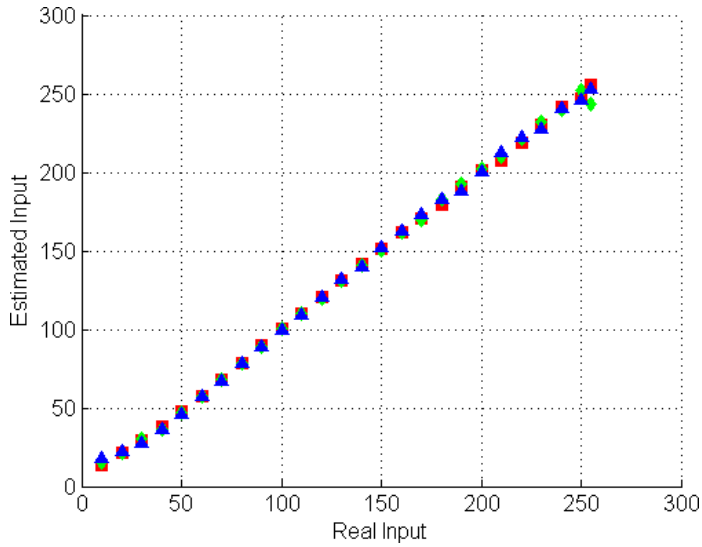



Figure 3.17: Asus VG248QE RGB correlation between the real and estimated RGB input values with the GOG model.

## 3.3 Effects of stereoscopic visualization on the color signal

Characterizing the response of the monitor is not enough to control the entire chain of visualization. As we saw earlier in Section 3.2.1, the use of stereoscopic visualization has a strong effect in the final image, both in terms of luminance and chromaticity, so it is necessary to account for this contribution. As a complement to the luminance graph shown in Figure 3.5, Figure 3.18 shows the normalized spectral transmittance of the Nvidia 3D shutter glasses used as the stereoscopic visualization device. This graph shows the percentage amount of the total incoming signal that is transmitted through the glasses; as we can see, the transmittance is very low, with a maximum value of around 15%, which explains the low luminance in Figure 3.5 with respect to the monoscopic mode.

In order to account for the contribution of stereoscopy, we have characterized our display using stereoscopic visualization throughout the entire process so that our characterization model already includes the contribution of stereoscopy. For this, we created a stereoscopic version of our 120 color samples, in order to force the monitor to enable the stereoscopic mode at all times. We then positioned a pair of shutter glasses in front of the spectroradiometer (Figure 3.19) and took the

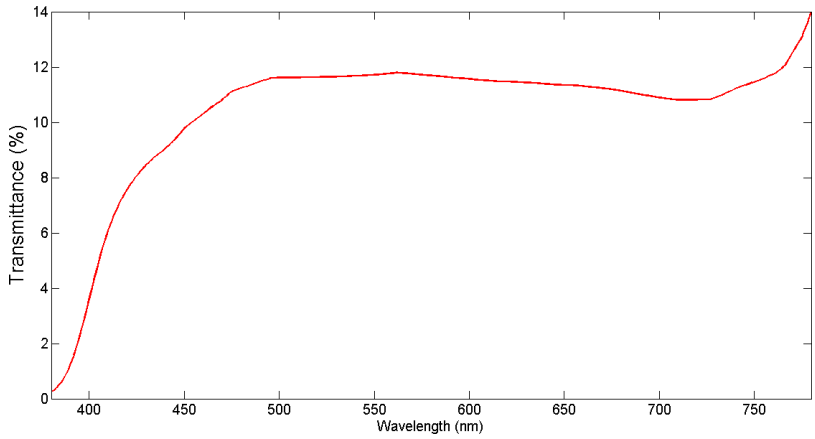



Figure 3.18: Spectral transmittance of the Nvidia 3D shutter glasses.

measurements as usual. This way, the spectrometer "sees" what a human observer would see in the same position.

Figure 3.19: Stereoscopic characterization with a spectrometer and the Nvidia 3D shutter glasses.

## 3.4 Conclusion

Chapter 2 showed the importance of controlling the output signal emitted by the display to avoid loosing perceptual information that might affect higher level perceptual processes when observers view the simulations. Although we focus our work on the use of a specific monitor model, this chapter presents an brief analysis of the most common display technologies, and which mechanisms they use to produce color, so that the methodologies used in this work can be easily adapted to different technologies without loss of generality. We have provided an analysis of the most important factors that must be taken into account when LCD technology is used to display the simulations, and how we can model their colorimetric output and apply the necessary corrections when need.

Although CRT displays have been used for a long time in psychophysical experiments, lately they have been replaced by LCD displays. LCD technology produces color in a different way than CRT does, which means that new characterizations models are required to account for the differences

between these two technologies. In terms of color reproduction, we have seen that the use of backlight illumination in LCDs introduces irregularities in the color signal, notably in terms of color channel independence. For this reason, simple characterizations where the color response of the display is modeled by a color-conversion matrix and the luminance response as a simple gamma function, do not produce good data estimations. Therefore, more complete models are required to account for the effect of additional factors like color gain and offset. The use of stereoscopic visualization in the observation tests introduces additional constraints in the characterization process. In this case, the colorimetric response of the stereoscopic visualization devices must be taken into account, as well as the implications of enabling the "3D mode" in the display. The filters used in the stereoscopic visualization devices ("3D glasses") are not always colorimetrically neutral, which means that they introduce modifications in the color signal emitted by the display. Additionally, when "3D mode" is enabled in the display, the frame refresh rate increases up to 2.5 times the refresh rate in monoscopic mode, but also reduces the overall display luminance. We take all of these effects into account in our methodology by performing the entire characterization process with stereoscopic color samples —so as to force the "3D mode" to stay enabled— and putting a pair of 3D glasses in front of the measuring device.

We have presented here two characterization methodologies. The first methodology estimates the model parameters directly from device-independent XYZ output values and device-dependent RGB inputs. We have seen that although this methodology produces good estimations in the average, the dependencies between color channels, most notably between the red and blue channels, reduces the goodness of the estimation for display stimuli where these two channels are involved. To correct for this problem, we propose a second characterization methodology where out XYZ values are fist converted to RGB so the model parameters are estimated between data sets in the same color space. This second methodology has shown to produce better results that the first one for our tests monitor.

We present our characterization methodology using XYZ as device-independent color space, although it is worth noting it could also be performed using the LMS color space. The absence of bimodal responsivity curves in the LMS cone responsivity curves —as opposed to XYZ— would simplify the modelization of the colorimetric response of the display; in this case the colorimetric response of the display could be considered as a special case of a chromatic adaptation. However, despite the latest advances, the absence of color models to account for factors such as contrast of color space uniformity, or the lack of standard objective transforms between RGB, XYZ, and LMS, are some of the reasons why we have decided to use XYZ instead. Indeed, XYZ has been accepted as the standard by the CIE for a long time, so there are objective color conversion between device-dependent RGB and XYZ. Furthermore, XYZ is the most commonly used independent color space in the field of physical rendering. Therefore, this choice simplifies the integration of our results in the simulation process.

Chapter 4

# Visual Psychophysics: Perception-guided display of sparkling materials

## Contents


**4.1 Description of the problem** . . . . . **60**
**4.2 Perceptual characteristics of sparkling materials** . . . . . **61**
**4.3 Modeling the human visual response with a digital camera** . . . . . **62**
4.3.1 Effects of camera optics on sparkle perception in photographs . . . . . 63
4.3.2 Obtaining radiometric information from the scene using photographs . . . . . 64
**4.4 Perceptually-aware acquisition of photographs of sparkling materials** . . **68**
4.4.1 Representing the real observation conditions in a photograph . . . . . 68
4.4.2 Adding binocular information to the photographs . . . . . 71
4.4.3 Image calibration: Adapting color information to the visualization environment 71
4.4.4 Effects of image calibration in the perception of sparkle in photographs . . . 73
**4.5 Perceptual reproduction of luminance dynamics in photographs of sparkling materials.** . . . . . **75**
4.5.1 Perceptual effects of image exposure in photographs of sparkling materials . . 75
4.5.1.1 Methodology . . . . . 76
4.5.1.2 Results . . . . . 77
4.5.2 Perceptual effects of dynamic range compression in photographs of sparkling materials . . . . . 80
4.5.2.1 Problem description . . . . . 80
4.5.2.2 Methodology . . . . . 82
4.5.2.3 Results . . . . . 83


We have described the mechanisms of the human visual system (HVS) to generate an initial representation of the objects in a scene, using sensory response information, as well as some of the physiological mechanisms that can influence this perception depending on observation and environmental aspects. We have also discussed some particularities associated with the amount of perceptual image realism when we use a display to present the simulations to human observers. We have shown a methodology to control the colorimetric signal of the display devices to minimize the loss of information during visualization, maximizing perceptual realism. In this chapter, Section 4.3 presents a methodology to capture the visual information from a scene, using a Digital Single-Lense Reflex (DSLR) camera to represent the behavior of the earlier stages of the HVS in terms of perceptually-aware sensory responses. This allows us to reproduce in the simulated images as much as possible of the perceptually relevant information present in the real scene.

Section 4.4 describes the effects of the imaging process in the perception of sparkles. The perceived aspect of the car paint materials that we study is strongly dependent on the light of

the scene and the material composition. In the case of sparkling materials, sparkling occurs when the light in the scene is reflected by the metallic flakes in the materials, as a result of their much higher specularity with respect to the rest of components in the material. In photographs, the light reflected by the material results in image luminance, and is a determining factor in the perceived size and depth position of the sparkles. Therefore, maintaining perceptual light information during the photograph acquisition process is a key factor to conserving the perceived aspect of the materials. Given the capabilities of camera sensors, it is not possible to capture the entire dynamic range of a real scene in the photographs. Thus, we must choose which information to select and which information to neglect in the photographs.

Section 4.5 presents two studies supported by results from visual psychophysical experiments. First, we try to find a correspondence between the sensitivity of the HVS to the light in a scene, and the sensitivity of the camera sensor, using image exposure as a measure of sensitivity (Section 4.5.1). We then present a second study where we use the results from the first one to analyze the effects of representing the image luminance with different dynamic ranges (Section 4.5.2).

## 4.1 Description of the problem

The visual aspect of paint coating materials is intimately linked to their material composition and physical properties, but also to the environment under which they are observed. As explained by Balcetis and Dunning [10], the material aspect perceived by an observer is the result of a bottom-up process where sensory organs and perceptual systems work automatically to produce a visual representation of the stimulus as the light in the environment arrives in the eyes of the observer after interacting with the materials. At the same time, this perception is also biased by top-down influences such as the cognitive and psychological state of the observer, or the observation environment. For this reason, physically-realistic images that ignore these properties are not correct when simulations are destined to be viewed by human observers; in this case, perceptual realism is needed, in order to account for the effect of human visual perception.

The rendering engine creates simulations of the real materials from a series of physical models and measurements that describe their microstructure. In order to link this structure with the macrometric aspect perceived by observers upon visualization, psychophysical studies must be performed to analyze the perceptual characteristics of the materials under a controlled observation environment. The case of sparkling materials is specially delicate. The micrometric metallic particles contained in this type of materials are distributed spatially on one of the material layers at random depths and orientations. Their high specular reflectivity produces strong reflections of the illuminating light, which are the origin of the sparkling effect. Therefore, an important factor to produce perceptually-realistic simulations of the sparkling effect is to understand how the human eye perceives luminance contrasts in the real scenes, and how this wide luminance range can be represented in the simulations.

When illuminated, a real scene is constantly emitting an almost infinite number of luminance levels, either directly from the light sources, or reflected from the objects in the scene; this is what we refer to as the dynamic range of the scene. When we look at an object in such a scene, the capabilities of our visual system —which is determined by the spatial distribution of the photoreceptors in the retina and their sensitivity to light— introduce a simplification in the perceived signal, compressing and reducing the dimensionality of the original information. When a scene is simulated, the images resulting from the simulations already contain information in a similar reduced response space; however, the complexity of the visualization devices may produce modifications in the information. Since we want to produce perceptually-realistic simulations, we must verify that the information remains perceptually accurate at each step of the process. In this work, we concentrate in the study of luminance perception in the visualization of real samples and images of sparkling materials, using stereoscopic visualization to provide additional depth information from binocular disparity. Since

the sparkle of the metallic flakes is perceived differently by each eye, this helps represent more accurately the material texture and, in turn, improve perceptual realism.

There are several ways to analyze the visual perception of the materials in a real and virtual context, the two most intuitive being the following:

- **Direct comparison:**

  The direct method is possibly the most straightforward. Intuitively, one could think of asking observers to compare material simulations directly with a real reference and give their opinion in terms of similarity. Nonetheless, although this might work for subjective analyses, there is no simple way to associate objective parameters between a simulation and a real scene. The visual response is directly dependent on the radiance values arriving on the retina of the eyes; if we had a way to measure at any given moment the radiance reaching the eyes of the observer, we could simply compare these values with those in the simulation. However, this is not practical, because it would imply measuring the radiance at every point of the scene for every image.

- **Indirect comparison:**

  To simplify the comparison process and introduce more objective criteria, we propose to replace the HVS with a DSLR camera. Cameras are based on an objective and validated model, whose correlation with the HVS is widely known. In this sense, the camera acts as a macrometric integrator of the real scene, and can also transform a radiometric signal into colorimetric information, the same way the human photoreceptors do.

This section presents the methodology and results of two observation experiments using the indirect comparison method with a DSLR camera. The first experiment analyzes the perceptual effects of acquiring sample images with varying image luminance via the comparison of sample photographs with a real reference scene; the second experiment studies the perceptual effects of different tone-reproduction methods for the display of sample photographs via the side-by-side comparison of sample photographs with variable dynamic ranges. Additionally, we will show how stereoscopy can improve the perceptual realism of the virtual images with respect to the real reference, supported by results from the conducted psychophysical experiments.

## 4.2 Perceptual characteristics of sparkling materials

The perceived visual aspect of a material is indeed strongly dependent on its optical properties and, in turn, on its composition, as we discussed earlier. As a result, the complexity of a material directly increases the number of optical phenomenon taking place as light hits its surface. Keeping this in mind, we distinguish two main types of paint coating materials in our studies: uniform and sparkling materials. Although our studies may apply to both types of materials, given the lack of perceptual effects resulting from uniform materials, we concentrate on the analysis of sparkling materials.

Automotive paint coatings are multi-layer composite materials with two basic purposes: protection and attractiveness [92]. Depending on their function and composition, each layer may have a variable thickness. Our materials are formed by four main layers: electrodeposition coating, primer surface, base coat, and clear coat (Figure 4.1). In the case of sparkling materials, the base coat layer consists of a pigment which contains micrometric-scale metallic (aluminum) particles, or flakes, of between 8 and 80 micrometers in size, with most particles around 10-40 micrometers [3]. These metallic flakes are deposited in the pigment following a pseudo-random distribution at different depths and orientations. In effect, as described by Streitberger and Dösel [92], during the application process an adjustable atomizer generates paint droplets of various sizes: small droplets contain

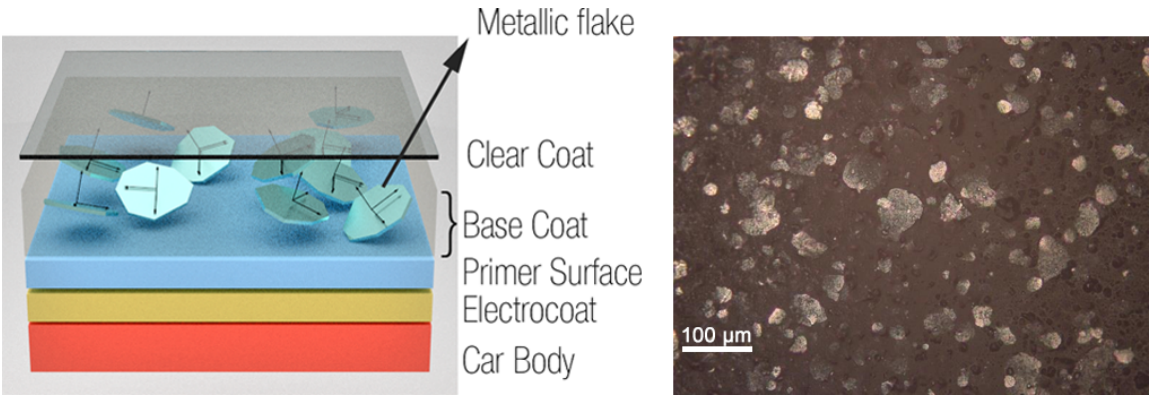



Figure 4.1: Material structure diagram (left), and microscopic closeup view of a paint layer of approximately 679 x 511.3 $\mu$m (right). Illustrations by Fernando Da Graça [20] (left) and Mona Ben Achour [3] (right).

a high portion of metallic flakes, which are distributed evenly and aligned parallel to the surface; on the other hand, the larger paint droplets contain more metallic flakes that form conglomerates, which prevents an even distribution and parallel alignment in the film. The sparkling effect is the result of light reflected at the surface of the metal particles. This effect will depend on many factors like particle size, shape, orientation, roughness, chemical composition —the lightness varies with the observation angle (lightness flop) in aluminum effect pigments, as does the color hue (color flop) in the case of mica effect or interference pigments— spatial distribution, or dispersion, which in turn are determined by the pigment manufacturing process and the paint making process.

As we saw in Section 2.4.1, a person with normal visual acuity can resolve elements of up to 8.86 mm in size from a 6-meter distance, which means that the largest particles (80 $\mu m$) should —if they could focus— be clearly visible to the eye when observed from a distance of around 5.5 cms. However, the sparkling effect is only the perceptual optical manifestation that results from the interaction of the metallic flakes with the light, so what we see when we look at sparkling materials is actually the sparkles and not the physical flakes. Given the high specularity of metals, the size of the sparkle —which is always larger that the flake— depends on the illumination and observation conditions, amongst other factors, as we will see later. Since flakes can be found at different depths and orientations, it is possible that the sparkle produced by large flakes makes them appear much bigger than they are [93, 94]. Similarly, shiny flakes located in deeper layers might appear the same size as duller ones that are closer to the surface. For all these reasons, it becomes necessary to characterize the materials properly from a perceptual point of view.

## 4.3 Modeling the human visual response with a digital camera

The perceptual characterization of this type of paint materials implies the analysis of the relationships between the physical attributes and the perceptual aspect of the material as a function of the illumination and observation conditions. Since the final goal is to simulate this behavior on a computer, we must also understand how these relationships work in the simulations. This can be done in many different ways. As mentioned earlier, the most straightforward way would be to compare directly the aspect of real and virtual samples, and use similarity metrics to evaluate the perceptual accuracy of each rendering parameter. For instance, we could produce images with different brightness values and ask a group of observers to select the image that looks the most similar to the reality; that way, we would find the brightness value that best approximates the real scene for a specific environment. However, this criterion would be purely subjective and there would be no objective metric to link brightness in the simulation with the perceived aspect of the real scene.

We know that the visual response perceived by an observer when viewing an image (real or virtual scene) depends ultimately on the radiometric signals that hits the retina of the observer's eyes. Therefore, if we could measure these radiometric signals at the moment of viewing the real scene in our example, we could associate those radiance values to the selected brightness value, using objective criteria. Using a calibrated DSLR camera in between the real scene and the simulation allows us to create an approximative "map" of the radiometric signals received by the camera's sensor and, by extension, by the photoreceptors in the retina of the observer.

As we explained in Section 4.2, cameras can simulate well the response of the human eye, integrating electromagnetic waves arriving onto the sensor, and translating it into perceptually relevant images of the stimuli. Therefore, it provides us with a flexible tool to study visual perception, giving us access to colorimetric and radiometric information, as well as giving us more control over the imaging process.

In general we can find many similarities between a DSLR camera and the HVS: both have a lens to focus the light; the amount of light passing through the lens is controlled by a diaphragm-like mechanism, which is the shutter in cameras and the pupil in the eye; and a virtual inversed image is projected on the sensor in a camera, and the retina in the eye. However, there are also important differences between them which must be considered in order to obtain accurate results with this approach. Notably, we must keep in mind that cameras can have optics of different quality to focus the light onto the sensor, and that each optical system may respond to light in a different manner. For this reason, it is necessary to fully characterize a camera before we can use it to study visual perception in psychophysical experiment and then use always the same camera.

For our experiments we chose a Nikon D800 DSLR camera, which is a model that offers a very high resolution (7360 x 4912 pixels) and sensitivity, and a large dynamic range, which is important given the amount of detail, shadows, and highlights (sparkles), in our images. We used a standard 50 mm FX lens[1], which we consider sufficient to obtain enough detail at the desired observation distance of around 70 cms, without introducing deformations. In this work, we have characterized two aspects of the camera: the effect of the optics on the incoming light, and the transfer function that converts the light's spectral power distribution into image pixel values.

### 4.3.1 Effects of camera optics on sparkle perception in photographs

As we discussed in Section 4.2, when we observe metallic-flake materials we must differentiate between the perceived sparkle size and the actual flake size. We explained that the sparkling effect is the perceptual manifestation of the physical flakes, which results from the interaction of the light with the material by means of different optical phenomena, most notably, specular reflection. The size of the sparkle resulting from these interactions is always larger than the flake producing it, as a consequence of several factors like the pixel size and resolution of the camera sensor, and the lens' point-spread function (PSF), which describes the magnitude of the diffraction that occurs when the light passes through the lens of the camera.

The sensor of our camera measures 35.9 x 24 mm, so a 50 mm lens would have a horizontal field of view (FOV) of nearly 40 degrees, and a vertical FOV of 27 degrees. The image size is 7360 x 4912 pixels, so assuming an acquisition distance of 70 cms, the entire scene falling in the camera's field of view will be imaged with a size of 50.4 x 33.6 cms. This corresponds to a resolution of roughly 146 image pixels per centimeter in the real scene. Since, as we will see later, we are imaging paint samples 15 cms wide by 10 cms high, that means that a sample will be imaged with a total size of roughly 2190 x 1460 pixels.

We mentioned earlier that the size of the perceived sparkle is always larger than the actual flake that produces it. Figure 4.2 shows a small area of a photograph of a paint sample, magnified with a

[1]Lens specifications: Nikon AF-S FX NIKKOR 50mm f/1.4G Lens with Auto Focus. Fix 50 mm focal length. Aperture values from F1.4 to F16.

700 factor, where the bright pixels correspond to the sparkle produced by a metal flake. As we can see, all the sparkles take up more than one pixel. Assuming a particle size between 5 and 80 $\mu m$, and a resolution of 146 image pixels per centimeter, most particles —specifically, any particle smaller than 68.5 $\mu m$— would have a size smaller than one pixel. Consequently, under some conditions —such as low illumination or particle orientations at low-reflection angles— it is possible that the sparkle produced by these particles would be imaged with one entire pixel, causing pixelation (as explained in Section 2.4.1) and losing image information like shape and color.

Figure 4.2: Zoom of a paint sample photograph taken from 70 cms. The area corresponds to an area of 80 x 80 pixels (0.55 cms) with a 700 magnification factor. The photograph has been taken with an exposure time of 1/15 seconds and then corrected (brightness and contrast) for illustration purposes. As we can see, all the sparkling flakes (brighter pixels) are imaged with more than one pixel.

Additionally, when the light in the scene passes through the optical elements of the camera, some diffraction may occur depending on the quality of the components. Good components will have a smaller diffraction and, therefore, will produce a much closer image to the reality. As we mentioned earlier, the amount of diffraction that occurs in a camera is modeled by its PSF, which describes the imaging response to a point. Bad lenses may produce images with an even larger sparkle size. The computation of the PSF is complex, but we can study its effect by analyzing the intensity profile of a sparkle point in the image. Figure 4.3 shows the intensity values[2] measured in a 8-pixel neighborhood around a sparkle point. The PSF determines how much the pixel's value is "spread" to the neighboring pixels. As we can see in the figure, the pixel intensity value is spread to a 4-pixel neighborhood (2 pixels on each side), and remains stable as the intensity value of the pixel changes. Therefore, since the dispersion does not change with the intensity, we can assume that it is caused by the actual specular reflection of the metal flake, and not by the optical elements in the camera.

### 4.3.2 Obtaining radiometric information from the scene using photographs

DSLR cameras create a color image of a real physical scene by converting electromagnetic radiation in the visible spectrum into color tristimulus values. The light from the scene is concentrated by the lens, then passes through a color filter array (CFA), before reaching the sensor, which stores light information as a scalar for each sensor pixel. The CFA uses spectrum-selective filters to integrate

[2] Note that the maximum intensity values in the graph decrease as the exposure time increases. This is because these values were measured on images that were calibrated to meet the dynamic range of a display, as we will see in Section 4.4.3. Nonetheless, this does not affect our observations about the variations in the intensity profiles.

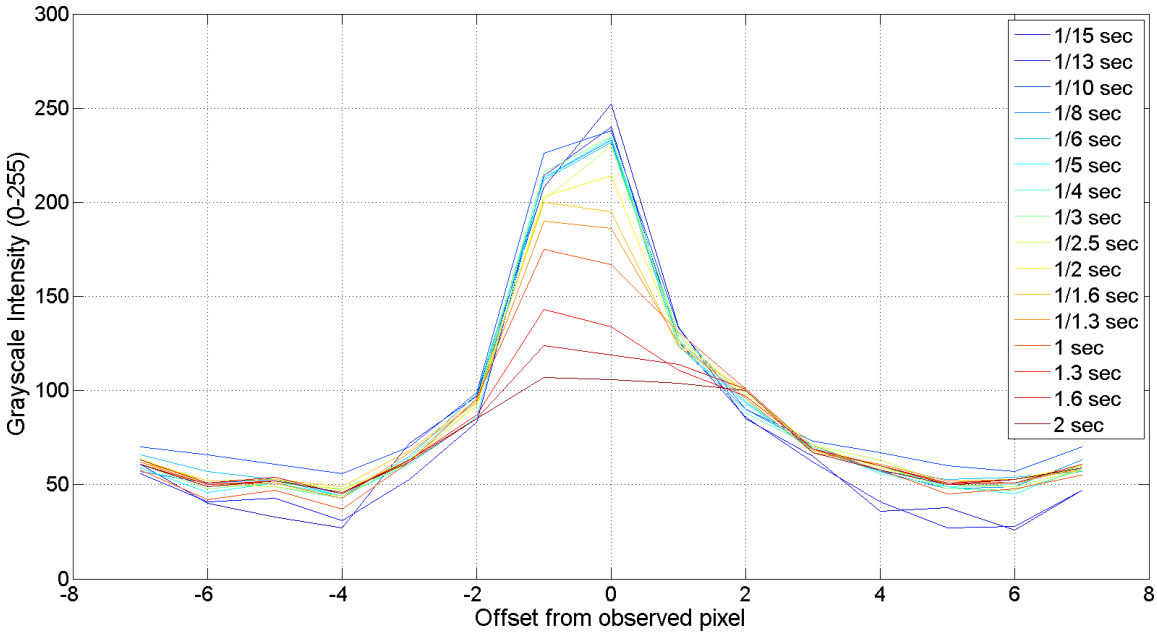



Figure 4.3: Intensity profile[2] of an image area around a sparkle pixel as a function of the total exposure. Only the shutter speed was changed, keeping all the other camera parameters fixed.

spectral light signals for each of the three color channels (red, green, and blue), emulating the function of the photoreceptors in the human retina [95].

Once a light signal has been integrated —i.e. converted into tristimulus value, either by the photoreceptors in the eye or by any other type of filter— the spectral information is lost, and cannot be recovered. All we can do is approximate the tristimulus response (or any of its metamers[3]) produced by the original spectral signal, in any type of physiologically-aware color space like XYZ or LMS. The HVS has been studied for a long time, and many experimental results have made possible the creation and normalization of direct and inverse transformations (typically, in the form of a transformation matrix) both between colorimetric spaces and physiological responses, and between different observation conditions. However, each camera produces images in a specific device-dependent tristimulus space, which depends on the spectral sensitivities of the imaging sensor's CFA. These sensitivities often differ from one camera to another and, in turn, are different from the device-independent RGB space based on the standard CIE color-matching functions (CMFs, see definition in A.33) [96, 97]. This means that the transformations must be calculated individually for each camera and each set of observation conditions.

The goal of the radiometric camera characterization is to estimate the transformation that predicts the tristimulus values of the raw sensor response —expressed in a device-independent color space, like XYZ— that produces a given device-dependent RGB color stimulus in the camera's color space, and viceversa [98]. This transformation is given by Expression 4.1:

$$\begin{aligned} X &= a_{11}R + a_{12}G + a_{13}B + v_1 \\ Y &= a_{21}R + a_{22}G + a_{23}B + v_2 \\ Z &= a_{31}R + a_{32}G + a_{33}B + v_3 \end{aligned} \tag{4.1}$$

where $a_{ij}$ are the elements of the color calibration matrix $A$, and $v_i$ are the elements of a vector $V$ that represents the contribution of the environment light to the scene. We can represent this same expression in matrix form as:

[3]See Section A.30 for a description of metamerism.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \times \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{4.2}$$

Furthermore, we can integrate matrix $A$ and light vector $V$ into one sole characterization matrix $M = (A|V)$, obtaining the following expression:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & v_1 \\ a_{21} & a_{22} & a_{23} & v_2 \\ a_{31} & a_{32} & a_{33} & v_3 \end{pmatrix} \times \begin{pmatrix} R \\ G \\ B \\ 1 \end{pmatrix} \tag{4.3}$$

The camera characterization process consists in finding a linear transformation between RGB and XYZ tristimuli given a set of N color correspondences. Let C denote an N x 3 matrix, where each row $c_i$ is the RGB tristimuli vector measured for color sample i, and let U denote another N x 3 matrix, where each row $u_i$ is the XYZ tristimuli vector corresponding to the radiance of sample i. We can then extend expression 4.3 as follows:

$$\begin{pmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \\ \cdots & \cdots & \cdots \\ X_N & Y_N & Z_N \end{pmatrix} = \begin{pmatrix} R_1 & G_1 & B_1 & 1 \\ R_2 & G_2 & B_2 & 1 \\ R_3 & G_3 & B_3 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ R_N & G_N & B_N & 1 \end{pmatrix} \times \begin{pmatrix} a_{11} & a_{12} & a_{13} & v_1 \\ a_{21} & a_{22} & a_{23} & v_2 \\ a_{31} & a_{32} & a_{33} & v_3 \end{pmatrix}^T \tag{4.4}$$

which is equivalent to:

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \cdots \\ u_N \end{pmatrix} = \begin{pmatrix} c_1 & 1 \\ c_2 & 1 \\ c_3 & 1 \\ \cdots & \cdots \\ c_N & 1 \end{pmatrix} \times \begin{pmatrix} a_{11} & a_{12} & a_{13} & v_1 \\ a_{21} & a_{22} & a_{23} & v_2 \\ a_{31} & a_{32} & a_{33} & v_3 \end{pmatrix}^T \tag{4.5}$$

and, more generally

$$U = C \times M^T \tag{4.6}$$

where $M^T$ is the transpose matrix of M, the 3 x 4 characterization matrix that maps device-dependent RGB stimuli to device-independent XYZ stimuli. We can solve this problem using a least-square fitting approach, by finding the matrix $M$ that minimizes the error $E$ between the measured XYZ tristimulus $u_i$ and the estimated XYZ tristimulus $u'_i$, as shown by the following expression:

$$E = \sum_{i=1}^{N} \left( {u_i}^T - {u'_i}^T \right) = \sum_{i=1}^{N} \left( {u_i}^T - M {c_i}^T \right) \tag{4.7}$$

The matrix $M$ that minimizes $E$ is given by

$$M = \left( C^T \times C \right)^{-1} \times \left( C^T \times U \right) \tag{4.8}$$

where $C^T$ is the transpose matrix of C, and $C^{-1}$ is the inverse.

Performing a target-based calibration carries many difficulties that must be taken into account. Some of these issues have to do with eye-camera metamerisms [97] or the specularity of the color target patches.

The problem of eye-camera metamerisms in target-based characterizations result from not taking into consideration the spectral responses of the camera's sensor. Several solutions exist to deal with this issue: Hong et al. propose to average color differences obtained with characterization matrices from different color sets [97]; alternatively, Alsam and Finlayson propose to find the set of all possible metamers by finding the set of convex combinations of the reflectance spectra from all the color samples in the target [99]. The presence of color patches with high specularity and saturation may translate in the presence of outliers which make difficult the fitting process. This is the case for color patches such as white or bright orange. Joshi et al. propose a solution based on averaging measurements from photographies of the color targets taken with varying illuminations [100]. Finally, a common solution to improve the results for target-based characterization is to use a multi-polynomial approach, such as the ones presented by Renani et al. [101] or Hong et al. [97].

For simplification reasons, we performed the characterization of the Nikon D800 DSLR camera using the target-based process – using the Macbeth color chart (see A.36) – since this turned out to be sufficient for our purposes. The characterization process consists of two parts, radiometric and colorimetric acquisition, both of which are performed in the same environment and under the same illumination conditions. The target was positioned at 70 cms from the measuring devices, the same observation distance used in the experiments, which we consider a reasonable distance for a computer display.

In the radiometric acquisition stage, we used a spectroradiometer[4] to measure the radiance emitted by each color field in the color chart under those illumination conditions, using a wide angle (1°) to integrate the measured area. In the colorimetric acquisition stage we took a raw photograph of the color chart under the same illumination conditions. This procedure results in a set of 24 correspondences between radiometric and colorimetric data.

The total luminance of a photograph depends mainly on four parameters: scene illuminance (L), film speed or sensitivity (ISO), f-number or lens aperture (N), and shutter speed (t). L relates to the illumination conditions in the environment, and the ISO is inherent to the film used -– although digital cameras can emulate different film speeds – whereas N and t can be set in the camera. The f-number and shutter speed combined result in the exposure value (EV), according to the following formula:

$$EV = \log_2 \frac{N^2}{t} \tag{4.9}$$

Since all the measurements in the characterization process are taken under the same illumination conditions, L will be constant throughout the process. Given the nature of our samples – which contain large amounts of small size shiny particles per unit area – we must pay attention to the sharpness of the image; both the ISO and the f-number have an effect over the image sharpness, the first one affecting image noise, and the second one affecting the depth of field (or the amount of image that is in focus). To simplify the process, we chose to keep the ISO and f-number constant during the characterization – set experimentally to numbers found to give acceptable results in terms of the amount of oversaturated and undersaturated pixels – so that the resulting image luminance be directly dependent on the shutter speed and, in turn, on the exposure value, as indicated by expression 4.9.

The camera was characterized for several exposure values, so that it could be used in later tests. The illumination was kept constant for short intervals of shutter speed values, reducing the intensity

[4] The spectroradiometer is a Konica Minolta CS-2000. It is a non-contact spectroradiometer that measures contrasts of 100,000:1 and low luminance levels down to 0.003 $cd/m^2$.

| | |
|---|---|
| Focal length: | 50 mm |
| Aperture: | f/8 |
| ISO: | 800 |

Table 4.1: Fixed camera parameters for the radiometric characterization of the Nikon D800.

| Exposure set | Macbeth white luminance ($cd/m^2$) | Shutter speed values (seconds) |
|---|---|---|
| 1 | 132.32 | 1/15, 1/13, 1/10 |
| 2 | 34.4 | 1/8, 1/6, 1/5, 1/4, 1/3 |
| 3 | 14.82 | 1/2.5, 1/2, 1/1.6, 1/1.3, 1 |
| 4 | 3.7 | 1.3, 1.6, 2 |

Table 4.2: Exposure sets with variable luminance for the characterization of the Nikon D800, indicating the shutter speeds in each set and the luminance of the white square in the Macbeth color chart.

as the exposure time increases to minimize oversaturation[5]. This allows to characterize using the same camera settings in all cases. In total we characterized for 16 different exposure values, divided into four different shutter speed value sets, as summarized in Table 4.2. The rest of the settings are summarized in Table 4.1.

## 4.4 Perceptually-aware acquisition of photographs of sparkling materials

### 4.4.1 Representing the real observation conditions in a photograph

We use photographs of paint samples as an intermediate step to determine the relationship between the HVS and the material rendering model. Because these photographs will be used in real-virtual comparisons, we must be able to control the acquisition conditions in order to reproduce them later on in the rendering engine.

The sample photographs were captured under a controlled light environment inside a dark room. We simulated solar daylight using a set of Solux incandescent lamps at 4700 Kelvins, running at 12 volts and 50 watts each, with a 36-degree illumination cone. We used a set of four lamps for global illumination, plus an additional directional light focused on the paint plate being photographed (Figure 4.4). The global illumination is used to increase the overall ambient light, and therefore the illuminance arriving on the paint plates; similarly, the directional lighting brings up the sparkling effect, resulting in a type of texture that authors Kirchner and Ravi refer to as *Glint impression* [102].

To capture the sample photographs we used a custom-built support (Figure 4.5) consisting of a T-shaped metal structure with a short bar (the T's crossbar) intersecting perpendicularly a longer bar on one of the edges. The longest bar holds the paint sample in front of the camera mount, allowing to regulate the observation distance; the shortest bar serves as the mount for the camera and is distance-marked, which permits stereoscopic image acquisition as well, by allowing different offsets in the camera viewpoint.

We took the photographs at a distance of roughly 70 centimeters, which we found to be a comfortable distance for an observer sitting at the observation post while allowing for the entire sample to stay within the standard 10-degree view angle. Using the stereoscopic acquisition bar described above, we took three photographs for each case in Table 4.2: centered view, left view, and

[5] As a result, sparkling areas reflect much less light in long-exposure photographs than in short-exposure ones, which translates into lower intensity values in the image, as we saw in Figure 4.3.

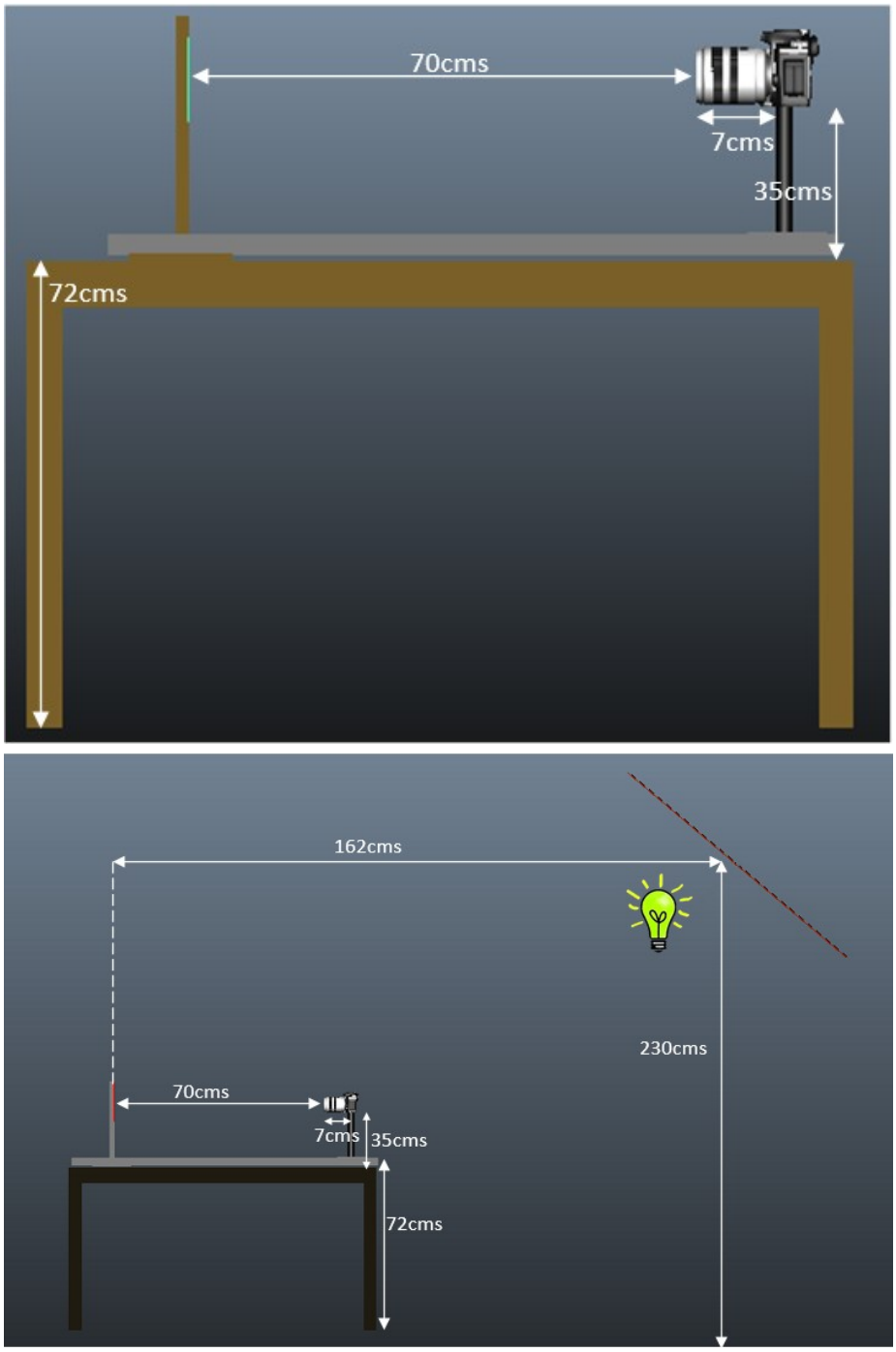



Figure 4.4: Acquisition setup for sample photographs.

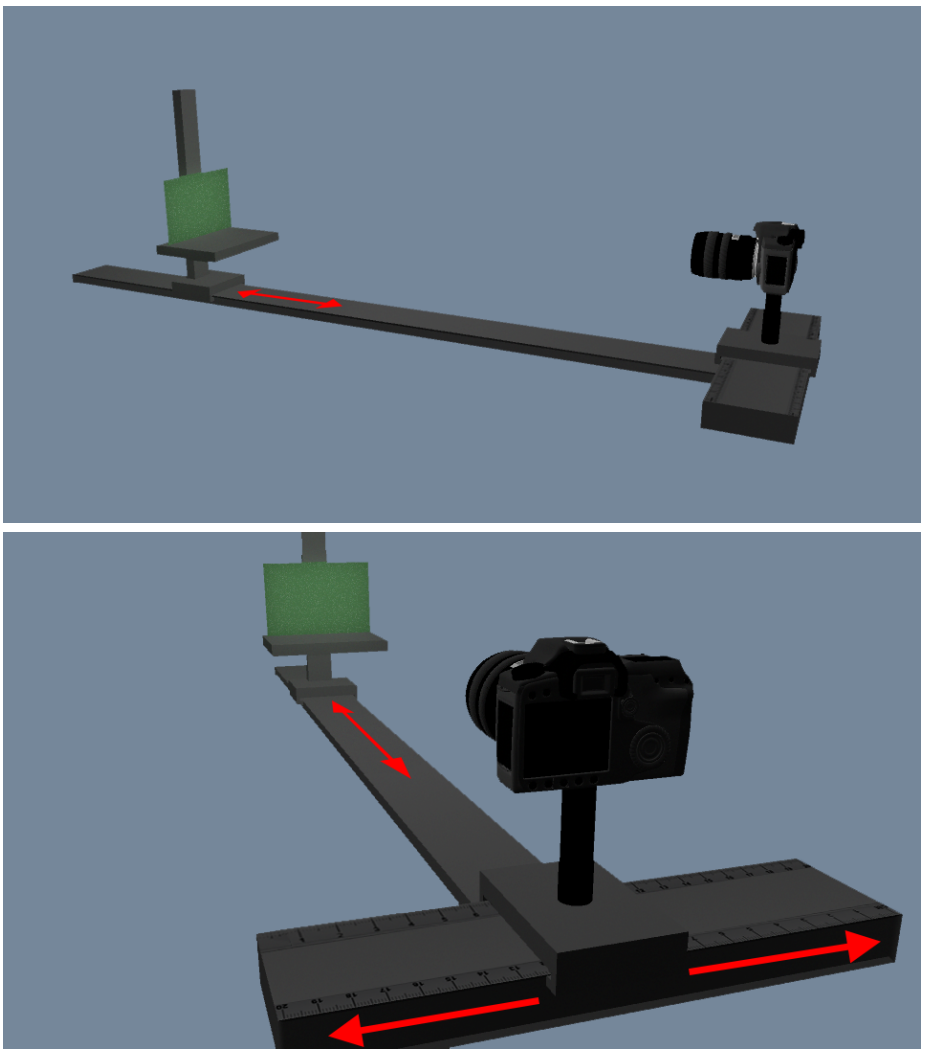


Figure 4.5: Custom support for stereoscopic sample acquisition. The red arrows indicate that the position can be regulated in that direction.

right view. The centered view was used to create monoscopic versions of the photographs, whereas the right and left views were combined to produce stereoscopic images of the same scene.

### 4.4.2 Adding binocular information to the photographs

Binocular disparity was recreated in the images by introducing a horizontal offset between the right and left photographs, using the generally accepted interpupillary distance (IPD) of 65 mm [103]. The photographs were taken using the *off-axis frustum* method, were the camera's field of view (or frustum) remains parallel to the projection plane for both views. This method of creating stereoscopic images is preferred over its alternative, the *toe-in* method —where the cameras rotate in each view, converging at the focal point— because it produces increased eyestrain [104, 105] (Figure 4.6).

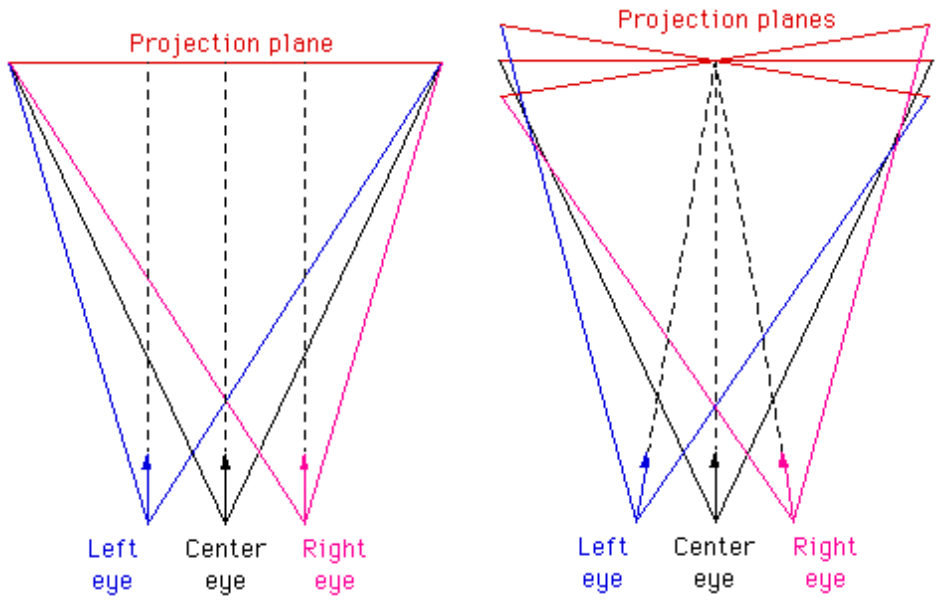



Figure 4.6: Methods for stereo-pair creation (top view): Off-axis frustum (left) and Toe-in (right) [105].

Since we are taking photographs of a real scene, the distance between the camera and the samples in each view results in what is known as a *positive parallax* —the object is behind the projection plane— when these images are displayed on the screen (Figure 4.7). To maximize the similarity with the real scene, positive parallax was eliminated by aligning both views with respect to a reference located at the same distance as the sample, obtaining *zero parallax* (Figure 4.8). This reduces the chances of visual fatigue in the observer and simplifies the stereoscopic image model, while maintaining local parallax (which is what interests us) within the sample's texture.

### 4.4.3 Image calibration: Adapting color information to the visualization environment

The HVS, as we have seen in Chapter 2, has the ability to maintain color and luminance constancy through varying observation conditions, and possesses a variable acuity that depends on factors like color, luminance, and contrast. Cameras, on the other hand, cannot do this on their own, so the raw image data sensed by the camera has to be transformed to a specific colorimetric system (a set of primary colors) other than the camera's, in order to obtain a meaningful image where these properties apply.

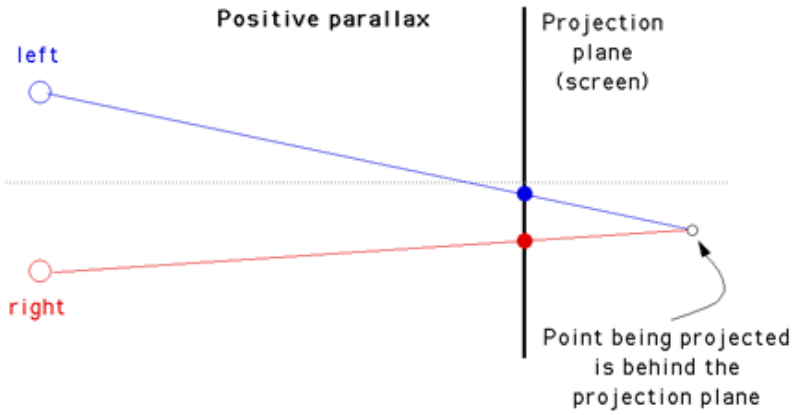



Figure 4.7: Positive parallax occurs when the projected object is behind the projection plane [105].

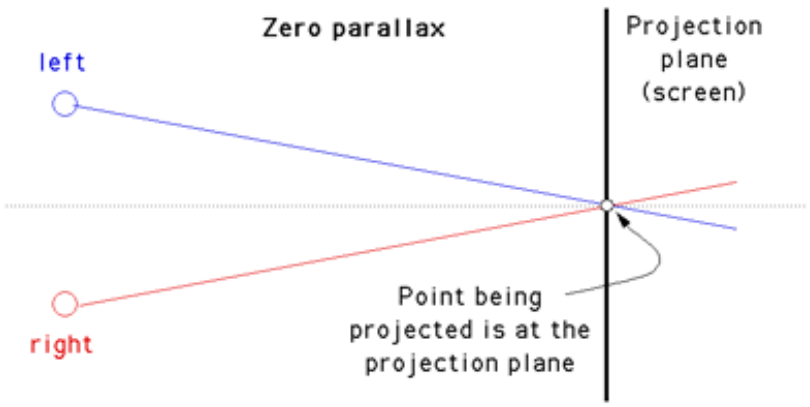



Figure 4.8: Zero parallax occurs when the projected object lies at the projection plane, and its projection onto the focal plane coincides for both eyes [105].

When we take a photograph of a paint sample, this photograph is stored in an uncompressed raw format where data is encoded according to the sensor's internal CFA. In order to obtain a color image, raw images must first be linearized and demosaiced, which results in a color image in the camera's local RGB color space [106]. Linearization is sometimes required to map raw values to the full 10-14 bit values through a lookup table (LUT) provided by the camera, and linearize and normalize the pixel values to the range [0,1] via an affine transformation. Demosaicing is the process of transforming 2D CFA pixel values into 3D RGB image values, interpolating the remaining two dimensions from neighboring pixels. Rob Sumner explains this process more in detail in his guide *Processing RAW Images in MATLAB* [106]. Once we have the RGB image in the camera's color space, we can apply the transformation obtained from the radiometric camera characterization process described in Section 4.3.2, in order to estimate the radiometric input signals sensed by the camera, represented in a physiologically-aware tristimulus space (e.g. XYZ or LMS).

In order to produce perceptually-realistic images, as defined in Section 1.2.1, we must ensure that the visual responses produced by the virtual images and the real scenes are the same. Given that the physiological camera input obtained above represents the real scene's visual response, this means that the physiological output produced by the display should be the same as the estimated camera input (i.e. the tristimulus values emitted by the display should match those "seen" by the camera in the real scene). Therefore, we need an additional step where the display's characterization model obtained in Section 3.2.2 is used to estimate the image RGB values that must be input to the display to produce the desired visual response. Figure 4.9 illustrates this process, where $OUT_{RGB}^{CAM}$ is the imaged RGB values in the camera's color space, $IN_{XYZ}^{CAM}$ is the camera's XYZ input values, which is the same as $OUT_{XYZ}^{DISP}$, the display's emitted XYZ output, and $IN_{RGB}^{DISP}$ is the display's input RGB image. Similarly, $M_{RGB2XYZ}^{CAM}$ is the camera's inverse transformation matrix computed in Section 4.3.2, and $M_{XYZ2RGB}^{DISP}$ is the display's direct transformation matrix computed in Section 3.2.2.

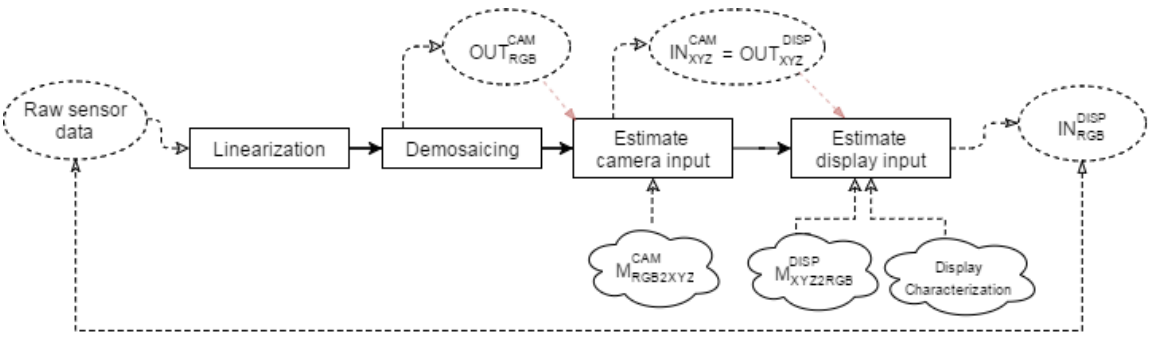



Figure 4.9: Raw image processing workflow from sensor data to display input.

### 4.4.4 Effects of image calibration in the perception of sparkle in photographs

The result from the image calibration process are photographs which all have a similar average luminance, but with different variance amongst the pixel values. Raw images with short exposure times present a wide contrast between shadows and highlights, whereas the contrast in long-exposure images tends to be much narrower. Sparkling is the result of specular reflections from metal flakes and, as such, they are imaged as pixels with very high-intensity values. As the amount of light reflected by the flakes increases, the resulting intensity values will overexpose. Eventually this intensity will exceed the saturation limits of the camera's sensor, which will clip values above this threshold to the upper limit of the color depth (number of bits) used to encode the image. For this

reason, it is important to control the luminance levels during acquisition, to reduce the amount of sensor overexposure and avoid losing color information, as explained in Section 4.3.2.

However, when we calibrate images, using the display's characterization model to adapt the intensity values in the image to its capabilities, we face an additional challenge: the magnitude of the radiance emitted by the display's screen is much lower than that emitted by our paint samples, due to the high reflectivity of the latter. Hence, some clipping may still occur after image calibration due to the dynamic range (range of intensity values) differences between the real scene and the displayed image. As we may remember from Section 1.2.1, this was indeed one of the big issues associated with perceptual realism. This problem occurs more frequently as the exposure time increases during photograph acquisition. Indeed, as the intensity of the high-intensity pixels corresponding to sparkles is typically near the saturation limits of the sensor, their intensity values will not increase much with exposure time; conversely, the intensity of lower-intensity pixels will do increase, reducing the contrast between them. This phenomenon is illustrated in Figures 4.10 and 4.11.

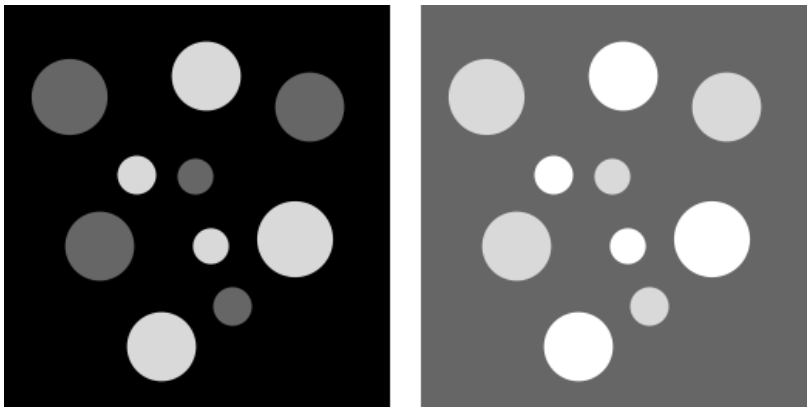


Figure 4.10: Contrast reduction in calibrated images as exposure time increases (exposure increases from left to right), due to the lower dynamic range of the display. As the exposure increases, the intensity of the brighter circles remains relatively similar due to clipping, whereas the intensity of the less bright pixels increases much more, reducing the contrast between them.

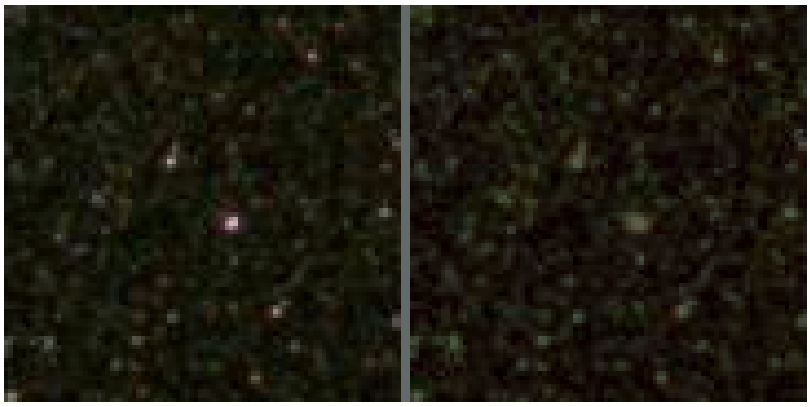


Figure 4.11: Two photographs of a green paint sample taken with a shutter speed of 1/8 s (left) and 1 s (right). We can see that the sparkles on the left image (shorter exposure time), are narrower than on the right image, but there is more contrast with the background.

## 4.5 Perceptual reproduction of luminance dynamics in photographs of sparkling materials.

The characteristic sparkling texture of some paint materials is produced, as we have seen, by the metallic particles, or flakes, contained in the materials. These particles, with a high specularity, reflect the light producing the perceptual phenomenon that is called sparkling. The luminance (amount of light, see Section A.18) emitted by the material produces a response in both the observer's visual system and the camera's sensor, which can be represented in some perceptual color space like XYZ or LMS. In this sense, producing a perceptually-realistic simulation of the materials consists, as described in Section 1.2.1, in ensuring a match of the chromaticity and luminance between the visual responses —expressed in terms of these two dimensions— produced by the real and virtual scenes.

We can distinguish two main steps in the conversion from real scene to virtual image: scene acquisition and image representation on a screen. Firstly, the physical reality is captured by a sensor, which results in an initial integration of the information by the sensor's CFA. The amount of information loss carried by this integration will depend on the quality of the camera characterization process, which gives us the transformation between image data in the camera-dependent RGB color space and the sensor response in a device-independent color space. Secondly, the sensed image must be adapted to the capabilities of the display, using a second transformation, given by the display's transfer function and characterization model. This transformation, which estimates the device-dependent RGB data required as input for the display, will also result in some loss of information. Additionally, a tone-reproduction transformation may be applied to improve image aspect.

Therefore, as discussed in Section 4.4.4, it is important to control the luminance range of the images during acquisition and display, to try to keep as much of the original perceptual information as possible. Here we present two experiments were we analyze the effects of the acquisition parameters (Section 4.5.1) and tone-reproduction methods (Section 4.5.2) on the perception of the aspect of the materials.

### 4.5.1 Perceptual effects of image exposure in photographs of sparkling materials

Typically, in our environment, we can find a vast array of different light radiation that affect our perception, not only in the visible spectrum, but also in the high ultraviolet and low infrared spectra that may also be seen by some people. Therefore, the amount of different luminance levels present in any real scene is almost infinite. However, when we try to capture a photograph of a real scene using some electronic equipment, due to its limitations, the acquisition process introduces many simplifications that may alter our perception of them. The purpose of this experiment is to study the sensitivity of human observers to this type of simplifications. We saw in Chapter 2 that the photoreceptors in the human retina work within a known limited range of illumination levels, and their response depends on the wavelength of the radiation. The same thing is true for a photographic camera, but because the camera response varies from one to another (see Section 4.3) the correlation between the sensitivity and the representation of luminance in both systems is unknown.

As we saw in Section 4.3.2, the total luminance of an image depends on the scene illuminance (the amount of light arriving on the scene), the camera's sensitivity, and the exposure time used for acquisition (that is, the time during which the camera's sensor is exposed to the light). We discussed that by fixing the illumination and the sensitivity, the luminance depends exclusively on the exposure time and aperture. Furthermore, by fixing also the aperture, we can study the relationship between perception and exposure time as a function of only one parameter, the shutter speed. Depending on the chosen value, the displayed photograph may vary from an image with a clearly perceptible depth and large overexposed areas, to a flatter image with smaller overexposed

areas. Hence, we can see that choosing the right exposure is essential to obtain the correct amount of perceived sparkle and texture depth throughout the image.

#### 4.5.1.1 Methodology

For this experiment we created a series of stereoscopic photographs of the same paint plate, captured at different exposures as described in previous sections of this chapter, and presented these photographs to a group of observers. The participants were asked to select the one image they perceived as being the most similar to a physical reference plate —with special attention to flake density, depth, sparkle-background contrast, and sparkle intensity.

To compensate for possible luminance disparities between the real conditions and the display, and to facilitate the comparisons, the luminance of the virtual images was adapted after calibration to a chosen target value —preserving the colorimetric ratios— to produce an overall luminance similar to that of the reference real samples. This luminance target value was chosen so that a Macbeth color chart displayed in the computer monitor would produce a similar luminance response as the real physical chart under the same illumination conditions. Since the same value was used for all the images, they all share a similar global luminance after the adaptation, but preserve the local texture contrast due to their different exposure times.

The images were presented to the observers inside a custom made visualization cabin (see Image 4.12). The cabin walls were all black and the back wall (the one facing the observer) contained two apertures the size of the paint plates. Observers could see the physical sample plate (reference) through the aperture on the left and a portion of a computer monitor (where the photographs, or test images, were displayed) through the aperture on the right. This was done to minimize the effects of surrounding stimuli. The illumination and observation conditions were as described in Section 4.4.1, so the observers were seated at roughly 70 cm from the cabin's back wall with their eyesight at the same height as the apertures. Both the test images and reference samples were viewed wearing a pair of active stereoscopic glasses (shutter glasses) to equalize the observation conditions in all the cases.

Figure 4.12: Visualization setup for the comparison of virtual images with real references.

We presented five different paint samples to a group of 15 observers, with the following respective paint codes: GM1, GM2, HI, BC and BR (Figure 4.13). Three different series were created for each sample, each one consisting of the same photograph taken with six increasing shutter speeds of 1.3, 1.6, 2, 2.5, 3, and 4 seconds. This resulted in a total of 90 sample plate photographs, that is, three series of six images per sample plate. The first series consisted of six stereoscopic color photographs. The second series consisted also of six stereoscopic color photographs, but taken with a longer exposure time, resulting in a higher luminanc (33% brighter); the purpose of this series is to study how changing the global image luminance affects their perceived aspect. The third series consisted of the same photographs in the first series, but converted into grayscale; in this case, the use of grayscale images served to study the role of color in the perception of the images.

Figure 4.13: Paint samples. From left to right: GM1, GM2, HI, BC and BR. The first three samples possess an overall green tint —HI being slightly brownish—, whereas BC and BR have a red and blue tint respectively. All samples are shiny except for GM2, which is matte.

For each series of photographs, observers were asked to select the image that they perceived as being the most similar to the reference plate, using criteria such as flake density, texture depth, sparkle-background contrast, and sparkle intensity. They were also asked to rate, in a scale from 1 to 10, the similarity of the chosen image to the reference; this parameter was later used as a measure of the reliability of each choice, as well as an indicator on the average similarity between the images in a given series and the reference plate. Finally, observers were asked to indicate the level of difficulty (from 1 to 10) to discern between the images within each series which, in addition, gives us a measure of the reliability of each choice. To interpret the test results we gave a score to each chosen photograph —that is, each photograph that has been chosen by an observer as being the closest to the reference— which we obtained by weighting each observation by its similarity to the reference; this way we penalize low-quality responses and reward high-quality ones.

#### 4.5.1.2 Results

The average scores for each series of images is shown in Figure 4.14, where we can see that the most chosen photograph for all the series is the one with the highest exposure. In the case of the first series (base images), the standard deviation is not very high, which we can see in the bar graph, where there is not much variation in the score for the rest of the photographs in the series. One possible explanation could be that the details in the images or the plates (or both) were not clear, and that made observers choose randomly. This would be consistent with that fact that many observers pointed out that the use of stereoscopic glasses reduced notably the amount of detail both in the images and the sample plates.

The participants tended to say that the images with a higher luminance, as well as the grayscale images, had very little texture contrast and sparkles, which resulted in duller, flatter images that resemble the reference less than the corresponding base images. This could be due to the contrast reduction effect discussed in Section 4.4.4. We believe that in these cases they chose higher-exposure images to compensate the increased luminance for the lack of detail and texture contrast. This would justify the fact that standard deviations in the image scores are generally larger in those two series

than for the base photographs, and that the chosen photographs —i.e. the one with the highest exposure— were given a much higher resemblance score that the rest of photographs, possibly because the observers could see the details much better.

The time increments between the chosen shutter speeds —namely, 0.3, 0.4, 0.5, 0.5, and 1 second, respectively— are not constant (the increments are determined by the camera), which means that the differences between any two consecutive images in the series do not have to be similar. In effect, observers noted that there were some images (namely, the first three) where differences were almost imperceptible, whereas differences started to be clearer starting from the fourth image. This could suggest that half a second of exposure is around the just noticeable difference (JND) for observers to start perceiving a difference between two of these images.

As we can see from the bar graph in Figure 4.15, the images in the first series (base images) are most often perceived to have a closer resemblance to the reference. We commented earlier that observers thought that the images in the higher-luminance and grayscale series were flat and little detailed, which in general made them look much more different from the references than the images in the base series; however, observers also noticed that this flatness made it easier to discern between the images in those series, as we can see in Figure 4.16 where the results for the first series are generally lower or similar than for the other two. The paint tint does not seem to affect the similarity between test images and reference plates, although it does seem to influence the difficulty for intra-series image discernment. Indeed we can see that, in the average, observers seem to find it easier to differentiate between series of green-tinted textures than red or blue, which would be consistent with the fact the HVS is more sensitive to green stimuli under photopic conditions (bright light levels).

We have discussed several times that the reason why we choose to use stereoscopic visualization in our tests is because, based on previous results by Da Graça et al [20, 4], we believe that the additional binocular cues provided by this technique of depth visualization do improve the perception of images of this kind of materials. In order to test this affirmation, and justify our choice, we started the experiment by showing each participant two versions of the first 6-image series —i.e. base-luminance photographs of sample plate GM1 with shutter speeds of 1.6 through 4 seconds— using monoscopic and stereoscopic images in a random order. We then asked them to say which images where closer to the reference. Surprisingly enough, many participants, despite being capable of seeing differences between a stereoscopic and monoscopic image if shown explicitly, could not distinguish between each type during the experiment. Nonetheless, they all chose the stereoscopic images as closer to the reference in each case.

In general, these results are an initial approach to understanding one of the many problems associated with the acquisition of sparkling materials, which is the choice of an appropriate exposure value. Although the results hereby presented are not very conclusive, we can draw very important ideas from them. We have observed that the use of stereoscopy poses additional visualization problems in case of low-luminance images, since many properties of the materials, such as the amount of perceived sparkling and the contrast, seem to be reduced. However, despite this observation, all the participants seem to agree that stereoscopic images are perceived as more similar to the reference than monoscopic ones. We have also learned that high-exposure values are typically preferred in the case of images with a low global luminance. We justify this observation by supposing that, although observers affirm that these images have less detail and contrast, the additional luminance increases the visibility of identifiable features to use as reference in the comparison with the reference. Finally, given the low similarity scores of the images in the grayscale series, we must conclude that color seems to play an important role in how in the perception of sparkling images, and textures in general. Nonetheless, when comparing grayscale images between themselves (see the intra-series discernment scores in Figure 4.16), they seem to be easier to distinguish from each other. This is probably explained by the fact that, although grayscale images may overall seem very different from the color reference for an observer, the human eye is in fact much more sensitive to details in the absence of color, as explained in Chapter 2.

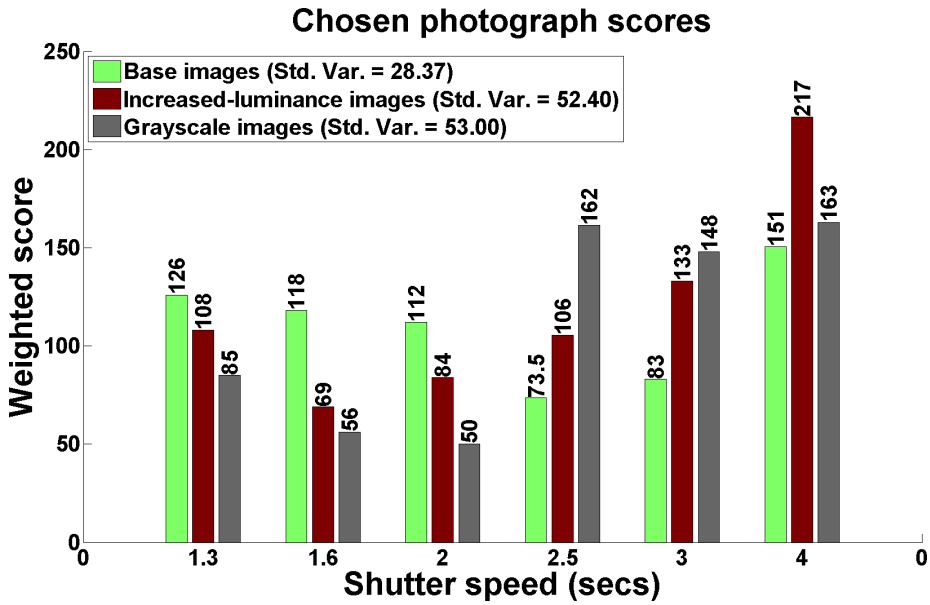



Figure 4.14: Average image similarity score with respect to the real reference for the base photograph series (left bar), the increased global image luminance series (middle bar), and the grayscale series (right bar), as a function of the shutter speed.

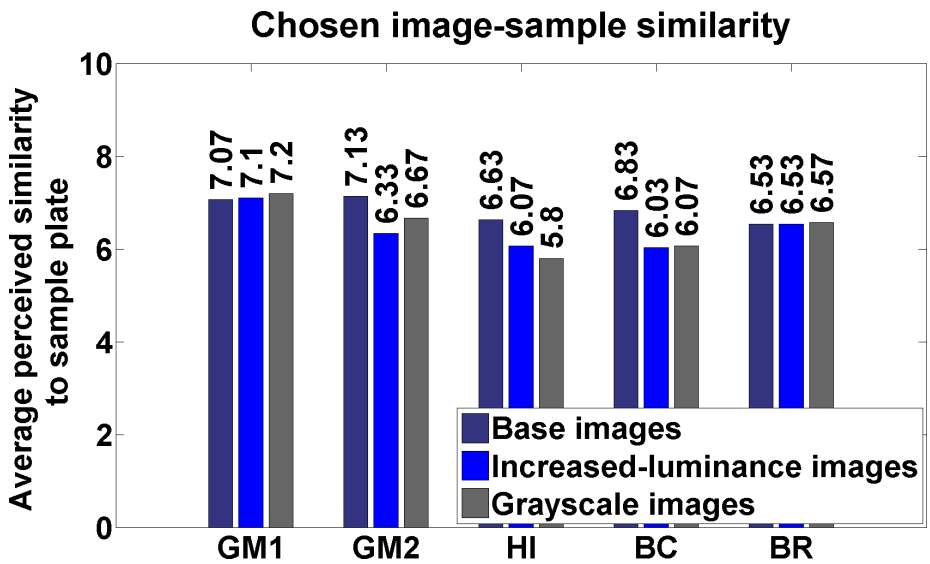



Figure 4.15: Similarity between chosen images and the corresponding reference sample (a high number indicates a high similarity).

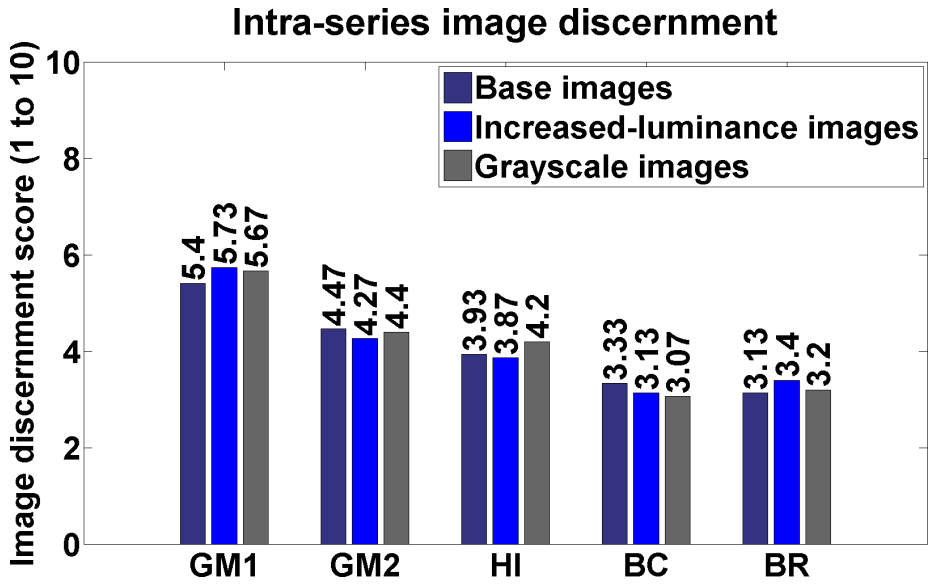



Figure 4.16: Difficulty to discern between the images in a series (a low number indicates a high difficulty to discern between the images). Each 3-bar group represents a paint sample, from left to right: GM1, GM2, HI, BC, BR; the leftmost bar (purple) represents the base images, the middle bar (blue) represents the image series with increased luminance, and the rightmost bar (gray) represents the grayscale images.

### 4.5.2 Perceptual effects of dynamic range compression in photographs of sparkling materials

The results from the previous section provide us with important knowledge about the perceptual effects of scene luminance on the image during the acquisition process, but also about the observation methodology itself. Indeed, we observed that using stereoscopic visualization in the tests results in much lower scene luminance levels due to the light filtering properties of the stereoscopic glasses, which complicates the observation and visual analysis of texture details. Additionally, luminance disparities were observed between the real and virtual scenes, due to the additional luminance introduced by the display in the virtual images. In this section we describe another observation experiment, where we propose a new comparison methodology that tries to resolve the issues in the previous one.

This experiment analyzes the perceptual effects of different methods to adapt the luminance of sample paint photographs to the low-dynamic range capabilities of our display. This problem, as we have described earlier, is known as *tone reproduction*. The previous section showed us the importance of acquiring scene luminance information correctly, maintaining the visual response of the real scene. The results from this experiment help us understand the effect of different tone-reproduction methods to represent the same luminance information on a display. Given the much lower dynamic range of conventional displays like ours, this process is crucial to ensure perceptually-realistic simulations, as discussed in Section 1.2.1.

#### 4.5.2.1 Problem description

Previous experience has shown us that the comparison of real and virtual samples implies three important constraints: firstly, the photographs must be taken under the same illumination conditions as those used in the visualization experiment; secondly, given the effect of contrast and luminance on perception in general [15] and material texture perception in particular (see Section 4.4.4), we

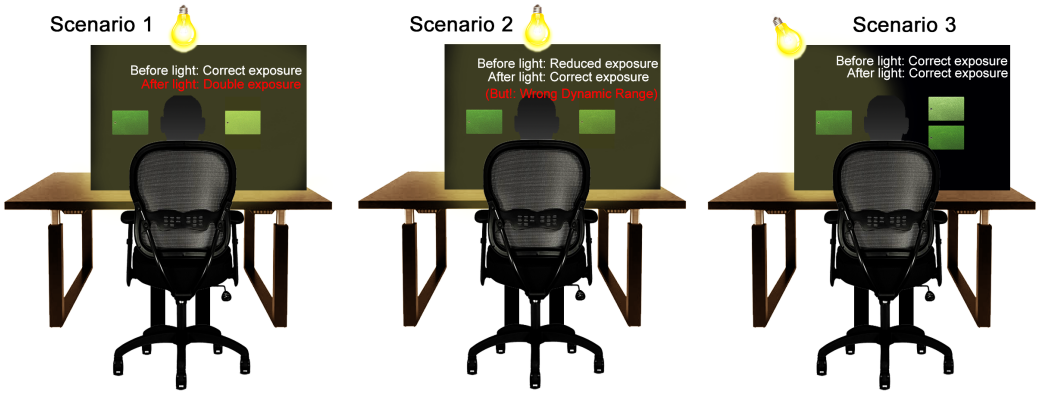



Figure 4.17: [a,b,c] Diagram of the setup in the three scenarios suggested. The test photographs are to the right and the real sample reference to the left. In the first two scenarios the illumination is common to both sides of the comparison, whereas in the third case it only affects the physical sample. Notice the important luminance difference between both sides of the comparison in the first scenario, due to the additional light emitted by the display. The use of paired comparisons in the third scenario simplifies the assessment of the images.

must ensure similar contrast levels in both the images and the physical samples; finally, if we want to include both stereoscopic and monoscopic images in the comparisons, observers must keep the stereoscopic glasses on at all times, regardless of the type of image they are looking at, to ensure similar colorimetric conditions.

The most straightforward scenario to compare a physical and virtual sample under the same observation conditions is to visualize the photographs and the physical reference together, under the same illuminant, as we did in our previous experiment (Figure 4.17.a). The idea behind this test scenario is to equalize as much as possible the observation conditions so that, if we hide everything outside the samples, we can somehow make the observers forget that they are looking at two samples on different supports (physical and digital). However this solution has an important downside to it: the display is an emissive surface and, as such, it introduces additional light into the virtual sample —the illumination already present in the photograph, plus the one in the visualization environment, plus that of the black point of the display. Although we can correct for some of the additional brightness —as we did in Section 4.5.1— we have seen that it is impossible to obtain similar luminance levels on both samples due to a common factor, the ambient illumination in the visualization environment, which always remains constant on both sides of the comparison.

A possible workaround to this scenario is to take the sample photographs with a lower exposure to compensate for the additional luminance that will later be introduced by the display (Figure 4.17.b); this solution would indeed produce images with a similar luminance as the physical samples when viewed under the same illumination. However, as discussed throughout this chapter, image exposure plays an important role in our perception when observing photographs of materials. Hence, by taking the photographs with a lower exposure we might inadvertently be favoring certain aspects of the material and, therefore, biasing our perception.

The solution that we propose is to visualize the real and virtual samples side by side, but using a different intensity on each side (Figure 4.17.c). This scenario is more versatile because it allows for independent tuning of the illumination. The experiment presented in this section shows that this methodology produces better results than those obtained with the previous comparison scenarios.

We are interested in obtaining perceptually-realistic results, and paired comparisons are a good way to assess the aspect of these sparkling metallic paint textures. Given the complexity of the sparkling effect, it is difficult for an observer to visually assess these images on an absolute scale, so introducing a reference that acts as ground truth simplifies the evaluation task. At the same time,

| Image | Exposure time (sec) | Luminance scaling factor | Maximum luminance | Saturated pixels (%) |
|---|---|---|---|---|
| 1.1 | Short (1/15) | 1 | 205.3 | R=0.158<br>G=0.034<br>B=0.032 |
| 1.2 | Short (1/15) | 0.51 | 104.7 | R=0.039<br>G=0.006<br>B=0.007 |
| 2.1 (Reference) | Medium (1/8) | 1 | 104.7 | R=0.145<br>G=0.028<br>B=0.030 |
| 3.1 | Long (1) | 1 | 15.65 | R=0.009<br>G=0.000<br>B=0.000 |
| 3.2 | Long (1) | 1.78 | 27.85 | R=0.300<br>G=0.000<br>B=0.000 |

Table 4.3: List of images used in the experiment. There is a base version of each image and a tone mapped version of Images 1 and 3. There is no tone mapped version of Image 2 because it is the reference image for luminance levels. Image 3.2 has a maximum luminance lower than the reference (104.7) because scaling it by a higher factor would result in a high amount of oversaturated pixels.

comparing the images side-by-side reduces the time to move the eyes from the test image to the reference and, therefore, the effects of low visual memory persistence times in the comparisons.

#### 4.5.2.2 Methodology

From the original multi-exposure series of photographs of the sample paint plates, we chose three different values: short (Image 1.1, 1/15 seconds), medium (Image 2.1, 1/8 seconds), and long (Image 3.1, 1 second). This choice tries to maximize contrast and minimize the amount of underexposure and overexposure present in the calibrated images. As we saw in Section 4.4, controlling the amount of saturation is crucial to maintain as much information as possible from the original scene and, in turn, ensure perceptually-realistic simulations.

This time we only studied one type of paint sample, commercially named *Manitoba Gray*. Using the medium exposure images as reference, we created tone-mapped versions of the short and long exposure images, resulting in Images 1.2 and 3.2, respectively. To maintain the perceptual properties of the images, we performed the transformations in the sensor-response domain —represented in the XYZ color space, where the Y channel carries the luminance information. The tone-mapping transformation is shown in Expressions 4.10 and 4.11, where $V_{in}$ and $V_{out}$ are the pixel luminance of the original and tone mapped version of the image, respectively, $L_{max}$ is the maximum luminance in the original image, $L_{ref}$ is the maximum luminance in the reference image, and $k_m$ is the tone-mapping operator.

$$V_{out} = V_{in} k_m \tag{4.10}$$

$$k_m = \frac{L_{ref}}{L_{max}} \tag{4.11}$$

Factor $k_m$ was chosen so that the maximum luminance (maximum Y) is as close as possible to that of Image 2.1 (the reference image). If using the definition in Expression 4.11 results in

excessive oversaturation[6] in the RGB image, then the maximum possible value before reaching that much oversaturation is used. In our case, the maximum luminance in the reference image is 104.7, and the maximum luminance in Images 2.1 and 3.1 is, respectively, 205.3 and 15.65 (see Table 4.3). In the case of Image 1.1, the tone-mapping operation consists in multiplying each pixel by $k_m = 104.7/205.3 = 0.51$. However, Image 3.1 was taken with a much longer exposure time, so scaling the pixels by $k_m = 104.7/15.65 = 6.7$ results in excessive oversaturation; therefore, we can only scale by a factor of 1.78.

Scaling the XYZ tristimulus values in this manner has several effects on the images. If $k_m$ is greater than one, the dynamic range is expanded, which results in increased overall image brightness; on the other hand, if $k_m$ is less than one, the dynamic range is compressed and the overall brightness reduced. Given the low concentration of sparkle pixels with respect to background pixels, we observe a decreasing contrast between sparkle pixels and background as the overall image brightness increases, which establishes an inversely proportional relationship between sparkle contrast and overall image brightness, as a function of the dynamic range.

For each of the five images listed in Table 4.3 we created two versions, monoscopic and stereoscopic, to also test whether or not the use of stereoscopy has any effect on the preference for a given sample. Altogether, this resulted in a set of 10 different images (see Table 4.4). We created series of 45 pairs of images from the list —i.e. all possible combinations of the 10 images, excluding comparisons with themselves— combined randomly, and asked a group of 30 observers to compare the two images in each pair (test images) and say which one they perceived as "closer" to the reference paint plate —i.e. the physical plate that we took the original photographs from. The test images were arranged vertically on the right-hand side, whereas the reference was positioned to their left, 42 centimeters away (Figure 4.17.c).

Both the test images and the reference were observed through an aperture on a black surface to isolate them from their surroundings. The reference was illuminated independently from the test images during the experiment, with the same illumination used to take the photographs. This was done to avoid the luminance differences observed in the previous experiment. The observers were seated in a position between the test images and the reference, wearing a pair of stereoscopic shutter glasses at a distance of approximately 70 centimeters.

#### 4.5.2.3 Results

From the responses obtained in the experiment —answering the question "which image from the test pair is closer to the reference?"— we computed the scores for each sample, according to the Bradley-Terry model [107]. These scores indicate the probability of each sample being chosen by an observer as closer to the reference over any other. Representing these scores in a logarithm scale —which represents better the relative nature of visual perception— we can group our ten samples into different similarity classes. Figure 4.18 shows five classes designated by colors green (1), blue (2), yellow (3), brown (4), and red (5).

In this experiment we are analyzing the effect of three different factors on the perception of images from our paint samples: the use of stereoscopy, exposure time, and dynamic range and contrast. From the information in Table 4.4 we can see a tendency from observers to choose samples giving these factors a clear order of preference. Indeed we can see that stereoscopic samples are clearly preferred over monoscopic ones; when both samples are of the same type, contrast comes into play. Base images —i.e. the original non-tone-mapped images— are preferred over the adjusted ones, with observers preferring medium-exposure images over short-exposure (excessive contrast) and long-exposure ones (low contrast).

[6] We consider that an RGB image is excessively oversaturated when more than a 0.3% of the pixels have the maximum gray level value. This saturation percentage was chosen experimentally to obtain a reasonably low amount of perceivable oversaturated pixels.

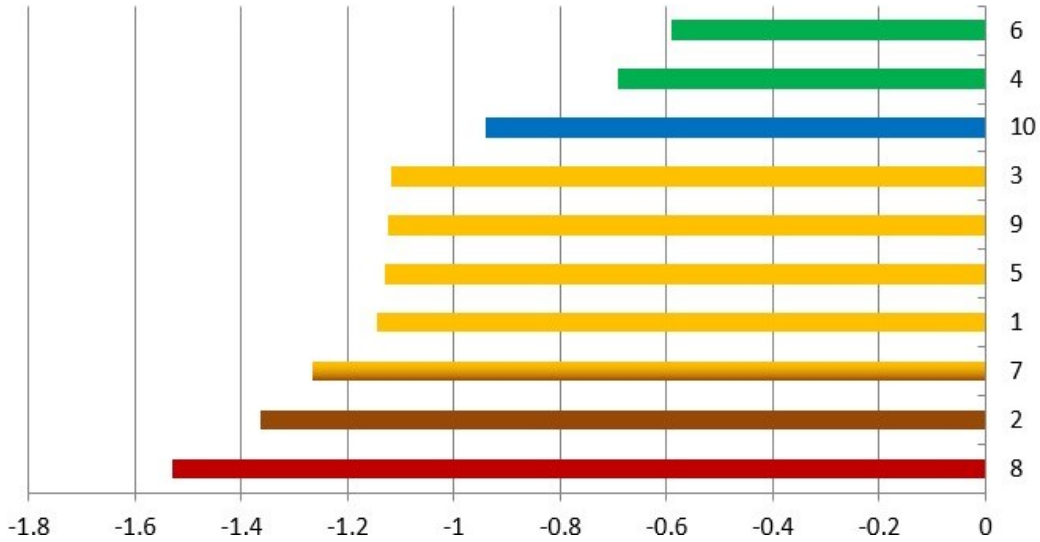



Figure 4.18: Experiment results. The ordinates indicate the sample number (see Table 4.4) and the abscissas indicate the Bradley-Terry scores in logarithmic scale.

| Sample | Image | Type | Score | LOG score | Perceptual class |
|---|---|---|---|---|---|
| 8 | 3.2 | M | 0.0296 | -1.5287 | 5 |
| 2 | 1.2 | M | 0.0432 | -1.3645 | 4 |
| 7 | 3.1 | M | 0.0543 | -1.2652 | 3+4 |
| 1 | 1.1 | M | 0.0718 | -1.1439 | 3 |
| 5 | 2.1 | M | 0.0740 | -1.1307 | 3 |
| 9 | 3.2 | S | 0.0751 | -1.1244 | 3 |
| 3 | 1.2 | S | 0.0762 | -1.1180 | 3 |
| 10 | 3.1 | S | 0.1149 | -0.9397 | 2 |
| 4 | 1.1 | S | 0.2036 | -0.6912 | 1 |
| 6 | 2.1 | S | 0.2574 | -0.5894 | 1 |

Table 4.4: Bradley-Terry scores and Log scores for each of the samples used in the experiment in ascending order. These scores represent the probability of each sample being judged closer to the reference over any other. Types M and S stand for Monoscopic and Stereoscopic respectively.

Based on these results, we can extract the following observations, regarding observer preference to match a sample image to its reference:

1. Stereoscopic images are always found more similar to their reference than any monoscopic version of the same image, regardless of the other parameters.

2. Medium-exposure images are preferred over short and long exposure ones. Between long and short exposure images, short exposures are preferred.

3. Tone-mapped images are never preferred over the non-tone-mapped ones, which suggests that, in the absence of overexposure, no information is actually lost during display when we acquire photographs with medium exposure values.

We mentioned earlier that contrast is inversely proportional to exposure time and dynamic range. In general, we can say that contrast increases as global image brightness decreases, and the other way around. Our results show that this affirmation is true, but only for intermediate values of image

brightness; below and above a certain threshold, perceptual similarity decreases. Although this contrasts with our results from the previous experiment, where we found that observers preferred bright images, the fact that high contrast images are preferred supports our assumption that it was due to the lack of contrast in darker images.

Based on previous observation results (Section 4.5.1), we can affirm that image brightness and contrast —image attributes determined by the luminance of the scene and the dynamic range of the display— are also strongly linked to the amount of light exposure during acquisition. The results shown in this section support our previous claims that image brightness and contrast can alter the perception of sparkling material images due to its effect on perceived flake size. Moreover, they suggest that, under the described observation and illumination conditions, close perceptual similarity with respect to real references can be achieved by ensuring device-independent XYZ visual responses with luminance values (Y channel) around 100 $cd/m^2$.

Chapter 5

# Objective validations of the rendering model against photographs of real scenes

## Contents


5.1 Introduction . . . . . 88
5.2 Setup and characterization of a controlled test scene . . . . . 89
5.3 Data acquisition from test scene photographs . . . . . 91
5.4 Simulation of the scene in the target rendering engine . . . . . 91
5.5 Objective comparison of photographs with simulations . . . . . 92
5.6 Conclusion . . . . . 99


System validation consists of gathering and analyzing data from the different stages of the system's workflow, to ensure that it produces quality results according to a predefined standard. In the case of a simulation system, this standard is typically the level of realism of the simulated scenes with respect to a chosen reference [108, 109]. Physically-based rendering is a complex process where many different physical phenomena must be reproduced in order to obtain accurate, or *realistic*, results. We have seen that before we can judge the level of realism of an image, we must define realism within the context of our application. Considering that our simulations will be viewed by human observers, we can use our knowledge of the human visual system (HVS) to improve the perceptual similarity between the simulations and the real objects under the observation conditions. At the same time, we can use this information to avoid computing redundant image information that will not be seen by the viewer, while ensuring that the visual response produced by the simulated scene will be the same as in reality. We have defined this type of realism as perceptual realism. In this context, the visuo-perceptual validation of our system is the process that verifies that the simulations generated by the system can induce in the observers the same visual response as if they viewed the original scene in reality.

The validation of a rendering engine requires two types of validations: objective and subjective. In order to perform a robust subjective (visuo-perceptual) validation, we must first make sure that the rendering engine produces valid results with respect to the real models [110]. This chapter presents a methodology that allows us to validate a simulated scene using pairs of photographs and rendered images of the real scene. The goal is to take photographs of the real conditions under different settings, and at the same time reproduce the real conditions in the rendering engine, so that the photographs and their corresponding rendered images are as close as possible to each other. We will say that the rendering engine is objectively validated when the images rendered from a scene are identical to the photographs from the same scene, or in the event that differences exist, we are capable of identifying the causes and justifying them.

We propose a complete workflow, including: scene setup, scene characterization, data acquisition, scene simulation, and data analysis. At each step we describe the most relevant aspects to consider

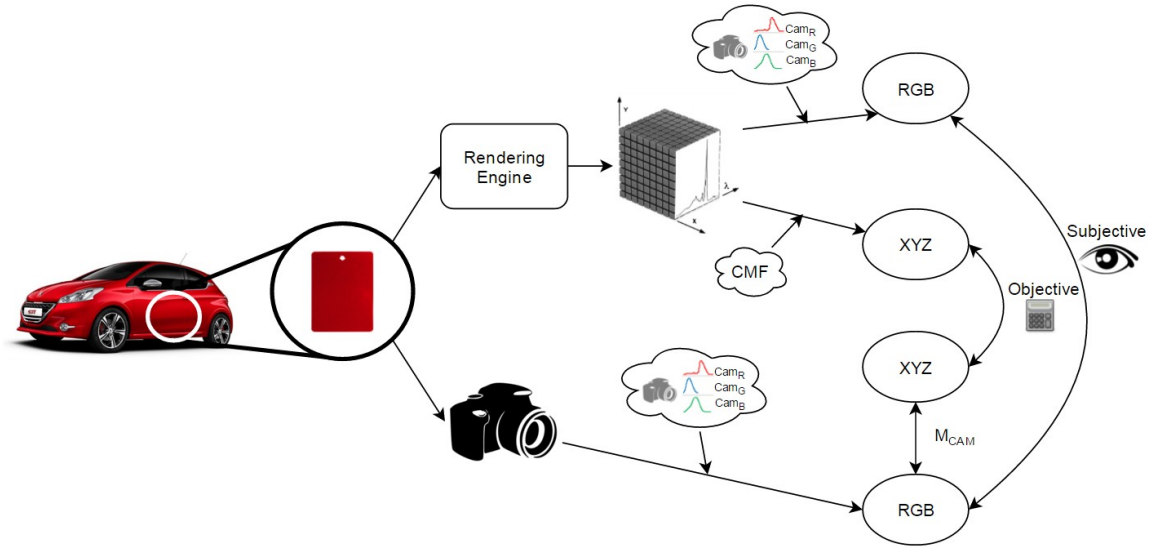



Figure 5.1: Visuo-perceptual validation methodology comparing photographs with simulations.

in order to maintain the visuo-perceptual realism of the simulation. Using objective data, we provide results that show that we can achieve high levels of visuo-perceptual accuracy following this methodology. We will also present the bases to perform subjective validations, leaving the actual experimentation results as a future work.

## 5.1 Introduction

The visuo-perceptual validation workflow that we propose for our simulation system consists of two parts: objective (colorimetric) and subjective (visual examination) validations. Objective validations use numerical comparisons to validate the simulations produced by the system against ground-truth data; subjective validations use human observers to validate the perceptual realism of the results. Since we cannot perform a direct comparison of a simulation with a real scene, we use photographs as an intermediate step. We can take advantage of the results from Chapter 4 to use these photographs as ground truth without any loss of perceptual accuracy.

We want to ensure that the visual response produced by the simulations is the same as in the real conditions. We know that the visual response depends ultimately on the response of a given sensor —i.e. human retina, imaging sensor, etc— for each type of light (tristimulus sensory quantities), and that we can represent and compare device-independent sensory quantities using some physiological color space like XYZ. Furthermore, we also know that we can characterize an imaging device to obtain the transformation function that estimates device-independent sensory quantities from device-dependent RGB color data. Consequently, device-independent sensory quantities can be used as objective radiometric ground-truth data in the validation process. In this context, the objective validation process ensures that the virtual images contain the same perceptual information as the real scene; the subjective validation process ensures that the visual response produced by said images on human observers is equally maintained. Figure 5.1, which we already presented in Chapter 1, illustrates the two parts of our methodology for visuo-perceptual validations.

An important phenomenon to keep in mind in any validation process is that of metamerism. Two color stimuli with different spectral power distributions (SPD) are said to be metamers under a given illuminant, and for a given set of spectral sensitivities, when they are represented by the same tristimulus values under those conditions. Given that metamers produce the same set of tristimulus sensory quantities under the metameric conditions, they are potentially a source of

discrepancies between the real and virtual scene. Although visual responses should not be affected by this phenomenon, because they are triggered by integrated tristimulus quantities, this is something that should be verified during the validation process.

The rendering engine produces 3-dimensional hyperspectral images where each pixel contains an array of spectral radiances for the specified wavelength range. By integrating these values with the averaged sensitivity curves of the standard colorimetric observer (the color matching functions, or CMF), the XYZ tristimulus response can be computed. Additionally, the subjective tests require that a displayable device-dependent color image also be computed. We could use the inverse camera characterization matrix to convert from the XYZ response to RGB pixel colors; however, we would be neglecting spectrally-related phenomena, most notably metamerisms. We can take into account the effect of potential metamerisms by reproducing the camera's color filtering process. Thus, instead of using the camera's inverse characterization matrix to estimate RGB values directly from sensory tristimuli, the real color response can be obtained from the spectral data by using the sensor's spectral sensitivities (Figure 5.1).

## 5.2 Setup and characterization of a controlled test scene

The response produced by our visual system when we observe an object in a given scene depends on the characteristics of the object itself, but also on its environment (surrounding objects, lighting, etc) and our position with respect to the scene (observation conditions). As we explained in previous chapters, what we see when we look at an object is the response of our visual system to the light arriving to our eyes from the object. This light is in turn the result of complex physical interactions between the light and all the elements in the scene that contains it. For this reason, objects cannot be simulated in isolation, without taking into account the contribution from additional elements in its environment; doing so would result in an incorrect simulation, since the real conditions will not be reproduced accurately. Therefore, when we speak of a simulated object, we are in fact referring to an object contained within a larger scene.

We consider a scene as a collection of objects, with some light sources, and an environment containing them. The light emitted by the direct light sources (primary light sources) will be reflected by the objects in the scene and the environment, resulting in additional lighting (secondary light sources). Any non-black surface in the scene is subject to act as a secondary light source, so we must ensure that all potential light sources, primary and secondary, are known and characterized (e.g. their photometric distribution, such as the IES[1] measurements, their spectral power distribution, etc.). Consequently, in order to simplify the simulation, it is advisable to reduce the complexity of the scene to a minimum.

To produce a minimal scene, we built our own light cabin, where a black cardboard box acts as environment (Figure 5.3). The box dimensions were large enough to allow for the desired distance between the objects and the observer —in this case, the camera— and the color was chosen black to reduce the amount of light reflected from the walls. All sides were closed except for one aperture, where the camera and lighting were placed. The sample was located in the middle of this box, to a distance of 80 cms from the camera —measuring at the bottom of the sample— inclined backwards an angle of $15°50'$ with respect to the vertical to enhance the perception of sparkles. It was illuminated by two direct Solux halogen lights —the same ones used in all of our other experiments[2]— one to each side of the box, oriented towards the sample. Figure 5.2 shows the SPD of the Solux lights, and Figure 5.3 shows some photographs of this setup.

[1] IES is the standard file format recommended by the Illuminating Engineering Society (IES) for electronic transfer of photometric data. The IES file defines the input electrical power (W) of the light, as well as the luminous flux and luminous intensity for multiple different angles.

[2] Solux incandescent lights at 4700 Kelvins, running at 12 volts and 50 watts each, with a 36-degree illumination cone

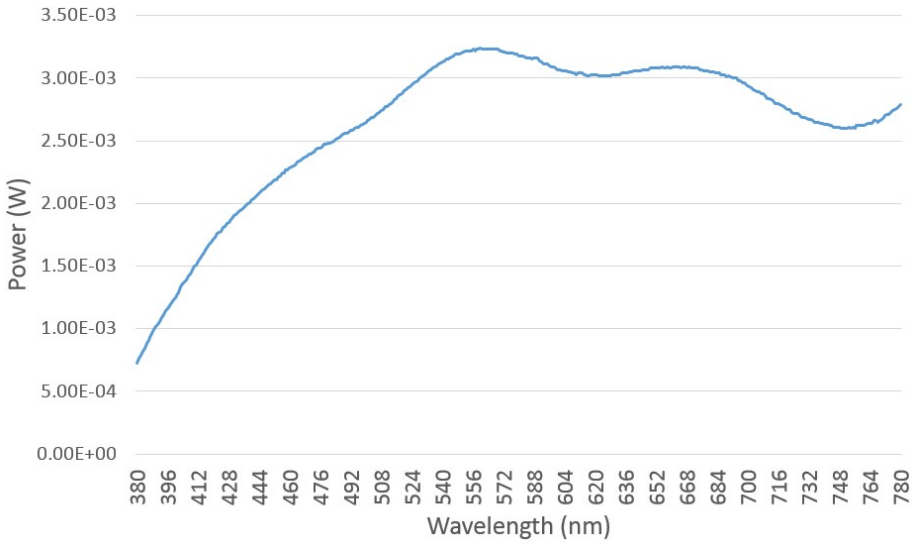



Figure 5.2: Spectral power distribution of the Solux lights at 4700K, 12v, 50w, 36°. Measured with a CS-2000 spectroradiometer with a 0.2-degree angle and an integration time of 0.401 seconds.

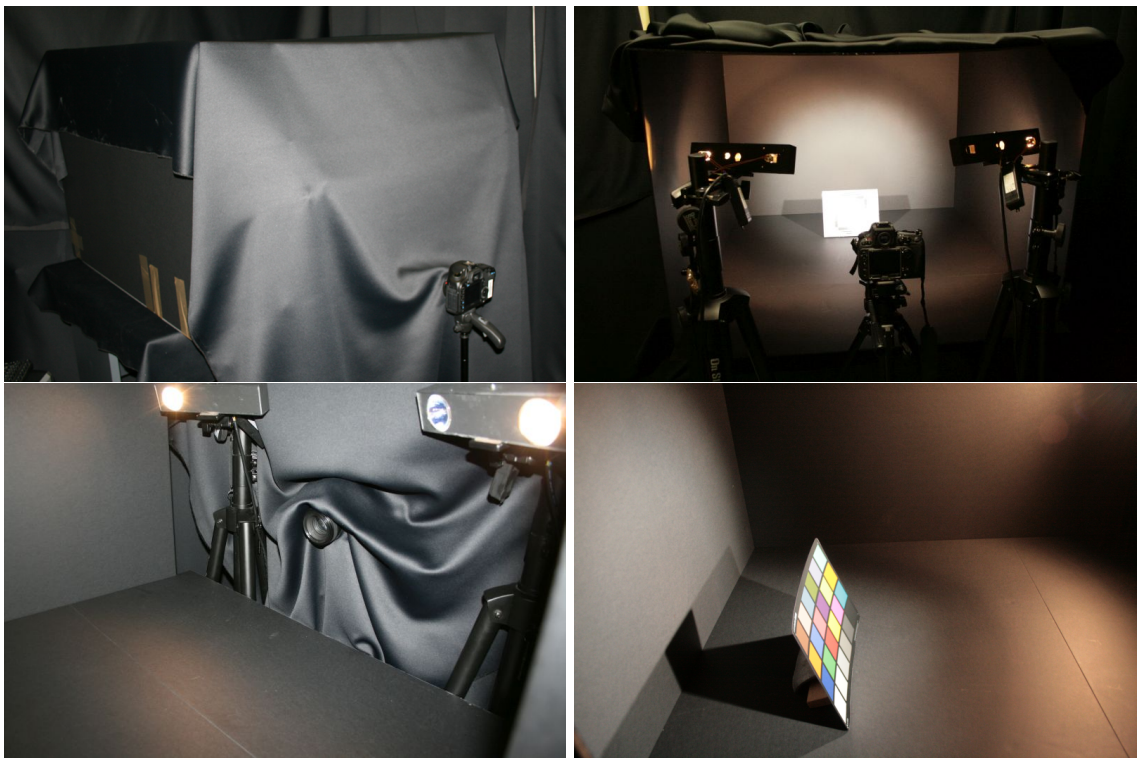


Figure 5.3: Acquisition light cabin.

| | |
|---|---|
| Focal length: | 50 mm |
| Aperture: | f/8 |
| ISO: | 800 |

Table 5.1: Fixed Nikon D800 parameters for the comparison of photographs and simulations.

In order to provide accurate comparisons, the real scene has to be reproduced as close as possible in the simulation —another reason why it is advisable to keep the scene simple— because any difference between them will result in different calculations and, therefore, different images, and possibly, wrong validation results. Therefore, we took measurements of every distance, size, and inclination angles in the scene, and did not change it until the entire validation process was finished. Similarly, it is also important to properly characterize the optical behavior of the materials used in the scene, in order to reproduce it in the simulation. We took BRDF measurements (as defined in Section A.15 of the glossary) from small samples of each material, using a beam diameter of 2.1° on the incident plane, and 4.2° on the vertical plane. The beam diameter is important because it determines the surface area for which the BRDF is computed. If the area is very narrow, the measurements might not contain enough variability, which will translate in lower texture detail —similar to the effect of an averaging filter. This is the case in Figure 5.12, where the background has a different texture in the photograph than the simulation because the material's BRDF was measured with a beam diameter smaller than the texture pattern.

If a complete characterization of the scene is not possible, we recommend performing a thorough analysis of the eventual repercussions of each measurement inaccuracy, in order to consider them in case of possible discrepancies between the images.

## 5.3 Data acquisition from test scene photographs

Once the scene has been fully characterized, we can proceed to take photographs of the scene. Since the position of the camera also needs to be reproduced in the simulation, it is advisable to perform this step before creating the virtual scene —especially in experiments where the position of the camera might change— and register each camera position used in the experiment to be simulated later on.

The photographs were taken with the same DSLR camera (Nikon D800) keeping all the parameters fixed (Table 5.1) except for the exposure times. In order to have a wide enough dynamic range, we took photographs with shutter speeds of 1/8, 1/15, 1/30, 1/60, 1/125, and 1/250 seconds. Since the samples are inclined, it is important to choose the right focus point, because otherwise the depth of field may produce blurry images —especially in the case of sparkling samples. We used a focus chart to focus on the center of the sample, so that there is the same amount of negative and positive offset with respect to the focus point. Similarly, we took photographs of a Macbeth color chart before each series for color control purposes.

The samples are inclined with respect to the camera and the illumination to increase the perception of sparkles. However, depending on the application and the position of the camera, the inclination of the samples and camera with respect to each other may vary; this is not an issue as long as the angles and distances are measured each time as we said earlier.

## 5.4 Simulation of the scene in the target rendering engine

The final task required before proceeding with the validations is the simulation of the scene to be used in the rendering engine. As we said, this simulation must reproduce as close as possible the

aspect of the real scene, in order to minimize the amount of discrepancies in the validation results. The materials used for each object must behave according to the corresponding BRDF function. Figure 5.4 shows wireframe views of the scene after modeling with the measurements from the real scene. Figure 5.5 shows a low-quality rendered preview of the same scene.

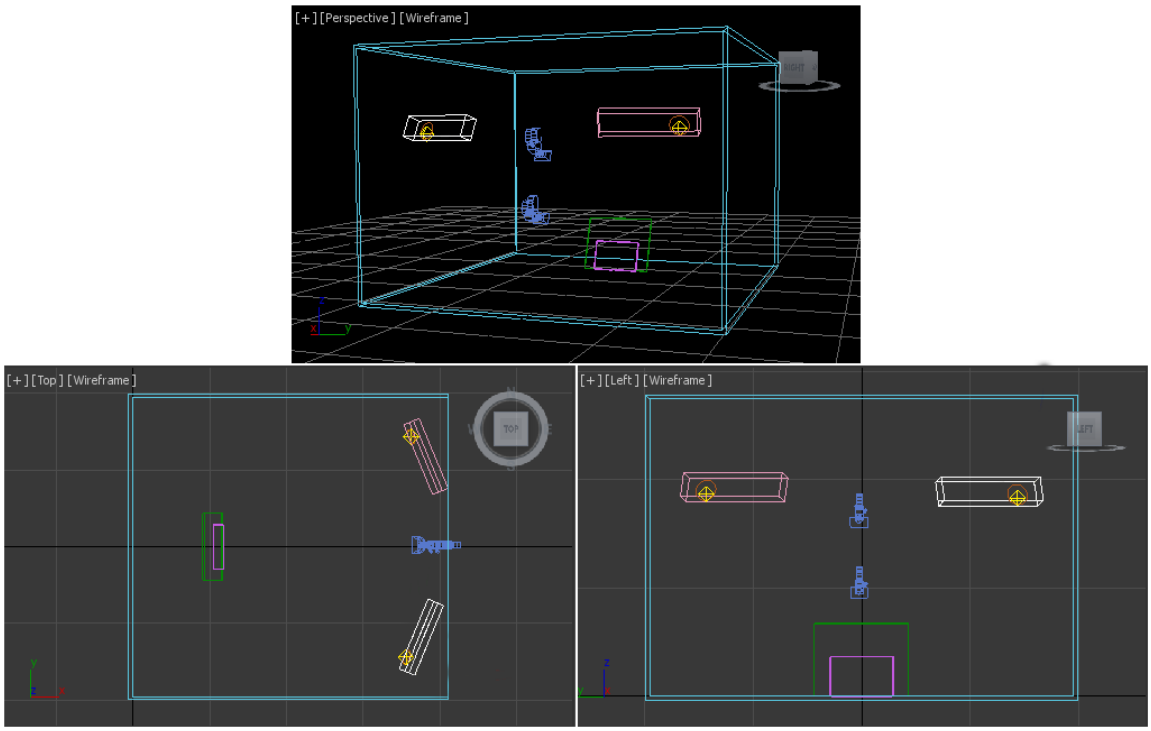



Figure 5.4: Clockwise from the top: perspective, front, and top wireframe views of the validation scene for comparing photographs with simulations. *Screenshots from Autodesk 3ds Max. Modeling and simulation by Thomas Muller.*

For the purpose of colorimetric validation, to ensure that the colorimetric values resulting from the radiometric XYZ data are identical, we have implemented a virtual model of the DSLR camera in the rendering engine. This way, we can evaluate the results visually as well as numerically, disregarding the effect of unaccounted phenomena such as metamerisms (see Section 4.3.2), and identify and isolate eventual errors, by reproducing the exact same setup used to take the photograph. The virtual camera mimics the spectral response of our real camera, the Nikon D800 (Figure 5.6) so, if there are no measurement errors in the scene simulation, and all the camera parameters are set to the same value, photograph and simulation should both contain the same information. We used Eclat Digital's Ocean simulation software (versions 2014 and 2015) to generate a validated ground-truth simulation of the scene. Since this is already a commercialized software, it provides a good foundation to analyze our results, ensuring that any eventual validation discrepancies are only produced by the validation methodology, and not the simulation software.

## 5.5 Objective comparison of photographs with simulations

The simulation process computes the radiance information in the images as a function of the totality of light rays arriving on each pixel, which has no specified range (i.e. 0 to infinity). Similarly, the sensor responses estimated by the camera's characterization matrix are also given on an unspecified range, which does not necessarily have to match that of the rendering engine. Therefore, it is not

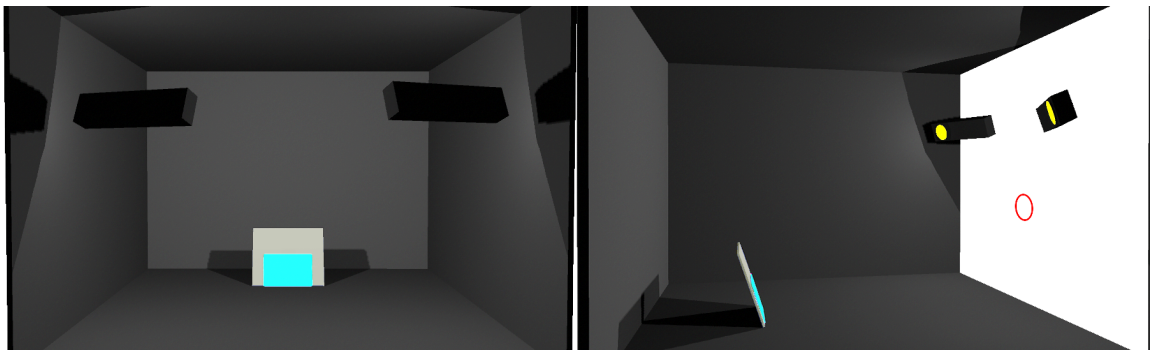


Figure 5.5: Render preview of the scene used for the validations. The red circle on the right image marks the position of the camera as a reference. Note that the brightness has been increased for visualization purposes.

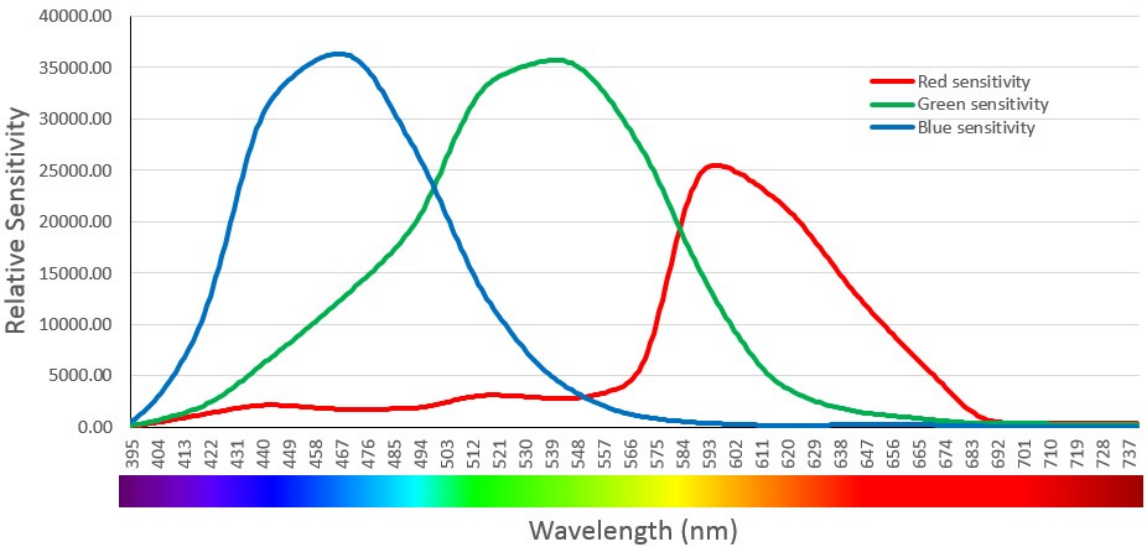
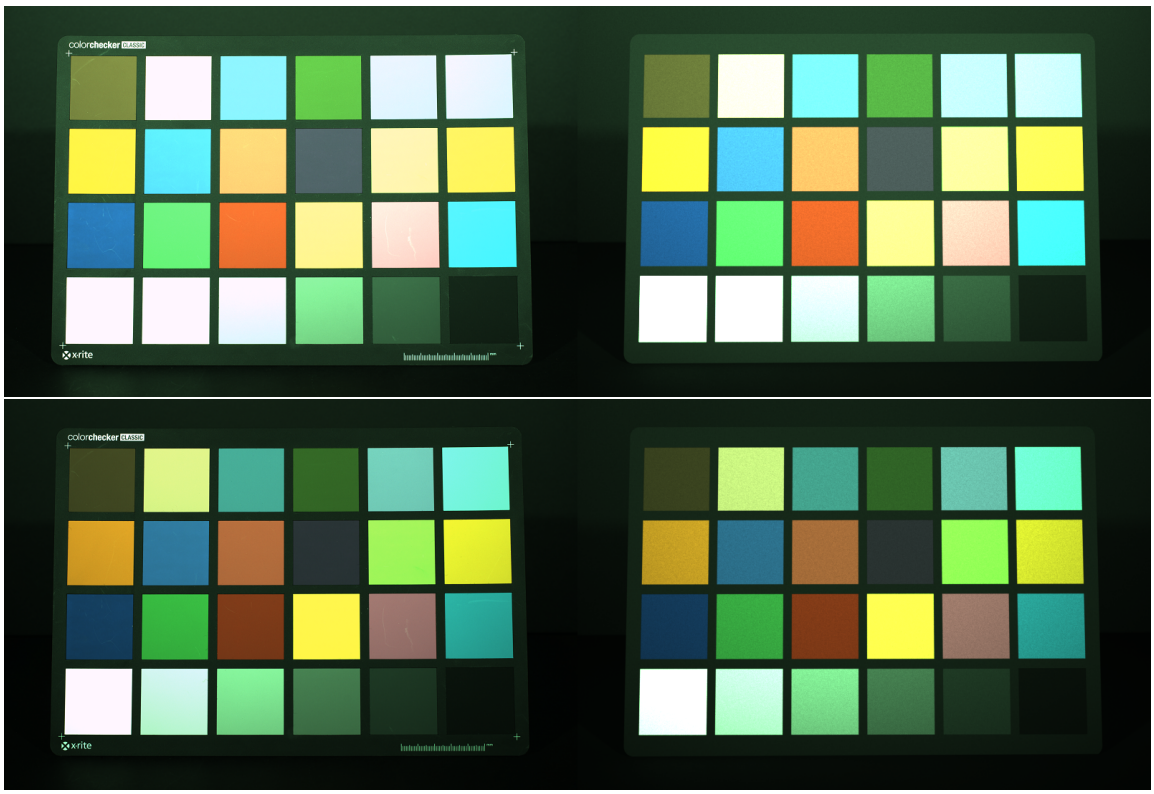



Figure 5.6: Spectral sensitivity of the Nikon D800 [111].

possible to perform a direct value comparison in the XYZ space between the simulations and the photographs.

We assume that, in order for a photograph and a rendered image to produce the same visual response on a human observer, their tristimulus sensory quantities must be proportional by a certain factor, which depends on the amount of light entering the sensor. We can think of any visual stimulus as formed by two independent components: chromaticity and luminance. Chromaticity depends on the SPD of the light, and luminance depends on its intensity; therefore, since varying the amount of light does not change its SPD, it will only affect the luminance component of the stimuli, but the chromaticity will be conserved. We can transform the XYZ tristimulus values into the derived CIE xyY color space, which expresses color stimuli in terms of their chromaticity and luminance —the two first components of this space (xy) represent the chromaticity coordinates (color) of the stimulus, whereas the third component (Y) represents its luminance (brightness). This allows us to compare visual responses by separating chromaticity from luminance.

Figure 5.7 shows two photographs (left) of a Macbeth color chart taken with shutter speeds of 1/8 and 1/15 seconds, and their corresponding simulations (right). Note that both the photographs and the simulations are represented in the camera's own RGB color space and have not been either calibrated nor white balanced. We can do this because we are comparing two images represented



| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 |

Figure 5.7: Photograph (left) and simulation (right) of a Macbeth color chart with total exposures corresponding to a shutter speed of: (a) 1/8 seconds (top row); and (b) 1/15 seconds (bottom row). The images are in the camera's own RGB color space. They have been neither calibrated nor white balanced. The table below indicates the reference number of each color square in the chart.

in the same color space, and under identical illumination conditions [112]; however, a calibration and white balance would be required if we were to perform visual comparisons on different media (e.g. a photograph with a physical reference). We can transform our data into the xyY color space, to represent luminance and chromaticity information separately. Typically, the xy space is not appropriate to represent chromaticity distances, because it is not perceptually uniform [113]. This means that: the same x-y distances do not correspond to the same perceptual differences to a human observer for different parts of the space; and a given difference in x is perceived differently by a human observer as the same difference in y. However, at this point we just want to compare chromaticity and luminance independently as a way to validate the similarity between the simulation and the photographs, so we can represent this relation using distances in the xyY color space.

Before we proceed any further, there are some things that we must consider. Firstly, it may occur that some test scenes have a higher dynamic range (i.e. difference between the highest and lowest luminance value) than the characterization scene, resulting in some oversaturation or undersaturation; in this case, it is advisable to remove these colors from the test set if possible or, at least, to keep in mind that they may produce unexpected results. Furthermore, since the camera's XYZ response was characterized under some illumination conditions that may differ — and in fact most times they do— from those of the actual scene where the test photographs are

taken, we may obtain some colors which fall out of the characterized range. If this happens, using the camera's characterization matrix to estimate sensor responses may produce unexpected results. For this reason, it is recommendable to characterize the devices with a set of samples with similar properties to those that will be used in the tests. However, since this is not our case, we must take this phenomenon into account, by omitting from the test set those samples which produce values outside the characterized range. Figure 5.8, shows the luminance values measured with a spectroradiometer during the camera characterization process, where oversaturated samples have already been removed for simplification. The red rectangle indicates the "safe zone" where luminance stimuli, estimated from a photograph, can be trusted according to the camera's characterization. This means that sample photographs with a estimated Y luminance under 9.41 and over 120.8 should be excluded from the comparisons or, at least, not be considered for making decisions about the results.

Table 5.2 represents the same colors from the Macbeth color chart images in Figure 5.7, using the xyY space, as well as the differences in the x ($\Delta$x) and y ($\Delta$y) chromaticity coordinates, and the total Euclidean distance between the points ($\Delta$xy). Oversaturated samples, as well as those outside the characterized range zone, have been excluded from this list to avoid confusion.

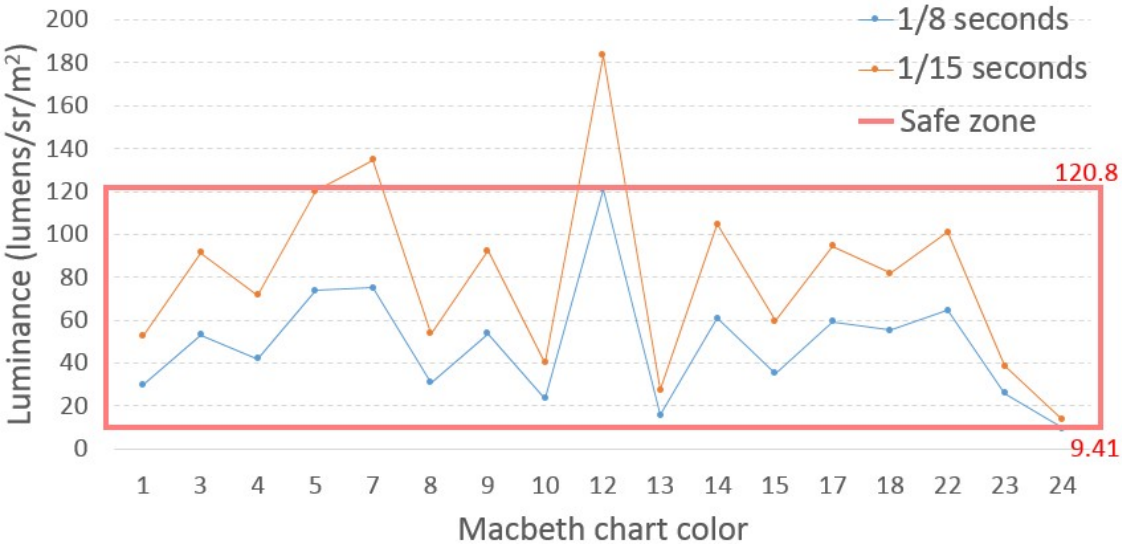
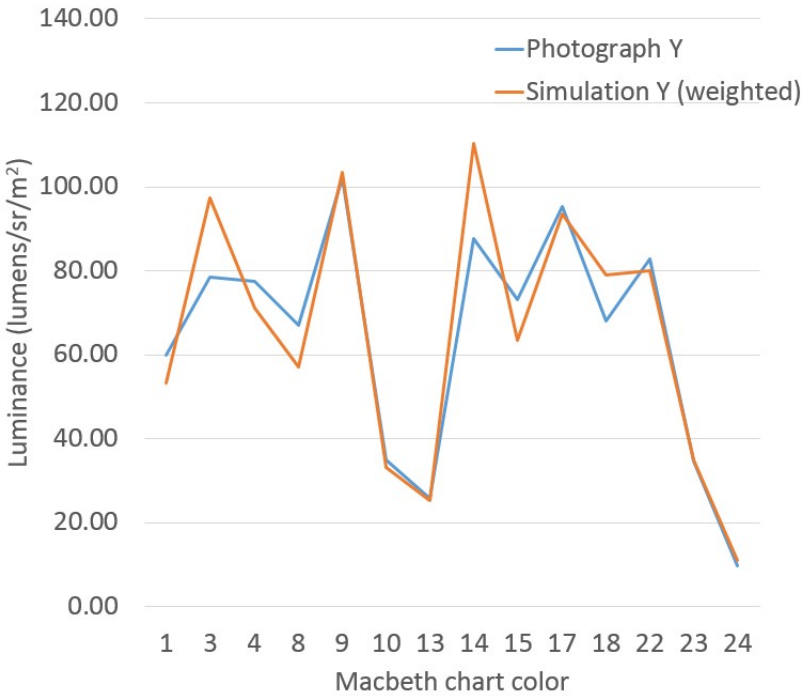



Figure 5.8: Luminance of the Macbeth color chart measured with a spectroradiometer for the camera characterization process. Luminance values outside the red rectangle may produce produce unexpected results.

Looking at these values, we can see that the chromaticity coordinates differ very little between the photographs and the simulations for both images; however, the luminance (Y) of the colors in the photographs is much higher than that of the corresponding colors in the simulation. We have rejected the idea that this is due to measurement errors during the characterization of the real scene, because in that case the magnitude of the values would be much closer between the photographs and the simulations. A more likely explanation is that it is related to the characteristics of the ray-tracing rendering process.

The rendering engine generates the simulations using a Ray-tracing algorithm where the total radiance in each image pixel is calculated from a series of "light rays" intersecting with the position of the pixel. If the radiant power of these rays is weaker than in reality, the simulation would produce radiance (and, thus, luminance) values at a different scale, differing by a multiplicative factor. This hypothesis is supported by some of our tests. Figures 5.9 and 5.10 compare the luminance values in the photographs with the luminance values from the same colors in the simulation, after scaling them by a multiplicative factor —obtained as the average of the ratio between each pair of luminance

values. As we can see, the luminance values in both images differ very little after scaling, which is supported by their high correlation. We attribute larger differences, like those in some colors in Figure 5.9, to the different color sensitivities of the filters in the camera sensor (Figure 5.6).



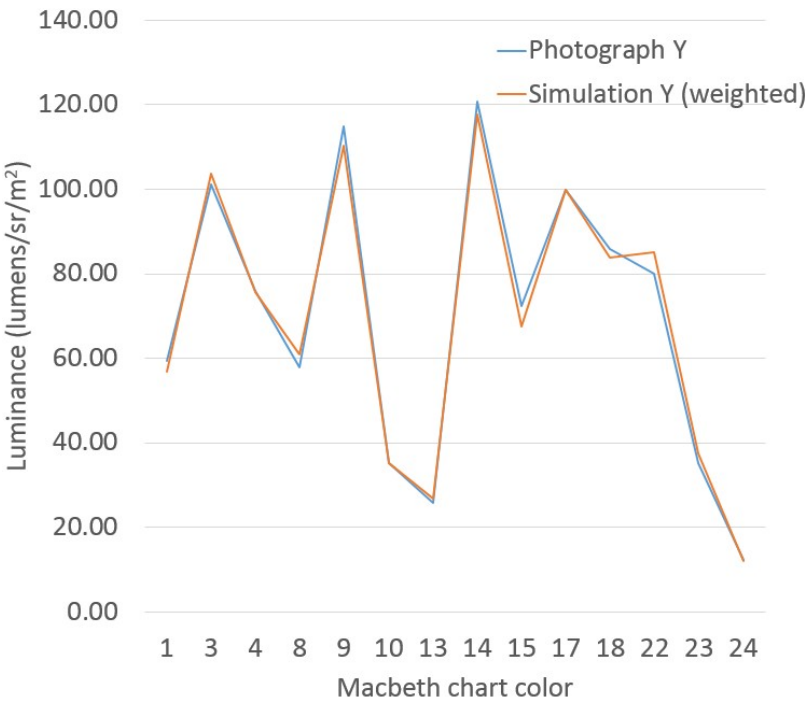

| Color | ΔY |
|---|---|
| 1 | 6.40 |
| 3 | 18.90 |
| 4 | 6.58 |
| 8 | 9.92 |
| 9 | 1.34 |
| 10 | 1.79 |
| 13 | 0.66 |
| 14 | 22.62 |
| 15 | 9.73 |
| 17 | 1.80 |
| 18 | 11.15 |
| 22 | 2.78 |
| 23 | 0.23 |
| 24 | 1.26 |
| Correlation | 0.95 |

Figure 5.9: Luminance comparison between a photograph (at a 1/8-second shutter speed) and its simulation, after scaling the latter by a multiplicative factor obtained as an average ratio.

In general, the results presented above suggest that we can achieve simulations which are very accurate in terms of chromaticity, and differ in luminance by some multiplicative (intensity) factor. However, these results are only numerical differences. Since we are concerned about visual perception, we must find the correspondence between these numerical quantities and perceptual differences. Therefore, color differences between the images must also be analyzed using a more perceptually uniform metric. We chose to use the industry standard CIEDE2000 metric (ΔE00), because it provides improved perceptual uniformity with respect to its predecessor, the CIE DeltaE 1994 metric, which in turn is based on RIT-DuPont results from automotive paint experiments [114, 115, 116]. CIEDE2000 operates on the CIE 1976 Lab (CIELAB) color model, which represents colors in terms of lightness and chrominance (chrominance-based color space) with respect to a given white point. As we explained earlier, the choice of the white point can be arbitrary, since both images have been obtained under the same environment; therefore, we used a standard ICC white point, which is Matlab's[91] default[3]. The color differences in Figure 5.7 were calculated for an averaged area around the center of each color square. They are shown numerically in Table 5.3 as percentages, and graphically in Figure 5.11, with the color scale range adapted to a 0-10 range to facilitate visual inspection.

The results in Table 5.3 show that generally the CIEDE2000 color difference increases as the shutter speed decreases (total exposure increases) as the luminance response approaches the oversaturation limit —which is why we recommend to remove them from the test set. That occurs because CIELAB is represented in a chrominance-based color space and, therefore (unlike in chromaticity-based color spaces), chromaticity and luminance are very much linked. The CIEDE2000 color

[3]As defined in Matlab's[91] implementation of this metric, the ICC white point is the *ICC standard profile connection space illuminant; a 16-bit fractional approximation of D50*




| Color | ΔY |
|---|---|
| 1 | 2.41 |
| 3 | 2.53 |
| 4 | 0.15 |
| 8 | 3.03 |
| 9 | 4.53 |
| 10 | 0.09 |
| 13 | 1.06 |
| 14 | 3.03 |
| 15 | 4.77 |
| 17 | 0.19 |
| 18 | 1.99 |
| 22 | 5.15 |
| 23 | 2.15 |
| 24 | 0.38 |
| Correlation: | 1.00 |

Figure 5.10: Luminance comparison between a photograph (at a 1/15-second shutter speed) and its simulation, after scaling the latter by a multiplicative factor obtained as an average ratio.

difference equation introduces corrections via a series of weighting functions, $S_L$, $S_C$, and $S_H$ [117], to compensate for the influence from the luminance; that is why, in the case of the 1/8 second shutter speed photograph, the highest values for the $S_L$ function occur for the samples which we have removed in previous tests (2, 5, 6, 7, 11, 12, 16, 19, 20 and 21).

So far we have only obtained color differences for the Macbeth color chart, but what is more interesting for us is to analyze these color differences in the case of our paint materials. For that we have chosen a sample green paint plate[4], and computed the same differences as before. It is worth noting that, since the plate was positioned with a certain inclination with respect to the camera and the light sources, there is some amount of vertical color gradient due to non-uniform lighting. For this reason, we have taken three sample colors at the top, middle and bottom positions of the plate, and compared the images at these three positions.

The tolerance of the HVS to color differences has been the subject of multiple studies, notably those by Alman, Snyder, Berns, and Melgosa [118, 119, 120, 114, 121], which provide a way to interpret color difference magnitudes in different fields of application. Most recently, Berns and Hou [115] provided some CIEDE2000 statistics for the RIT-DuPont dataset, indicating a mean CIEDE2000 difference of around 1, and a standard deviation of around 0.4. Since these results are based on automotive paint datasets, we can use them to judge the goodness of our color difference results. We can then consider that, in our context of paint colors, any color pair which differ by a CIEDE2000 of 1.4 or less, will be considered as equal by a human observer.

As we can see in Table 5.4, the differences in the xyY color space are much bigger than they were for the Macbeth chart. The XYZ tristimulus values are estimated from the photograph using the camera's characterization matrix, which in turn was obtained using samples from the Macbeth color chart. Therefore, we can expect estimations for colors outside the characterized gamut to be of lower quality. This is why, as we mentioned earlier, the characterization should be performed with

[4]Commercial name: Vaillant Green; Grain code: A001; Color code: KRM (7M)

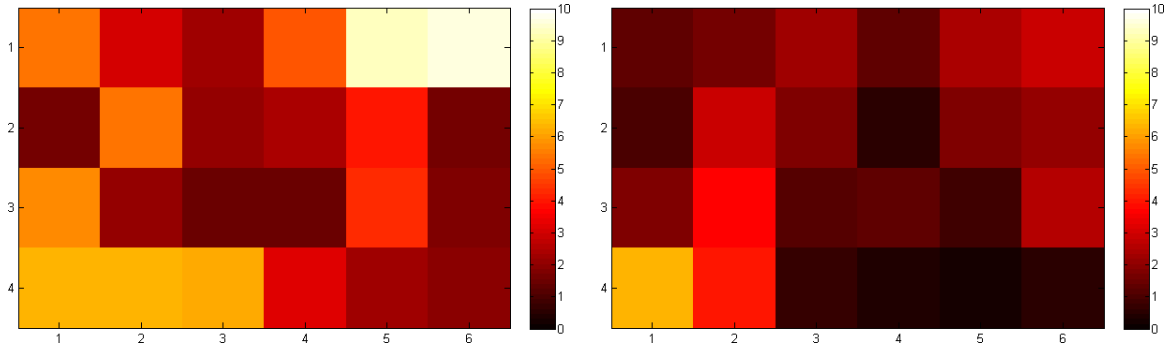



Figure 5.11: Graphical representation of the CIEDE2000 color differences between the photograph and the simulated image in Figures 5.7(a) (1/8 seconds, left), and 5.7(b) (1/15 seconds, right), using the standard ICC white point for the color space conversions to CIELAB. The images show a color representation of the differences, with the color scale range adapted to a 0-10 range to facilitate visual inspection.

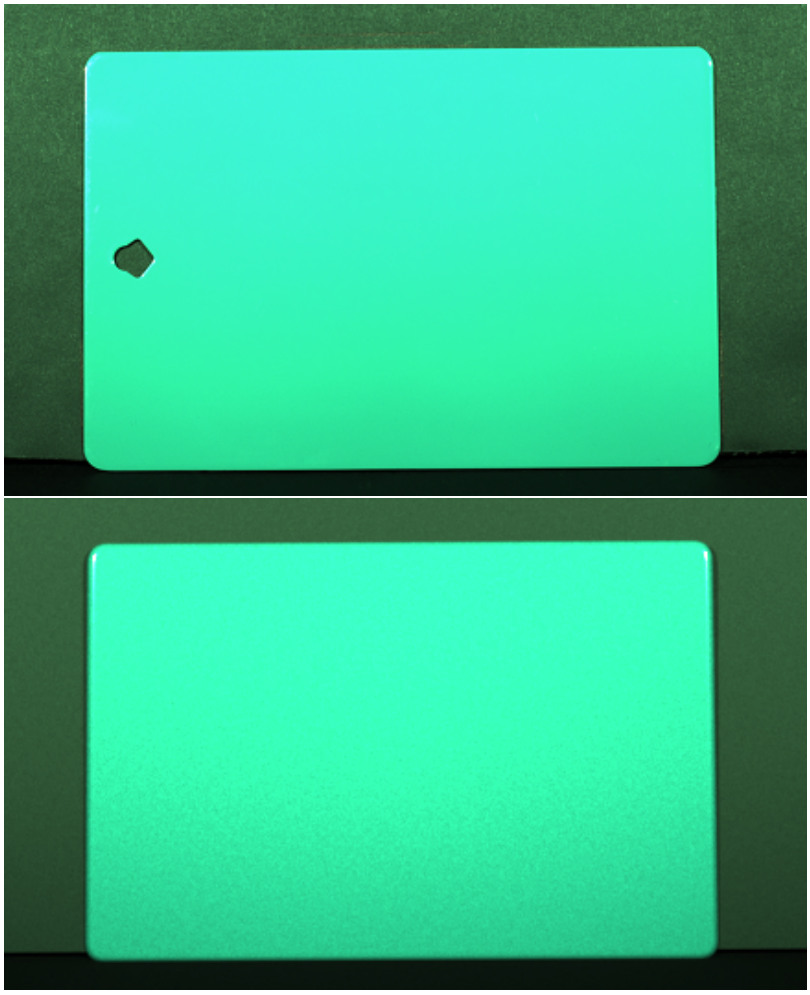


Figure 5.12: Photograph (top) and simulation (bottom) of a sample green paint plate, with a total exposure corresponding to a shutter speed of 1/8 seconds. The images are in the camera's own RGB color space and have not been calibrated nor white balanced.

color samples of similar optical properties to those used in the tests. Nonetheless, the perceptual behavior is different. As we can see in Table 5.5, when we consider selected areas individually, the CIEDE2000 color differences are very high; however, this difference is drastically reduced when we compare a larger averaged area. If we look again at Table 5.2 we can see that, for an exposure of 1/8 seconds, a much larger luminance difference is obtained for the green square (number 14) of the Macbeth chart. Regardless of the fact that the images have not been calibrated —which means that they have double number of green pixels from the camera's CFA— this shows that the camera's green filters are more sensitive to luminance. Looking at Table 5.5, we can see that the luminance differences ($\Delta$L) are indeed very large in each individual area, but with negative symbol. This means that, despite the colors being very different locally, the contribution of each one to the globally perceived color compensates for local differences. This phenomenon shows that, although numerical results are often useful, it is very important to perform visual tests as well, to support or refute the numerical quantities. In fact, although the color differences under xyY were quite big, we have obtained a perceptual color difference of 0.85 for the averaged area, which is smaller than the mean value of 1.00 provided by Berns and Hou [115]. Therefore we can consider that, in the case of this paint plate, when observed globally under the defined observation conditions, a human observer would not discern the difference between the photograph and the simulation.

## 5.6 Conclusion

In order to perform the objective validation of a rendering engine, we need ground-truth information about the scene that we are rendering. We propose to use radiometrically-calibrated photographs of the scene, under a set of well known conditions, as ground-truth information to measure the amount of realism of the results. Although similar methodologies have been proposed in the past [122, 123], only the results and a brief description of the process are usually given. Furthermore, we have only found cases where this methodology is used for colorimetric and photometric comparisons, but never for radiometric comparisons.

In this chapter, we have presented a complete methodology that allows us to render a computer simulation of a real scene with a minimum number of perceptual differences. Since the photographs are taken with a radiometrically calibrated camera, as described in Section 4.3.2, we can obtain both colorimetric and radiometric information from them. Our methodology describes every step, from the setup of the physical scene to the generation of the rendered images, presenting the most important factors to keep in mind at each step to minimize the amount of discrepancies between the real scene, the photograph, and the simulation.

Following the proposed methodology, we can obtain two sets of data from each image: radiometric and colorimetric. The radiometric data are used for numerical comparisons, and the colorimetric data are the images used for visual comparison and color difference metrics, which allows for a double validation.

Our results show that in general we can obtain simulations with very small differences from the photographs, and have highlighted the need to characterize the camera with sample color surfaces whose optical behavior resembles that of the test materials. Otherwise, we may obtain good results for images of the same type of materials used for characterization, but the differences might be larger for materials with different spectral reflectances, since we may predict the wrong radiometric responses from the color images.

Although there is still a lot of work left to do in this area, being able to simulate a real scene with such level of perceptual fidelity supposes an important step in the right direction for the validation of a rendering model.

| | 1/8 seconds | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Photograph** | | | **Simulation** | | | **Difference (Δ)** | | |
| **Color** | x | y | Y | x | y | Y | Δx | Δy | Δxy |
| 1 | 0.44 | 0.38 | 59.81 | 0.44 | 0.39 | 0.07 | 0.00 | 0.01 | 0.01 |
| 3 | 0.31 | 0.27 | 78.39 | 0.29 | 0.32 | 0.13 | 0.02 | 0.05 | 0.05 |
| 4 | 0.39 | 0.46 | 77.61 | 0.38 | 0.46 | 0.09 | 0.01 | 0.00 | 0.01 |
| 8 | 0.26 | 0.26 | 67.12 | 0.25 | 0.23 | 0.08 | 0.02 | 0.03 | 0.04 |
| 9 | 0.46 | 0.35 | 102.17 | 0.51 | 0.34 | 0.14 | 0.05 | 0.01 | 0.05 |
| 10 | 0.35 | 0.27 | 34.87 | 0.35 | 0.27 | 0.04 | 0.00 | 0.00 | 0.00 |
| 13 | 0.21 | 0.18 | 25.84 | 0.21 | 0.19 | 0.03 | 0.01 | 0.00 | 0.01 |
| 14 | 0.35 | 0.42 | 87.70 | 0.34 | 0.51 | 0.15 | 0.01 | 0.09 | 0.10 |
| 15 | 0.55 | 0.34 | 73.11 | 0.58 | 0.34 | 0.08 | 0.03 | 0.00 | 0.03 |
| 17 | 0.41 | 0.28 | 95.43 | 0.44 | 0.28 | 0.13 | 0.03 | 0.00 | 0.03 |
| 18 | 0.25 | 0.27 | 67.94 | 0.23 | 0.32 | 0.11 | 0.02 | 0.05 | 0.05 |
| 22 | 0.36 | 0.37 | 82.77 | 0.36 | 0.37 | 0.11 | 0.00 | 0.01 | 0.01 |
| 23 | 0.35 | 0.37 | 34.72 | 0.36 | 0.37 | 0.05 | 0.01 | 0.01 | 0.01 |
| 24 | 0.34 | 0.37 | 9.87 | 0.36 | 0.37 | 0.01 | 0.02 | 0.00 | 0.02 |
| | | | | | | Average: | 0.02 | 0.02 | 0.03 |

| | 1/15 seconds | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Photograph** | | | **Simulation** | | | **Difference (Δ)** | | |
| **Color** | x | y | Y | x | y | Y | Δx | Δy | Δxy |
| 1 | 0.44 | 0.38 | 59.31 | 0.44 | 0.39 | 0.04 | 0.00 | 0.01 | 0.01 |
| 3 | 0.29 | 0.31 | 101.17 | 0.29 | 0.32 | 0.07 | 0.00 | 0.01 | 0.01 |
| 4 | 0.39 | 0.46 | 75.87 | 0.38 | 0.46 | 0.05 | 0.01 | 0.00 | 0.01 |
| 8 | 0.24 | 0.22 | 57.97 | 0.25 | 0.23 | 0.04 | 0.01 | 0.01 | 0.02 |
| 9 | 0.51 | 0.33 | 114.97 | 0.51 | 0.34 | 0.07 | 0.01 | 0.01 | 0.01 |
| 10 | 0.35 | 0.27 | 35.23 | 0.35 | 0.27 | 0.02 | 0.00 | 0.00 | 0.00 |
| 13 | 0.21 | 0.18 | 25.73 | 0.21 | 0.19 | 0.02 | 0.01 | 0.00 | 0.01 |
| 14 | 0.34 | 0.51 | 120.68 | 0.34 | 0.51 | 0.08 | 0.00 | 0.00 | 0.00 |
| 15 | 0.55 | 0.34 | 72.42 | 0.58 | 0.34 | 0.05 | 0.03 | 0.00 | 0.03 |
| 17 | 0.43 | 0.28 | 99.79 | 0.44 | 0.28 | 0.07 | 0.00 | 0.00 | 0.00 |
| 18 | 0.23 | 0.31 | 85.94 | 0.23 | 0.32 | 0.06 | 0.00 | 0.02 | 0.02 |
| 22 | 0.36 | 0.37 | 80.07 | 0.36 | 0.37 | 0.06 | 0.00 | 0.01 | 0.01 |
| 23 | 0.35 | 0.36 | 35.23 | 0.36 | 0.37 | 0.02 | 0.01 | 0.01 | 0.01 |
| 24 | 0.35 | 0.36 | 12.31 | 0.36 | 0.37 | 0.01 | 0.01 | 0.02 | 0.02 |
| | | | | | | Average: | 0.01 | 0.01 | 0.01 |

Table 5.2: xyY tristimulus values of the Macbeth color samples and non-perceptual chromaticity distances in the CIE xy color space.

| | 1/8 seconds | | | | 1/15 seconds | | | |
|---|---|---|---|---|---|---|---|---|
| **Color** | **ΔE00** | $S_L$ | $S_C$ | $S_H$ | **ΔE00** | $S_L$ | $S_C$ | $S_H$ |
| 1 | 5.38 | 1.21 | 2.72 | 1.53 | 1.40 | 1.37 | 2.04 | 1.35 |
| 2 | 3.00 | 1.75 | 1.28 | 1.06 | 1.66 | 1.66 | 3.50 | 1.83 |
| 3 | 2.34 | 1.65 | 2.63 | 1.57 | 2.23 | 1.26 | 2.59 | 1.54 |
| 4 | 4.90 | 1.36 | 4.12 | 2.37 | 1.37 | 1.25 | 2.84 | 1.81 |
| 5 | 9.35 | 1.71 | 1.78 | 1.31 | 2.49 | 1.41 | 2.50 | 1.50 |
| 6 | 9.62 | 1.72 | 1.68 | 1.29 | 2.96 | 1.62 | 3.09 | 1.74 |
| 7 | 1.61 | 1.72 | 4.67 | 1.86 | 1.08 | 1.32 | 3.94 | 1.57 |
| 8 | 5.44 | 1.53 | 2.88 | 1.88 | 2.82 | 1.21 | 2.30 | 1.68 |
| 9 | 2.13 | 1.58 | 3.41 | 1.47 | 1.77 | 1.21 | 3.00 | 1.41 |
| 10 | 2.37 | 1.22 | 1.53 | 1.17 | 0.59 | 1.47 | 1.31 | 1.10 |
| 11 | 3.95 | 1.73 | 2.75 | 1.42 | 1.86 | 1.62 | 4.67 | 2.49 |
| 12 | 1.67 | 1.72 | 4.33 | 1.79 | 2.13 | 1.61 | 4.67 | 1.95 |
| 13 | 5.77 | 1.21 | 2.78 | 1.65 | 1.80 | 1.43 | 2.06 | 1.43 |
| 14 | 2.13 | 1.60 | 4.42 | 2.58 | 3.69 | 1.26 | 4.05 | 2.39 |
| 15 | 1.43 | 1.26 | 4.42 | 1.74 | 1.23 | 1.30 | 3.00 | 1.43 |
| 16 | 1.55 | 1.73 | 3.22 | 1.53 | 1.38 | 1.72 | 4.57 | 1.84 |
| 17 | 4.33 | 1.64 | 1.83 | 1.18 | 0.87 | 1.22 | 1.96 | 1.24 |
| 18 | 1.78 | 1.62 | 3.11 | 1.72 | 2.62 | 1.22 | 2.63 | 1.51 |
| 19 | 6.31 | 1.75 | 1.16 | 1.08 | 6.31 | 1.75 | 1.16 | 1.08 |
| 20 | 6.31 | 1.75 | 1.16 | 1.08 | 3.96 | 1.68 | 2.22 | 1.48 |
| 21 | 6.13 | 1.73 | 1.40 | 1.20 | 0.65 | 1.50 | 3.43 | 2.13 |
| 22 | 3.14 | 1.51 | 3.44 | 2.13 | 0.37 | 1.21 | 2.47 | 1.68 |
| 23 | 2.22 | 1.25 | 2.33 | 1.61 | 0.25 | 1.49 | 1.94 | 1.41 |
| 24 | 2.02 | 1.61 | 1.66 | 1.29 | 0.51 | 1.71 | 1.23 | 1.10 |

Table 5.3: CIEDE2000 color differences between the photographs and the simulated images in Figure 5.7, using the standard ICC white point for the color space conversions to CIELAB. $S_L$, $S_C$, and $S_H$ indicate the value of the weighting functions for lightness, hue, and chroma, respectively.

| | **Photograph** | | | **Simulation** | | | **Difference (Δ)** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Area** | x | y | Y | x | y | Y | Δx | Δy | Δxy |
| Top | 0.3457 | 0.3689 | 9.7939 | 0.2353 | 0.4105 | 0.1085 | 0.11 | 0.04 | 0.12 |
| Middle | 0.3453 | 0.3688 | 9.6353 | 0.2350 | 0.4109 | 0.0985 | 0.11 | 0.04 | 0.12 |
| Bottom | 0.3444 | 0.3691 | 9.5153 | 0.2347 | 0.4112 | 0.0868 | 0.11 | 0.04 | 0.12 |
| Average | 0.3451 | 0.3689 | 9.6347 | 0.2350 | 0.4109 | 0.0989 | 0.11 | 0.04 | 0.12 |

Table 5.4: xyY tristimulus values, and non-perceptual chromaticity distances in the CIE xy color space, of the photograph and the simulated images in Figure 5.12.

| | **Photograph** | | | **Simulation** | | | **CIEDE2000** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Area** | L | a | b | L | a | b | ΔL | $\Delta H_{ab}$ | ΔE00 | $S_L$ | $S_C$ | $S_H$ |
| Top | 87.53 | -52.90 | 2.71 | 89.91 | -58.70 | 13.25 | 2.38 | -9.63 | 2.52 | 1.60 | 3.55 | 1.91 |
| Middle | 87.08 | -56.92 | 11.84 | 89.30 | -61.23 | 19.49 | 2.22 | -6.29 | 2.14 | 1.60 | 3.76 | 2.09 |
| Bottom | 84.69 | -60.39 | 22.81 | 80.00 | -56.68 | 20.02 | -4.69 | 1.34 | 3.32 | 1.51 | 3.81 | 2.18 |
| Average | 86.40 | -57.04 | 12.93 | 86.53 | -59.01 | 17.71 | 0.13 | -4.11 | 0.85 | 1.57 | 3.70 | 2.07 |

Table 5.5: CIEDE2000 color differences between the photographs and the simulated images in Figure 5.12, using the standard ICC white point for the color space conversions to CIELAB. $S_L$, $S_C$, and $S_H$ indicate the value of the weighting functions for lightness, hue, and chroma, respectively.

Chapter 6

# Discussion and future work

Industrial design simulations reproduce processes that will eventually be implemented in reality, so the accuracy of a simulation will depend on the specific process that it simulates. In the context of computer graphics, accuracy is not always necessarily synonymous with visual realism, but it depends on the requirements of the specific application. With this in mind, our aim is to obtain perceptually-realistic simulations, in the sense of photo-realism given in Ferwerda's classification [5]. We have proposed several contributions to perceptual realism. First, we propose a methodology for the characterization of the visualization devices, to ensure a controlled reproduction of the simulated color signal at display time. We also propose the use of photographs as ground truth representations of perceptual information from the real scene. A radiometrically-calibrated DSLR camera is used to acquire real scene information, and represent it in terms of human visual sensory responses; this way, the photographs can be calibrated according to perceptual criteria, and used in visual psychophysics experiments. Then, we present results from two visual psychophysics experiments where we analyze the perceptual realism of different methods for representing luminance information from a real scene in images.

Visual perception is the result of a series of very complex mechanisms that take place in the human visual system (HVS), as the light (radiant energy) in a scene arrives into our eyes. The goal of physico-realistic rendering is to create simulations that produce the same radiometric information as the real scene. In some cases (e.g. remote sensing) it is important to compute more information than what is visible by the human eye [124]; however, since our images will be seen by a human observer, we can take advantage of the limitations of the human visual system to reduce the amount of computation. Perceptual realism takes into account these limitations, as well as others like the inability of conventional displays to reproduce the radiometric richness of physico-realistic images. In this case, accuracy is defined in terms of perceptually-realistic rendering, which means that the simulation can trigger the same response as the real scene in the visual system of the observer.

We have focused this research on the case of automobile sparkling paint materials containing a layer of metal particles, or flakes. These materials produce a sparkling effect when they reflect the light from an illuminating source. The perceived aspect of the sparkles depends on the amount of reflected light that arrives on the observer's eye (sparkle luminance), which in turn depends on the depth and orientation of the underlying metal flake within the paint coating, and its position with respect to the light source and the observer. Since the use of 2D visualization eliminates many of the cues that allow us to perceive depth in this type of images, we try to improve perceptual realism by using stereoscopic visualization, which provides additional depth information from binocular disparity.

Given the strong influence of sparkle luminance towards the perceived aspect of these materials, we have performed extensive psychophysics studies to analyze the perception of luminance values in real and virtual scenes. To perform these studies, a perceptually-realistic image validation approach has been defined, using a DSLR camera as an intermediate device to emulate the trichromatic response of the human eye from input spectral power distributions. This allows us to obtain a radiometric map of the scene, expressed in a perceptual tristimulus (XYZ) space analogous to that of visual responses. Image realism at any point of this process is ensured by a full characterization and

calibration methodology that gives us information and control over any transformation undergone by the radiometric signals throughout the entire chain of image acquisition and visualization.

We present our results from two observation experiments. The first experiment has the following aims: to confirm whether or not the use of stereoscopy improves the realism of the images, to try to find a relationship between image exposure and real luminance, and to study how these properties affect the perceptual aspect of the materials in the images. In this experiment, a series of stereoscopic photographs of material samples at different exposures was shown to a group of observers; they were asked to compare them with their corresponding real objects, under the same observation conditions, and select the closest photograph. The results from this experiment corroborate that stereoscopy does indeed improve perceptual realism in the images, by improving depth perception and emphasizing the sparkling effect; it also confirms our previous hypothesis that image luminance has an effect on its brightness and contrast, which in turn affects the perceived size of the sparkles, texture depth, and sparkle-background contrast. They also suggest that, under those conditions, perceptually-close results are obtained for images with a maximum luminance of around 100 $cd/m^2$.

The second experiment analyzes the perceptual effects of different tone reproduction methods. Using the results from the previous experiment, we take as reference exposure the one that produces the images with a maximum luminance closest to 100 $cd/m^2$. A group of observers were asked to compare side by side a series of stereoscopic images with three different exposures —the reference, and the one immediately above and below— and tone-mapped versions of each one of them where the luminance values were scaled up or down to approach as much as possible the target luminance of 100 $cd/m^2$. This experiment shows that sparkle contrast increases as global image brightness decreases, and the other way around, but only for intermediate brightness values, with perceptual similarity decreasing below and above a certain threshold. Moreover, these results reveal interesting facts about tone reproduction: as long as the amount of overexposed pixels stays below a minimum, no information is lost upon display and, in fact, tone-mapped images differ more from the reference than the original ones.

The findings presented above provide important information to validate the accuracy of the simulations. Since each simulation engine can implement a different rendering model, the relationship between the rendering model parameters and the real radiometric values is a priori unknown. Therefore, photographs can be used as a perceptually valid colorimetric and radiometric ground-truth to adjust the parameters of the simulation and validate its results. Following this line of work, we have designed a generic validation methodology that allows to create a simulation of a real scene with perceptually-realistic accuracy, pointing out the most important factors that may influence this accuracy. We have also presented a methodology to compare these simulations with photographs of the same scene, taken under the same conditions.

Initial results with this methodology reveal that, when comparing images of materials with similar optical properties to those used in the camera characterization process, both the chromaticity and perceptual CIELAB2000 differences between the simulation and the photographs, are very small in the absence of oversaturation. Luminance values differ between the simulations and the photographs, but they are related by a multiplicative factor, hinting that they could be in a different scale which we attribute to differences in the radiometric properties of the ray-tracing rendering process. When comparing images of materials with different optical properties than those used for the camera characterization, the absence of reliable radiometric data prediction results in relatively large colorimetric differences. Nonetheless, our results show that perceptual CIELAB2000 differences remain below the accepted tolerance values when large enough areas are compared. Typically, tolerance indices are given to interpret the magnitude of perceptual color differences, but these are defined for material sets with specific properties. We have used existing statistics from the RIT-DuPont automotive paint dataset, given the similarity with our materials; however, the particular properties of this type of sparkling materials might justify the use of its own specific dataset.

As we mentioned in Chapter 1, the LIMA project proposes a double approach, where the aspect

of the materials is to be predicted from their composition, and vice versa. The aim is to find the relationship between optical behavior and aspect perception without having to produce the physical material, using only their simulations. The work produced in this PhD thesis is only a small part of what is certainly a very extensive research area. In this work, we propose some foundations in the field of visuo-perceptual validations of physically correct simulations of sparkling materials. We have concentrated on what we consider one of the most important factors in the simulation of sparkling materials, which is the reproduction of luminance values, and have provided important psychophysics results regarding the effects of luminance over the perception of images of sparkling materials. After some further testings, the validation methodology —presented at the end of this work— to compare material simulations with photographs already permits to integrate all these psychophysics results into the rendering model. However, a lot more work is necessary to ensure a fully perceptual validation for this type of simulations.

Possible areas for future work include, for example: the design of a material-specific characterization method to improve radiometric estimations and, in turn, perceptual image realism; a further analysis of color-related perception; or a study of perceptual thresholds in metal-flake density. We know that the accuracy of human vision varies for different parts of the color spectrum, so the perception of color paints might vary for different hues. In this regard, initial results presented in this thesis suggest that observers could distinguish more easily between different image exposures for green paints than red or blue. We propose to use non-sparkling paint samples to carry out additional analyses of visual perception. The absence of sparkles results in uniform colors that can be used to compute color difference metrics in sample photographs. These color differences can then be compared to the results from psychophysics experiments. Since the camera is calibrated, we can extract radiometric information from the photographs, in order to establish a relation between perceptual tolerance and physical values for selected attributes. We mentioned that one of the reasons why we performed our studies using the XYZ color space instead of LMS is because of the lack of perceptual attributes in the existing LMS color appearance models; performing these additional studies in the LMS color space could provide us with important information to expand current color appearance models in the context of our particular type of materials and observations conditions. Finally, additional work directions include testing the consistency of our results for different visualization devices and under different illumination and observation environments.

Appendix A

# Appendix 1: Glossary

Many of the terms used in this PhD thesis may have different interpretations depending on factors such as the field of study (e.g. computer graphics, design, optics, etc.) or the language. For this reason, the multidisciplinary nature of this work and the fact that it is presented in a french speaking environment —we can find several false friends amongst technical terms in French and English— demand that we ensure uniformity in the definition of the terms. This sections presents a series of definitions that allow for a more objective interpretation of the terms used throughout this work.

## A.1 Real Scene

We will use the term "real" to refer to any physical tangible object, or collection of objects, that exist in real life.

## A.2 Virtual scene

As opposed to a real scene, the term "virtual" will refer to any object, or collection of objects, that is displayed on any type of screen (computer display, projector, etc).

## A.3 Simulated scene

The term "simulation" will refer only to computer-generated virtual scenes, resulting from a simulation rendering process.

## A.4 Radiometry

Set of techniques to measure electromagnetic energy, including the visible light, in terms of absolute power.

## A.5 Photometry

Science that measures visible light in terms of psychophysical quantities that reflect the psychological impressions perceived by a human observer. Photometric quantities are typically normalized with respect to the luminous efficiency curve of the human visual system, which indicates its peak sensitivity to each wavelength under specific illumination conditions (photopic or scotopic). The ratio between photometric and radiometric units is given by the peak sensitivity (at 555 nm), called the luminous efficacy (683 lm/W). Therefore, photometry only applies to visible light measurements, that is, to electromagnetic energy within the visible spectrum.

## A.6 Radiant flux (or radiant power)

Amount of radiant energy transferred per unit time. It is measured in watts(W) or Joules per second (J/s), its equivalent SI base unit. We speak of spectral radiant flux (or spectral radiant power) when the flux is given for each wavelength (watts/nanometer or W/nm).

## A.7 Luminous flux (or luminous power)

Photometric quantity, equivalent to radiant power in the visible spectrum. It is expressed in lumen (lm). To convert between radiometric and photometric units keep in mind that, at 555 nm, a luminous power of 1 lm corresponds to a radiant power of 1/683 W.

## A.8 Irradiance

Radiant power incident on a surface per unit area. It is measured in watts per square meter ($W/m^2$) and it is inversely proportional to the squared distance between the illuminating source and the illuminated surface (inverse square law).

## A.9 Illuminance

Luminous power incident on a surface per unit area. It is measured in lumen per square meter ($lm/m^2$).

## A.10 Solid Angle

A unit solid angle, measured in steradian (sr) is defined as the angle subtended at the center of a sphere with a 1-meter radius by an area of 1 $m^2$ located on its surface. It is a measure of the apparent size of an object for an observer located at a given point away from it, independent of the object's size and its distance from the observer.

## A.11 Radiance

Radiant power that leaves a surface per unit solid angle and unit projected area of that surface. For point sources the term intensity is equivalent to radiance – in practice, any source with a diameter smaller than 1/20th of its distance to the irradiated surface is considered a point source. It is expressed in watts per steradian per square meter ($W/sr/m^2$).

*Language disambiguation:* The word radiance does not exist in french (it is called "luminance énergétique"), although it is sometimes used as an anglicism.

## A.12 Spectral Power Distribution (SPD)

The SPD of a given electromagnetic radiation is the contribution of that illumination to a radiometric or photometric quantity, for each wavelength of a given spectrum range. Typically, we give SPDs for a wavelength range covering the visible portion of the electromagnetic spectrum, roughly

between 400 and 700 nm. It is expressed in $W/sr/m^2$ or $lm/m^2$ per unit wavelength (typically meters or nanometers), respectively for spectral radiant flux and spectral luminous flux.

## A.13 Reflection, Refraction, Transmittance and Absorption

Reflection is caused when a given light wave bounces on the interface between two media (in our case, the air and the material), returning back through the same medium. If the wave is not perfectly reflected, part of the light wave gets refracted through the second material. In turn, the refracted light can be transmitted through and/or absorbed by the second material. The amount of light that gets reflected, refracted, transmitted and absorbed by the material depends, respectively, on the reflectance, refraction, transmittance and absorption indexes of the materials. Similarly, when the light wave is given as a function of the wavelength, we speak of spectral reflectance, spectral refraction, spectral transmittance and spectral absorption.

## A.14 Reflectance factor

Ratio of light reflected from an object on a surface to the amount of light reflected from a perfect diffuser on the same surface, under the same illumination conditions.

## A.15 Bidirectional Reflectance Distribution Function (BRDF)

The BRDF (with units inverse steradians, $sr^{-1}$) describes the amount (proportion) of light reflected when light makes contact with a material, as a function of the illumination and observation angles, relative to the surface normal and tangent, and the wavelength of the light [12]. This is indicated by Expression A.1, where $\omega_i$ is a vector representing the incoming direction of the light, $\omega_r$ represents the outgoing or reflection direction of the light, L is radiance ($W \cdot sr^{-1} \cdot m^{-2}$), and E is irradiance ($W \cdot m^{-2}$).

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{dE_i(\omega_i)} \tag{A.1}$$

## A.16 Snell-Descartes law

When light passes through a boundary between two isotropic media with respective refractive indices $n_1$ and $n_2$, Snell-Descartes' law states that the ratio between the sines of the incidence and refraction angles ($\theta_1$ and $\theta_2$) is equivalent to the ratio of phase velocities ($\nu_1$ and $\nu_2$) and light wavelengths ($\lambda_1$ and $\lambda_2$) in the two media, or the inverse ratio of the refraction indices. This is indicated by Expression A.2.

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{\nu_1}{\nu_2} = \frac{\lambda_1}{\lambda_2} = \frac{n_1}{n_2} \tag{A.2}$$

## A.17 Brightness / Lightness

The quantity of light in a color is expressed with the terms brightness and lightness. Brightness is measured in an absolute scale from dark to bright, which depends on adaptation and may vary over time for different conditions; it is typically used for the colors of light sources, and it expresses the

intensity of the light source in terms such as dark, dim, bright, etc. Lightness, on the other hand, is measured in a relative scale, so it does not change with adaptation and conditions; it is typically used for the colors of reflecting objects, expressing the appearance of the surface in terms such as black, dark gray, light gray, white, etc.

## A.18 Luminance / Luminous reflectance

The photometric or psychophysical quantities corresponding to brightness and lightness are, respectively, luminance and luminous reflectance. Luminance can be seen as the amount (absolute scale) of visible light that arrives in the eye (or a sensor) from a given surface or light source in candles per square meter or, simply, $cd/m^2$; luminous reflectance is the proportion (relative scale) of light that is reflected from a surface from an incident light source. Luminance is an important factor to determine the relationships between colors because it indicates the contrast between them (Figure A.1).

*Language disambiguation:* Be aware that, while the french word "luminance" refers to its english homonym "luminance", the term "luminance énergétique" is used in french to refer to the english word "radiance".

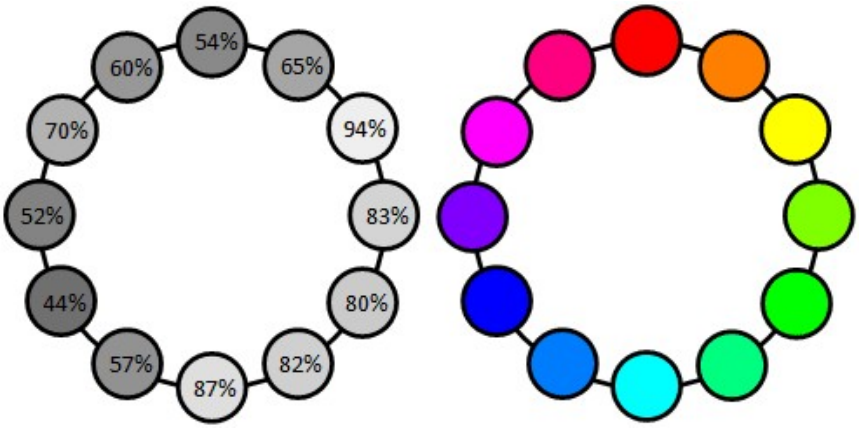



Figure A.1: Main hues of the color wheel with constant saturation (right) and their corresponding lightness levels converted to grayscale (left). When the color saturation level decreases, the luminance increases for hues with natural luminance above 50%, and decreases for hues with natural luminance below 50%. That means that, when increasing the saturation, dark hues become darker and light hues become lighter. [125]

| Quantity | Symbol | SI units | Notes |
|---|---|---|---|
| Luminous energy | $Q_v$ | $lm{\cdot}s$ | Units are sometimes called Talbots. |
| Luminous flux | F | lm | Also called luminous power. |
| Luminous intensity | $I_v$ | cd | An SI base unit. |
| Luminance | $L_v$ | $Cd/m^2$ | Units are sometimes called nits. |
| Illuminance | $E_v$ | lx | Used for light incident on a surface. |
| Luminous emittance | $M_v$ | lx | Used for light emitted from a surface. |
| Luminous efficacy | - | lm/W | Ratio of luminous flux to radiant flux; theoretical maximum is 683. |

Table A.1: SI Photometry Symbols and Units [126, 127]

| Quantity | Symbol | SI units | Notes |
|---|---|---|---|
| Radiant energy | Q | J | Energy. |
| Radiant flux | $\phi$ | W | Radiant energy per unit time, also called radiant power. |
| Radiant intensity | I | $W{\cdot}sr^{-1}$ | Power per unit solid angle. |
| Radiance | L | $W{\cdot}sr^{-1}{\cdot}m^{-2}$ | Power per unit solid angle per unit projected source area. Sometimes confusingly called "intensity". |
| Irradiance | E | $W{\cdot}m^{-2}$ | Power incident on a surface. Sometimes confusingly called "intensity". |
| Radiant emittance or exitance | M | $W{\cdot}m^{-2}$ | Power emitted from a surface. Sometimes confusingly called "intensity". |
| Spectral radiance | $L_\lambda$ ($L_\nu$) | $W{\cdot}sr^{-1}{\cdot}m^{-3}$ ($W{\cdot}sr^{-1}{\cdot}m^{-2}{\cdot}Hz^{-1}$) | Commonly measured in $W{\cdot}sr^{-1}{\cdot}m^{-2}{\cdot}nm^{-1}$. |
| Spectral irradiance | $E_\lambda$ ($E_\nu$) | $W{\cdot}m^{-3}$ ($W{\cdot}m^{-2}{\cdot}Hz^{-1}$) | Commonly measured in $W{\cdot}m^{-2}{\cdot}nm^{-1}$. |

Table A.2: SI Radiometry Symbols and Units [126, 127]

## A.19 Spectroradiometer

Measures the SPD emitted by a source. It is the most complete measuring device because from the SPD we can also compute other radiometric, photometric, and colorimetric quantities.

## A.20 Spectrocolorimeter

Measures the spectral reflectance, transmittance, or relative irradiance of a color sample [11] and calculate its tristimulus values.

## A.21 Tristimulus colorimeter

Measures the chromaticity coordinates of a color surface.

## A.22 CIE

The International Commission on Illumination or CIE (french acronym for Commission internationale de l'éclairage) is the international regulatory body on illumination and color.

## A.23 Color Stimulus

Represents the psychophysical specifications of the light that causes the sensation of color in our eyes when observing a given surface. The perceived color is represented by a tuple or vector of (typically) three values, known as tristimulus values.

## A.24 Tristimulus Values

The amount of each of the three reference stimuli, under a given representation system, that needs to be in the mixture to produce a given color stimulus. Depending on the model, the tristimulus values may be more than three, but the same word is used by extension.

## A.25 Color response

The color response – or color output – of an output device describes the set of colors (color gamut) that can be reproduced by a given device.

## A.26 Transfer function

Mathematical representation that describes the response of a device, that is, the relation between the device's input and output signal.

## A.27 Calibration

The process of modifying the response of a device to meet a given target.

## A.28 Illuminant

The source of visible light that generates the sensation of colors after bouncing on an object. An illuminant is generally specified by its SPD. There is a set of known theoretical illuminants, each one with a name and a known predefined SPD; the illuminants in this group are called standard illuminants.

*Language disambiguation:* Be aware that the french word "éclairage" refers to any illuminant in general, whereas the french word "illuminant", unlike the english one, refers only to a standard illuminant.

## A.29 Reference white

These are the coordinates of the white color used as a reference. It is chosen so that it is close the basic white stimulus, also known as the equi-energy white, established by the CIE at coordinates (0.33,0.33). The tristimulus values of the reference white are considered the tristimulus values of a perfect reflecting diffuser under the same illuminant as the rest of the objects under consideration.

## A.30 Metamerism

The phenomenon under which two color stimuli, with different spectral distributions, match (are perceived as being the same color) under specified observation conditions. The two color stimuli are called metamers.

## A.31 Viewing conditions

The viewing conditions are the conditions under which a certain scene is observed. These conditions include the viewing angle with respect to the surface normal, the illuminant under which the scene is observed and several other factors that may influence the aspect of the perceived colors.

## A.32 Vision adaptation / Color constancy

When the viewing illumination varies, the human visual system has the ability to adapt to the changing conditions without being greatly influenced by the change. When the level of the ambient illumination changes, we talk about Luminance adaptation; when the ambient illumination changes its color, we talk about chromatic adaptation. Because of these two mechanisms, we can perceive the color of objects without many changes under different illuminants (this is known as color constancy).

## A.33 Color Matching Functions (CMF)

The color matching functions are three functions, $\hat{x}$, $\hat{y}$, and $\hat{z}$, that represent the chromatic sensitivity alongside the visual spectrum of wavelengths, of a reference human observer (the standard observer) for the three types of color-sensitive cones in the human retina. They were computed by the CIE (International Commission on Illumination) as the averaged response of a series of real observers, obtaining functions measured at 2° (for viewing angles less than 2° wide) and 10° (for and viewing angles larger than 4°). The averaged set of responses obtained for each viewing angle are referred to as the *CIE 1931 Standard Observer* and the *CIE 1964* 10° *Standard Observer*, respectively. Alternatively, an additional set of CMFs was defined for the CIE 1931 RGB color space, which are expressed as $\hat{r}$, $\hat{g}$, and $\hat{b}$. See Section **??** for more information.

## A.34 Color Space

The mathematical space that defines the interval (or locus) for the possible tristimulus values that specify a given color stimulus.

## A.35 Color (Appearance) Model

As color vision models get more complex, it becomes necessary to define colors in a better way. Sometimes it is not enough to represent a color by its set of tristimulus values, so color spaces must then be extended, defining equations that describe how color stimuli relate with the appearance attributes of the color —such as hue, saturation, lightness, brightness, etc.— as well as taking other factors into account —like the observer's adaptation to the viewing conditions, or the surround and ambient light— to predict the color appearance. This is known as a Color Model, or Color Appearance Model.

## A.36 Color calibration target

A color calibration target, or color calibration chart, is a physical swatchbook-like cardboard chart formed by a set of color fields arranged in a grid. The number of color fields depends on the size of the gamut of the device being calibrated, with different standard calibration targets for different

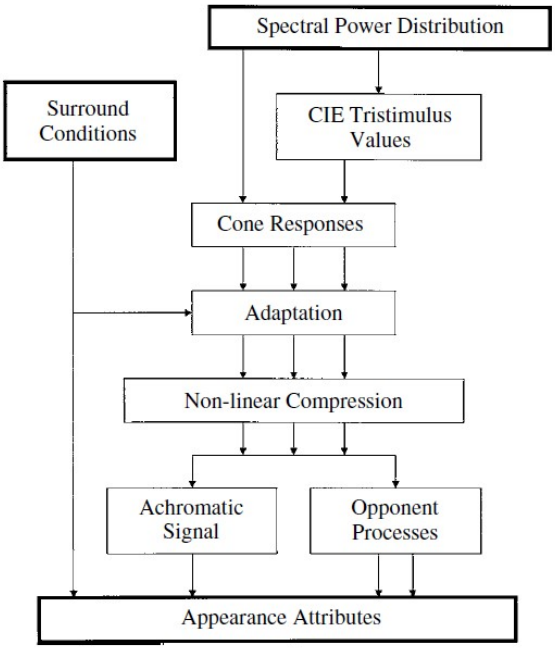



Figure A.2: Flowchart of a typical color appearance model.[34]

devices. A typical color calibration target is the Macbeth ColorChecker [128], or Macbeth chart, which consists of 24 color fields arranged in 4 rows and 6 columns (Figure A.3).

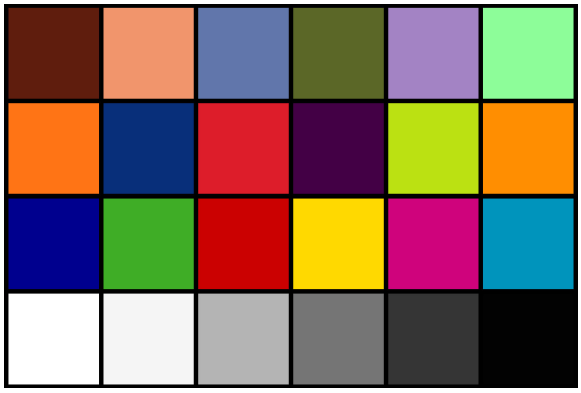


Figure A.3: Macbeth color calibration target.

Appendix B

# Appendix 2: Scientific production

## Contents


B.1 International conference articles . . . . . 115
B.2 Oral presentations . . . . . 116
B.3 Poster presentations . . . . . 116


The research work performed throughout the duration of this PhD thesis have been presented to the scientific community in the following forms of posters, conference articles, and oral presentations.

## B.1 International conference articles

- Physically based image synthesis of materials: a methodology towards the visual comparison of physical vs. virtual samples.
  - Authors:
    * Victor Medina (Mines ParisTech, PSA Peugeot-Citröen)
    * Dominique Lafon-Pham (C2MA mines d'Alès)
    * Alexis Paljic (Mines ParisTech)
    * Emmanuelle Diaz (PSA Peugeot-Citröen)
  - Publisher: Institute of Electrical and Electronics Engineers (IEEE)
  - Format: Proceedings
  - Conference: IEEE Colour and Visual Computing Symposium (CVCS-2015)
  - Dates: 25-26 August 2015
  - Location: Gjøvik (Norway)
  - Publication URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7274878
- A study of image exposure for the stereoscopic visualization of sparkling materials.
  - Authors:
    * Victor Medina (Mines ParisTech, PSA Peugeot-Citröen)
    * Alexis Paljic (Mines ParisTech)
    * Dominique Lafon-Pham (C2MA mines d'Alès)
  - Publisher: International Society for Optics and Photonics (SPIE)
  - Format: Proceedings
  - Conference: IS&T/STIE Electronic Imaging Conference 2015
  - Dates: 08-12 February 2015
  - Location: San Francisco (USA)
  - Publication URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=2109943

## B.2 Oral presentations

- Physically based image synthesis of materials: a methodology towards the visual comparison of physical vs. virtual samples.
  - Event: IEEE Colour and Visual Computing Symposium (CVCS-2015)
  - Dates: 25-26 August 2015
  - Location: Gjøvik (Norway)
- A study of image exposure for the stereoscopic visualization of sparkling materials.
  - Event: IS&T/STIE Electronic Imaging Conference 2015
  - Dates: 08-12 February 2015
  - Location: San Francisco (USA)
- Visuo-perceptual validation methods for physically-based image synthesis.
  - Event: Virtual reality international summer school 2014
  - Dates: 25-29 August 2014
  - Location: Biarritz (France)

## B.3 Poster presentations

- 5th European Workshop on Visual Information Processing (EUVIP)
  - Dates: 9-12 December 2014
  - Location: Paris (France)
- Virtual reality international summer school 2014.
  - Dates: 25-29 August 2014
  - Location: Biarritz (France)
- PRISM 2 workshop: The science of light and shade
  - Dates: 8-11 October 2013
  - Location: Bordeaux (France)

Appendix C

# Appendix 3: Additional data

## C.1 Color set used for the display characterization.

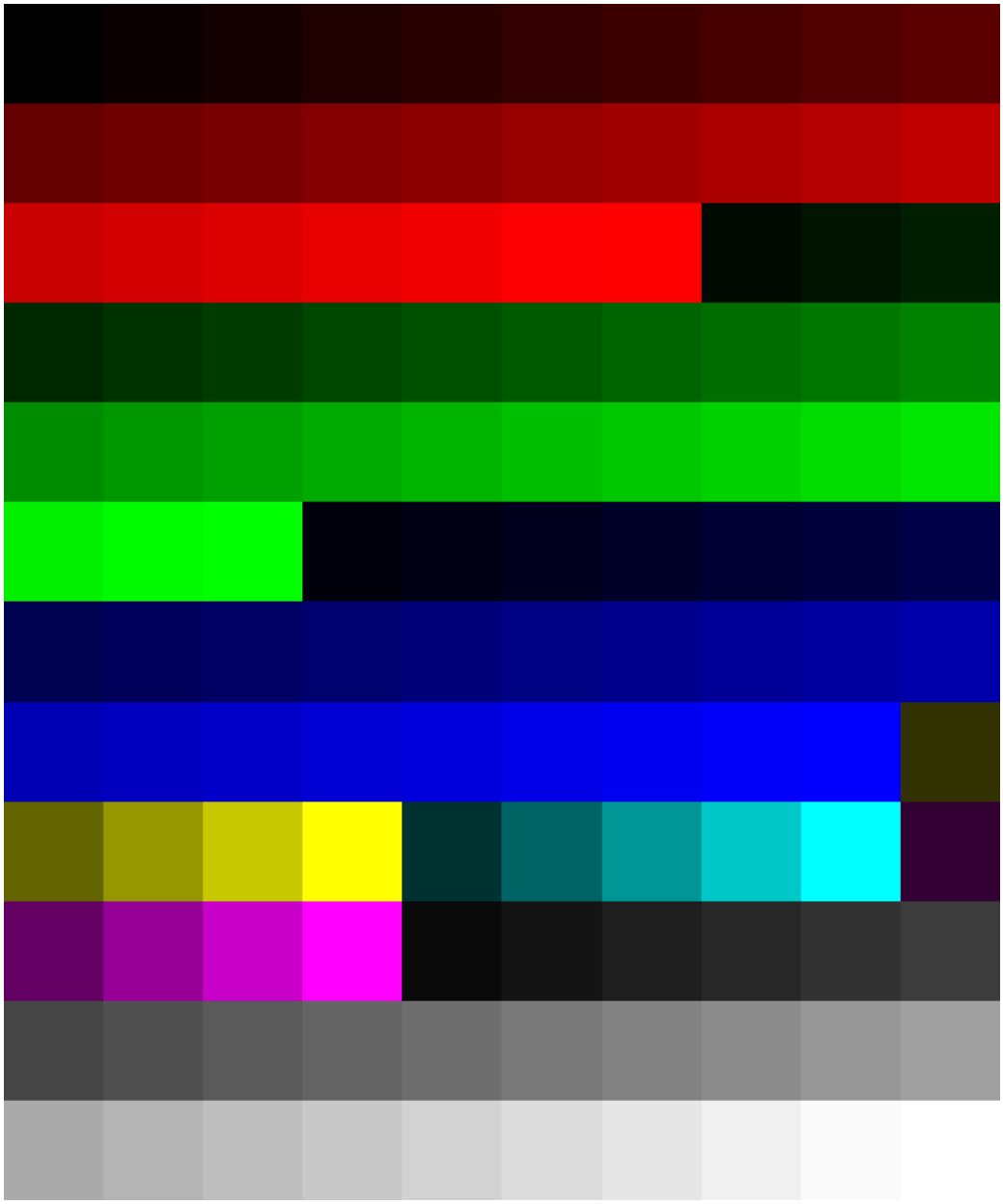


Figure C.1: Characterization color samples.

| Color | R | G | B |
|---|---|---|---|
| black | 0 | 0 | 0 |
| red10 | 10 | 0 | 0 |
| red20 | 20 | 0 | 0 |
| red30 | 30 | 0 | 0 |
| red40 | 40 | 0 | 0 |
| red50 | 50 | 0 | 0 |
| red60 | 60 | 0 | 0 |
| red70 | 70 | 0 | 0 |
| red80 | 80 | 0 | 0 |
| red90 | 90 | 0 | 0 |
| red100 | 100 | 0 | 0 |
| red110 | 110 | 0 | 0 |
| red120 | 120 | 0 | 0 |
| red130 | 130 | 0 | 0 |
| red140 | 140 | 0 | 0 |
| red150 | 150 | 0 | 0 |
| red160 | 160 | 0 | 0 |
| red170 | 170 | 0 | 0 |
| red180 | 180 | 0 | 0 |
| red190 | 190 | 0 | 0 |
| red200 | 200 | 0 | 0 |
| red210 | 210 | 0 | 0 |
| red220 | 220 | 0 | 0 |
| red230 | 230 | 0 | 0 |
| red240 | 240 | 0 | 0 |
| red250 | 250 | 0 | 0 |
| red255 | 255 | 0 | 0 |
| green10 | 0 | 10 | 0 |
| green20 | 0 | 20 | 0 |
| green30 | 0 | 30 | 0 |
| green40 | 0 | 40 | 0 |
| green50 | 0 | 50 | 0 |
| green60 | 0 | 60 | 0 |
| green70 | 0 | 70 | 0 |
| green80 | 0 | 80 | 0 |
| green90 | 0 | 90 | 0 |
| green100 | 0 | 100 | 0 |
| green110 | 0 | 110 | 0 |
| green120 | 0 | 120 | 0 |
| green130 | 0 | 130 | 0 |

| Color | R | G | B |
|---|---|---|---|
| green140 | 0 | 140 | 0 |
| green150 | 0 | 150 | 0 |
| green160 | 0 | 160 | 0 |
| green170 | 0 | 170 | 0 |
| green180 | 0 | 180 | 0 |
| green190 | 0 | 190 | 0 |
| green200 | 0 | 200 | 0 |
| green210 | 0 | 210 | 0 |
| green220 | 0 | 220 | 0 |
| green230 | 0 | 230 | 0 |
| green240 | 0 | 240 | 0 |
| green250 | 0 | 250 | 0 |
| green255 | 0 | 255 | 0 |
| blue10 | 0 | 0 | 10 |
| blue20 | 0 | 0 | 20 |
| blue30 | 0 | 0 | 30 |
| blue40 | 0 | 0 | 40 |
| blue50 | 0 | 0 | 50 |
| blue60 | 0 | 0 | 60 |
| blue70 | 0 | 0 | 70 |
| blue80 | 0 | 0 | 80 |
| blue90 | 0 | 0 | 90 |
| blue100 | 0 | 0 | 100 |
| blue110 | 0 | 0 | 110 |
| blue120 | 0 | 0 | 120 |
| blue130 | 0 | 0 | 130 |
| blue140 | 0 | 0 | 140 |
| blue150 | 0 | 0 | 150 |
| blue160 | 0 | 0 | 160 |
| blue170 | 0 | 0 | 170 |
| blue180 | 0 | 0 | 180 |
| blue190 | 0 | 0 | 190 |
| blue200 | 0 | 0 | 200 |
| blue210 | 0 | 0 | 210 |
| blue220 | 0 | 0 | 220 |
| blue230 | 0 | 0 | 230 |
| blue240 | 0 | 0 | 240 |
| blue250 | 0 | 0 | 250 |
| blue255 | 0 | 0 | 255 |
| yellow50 | 50 | 50 | 0 |

| Color | R | G | B |
|---|---|---|---|
| yellow100 | 100 | 100 | 0 |
| yellow150 | 150 | 150 | 0 |
| yellow200 | 200 | 200 | 0 |
| yellow255 | 255 | 255 | 0 |
| cyan50 | 0 | 50 | 50 |
| cyan100 | 0 | 100 | 100 |
| cyan150 | 0 | 150 | 150 |
| cyan200 | 0 | 200 | 200 |
| cyan255 | 0 | 255 | 255 |
| magenta50 | 50 | 0 | 50 |
| magenta100 | 100 | 0 | 100 |
| magenta150 | 150 | 0 | 150 |
| magenta200 | 200 | 0 | 200 |
| magenta255 | 255 | 0 | 255 |
| gray10 | 10 | 10 | 10 |
| gray20 | 20 | 20 | 20 |
| gray30 | 30 | 30 | 30 |
| gray40 | 40 | 40 | 40 |
| gray50 | 50 | 50 | 50 |
| gray60 | 60 | 60 | 60 |
| gray70 | 70 | 70 | 70 |
| gray80 | 80 | 80 | 80 |
| gray90 | 90 | 90 | 90 |
| gray100 | 100 | 100 | 100 |
| gray110 | 110 | 110 | 110 |
| gray120 | 120 | 120 | 120 |
| gray130 | 130 | 130 | 130 |
| gray140 | 140 | 140 | 140 |
| gray150 | 150 | 150 | 150 |
| gray160 | 160 | 160 | 160 |
| gray170 | 170 | 170 | 170 |
| gray180 | 180 | 180 | 180 |
| gray190 | 190 | 190 | 190 |
| gray200 | 200 | 200 | 200 |
| gray210 | 210 | 210 | 210 |
| gray220 | 220 | 220 | 220 |
| gray230 | 230 | 230 | 230 |
| gray240 | 240 | 240 | 240 |
| gray250 | 250 | 250 | 250 |
| white | 255 | 255 | 255 |

Table C.1: RGB tristimulus values of the 120 characterization color samples shown in Figure C.1.

# Index


absorption, 6, 7, 14, 109
  absorption index, 109
  spectral absorption, 109
accommodation, 16, 31, 35, 36
additive color synthesis, 6
alternate frame-sequencing, 36
autostereoscopy, 36
axial chromatic aberration, 30

backlight, 8, 35, 36, 42, 45
barium white, 6
binocular disparity, *see* parallax
binocular summation, 31
binocular vision, 30
blind spot, 19
Bradley-Terry score, 83, 84
BRDF, 7, 109
brightness, 27, 34, 35, 39, 40, 48, 62, 81, 83–85, 93, 109, 110, 113

calibration, 112
cathodic raytube display (CRT), 34, 37, 39, 41, 46
center-surround inhibition, 17
characterization, 40, 41, 53
  camera characterization, 63–68, 73, 75
  camera characterization matrix, 66, 67
  CRT characterization model, 39, 46
  display characterization, 9, 37–42, 48, 50, 73
  display characterization model, 34, 38, 39, 41, 49, 50, 74, 75
  GGO model, 48–50
  GOG model, 48–50, 53, 55
  GOGO model, 48–50, 52
  material characterization, 62
  Simple gamma model, 49
choroid, 16
chromatic adaptation, 25
  adaptive color shift, 25
  color shift, 25
  corresponding color, 25
  illuminant color shift, 25
chromatic adaptation matrix (CAT), 27
chromaticity, 93
chrominance, 96
CIE, 111
CIE 1931 RGB color space, 24, 113
CIE 1931 standard observer, 24, 89
CIE 1931 XYZ color space, 9, 24, 27, 75
CIE 1964 standard observer, 24
CIE 1976 L*a*b* color model, 96
CIE DeltaE 1994 color difference, 96
CIE DeltaE 2000 color difference, 96, 97
CIE xyY color space, 93, 94
CIELAB, *see* CIE 1976 L*a*b* color model
color, 13, 24, 61, 63, 93, 111
  color flop, 62
  object color, 13
color appearance model (CAM), 27
color calibration target, 41, 67, 68, 113, 114
color constancy, 25, 113
color depth, 35
color filter array (CFA), 64, 65, 73, 75
color gamut, 34, 37, 112
color matching function (CMF), 23–25, 65, 89, 113
color model, 24, 113
color space, 24, 73, 113
colorimeter, *see* tristimulus colorimeter
colorimetry, 23
cones, 18, 19, 24, 25, 27
  blue cones, *see* S cones
  green cones, *see* M cones
  L cones, 19
  M cones, 19
  red cones, *see* L cones
  S cones, 19
contrast, 8, 17, 27, 29, 40, 71, 74, 77, 80, 83–85, 110
  contrast sensitivity function (CSF), 29, 31
  simultaneous contrast, 29, 30
  Weber contrast, 29
convergence, 31, 35
cornea, 16

depth perception, 30, 36
  binocular cues, 31, 36
  monocular cues, 31
  true 3D, 31
diffraction, 14, 63, 64
diffuser, 6
digital to analog converter (DAC), 41, 50

digital video input (DVI), 41
direction-multiplex, 36
dispersion, 62, 64
dithering, 8, 35
dynamic range, 9, 33, 61, 63, 64, 74, 83, 84

electromagnetic spectrum, 15
exposure, 65, 67, 68, 73–77, 81, 83–85
    exposure time, 64, 68, 73–76, 83, 84
    overexposure, 74, 82
    underexposure, 82
eye lens, 16

field of view, 63, 71
flickering, 35
fovea, 16, 19, 24
foveola, 16, 19
frame rate control (FRC), 35

gamma function, 46, 48
Grasmmann laws, 23

holography, 37

illuminance, 67, 68, 75, 108
illuminant, 62, 112
image demosaicing, 73
image linearization, 73
image persistence, 35
in plane switching LCD (IPS-LCD), 34, 35
infrared light, 15
interpupillary distance, 30
iris (eye), 16
irradiance, 6, 108, 111

lambertian reflectance, 6, 7
lateral geniculate nucleus (LGN), 16
lateral inhibition, 17
light, 13, 14
light measurement, 6
    incident light measurement, 6
    reflection light measurement, 6
light particle, 14
light wave, 14
light-source color, 6
lightness, 62, 96, 109
    lightness flop, 62
liquid crystal display (LCD), 34, 35, 37, 39, 42, 46
LMS color space, 9, 24, 27, 75
logMAR chart, 27
luminance, 6, 22, 23, 33, 42, 48, 67, 71, 75, 80, 82, 85, 110
    luminance adaptation, 25, 71
luminous flux, 108
luminous power, *see* luminous flux
luminous reflectance, 110

Macbeth color chart, *see* color calibration target
mesopic vision, 18, 25
metal flake, 10, 61–64, 67, 73, 75–77, 85
metal particle, *see* metal flake
metamerism, 65, 67, 88, 112

object color, 6
opponent color theory, 21
opponent color theory of vision, 21
optic nerve, 16, 19
optical axis, 16

parallax, 30, 31, 35–37, 71
    motion parallax, 31, 35, 36
    positive parallax, 71, 72
    zero parallax, 71, 72
parallax barrier, 37
particle theory of light, 14
photometry, 107
photon, 14
photopic vision, 18, 25, 27
photoreceptors, 13, 15–19, 21–23, 29, 63, 65, 75
physically-based rendering (PBR), 4, 5, 8, 87
pixel independence, 41
pixelation, 28, 64
plausible rendering, 4
point-spread function (PSF), 49, 63, 64
predictive rendering, 4, 8
primary color, 23, 24, 34, 39, 41, 52
Purkinje effect, 19

radiance, 6, 61, 63, 66, 74, 108
radiant flux, 108
radiant power, *see* radiant flux
radiometry, 107
re-imaging display, 37
real scene, 107
realism, 4
    functional realism, 4, 5
    perceptual realism, 5, 8, 37, 38, 73–75, 80–82, 87, 103, 104
    photo-realism, 4, 5
    physical realism, 4, 5
    plausible realism , *see* plausible rendering
    predictive realism , *see* predictive rendering
    visually-rich realism , *see* visually-rich rendering
receptive field, 17
reference white, 112

reflectance
  reflectance factor, 109
  reflectance index, 109
  spectral reflectance, 6, 109, 111
reflection, 6, 7, 62, 67, 68, 73, 74, 109
  diffuse reflection, 6
  specular reflection, 6, 62–64, 67, 73, 75
refraction, 6, 109
  refraction index, 109
  spectral refraction, 109
refraction index, 6
refractive index, 16
rendering, 2, 4, 8
retina, 6, 9, 16–19, 22, 35, 61, 63, 65, 75
retinal cells
  amacrine, 17, 18
  bipolar, 17, 18
  ganglion, 17, 18
  horizontal, 17
  On-center/Off-center ganglion, 18
  ON/OFF bipolar, 17
RGB color space, *see* CIE 1931 RGB color space
RIT-DuPont dataset, 96, 97
rods, 18, 19, 27

saccades, 31
  microsaccades, 31
  pursuit movements, 31
schlera, 16
scotopic vision, 18, 25, 27
shutter glasses, *see* stereoscopy>shutter stereocopy
shutter speed, 65, 67, 68, 74, 75, 77
simulation, 61–63, 75, 107
Snell-Descartes law, 109
Snellen chart, 27
Snellen fraction, 27, 28
solid angle, 108
sparkle, 1, 10, 61–65, 68, 74–77, 83, 85
spectral power distribution (SPD), 6, 7, 23, 39, 49, 108
spectrocolorimeter, 6, 111
spectroradiometer, 6, 42, 49, 111
stage theory, 22
Standard 10° Observer , *see* CIE 1964 standard observer
Standard 2° Observer, *see* CIE 1931 standard observer
stereopsis, 30
stereoscopy, 36, 37, 61, 68, 70, 71, 76, 77, 80, 83
  active stereoscopy, 36
  anaglyph, 36
  off-axis frustum method, 71
  passive stereoscopy, 36
  polarization, 36
  shutter stereoscopy, 36, 37, 40, 41, 55, 56, 76
  stereo-pair stereoscopy, 36
  stereoscopic glasses, 36
  toe-in method, 71
stripping, 35
subtractive color synthesis, 6, 7

temporal integration, 31
tone reproduction, 5, 9, 61, 75, 80, 84
  tone mapping operator, 33, 83
transfer function, 33, 38, 48, 75, 112
transmittance, 7, 109, 111
  spectral transmittance, 6, 109
  transmittance index, 109
trichromatic theory of vision, 8, 21
tristimulus colorimeter, 6, 111
tristimulus values, 6, 23–25, 27, 65, 111–113, 118
twisted nematics LCD (TN-LCD), 34, 35, 37

ultraviolet light, 15

V1 brain area, *see* visual cortex
virtual scene, 107
visible light, 14, 15
visible spectrum, 15
vision
  color vision, 18, 22
  normal vision, 28
visual acuity, 27, 28, 30, 62, 71
  spatial acuitity, 29
visual axis, 16, 19
visual cortex, 16, 22
  primary (V1), 16
visual field, 16, 18, 19, 24, 28, 30
visual response, 61, 63, 73, 75
visually-rich rendering, 4
von Kries
  model, 27
  transform, 27

wave theory of light, 14, 15
wave-particle duality, 14
Weber–Fechner law, 29
white balance, 26

XYZ color space, *see* CIE 1931 XYZ color space

# Bibliography


[1] S. Chen and A. Beghdadi, "Natural enhancement of color image," *Journal on Image and Video Processing*, vol. 2010, p. 2, 2010.

(Cited in page 1.)

[2] E. Couka, *Modélisation des propriétés optiques de peintures par microstructures aléatoires et calculs numériques FFT*. PhD thesis, École nationale superieure des mines de Paris, November 2015.

(Cited in page 2.)

[3] M. Ben-Achour, *Synthèse et caractérisation multi-échelle de nanoparticules pour revêtements automobiles*. PhD thesis, École nationale superieure des mines de Paris, December 2015.

(Cited in pages 2, 61 and 62.)

[4] F. da Graça, *Validations perceptives stéréoscopiques d'Images physico réalistes de matériaux basées sur des modèles de microstructure*. PhD thesis, École nationale superieure des mines de Paris, December 2015.

(Cited in pages 2, 9 and 78.)

[5] J. A. Ferwerda, "Three varieties of realism in computer graphics," in *Electronic Imaging 2003*, pp. 290–297, International Society for Optics and Photonics, 2003.

(Cited in pages 4, 5 and 103.)

[6] P. Shirley, R. K. Morley, P.-P. Sloan, and C. Wyman, "Basics of physically-based rendering," in *SIGGRAPH Asia 2012 Courses*, p. 2, ACM, 2012.

(Cited in page 4.)

[7] A. Chalmers, K. Debattista, and L. P. dos Santos, "Selective rendering: computing only what you see," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pp. 9–18, ACM, 2006.

(Cited in page 4.)

[8] Reader's Digest, *The Reader's Digest Complete Do-It Yourself Manual*. Reader's Digest Association, 1973.

(Cited in page 5.)

[9] J. M. Henderson and A. Hollingworth, "High-level scene perception," *Annual review of psychology*, vol. 50, no. 1, pp. 243–271, 1999.

(Cited in page 5.)

[10] E. Balcetis and D. Dunning, "See what you want to see: motivational influences on visual perception.," *Journal of personality and social psychology*, vol. 91, no. 4, p. 612, 2006.

(Cited in pages 5, 31 and 60.)

[11] J. Schanda, G. Eppeldauer, and G. Sauter, "Tristimulus color measurement of self-luminous sources," *Colorimetry: Understanding the CIE system*, pp. 135–157, 2007.

(Cited in pages 6 and 111.)

[12] C. Wynn, "An introduction to brdf-based lighting," *Nvidia Corporation*, 2000.

(Cited in pages 7 and 109.)

[13] E. Peli, L. Arend, and A. T. Labianca, "Contrast perception across changes in luminance and spatial frequency," *JOSA A*, vol. 13, no. 10, pp. 1953–1959, 1996.

(Cited in page 8.)

[14] K.-H. Bäuml *et al.*, "Simultaneous color constancy: how surface color perception varies with the illuminant," *Vision research*, vol. 39, no. 8, pp. 1531–1550, 1999.

(Cited in page 8.)

[15] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," *Handbook of image and video processing*, pp. 669–684, 2000.

(Cited in pages 8 and 80.)

[16] L. T. Maloney and D. H. Brainard, "Color and material perception: Achievements and challenges," *Journal of Vision*, vol. 10, no. 9, 2010.

(Cited in page 8.)

[17] R. W. Fleming, "Visual perception of materials and their properties," *Vision research*, vol. 94, pp. 62–75, 2014.

(Cited in page 8.)

[18] P. Irawan, J. A. Ferwerda, and S. R. Marschner, "Perceptually based tone mapping of high dynamic range image streams.," in *Rendering Techniques*, pp. 231–242, 2005.

(Cited in page 9.)

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

(Cited in page 9.)

[20] F. da Graça, A. Paljic, D. Lafon-Pham, and P. Callet, "Stereoscopy for visual simulation of materials of complex appearance," in *IS&T/SPIE Electronic Imaging*, vol. 90110X, SPIE, 2014.

(Cited in pages 9, 62 and 78.)

[21] A. Sawadogo, *Outils d'évaluation de la qualité d'un paramétrage de propriétés visuelles: cas des textures couleur*. PhD thesis, Université de Pau et des Pays de l'Adour, 2009.

(Cited in page 13.)

[22] A. McNamara, "Visual perception in realistic image synthesis," in *Computer Graphics Forum*, vol. 20, pp. 211–224, Wiley Online Library, 2001.

(Cited in pages 14, 27, 29 and 30.)

[23] R. McCann, "Visible light- what the human eye sees." Photovoltaic Lighting Group, February 2013. Accessed: 2015-10-30.

(Cited in page 15.)

[24] G. Wyszecki and W. S. Stiles, *Color science*, vol. 8. Wiley New York, 1982.

(Cited in page 16.)

[25] R. Sève, *Science de la couleur - Aspects physiques et perceptifs*. Chalagam Édition, 1 ed., 2009.

(Cited in pages 16, 17, 18, 19 and 23.)

[26] E. M. Valero and R. Huertas, "Color science lecture notes." 2010.

(Cited in pages 16, 17, 20, 21 and 22.)

[27] D. Purves, E. M. Brannon, R. Cabeza, S. A. Huettel, K. S. LaBar, M. L. Platt, and M. G. Woldorff, *Principles of cognitive neuroscience*, vol. 83. Sinauer Associates Sunderland, MA, 2008.

(Cited in page 17.)

[28] Wikipedia, "Receptive field — wikipedia, the free encyclopedia," 2015. [Online; accessed 5-November-2015].

(Cited in page 19.)

[29] Wikipedia, "Photoreceptor cell — wikipedia, the free encyclopedia," 2015. [Online; accessed 10-November-2015].

(Cited in page 20.)

[30] M. D. Fairchild, *Color appearance models*. John Wiley & Sons, 2nd ed., 2005.

(Cited in pages 19, 21 and 22.)

[31] R. M. Boynton, "A system of photometry and colorimetry based on cone excitations," *Color Research & Application*, vol. 11, no. 4, pp. 244–252, 1986.

(Cited in page 22.)

[32] A. Ford and A. Roberts, "Colour space conversions," *Westminster University, London*, vol. 1998, pp. 1–31, 1998.

(Cited in page 23.)

[33] D. Malacara, *Color vision and colorimetry: theory and applications*. SPIE Press, Bellingham, WA, 2 ed., 2011.

(Cited in page 23.)

[34] N. Ohta and A. Robertson, *Colorimetry: fundamentals and applications*. Wiley, 2006.

(Cited in pages 23, 24, 25, 26 and 114.)

[35] Wikipedia, "Cie 1931 color space — wikipedia, the free encyclopedia," 2015. [Online; accessed 22-November-2015].

(Cited in page 25.)

[36] 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.

(Cited in page 24.)

[37] M. H. Brill, "How the cie 1931 color-matching functions were derived from wright-guild data (erratum)," *Color Research & Application*, vol. 23, no. 4, pp. 259–259, 1998.

(Cited in page 24.)

[38] M. Kalloniatis and C. Luu, "Light and dark adaptation," *Webvision: The Organization of the Retina and Visual System H. Kolb, E. Fernandez, and R. Nelson, Salt Lake City (UT): University of Utah Health Sciences Center*, 2007. [Accessed on February 2016].

(Cited in page 26.)

[39] W. Strouse-Watt, "How visual acuity is measured," *Retrieved June*, vol. 16, p. 2004, 2004.

(Cited in page 27.)

[40] M. Kalloniatis and C. Luu, "Visual acuity," *Webvision: The Organization of the Retina and Visual System H. Kolb, E. Fernandez, and R. Nelson, Salt Lake City (UT): University of Utah Health Sciences Center*, 2007. [Accessed on February 2016].

(Cited in pages 28 and 29.)

[41] Wikipedia, "Contrast (vision) — wikipedia, the free encyclopedia," 2015. [Online; accessed 22-November-2015].

(Cited in page 29.)

[42] G. W. Meyer and A. Liu, "Color spatial acuity control of a screen subdivision image synthesis algorithm," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pp. 387–399, International Society for Optics and Photonics, 1992.

(Cited in pages 29 and 30.)

[43] W. K. Middleton and M. C. Holmes, "The apparent colors of surfaces of small subtense–a preliminary report," *Josa*, vol. 39, no. 7, pp. 582–591, 1949.

(Cited in page 30.)

[44] O. H. Schade, "On the quality of color-television images and the perception of color detail," *Journal of the SMPTE*, vol. 67, no. 12, pp. 801–819, 1958.

(Cited in page 30.)

[45] K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings.," *The Journal of Physiology*, vol. 359, p. 381, 1985.

(Cited in page 30.)

[46] P. J. Savino and H. V. Danesh-Meyer, *Color Atlas and Synopsis of Clinical Ophthalmology–Wills Eye Institute–Neuro-Ophthalmology*. Lippincott Williams & Wilkins, 2 ed., 2012.

(Cited in page 30.)

[47] G. Dagnelie, *Visual prosthetics: physiology, bioengineering, rehabilitation*. Springer Science & Business Media, 2011.

(Cited in page 30.)

[48] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, and J. T. McConville, "1988 anthropometric survey of us army personnel: methods and summary statistics," tech. rep., DTIC Document, 1989.

(Cited in page 30.)

[49] D. Stidwill and R. Fletcher, *Normal binocular vision: Theory, investigation and practical aspects*. John Wiley & Sons, 2010.

(Cited in page 31.)

[50] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," in *Advances in Neural Information Processing Systems*, pp. 1161–1168, 2005.

(Cited in page 31.)

[51] D. F. McAllister, *Stereo and 3-D Display Technologies*. John Wiley & Sons, Inc., 2002.

(Cited in page 31.)

[52] M. Kalloniatis and C. Luu, "Temporal resolution," *Webvision: The Organization of the Retina and Visual System H. Kolb, E. Fernandez, and R. Nelson, Salt Lake City (UT): University of Utah Health Sciences Center*, vol. 5, 2007. [Accessed on February 2016].

(Cited in page 31.)

[53] J.-F. Le Gargasson, "Ii-l'œil et la vision," *Œil et physiologie de la vision*, 2010.

(Cited in page 31.)

[54] B. Heider, L. Spillmann, and E. Peterhans, "Stereoscopic illusory contours–cortical neuron responses and human perception," *Journal of Cognitive Neuroscience*, vol. 14, no. 7, pp. 1018–1029, 2002.

(Cited in page 32.)

[55] T.-N. Lin, J. Shu, and J. W. Swic, "Frame rate modulation for liquid crystal display (lcd)," May 16 2000. US Patent 6,064,359.

(Cited in pages 34 and 35.)

[56] G. Sharma, "Lcds versus crts-color-calibration and gamut considerations," *Proceedings of the IEEE*, vol. 90, pp. 605–622, Apr 2002.

(Cited in page 34.)

[57] D. H. Brainard, D. G. Pelli, and T. Robson, "Display characterization," *Encyclopedia of imaging science and technology*, 2002.

(Cited in pages 34, 37 and 39.)

[58] T. Elze and T. G. Tanner, "Temporal properties of liquid crystal displays: Implications for vision science experiments," 2012.

(Cited in pages 34, 37 and 38.)

[59] S. Baker, "Panel technologies." TNT Central, December 2013.

(Cited in pages 34 and 35.)

[60] H. B. Gerhard, "Twisted nematic liquid crystal displays (tn-lcds), an invention from basel with global effects," 2005.

(Cited in page 35.)

[61] G. W. Gray and S. M. Kelly, "Liquid crystals for twisted nematic display devices," *Journal of Materials Chemistry*, vol. 9, no. 9, pp. 2037–2050, 1999.

(Cited in page 35.)

[62] J. Klein, M. Zlatkova, J. Lauritzen, and B. Pierscionek, "Photometric and colorimetric measurements of crt and tft monitors for vision research," *Journal of Modern Optics*, vol. 60, no. 14, pp. 1159–1166, 2013.

(Cited in page 35.)

[63] Japan Display Inc., "Ips lcd," 2015. [Online; accessed 27-November-2015].

(Cited in page 35.)

[64] LG Display, "Ips technology," 2015. [Online; accessed 27-November-2015].

(Cited in page 35.)

[65] J. Lei and X. Zou, "Dynamic dithering algorithm and frame rate control technique for liquid crystal display controller," in *ASIC, 2003. Proceedings. 5th International Conference on*, vol. 2, pp. 740–743 Vol.2, Oct 2003.

(Cited in page 35.)

[66] S.-W. Lee and H. Nam, "A new dithering algorithm for higher image quality of liquid crystal displays," *Consumer Electronics, IEEE Transactions on*, vol. 55, pp. 2134–2138, November 2009.

(Cited in page 35.)

[67] O. Artamonov, "X-bit's guide: Contemporary lcd monitor parameters and characteristics," 2004.

(Cited in page 35.)

[68] N. A. Dodgson, "Autostereoscopic 3d displays," *Computer*, no. 8, pp. 31–36, 2005.

(Cited in page 36.)

[69] S. Pastoor and M. Wöpking, "3-d displays: A review of current technologies," *Displays*, vol. 17, no. 2, pp. 100–110, 1997.

(Cited in page 36.)

[70] M. Halle, "Autostereoscopic displays and computer graphics," in *ACM SIGGRAPH 2005 Courses*, p. 104, ACM, 2005.

(Cited in page 37.)

[71] T. Kozacki and M. Chlipala, "Color holographic display with white light led source and single phase only slm," *Optics Express*, vol. 24, no. 3, pp. 2189–2199, 2016.

(Cited in page 37.)

[72] E. Moon, M. Kim, J. Roh, H. Kim, and J. Hahn, "Holographic head-mounted display with rgb light emitting diode light source," *Optics express*, vol. 22, no. 6, pp. 6526–6534, 2014.

(Cited in page 37.)

[73] T. Kozacki, G. Finke, P. Garbat, W. Zaperty, and M. Kujawińska, "Wide angle holographic display system with spatiotemporal multiplexing," *Optics express*, vol. 20, no. 25, pp. 27473–27481, 2012.

(Cited in page 37.)

[74] H. Gao, X. Li, Z. He, Y. Su, and T.-C. Poon, "Real-time holographic display based on a super fast response thin film," in *Journal of Physics: Conference Series*, vol. 415, p. 012052, IOP Publishing, 2013.

(Cited in page 37.)

[75] Y. Sando, D. Barada, and T. Yatagai, "Holographic 3-d display viewable from all horizontal directions by using a single high-speed slm," in *Information Optics (WIO), 2015 14th Workshop on*, pp. 1–2, IEEE, 2015.

(Cited in page 37.)

[76] S. Reichelt, R. Haussler, N. Leister, G. Futterer, H. Stolle, and A. Schwerdtner, "Holographic 3-d displays-electro-holography within the grasp of commercialization," 2010.

(Cited in page 37.)

[77] E. Couka, F. Willot, and D. Jeulin, "A mixed boolean and deposit model for the modeling of metal pigments in paint layers," *Image Analysis & Stereology*, vol. 34, no. 2, pp. 125–134, 2015.

(Cited in page 37.)

[78] E. Couka, F. Willot, D. Jeulin, M. B. Achour, A. Chesnaud, and A. Thorel, "Modeling of the multiscale dispersion of nanoparticles in a hematite coating," *Journal of Nanoscience and Nanotechnology*, vol. 15, no. 5, pp. 3515–3521, 2015.

(Cited in page 37.)

[79] C. Eberle, "Asus vg248qe: A 24-inch, 144 hz gaming monitor under $300." Tom's Hardware, October 2013.

(Cited in pages 37 and 40.)

[80] T. Elze, "Achieving precise display timing in visual neuroscience experiments," *Journal of neuroscience methods*, vol. 191, no. 2, pp. 171–179, 2010.

(Cited in page 38.)

[81] S. Tourancheau, P. Le Callet, and D. Barba, "Image and video quality assessment using lcd: comparisons with crt conditions," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 91, no. 6, pp. 1383–1391, 2008.

(Cited in page 38.)

[82] C. Mantel, J. Korhonen, J. M. Pedersen, S. Bech, E. Nadernejad, N. Burini, and S. Forchhammer, "Modeling the leakage of lcd displays with local backlight for quality assessment," in *IS&T/SPIE Electronic Imaging*, pp. 90140M–90140M, International Society for Optics and Photonics, 2014.

(Cited in page 38.)

[83] H. E. Lagroix, M. R. Yanko, and T. M. Spalek, "Lcds are better: Psychophysical and photometric estimates of the temporal characteristics of crt and lcd monitors," *Attention, Perception, & Psychophysics*, vol. 74, no. 5, pp. 1033–1041, 2012.

(Cited in page 38.)

[84] M. Bass, E. Stryland, D. Willians, and W. Woffe, *Handbook of Optics, vol. 1*. 1995.

(Cited in page 38.)

[85] H. Widdel and D. L. Post, *Color in electronic displays*, vol. 3. Springer, 1992.

(Cited in page 38.)

[86] Z.-L. Lu and B. Dosher, *Visual Psychophysics: From Laboratory to Theory*. MIT Press, 2013.

(Cited in page 41.)

[87] D. G. Pelli, "Pixel independence: Measuring spatial interactions on a crt display," *Spatial vision*, vol. 10, no. 4, pp. 443–446, 1997.

(Cited in page 41.)

[88] J.-B. Thomas, J. Y. Hardeberg, I. Foucherot, and P. Gouton, "The plvc display color characterization model revisited," *Color Research & Application*, vol. 33, no. 6, pp. 449–460, 2008.

(Cited in pages 48 and 49.)

[89] B. Bastani, B. Cressman, and B. Funt, "Calibrated color mapping between lcd and crt displays: A case study," *Color Research & Application*, vol. 30, no. 6, pp. 438–447, 2005.

(Cited in page 48.)

[90] R. Byshko and S. Li, "Characterization of iphone displays: a comparative study," 2012.

(Cited in page 48.)

[91] The Mathworks, Inc., Natick, Massachusetts, United States, *MATLAB version 8.0 (R2012b)*, 2012.

(Cited in pages 50 and 96.)

[92] H.-J. Streitberger, W. Kreis, G. Decher, and J. Schlenoff, *Automotive Paints and Coatings*. Wiley Online Library, 2006.

(Cited in page 61.)

[93] C. McCamy, "Observation and measurement of the appearance of metallic materials. part i. macro appearance," *Color Research & Application*, vol. 21, no. 4, pp. 292–304, 1996.

(Cited in page 62.)

[94] C. McCamy, "Observation and measurement of the appearance of metallic materials. part ii. micro appearance," *Color Research & Application*, vol. 23, no. 6, pp. 362–373, 1998.

(Cited in page 62.)

[95] D. Alleysson and N. Guyader, "Retinal processing: From biology to models and applications," *Biologically Inspired Computer Vision: Fundamentals and Applications*, 2015.

(Cited in page 65.)

[96] S.-W. Jang, E.-S. Kim, S.-H. Lee, and K.-I. Sohng, "Adaptive colorimetric characterization of digital camera with white balance," in *Image Analysis and Recognition*, pp. 712–719, Springer, 2005.

(Cited in page 65.)

[97] G. Hong, M. Luo, and P. Rhodes, "A study of digital camera colorimetric characterization based on polynomial modeling," *Color: Research and applications*, vol. 26, no. 1, pp. 76–84, 2001.

(Cited in pages 65 and 67.)

[98] H. Lin, S. J. Kim, S. Susstrunk, and M. S. Brown, "Revisiting radiometric calibration for color computer vision," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 129–136, IEEE, 2011.

(Cited in page 65.)

[99] A. Alsam and G. Finlayson, "Metamer sets without spectral calibration," *JOSA A*, vol. 24, no. 9, pp. 2505–2512, 2007.

(Cited in page 67.)

[100] N. Joshi, B. Wilburn, V. Vaish, M. L. Levoy, and M. Horowitz, *Automatic color calibration for large camera arrays*. [Department of Computer Science and Engineering], University of California, San Diego, 2005.

(Cited in page 67.)

[101] S. A. Renani, M. Tsukada, and J. Y. Hardeberg, "Compensating for non-uniform screens in projection display systems," in *IS&T/SPIE Electronic Imaging*, pp. 72410F–72410F, International Society for Optics and Photonics, 2009.

(Cited in page 67.)

[102] E. Kirchner and J. Ravi, "Predicting and measuring the perceived texture of car paints," in *Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance*, p. 25, 2012.

(Cited in page 68.)

[103] P. Thompson, "Eyes wide apart: Overestimating interpupillary distance," *PERCEPTION-LONDON-*, vol. 31, no. 6, pp. 651–656, 2002.

(Cited in page 71.)

[104] W. A. IJsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: effects of camera parameters and display duration," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 2, pp. 225–233, 2000.

(Cited in page 71.)

[105] P. Bourke, "Calculating stereo pairs," *Course notes. Published by The Swinburne Centre for Astrophysics and Supercomputing: Australia.*, 1999.

(Cited in pages 71 and 72.)

[106] R. Sumner, "Processing raw images in matlab," 2014.

(Cited in page 73.)

[107] K. Tsukida and M. R. Gupta, "How to analyze paired comparison data," tech. rep., DTIC Document, 2011.

(Cited in page 83.)

[108] J. Mardaljevic, "Validation of a lighting simulation program under real sky conditions," *Lighting research and Technology*, vol. 27, no. 4, pp. 181–188, 1995.

(Cited in page 87.)

[109] F. Drago and K. Myszkowski, "Validation proposal for global illumination and rendering techniques," *Computers & Graphics*, vol. 25, no. 3, pp. 511–518, 2001.

(Cited in page 87.)

[110] D. P. Greenberg, K. E. Torrance, P. Shirley, J. Arvo, E. Lafortune, J. A. Ferwerda, B. Walter, B. Trumbore, S. Pattanaik, and S.-C. Foo, "A framework for realistic image synthesis," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 477–494, ACM Press/Addison-Wesley Publishing Co., 1997.

(Cited in page 87.)

[111] M. Melis, M. Miccoli, and D. Quarta, "Variante full range su fotocamere nikon: hardware e software fanno sistema e misurano la riflettanza spettrale." Nikon eXperience, December 2013. [Accessed on September-2015].

(Cited in page 93.)

[112] S. Upton, "Delta-e: The color difference," *CHROMiX ColorNews*, vol. 17, February 2005.

(Cited in page 94.)

[113] D. A. Kerr, "A metric for chromaticity difference." dougkerr.net, February 2008.

(Cited in page 94.)

[114] M. Melgosa, E. Hita, A. Poza, D. H. Alman, R. S. Berns, *et al.*, "Suprathreshold color-difference ellipsoids for surface colors," *Color Research and Application*, vol. 22, no. 3, pp. 148–155, 1997.

(Cited in pages 96 and 97.)

[115] R. S. Berns and B. Hou, "Rit-dupont supra-threshold color-tolerance individual color-difference pair dataset," *Color Research & Application*, vol. 35, no. 4, pp. 274–283, 2010.

(Cited in pages 96, 97 and 99.)

[116] P. J. Baldevbhai and R. Anand, "Color image segmentation for medical images using l* a* b* color space," *IOSR Journal of Electronics and Communication Engineering (IOSRJECE)*, vol. 1, no. 2, pp. 24–45, 2012.

(Cited in page 96.)

[117] M. R. 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.

(Cited in page 97.)

[118] D. H. Alman, R. S. Berns, G. D. Snyder, and W. A. Larsen, "Performance testing of color-difference metrics using a color tolerance dataset," *Color Research & Application*, vol. 14, no. 3, pp. 139–151, 1989.

(Cited in page 97.)

[119] R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. R. Balonon-Rosen, "Visual determination of suprathreshold color-difference tolerances using probit analysis," *Color Research & Application*, vol. 16, no. 5, pp. 297–316, 1991.

(Cited in page 97.)

[120] G. Snyder, "Visual determination of industrial color-difference tolerances using probit analysis," 1991.

(Cited in page 97.)

[121] S. Shen and R. S. Berns, "Evaluating color difference equation performance incorporating visual uncertainty," *Color Research & Application*, vol. 34, no. 5, pp. 375–390, 2009.

(Cited in page 97.)

[122] K. F. Karner and M. Prantl, "A concept for evaluating the accuracy of computer generated images," in *Proceedings of the Twelfth Spring Conference on Computer Graphics (SCCG'96).[www. cg. tuwien. ac. at/˜ wp/SCCG96-proceedings/papers/Karner. ps. gz]*, Citeseer, 1996.

(Cited in page 99.)

[123] J. Bärz, N. Henrich, and S. Müller, "Validating photometric and colorimetric consistency of physically-based image synthesis," in *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2010, pp. 148–154, Society for Imaging Science and Technology, 2010.

(Cited in page 99.)

[124] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting.* Morgan Kaufmann, 2010.

(Cited in page 103.)

[125] WorkWithColor.com, "Color luminance: perceived brightness, luminance and contrast." Accessed January, 2014., July 2008.

(Cited in page 110.)

[126] N. A. Schellart, *Compendium of Medical Physics, Medical Technology and Biophysics.* Department of Biomedical Engineering and Physics (Academic Medical Center, University of Amsterdam): Academic Medical Center, University of Amsterdam, 2009.

(Cited in pages 110 and 111.)

[127] J. R. Meyer-Arendt, "Radiometry and photometry: Units and conversion factors," *Applied Optics*, vol. 7, no. 10, pp. 2081–2081, 1968.

(Cited in pages 110 and 111.)

[128] C. S. McCamy, H. Marcus, and J. Davidson, "A color-rendition chart," *Journal of Applied Photographic Engineering*, vol. 2, no. 3, pp. 95–99, 1976.

(Cited in page 114.)

## Résumé

La simulation de matériaux physico-réalistes est un processus demandant beaucoup de calcul. Les images de synthèse étant destinées aux observateurs humains, nous pouvons utiliser les limitations de notre système visuel pour simplifier le modèle de rendu, en évitant le calcul d'information invisible. Cela s'appelle le réalisme perceptif. Nous nous intéressons à la simulation de peintures d'automobiles, en particulière aux peintures scintillantes à paillettes métalliques. Nous essayons d'améliorer le réalisme perceptif de deux manières : en utilisant de la visualisation stéréoscopique pour apporter de l'information de profondeur additionnelle à partir de la disparité binoculaire ; et en conservant autant d'information perceptive de la luminance originale que possible. La gamme dynamique illimitée d'une scène réelle est réduite lorsqu'une image est traitée dans des dispositifs à une gamme dynamique plus basse dans la chaîne d'acquisition et de visualisation. Pour assurer un réalisme perceptive, nous proposons une méthodologie reposant sur la caractérisation des dispositifs, l'acquisition d'information radiométrique, et des validations visuo-perceptives. En remplaçant l'œil humain par un appareil photo numérique, en tant qu'intégrateur tristimulaire d'information radiométrique, nous réalisons des comparaisons visuelles entre des échantillons réels et des photographies pour estimer la valeur d'exposition qui maximise le réalisme perceptif dans un environnement d'observation contrôlé. Ces résultats sont ensuite contrastés avec plusieurs méthodes de reproduction tonale, afin d'analyser les effets perceptifs de certains attributs d'image tels que l'exposition, la gamme dynamique, la brillance, et le contraste. Nous proposons également une méthodologie complète pour simuler des scènes réelles qui soient comparables, d'un point vue radiométrique et colorimétrique, aux photographies de la même scène. En assurant des images simulées correctes, cette méthodologie établie les bases face à une future intégration de nos observations dans le moteur de rendu.



## Mots Clés

Perception visuelle humaine, réalisme perceptive, validation d'images, modèles physiques de matériaux, peintures scintillantes automobile, psychophysique.



## Abstract

The simulation of physico-realistic materials is a process that requires a lot of computation. Since the images are meant to be seen by human observers, we can use the limitations of their visual system to simplify the rendering model, avoiding redundant information that will not be seen. This is known as perceptual realism. Focusing on the simulation of automobile paint coatings, with special attention to metallic-flaked coatings with a sparkling appearance, we try to improve perceptual realism in two ways: using stereoscopic visualization, to provide additional depth information from binocular disparity; and preserving as much of the original perceptual luminance information as possible. The unlimited luminance levels, or dynamic range, of a real scene must be reduced as an image is processed by lower-dynamic range media throughout the acquisition and visualization chain. To ensure perceptual accuracy throughout this process, we propose a methodology consisting of device characterization, radiometric acquisition, and visuo-perceptual validations. Replacing the human eye by a DSLR camera, as a trichromatic color integrator of radiometric information, we perform visual comparisons of real samples and photographs to estimate the image exposure that maximizes perceptual accuracy under a controlled observation environment. These results are then contrasted with different tone reproduction methods, in order to analyze the effects on texture perception of specific image attributes like exposure, dynamic range, brightness, and contrast. We also propose a full methodology to produce simulations of a real scene, which are radiometrically and colorimetrically comparable to photographs of the same scene. By ensuring that the simulation produces correct images, this methodology lays the foundations for a future integration of our observations into the rendering engine.



## Keywords

Human visual perception, perceptual realism, image validation, physical material models, sparkling car paint coatings, psychophysics.