Digital Camera Colour Processing Pipeline for High Dynamic Range Imaging and Colour Stabilisation for Cinema

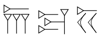
Raquel Gil Rodríguez

DOCTORAL THESIS UPF / YEAR 2018

Thesis Advisors: Marcelo Bertalmío and Javier Vazquez-Corral
Department of Information and Communication Technologies

Universitat
Pompeu Fabra
Barcelona

IP4EC
Research Group on Processing
for Enhanced Cinematography

To my family *ba* 

Acknowledgments

First of all, I would like to thank my supervisors Marcelo Bertalmío and Javier Vazquez-Corral for their continuous support and encouragement through all these Ph.D years, and for their patience and motivation. Their guidance helped me in all the time of research and writing of this thesis.

I would like to thank my dear Ph.D comrades, without them this journey would have been dull. I am glad we shared our failures and success among these years. You are and always be my Ph.D family: Gabriela, Waqas, Praveen, Maria, Antoine, Jiji, Itziar, Adrián, Javier, and many others that make this family huge.

Last but not the least, I would like to thank my family and friends: my parents and sister for supporting me no matter what, and the help of my dear friends, even when they could not understand me. Words cannot express how thankful I am to Arash *koochooloo-ye man*.

‘Nothing is so painful to the human mind as a great and sudden change.’

Mary Wollstonecraft Shelley, Frankenstein

Abstract

In this thesis we focus on two different image processing challenges: high dynamic range (HDR) image/video generation, and colour matching. In both cases, we redefine these tasks by taking into account the prior knowledge from the different processes the camera performs when capturing the image.

High dynamic range techniques have recently become very popular, thanks to the emergence of HDR recording and displaying technology. We propose two different approaches for HDR creation, one for still images, and one for video creation. In case of still images, most of previous methods follow a multiple exposure combination approach. These methods share a set of building assumptions: i) the full dynamic range of the scene can be recovered, ii) colour channels are independent, and iii) the camera response function remains constant while varying the exposure time. First, we highlight how these assumptions do not apply in general for digital cameras, and then, we propose a method to improve multiple exposure combination. Our results outperform state-of-the-art multiple exposure techniques.

In case of HDR video, we present a simple and affordable method to generate high quality video from an HDR scene. Our input is an interlaced video alternating row pairs with different ISO values, as some DSLR camera models can provide. The proposed algorithm involves two main steps: i) the computation of two single-ISO full-frame images (one for each ISO value) using an inpainting-based deinterlacing method, and ii) the linear combination of the full-frame ISO’s into a single HDR frame. Finally, the results are tone-mapped into an LDR frame ready for display. Results are free of ghosting artefacts and present little noise.

Colour matching approaches attempt to transfer the colours of a reference image, to another source image. In this context, we focus on colour matching two images taken from the same scene. First, we propose a method that modifies logarithmic encoded images, used in cinema for HDR content, in order to behave as gamma encoded images, used in most digital cameras. Then, we extend a previous approach defined just for gamma encoded images. We redefine the transformation between the two images, by considering a projective transform, and we estimate the method parameters in a single optimisation step. The method outperforms the state-of-the-art methods and can handle real-life examples.

Resum

En aquesta tesi ens centrem en dos problemes de processament d'imatges diferents: generació d'imatge/vídeo d'alt rang dinàmic (HDR) i coloració. En tots dos casos, redefinim aquestes tasques tenint en compte el coneixement previ dels diferents processos que realitza la càmera en capturar la imatge.

Actualment, les tècniques d'alt rang dinàmic s'han tornat molt populars, gràcies a l'aparició de la tecnologia per capturar i visualitzar HDR. Proposem dos enfocaments diferents per a la creació d'HDR, un per a imatges i un altre per a la creació de video. En el cas d'imatges, la majoria de mètodes combinen múltiples exposicions. Aquests mètodes comparteixen un conjunt d'hipòtesis: i) la recuperació del rang dinàmic complet de l'escena, ii) els canals de color són independents, i iii) la funció de resposta de la càmera es manté constant mentre es varia el temps d'exposició. En primer lloc, destaquem com aquestes suposicions no s'apliquen, en general, a les càmeres digitals, i després proposem un mètode per millorar aquesta tècnica. Els nostres resultats superen l'estat de l'art.

En el cas de vídeo HDR, presentem un mètode senzill i assequible per generar vídeos d'alta qualitat d'una escena HDR. El nostre input és un vídeo entrellaçat alternant parells de fileres amb diferents valors d'ISO, com alguns models de càmeres DSLR poden proporcionar. L'algorisme inclou dos passos principals: i) el càlcul de dues imatges full-frame ISO (una per a cada valor d'ISO) utilizant un mètode de desentrellaçat basat en inpainting, ii) la combinació lineal dels ISOs full-frame en un HDR únic. Finalment, els resultats es mapegen tonalment per obtenir un LDR per mostrar per pantalla. Els resultats no tenen artefactes de ghosting i presenten poc soroll.

Els mètodes d'igualació de colors intenten transferir els colors d'una imatge de referència, a una altra imatge d'origen. En aquest context, ens centrem en el cas de dues imatges capturades a la mateixa escena. En primer lloc, proposem un mètode que modifica imatges codificades logarítmicament, utilitzades en el cinema per a continguts HDR, per tal de comportar-se com imatges gamma codificades, que s'utilitzen en la majoria de les càmeres digitals. A continuació, extenem un mètode definit prèviament només per imatges gamma codificades, redefinint la transformació entre les dues imatges, considerant una transformació projectiva i estimant els paràmetres del mètode en un únic pas d'optimització. El mètode supera l'estat de l'art i pot tractar exemples de la vida real.

Contents

Nomenclature **xix**

List of Figures **xxvi**

List of Tables **xxviii**

1 INTRODUCTION **1**
- 1.1 Contributions . . . 7
- 1.2 Publications . . . 10
- 1.3 Thesis Outline . . . 10

2 HUMAN VISUAL SYSTEM: FROM LIGHT TO COLOUR VISION **13**
- 2.1 Light . . . 13
  - 2.1.1 Biology of the human eye: optics and retina . . . 16
  - 2.1.2 Dynamic Range . . . 20
    - Visual adaptation: Photoreceptor Response . . . 22
    - Dark and Light Adaptation . . . 23
- 2.2 Colour matching experiments . . . 24
  - 2.2.1 Chromatic Adaptation . . . 26
- 2.3 Colour spaces . . . 27
  - 2.3.1 Standard colorimetric colour space . . . 27

2.3.2 Perceptually uniform colour spaces 32
2.3.3 Other colour spaces 34
2.3.4 Colour metrics 35

3 DIGITAL IMAGE FORMATION 39
3.1 Image acquisition: Exposure Control 40
3.1.1 Exposure Control 40
3.2 Camera Sensor 41
3.2.1 Dynamic Range 44
3.3 In-Camera Colour Processing Pipeline 45
3.3.1 White balance 46
3.3.2 Demosaicing 47
3.3.3 Colour correction 47
3.3.4 Encoding techniques 48
Gamma correction 49
Logarithmic encoding 52
3.3.5 Post-processing steps 52
3.3.6 Compression and image formats 52
3.4 A general approximation for in-camera colour processing pipeline 53

4 HIGH DYNAMIC RANGE ACQUISITION AND GENERATION 55
4.1 Dynamic Range 55
4.2 Multiple LDR exposures for HDR reconstruction 55
4.2.1 Camera response function 57
4.2.2 Noise models 59
4.2.3 Exposure fusion 60
4.2.4 Camera or scene motion 61
LDR alignment and HDR reconstruction 61
Joint LDR alignment and HDR reconstruction 62
Ghost detection 62

4.2.5 Colour matching for HDR 64
4.3 High dynamic range sensors and video 65
4.4 High dynamic range storage 68
4.4.1 Linear encodings 68
4.4.2 Non-linear encodings 69
4.5 HDR datasets 70
4.6 HDR colour spaces 70
4.7 HDR metrics 71
4.8 HDR visualisation: Tone Mapping Operators 72
4.8.1 Global TMOs 73
4.8.2 Local TMOs 74
4.8.3 Joint global and local TMOs 74
4.8.4 Video TMO methods 74
4.8.5 TMO metrics 75
4.9 HDR reconstruction from single LDR images: inverse Tone Mapping Operators 76

5 CONSIDERING THE DIGITAL CAMERA PIPELINE FOR HDR CREATION 79

5.1 Multiple exposures assumptions for HDR 79
5.2 The effect of optical scattering on the dynamic range 81
5.3 The response function of digital cameras 84
5.4 Proposed method to make the CRF constant 90
5.4.1 Step 1 91
5.4.2 Step 2 92
5.5 Results and comparisons 94
5.5.1 Database 94
5.5.2 Ground-truth generation 94
5.5.3 Evaluation 95
Dynamic scenes 97

5.6 Conclusion 99

6 HDR VIDEO FROM DUAL-ISO SENSOR 101

6.1 High quality video in high dynamic range scenes from interlaced dual-ISO footage 101

6.1.1 Generation of single-ISO full-frame images 102

6.1.2 Row interpolation by deinterlacing 103

6.1.3 Refinement step 104

6.1.4 HDR creation 105

6.1.5 Final LDR video 106

6.2 Experiments and Evaluation 107

6.2.1 Quantitative evaluation 107

6.2.2 Qualitative evaluation 108

6.3 Analysis and optimisation of the proposed method 112

6.4 Conclusion 114

7 COLOUR MATCHING: RELEVANCE AND METHODS 117

7.1 Motivation 117

7.2 Current solution on cinema and TV 118

7.3 Colour matching in academia 119

7.3.1 Colour transfer 119

7.3.2 Colour stabilisation 121

Colour consistency among multiple views 122

8 COLOUR STABILISATION FOR CINEMA CONTENT 125

8.1 Motivation: encoding techniques for cinema content 125

8.2 The proposed framework 127

8.2.1 From log-encoded to gamma-corrected images 127

8.2.2 Colour stabilisation 129

8.2.3 Undo power 10 function 130

8.3 Results and Discussions 130

8.3.1 Dataset 131
8.3.2 Analysis of our colour stabilisation model 132
8.3.3 Experiments versus state-of-the-art 134
Gamma-corrected inputs 136
Log-encoded inputs 136
Log-encoded reference and gamma-corrected source . . 136
Gamma-corrected reference and log-encoded source . . 138
8.3.4 Experiments with power 10 139
Log-encoded inputs 139
Log-encoded reference and gamma-corrected source . . 140
Gamma-corrected reference and log-encoded source . . 140
8.3.5 Beyond gamma and log-encoded images: HLG and PQ . 142
8.4 Conclusion 145

9 CONCLUSIONS AND FUTURE WORK 147

Nomenclature

3D Three Dimensions

CAM Colour Appearance Model

CCD Charge Coupled Device

CCM Colour Correction Matrix

CFA Colour Filter Array

CID Colour Image Difference

CIE Commission Internationale de l’ Éclairage

CMOS Complementary Metal Oxide Semiconductor

CRT Cathode Ray Tube

DR Dynamic Range

DRIM Dynamic Range Inpendent Metric

DSLR Digital Single-Lens Reflex Camera

HDR High Dynamic Range

HVS Human Visual System

HLG Hybrid Log-Gamma

iTMO	Inverse Tone Mapping Operator
LCD	Liquid Crystal Display
LDR	Low Dynamic Range
LUT	Look Up Table
MSE	mean square error
PQ	Perceptual Quantizer
PSF	Point Spread Function
PSNR	Peak Signal to Noise Ratio
RGB	Red Green Blue
SMPTE	Society of Motion Picture and Television Engineers
SSIM	Structural Similarity Index Measure
TMO	Tone Mapping Operator
VDP	Visible Difference Predictor

List of Figures

1.1 Images acquired with different exposure times and the reconstructed HDR image. 2
1.2 Interlaced image with rows alternating ISO values of 100 and 1600. Image from [Lantern, 2013]. 4
1.3 Images acquired with a Nikon D3100 changing the white balance of the camera. 5

2.1 Electromagnetic spectrum. Image from [Bertalmío, 2014b]. . . 14
2.2 Relation between radiance, luminance and brightness. Image from [MathWorks, 2010]. 15
2.3 Structure of the eye and its components. Image from [Hunt and Pointer, 2011]. 16
2.4 Schema of the retina and its photoreceptors. Images from [Fairchild, 2013]. 17
2.5 The distribution of the photoreceptors in the retina. Image from [Wandell, 1995a]. 18
2.6 Centre-surround of receptive field: on-centre and off-centre field. Image from [Fairchild, 2013]. 19
2.7 Spectral sensitivity of the cones and luminosity functions. Images adapted from [Wikipedia, 2017] and [Wikipedia, 2018e]. 19
2.8 The tristimulus that we perceive are the integral of the product of the radiance and the spectral sensitivity at each cone type. . . . 20

2.9 Opponent colour signals of the tristimulus of the cones. Image from [Wikipedia, 2018d]. 21

2.10 Luminance range of the real world, and our HVS. Image from [Kunkel et al., 2016]. 21

2.11 An approximation of high dynamic range scene perception. Image from [Cambridge, 2005]. 22

2.12 Photoreceptor response to luminance. Image adapted from [Reinhard et al., 2010], and [Kunkel and Reinhard, 2010]. 23

2.13 The time course of dark and light adaptation. 24

2.14 Experiment set-up and colour matching functions. Images from [Wong, 2018] and [Wikipedia, 2018f]. 26

2.15 Chromatic adaptation. Two pictures of an opened book with white pages, under two different illuminants. 27

2.16 Chromatic adaptation. Image from [Fairchild, 2013]. 28

2.17 The colour matching functions $\{\bar{x}, \bar{y}, \bar{z}\}$. Image from [Wikipedia, 2018a]. 29

2.18 The chromaticity diagram. Image from [Solis et al., 2010]. . . . 30

2.19 The geometric transformations from the RGB cube to HSV. Image adapted from [Wikipedia, 2018c, Popov et al., 2018]. . . 31

2.20 CIELAB and CIELCh colour space representations. Image from [Bertalmío, 2014a]. 33

2.21 The chromaticity diagram of CIELUV (u', v'). Image from [Wikipedia, 2018b]. 33

2.22 The colour spaces presented and their connections. Image adapted from [Poynton, 2003]. 35

2.23 Two images I_1 and I_2 which differ in terms of colours, with colour metric values: $CID = 0.225$, $\Delta E^*_{00} = 8.50$. 36

3.1 Digital camera processing pipeline. Image from [Ramanath et al., 2005]. 39

3.2 Exposure triangle: aperture, ISO value and shutter speed. Image adapted from [PhotographyLife, 2018]. 41

3.3 Two camera sensor types: on the left a CCD, and on the right a CMOS sensor. Image from [Nakamura, 2005]. 42
3.4 Three sensor system and colour filter array (CFA). Image from [Bertalmío, 2014a]. 43
3.5 Comparison of the spectral sensitivity of the cones, and two DSLR camera sensors. Image adapted from [Wikipedia, 2017] and [Photography, 2018]. 43
3.6 The tristimulus as the integral of the product of radiance and spectral sensitivity of the filter sensor. 44
3.7 The dynamic range of a camera pixel. Image from [Nakamura, 2005]. 45
3.8 Demosaicing framework. 48
3.9 Colour transformation from RGB colour space of camera sensor to standard RGB colour space. 49
3.10 Opto-electro transfer function. Image adapted from [Borer and Cotton, 2017]. 49
3.11 Comparison of CRT and gamma function versus input voltage . 50
3.12 Results of applying different gamma correction values to *Lena*. . 51
3.13 Graph of transfer functions, and lightness; and plot of quantisation vs brightness of linear and gamma responses. 51
3.14 In-camera colour processing pipeline. 53

4.1 Pictures taken with a DSLR by fixing aperture and ISO value, and only varying the exposure time. 56
4.2 'The Brig', Photograph by Gustave Le Gray, 1856. Image from [Le Gray, 1856]. 57
4.3 The camera response function. The chain of steps from light scene to camera output. Picture from [Debevec and Malik, 1997]. 57
4.4 The characteristic curve of film: relative exposure versus optical density. Image adapted from [Kodak, 2018]. 58
4.5 Ghosting artefacts in HDR imaging. Image from [Srikantha and Sidibé, 2012]. 62

4.6 Dual-camera set-up in stereo rig composition. Image from [Froehlich et al., 2014]. 66

4.7 Camera prototype from [Tocci et al., 2011]. Image from [Tocci et al., 2011]. 66

4.8 Colour representation of colour space IC_TC_P. Image from [Dolby, 2016]. 71

4.9 Display of an HDR image: linear rescaling, and tone-mapped [Mantiuk et al., 2008]. 73

5.1 Point spread function of a CCD sensor, and the median versus DR of HDR data [Fairchild, 2007]. 82

5.2 Sixteen individual exposures used by M. Fairchild in [Fairchild, 2007] to create the *LuxoDoubleChecker* HDR image. 85

5.3 Plots from points in the coloured squares in the dark colour checker in Figure 5.2, using the RAW, and the JPEG values. . . 86

5.4 Analysis of non-diagonal elements of A and gamma-correction value for different cameras and varying exposure time. 88

5.5 Comparison of different HDR methods applying different TMOs. 89

5.6 Diagram of the proposed HDR method. 90

5.8 Intermediate HDR images taken different references, and the final HDR image. 94

5.9 Comparison of the proposed HDR method and other approaches. 98

5.10 Comparison of different HDR methods under different TMOs. . 99

5.11 HDR results on a dynamic scene applying the HDR creation method of [Sen et al., 2012]. 100

6.1 Flowchart of dual-ISO proposed method for HDR video generation. 102

6.2 The dual-ISO input frame, which is split into two half-size images, each one with the rows corresponding to a single ISO value. . . 103

6.3 Generated full-frame images I_l (left) and I_h (right). 103

6.4 Inpainting-based deinterlacing algorithm. Figure adapted from [Ballester et al., 2007]. 104

6.5 Extended deinterlacing algorithm applied to an CFA image. . . . 105
6.6 Refinement step after deinterlacing is applied. 106
6.7 Left: dual-ISO input. Right: tone-mapped output from our method. 106
6.8 Three scenes from Fairchild dataset. The first row represents the simulated dual-ISO input, and the second row the final tone-mapped output. 108
6.9 Evaluation diagram of dual-ISO approach. 109
6.10 Comparison results against ML [Lantern, 2013] and Hajisharif *et al.* [Hajisharif et al., 2014]. 110
6.11 Comparison results against ML [Lantern, 2013]. 110
6.12 Frames from dual-ISO video and our results. 111
6.13 Tone-mapped results from two video sequences. Results of ML versus results of our approach. 111
6.14 Results from using different parameter values in our method. . . 113

7.1 Colour matching a pair of images taken from the same scene. . . 117
7.2 Camera control unit, DaVinci Resolve tool and Imatest sofware. 119
7.3 Colour transfer example: reference, source and colour transferred result. 120
7.4 Colour stabilisation example: reference, source and resulting image. Image from [Vazquez-Corral and Bertalmío, 2014a]. . . 122
7.5 Example of colour consistency application. Image from [HaCohen et al., 2013]. 123

8.1 Linear response versus gamma-corrected and logarithmic response. 126
8.2 Flowchart for colour stabilisation using power 10 approach given two log-encoded images. 128
8.3 Graph of 3 standard logarithmic encoded curves. 129
8.4 Evaluation framework: data acquisition, data creation and evaluation. 131
8.5 Results of varying matrix size in our colour stabilisation method. 134
8.6 Results of all the methods for the four comparisons. 137

8.7 Results from applying power 10 to log-encoded inputs in all the methods. 141
8.8 Results of our algorithm using ARRI dataset. 144

List of Tables

2.1 Photometric and radiometric measurements. 15

4.1 Different dynamic range definitions depending on the application or context were they are introduced. 56
4.2 Characteristics of multiple exposure methods of static scenes for HDR reconstruction. 61
4.3 Characteristics of multiple exposure methods for camera motion and/or dynamic scenes. 64
4.4 Image and video encodings presented in this section. 68
4.5 High dynamic range datasets. 70

5.1 Mean and median results from all the presented methods versus GT among the 104 scenes. The last metric is computed only among the 42 images. 97

6.1 Quantitative evaluation. Results for proposed method and for ML [Lantern, 2013]. 109
6.2 Comparison of the total execution time of the original version, and the two versions proposed in [Sintes, 2017]. 114

8.1 Study of our approach considering different matrix size. 133
8.2 Results from the comparison among 35 image pairs. 135
8.3 Results show mean and median among 22 pairs, where reference and source images are encoded using logarithmic curves. 138

8.4 Results from the comparison among 35 image pairs applying power 10 to the methods. 139
8.5 Results show mean and median among 10 pairs, where reference and source images are encoded using HLG, PQ and logarithmic curves. 143

CHAPTER 1

Introduction

In this thesis we focus on three different image processing challenges arising in cinematography. First, we present two works in the topic of high dynamic range (HDR) imaging. In particular, we propose efficient solutions for HDR image generation and HDR video creation. The third problem in which we work is colour stabilisation of image pairs. In this case, we focus on images encoded using logarithmic curves, which are becoming the standard for professional cinema cameras. Our solutions to these problems exploit the prior knowledge we do have from the different processes the camera performs when capturing an image.

Dynamic range is defined as the ratio between the brightest and the darkest intensity levels within a scene. The human visual system (HVS) is able to adjust to world scenes where the light intensity values cover a very wide range varying from 10^{-6} cd/m^2 for starlight to 10^{8} cd/m^2 for sunlight. Moreover, it can capture details in dark and bright areas simultaneously because of its capacity of adaptation. This is not the case of standard cameras/sensors. While in common situations the light coming from a scene is of HDR, the vast majority of camera sensors (and displays) are of low dynamic range (LDR).

We first focus on the HDR reconstruction for still images. In digital cameras, the 12-bit or 14-bit values captured at sensor level which are proportional to light intensity (this is the linear data, stored in RAW format) go through a chain of colour-correction transformations culminating in a non-linear transform (gamma correction) followed by quantisation in 8 bits per channel. The net result is that standard cameras are only able to capture well different intervals of the luminance range at different exposure times. In particular, bright areas are better captured at short exposure times, while longer exposure times are required for dark areas. In Figure 1.1 is presented a set of images taken varying exposure time (top row), and the HDR image obtained from the linear RAW data (bottom). On the reconstructed HDR, there are shown the luminance values in cd/m^2 of some regions in the scene, which has a dynamic range of almost 5 orders of magnitude.

Figure 1.1: On top row, images acquired with different exposure times using a Nikon D3100, from $[1/160, 1/10, 1.6]$. The image on the bottom row corresponds to the reconstructed HDR image from the set above. For display purposes, the image has been tone-mapped.

There is a vast literature on methods for creating HDR images using regular, LDR sensors, that started with the seminal approaches of [Mitsunaga and Nayar, 1999], and [Debevec and Malik, 1997]. In those works, several LDR pictures of the same HDR scene are taken only varying exposure time, so that the short exposures capture details in the bright regions, and long exposures capture details in the dark regions. These two approaches, and the vast majority of multiple exposure approaches, estimate the so called camera response function (CRF), which is the function that maps the non-linear values of the final camera output, to the linear RAW values at camera sensor. Once the set of images is 'linearised' using the CRF, they are all combined into a single HDR image with overall detail visibility. In order to build the CRF, some assumptions are taken into account:

1. Different colour channels are independent.
2. The CRF remains constant while changing the exposure.
3. The full dynamic range of the scene can be recovered.

The first condition indicates that given an RGB colour image, a curve will be computed for each colour channel. The second condition shows that the CRF is constant and unique. The last condition refers to the final computed 'radiance map', which is proportional to the luminance in the scene. These three assumptions apply for film photography, but they are not an accurate model of how digital cameras work. In this context, our main contributions are: i) to highlight the reasons why these three assumptions do not apply for digital cameras based on experiments and analysis, and ii) to propose a new method to improve multiple exposure techniques.

The second problem we address is the creation of HDR video. In this case, it is not possible to use the previous approaches due to artefacts that might appear when there are moving objects in the scene, and/or camera motion. These artefacts are called ghosting artefacts. Currently in the movie industry, the interest in HDR imaging is becoming more widespread, although the challenge of shooting HDR scenes using LDR equipment exists since the beginning of cinema. The way to address this limitation is adding artificial lights, in order to raise the intensity levels of the darkest areas of the image therefore reducing the dynamic range of the scene, fitting it into the reduced range of the capture medium (film or digital). This procedure is complicated and expensive, and it requires a large number of human and material resources that might affect significantly the cost of the production. In this framework, there exist some alternatives but they are not completely practical. For example, some digital cinema camera models are able to alternate exposure times on consecutive frames, creating pairs of different-exposure images that are then fused following the approach of [Debevec and Malik, 1997], but camera and/or object motion produces ghosting artefacts on the fusion results. A recent alternative is to use a dual-camera set-up [Froehlich et al., 2014], with two synchronised, perfectly registered cameras on an orthogonal rig so that a semi-transparent mirror sends most of the light intensity to one of the cameras, and the rest to the other camera. These images can be fused without any problem because they are fully aligned and synchronised, so there is no risk of ghosting artefacts. Nevertheless, the dual-camera set-up has some limitations:

cost and practicality considerations arising from the use of two cameras, image problems caused by the mirror imperfections, and the need to perform tone mapping to the output. For more details we refer the reader to [Bertalmío, 2014b].

Figure 1.2: Interlaced image with rows alternating ISO values of 100 and 1600. Image from [Lantern, 2013].

More recently, we find works that perform HDR reconstruction from a single interlaced image. Gu *et al.* [Gu et al., 2010] combine rows taken with different exposures times, and since the rows are not captured simultaneously this method produces ghosting artefacts as well, which need to be reduced by estimating and compensating for the motion-blur. The camera software Magic Lantern (ML) [Lantern, 2013] allows some camera models to capture image/video with dual-ISO values that alternate between consecutive image line pairs, see Figure 1.2. It provides an implementation to interpolate a full-frame low-ISO image, containing less noise on shadow areas. The method follows a chain of steps: separate the two ISO frames, interpolate the missing lines to get the full images, and combine information from both interpolated frames to highly reduce the noise in dark regions. Hajisharif *et al.* [Hajisharif et al., 2014] perform at the same time demosaicing, denoising, re-sampling and HDR reconstruction, starting from the interlaced input provided by the ML software [Lantern, 2013]. This method requires a previous radiometric calibration process, therefore it cannot be used when the camera is not available. A similar idea was developed by Heide *et al.* [Heide et al., 2014], who propose a single optimisation step using image priors and regularisers of the different stages going on in the camera colour pipeline (denoising, demosaicing, etc). The selection of image priors is crucial for the optimisation process, and the values are highly dependent on the set of images selected for learning the best weights. These latter two methods have the advantage of working directly with the RAW data without following a staged pipeline, therefore no cumulative errors are carried out from one process to the next; nevertheless, this integration makes it difficult to further extend the processes involved, since they are not independent, plus it is also challenging to locate and rectify errors in the pipeline.

In order to avoid ghosting artefacts and flickering, we decide to work with the software [Lantern, 2013] that allows us for capturing interlaced frames with different ISO values using common digital cameras. On this account, we follow a strategy that involves the same idea used in video deinterlacing but without taking into account temporal information.

Our last application focus on colour matching pair of logarithmic-encoded images acquired from the same scene. Colour matching techniques aim to map the colours of one image, defined as source, to those of a second image, defined as reference. A particular case is colour stabilisation, where the two pictures are taken from the same scene and differ in terms of colour. These colour differences are due to the use of different camera models, or even when the same camera model is used under different settings (white balance, exposure time, aperture, etc.). We can see an example in Figure 1.3, in where two images from the same scene are presented. We set the same exposure time, aperture and ISO value for both images, and only vary the white balance.

Figure 1.3: Images acquired with a Nikon D3100 changing the white balance of the camera.

As mentioned above, digital cameras perform typically the following in-camera steps: demosaicing, white balance, colour correction (from RGB camera sensor to device independent colour space), encoding standard (usually a gamma correction), and compression. Bianco *et al.* [Bianco et al., 2012] summarised a generic colour processing pipeline for digital cameras as

$$I_{out} = (A \cdot I_{lin})^{1/\gamma}, \tag{1.1}$$

where I_{out} is the output image, A is a 3×3 matrix which carries colour information and white balance, γ value defines a power law function, and I_{lin} is the linear image read by the camera sensor after demosaicing. This model is a simplification of the pipeline, since other post-processing techniques, like denoising or contrast enhancement might be applied.

Gamma correction has been the most used encoding technique in digital cameras, but it encounters issues when dealing with high dynamic range imaging, since in low luminance areas it might introduce quantisation artefacts. Current professional cinema cameras are able to capture a wide range of light intensities, and therefore, a compression of this range is needed for storage, while preserving all the details and appearance. Cinema cameras substitute gamma correction ($1/\gamma$) with a logarithmic function, with general form as,

$$I_{out} = c \log_{10}(a \cdot A \cdot I_{lin} + b) + d, \tag{1.2}$$

where I_{out} and I_{lin} are defined as above, and the parameters $a, b, c,$ and d are constant real values. These parameters might differ from camera manufacturers and camera settings.

The vast majority of methods in the literature consider gamma corrected inputs for colour matching. HaCohen *et al.* [HaCohen et al., 2011] presented a method to compute dense correspondences between the images, combined with a global colour mapping model. Vazquez-Corral and Bertalmío [Vazquez-Corral and Bertalmío, 2014a] proposed a colour stabilisation algorithm that consists of estimating a power law (γ value) for each of the images, and a single 3×3 matrix, to colour match the source image to the reference. It is built on the assumption that in digital cameras the colour encoding can be expressed as a matrix multiplication followed by a power law (gamma correction), see Equation (1.1). In Frigo *et al.* [Frigo et al., 2016], the authors presented a method to colour stabilise video sequences, based on the estimation of a non-linearity and channel-based scaling.

In the industry, there exist several solutions for bringing consistency across shots. They normally involve very skilled manual work, done by colourists during colour grading in movie post-production and by technicians using camera control units (CCU) [MediaCollege, 2012] in live TV broadcasts; require a proper characterisation of the cameras used and their settings like with the ACES framework [Postma and Chorley, 2015], or the presence of colour-charts in the shots.

To the best of our knowledge, the method of Vazquez-Corral and Bertalmío [Vazquez-Corral and Bertalmío, 2016] is the only work in the literature dealing with colour stabilisation for logarithmic-encoded images. Their algorithm relies on finding a sufficiently large number of achromatic matches among source and reference, in order to estimate the logarithmic curves and colour-stabilise the images. This detection of achromatic matches in some situations may be a challenging limitation.

In this framework, we propose to transform logarithmic-encoded images, in order to make them behave as gamma corrected ones. Next, we extend the previous work of [Vazquez-Corral and Bertalmío, 2014a], on colour matching pairs of gamma corrected images, to improve the results. Moreover, the transformation of logarithmic-encoded images can also be applied before utilising current gamma corrected methods. This step boosts the performance of the methods when dealing with these new inputs.

Contributions

Our first contribution in this thesis is the disproval of the assumptions used when creating an HDR image from a stack of different exposed images. These assumptions are: i) camera response function (CRF) does not change when varying exposure time, ii) colour channels are independent and iii) the full dynamic range of the scene (radiance map) can be recovered. Majority of the methods we find in the literature propose to compute a single and unique CRF from the non-linear multiple exposure images. We show that only varying the exposure time of the camera might cause changes in some other internal parameters of the camera, thus breaking the assumption of a *unique* CRF. The second assumption is related to the dependency of the colour channels. It is well known the correlation of the R, G, and B channels, thus when colour correction matrices are applied as an inner camera stage, those are not just diagonal matrices (which would mean that the channels are independent), rather those matrices are diagonal dominant with values different of zero outside it. Finally, we show that cameras cannot recover the full dynamic range of the scene due to the veiling glare of the camera optics. In addition, we propose a method that aims to improve quantitatively and qualitatively, multiple exposure techniques. At this point, given a stack of exposed non-linear images as inputs, our method considers the camera colour processing pipeline presented in [Bianco et al., 2012]. First, we set one the images in the stack as the reference, and then transform the rest of the images in the stack into the ‘space’ (in terms of colour and time) where the reference image is defined. The transformed images are combined linearly to obtain the final HDR image. We also proposed the option of selecting several images as reference, in order to better preserve details, and make the algorithm more robust and independent from only selecting one reference image. The results are compared against classical and state-of-the-art approaches using several metrics. Our analysis shows that our algorithm outperforms the rest of the methods.

The second contribution in this thesis is the creation of high quality HDR video. The approach we propose for HDR video reconstruction uses the output of a software designed to capture RAW frames in current DSLR cameras. Each frame is captured alternating two pairs of rows with two different ISO values. In this way, the camera output is an image with half of the rows acquired with an ISO low, and other half with ISO high. Then, we adapt a deinterlacing algorithm [Ballester et al., 2007] in the spatial domain to reconstruct the missing rows of each ISO half rows images. The reconstructed full size ISO images are then combined linearly, and finally a TMO is applied taking into account temporal domain. The result is an HDR video with no flickering neither or visible artefacts. We also compare our results with state-of-the-art algorithms that use dual-ISO.

Our last contribution is related to colour stabilisation on image pairs that are encoded either with gamma correction, or as many professional cameras now, with logarithmic functions. The vast majority of the literature assumes that the images are given in gamma corrected sRGB colour space. Due to the fact that logarithmic functions are becoming popular for HDR encoding in professional cameras, we introduce a modification that transforms log-encoded images, in order to be treated as gamma corrected ones. Then, we present an extension of a previous colour stabilisation work [Vazquez-Corral and Bertalmío, 2014a]. We consider a projective transformation instead of a linear mapping in the relation between the two images. Thus, we improve previous results, and moreover we outperform state-of-the-art methods. In addition, we show that the transformation of log-encoded images can be applied as well to other colour transfer algorithms, and the results show great improvement in most of the algorithms we considered for comparison, both qualitative and quantitatively. Finally, we show that our method can be applied to images captured with professional cinema video cameras, whose use current encoding techniques, like PQ and HLG.

Publications

Journals:

- R. Gil Rodríguez, J. Vazquez-Corral, M. Bertalmío. Issues with common assumptions in HDR imaging from multiple exposures. In preparation, 2018.
- R. Gil Rodríguez, J. Vazquez-Corral, M. Bertalmío. Color matching images with unknown non-linear encodings. In preparation, 2018.

Conferences:

- R. Gil Rodríguez, J. Vazquez-Corral, M. Bertalmío. The Intrinsic Error of Exposure Fusion for HDR Imaging, and a Way to Reduce it. British Machine Vision Conference (BMVC), Swansea, United Kingdom, 2015.
- R. Gil Rodríguez, M. Bertalmío. High quality video in high dynamic range scenes from interlaced dual-ISO footage. IS&T Electronic Imaging Conference, 2016.
- A. Akbarinia, R. Gil Rodríguez, C. A. Párraga. Colour Constancy: Biologically-inspired Contrast Variant Pooling Mechanism, in British Machine Vision Conference (BMVC), London, United Kingdom, 2017.
- R. Gil Rodríguez, J. Vazquez-Corral, M. Bertalmío. Color-matching Shots from Different Cameras Having Unknown Gamma or Logarithmic Encoding Curves. SMPTE Annual Technical Conference & Exhibition, 2017.

Thesis Outline

The rest of the thesis is structured as follows. In Chapter 2, we present briefly basic concepts of our human visual system and colour science: from the light reaching our eyes to our colour perception. Then in Chapter 3, we introduce how images are created in current digital cameras: from light reaching the sensor to the final colour image ready to be displayed. We emphasise the connections between these two chapters, since cameras try to emulate how scenes are perceived by humans. Chapter 4 presents the concept of high dynamic range, and methods and techniques for acquisition and creation of HDR imaging. We also introduce HDR visualisation techniques, and HDR reconstruction methods from a single image. In Chapter 5, we explain our proposed HDR method and discuss about the issues of most common building assumptions in HDR reconstruction from a stack of multiple exposure images. Next, Chapter 6 presents a method for HDR video creation from common DSLR cameras. It explains how to combine information from an original image that conveys information from two different ISO values. In Chapter 7, we review colour matching techniques used in image processing and computer vision, and we show its relevance to cinema industry. In Chapter 8, we explain our proposed method for colour matching a pair of images encoded using different techniques for cinema content. Finally, Chapter 9 gathers the conclusions of this thesis and suggests some further work regarding HDR and colour matching.

CHAPTER 2

Human Visual System: From light to colour vision

In our physical world, we can measure the light emitted from a source and reflected by an object using different sensors and tools. Our Human Visual System (HVS) works with its own sensors in order to perceive the light that reaches our eyes in a particular way. In this section, we will introduce some very basic knowledge on HVS, and how do we perceive colours.

Light

Electromagnetic radiations are present in our world. They are electromagnetic waves emitted by natural or human sources. These waves can be determined by their wavelength (λ) and they are represented in a range (spectrum) measured in meters (m). We encounter waves from short wavelengths, such as γ-rays and X-rays, to very long wavelengths such as microwaves and radio signals. The light that we can see corresponds to a very narrow range of the electromagnetic spectrum. The wavelengths of visible light vary only from 380 nanometres (nm) to 740 nm, see Figure 2.1.

It is important to introduce some definitions before going any further. On the one hand, *radiometry* measures physically wavelengths along the visible light and the whole electromagnetic spectrum. In particular, the *radiance* $E(\lambda)$ is defined as the product of the irradiance, which is the incident light $I(\lambda)$, and the

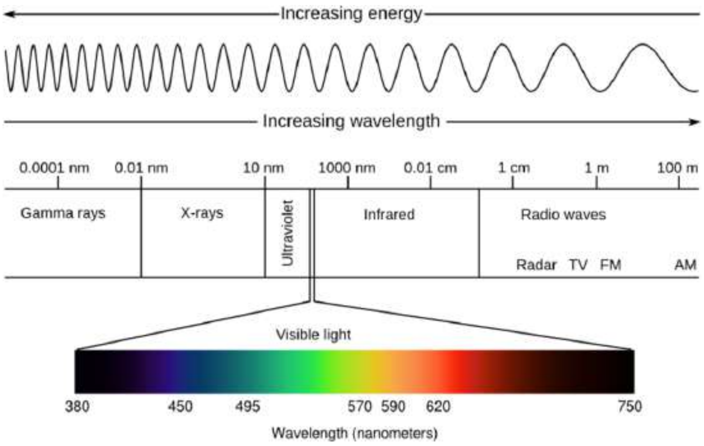


Figure 2.1: Electromagnetic spectrum. Image from [Bertalmío, 2014b].

reflectance of the object’s surface $R(\lambda)$, as functions of wavelength, as:

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

On the other hand, *photometry* measures the light based on the response of the human eye. In particular, the *luminance* Y is the effect of radiance in our eyes, and it is the integral of a weighted radiance over the visible spectrum,

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

The curve $V(\lambda)$ is a *luminosity function*, and it represents the efficiency of activating the cells in our retina during daylight at different wavelengths. As Figure 2.2 shows in the graph on the right, $V(\lambda)$ has a higher response on wavelengths in the middle of visible light, and the response decreases as it gets closer to the extremes of the visible spectrum. Note that this luminosity function converts radiometric measurements into photometric ones. In the next subsection, we will refer again to luminosity functions. The luminance values are measured in candelas per square meter (cd/m^2). We categorise our vision given different levels of luminance as:

- *scotopic* vision for low ranges of luminance,
- *mesopic* vision in low-medium ranges,

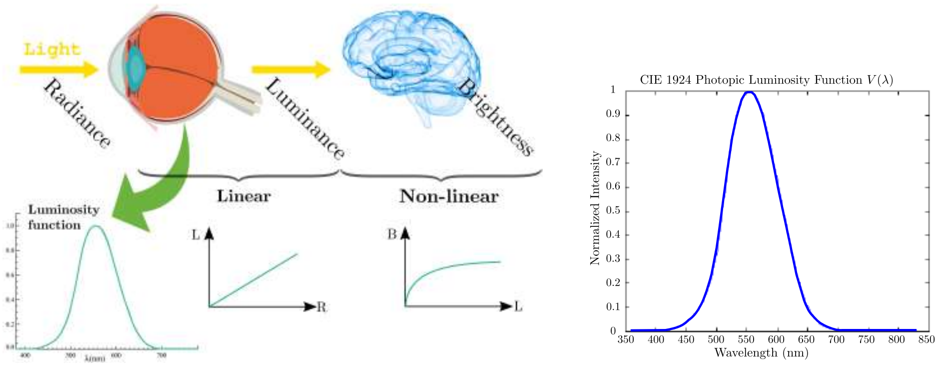


Figure 2.2: On the left: the relation between radiance, luminance and brightness. Image adapted from [Kindlmann, 2002]. On the right: the luminosity function $V(\lambda)$. Image adapted from [MathWorks, 2010].

Table 2.1: Photometric and radiometric measurements.

Photometric	*Radiometric*	*Description*
Luminous flux lm	Radiant flux W	power of light
Luminous intensity lm/sr	Radiant intensity W/sr	amount of flux emitted through solid angle
Illuminance lm/m^2	Irradiance W/m^2	power received on a known surface area
Luminance $lm/(sr \cdot m^2)$	Radiance $W/(sr \cdot m^2)$	intensity emitted from a known unit area of source area

- *photopic* vision in high-luminance ranges, where $V(\lambda)$ describes the eye sensitivity.

It is worth mentioning the concept of *brightness*, which is a subjective measure of the luminance from higher processing levels in our brain. The relation between brightness and luminance is non-linear, and it can be approximated by a cube root function [Stevens and Stevens, 1963]. Figure 2.2 shows the relation between radiance, luminance and brightness. In Table 2.1, we present the most common photometric and radiometric measurements, together with their units and a brief explanation. The term steradian (sr) is the unit of solid angle and it is the analogous in 3D of the radians. The term lm stands for lumens and $lm = cd \cdot sr$, and W for watts.

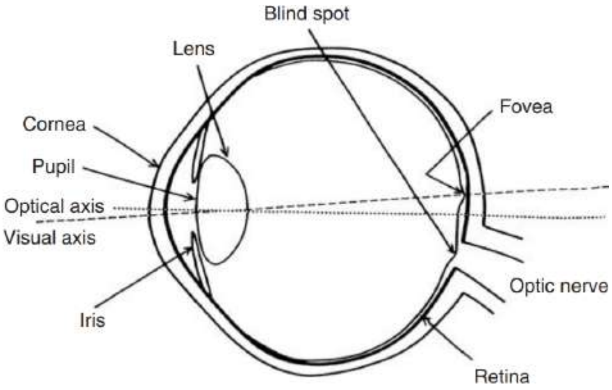


Figure 2.3: Structure of the eye and its components. Image from [Hunt and Pointer, 2011].

Biology of the human eye: optics and retina

The light passes through the optics of our eyes and reaches the sensors in our retina, where the light is converted into impulse signals that will be processed by our brain. In this section, we will first focus on the optical components of the eye. Then, we will move on to the sensor part of the eye, the retina, and the photoreceptors in it. In Figure 2.3 it is shown the structure of the eye, and its main components. The optical system of the human eye controls the amount of light that goes through it, and its direction. We present some of the elements involved in this process: cornea, pupil, iris, and lens.

- *Cornea* is the outer transparent layer through which the light passes the eye. The cornea absorbs most of the short wavelengths in order to protect the eye [Packer and Williams, 2003].
- *Pupil* can change its size (diameter), and it determines the aperture of the eye, controlling the amount of light going through.
- *Iris* controls and determines the maximum aperture of the pupil size.
- *Lens* is layered and flexible, and it can change its shape in order to focus on near or far objects.

In the retina, the light is projected into the fovea, the central region, since it is the most sensitive area to colour and spatial vision. The photoreceptors in the

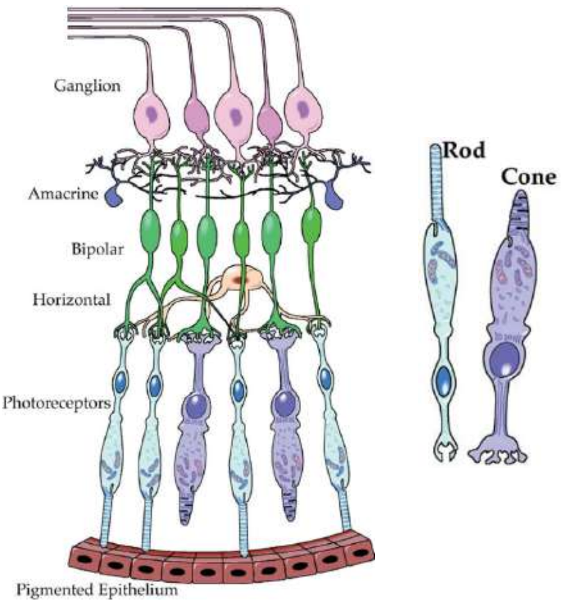


Figure 2.4: Schema of the retina. On the left, the neural cells: photoreceptors, horizontal, bipolar, amacrine and ganglion. On the right, the two types of photoreceptors: cones and rods. Images from [Fairchild, 2013].

retina absorb the photons, then generate chemical reactions that will become electrical impulses to be processed by our brain. We find two types of photoreceptors in the retina: rods and cones. The former are not sensitive to colour, and the latter are the ones which give us our colour vision. The sensitivity of rods and cones depends on the luminance level of the light. In Figure 2.4 it is shown a diagram of the retina, and the cells involved on capturing and transmitting the processed light.

- *Rods* are sensitive to low and mid-low luminances values (from 10^{-6} to 10 cd/m^2), and they give grayscale information. In scotopic and mesopic vision, rods are activated, and they saturate in photopic vision.
- *Cones* contain pigments less sensitive to light than rods (from 0.01 to 10^8 cd/m^2). Cones are the responsible for colour vision and they are divided in three different types, S-cones, M-cones and L-cones, that stand for short, medium and long wavelengths, respectively [Dartnall et al., 1983]. The

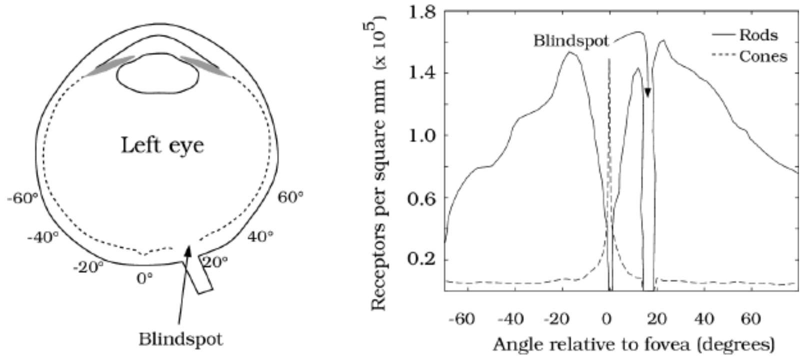


Figure 2.5: The distribution of the photoreceptors in the retina. On the left: the angles relative to fovea in degrees. On the right: concentration of rods and cones with respect to the angle relative to fovea. Image from [Wandell, 1995a].

cones are activated in photopic and mesopic visions.

Figure 2.5 shows the distribution of rods and cones with respect to the fovea area. Notice that the concentration of cones in the fovea is very high, and it drops drastically as it goes away from fovea. As opposite, rods are more concentrated in the peripheral areas. The number of rods in our eyes is approximately 120 millions, and the number of cones, approximately 6 millions.

The absorbed light by the photoreceptors is then transmitted to the neural cells (horizontal, bipolar, amacrine and ganglion), and finally the signal is sent to the optic nerve. This process within the retina is extremely complex, since different layers are involved. Notice that there are approximately 1 million optic nerves fibres, versus the roughly 126 millions of photoreceptors. This means that the signal need to be compressed. At each ganglion cell, there is a region, called the *receptive field*, in which all the photoreceptors connected to the cell will send their response. The structure of a receptive field is defined by its centre and surround. There are two types, on-centre field and off-centre field, as shown in Figure 2.6.

In the previous subsection, we introduced the curve that measured the sensitivity of our eyes during daylight which applies for photopic vision, $V(\lambda)$. There exists also another luminous function $V'(\lambda)$ that measures the sensitivity of our eyes during low light conditions, in scotopic vision. On the left plot of Figure 2.7, both curves are shown, $V(\lambda)$ in black, and $V'(\lambda)$ in green. These functions were adopted by the Commission Internationale de l'Éclairage (CIE) in 1924. Notice

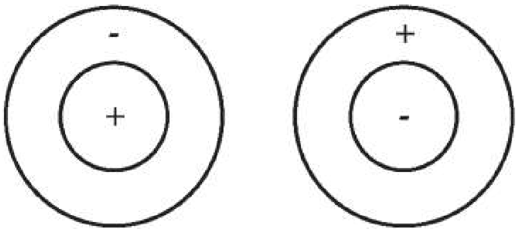


Figure 2.6: Centre-surround of receptive field. On the left: the on-centre field, and on the right: off-centre field. Image from [Fairchild, 2013].

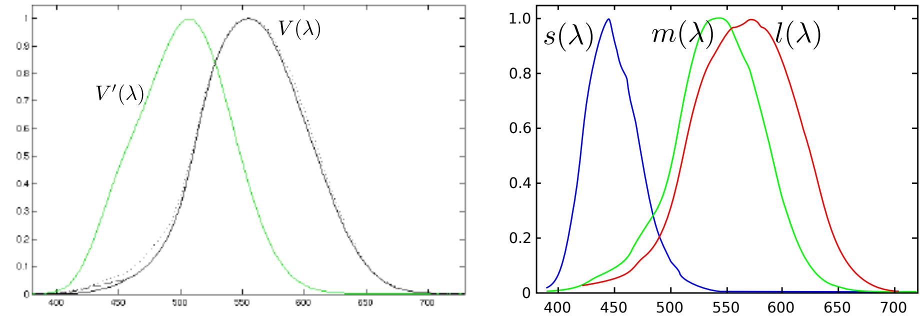


Figure 2.7: On the left: the luminosity functions *photopic* ($V(\lambda)$)curve in black, and *scotopic* ($V'(\lambda)$) curve in green. On the right: the spectral sensitivity (normalised) of S, M, and L-cones as functions of wavelengths. Images adapted from [Wikipedia, 2018e] and [Wikipedia, 2017], respectively.

that $V(\lambda)$ represents the combined sensitivity of the three different types of cones, and $V'(\lambda)$ represents exactly the response of the rod cells [Wyszecki and Stiles, 1982].

Let us focus on the three types of cones. Each of them has a different sensitivity to different wavelengths along the visible spectrum. On the right plot of Figure 2.7, we present the spectral sensitivity of S, M and L-cones. They are represented as functions of wavelength as $s(\lambda)$, $m(\lambda)$ and $l(\lambda)$. Notice that short cones reach their maximum spectral sensitivity at 420 nm, medium cones at 533 nm, and long cones at 584 nm. At the end, the response of the cones is expressed as *tristimulus* values, which are defined as the integral of the product of radiance and

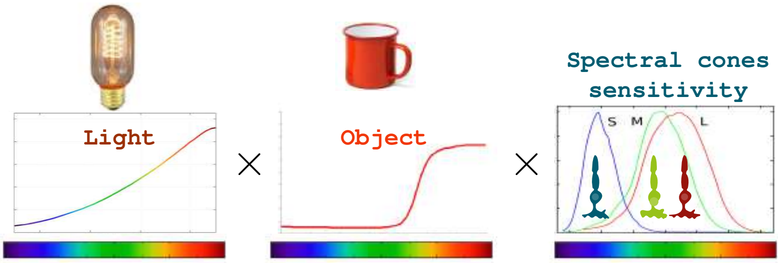


Figure 2.8: The tristimulus that we perceive are the integral of the product of the radiance and the spectral sensitivity at each cone type.

the sensitivity of each cone,

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

In Figure 2.8 it is shown the tristimulus as the product of: i) a light source (tungsten light), ii) the reflectance of a red object, and finally iii) the cones sensitivity. In this particular case, the tristimulus with higher value will be L.

The tristimulus L, M, S values are then processed in the receptive fields of higher neural cells, and as a result, three opposing pairs are defined [Shapley and Hawken, 2011]: 1) *intensity* which is the sum of all tristimuls $L + M + S$, equivalent to $V(\lambda)$, 2) *red-green* which is represented by $L - M$, and 3) *blue-yellow* by $L+M-2S$, see Figure 2.9. The response of the photoreceptors L, M, S is transformed into opponent signals in higher levels in the retina, and thus it provides an efficient way of compressing the signal for its transmission, [Jameson and Hurvich, 1955].

Dynamic Range

The dynamic range of a scene is the ratio between the highest luminance and the lowest luminance value. Our world presents a high dynamic range, and the

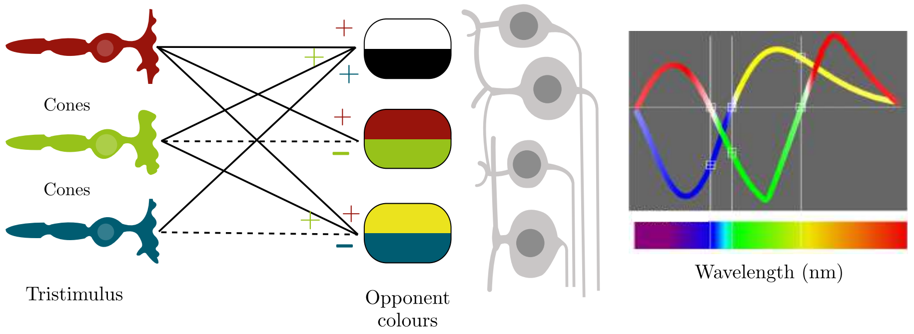


Figure 2.9: The tristimulus of the cones L, M, S are transformed into opponent colour signals, in higher neural cells. The graph on the right represents the *red-green*, and *blue-yellow* response as functions of wavelength. Image from [Wikipedia, 2018d].

luminance values can vary from $10^{-6}\ cd/m^2$ (starlight) to $10^8\ cd/m^2$ (sunlight), thus a contrast ratio of $10^{14} : 1$. This large range is often presented in $\log_{10}$ scale, for example, a scene during daylight, with luminance values varying from 10^0 to 10^5, will have a dynamic range of 5 orders of magnitude. In Figure 2.10, it is shown the range of luminance values that exist in the world. Notice that for scotopic and mesopic vision, rods can adapt to a range of 6 orders of magnitude, and in the case of photopic and mesopic vision, where cones are activated, those can adapt within a range of 10 orders of magnitude [Ferwerda et al., 1996].

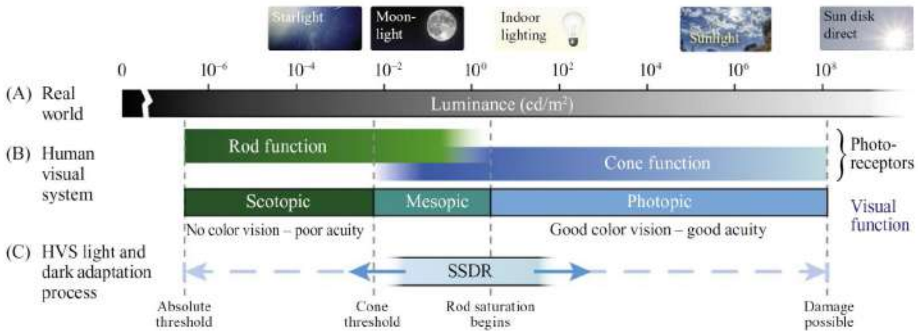


Figure 2.10: Luminance range of the real world, our HVS, and the adaptability of the photoreceptors to different light conditions. Image from [Kunkel et al., 2016].

Figure 2.11: High dynamic range scene, from left to right, an approximation of the our perception when we focus on the sky, when we focus on the rocks, and the overall perception. Image from [Cambridge, 2005].

Visual adaptation: Photoreceptor Response

Our HVS is able to function over a large range of luminance values during daylight. Although the contrast ratio between moonlight and sunlight is about $10^8 : 1$, it is difficult to encounter both in the same scene. Thus, the dynamic range of a scene is usually smaller. The HVS has the ability to adapt within different luminance conditions. Figure 2.11 shows a high dynamic range scene of the sea during sunset. The left image shows an approximation of what we perceive when we focus on the sky, on the middle it is shown the perception we get from the scene when focusing on the rocks in the front, and finally on the right the overall perception of the sea scene during sunset. Many different and complex factors influence our perception of luminance and colour. In this section we focus only on light, dark and chromatic adaptation of the photoreceptor cells in our retina.

Our HVS can operate at any given time in the range of approximately 4 orders of magnitude (when the pupil size remains unchanged), and it is called the *steady-state dynamic range*. The photoreceptors react within a small range of luminance, and their response to luminance values is given by the following equation,

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

where σ is the semi-saturation value, Y is the input luminance, R is the output response of the photoreceptor, R_{max} is the maximum photoreceptor response, and n is a sensitivity-control exponent, with ranges between 0.7 and 1 [Valeton and Norren, 1983]. Equation (2.4) is called *Naka-Rushton* equation, and for

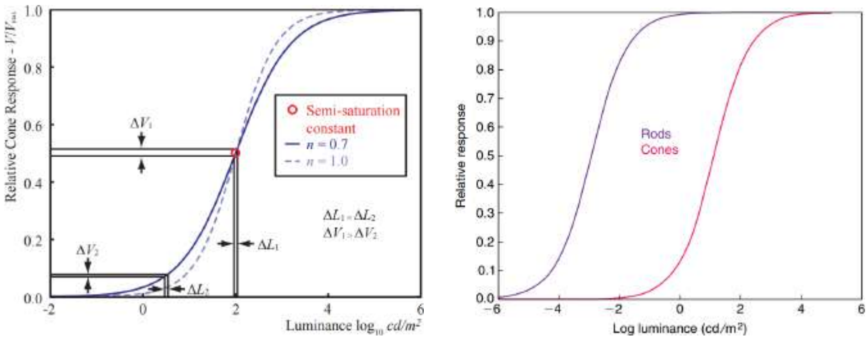


Figure 2.12: Photoreceptor response to luminance. On the left: luminance versus cone response represented by Equation (2.4). On the right: the response of rods and cones. Image adapted from [Reinhard et al., 2010], and [Kunkel and Reinhard, 2010].

$n = 1$, it is known as the *Michaelis-Menten* equation. In Figure 2.12, it is shown the luminance versus the response function of rods and cones, where the luminance axis is logarithmic. The semi-saturation value σ shifts the curve along the luminance access, therefore the σ value for rods is smaller than for cones, since rods are more sensitive to light.

Dark and Light Adaptation

One the one hand, dark adaptation happens when the illumination is reduced. As an example, when we enter into a cinema the lights are already off, and only the light from the screen can be seen. After a short time, ours eyes can adapt and we are able to see the people and objects inside, while still seeing the image projected on the screen. Figure 2.13 presents in the first row, the dark adaptation of rods and cones in a graph of time versus luminance. It takes up to 30 minutes to fully adapt to dark environment.

On the other hand, light adaptation happens when the level of luminance increases. As an example, when we leave the cinema, our eyes need to adapt from dark environment to sunlight. Figure 2.13 presents in the second row, the light adaptation of rods (left) and cones (right) system in a graph of time versus luminance. From the graphs, light adaptation takes around 5 minutes in the cones system, and in case of rods, light adaptation is achieved in only few seconds. For more details, we refer the reader to [Ferwerda et al., 1996]. It is interesting to

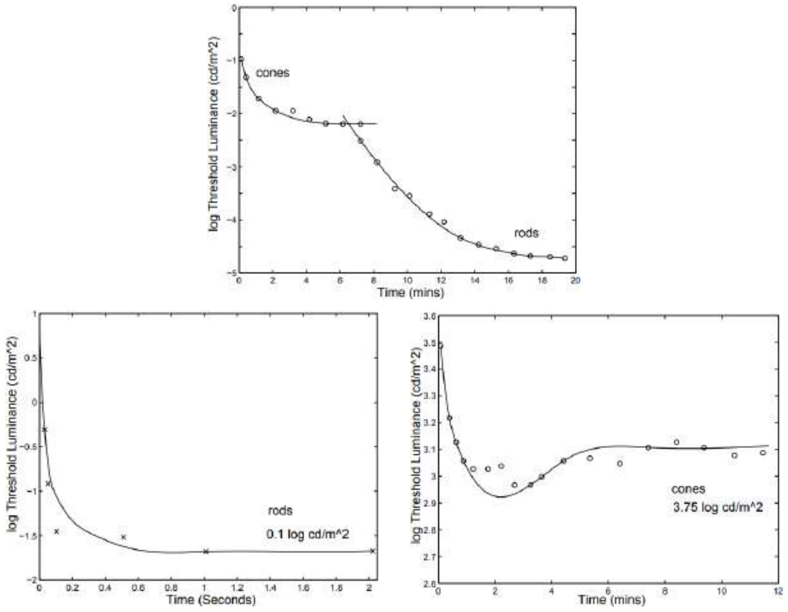


Figure 2.13: First row: the time course of dark adaptation. Second row: the time course of light adaptation in the rods (left), and cones (right) system. Image adapted from [Ferwerda et al., 1996].

mention that light/dark adaptation can be the equivalent of the exposure triangle in cameras (shutter speed, aperture and ISO value) that we will introduce in the next Chapter.

Colour matching experiments

Colorimetry is the branch of colour science that studies how to quantify the colours that we perceive from a physical stimulus, in a way that another physical stimulus under the same conditions should match [Wyszecki and Stiles, 1982]. In this section, we will describe the experiments done to quantify light as a triplet of *primary* values. Moreover, we will introduce the definition of different colour spaces.

The trichromacy property tells us that any colour can be defined as the combination of three primaries, by varying the amount of each of them. This is a property of the HVS and not a property of light, [Young, 1802]. It comes from the fact that in the retina we do have three different types of cones. Three primaries are enough and sufficient for describing a specific colour, this is known as Grassman's first law of additive colour mixture (1853). In 1850s Helmholtz and Maxwell continued the study of the trichromacy theory based on experiments, but it was not until 1890s that the first experiments were perform in order to measure the spectral absorption of the photoreceptors by König and Dieterici [König and Dieterici, 1892]. Experiments and studies for quantifying the spectral sensitivity of the three types of cones have been carried out until late 20th century years. For more detail we refer the reader to [Stockman and Sharpe, 2001]. This colour vision theory is known as *trichromatic theory*.

At the time, it was not proven the existence of exactly three different cones, nor their sensitivity responses. Experiments were performed to quantitatively show the existence and to analyse trichromacy. W. David Wright [Wright, 1929] and John Guild [Guild, 1931] carried a similar experiment independently, with 10 and 7 subjects, respectively. Subjects were asked to adjust the intensity of three monochromatic lights (460, 530, and 650 nm in Wright experiment), in order to colour-match a given target monochromatic light. The 3 monochromatic lights and the target light were projected on a screen side-to-side, as shown in Figure 2.14. At the end, the average among all the subjects was taken to define the so-called colour matching functions (CMF): $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$. Notice that some negative values appeared in the graph. It means that particular wavelengths cannot be matched, and thus for matching a fixed light (e.g. 500 nm), the red primary will be added and adjusted with the target light. The tristimulus values (R, G, B) for a given light of radiance $E(\lambda)$ are given by the integral over the visible light of the product of each CMF with the radiance,

$$\begin{aligned} R &= \int_{380}^{740} \bar{r}(\lambda)E(\lambda)\mathrm{d}\lambda, \qquad (2.5)\\ G &= \int_{380}^{740} \bar{g}(\lambda)E(\lambda)\mathrm{d}\lambda,\\ B &= \int_{380}^{740} \bar{b}(\lambda)E(\lambda)\mathrm{d}\lambda. \end{aligned}$$

Summarizing, given any set of primary lights $\mathcal{R}$, $\mathcal{G}$, and $\mathcal{B}$ (combination of two of them cannot result in the other one), a perceived colour given by a power light of radiance $E(\lambda_c)$ can be defined as a linear combination of these primaries

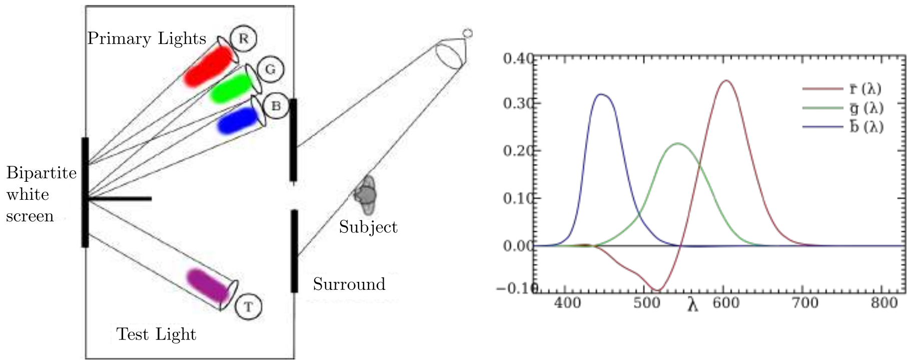


Figure 2.14: On the left, the experiment set-up. On the right, the colour matching functions. Images adapted from [Wandell, 1995b] and [Wikipedia, 2018f], respectively.

and their corresponding CMF as weighting functions $\bar{r}(\lambda_c)$, $\bar{g}(\lambda_c)$, and $\bar{b}(\lambda_c)$,

$$E(\lambda_c) \sim \bar{r}(\lambda_c)\mathcal{R} + \bar{g}(\lambda_c)\mathcal{G} + \bar{b}(\lambda_c)\mathcal{B}. \tag{2.6}$$

The colour matching experiments showed that only three primaries are needed to match all colours, thus our colour vision should be trichromatic.

Chromatic Adaptation

Chromatic adaptation takes place when one object is perceived with the same colour, under different illuminant conditions. For example, a white paper will have a flat radiance under an illuminant with uniform power spectra, whereas the same paper will have a radiance with more power on long wavelengths under a tungsten illuminant. Still in both situations, we perceive this paper as white. This is also known as *colour constancy*. In Figure 2.15, we show the pages of an opened book, with an orange pen on top, and a black keyboard at the background, under two different illuminant. Notice that, although the pages of the book appear greenish in the right picture, we still perceive them as white.

In 1902, von Kries described his hypothesis on the independence of the elements involved in our colour vision. He stated that each component is adapted based on its own response. Von Kries ideas were formulated later on into equa-

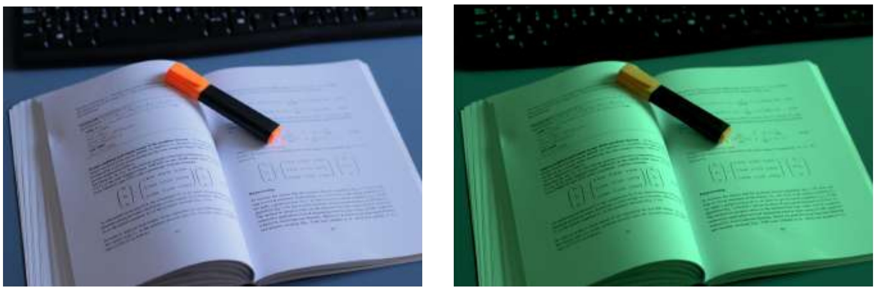


Figure 2.15: Chromatic adaptation. Two pictures of an opened book with white pages, under two different illuminant. In both cases, the book pages would be perceived as white.

tions, known as the *von Kries coefficient law*, as follows

$$\begin{aligned} L_2 &= k_L L_1, \\ M_2 &= k_M M_1, \\ S_2 &= k_S S_1, \end{aligned} \tag{2.7}$$

where k_L, k_M, and k_S are real values, L_1, M_1, S_1 are the initial tristimulus under one illumination, and L_2, M_2, S_2 are the new tristimulus under a second illumination (usually a equi-illuminant illumination, i.e. achromatic). This means that the cones spectral sensitivity can vary their maximum peaks independently, as multiplying each response by a different scalar. This is shown in Figure 2.16. Chromatic adaptation can be equivalent to automatic white balance in cameras.

Colour spaces

A colour space is a three-dimensional representation of colour. In the following sections, we will present different colour spaces definitions.

Standard colorimetric colour space

In 1931 the CIE proposed two colour matching functions: $\{\bar{r}, \bar{g}, \bar{b}\}$ and $\{\bar{x}, \bar{y}, \bar{z}\}$, from the data collected from the experiments of Wright and Guild. The first set of colour matching functions $\{\bar{r}, \bar{g}, \bar{b}\}$, see Figure 2.14, define the colour system

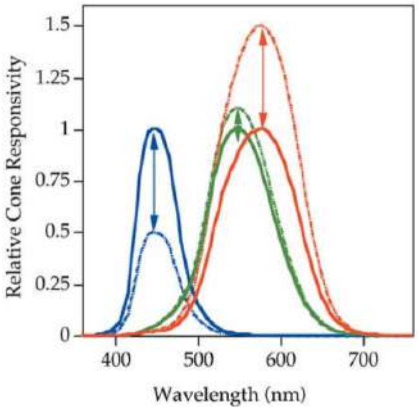


Figure 2.16: Chromatic adaptation. Each cone spectral sensitivity can vary its response independently by scalar multiplication. Image from [Fairchild, 2013].

CIE RGB, where the (R, G, B) tristimulus are expressed as the integral over the visible light of the product of each CMF with the radiance,

$$
\begin{aligned}
R &= \int_{380}^{740} \bar{r}(\lambda)E(\lambda)\mathrm{d}\lambda, \qquad (2.8)\\
G &= \int_{380}^{740} \bar{g}(\lambda)E(\lambda)\mathrm{d}\lambda,\\
B &= \int_{380}^{740} \bar{b}(\lambda)E(\lambda)\mathrm{d}\lambda.
\end{aligned}
$$

The second set of CMF’s $\{\bar{x}, \bar{y}, \bar{z}\}$, see Figure 2.17, are defined as linear combination of $\{\bar{r}, \bar{g}, \bar{b}\}$ under some constraints:

- $\{\bar{x}, \bar{y}, \bar{z}\}$ are always positive,
- $\bar{y}$ is equal to luminosity function $V(\lambda)$, see Figure 2.2.
- $\{\bar{x}, \bar{y}, \bar{z}\}$ are normalised.

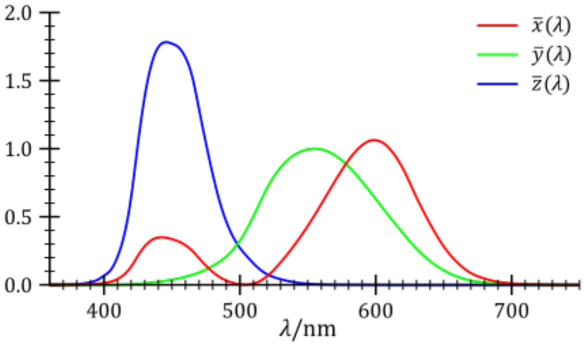


Figure 2.17: The colour matching functions $\{\bar{x}, \bar{y}, \bar{z}\}$. Image from [Wikipedia, 2018a].

Then, the system CIE XYZ is introduced, where the tristimulus values (X, Y, Z) are defined as,

$$\begin{aligned} X &= \int_{380}^{740} \bar{x}(\lambda)E(\lambda)\mathrm{d}\lambda, \\ Y &= \int_{380}^{740} \bar{y}(\lambda)E(\lambda)\mathrm{d}\lambda, \\ Z &= \int_{380}^{740} \bar{z}(\lambda)E(\lambda)\mathrm{d}\lambda. \end{aligned} \tag{2.9}$$

Notice that CIE XYZ system is very common in the industry, since it is device independent. Usually, the perceived colours are defined in terms of *luminance* and *chromaticity*, which it is the concept of colour after discarding the intensity. In this way, CIE XYZ colour space can be presented as CIE xyY by normalising each component by the total intensity,

$$\begin{aligned} x &= \frac{X}{X+Y+Z}, \\ y &= \frac{Y}{X+Y+Z}, \\ Y &= Y, \end{aligned} \tag{2.10}$$

where Y represents the luminance, and x and y the chromaticity. Notice that, if we consider the normalised component for Z, by construction $z = 1 - x - y$,

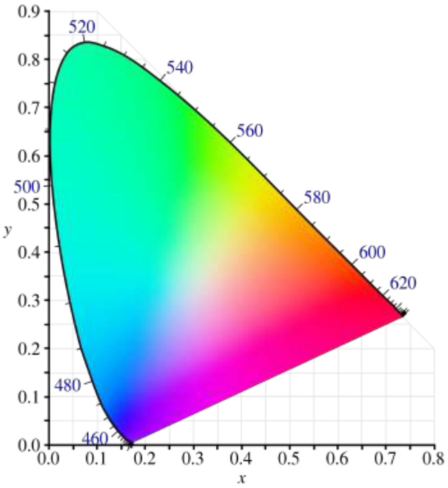


Figure 2.18: The chromaticity diagram. Image from [Solis et al., 2010].

which is a linear combination of the other two components. We present in Figure 2.18 the CIE xy *chromaticity diagram*, which represents all the colours perceived by humans (approximately 10 millions) in a plane. The curve on the boundary represents monochromatic colours (pure spectrum), and it is called *spectral locus* [Sharma, 2002]. The line that closes the tongue-shaped is called the line of purples. Notice that in this line there is no wavelengths notation, since those colours are not spectral colours. A intuitive way of describing colours based on human perception is given by three different characteristics:

- *Hue* is referred to the actual colour, e.g. red, blue, yellow, etc.

- *Saturation* defines how ‘pure’ the colour is with respect to white.

- *Brightness/Value* represents the intensity, the amount of light emitted, and it gives the information of how dark or bright is the colour.

In these terms, we can define a new colour space HSV. This colour space is defined from CIE RGB, by a set of non-linear transformations. Let us define

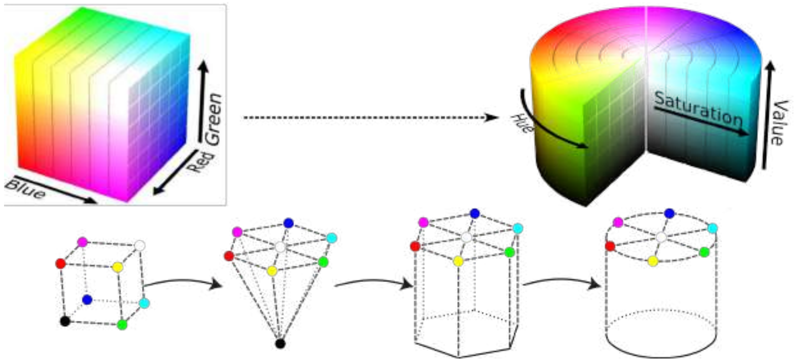


Figure 2.19: The geometric transformations from the RGB cube to HSV representation. Image adapted from [Wikipedia, 2018c, Popov et al., 2018].

$M = \max(R, G, B)$, $m = \min(R, G, B)$, and H' as

$$H' = \begin{cases} 0 & \text{if } M - n = 0. \\ \frac{G - B}{M - m} (\mod 6) & \text{if } M = R. \\ \frac{B - R}{M - m} + 2 & \text{if } M = G. \\ \frac{R - G}{M - m} + 4 & \text{if } M = B. \end{cases} \tag{2.11}$$

Then, the HSV coordinates are defined as

$$\begin{aligned} H &= 360 \cdot H'. \\ S &= \frac{M - m}{M}. \\ V &= M. \end{aligned} \tag{2.12}$$

Notice that HSV description is a chain of geometric transformations from CIE RGB into cylindrical coordinates, as shown in Figure 2.19.

Perceptually uniform colour spaces

The need of defining new colour spaces comes from the fact that neither CIE XYZ, nor CIE xyY are perceptually uniform. This means that the difference between two colours in any of these colour spaces does not correspond to our perceptual difference [Judd, 1979]. In 1976, CIE introduced two colour spaces, the CIE $L^*a^*b^*$ (CIELAB) and CIE $L^*u^*v^*$ (CIELUV). These new colour spaces are defined as follows starting out from the tristimulus values XYZ,

$$L^* = \begin{cases} \left(\frac{29}{3}\right)^3 \frac{Y}{Y_n}, & \text{if } \frac{Y}{Yn} \leq 0.008856 \\ 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16, & \text{otherwise.} \end{cases} \tag{2.13}$$

$$\begin{aligned} a^* &= 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right), \\ b^* &= 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right), \end{aligned} \tag{2.14}$$

where (X_n, Y_n, Z_n) is the tristimulus value of a reference white, and function f,

$$f(x) = \begin{cases} x^{1/3}, & \text{if } x > 0.008856 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 x + \frac{4}{29}, & \text{otherwise.} \end{cases} \tag{2.15}$$

The chrominance channels in the CIELAB definition, (a^*, b^*), can be represented in cylindrical coordinates as,

$$C^* = \sqrt{a^{*2} + b^{*2}},\ h^\circ = \arctan\left(\frac{b^*}{a^*}\right), \tag{2.16}$$

where C^* represents the chroma which is the distance from the origin, and h° the hue angle, which $h^\circ = 0$ represents red, $h^\circ = 60^\circ$ represents yellow, and so on. The colour space defined with the L luminance as in CIELAB, and new coordinates C^* and h° is known as the CIELCh, see Figure 2.20.

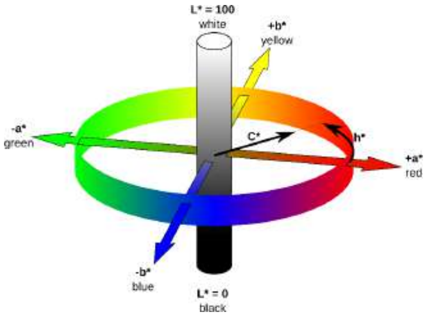


Figure 2.20: CIELAB and CIELCh colour space representations. Image from [Bertalmío, 2014a].

The colour space CIE $L^*u^*v^*$, see Figure 2.21, is defined as the L^* channel as before in Equation (2.13), and the chromaticities u^* and v^* as

$$\begin{aligned} u^* &= 13L^*(u' - u'_n), \text{ where } u' = \frac{4X}{X + 15Y + 3Z} \\ v^* &= 13L^*(v' - v'_n), \text{ where } v' = \frac{9Y}{X + 15Y + 3Z}. \end{aligned} \tag{2.17}$$

The values of u'_n and v'_n correspond to the chromaticity coordinates of the reference white.

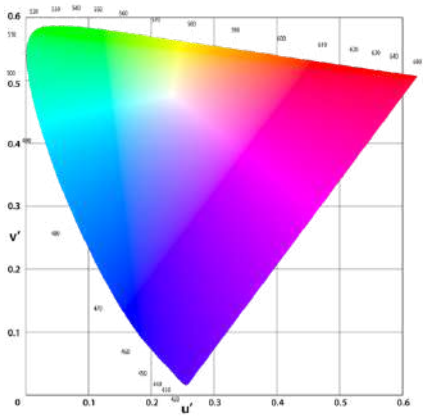


Figure 2.21: The chromaticity diagram of CIELUV (u', v'). Image from [Wikipedia, 2018b].

Although these colour spaces were introduced to overcome the non-uniformity of the previous ones, still they are not fully uniform. Therefore, they are not useful when we want to compare stimuli under different adaptation conditions. For this reason, more complex colour appearance models have been proposed [Fairchild, 2013].

Other colour spaces

The colour space $Y'C_bC_r$ is defined from a non-linear $R'G'B'$ colour space, for example by applying a power law function to each RGB values. Let us introduce the mapping from non-linear (R', G', B') triplet to (Y', C_b, C_r) values

$$\begin{aligned} &Y' = 0.2126 \cdot R' + 0.7152 \cdot G' + 0.0722 \cdot B', \\ &C_r = R' - Y', \\ &C_b = B' - Y', \end{aligned}$$

where Y' is defined as *luma*, and C_r, C_b are the red-difference and blue-difference chroma channels. The definition of these components is more efficient in terms of storage and data transmission. It allows for encoding video sequences, by transferring the luma channel, and compressing only the chroma channels.

The IPT colour space was introduced by Ebner in his thesis in 1998, [Ebner, 1998], and it is defined from CIE XYZ as follows

$$\begin{aligned} \begin{bmatrix} L \\ M \\ S \end{bmatrix} &= \begin{pmatrix} 0.4002 & 0.6075 & -0.0807 \\ -0.2280 & 1.15 & 0.0612 \\ 0 & 0 & 0.9184 \end{pmatrix} \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix}, \\ L' &= \begin{cases} L^{0.43} & \text{if } L \geq 0 \\ -(-L)^{0.43} & \text{if } L < 0 \end{cases}, \text{ idem for } S', M', \\ \begin{bmatrix} I \\ P \\ T \end{bmatrix} &= \begin{pmatrix} 0.40 & 0.40 & 0.20 \\ 4.4550 & -4.8510 & 0.3960 \\ 0.8056 & 0.3572 & -1.1620 \end{pmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix}, \end{aligned}$$

where $(X_{D65}, Y_{D65}, Z_{D65})$ triplets correspond to the normalisation of the CIE XYZ with respect to the white reference point of the standard illuminant *D65*, which refers for average midday light.

In Figure 2.22, we present the different colour spaces introduced above and the connections between them. For example, the mapping from CIE XYZ to CIE xy is defined as a projection; a change of coordinates define the relation between CIELAB and CIELCh. The linear transformations are represented with continuous lines, and the non-linear transformations with dashed lines.

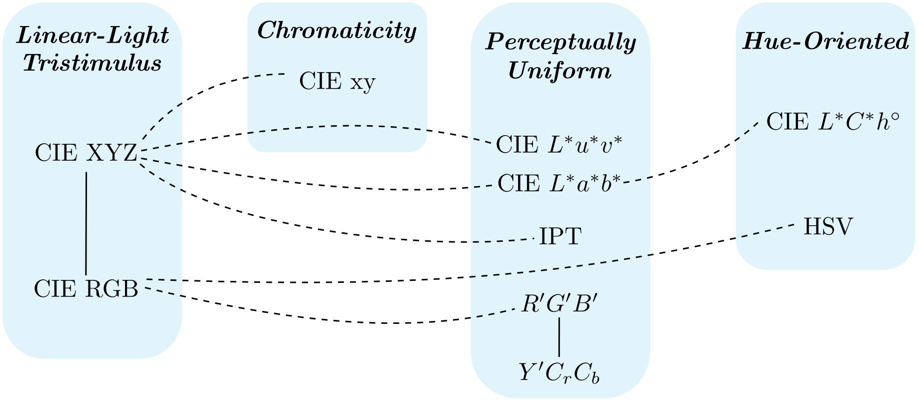


Figure 2.22: The colour spaces presented and their connections. Image adapted from [Poynton, 2003].

Colour metrics

In order to define a distance or similarity between two colours researchers have proposed different metrics. They quantify the notion of differences between colours, and they should account for colour perception in our HVS. Here we present two colour metrics that will be use in the evaluation of following projects.

The CIE first introduced the colour metric ΔE^*_{ab} [Backhaus et al., 1998], as the euclidean distance in the CIELAB colour space. Given two colour points in CIELAB space (L^*_1, a^*_1, b^*_1) and (L^*_2, a^*_2, b^*_2),

$$\Delta E^*_{ab} = \sqrt{(L^*_1 - L^*_2)^2 + (a^*_1 - a^*_2)^2 + (b^*_1 - b^*_2)^2}. \tag{2.18}$$

Since this colour space is not completely perceptually uniform, they introduced modifications on the initial ΔE definition, in order to account for differences in lightness L, chroma C and hue H components. Then, two new metrics were presented: ΔE^*_{94} [Berns, 1993] and ΔE^*_{00} [Luo et al., 2001]. Let us introduce the latter colour metric ΔE^*_{00} given two colour points (L^*_1, a^*_1, b^*_1) and (L^*_2, a^*_2, b^*_2),

$$\Delta E^*_{00} = \sqrt{\left(\frac{L^*_1 - L^*_2}{k_L S_L}\right)^2 + \left(\frac{C_1 - C_2}{k_C S_C}\right)^2 + \left(\frac{H_1 - H_2}{k_H S_H}\right)^2 + \Delta RCH}, \tag{2.19}$$

where $\Delta RCH = R_T \dfrac{C_1 - C_2}{k_C S_C} \dfrac{H_1 - H_2}{k_H S_H}$,

where C and H stand for chroma and hue, and R_T is defined in terms of hue and chroma. For more details on the equation parameters, we refer the reader to [Luo et al., 2001] and [Sharma et al., 2005].

A more recent metric, known as the *Colour-Image-Difference* (CID) was introduced by [Lissner et al., 2013]. It is the colour extension of another metric that accounts for distortions between two gray-level images, the so called *structural similarity index* (SSIM) [Wang et al., 2004]. In CID definition, the colour differences are computed in a colour space introduced previously by the same authors, named LAB2000HL [Lissner and Urban, 2012]. This colour space was defined to be perceptually uniform. The metric accounts for five *image difference features* (IDFs): i) lightness, ii) lightness-contrast, iii) lightness-structure, iv) chroma, and v) hue. For more details on the features definition, we refer the reader to [Lissner et al., 2013].

Figure 2.23: Two images I_1 and I_2 which differ in terms of colours. The corresponding colour metric values setting I_1 as reference: $CID = 0.225$, $\Delta E^*_{00} = 8.50$.

These colour metrics tell us how different the images are in terms of colour, thus the bigger the difference, the bigger the metric value. In case of comparing an image with itself, both metrics would return 0 as value. In Figure 2.23 we show two images that are the same in terms of structure, but differ in colour. The values of the metrics are $CID = 0.225$, and $\Delta E^*_{00} = 8.50$. Notice that the CID metric is bounded by 1, but that ΔE^*_{00} errors can be very large (more than 50).

CHAPTER 3

Digital Image Formation

In this chapter we will describe the functioning of digital cameras, from the light captured in the scene to the final displayed picture. We coarsely divide this process into three main stages: 1) image acquisition, which explains how light reaches the sensor, 2) the transformation of the photons into electrical signals at sensor level, and 3) the in-camera colour processing pipeline, which is a set of steps performed inside the camera in order to obtain the final image.

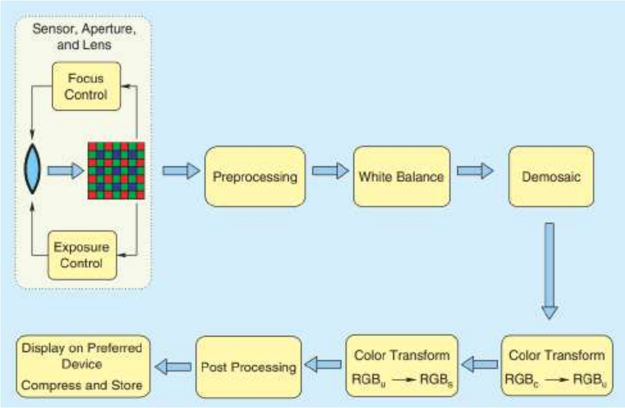


Figure 3.1: Digital camera processing pipeline. Image from [Ramanath et al., 2005].

Figure 3.1 shows 1) and 2) in a single rectangle (pale yellow), and stage 3) is represented in the blue rectangle. The order and techniques used for each step in 3) might differ for different cameras. Notice that, we will focus on the steps that generally cameras performed, since manufacturers do not make available the exact processes of their pipelines.

Image acquisition: Exposure Control

Cameras emulate the process of image formation in our visual system by considering its characteristics. As mentioned in Section 2.1.2, our vision is adapted to luminance conditions (dark and light) to be able to get all the details and information from the scene. Similarly, digital cameras have the option (automatic or manually) to control the amount of light reaching the sensor in order to capture the best possible picture, which is called *exposure control*. Then, once the light passes through the optics of the camera, it reaches the sensor, which it is the equivalent of the retina in our HVS.

Exposure Control

In the case of digital cameras, there are three camera settings that could determine which regions of the scenes will be better displayed in the final image. These settings are the aperture, the ISO value, and the shutter speed or exposure time. They are the so called exposure triangle. In DSLR cameras the metering (light meter) will determine the best values for these parameters in an automatic manner, i.e. the optimised values of these parameters in order to obtain the best well-exposed image. Even so, in manual mode the user is able to set the values for the preferred outcome. With the purpose of understanding their influence in the output image let us define each parameter:

- The aperture controls the amount of light reaching the sensor. It is a part of the lenses, and it works as the iris in the human eyes. The aperture is defined in f-stops that can range from 2 to 22. The larger the number, the smaller quantity of light will pass. Also, it affects the focus, so when the value is large, we will appreciate details that are far from the camera.
- The ISO value determines the sensitivity level of the camera to the light that reaches the sensor. Lower values mean less sensitivity, higher values more sensitivity. The range of values depends on the camera, and it can

vary from 100 to 3200. The larger the ISO value the more noise will appear in the image, which will also affect the dynamic range.

- The exposure time determines how long the sensor is exposed to the light. It is measured in seconds and it can range from $1/500$s to 30s. Larger values of exposure time may show moving objects in the scene as blurred in the captured image.

Figure 3.2 shows the influence in the captured image depending on the settings of these three camera parameters.

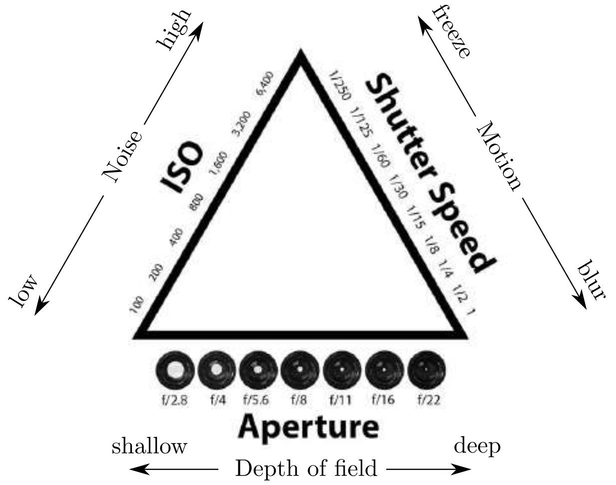


Figure 3.2: Exposure triangle: aperture, ISO value and shutter speed. Image adapted from [PhotographyLife, 2018].

Camera Sensor

The sensor of the camera is a semiconductor device that transforms light (photons) into electrical signals (electrons). It is formed by a matrix of cells, where each cell is known as picture element, *pixel*. These pixels are built using material (silicon) that is sensitive to the visible spectrum of light (380 to 740 nm). In addition, a filter in the optics is used to avoid wavelengths larger than 1100 nm to reach the sensor. The relation between the photons absorbed at pixel level, and the

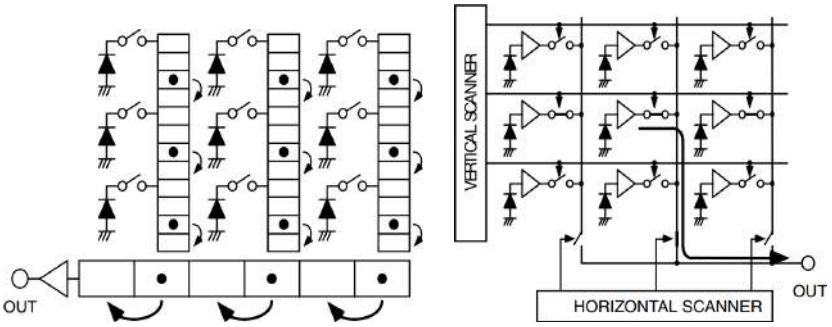


Figure 3.3: Two camera sensor types: on the left a CCD, and on the right a CMOS sensor. Image from [Nakamura, 2005].

electrons generated is linear. Nonetheless, the proportion of absorbed photons decreases under longer wavelengths. The generated electrons in each pixel need to be transferred through a scanning process. Depending on the way of reading the information, we found two types of sensors: charged coupled device (CCD) and complementary metal-oxide semiconductor (CMOS) [Nakamura, 2005].

- CCD sensors transfer the signal vertically all the way down from each column of pixels, then those are transferred horizontally to the amplifier. Finally, the electrical signal is converted to voltage.
- CMOS sensors have an amplifier at each cell location, and the conversion to voltage is done at each pixel at the same time.

Finally the analog voltage is converted into digital values. The range of integer values is defined by the *bit depth* number, e.g. a 12-bit image represents values from 0 to 4095 (in total 2^{12} different values), and a 16-bit image represents values from 0 to 65535 (2^{16}), and so on.

At this point, pixels represent a digital value and do not give any information about colour. In order to represent colour information there exist two main strategies: i) a three sensor system and ii) a colour filter array (CFA) sensor. The former is able to separate the light that reaches the sensor depending on its wavelengths using a beam splitter. Thus, three different sensors are designated for recording short, medium and long wavelengths, see left picture in Figure 3.4. The latter is very popular since it needs only a single sensor. At each pixel location,

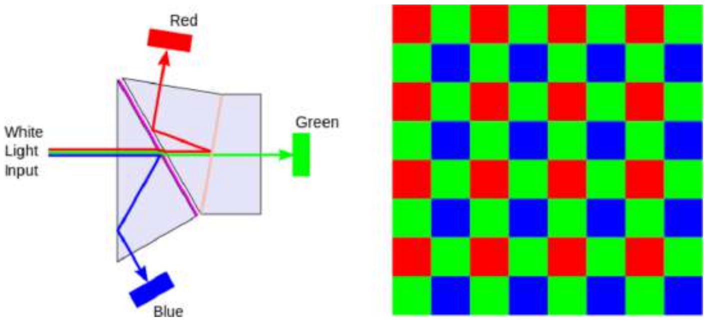
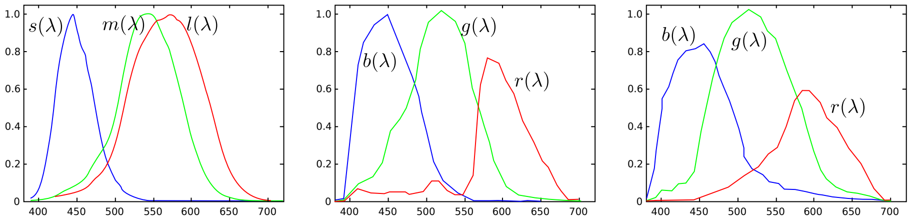


Figure 3.4: On the left: a three-sensor system, and on the right: a colour filter array that follows a Bayer pattern ‘RGGB’. Image from [Bertalmío, 2014a].



a) Cones b) Nikon D5000 c) Canon EOS 500D

Figure 3.5: Comparison of the spectral sensitivity of the cones, and two DSLR camera sensors. On the left: spectral sensitivity of the cones, on the middle: the spectral sensitivity of Nikon D5000 camera sensor, and on the right: the spectral sensitivity of Canon EOS 500D camera sensor. Image adapted from [Wikipedia, 2017] and [Photography, 2018].

a different filter is used to read short, medium and long wavelengths, see right picture in Figure 3.4. The way of arranging the filters follows generally a *Bayer Pattern* [Bayer, 1975]. It is described as a 2×2 mosaic of colour filters, and this mosaic is repeated all over the sensor, e.g. ‘$RGGB$’ would be the Bayer pattern in Figure 3.4. Notice that the green filter is repeated in half of total number of pixels, since HVS is more sensitive to green details, and then red and blue filters are arranged evenly on the rest of pixels. From now on, we focus on CFA sensors.

Each filter in a CFA distribution has its own spectral sensitivity depending on the wavelength, and are usually denoted by $r(\lambda)$, $g(\lambda)$, and $b(\lambda)$. In Figure 3.5 we can compare the cones spectral sensitivity with the spectral sensitivity of two commercial cameras.

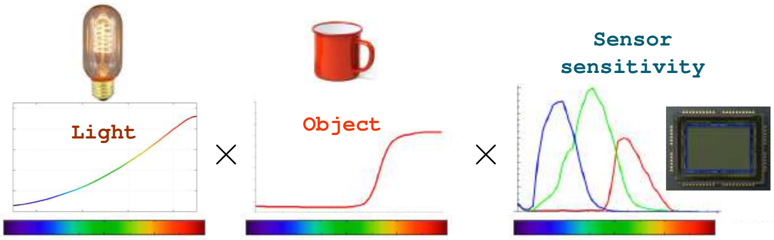


Figure 3.6: The tristimulus that we perceive are the integral of the product of the radiance and the spectral sensitivity of each filter sensor.

The response of the camera sensor can be expressed as a triplet of values (R, G, B), which are defined as the integral of the product of radiance and the filters sensitivity,

$$\begin{aligned} R &= \int_{380}^{740} r(\lambda)E(\lambda)\mathrm{d}\lambda, \\ G &= \int_{380}^{740} g(\lambda)E(\lambda)\mathrm{d}\lambda, \\ B &= \int_{380}^{740} b(\lambda)E(\lambda)\mathrm{d}\lambda. \end{aligned} \tag{3.1}$$

where $r(\lambda)$, $g(\lambda)$, and $b(\lambda)$ are the spectral sensitivity of the filters, and $E(\lambda)$ as mentioned in Chapter 2 is the radiance, the product of the incident light $I(\lambda)$, and the reflectance of the object's surface $R(\lambda)$, see Figure 3.6.

Dynamic Range

The camera sensor dynamic range is defined as the ratio of the maximum number of electrons that can be accumulated in one pixel (full-well capacity N_{sat}), and the read-out noise (n_{read}), which is the noise introduced by the electronics of the sensor (amplifier and analogue-to-digital converter). The DR of the camera sensor is expressed in decibels (dB), and it is defined as follows

$$DR = 20\log_{10}\left(\frac{N_{sat}}{n_{read}}\right) \text{ dB}. \tag{3.2}$$

In Figure 3.7, we present a plot of the relation between the input photons versus the number of electrons produced. In this particular example, Equation (3.2)

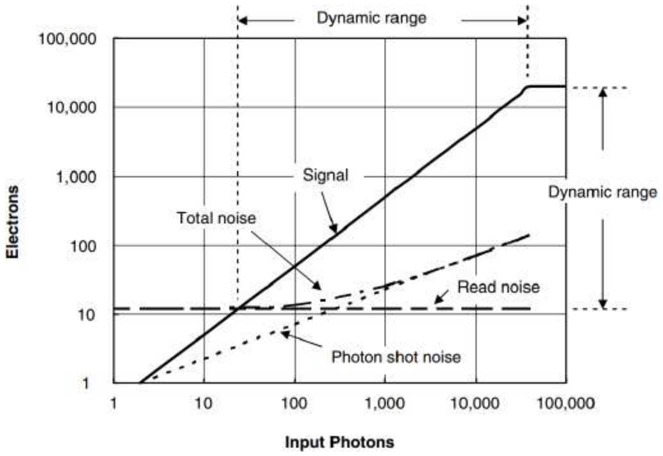


Figure 3.7: The dynamic range of a camera pixel. Image from [Nakamura, 2005].

would be

$$DR = 20\log_{10}\left(\frac{20000}{12}\right) = 64.4 \text{ dB}. \tag{3.3}$$

In order to increase the DR of a camera, we should decrease the read-out noise and/or increase the full-well capacity of the pixel. A high dynamic range scene would need above 100 dB to be able to present in a single picture details in the bright and the dark areas, and this becomes a challenge for most of conventional cameras in which DR is around 70 dB. We will discuss in detail about DR in next Chapter, and the techniques used to increase the DR captured by the camera (both in terms of hardware and software).

In-Camera Colour Processing Pipeline

At this point, the digital values of the sensor can be stored before applying any processing to them. This image can be stored in RAW format, specific for each camera manufacturer, containing the ‘linear’ values read by the sensor. In this way, the RAW image can be processed offline, and it gives more freedom to photographers in order to provide their artistic intent. Nonetheless, not all the cameras have the option of storing RAW information. Hereafter, the image goes through a chain of steps in order to get the final picture to be displayed. These steps might differ from different manufacturers, as well as the order in which they are applied in the pipeline, but we will introduce them as they are shown in Figure 3.1: preprocessing, white balance, demosaicing, colour transform, post-processing, gamma correction and compression.

The preprocessing step refers to *linearise* the data by taking into account the black, as well as the saturation level. For example, a 12-bit RAW image will contain values from 0 to 4095. In reality, those are theoretical borders. We can account for the real black level threshold by taking a picture with the cover of the lens. Ideally, no light is reaching the sensor, thus the image should be completely black (all values 0), although in reality, the image will contain some higher values that will define the black level threshold (we can use a statistical analysis to set the value). In order to obtain the saturation level, we can take several images and determine the maximum value that the images can get. In general this value is below the theoretical value, for example for a 14-bit image (16384 different values), the saturation level might be close to 13800.

White balance

As mentioned already in Section 2.1.2, our HVS does chromatic adaptation in order to perceive the colours of an object the same, even under different illumination conditions. As an example, a white paper under different illumination conditions will still be perceived as white colour. For this reason, cameras need to account and correct the illuminant in order to reproduce white colour objects as white. This process performed inside the camera is called white balance, and in computational photography it is called *colour constancy*. Theoretically, it is a 9 parameter modification (3×3 matrix for a 3 colour filter sensor), but in practice, the estimation of the illuminant is an array of three elements (for each red, green, and blue filters), in order to scale each colour channel in the CFA [Finlayson et al., 1994] and [Vazquez-Corral and Bertalmío, 2014b]. In automatic mode the cameras can estimate the illuminant using different techniques, e.g. colour by correlation from [Finlayson et al., 2001]. In the literature, we find classical approaches such as grey world [Buchsbaum, 1980] or white patch [Judd, 1979], and more complex colour constancy algorithms like [Lam and Fung, 2008], [Vazquez-Corral et al., 2012], and [Akbarinia and Párraga, 2018].

Demosaicing

In the CFA image each pixel gives information of red, green or blue, and in order to obtain a colour image we need to get, for every pixel, the colour information of the red, green, and blue values. Demosaicing techniques produce a 3-channel image from the original 1-channel CFA, as shown in Figure 3.8. The first step is to split the information based on each colour channel, and then apply different techniques to fill the missing colour information. There exist a vast literature on demosaicing algorithms. For a survey of demosaicing methods, we refer the reader to [Li et al., 2008], and [Menon and Calvagno, 2011].

Let us introduce some spatial domain approaches for demosaicing. The authors in [Chang et al., 1999] presented an algorithm based on variable number of gradients (VNG). For each pixel, the gradients for 8 different directions are computed, and a threshold determines which directions to keep for reconstructing the missing colour information. The method of [Malvar et al., 2004] proposed different linear filters depending on the colour to be filled, and its location with respect to the rest of CFA distribution. The filters are defined in a centred 5×5 pixel neighbourhood. In [Zhang and Wu, 2005], the authors used linear minimum mean-square error (LMMSE) estimation in vertical and horizontal directions. Then, both directions are combined to reduce noise, and finally the missing colour information is filled. Due to the fact that demosaicing techniques that take into account directions for interpolation might introduced artefacts, [Hirakawa et al., 2005] presented a method that is able to account for these artefacts and remove them. It is based on homogeneity maps, and those are adaptive and dependent on the image. It is known as the adaptive homogeneity-directed (AHD) demosaicing.

Colour correction

From the previous step, we get a colour image defined in the RGB colour space of the camera sensor. The goal is to obtain an image defined in a standard RGB colour space. As shown in Figure 3.5, the spectral sensitivity response of the colour filters do not match the cones sensitivity. Thus, the (R, G, B) triplet do not match the tristimulus that our HVS would have perceived. This happens because in practice, it is not feasible for the sensors to obtain the responses with the overlap that exists in medium and large cones [Hubel et al., 1997]. For that reason, we need to transform the colours captured by the camera to the perceptual-based

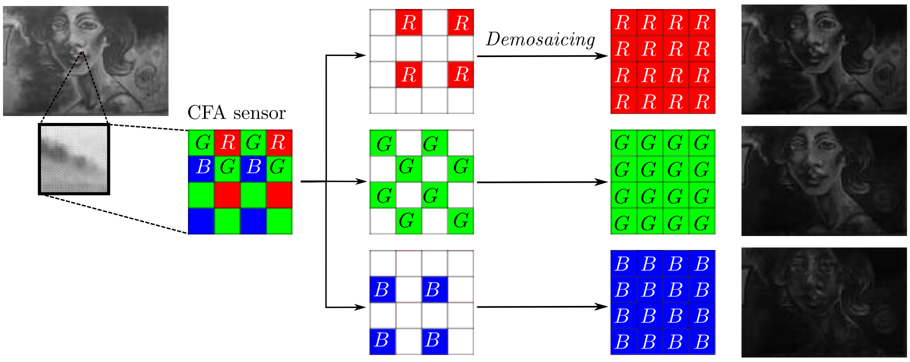


Figure 3.8: Demosaicing framework: split the three channels information, and estimate the missing values for each of the channels. The result is a colour image represented by three arrays of the same resolution as the initial CFA.

colour space CIE XYZ. This colour transformation defines the triplet (R, G, B) as the tristimulus values (X, Y, Z) in CIE XYZ colour space. In Figure 3.9 it is shown, on the left, the image in its RGB camera sensor colour space, and in the middle, the image in CIE XYZ colour space after the transformation. Finally, the tristimuls (X, Y, Z) are transformed to a standard RGB colour space for display purposes. In Figure 3.9, on the right, it is presented the final colour space that will be used for TV displays, monitors, etc.

Summarising, colour correction is defined as a two steps colour transformations: 1) the first one transforms the (R, G, B) values from the camera sensor to the perceptual (X, Y, Z) tristimulus. There exist several approaches to compute this transformation, for example the methods from [Bianco et al., 2007] and [Vazquez-Corral et al., 2014] estimate a 3×3 matrix, and [Finlayson et al., 2015] computes a polynomial to define the mapping. Then, 2) the colour transformation from the CIE XYZ colour space to a standard RGB.

Encoding techniques

The camera encodes the ‘linear’ data by applying different standard transfer functions. In this way, dark areas of the images are expanded, while bright regions are compressed, since HVS is more sensitive to changes in dark areas than in bright ones. These functions, that are applied at camera level and which

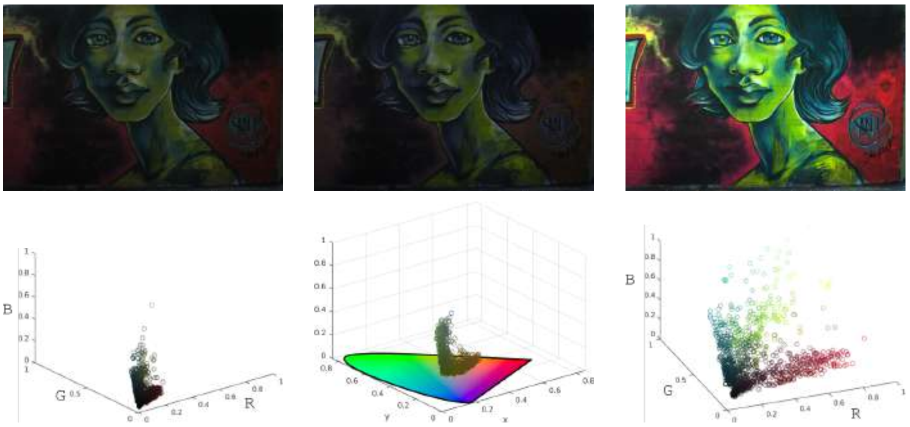


Figure 3.9: Colour transformation from RGB colour space of camera sensor to standard RGB colour space. On the left: the image right after demosaicing, in the middle: the colour image in CIE XYZ colour space, and on the right: the image in $sRGB$ colour space (without gamma correction).

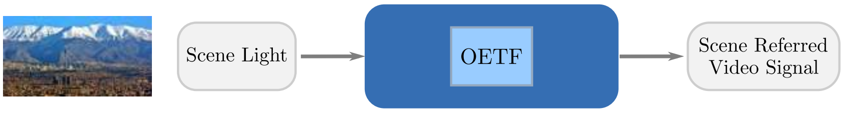


Figure 3.10: Opto-electro transfer function. Image adapted from [Borer and Cotton, 2017].

are scene-referred, are known as *opto-electro transfer function* (OETF), see Figure 3.10. In this subsection we present the *gamma correction* transfer function, and *logarithmic encodings* used in current professional video cameras.

Gamma correction

The original reason for using gamma correction was due to the cathode ray tube (CRT) displays. The relation between the device input voltage and the luminance of the screen for these displays was defined as

$$L = \alpha V^{\gamma}, \tag{3.4}$$

where L is the luminance, α is the proportionality coefficient, V is the voltage, and γ is a power function of a value approximately of 2.4, which is called decoding gamma. For that reason, the output from the camera (L_c) was corrected before

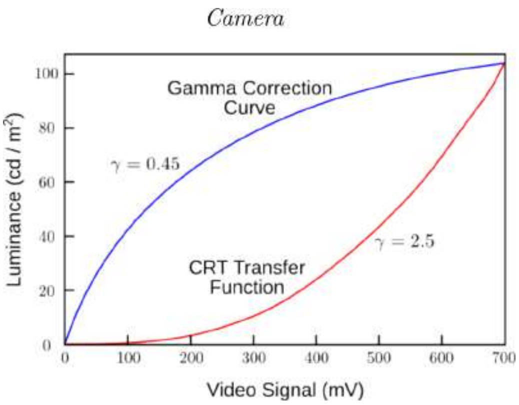


Figure 3.11: The CRT transfer function and the gamma correction curve graphs as functions of the input voltage versus luminance. Image from [Bertalmío, 2014a].

being displayed on the CRT monitor, by applying the inverse of the γ value $V = \beta L_c^{1/\gamma}$, which it is known as encoding gamma [Poynton, 2003]. The graph in Figure 3.11 shows in red the response function of a CRT display, and in blue the gamma correction curve.

Although gamma correction is attributed to the CRT non-linearity, it was also well known that our perception of lightness follows a non-linearity with respect to the luminance in the scene. This non-linear relation is a power law function of value approximately 0.42. This means that our HVS is more sensitive to differences in luminance in dark areas, than in bright ones. While CRT displays are obsolete, gamma correction is still used in the camera output to emulate the perception of luminance in HVS. Figure 3.12 presents the linear output of an RGB image (middle), together with the results of applying a gamma correction of $1/2.2$ (left), and a gamma value of 2.2 (right).

There exist different transfer functions to implement gamma correction. For example, the BT.709 is used for high definition television (HDTV),

$$V' = \begin{cases} 4.5T, & 0 \leq T \leq 0.018 \\ 1.099T^{0.45} - 0.099, & 0.018 \leq T \leq 1, \end{cases} \tag{3.5}$$

where V' is defined as the non-linear gamma corrected values R', G' and B', and T represents the linear R, G and B values. The standard used for screen monitors

Figure 3.12: In the middle the linear RGB image of *Lena*. On the left, the result after applying a gamma correction of $1/2.2$. On the right, the output of using a gamma value of 2.2.

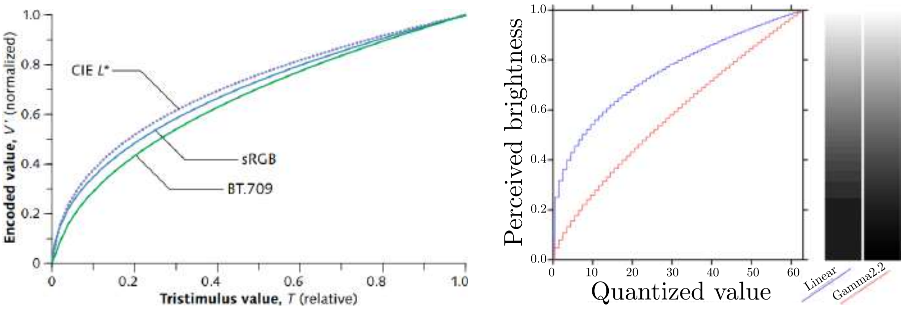


Figure 3.13: On the left the graph of transfer functions BT.709 and sRGB, and lightness from CIE L^*. On the right, the comparison of linear (red) and gamma correction (blue) in their response of quantisation versus perceived brightness. Image from [Poynton, 2003] and image adapted from [Ward,].

and internet is called sRGB,

$$V' = \begin{cases} 12.92T, & 0 \leq T \leq 0.0031308 \\ 1.055T^{1/2.4} - 0.055, & 0.0031308 \leq T \leq 1, \end{cases} \tag{3.6}$$

where V' and T are defined as in the above equation. The left graph in Figure 3.13, we show the transfer functions defined by BT.709 and sRGB, as well as the lightness in CIE L^*.

We should also consider that the output image is encoded using a limited number of bits, for that reason, it would be more convenient to allocate more bits for the dark values, as our HVS is more sensitive to changes in those values, rather than in bright ones. In this way, quantisation errors and noise are minimised. On the right side of Figure 3.13, we show the plots of considering linear encoding (red) and gamma encoding (blue) when the quantisation step is done, versus the perceived brightness. On the intensity bars on the right side of the graph, we can notice jumps on the intensity values in the linear case, rather than on the bar corresponding to gamma, the intensities from black to white follow a smoother transition.

Logarithmic encoding

Log-encoded techniques are commonly used for high dynamic range content. They are defined as logarithmic curves, and each camera manufacturer may have its own definition. In a general form,

$$I_{out} = c \cdot \log_{10}\left(aI_{in} + b\right) + d, \tag{3.7}$$

where I_{out} is the output image, I_{in} is the colour processed image, and the parameters $\{a, b, c, d\}$ are real values. In the next Chapter, we will present in more detail different logarithmic encoding techniques.

Post-processing steps

Each camera manufacturer performs some post-processing operations like: denoising, contrast enhancement, colour artefact removal, etc. In principle, denoising techniques are applied in early stages of the pipeline, since the chain of steps might increase and alter the nature of the noise. All these operations are usually not linear.

Compression and image formats

In order to optimise memory space, the images are compressed in order to reduce their size. There are two types of compression: lossless and lossy. In case of storing the RAW data, lossless compression is used, and the images are presented in TIFF format, or the specific format for each camera, e.g. '.NEF' or '.CR2' for Nikon and Canon cameras, respectively. As a final output, most of digital cameras present a JPEG image, which it is a lossy compression in the frequency domain.

Summarising, the steps in the in-camera colour processing pipeline can be represented by: linearisation, white balance, demosaicing, colour correction and finally, gamma correction. The output of each of these processes are shown in Figure 3.14.

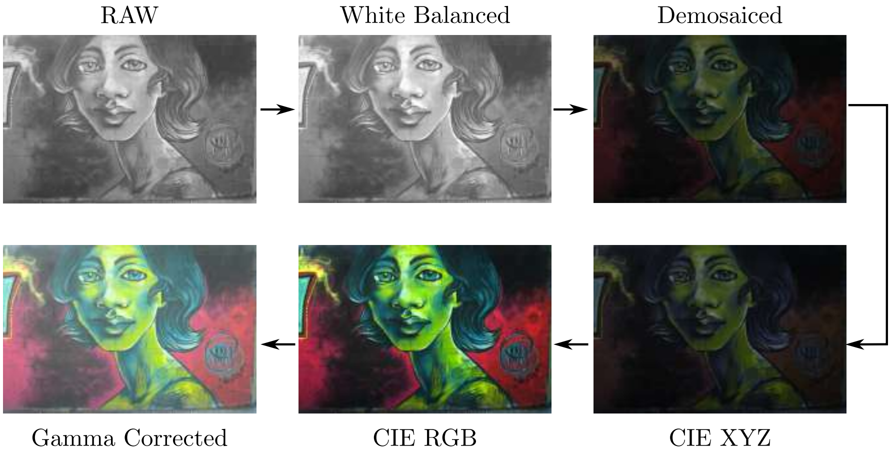


Figure 3.14: In-camera colour processing pipeline. The chain of processes from linear RAW image to the final gamma corrected image.

A general approximation for in-camera colour processing pipeline

Let us consider the generic colour processing pipeline in regular digital cameras as proposed in [Bianco et al., 2012],

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{out} = OETF \left(A \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{in} \right), \tag{3.8}$$

where $[R, G, B]^t_{in}$ corresponds to a raw triplet read by the camera sensor, A is a 3×3 matrix containing information about white balance and colour information, the $OETF(\cdot)$ is applied to the linear coloured image, and $[R, G, B]^t_{out}$ is the output pixel value. We are omitting some steps (like quantisation, contrast enhancement or compression), but it is a good enough approximation: for details see [Vazquez-Corral and Bertalmío, 2014a]. Notice the general definition for $OETF(\cdot)$, which can refer to the non-linearity power exponent $1/\gamma$, or to a logarithmic encoding curve.

CHAPTER 4

High dynamic range acquisition and generation

Dynamic Range

In a real world scene the dynamic range is defined as the ratio between its highest and its lowest luminance values, and may easily span five orders of magnitude or more. This is way above what digital cameras are able to represent. In cameras, the size of the photosite and the way of measuring its content will determine the actual dynamic range of the sensor.

The dynamic range is defined as a ratio, and it can be expressed in f-stops, orders of magnitude, contrast ratio, and decibels (dB). The relation between stops and contrast ratio is a power 2 function (e.g. DR of 10-stops is equivalent to $2^{10} = 1024 : 1$). Orders of magnitude, as mentioned in Chapter 2, are obtained applying a $\log_{10}$ function to the luminance values. Also the common used measure dB is related with the peak signal-to-noise-ratio (PSNR), and it was introduced in Chapter 3. These definitions of the dynamic range differ just in the context they are introduced, as Table 4.1 shows.

Multiple LDR exposures for HDR reconstruction

High dynamic range is an ongoing research, with many references for digital still images. One of the most extended solutions is the combination of a set of LDR images. This idea is based on the fact that varying the exposure time and

Table 4.1: Different dynamic range definitions depending on the application or context were they are introduced. L_{max} and L_{min} refer to maximum and minimum luminance values, N_{sat} stands for the full-well capacity of the sensor, and n_{noise} for the amount of read-out noise.

	Formula	*Application*
f-stops	$\log_2 (L_{max}/L_{min})$	Photography
Orders of magnitude	$\log_{10} (L_{max}/L_{min})$	HDR imaging
Contrast ratio	$L_{max}/L_{min} : 1$	Displays
Decibels	$20 \log_{10} (N_{sat}/n_{noise})$	Digital cameras

fixing ISO and shutter speed, we can capture a stack of images containing details present in bright areas (short exposures), and those present in dim areas (long exposures), as shown in Figure 4.1. However, the idea of combining different ‘exposed images’ was already introduced by the french photographer Gustave Le Gray in the 19th century utilising two negative films. At the time, it was not possible to capture the details of the clouds and the waves of the sea, all in a single image. In many of his photographs, he took two different negatives. For example, in a scene on the beach during the sunset, he would take one negative of the sky, and another of the sea. Then, he would combine both negatives in a single picture by adjusting the exposure to light during development. In Figure 4.2, we can see the final picture of combining two negatives, one for the sky and one for the sea. In a single picture of a high dynamic range scene, both the clouds and the waves appeared in detail.

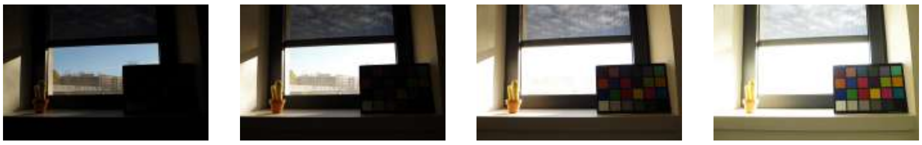


Figure 4.1: Pictures taken with a DSLR (PentaxK50), by fixing aperture and ISO value, and only varying the exposure time. From left to right, exposure time goes from the lowest value, to the largest one $[1/250, 1/60, 1/15, 1/4]$.

In the case of digital cameras, all the methods presented in this section consider the multiple exposure approach to compute an HDR image. The main goal is to estimate the irradiance (the values read by the sensor) for each photograph, since

Figure 4.2: ‘The Brig’, Photograph by Gustave Le Gray, 1856. Image from [Le Gray, 1856].

those values are proportional to the radiance in the scene. Finally, irradiances are combined, e.g. using a weighting function, to build the final HDR image.

Camera response function

The camera response function (CRF), usually referred to as $f(\cdot)$, is defined as the inner camera process that starts from the sensor values (linear) to the 8-bit final image (non-linear), see Figure 4.3 and Equation (4.1).

$$\begin{aligned} Z = f(X), \ & \text{where } X \text{ is the sensor exposure,} \\ & \text{and } Z \text{ are the digital values.} \end{aligned} \tag{4.1}$$

The sensor exposure X is the product of the irradiance and the exposure time $X = E \cdot \Delta t$. Once the CRF is computed, the X values (linear) can be recovered by linearising Z values (non-linear): $X = f^{-1}(Z)$. Usually, we denote $g = f^{-1}$. Then, the irradiance values of each picture E_i can be recovered, this process is

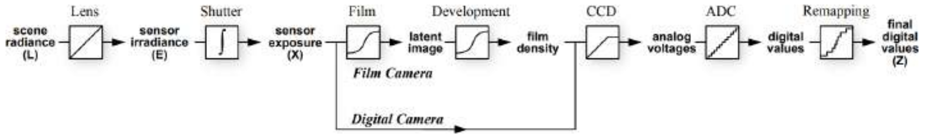


Figure 4.3: The camera response function. The chain of steps from light scene to camera output. Picture from [Debevec and Malik, 1997].

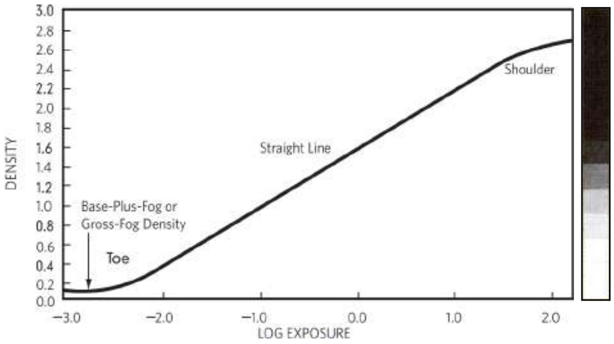


Figure 4.4: The characteristic curve of film: relative exposure versus optical density. Image adapted from [Kodak, 2018].

also known as *radiometric calibration*. Finally combined in a single HDR image by considering a weighted sum of them.

In HDR reconstruction by CRF estimation, the main assumption is that the CRF is *unique*, thus it does not change when the exposure time varies. The uniqueness assumption applies for film cameras. In this case, the CRF is known as *characteristic curve* of the film (exposure vs. film density) and it remains constant when the film is developed under the same set of conditions. In Figure 4.4, we present a characteristic curve of a specific film. In case of digital cameras, the assumption does not apply. One of the main contributions of this thesis is a method that is able to overcome this limitation (see Chapter 5).

The following methods estimate the camera response function for HDR creation. The algorithms can be categorised depending on its CRF estimation. Mann and Picard, in their seminal work [Mann and Picard, 1995], introduced the idea of creating a high dynamic range (HDR) picture of a static scene by combining a set of low dynamic range (LDR) images taken with different exposure times, proposing a parametric method to estimate the CRF. Once the set of images are linearised, the images are combined in a weighted sum, in which the weighting functions are defined from the derivative of the CRF.

The work of Debevec and Malik [Debevec and Malik, 1997] built an HDR *radiance map* from a sequence of multi-exposed images, by estimating the CRF as a non parametric curve. From the definition $X = E \cdot \Delta t$, the goal is to recover the exposure values X, i.e. the camera sensor values, in order to obtain the irradiance values E which are proportional to the radiance of the scene. The CRF is different for each colour channel, and it remains constant over the whole stack. Once each image is linearised, they are merged using a weighting hat function for reducing noise on the irradiance estimation.

[Mitsunaga and Nayar, 1999] proposed to compute the CRF as a polynomial of a maximum order of 10. Once the CRF is computed, the new linearised images are then fused using a weighting function based on signal-to-noise-ratio to obtain the final HDR (they assume the noise is independent of the measured intensity values). For colour images, one response function is computed per channel, and then colours are corrected by using a chromaticity constraint.

In [Robertson et al., 2003], the authors estimated the CRF based on a probabilistic approach. They introduced a noise term in the measured exposure values. This noise term follows a 0-mean Gaussian. The inverse of the noise variance is defined as the weighting function, meaning the highest the confidence, the highest the accuracy in the data.

The work of [Lee et al., 2013] computed radiometric calibration. It assumes the linearity of sensor irradiances to describe the problem as a rank minimisation. They represented the images in a matrix D. They formulate,

$$g \circ D = A, \tag{4.2}$$

where g is the inverse of the CRF and A the matrix containing the irradiances, which should have rank 1. For minimising, they used nuclear norm on matrix A, but only focusing on the second singular value, since this should be 0. Also in the computation, they take care of outliers, as well as exponential ambiguity.

Noise models

The methods in this subsection do not perform an estimation of the CRF, since they work with directly with linear data. They focus on the definition of the weighting functions used for the combination of the multi-exposed images. These weightings are based on noise models in order to discard or not some pixels in the weighting formula. Notice that in these cases, a calibration beforehand is needed to determine the parameters of the noise model.

The paper of [Kirk and Andersen, 2006] introduced a new weighting approach defined by the sensor gain, and the variance of read-out and quantisation noise.

The model works for linear sensors, i.e the recorded values are linearly proportional to the light in the scene. The parameters defined in their formulation need to be estimated in advance by taking several pictures. The main idea is to give more weight to pixel values with lower variances based on the noise model. It compares the performance of different weightings with respect to noise.

Granados *et al.* [Granados et al., 2010] extended the work of [Kirk and Andersen, 2006] by taking into account the spatial noise as well as the temporal noise. The weighting function, defined as the inverse of the exposure values variance, depends on the following parameters: gain factor, gain factor per pixel, mean of dark current noise and variance of read-out noise. If the parameters are not given by the camera manufacturer, it estimates them based on [Janesick, 2001]. It also evaluates the quality of different weighting functions for the HDR reconstruction.

In [Aguerrebere et al., 2014], the authors analysed the performance of different HDR radiance maps generators based on the limits of irradiance estimation. They noticed that almost all methods discard saturated pixels in their HDR reconstruction. For this reason, they studied the utility of considering those pixels, and how they should be considered. Finally, they pointed out the importance of introducing uncertainty for the estimation of the parameters in the noise model, and the authors proposed a new approach.

Exposure fusion

The work of [Mertens et al., 2007] built an LDR image from a set of multi-exposed images. This process is called *exposure fusion*. For each pixel, in every image sequence, three quality measures are computed: contrast, saturation and well-exposed pixels. These measures are used to define a weighting map. The resulting LDR image is a linear combination of these weighted maps together with the respective images using a coarse-to-fine approach. Notice that this method does not compute an HDR image.

In Table 4.2 we present an overview of the above techniques, by terms of colour (C) dependency (independent(I)/dependent(D)), the input linearity (linear/non-linear), exposure time requirement (t_i), the need of a reference image (Ref), and the estimation of the camera response function (CRF). Notice that all the methods

Table 4.2: Characteristics of multiple exposure methods of static scenes for HDR reconstruction.

	C	**Input**	t_i	**Ref**	**CRF**
[Debevec and Malik, 1997]	I	non-linear	✔	✗	✔
[Robertson et al., 2003]	I	non-linear	✔	✗	✔
[Mitsunaga and Nayar, 1999]	I	non-linear	✔	✗	✔
[Lee et al., 2013]	I	non-linear	✔	✗	✔
[Kirk and Andersen, 2006]	I	linear	✔	✗	✗
[Granados et al., 2010]	I	linear	✔	✗	✗
[Aguerrebere et al., 2014]	I	linear	✔	✗	✗
[Mertens et al., 2007]	I	non-linear	✗	✗	✗

working with colour images consider the colour channels independently, in both CRF estimation and for the final HDR reconstruction.

Camera or scene motion

All the above methods considered static scenes and camera motionless. More difficulties arise when there is movement during the acquisition of the multiple exposure scheme. One of the main problems is the appearance of *ghosting artefacts* in the final HDR fused image, for those objects that present movement, see Figure 4.5. In the first row, we can see a person cycling, and in different positions at each exposed image. The last row presents the final HDR (tone-mapped), in which we can appreciate the different positions of the person during the acquisition times. Here we present algorithms that take into account the misalignments between the different exposed images, and reconstruct the final HDR avoiding ghosting artefacts. These approaches can be categorised in: methods that detect explicitly the moving objects/areas, and methods that require only to register the images in the stack.

LDR alignment and HDR reconstruction

Granados *et al.* [Granados et al., 2013] extended their previous work in [Granados et al., 2010] considering moving objects in the scene. It also assumed linear

Figure 4.5: Ghosting artefacts in HDR imaging. A set of multiple exposures in a scene with moving objects (top row) and the resulting HDR image (bottom). Image from [Srikantha and Sidibé, 2012].

sensors and constant illumination. The first step is to globally align the images using SURF [Bay et al., 2008] and RANSAC [Fischler and Bolles, 1981], and to estimate the noise parameters: read-out noise variance, saturation and gain. It computed irradiances values within consistent pixels. The definition of consistency is based on the distribution of the differences between pixel irradiances.

Joint LDR alignment and HDR reconstruction

In the paper [Sen et al., 2012], the authors performed in one step a patch-based energy minimisation to align the LDR images (one image is taken as reference) and reconstruct a final HDR image across multiple scales. It used information from the well-exposed pixels from the reference, and in under/over-exposed areas it considers the information from the other images. They made use of a hat weighting function for the HDR merging in the energy equation. Although the algorithm requires linear data, they apply gamma correction as a preprocessing step.

Ghost detection

In the work of [Gallo et al., 2009], the authors selected a reference image and quoting from the paper ‘the final HDR is the HDR version of the reference’. To do so, they extended the dynamic range of the reference in consistent parts from the information present in the rest of the images in the sequence. Consistency between patches is based on a linear relation between exposure values. When the linear relation is broken it might indicate ghosting values. Then, the information from consistent patches is used to compute the irradiance using a hat weighting function. At the end, they used Poisson blending for each channel independently to obtain the final image.

The authors of [Heo et al., 2011] selected a reference image and globally aligned the rest to it using SIFT [Lowe, 1999] and RANSAC [Fischler and Bolles, 1981]. Ghost detection is done by estimating a global intensity transfer function using probability density functions and energy minimisation (one per colour channel). Then, CRF is computed using [Debevec and Malik, 1997] and refined over the non-ghosting pixels. The final HDR is constructed by combination of three weighting functions: the hat function, one based on colour differences and one based on pixel locations.

The work of [Lee et al., 2014] detected ghosting regions as a rank minimisation problem. It computed the CRF from the method of [Debevec and Malik, 1997] to get the irradiances. The irradiances of each image are represented in a matrix D. They estimated a matrix A which describes the background (static), as well a binary matrix B for describing ghosting artefacts. The final HDR is computed combining the hat weighting function and another weighting function based on matrix B. In case of static scenes, matrix B would be 0, and the method becomes [Debevec and Malik, 1997] approach. Nonetheless, they computed the algorithm for their best performance.

In the paper [Oh et al., 2015], the authors reconstructed an HDR image taking into account camera motion, moving objects and saturated regions. It assumes linear camera sensor response. The following intensity observation model is optimised using rank minimisation,

$$O \cdot h = A + E, \tag{4.3}$$

where O are the observations, h is the homography transformation which corresponds to camera motion, A are the irradiances, and E the moving objects. Then, matrix completion is used to repair missing values from saturated areas in the computed irradiances. Notice that if input images are non-linear, it follows the approach from [Lee et al., 2013].

Table 4.3: Characteristics of multiple exposure methods for camera motion and/or dynamic scenes. τ refers to colour transfer function, (*1) [Debevec and Malik, 1997] method, and (*2) [Lee et al., 2013] method.

	C	**Input**	**t_i**	**Ref**	**Motion**	**CRF**
[Granados et al., 2013]	I	linear	✔	✗	✗	✗
[Sen et al., 2012]	I	linear	✔	✔	✗	✗
[Hu et al., 2012]	I	non-linear	✔	✔	✗	✗ (τ)
[Hu et al., 2013]	I	non-linear	✔	✔	✗	✗ (τ)
[Oh et al., 2015]	I	linear	✔	✔	✔	✗(*2)
[Gallo et al., 2009]	I	linear	✔	✔	✔	✗(*1)
[Heo et al., 2011]	I	non-linear	✔	✔	✔	DM
[Lee et al., 2014]	I	linear	✔	✗	✔	✗(*1)

Colour matching for HDR

The next two references present HDR reconstruction techniques based on colour matching pairs of images as in [HaCohen et al., 2011], followed by exposure fusion of the new computed stack of images.

The authors of [Hu et al., 2012] selected one image as the reference and aligned the rest to it using dense correspondences [HaCohen et al., 2011]. Then, a colour transfer function τ_c is estimated for each pair reference and source, that minimises differences in colour within the calculated correspondences. For missing correspondences they computed a homography from a bounding box discarding the 'unknown area'. Finally, they use Poisson blending to fit all the information without artefacts.

The work of [Hu et al., 2013] selected one image as the reference (R). Then, they computed a new stack of registered images (L_i) that look like R, but as taken under different exposure times. They computed each new image by taken the reference and a source image and estimated a τ_c (intensity mapping function as in [HaCohen et al., 2011] and [Hu et al., 2012]), and a mapping function u (pixel locations) based on Generalised PatchMatch [Barnes et al., 2010]. Notice that τ is different for each channel. This approach does not generate the final HDR image, instead the authors used exposure fusion [Mertens et al., 2007] for presenting the final LDR image.

In Table 4.3 we show the main characteristics of the above algorithms. The notation is the same as in Table 4.2, and in addition, we account for motion detection (Motion). A similar situation is to be found with regards to the previous table: in colour images the channels are computed independently. Majority of the methods for dynamic scenes set up a reference image. Notice that methods which detect motion in the scene, also estimate a CRF. The methods from [Lee et al., 2014], [Gallo et al., 2009], and [Oh et al., 2015] expect linearised inputs, nonetheless they proposed to use [Debevec and Malik, 1997] (the first two), and [Lee et al., 2013] the latter one in case the inputs are non-linear.

High dynamic range sensors and video

In the movie industry there is a growing interest in HDR imaging, but the challenge of shooting HDR scenes using LDR equipment exists since the inception of cinema. The current way to address it relies on adding artificial lights, in order to raise the intensity levels of the darkest parts of the image. In this way the dynamic range of the scene is reduced, fitting it into the range of the capture medium (film or digital). This is a cumbersome, expensive procedure requiring very significant human and material resources that greatly affect the cost of the production. Some alternatives exist but they are not fully practical: some digital cinema camera models are able to alternate exposure times on consecutive frames. As an example, HDRx for RED EPIC and SCARLET video cameras [RED, 2018]. HDRx enables for the acquisition of two frames of different exposure times, one for under and well exposed regions, and another using longer exposure time for highlights. The main drawback of this technique is the appearance of ghosting artefacts when objects move within the scene. Although, they propose the use of optical flow to compensate for motion blur, results might not be as good as expected. A recent possibility is to use a dual-camera set-up [Froehlich et al., 2014], with two synchronised, perfectly registered cameras on an orthogonal rig so that a semi-transparent mirror sends most of the light intensity to one of the cameras, and the rest to the other camera, see Figure 4.6. These images can be fused without problem because they are fully aligned, so there is no risk of ghosting artefacts, but the dual-camera process has limitations: cost and practicality considerations

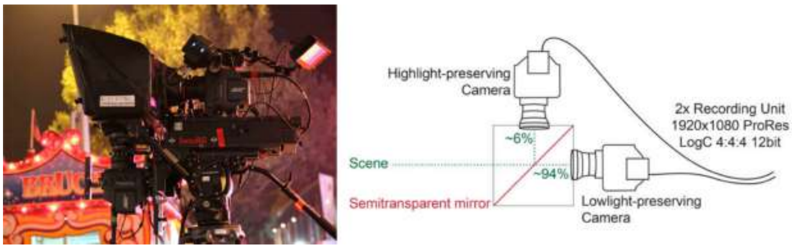


Figure 4.6: Dual-camera set-up in stereo rig composition. Image from [Froehlich et al., 2014].

Figure 4.7: Camera prototype from [Tocci et al., 2011]. On the left the optical set-up, and on the right the camera final prototype. Image from [Tocci et al., 2011].

stemming from the use of two cameras, image problems caused by imperfections in the mirror, and the need to perform tone-mapping to the output. For more details we refer the reader to [Bertalmío, 2014b].

A prototype of an HDR capture system was proposed by [Tocci et al., 2011], see Figure 4.7. The authors presented a three-sensor architecture, each of the sensors recording high, medium, and low exposure images. In the same line, [Kronander et al., 2012] worked with multi-sensors for HDR acquisition, and in one step it performed alignment between sensors, demosaicing and denoising.

In the paper [Zhao et al., 2015], it is presented a camera hardware called unbounded high dynamic range (UHDR). The authors proposed the use of a modulo sensor in order to be able to store high radiance levels, and then recover those values in a post-processing step. They studied two cases: 1) HDR reconstruction from a single image by using graph-cuts-based unwrapping, and 2) HDR from a set of multiple exposures.

[Kang et al., 2003] presented an HDR video generation as a three step problem. The method automatically selected two exposure times to recover the whole scene radiance. The images are registered and the radiance map is computed by estimating CRF using [Mitsunaga and Nayar, 1999]. In [Guthier et al., 2013], the authors presented an HDR video pipeline: acquisition of multiple exposure frames, registration of the multi-exposed images, reconstruction of the HDR frame, and finally the use of a tone mapping operator for video. Based on this HDR video system presentation [Guthier et al., 2013], the authors proposed in [Guthier et al., 2014] a parallelisation of the steps involved in HDR video reconstruction after the multi-exposed images are captured.

More recently, we find a very small number of works that perform HDR reconstruction from a single interlaced image. Gu *et al.* [Gu et al., 2010] combined rows taken with different exposures times, and since the rows are not captured simultaneously, this method produces ghosting artefacts, which need to be reduced by estimating and compensating for the motion-blur.

The camera software Magic Lantern (ML) [Lantern, 2013] allows some camera models to capture image/video with dual-ISO values that alternate between consecutive image line pairs. It provides an implementation to interpolate a full-frame low-ISO image, containing less noise on shadow areas. It does not claim to compute an HDR image, though the final picture is the result of combining the information from both the low-ISO and the high-ISO full-frame images. The method follows a chain of steps: separate the two ISO frames, interpolate the missing lines to get the full images, and combine information from both interpolated frames to highly reduce the noise in dark regions.

Hajisharif *et al.* [Hajisharif et al., 2014] performed at the same time demosaicing, denoising, re-sampling and HDR-reconstruction, starting from the interlaced input provided by the ML software [Lantern, 2013]. This method required a previous radiometric calibration process, therefore it cannot be used when the camera is not available.

A similar idea was developed by Heide *et al.* [Heide et al., 2014], who proposed a single optimisation step using image prior and regularisers of the different inner stages of the camera pipeline (e.g. denoising, demosaicing, etc). The selection of the image prior is crucial for the optimisation process, and the values are highly dependent on the set of images.

These latter two methods have the advantage of working directly with the RAW data without following a staged pipeline, therefore no cumulative errors are carried out from one process to the next; nevertheless, this integration makes it difficult to further extend and add more processes in a single optimisation.

There exist HDR video cameras, like HDRC VGAx from IMS-CHIPS [Hoef-

Table 4.4: Image and video encodings presented in this section.

Images	OpenEXR	linear
	Radiance RGBE	linear
	LogLuv	non-linear
Video	PQ	non-linear
	HLG	non-linear

flinger, 2007], which has a logarithmic response, and Silicon Vision Lars III [Lulé et al., 1999] which exposure time varies for each pixel. Although these cameras can capture a wider dynamic range of the scene, they lack in resolution in comparison with the rest of the techniques to achieve HDR images, for more details see [Myszkowski et al., 2008].

For a deeper analysis, we refer the reader to the book of [Dufaux et al., 2016].

High dynamic range storage

There exist different encoding techniques to store HDR images and video information. In this section we present briefly linear and non-linear encoding techniques for HDR imaging. Table 4.4 details the encodings we present in next subsections.

Linear encodings

The next two linear encodings are used for HDR still images. In 1998, Ward proposed the format *Radiance RGBE*, with extension ‘.hdr’. The encoding is done in 32-bit per pixel. This encoding has also another variant, in which it uses CIE XYZ colour model.

The *OpenEXR* format was introduced by Industrial Light and Magic (ILM) in 2003. The extension of the files is ‘.exr’ and it encodes the channels using 64-, 32- and 16-bit floating point values.

In both formats, floating points are defined using mantissa and exponent, each of them assigned with different number of bits. For this reason, these encodings are not useful for compression.

Non-linear encodings

The relation between our perception of brightness and the light within a scene is *non-linear*, due to the fact that our HVS is more sensitive to changes in low luminance levels, rather than in high luminance levels. This is the reason for introducing non-linear encodings, since they will consider the non-linear behaviour of the HVS and will help to minimise the quantisation errors that might appear in low luminance levels.

Ward introduced the *LogLuv* format for still images. This method separates the luminance and chrominance channels using CIELUV (u', v'), and then a logarithmic function is applied to the luminance channel only. The extension of the files is '.TIFF'. It allocates 16-bit for luminance, and 8-bit for each chrominance channel.

The rest of the encodings presented below are used for encoding HDR video sequences. The *Perceptual Quantizer* (PQ) curve was introduced by [Miller et al., 2012]. It is a display-referred decoding curve, this means that the absolute luminance of the display should be known. It is represented by a modification of the Naka-Rushton Equation (2.4) that was already introduced in Chapter 2,

$$Y = L\left(\frac{V^{1/m} - C_1}{C_2 - C_3 V^{1/m}}\right)^{1/n}, \tag{4.4}$$

where Y is the luminance of the display, V is the input signal with values between 0 and 1, and $L = 10000, m = 78.8438, n = 0.1593, C_1 = 0.8359, C_2 = 18.8516, C_3 = 18.6875$. It was then standardised by the Society of Motion Picture Engineers (SMPTE) [SMPTE, 2014]. The PQ system was defined for cinema, and although it works good for this environment, it does not perform the same in different situations like in TV. One of the main objective of PQ is to minimise the visible artefacts in the quantisation step.

The *Hybrid Log-Gamma* (HLG) was proposed by BBC together with NHK [Borer and Cotton, 2016]. It combines a gamma encoding for small luminance values, and then a logarithmic function for larger values as follows,

$$E' = \begin{cases} rE^{0.5} & \text{if } 0 \leq E \leq 1, \\ a \cdot \log(E - b) + c & \text{if } E > 1, \end{cases} \tag{4.5}$$

where E' is the signal output, E is the input signal, and $r = 0.5$, $a = 0.17883277$,

Table 4.5: High dynamic range datasets.

	# of scenes	Format	RAW	Abs. lum.
SYNS Database [Adams et al., 2016]	92	hdr	✗	✗
Empa HDR [Nemoto et al., 2015]	33	exr & hdr	✔	✗
Fairchild database [Fairchild, 2007]	106	exr	✔	✔ (43)

$b = 0.28466892$, $c = 0.55991073$. The use of a gamma function for low luminance values makes this encoding suitable for standard LDR displays. In addition, HLG is scene-referred and does not contain metadata.

There exist as well different encodings based on logarithmic functions, that are camera manufacturer-based, for example *Log C* from ARRI, *Canon Log* from Canon, *S-Log* from Sony and *Panalog* from Panasonic.

HDR datasets

In Table 4.5 we present some of the most recent HDR datasets. Although there exist many different databases, in most cases it is difficult to get all the information from the scene and acquisition details.

There exists as well an HDR video dataset. In [Froehlich et al., 2014] the authors proposed a dataset acquired with an ARRI Alexa following the dual-camera set-up proposed in [Froehlich et al., 2014]. The content is graded for $4000\ cd/m^2$ displays. They also provided the code needed to transform the linear RAW data to the final logarithmic-encoded data.

HDR colour spaces

The *hdr-CIELAB* and the *hdr-IPT* [Fairchild and Wyble, 2010] colour spaces were defined to overcome the problems that the original versions encountered with HDR images. This is due to the fact that the original spaces require the value of a reference white, see Sections 2.3.2 and 2.3.3. In an HDR image, which can have

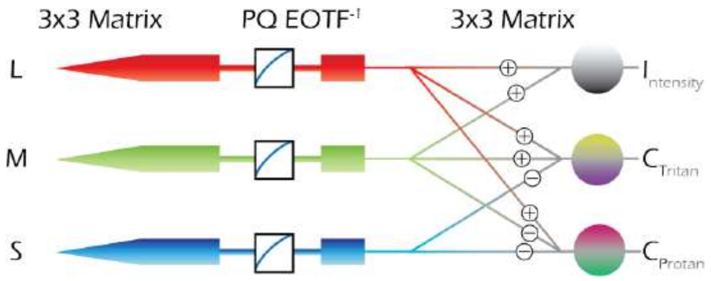


Figure 4.8: Colour representation of colour space IC_TC_P. Image from [Dolby, 2016].

light sources or highlights, this value is difficult to set. These two colour spaces differ from the previous CIELAB and IPT in the definition of the non-linearity applied to them. In the original spaces they were power function laws (L^* in CIELAB, and the cones response tristimulus in IPT). In this case, the authors replaced these power functions by the *Michaelis-Menten* equation, with optimised parameters for each colour space.

The $IC_T C_P$ colour space was introduced by Dolby in 2016. It follows three steps: 1) compute colour transform from Rec. 2020 to L, M, S values, 2) apply non-linearity PQ or HLG to L, M, S, and 3) apply the colour transformation from the non-linear L', M', S' to IC_TC_P colour space, which is inspired by IPT, see Figure 4.8. Notice that, these steps try to emulate the way our HVS processes colours, as explained in Chapter 2.

HDR metrics

In general, the metrics defined to quantify differences between an image pair expect non-linear 8-bit RGB images (LDR) inputs. In the case of HDR evaluation, this assumption is not fulfil, since the images might be encoded using a different non-linearity, or the images might contain linear values in a larger range of luminance values. For these reasons, some HDR quality metrics have appeared for still images and video.

The authors in [Aydin et al., 2008a] proposed to extend the LDR measures PNSR and SSIM [Wang et al., 2004] to HDR images. This is performed by a perceptually uniform (PU) encoding definition. Let us also notice that, non-linear encodings like PQ can also be used to map HDR images to a perceptual uniform space. In 2011, Mantiuk *et al.* [Mantiuk et al., 2011] proposed a new metric for HDR imaging known as HDR-VDP-2. This objective metric accounts for visibility and quality of the images based on a visual model. HDR-VDP-2.2 [Narwaria et al., 2015a] is a newer version of their previous work. Also, HDR-VQM [Narwaria et al., 2015b] metric was introduced for HDR video.

Furthermore, there are in the literature some studies about the performance of different HDR metrics for images and video regarding compression. Hanhart *et al.* [Hanhart et al., 2015a] presented a benchmark of 35 quality metrics, in order to evaluate HDR images compression. The authors performed their study by using 4 different rates of compression in 20 HDR images, and using as ground truth subjective scores. They concluded that the best metrics were HDR-VQM, HDR-VDP-2 and MS-SSIM [Wang et al., 2003].

There exist studies on HDR video compression, as [Hanhart et al., 2015b], [Narwaria et al., 2015c] and [Vigier et al., 2016] to name a few. It is interesting to note that in [Hanhart et al., 2015b], their results on the objective evaluation showed that HDR-VQM has a low performance, whereas PSNR and HDR-VDP-2 are more reliable metrics. In a recent study, [Sugito et al., 2017] analysed the performance of HDR-VQM and BD-rate in HDR video compression. The authors concluded that the ranking of the metrics depends greatly in the non-linearity used for encoding. In addition, they proposed a new encoding definition which performs better than PQ and HLG. This new encoding is based on a tone-mapping curve and it is called natural image statistics transfer function (NISTF).

HDR visualisation: Tone Mapping Operators

HDR images are generally shown in displays with much lower dynamic range, of around 2-3 orders of magnitude. Therefore, it exists the need for compressing the dynamic range of those HDR images to match the dynamic range of the display, and at the same time preserve the details in bright and dark areas, as well as its appearance. Tone mapping operators (TMO) perform this task, see Figure 4.9. There exist vast number of TMOs in the literature, in [Reinhard et al., 2010] we find a study of the algorithms proposed by that time. We can categorise them in global, local, and joint global and local operators.

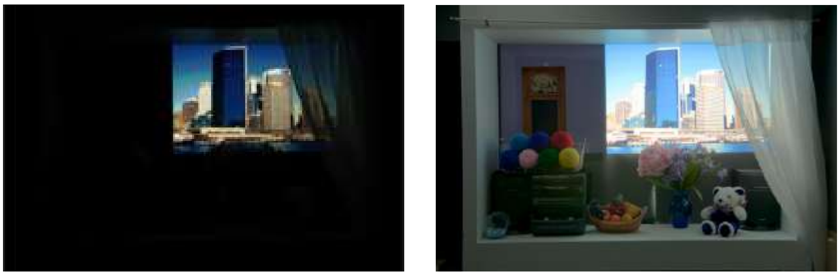


Figure 4.9: Display of an HDR image: linear rescaling, and tone-mapped [Mantiuk et al., 2008].

Global TMOs

Global TMOs are applied to the whole image, and give different values depending on the intensity of each pixel value. The one proposed by Drago *et al.* [Drago et al., 2003] applies different logarithmic curves (varying the base), depending on the luminance values, in order to compress the range of values. For darkest regions, a logarithm of base 2 is considered, and for brightest areas, they consider a base 10 logarithm. The values in between dark and bright values are tone mapped by adjusting the logarithmic curves based on a bias function (power law function).

The method of Mantiuk *et al.* [Mantiuk et al., 2008] defined a piece-wise linear tone mapping curve, and it is based on a HVS model that accounts for visible compression distortions. It can be adapted to different displays and environment conditions. Following the same idea, and considering a more simple HVS model, Mai *et al.* [Mai et al., 2011] proposed to estimate the best tone curve that minimises the difference between the original HDR, and the reconstructed HDR image from the computed tone-mapped image.

The author in [Oskarsson, 2017] redefined the problem of tone-mapping as a clustering problem in the luminance channel. An HDR image contain a large range of luminance values, and thus they estimate the optimal number of clusters (new luminance levels) minimising the differences between histogram bins of the estimated LDR and the given HDR image. The method is extended to colour images and HDR video.

Local TMOs

Local TMOs process each pixel taking into account the pixel’s neighbourhood. Although these TMOs are computationally more expensive than global ones, and results might show finer appearance, some artefacts and halos can appear around the edges of the objects.

In Reinhard *et al.* [Reinhard et al., 2002] the scene luminances are mapped to a specific intensity range, following the idea of the ‘Zone System’ introduced by Adams and Archer in 1939, a system that uses information from the real scene in order to get the final printed version.

In [Ma et al., 2014], the authors estimated an LDR image that best optimises the metric tone mapping index quality index (TMQI) [Yeganeh and Wang, 2013]. This means that no explicit TMO is estimated. In addition, they performed their optimisation getting as an input an already tone-mapped image, and they showed an improvement in terms of TMQI metric.

Joint global and local TMOs

Ferradans *et al.* [Ferradans et al., 2011] presented a two stages operator. The first stage is global and it is based on visual adaptation (human perception); and the second stage is a local approach which performs contrast enhancement. Following the same idea, the work of Cyriac *et al.* [Cyriac et al., 2015a] presented a TMO performing first a global approach and then a local contrast enhancement step.

The authors in [Ahn et al., 2013] proposed a TMO based on the retinex theory. They also performed a two stages method: i) global adaptation for compression, and ii) local adaptation based on edge-preserving filters that reduce the artefacts that might appear from the global adaptation.

Video TMO methods

The algorithms presented above are used in still images, and they cannot be used directly for HDR videos. This is because the algorithms estimate a set of parameters depending on the image, which might be different from one frame to another causing some artefacts like *flickering*. For this reason, algorithms for video will take into account the temporal dimension. For more details on video TMO taxonomy, we refer the reader to [Eilertsen et al., 2017].

In [Aydin et al., 2014], the authors presented a local TMO for video that uses a spatio-temporal filter. They decomposed each frame in two layers: base and detail. The former layer is estimated using a spatio-temporal filter, and the latter by using a temporal filtering.

Eilertsen *et al.* [Eilertsen et al., 2015] proposed a real-time TMO for video, which controls the noise and minimises the distortions. It can be adapted to different displays and environment conditions.

The authors in [Li et al., 2016] presented a global TMO for processing HDR images and video from logarithmic CMOS sensors. It is based on histogram equalisation and a noise camera model. In case of video, they proposed a temporal adaptation in the histogram definitions.

TMO metrics

The quantitative evaluation of tone mapped images is still a challenging problem, since there is not a reference image to compare with. Nonetheless, there exist TMO metrics that account for an objective evaluation.

The tone mapping quality index (TMQI) was introduced by [Yeganeh and Wang, 2013], and it is based on the SSIM [Wang et al., 2004] LDR metric. In [Aydin et al., 2008b], the authors defined the dynamic range independent metric (DRIM) of a pair of images called reference (HDR) and test (LDR). The metric computes three different distortion maps: loss of visible contrast, amplification of invisible contrast, and contrast reversal. The latter refers to contours that have opposite gradient directions in both images.

A recent metric was introduced by [Kundu et al., 2017b]. It is a no-reference image quality metric and it is built on natural scene statistics. The validation of the metric is done using the ESPL-LIVE HDR database for image quality [Kundu et al., 2017a]. In addition, the authors showed the utility of the metric not only for TMO, but also for exposure fusion methods, and post-processing techniques.

HDR reconstruction from single LDR images: inverse Tone Mapping Operators

Displays are increasing their dynamic range capabilities, and the problem of extending LDR content to HDR monitors is becoming popular. Inverse tone mapping operators (iTMO) tackle the problem by increasing the dynamic range of image/video, in order to show it in HDR displays, while preserving the same look and appearance of the image in an LDR display.

In Banterle et al. [Banterle et al., 2006], the authors pointed out limitations of iTMO at the time, the artefacts that methods might introduce, like colour changes, noise enhancement, etc., since no spatial information is taken into account and it is difficult then to identify the light sources. To address these issues, they proposed a method consisting on four steps: 1) to define an initial extended HDR image from the original LDR, 2) to find the highlights and the most illuminated areas from the LDR, 3) to create an expand map from the areas estimated in the previous step, and finally 4) to combine the original LDR and the extended HDR linearly using the expand map.

Meylan et al. [Meylan et al., 2007] presented a method that takes into consideration the light sources and specularities of the image. They first detect the specular highlights in order to compute a mask. Then, a tone scale function is applied in two different ways: 1) globally, using the same curve for all pixels in the image, or 2) taking into account the computed mask and applying the tone curve locally.

In 2009, Masia et al. [Masia et al., 2009] performed two psychophysical experiments. From the first perceptual experiment, they showed that current iTMOs were performing correctly in under-exposed regions, in contrast with over-exposed areas. Thus, they proposed to apply a global gamma value before increasing the DR of the image. The γ-curve plays the role of undoing gamma correction, in order to darken the image and enhance contrast. The γ value can be set automatically. In the second study, they focused on which errors introduced by the iTMOs are perceptually more unpleasant. From [Masia et al., 2009] conclusions, Masia et al. [Masia et al., 2017] presented a global method to compute the gamma value according to image statistics. This new method for computing gamma automatically gets similar results in over-exposed regions to previous work [Masia et al., 2009], while boosting the results in under-exposed regions.

Bist et al. [Bist et al., 2016] presented a model also based on applying a gamma correction. The use of a gamma function model is based on a set of subjective studies they performed. One of their main objectives and contributions is to maintain the aesthetic style of the given content.

CHAPTER 5

Considering the digital camera pipeline for HDR creation

Two main contributions are presented in this section: i) the analysis of the building assumptions on HDR image techniques that consider multiple exposed images, and ii) the proposal of a new method for HDR image creation. As mentioned in the previous chapter, majority of the techniques, that tackle the problem on HDR imaging given a set of multiple exposure images, aim to estimate the camera response function (CRF) that maps the linear values of the camera sensor (RAW) to the final non-linear values of the output image (e.g. JPEG). We will show that the suppositions for its estimation might not apply in case of digital images. For this reason, we propose an HDR method from the definition of the camera colour processing pipeline.

This chapter is based on our work ‘The Intrinsic Error of Exposure Fusion for HDR Imaging, and a Way to Reduce it’, published in *British Machine Vision Conference* (BMVC) [Gil Rodríguez et al., 2015] and its journal extension entitled ‘Issues with common assumptions in HDR imaging from multiple exposures’, which is currently in preparation.

Multiple exposures assumptions for HDR

The dynamic range of light intensities in a natural scene is defined as the ratio between the highest and the lowest luminance values, and it may easily span five orders of magnitude or more. This is way above what digital cameras are able to represent. As mentioned in previous chapters, the camera sensors store values whose are proportional to the light in the scene. Then, these ‘linear’ values go through a chain of processes to get the final non-linear image. As a result, standard cameras are only able to capture distinct luminance ranges by using different exposure times. Thus, short exposure times are used to capture bright areas, while long exposure times are better for capturing dim areas.

Multiple exposure approaches that use non-linear input pictures assume the following image formation model:

$$J(p) = f\left(E(p)\Delta t\right), \tag{5.1}$$

where Δt is the exposure time, p is a pixel location, $E(p)$ is the scene radiance value at p, f is a non-linear transform usually denoted as the camera response function (CRF), and finally $J(p)$ is the stored 8-bit image value, corresponding to one colour channel. Analogous expressions hold for each of the three colour channels, for which the function f might be different. In a static scene the values $E(p)$ remain constant, so taking a stack of N pictures by varying the exposure times gives us for each image

$$J_i(p) = f\left(E(p)\Delta t_i\right), \;\; i = 1, \ldots, N, \tag{5.2}$$

where the subindex i denotes the different exposures *and it is also assumed that the function f remains constant as Δt_i changes*. Multiple exposure methods estimate the inverse g of the CRF f, $g \equiv f^{-1}$, apply it to the image values $J_i(p)$ and then divide by the exposure time Δt_i so as to obtain one estimate of $E(p)$ for each image i in the stack:

$$\frac{g(J_i(p))}{\Delta t_i} = E(p), \;\; g \equiv f^{-1}. \tag{5.3}$$

These N estimates of $E(p)$ are then averaged in order to provide the final output, the HDR value for pixel p.

We can see then how all multiple exposure approaches share a set of building assumptions for the camera capture:

1. Different colour channels are independent.
2. The camera response remains constant while changing the exposure.

3. If the exposure times vary sufficiently, we can recover the full dynamic range of the scene and the resulting HDR image will be a *"radiance map"*, as Debevec and Malik put it [Debevec and Malik, 1997], i.e. the HDR image values will be proportional to the scene radiance values.

The contributions of this chapter are, on one hand, to highlight how these assumptions do not hold in general, and then to propose a method to improve multiple exposure combination, compensating for the violations of assumptions (1) and (2) above and allowing us to obtain more precise HDR images from non-linear LDR inputs. Our proposed method matches the colour and gamma correction transforms of all pictures to those of a reference picture in the stack, adapting the techniques introduced in [Vazquez-Corral and Bertalmío, 2014a] for colour matching and [Vazquez-Corral and Bertalmío, 2015] for gamma estimation, and merges all the linearised colour-corrected images with the linearised reference image obtaining an intermediate HDR result. This is done for several reference pictures and the resulting HDR images are averaged to produce the final output, which is shown to be more accurate to the ground truth HDR image (derived directly from RAW pictures) than those obtained with state-of-the-art multiple exposure combination methods, according to several image metrics. Also, the tone-mapped versions of the HDR pictures obtained with our method show no hue shifts, colour artefacts or contrast problems, which are issues that commonly crop up with other multiple exposure approaches.

Finally, let us also note that although the experiments on this chapter deal with static scenes, our method can potentially be extended to more generic situations like non-static scenes with moving cameras, as the framework used for colour and gamma stabilisation does not require image registration.

The effect of optical scattering on the dynamic range

The effect of scatter or glare in the context of HDR imaging was pointed out and extensively studied by McCann and Rizzi, see [McCann and Rizzi, 2011] and references therein, where they show through practical examples how it is not fully appropiate to refer to the outputs of multi-exposure HDR imaging techniques as *radiance maps*, since they are not proportional to the radiance values of the scene, and also it is wrong to assume that we can simply, without any further considerations, use multi-exposure HDR methods to recover the full dynamic

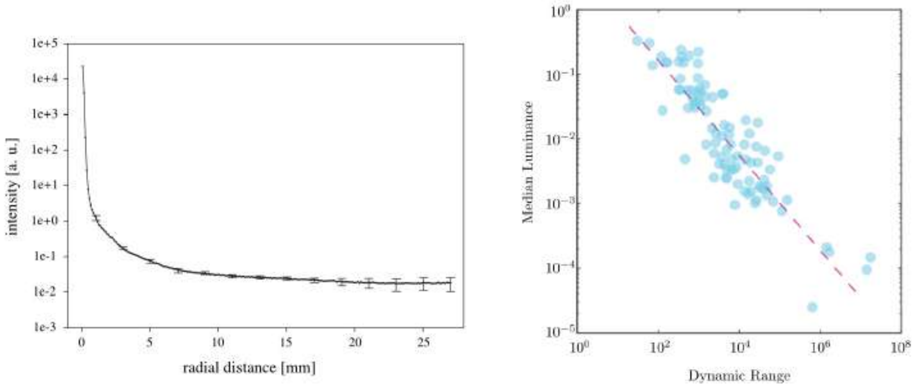


Figure 5.1: Left: example PSF of the optics of a CCD camera, from [Pilz et al., 2008]. Right: linear fit of normalised median vs. DR in the HDR Survey data [Fairchild, 2007], in log-log coordinates; this fit is consistent with Equation (5.8) and implies a value of $c = 0.48\%$ for the veiling glare coefficient.

range of any given scene, because veiling-glare limitations cannot be overcome by taking any number of different exposures.

Here our contribution is to propose a simple yet effective mathematical model, presented later in Equation (5.8), that given an HDR scene and a basic property of the camera lens estimates what is the effective dynamic range that can be recovered.

Consider an image I representing actual luminance values in a scene, as measured with a glare-free photometer. We define the dynamic range DR of the scene as the ratio

$$DR(I) = \frac{I_M}{I_m}, \tag{5.4}$$

where I_M, I_m are the maximum and minimum luminance values respectively. Now if we take a picture of the scene using a photo camera, the effects of reflections and scattering in the optic system can be approximately modelled by a convolution with a point spread function (PSF) S having a shape as shown in Figure 5.1 (left), so that the light $\hat{I}$ that actually reaches the sensor has the form

$$\hat{I} = I * S. \tag{5.5}$$

The slow decay of S implies that the light reaching each image point is effectively contributing a small fraction of its intensity to every other point in the image, regardless of how far apart they are: this is the *veiling glare*. Clearly, the brighter an image is, the higher the glare it will have. Given the linearity of the convolution operation and assuming a very long tail for S, $\hat{I}$ can be roughly approximated as the original irradiance I plus a homogeneous veiling glare component vg that can be estimated as a percentage c of the mean value μ of the image:

$$\hat{I} \simeq I + vg = I + c\mu, \tag{5.6}$$

and therefore the dynamic range for $\hat{I}$ is:

$$DR(\hat{I}) = \frac{max(\hat{I})}{min(\hat{I})} = \frac{I_M + c\mu}{I_m + c\mu}. \tag{5.7}$$

We see that glare is increasing the minimum level of the original scene to a value above $c\mu$ that in practice can be quite large[1]: it increases with the mean luminance μ, and c is no smaller than 0.25% [Tomić et al., 2014]. The net result is that *the DR of $\hat{I}$ is usually much lower than that of the original scene data I, and the darkest areas in $\hat{I}$ will have values that will usually be way off from the actual values as measured on the original scene.* This is due to the optics, it has nothing to do with the dynamic range capabilities of the camera sensor, and it does not depend either on the exposure time that is actually used to capture the picture, so these problems cannot be solved by taking different exposures, regardless of the number of them. We remark these implications because we feel that this very important information is usually missing from works on HDR imaging, and in particular from the multi-exposure literature.

From Equation (5.7), we propose the following practical model to estimate (and bound) the dynamic range for $\hat{I}$, i.e. the effective dynamic range that can be captured:

$$DR(\hat{I}) \simeq \frac{I_M + c\mu}{c\mu} \simeq \frac{I_M}{c\mu}. \tag{5.8}$$

To validate this model we have tested it on the HDR Survey dataset by Mark Fairchild [Fairchild, 2007], containing 105 different scenes acquired with the same camera. Equation (5.8) implies that a plot of normalised mean vs. DR, in log-log coordinates, could be fit by a line of slope -1, and this is indeed the

---

[1] Furthermore, this is a conservative estimate because the model in Equation (5.5) is the glare contribution from light sources outside the field of view.

case as Figure 5.1 (right) shows[2]. Furthermore, the linear fit provides us with an estimate for the veiling glare coefficient: $c = 0.48\%$. This value is very close to the one that can be estimated directly on the *LuxoDoubleChecker* image (for which M. Fairchild provides photometer readings for all regions in the picture), that yields $c = 0.44\%$.

With this very simple model, knowing the veiling glare coefficient c of a lens [Wüller, 2016] and having estimates for the maximum and mean luminance values of the scene, we can roughly estimate the effective DR than can be captured. For instance, for the *LuxoDoubleChecker* image, while the actual DR of the scene as measured with a photometer is 4×10^6, the effective DR as measured in the captured RAW data is of 3.3×10^5, and the estimate given by Equation (5.8) is very close, of 3.05×10^5. It may come as a surprise that, in many situations where we want to photograph an HDR scene (e.g. backlit images, scenes with high global contrast, etc.) and where it is assumed that a multi-exposure HDR method is required, the effective DR as estimated with Equation (5.8) quite often turns out to be enough for single-exposure photography. For instance, consider this numerical example: an indoors scene with some shadows where the minimum luminance is 1 cd/m^2, a window that covers 10% of the image area and through which outdoor light comes in with a maximum luminance of 10^5 cd/m^2, while in the remaining 90% of the image the average luminance is 100 cd/m^2. The camera lens has a veiling glare coefficient of $c = 0.25\%$. The dynamic range of the scene is therefore 10^5, but the dynamic range of the image produced by the optics is only:

$$DR(\hat{I}) \simeq \frac{10^5}{0.0025 \times (0.9 \times 100 + 0.1 \times 10^5)} = 4 \times 10^3, \tag{5.9}$$

so the scene is apt to be captured with a single exposure using a regular camera sensor, having at least 12-stops of latitude, as is the case with most current cameras.

The response function of digital cameras

The first and second assumptions enumerated above, namely, that (1) different colour channels are independent, and (2) the camera response remains constant

---

[2] For this plot we have used the median instead of the mean so as to avoid outliers and obtain a less biased fit.

while changing the exposure, made sense for film photography, but are not an accurate model of how digital cameras work. Digital cameras follow a typical camera colour processing pipeline [Bianco et al., 2012] that can be expressed as

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{out} = \left(A \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{in} \right)^{\gamma}, \tag{5.10}$$

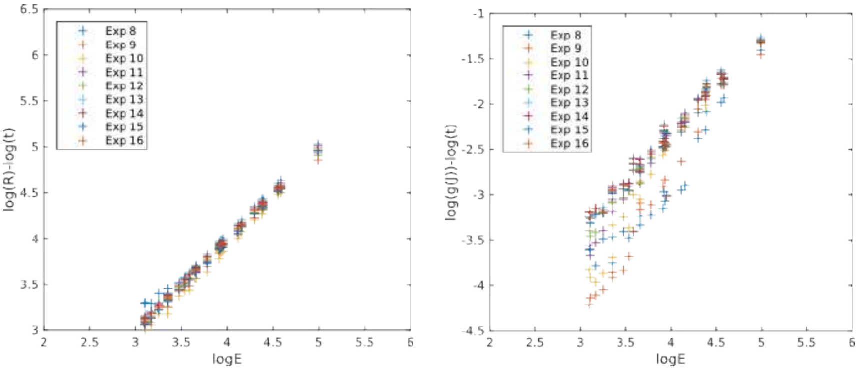
where $[R, G, B]^t_{in}$ is the sensor triplet (usually in 12 or 14 bits), $[R, G, B]^t_{out}$ is the pixel value at the end of the pipeline (in 8 bits per channel), A is a 3×3 matrix that combines the different colour channels taking into account white-balance, colour encoding, colour characterisation and a gain value, and γ is a value, typically between $1/1.8$ and $1/3$, performing gamma correction (notice that we omit demosaicing, denoising, compression, etc: for a complete explanation of these pipeline processes see [Bertalmío, 2014b]).

The implications, for the above assumptions, of using this more realistic model are the following:

1. The three channels R, G, B are not independent, because the matrix A is not diagonal as it incorporates colour processing steps like colour characterisation that involve all channels.
2. If, while taking the N differently exposed pictures, there are changes in the A matrix (e.g. because of changes in gain or white balance) or changes in the γ value, the pictures will be affected as if the non-linear transform f in Equation (5.2) changed from image to image in the stack, i.e. changes in A or γ imply that f is not constant.

Let us elaborate a little on this latter point.

Figure 5.2: Sixteen individual exposures used by M. Fairchild in [Fairchild, 2007] to create the *LuxoDoubleChecker* HDR image. Figure adapted from [Fairchild, 2007].



(a) (b)

Figure 5.3: Plots computed for points in the coloured squares in the dark colour checker in Figure 5.2, using (a) the RAW values, and (b) the JPEG values. While in both cases the plots should theoretically be a single line of slope one, it can be seen that in the JPEG case the points are more dispersed than in the RAW case.

Figure 5.2 reproduces the 16 individual exposures used by M. Fairchild in [Fairchild, 2007] to create the *LuxoDoubleChecker* HDR image. The images were captured in RAW format, alongside the non-linearly corrected counterparts. For a stack of N RAW pictures R_i the image formation model is:

$$R_i(p) = E(p)\Delta t_i, \;\; i = 1, \ldots, N, \tag{5.11}$$

and this equation is valid for the range of luminances for which the sensor operates in the linear range, above the black pedestal and below saturation. This is why, when creating an HDR image through multiple exposure combination, professional users prefer to take RAW pictures; in this way, there is no need to estimate and invert the CRF that is applied to the non-linearly modified pictures stored in 8 bits per channel form. Applying the logarithm to both sides of Equation (5.11) and leaving only the exposure term on the right we get

$$\log\left(\frac{R_i(p)}{\Delta t_i}\right) = \log(E(p)), \;\; i = 1, \ldots, N, \tag{5.12}$$

therefore if we plot $\log\left(\frac{R_i}{\Delta t_i}\right)$ versus $\log(E)$ we should get a single line of slope one. This is indeed approximately the case, as we can see in Figure 5.3(a).

In principle the same could be said in the non-linear case when applying the logarithm to Equation (5.3):

$$\log\left(\frac{g(J_i(p))}{\Delta t_i}\right) = \log(E(p)), \;\; i = 1, \ldots, N, \tag{5.13}$$

because if we plot $\log\left(\frac{g(J_i)}{\Delta t_i}\right)$ versus $\log(E)$ we should also get a single line of slope one. In practice, though, this does not always happen, as Figure 5.3(b) shows. The fact that the values for $\log\left(\frac{g(J_i)}{\Delta t_i}\right)$ are rather spread implies that it was wrong to assume that f (as well as its inverse g) was constant, and therefore the conclusion is that the camera must have modified the values for some of its parameters, $A, \; \gamma$, when the exposure time Δt_i is changed.

To support this claim and highlight how generalised this camera behaviour is, we have performed tests on multiple exposure sequences coming from four different camera models, where during capture only the exposure time changed, and with results recorded both in linear (RAW) and non-linear (JPEG) form. Having the same picture in these two versions allows us to estimate the values for γ and the matrix A with Equation (5.10), using the RAW data for the $[R, G, B]^t_{in}$ values and the JPEG data for the $[R, G, B]^t_{out}$ values.

The first column of Figure 5.4 plots, for different sequences taken with four different camera models, a value that measures how far the matrix A of each image in the sequence departs from being diagonal: we have chosen for this to compute the average of the absolute value of the non-diagonal elements of A normalised by its maximum value. The fact that these values are consistently above 0.1 shows that the three channels $R, \; G, \; B$ are not independent. The second column of Figure 5.4 plots the value of $1/\gamma$ for each image in the sequence, which is ordered from shortest to longest exposure time. We can see that, for all sequences, as the exposure time increases the value of $1/\gamma$ also increases, and the change is quite substantial. The third column of Figure 5.4 plots the difference between the matrix A of each image in the sequence with respect to the matrix A of the middle-image in the stack[3]. Again we see that the cameras are changing A from one exposure to the next.

In Figure 5.5, each column corresponds to a camera model from a different camera maker. The rows show tone-mapped results of the HDR pictures obtained

---

[3] We compute this difference as the Frobenius norm $||\cdot||_F$ of the difference between the matrices.

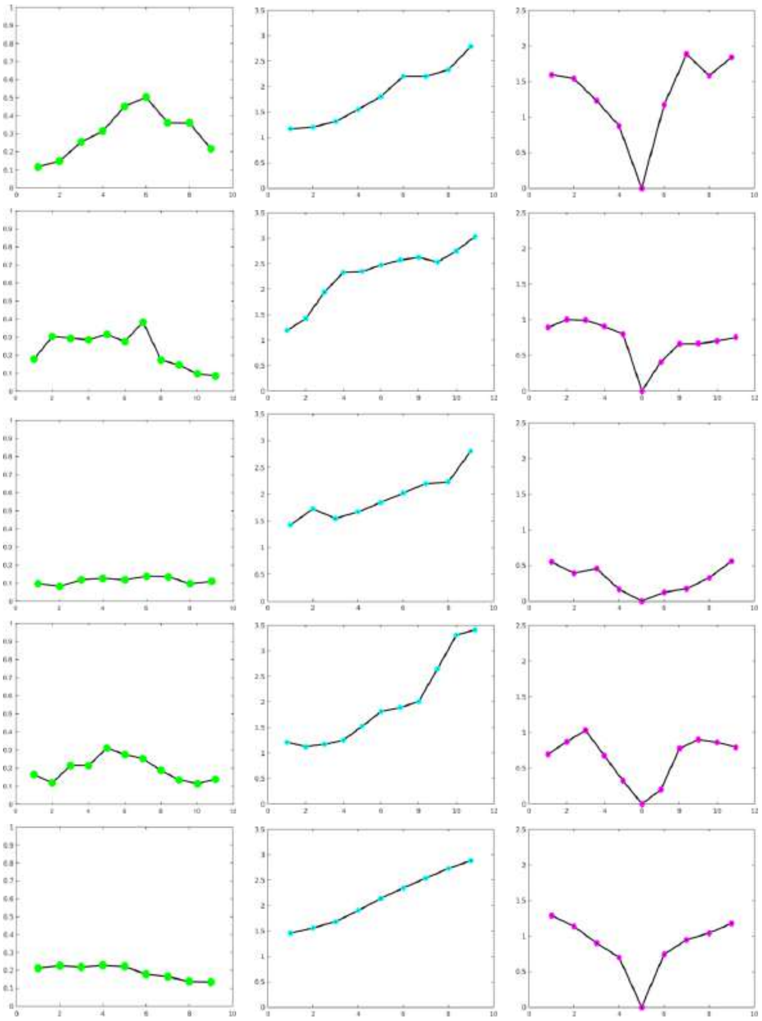


Figure 5.4: Rows 1 to 4 correspond to different sequences taken with different cameras, shown in columns 1 to 4 of Figure 5.5. The last row corresponds to the average over the HDR Survey database [Fairchild, 2007]. First column: average value of non-diagonal elements of A for each image in the sequence. Middle column: gamma-correction value for each image in the sequence. Last column: difference between colour correction matrix of each image in the sequence with respect to colour correction matrix of reference image (middle-exposure).

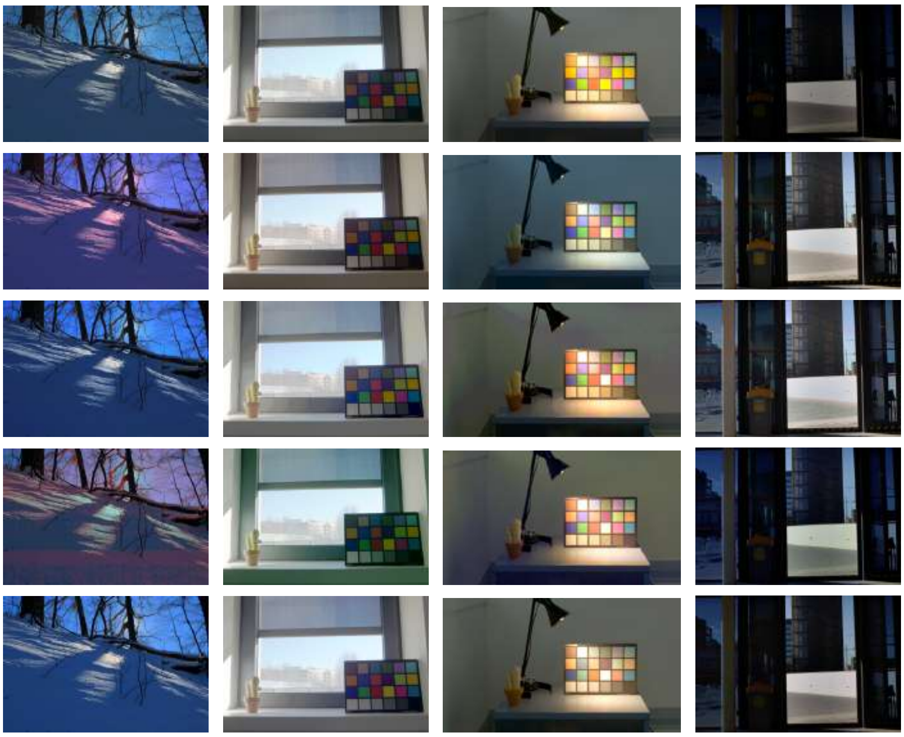


Figure 5.5: From top to bottom: tone-mapped HDR result obtained from RAW pictures (i.e. ground truth) (first row), tone-mapped HDR result obtained with the multiple exposure combination methods of Debevec and Malik [Debevec and Malik, 1997] (row 2), Mitsunaga and Nayar [Mitsunaga and Nayar, 1999] (row 3), Lee *et al.* [Lee et al., 2013] (row 4), and with our proposed approach (last row). Cameras used: Nikon (first column, from the HDR Survey [Fairchild, 2007]), Pentax (second row), Panasonic (third column), Canon (fourth column). Tone mapping methods used: Drago *et al.* [Drago et al., 2003] (column 1), Mai *et al.* [Mai et al., 2011] (column 2), Mantiuk *et al.* [Mantiuk et al., 2008] (column 3), Ferradans *et al.* [Ferradans et al., 2011] (column 4).

with different multi-exposure combination methods: from the RAW pictures, which would be the 'ground truth' or the best result we can aim for (first row), from the JPEG pictures using the multiple exposure combination methods of Debevec and Malik [Debevec and Malik, 1997] (row 2), Mitsunaga and Nayar [Mitsunaga and Nayar, 1999] (row 3), Lee *et al.* [Lee et al., 2013] (row 4), and the method proposed in this chapter (last row). We can see that the previous multiple exposure

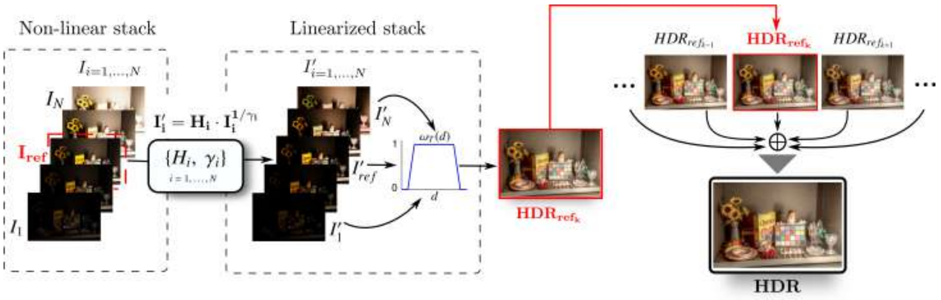


Figure 5.6: Diagram of our method. The input is a set of N non-linear images acquired under different exposure times. We select P reference images; for each reference image I_{ref_k}, all images in the stack are colour-matched and linearised with respect to I_{ref_k} (colour stabilisation [Vazquez-Corral and Bertalmío, 2014a]) and then averaged to produce an intermediate HDR picture HDR_{ref_k}. Then, these P intermediate HDR pictures are averaged to produce the final output.

methods that take the non-linear JPEG inputs produce results which have visible problems, like hue shifts and colour artefacts; to underline that these artefacts are not due to the particular tone-mapping method used, each column employs a different, state-of-the-art tone-mapping algorithm.

In the next section we will introduce a method that, considering all three channels simultaneously, removes the fluctuations in γ and A, effectively making the CRF constant for the whole sequence. Its results for the above sequences are shown in the last column of Figure 5.5.

Proposed method to make the CRF constant

A schematic of our method is presented in Figure 5.6. The input is a set of N non-linear LDR images, static or dynamic, acquired under different exposure times, and we select P of them (normally centred around the middle exposure) as reference images. Our algorithm consists of two steps:

- **Step 1**. For each reference image I_{ref_k}, all images in the stack are colour-matched and linearised with respect to it.
- **Step 2**. The linearised images are averaged to produce an intermediate HDR picture HDR_{ref_k}. The P intermediate HDR images HDR_{ref_k} for

$k = 1, ..., P$ are combined to produce the final HDR output.

We have made the code for our implementation available at `http://ip4ec.upf.edu/HDR_code`. Let us now see in detail the steps of the proposed method.

Step 1

The colour-matching process is key to our method, because it is the one that removes fluctuations in γ and A, and in practice turns the CRF constant for all images in the multi-exposure sequence. The basic idea was introduced in [Vazquez-Corral and Bertalmío, 2014a] for the particular problem of colour stabilisation among different shots of the same scene taken with different cameras: using the model of the camera colour processing pipeline of Equation (5.10), find the parameter values so as to obtain an optimal match to a reference image. While there are works on multi-exposure HDR creation that match geometry and/or colour to a reference image in the stack (e.g. see [Hu et al., 2013] and references therein), the novelty of our approach lies in the use of a camera processing model, which allows our results to be more accurate, as we will see in Section 5.5.3.

Given the reference image I_{ref}, for each other image I_i in the sequence we do the following. Let p be a scene point appearing both in I_{ref} and I_i, which therefore produces in the sensor the same irradiance triplet $[R, G, B]_p^t$ in both pictures. Allowing for camera and/or object motion, point p may appear at different locations p_{ref} and p_i in both pictures, and the pixel values $[R, G, B]_{p_{ref}}^t$ and $[R, G, B]_{p_i}^t$ are also different:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{p_{ref}} = \left(A_{ref} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_p \right)^{\gamma_{ref}} ; \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{p_i} = \left(A_i \begin{bmatrix} R \\ G \\ B \end{bmatrix}_p \right)^{\gamma_i}$$

These equalities can be combined into a single equation:

$$\left(\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{p_{ref}} \right)^{1/\gamma_{ref}} - H_i \left(\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{p_i} \right)^{1/\gamma_i} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (5.14)$$

where H_i is the unknown 3×3 matrix $H_i = A_{ref} A_i{}^{-1}$. Thus, each pair p_{ref}, p_i of pixel correspondences gives us an equation of the form of Equation (5.14), and we can find $H_i, \gamma_i, \gamma_{ref}$ by solving a system of equations. We only have 9 unknowns for H_i and two for γ_i, γ_{ref}, but we have many pixel correspondences: most of the image if there is motion, or the whole image if the scene is static. Accordingly the system is overdetermined, and the solution can be found by an optimisation procedure. This is in contrast with the approach taken both in [Vazquez-Corral and Bertalmío, 2014a] and in the preliminary version of our method [Gil Rodríguez et al., 2015], where the optimisation is done in two stages, first finding estimates for γ_i, γ_{ref}, and then for H_i. Our new single shot optimisation allows for a significant improvement in accuracy, as will be shown in Section 5.5.3.

Once H_i, γ_i have been found we can produce I'_i, the linearised and colour-corrected (w.r.t. the reference image) version of I_i:

$$I'_i = H_i I_i^{1/\gamma_i}. \tag{5.15}$$

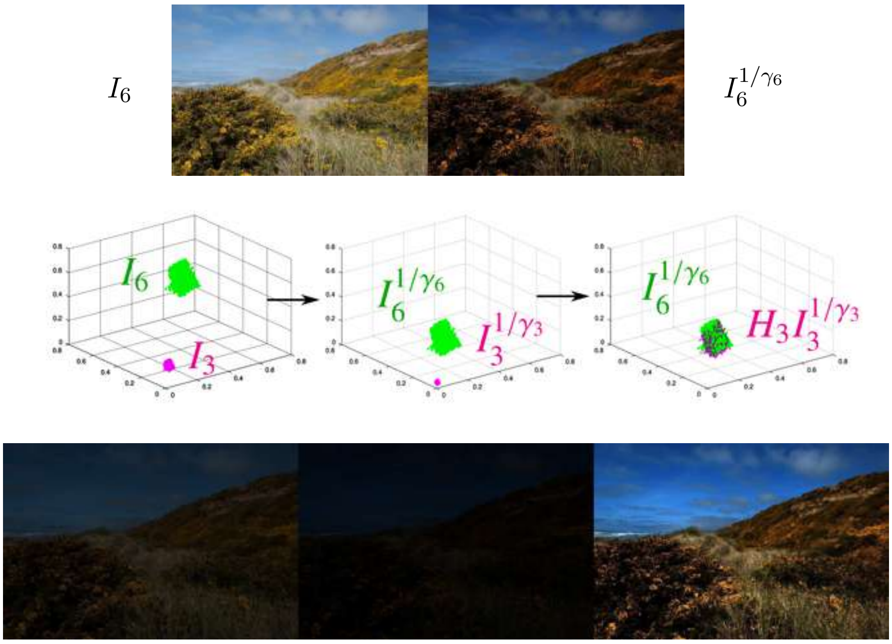
In Figure 5.7 we show an example of this procedure.

Step 2

After we have linearised and colour-corrected all images in the sequence, obtaining $I'_i, i = 1, ..., N$, we produce an intermediate HDR result HDR_{ref} by performing a weighted average with a trapezoidal weighting function ω_T, in the range $[0, 1]$, that discards extreme pixel values:

$$HDR_{ref} = \frac{\sum_{i=1}^{N} \omega_T^i I'_i}{\sum_{i=1}^{N} w_T^i}. \tag{5.16}$$

The intermediate HDR results present differences, as Figure 5.8 shows. In the top row, the leftmost image correctly captures the bright colour checker while missing out details on the dark colourchecker, and the reverse situation occurs with the middle and rightmost intermediate HDR results. Thus, we propose to combine the different $\{HDR_{ref_i}\}_{i=1,...,P}$ images to produce the final HDR output. We scale each of them so that they are all in the same range, since they have been computed from different reference images captured with different exposures. In order to do this, for each HDR_{ref_i} we compute the trimean, defined



(a) I_3 (b) I_3^{1/γ_3} (c) $H_3 I_3^{1/\gamma_3}$

Figure 5.7: Top row: the reference image I_6 on the left, and its linearised version I_6^{1/γ_6} on the right. Middle row: 3D point clouds representing the (R,G,B) values for I_6 in green and for I_3 in magenta. Bottom row: the corresponding images of I_3 for each plot in the middle row.

as $\frac{1}{4}(Q_1 + 2Q_2 + Q_3)$ where $Q_1,\ Q_2,\ Q_3$ are the quartiles; we choose the trimean in order to avoid outliers and take into account the distribution of the image data. Once all trimeans are obtained, we scale the values of each HDR_{ref_i} so that the resulting image has the same trimean as the one of a selected $HDR_{ref_{sel}}$, which in our case has been the middle-exposure reference (image 5 in a 9-image sequence). Finally, we average the scaled set $\{HDR_{ref_i}\}_{i=1,\ldots,P}$ to obtain the final HDR image, as shown in Figure 5.6. By fusing them, we achieve a final HDR image with more details in both bright and dark areas. This second step is not present in the preliminary version of our method [Gil Rodríguez et al., 2015].

Figure 5.8: Top row: intermediate HDR images HRD_{ref_6}, HRD_{ref_9}, and $HDR_{ref_{12}}$ computed taking as reference exposures $6, 9$ and 12, respectively. Bottom row: final HDR output of proposed method, after linear combination of intermediate results. All images have been tone-mapped using [Mantiuk et al., 2008].

Results and comparisons

Database

We performed our experiments using the HDR Survey dataset by Mark Fairchild [Fairchild, 2007]. The online public-domain database contains 105 different scenes acquired using a *Nikon D2x* DSLR camera. The data consists of corresponding JPEG and RAW images for different exposures. In each scene, images in the sequence are numbered going from shortest to longest exposure time. All the scenes except two are composed of 9 images; the other two have respectively 8 and 18 images. For the experiments and evaluation we reduce all images by a factor of $1/4$, so the image size equals 1072×712.

Ground-truth generation

Let us consider N RAW images acquired with different exposure times Δt_i. From the header of the RAW file we read the following parameters: 1) dark and saturation values, which are the minimum and maximum values that the camera produces, 2) a 3×1 array containing the white balance values for each channel and 3) the CFA Bayer pattern, e.g. *‘rggb’*. The ground-truth (GT) construction is defined in two stages. The first one is the merging step, where the N RAW images are combined to obtain a RAW HDR image HDR_L,

$$HDR_L = \frac{\sum_{i=1}^{P} \omega_i (RAW_i/\Delta t_i)}{\sum_{i=1}^{P} \omega_i}, \tag{5.17}$$

where ω_i is a weighting function. We have chosen to use as weighting function t^2 , where t is the exposure time, for its simplicity and good performance, as shown in [Robertson et al., 2003].

The second stage of GT creation converts the obtained HDR_L into a colour image by applying a number of steps: first HDR_L is linearly scaled to range $[0, 1]$, next white balance is applied, then demosaicing using the method proposed by Zhang and Xiaolin [Zhang and Wu, 2005], and finally the colour transformation in which each pixel triplet $[R, G, B]^t$ is multiplied by the 3×3 matrix $M_{colour} = E \cdot C$, where E is the matrix converting XYZ values into sRGB values, and C is the sensor characterisation matrix that transforms RGB sensor values into standard XYZ values and is the one described in [Fairchild, 2007] for the camera used to acquire the data.

Evaluation

We compare our approach against seven multiple exposure HDR methods. Four of them operate only on static scenes: the classical methods of Debevec and Malik (DM) [Debevec and Malik, 1997] and Mitsunaga and Nayar (MN) [Mitsunaga and Nayar, 1999], and the recent methods by Lee *et al.* (Lee13) [Lee et al., 2013] and by Gil *et al.* [Gil Rodríguez et al., 2015]. The other three work also on dynamic scenes: Lee *et al.*(Lee14) [Lee et al., 2014], Sen *et al.*(Sen) [Sen et al., 2012], and Hu *et al.*(Hu) [Hu et al., 2013] (linearisation performed with Debevec and Malik [Debevec and Malik, 1997]). We use public available codes either from the authors (Sen *et al.* [Sen et al., 2012], Lee *et al.* [Lee et al., 2014], Lee *et al.* [Lee et al., 2013], Hu *et al.* [Hu et al., 2013]), or by the *HDR Toolbox* from Banterle *et al.* [Banterle et al., 2011] for Debevec and Malik [Debevec and Malik, 1997], and Mitsunaga and Nayar [Mitsunaga and Nayar, 1999]. For our approach, we select for each sequence the central three exposures as references: images 4, 5 and 6.

Starting with quantitative, objective evaluation we compare the HDR outputs of each algorithm (with JPEG sequences as input) versus the computed GT using six standard metrics suggested in Hanhart *et al.* [Hanhart et al., 2015a] for this purpose: for luminance, peak signal-to-noise ratio (PSNR), structural similarity metric (SSIM) [Wang et al., 2004], and HDR quality assessment HDR-VDP-2 [Mantiuk et al., 2011]; for colour, the colour version of PSNR (CPSNR), the colour extension of SSIM, called CID [Lissner et al., 2013], and the colour difference measure $CIEDE2000$ (ΔE^*_{00}) [Sharma et al., 2005]. Finally, we also compute the l^2-norm on RGB space of the difference between a given method and the GT. The results, averaged over the dataset[4] are presented in Table 5.1. We can see that our method outperforms the others according to all metrics, except for HDR-VDP-2, where DM [Debevec and Malik, 1997] performs better and our algorithm comes second. Let us note that for HDR-VDP-2 the average is done over the 42 images for which photometer readings exist for the minimum and maximum absolute luminances of the scene, as these values are required by the metric; nonetheless, as we pointed out in Section 5.2, in general the minimum luminance value of the scene does not accurately correspond to the minimum luminance value of the HDR picture due to light scatter in the optics.

To show that the errors due to fluctuations in camera parameters can result in very visible artefacts, in Figure 5.9, top to bottom, we compare the outputs of DM [Debevec and Malik, 1997], Lee13 [Lee et al., 2013], Sen [Sen et al., 2012], and our approach, for the scenes *RITTiger*, *HancockKitchenInside*, *TheNarrows2* and *MasonLake1*, from left to right. All results have been tone-mapped with the method in [Mantiuk et al., 2008]. For the scene *RITTiger* we can see that DM [Debevec and Malik, 1997] (first row) presents a red cast in the image, therefore Sen [Sen et al., 2012] (third row) has the same colour cast since it uses DM as input. In *TheNarrows2*, Lee13 [Lee et al., 2013] (second row) shows very noticeable colour issues. Finally, for the *MasonLake1* scene, the method of Lee13 *et al.* presents a blue cast, while DM presents a reddish cast, and the method of Sen shows a banding artefact effect on the sky region.

To highlight that the visual problems described before are not due to a particular choice of tone-mapping algorithm, Figure 5.10 shows the same HDR

---

[4] For all metrics apart from HDR-VDP-2 we use all the images except *Zentrum*, since MN [Mitsunaga and Nayar, 1999] is not able to produce a reliable result for this image. Also, the algorithm of Hu *et al.* [Hu et al., 2013] only produces an output for 64 out of the total 105 scenes, so we have chosen to present just a qualitative comparison, in Figure 5.10.

Table 5.1: Mean ($\bar{x}$), and median ($\breve{x}$) results from all the presented methods versus GT among the 104 scenes. The last metric HDR-VDP-2 is computed only among the 42 images containing the values for the minimum and maximum absolute luminances.

		PSNR L	**CPSNR**	**SSIM L**	**CID**	**L2**	**ΔE^*_{00}**	**HDR-VDP-2**
DM	$\bar{x}$	29.48	28.77	0.890	0.179	0.052	4.41	**60.37**
	$\breve{x}$	27.44	27.29	0.912	0.150	0.032	3.87	**61.32**
MN	$\bar{x}$	26.50	25.91	0.863	0.194	0.071	5.32	57.75
	$\breve{x}$	25.01	24.79	0.872	0.175	0.071	5.02	57.79
Lee13	$\bar{x}$	30.91	30.24	0.905	0.179	0.043	4.56	59.48
	$\breve{x}$	29.29	28.94	0.929	0.125	0.030	3.37	59.88
Sen	$\bar{x}$	28.59	27.90	0.860	0.208	0.058	4.89	56.96
	$\breve{x}$	27.11	26.79	0.899	0.189	0.034	4.25	57.17
Lee14	$\bar{x}$	28.66	28.01	0.868	0.169	0.048	4.45	59.00
	$\breve{x}$	27.89	27.22	0.927	0.123	0.037	3.72	58.23
Gil	$\bar{x}$	30.37	29.43	0.899	0.164	0.044	4.74	58.44
	$\breve{x}$	29.97	29.21	0.924	0.141	0.032	4.19	60.16
Proposed	$\bar{x}$	**33.45**	**32.68**	**0.930**	**0.110**	**0.033**	**3.39**	60.00
	$\breve{x}$	**34.02**	**32.28**	**0.942**	**0.098**	**0.025**	**3.36**	60.98

results but tone-mapped with two different methods, [Mantiuk et al., 2008] for the first two columns and [Drago et al., 2003] for the last ones. The scenes are *AirBellowsGap* (columns 1 and 3) and *LabWindow* (columns 2 and 4), while the multiple exposure HDR methods to compare are, from top to bottom: MN [Mitsunaga and Nayar, 1999], Lee14 [Lee et al., 2014], Hu [Hu et al., 2013], and our approach. We can see how the previously existing methods produce colour artefacts in the sky and sun of the *AirBellowsGap* scene, and in the curtains, sky and background of the *LabWindow* scene, which are apparent for both of the tone-mapping methods used.

Dynamic scenes

It is worth emphasising that the proposed algorithm does not require image registration, only a set of pixel correspondences. Therefore, it can be used on dynamic scenes as well: in particular, Step 1 of our method can be employed as a pre-processing step to colour-stabilise the inputs of HDR methods operating on dynamic scenes, enhancing their performance. To illustrate this, we consider the algorithm of Sen *et al.* [Sen et al., 2012], which receives linearised images

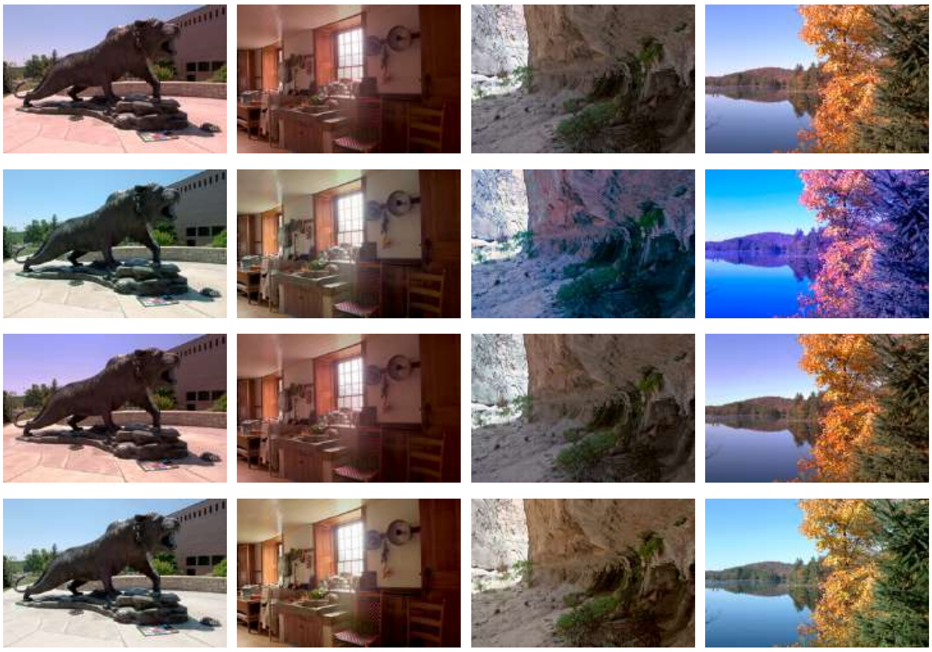


Figure 5.9: From left to right: scenes *RITTiger*, *HancockKitchenInside*, *TheNarrows2* and *MasonLake1* from the HDR Survey. From top to bottom: results from DM [Debevec and Malik, 1997], Lee13 [Lee et al., 2013], Sen [Sen et al., 2012] and finally our approach. All images tone-mapped with [Mantiuk et al., 2008].

as input. Consequently, given a stack of non-linear images, we compare three linearising approaches: 1) CRF computed by Debevec and Malik [Debevec and Malik, 1997], 2) radiometric calibration by Lee *et al.* [Lee et al., 2013], and 3) applying Step 1 of our method, using as a reference the image in the mid-point of the sequence and finding pixel correspondences with SIFT [Lowe, 1999]. We conducted this experiment on a stack of five images from the dataset presented in [Sen et al., 2012], the *Skater* sequence (that comes in JPEG format). In Figure 5.11 we present the HDR outputs obtained using the three different linearisation approaches. The zoomed-in details allow us to see how linearisation by [Debevec and Malik, 1997] (left) produces artefacts in overexposed areas, whereas linearisation with [Lee et al., 2013] produces results that, although free from artefacts, have lower contrast and less saturated colours than what can be obtained with our method.

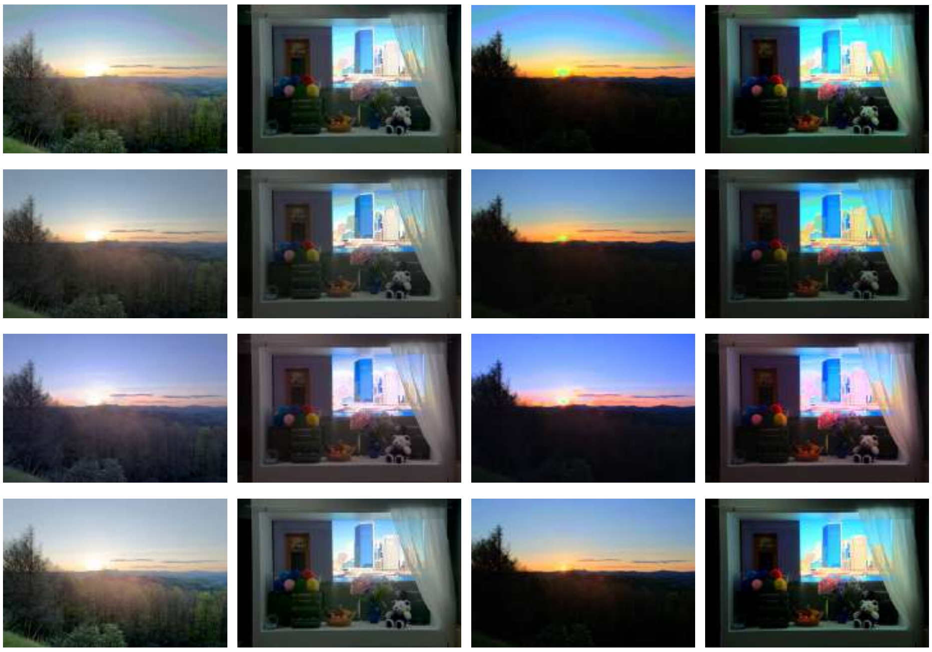


Figure 5.10: First and third column: *AirBellowsGap* scene. Second and fourth column: *LabWindow* scene. Columns 1 and 2: results tone-mapped with [Mantiuk et al., 2008]. Columns 3 and 4: results tone-mapped with [Drago et al., 2003]. From top to bottom: results from MN [Mitsunaga and Nayar, 1999], Lee14 [Lee et al., 2014], Hu [Hu et al., 2013] and finally our approach.

Conclusion

Our experiments show that the camera response function changes with the exposure and depends on the three colour channels simultaneously. For this reason, multi-exposure HDR approaches based on estimating and inverting a CRF that is supposed to be constant may have substantially more error than if computed directly from the linear data, and when tone-mapped they commonly show hue shifts, colour artefacts or contrast problems. In this chapter we have introduced a very simple model to estimate the effective dynamic range than can be captured with multi-exposure techniques, and proposed a method for removing the fluctuations in the internal settings that the camera has automatically modified, so that our approach effectively makes the CRF constant for the whole sequence. It can

Figure 5.11: HDR results on a dynamic scene applying the HDR creation method of Sen *et al.* [Sen et al., 2012] with three different linearisation techniques: Debevec and Malik [Debevec and Malik, 1997] (left), Lee *et al.* [Lee et al., 2013] (middle), and Step 1 of our proposed method (right), taking as reference the image in the mid-point of the sequence. HDR results tone-mapped with Mantiuk *et al.* [Mantiuk et al., 2008].

be applied both to static and dynamic scenes. Our results are more accurate than those obtained with state-of-the-art methods and show no visual problems.

CHAPTER 6

HDR video from dual-ISO sensor

Our main contribution in this section is to propose a simple and effective method to shoot high quality video in HDR scenarios using a single camera capable of recording interlaced dual-ISO footage. The interlaced input guarantees that the result will be free of ghosting artefacts, because the low and the high ISO lines are recorded simultaneously. Our implementation pipeline incorporates a set of stages: calculating the full-frame single-ISO images applying a deinterlacing algorithm, combining these full-frame images into a single HDR picture, and finally applying a tone-mapping operator to produce a LDR output. For these stages we adapt to our setting state-of-the-art algorithms which produce high-quality results. Tests and comparisons show that our method outperforms other approaches both quantitatively (in terms of PSNR) and qualitatively (no spurious colours, better edge preservation, less noise).

This chapter is based on our work ‘High quality video in high dynamic range scenes from interlaced dual-ISO footage’ published in *IS&T International Symposium on Electronic Imaging* [Gil Rodríguez and Bertalmío, 2016].

High quality video in high dynamic range scenes from interlaced dual-ISO footage

The input to our algorithm will be a video sequence in RAW format, where each frame alternates row pairs with different ISO values, and the output will be a

LDR video sequence with simultaneous detail visibility in the dark and bright zones of the picture, see scheme in Figure 6.1. All the stages of our method are applied on a frame-by-frame basis except for the final tone-mapping, which imposes temporal consistency on the output by considering several input frames at the same time. Let us describe our proposed method in detail.

Generation of single-ISO full-frame images

The first stage consists in the computation of two single-ISO full-frame images from an input dual-ISO frame, using an inpainting-based deinterlacing method. In order to obtain both these full-frame pictures I_l (for the low ISO value) and I_h (for the high ISO value) we proceed as follows:

- Split the dual-ISO input into two half-size images $I_{l/2}$ and $I_{h/2}$, each one with the rows corresponding to a single ISO value, see Figure 6.2.

- Using an adapted version of the inpainting-based deinterlacing method [Ballester et al., 2007], generate full-frame images I_l and I_h from $I_{l/2}$ and $I_{h/2}$ respectively.

- Perform demosaicing using [Zhang and Wu, 2005], and apply a refinement step to improve the interpolated results, see Figure 6.3.

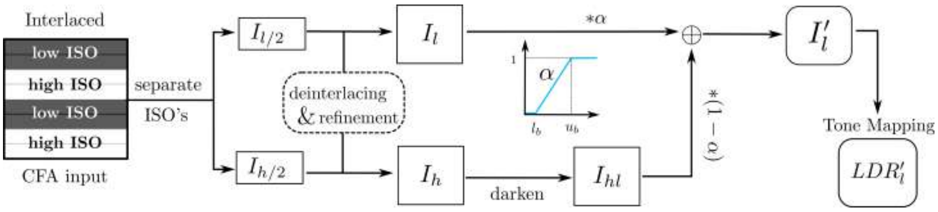


Figure 6.1: Schematic of the proposed method. First the dual-ISO input is split into two half-size images $I_{l/2}$ and $I_{h/2}$, each one with the rows corresponding to a single ISO value. Next a deinterlacing method is used to generate full-frame images I_l and I_h from $I_{l/2}$ and $I_{h/2}$ respectively. These full-frame images are linearly combined and tone-mapped to produce the final output.

Figure 6.2: The dual-ISO input frame (left), which is split into two half-size images $I_{l/2}$ and $I_{h/2}$ (right), each one with the rows corresponding to a single ISO value.

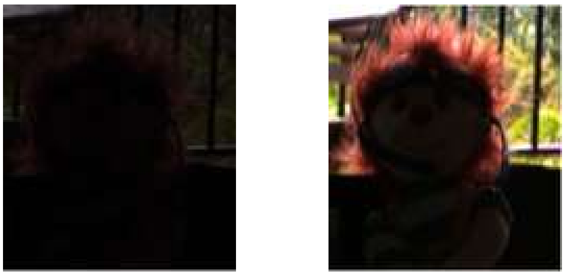


Figure 6.3: Generated full-frame images I_l (left) and I_h (right).

Row interpolation by deinterlacing

In order to interpolate the missing rows and generate I_l from $I_{l/2}$ and I_h from $I_{h/2}$, we adapt the deinterlacing method of Ballester *et al.* [Ballester et al., 2007]. This is a state-of-the-art technique that follows the dense stereo matching approach of Cox *et al.* [Cox et al., 1996] and fills-in a missing line $\mathcal{L}_0$ between two given lines $\mathcal{L}_-$ and $\mathcal{L}_+$ by (see Fig. 6.4):

- First, performing a global matching between lines $\mathcal{L}_-$ and $\mathcal{L}_+$ by computing the correlation matrix between their image values and finding the matches (optimal path) through dynamic programming. In practice only a band around the diagonal is considered for the search for the optimal path, and the matches are estimated assuming a certain noise variance for the image values.

- Second, each matching pair of pixels determines a segment that crosses the missing line $\mathcal{L}_0$ at a pixel location that is filled-in with the average of the matching pair. For those points that were not matched, their values are computed by bilinear interpolation from the neighbouring correspondences.

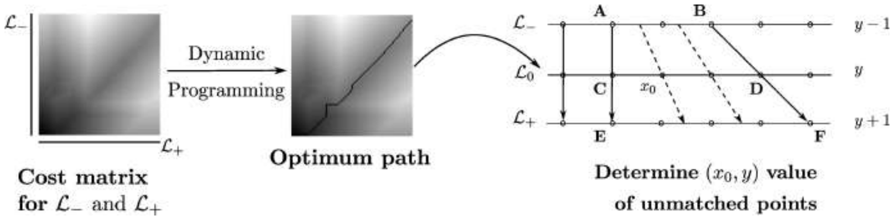


Figure 6.4: Inpainting-based deinterlacing algorithm. Left: global matching between two given lines $\mathcal{L}_-$ and $\mathcal{L}_+$. Right: each matching pair of pixels determines a segment that crosses the missing line $\mathcal{L}_0$ at a pixel location that is filled-in with the average of the matching pair. Figure adapted from [Ballester et al., 2007].

The images $I_{l/2}$ and I_h are colour filter array (CFA) RAW pictures with the 2×2 Bayer pattern ‘RGGB’; with the R and B values we create half-size channels that can be deinterlaced directly with [Ballester et al., 2007], while for the G channel we first demosaic the values using [Zhang and Wu, 2005] and then apply the deinterlacing method [Ballester et al., 2007] but with the modification of filling-in two consecutive lines simultaneously given the upper and lower neighbouring lines of the pair; see Figure 6.5.

Refinement step

In addition, a refinement step is required to improve the interpolated values in both the I_l and I_h full-frame images, see Figure 6.6. We describe the process for I_l, since it will be analogous for I_h. First, we demosaic the frame applying [Zhang and Wu, 2005]. Second, we perform deinterlacing once again using the same method [Ballester et al., 2007] as before. For simplification we focus on a small row neighbourhood defined by four rows r_{s_1}, r_{d_1}, r_{d_2} and r_{s_2}. The subscript s denotes a row with values given by the sensor, and d denotes a row calculated from deinterlacing in the previous step:

1. Consider $r_{s_1}, r_{d_1}, r_{d_2}$ and apply deinterlacing assuming that r_{d_1} is unknown. Store the new computed row in r_{d_1}.

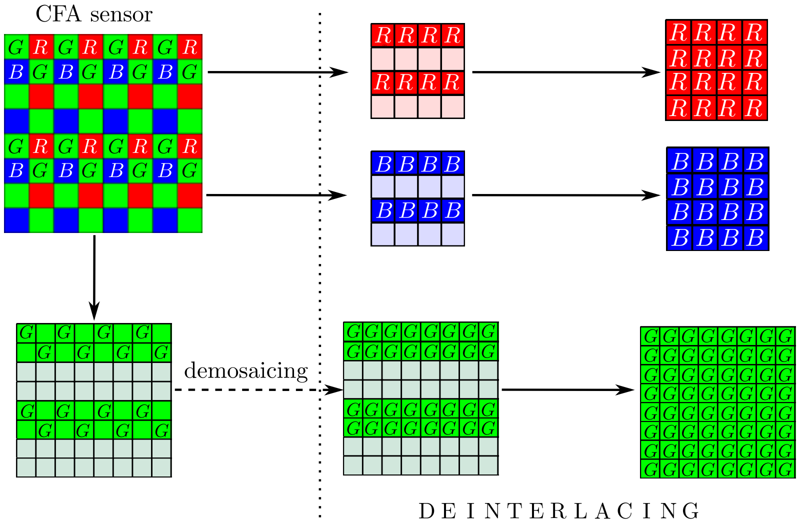


Figure 6.5: Extended deinterlacing algorithm. It is shown how RGB values are read from the CFA image before applying deinterlacing, and the final interpolated RGB planes.

2. Consider r_{d_1}, r_{d_2}, r_{s_2} and apply deinterlacing assuming that r_{d_2} is unknown. Store the new computed row in r_{d_2}.

As a result, we obtain the final full-frame images I_l and I_h. This refinement step produces more accurate and smoother edges, as Figure 6.6 shows.

HDR creation

In order for the final result to have simultaneous detail visibility in the dark and bright zones of the picture, we need to combine the information from I_l and I_h. We do this by first bringing the high-ISO image I_h to the range of I_l (simply dividing the values of I_h by the ratio between ISO values), obtaining an image we call I_{hl}, and then generating an HDR low-ISO image $I_{l'}$ by a linear combination of I_l and I_{hl}:

$$I_{l'}(p) = (1-\alpha)I_{hl}(p) + \alpha I_l(p), \tag{6.1}$$

where p denotes pixel location and the weighting parameter α depends on the value of $I_{hl}(p)$. Let us note that, since the reconstructed HDR I'_l is a CFA RAW image, we convert it into an RGB image following the main stages in

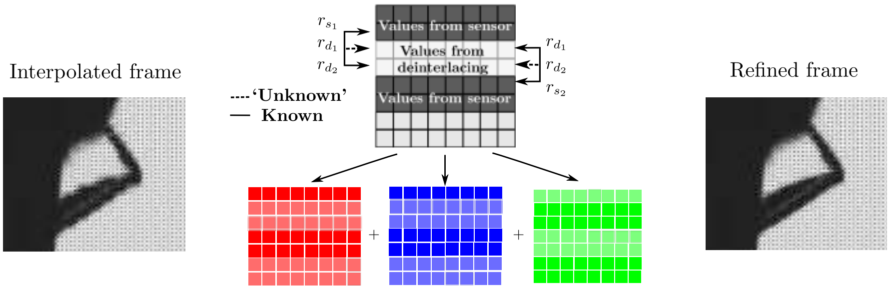


Figure 6.6: The diagram shows the rows considered for the refinement step: r_{s_1} and r_{s_2} from sensor readout, r_{d_1} and r_{d_2} from the previous deinterlacing step. The images show a crop of an I_l frame, before and after applying the refinement step (in each channel separately). Notice the smoothness of the resulting contours.

the colour processing pipeline described in [Bianco et al., 2012], without applying any gamma correction nor non-linearity: we perform denoising on CFA images [Akiyama et al., 2015], then apply demosaicing [Zhang and Wu, 2005], then multiply by the corresponding colour encoding and colour characterisation matrices.

Final LDR video

The final LDR image LDR'_l is obtained using the state-of-the-art tone mapping operator (TMO) [Cyriac et al., 2015b], which can be applied to single pictures, as well as video, see Figure 6.7.

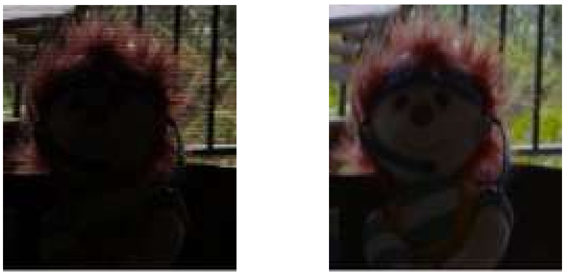


Figure 6.7: Left: dual-ISO input. Right: tone-mapped output from our method.

Experiments and Evaluation

The only parameters of our method are the noise standard deviation σ and the band size B (both parameters required by the deinterlacing step), plus the shape of the weighting function α. For all our tests, carried out on 12-bit input images, we use as default parameter values $\sigma = 100$, $B = 50$, and for the function α:

$$\begin{cases} \alpha(x) = 0 & x < l_b \\ \alpha(x) = \frac{x - l_b}{u_b - l_b} & l_b \leq x \leq u_b \\ \alpha(x) = 1 & x > u_b \end{cases}$$

where $l_b = 2048$ and $u_b = 0.8 * max(I_{hl})$ (see plot of α in Figure 6.1).

We start by showing the effectiveness of our proposed approach. We take image data from the HDR survey of M. Fairchild [Fairchild, 2007] and generate our dual-ISO input pictures by interlacing pairs of lines coming from images with a 4-stop difference in exposure (the scenes are static, therefore different exposure times are equivalent to different ISO values). Applying our method to these pictures produces the results that appear in Figure 6.8, with simultaneous visibility of details in both dark and bright zones, and absence of noise or visual artifacts like spurious colours or halos.

Quantitative evaluation

For a quantitative evaluation we take our results and those obtained with the camera software Magic Lantern (ML) [Lantern, 2013] and compare against the ground truth which can be obtained from the Fairchild database. We choose ML because its code is openly available, it seems to be the most widely used method for this application, and it provides intermediate outputs equivalent to our images I_l and I'_l.

For the ground truth we select a sample of 20 scenes from Fairchild’s HDR survey so that it includes night and daytime images, outdoor and indoor environments, varying from very high to low dynamic ranges: *507*, *AmikeusBeaverDamPM1*, *BenAndJerrys*, *CanadianFalls*, *DevilsBathtub*, *Flamingo*, *GoldenGate1*, *HancockKitchenOutside*, *JesseBrownsCabin*, *LabBooth*, *LetchworthTeaTable1*, *MammothHotSprings*, *MirrorLake*, *OtterPoint*, *SmokyTunnel*, *SunsetPoint1*, *TheNarrows2*,

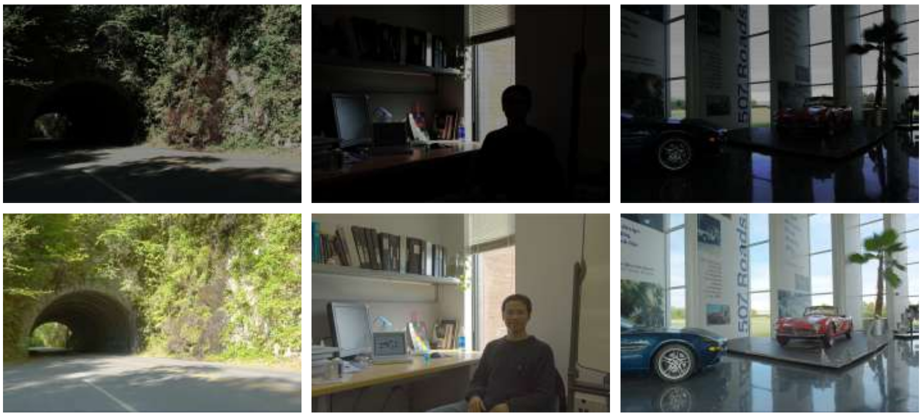


Figure 6.8: Three scenes from Fairchild dataset are shown: *SmokyTunnel*, and *WillyDesk*. The first row represents the simulated dual-ISO input, and the second row shows the final tone-mapped output of our method.

WaffleHouse, *WillyDesk*, and *Zentrum*. All scenes consist of 9 RAW pictures taken with different exposure times. For each scene we generate our dual-ISO input by interlacing pairs of lines coming from pictures RAW_1 and RAW_2 that have a 4-stop difference in exposure. We apply ML and our method to these interlaced inputs, each algorithm providing intermediate outputs I_l (interpolated full-frame low-ISO image) and I_l' (that combines I_l and I_h). We have three ground truth images: the low-ISO (actually short-exposure) LDR image RAW_1, the HDR image GT_{HDR} (obtained combining the 9 differently-exposed pictures with the method [Gil Rodríguez et al., 2015]), and the HDR image Lum which is the luminance of GT_{HDR}. The quantitative comparisons are the following: PSNR between I_l and RAW_1, PSNR between I_l' and GT_{HDR}, and the HDR visibility difference metric HDR-VDP-2 [Mantiuk et al., 2011] between I_l' and Lum. See Figure 6.9. Table 6.1 shows the results, averaged over the 20 images of our sample of Fairchild’s database; we can see that our approach outperforms ML for all three measures considered.

Qualitative evaluation

Figure 6.10 compares our (tone-mapped) result to those of Hajisharif *et al.* [Hajisharif et al., 2014] and ML [Lantern, 2013] on an actual dual-ISO picture (ISO

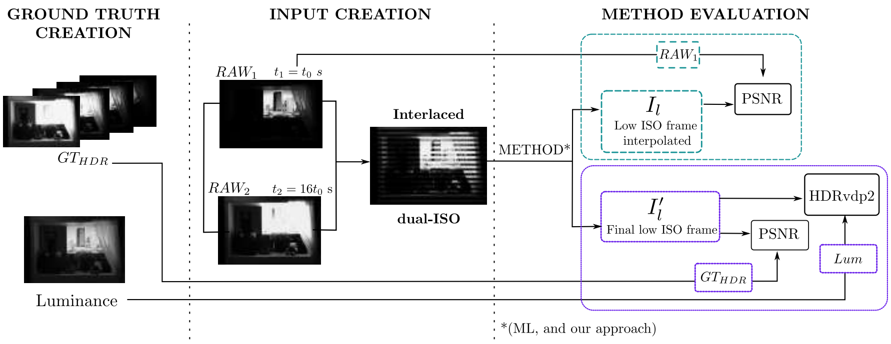


Figure 6.9: Evaluation diagram. For each scene we select two images RAW_1 and RAW_2 of 4-stops difference in exposure time, and interlace them to create the dual-ISO input. We compute as well the ground truth HDR image GT_{HDR} from the set of all the differently-exposed pictures that comprise the scene; Lum is the luminance of GT_{HDR}. We compute: PSNR between I_l and RAW_1, PSNR between I_l' and GT_{HDR}, HDR-VDP-2 between I_l' and Lum.

Table 6.1: The averaged results for the proposed method and for ML [Lantern, 2013].

	PSNR		**HDR-VDP-2**
	$\mathbf{I_l}$ vs $\mathbf{RAW_1}$	$\mathbf{I_l'}$ vs $\mathbf{GT_{HDR}}$	$\mathbf{I_l'}$ vs **Lum**
ML	47.71	29.89	54.17
Ours	**48.79**	**30.58**	**57.43**

values 100 and 1600) taken with a *Canon EOS 7D Mark III* camera (access to the camera was required for the application of the method of Hajisharif *et al.* [Hajisharif et al., 2014], so those authors kindly provided the original image and their result for comparison). The first row uses as a tone-mapping operator our default method [Cyriac et al., 2015b], while for the second row we have used another state-of-the-art TMO, that of Mantiuk *et al.* [Mantiuk et al., 2008]. We can see that ML produces a result with severe colour problems, due to the fact that the output of ML is not really an HDR image so the application of a TMO increases the overall brightness at the prize of creating unrealistic colours. On the other hand, the result of Hajisharif *et al.* [Hajisharif et al., 2014] is of very good quality, but zooming-in on a textured area we can notice some colour artifacts.

Figure 6.11 compares ML and our method for the Fairchild database scenes *HancockKitchenOutside* and *JesseBrownsCabin*. In this case the tone-mapping is not problematic for ML, but we can appreciate in the zoomed-in details (bottom

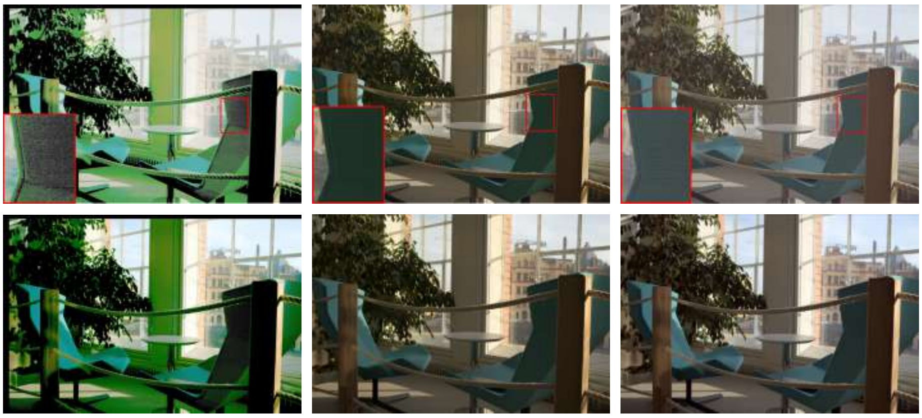


Figure 6.10: From left to right: the tone-mapped results of ML [Lantern, 2013], Hajisharif *et al.* [Hajisharif et al., 2014] and ours. Top row: results tone-mapped with [Cyriac et al., 2015b] (the default in our method). Bottom row: results tone-mapped with [Mantiuk et al., 2008].

Figure 6.11: Final tone-mapped images of *HancockKitchenOutside* and *JesseBrownsCabin* scenes from Fairchild data. For each pair of images we show the output of ML (left), and of our approach (right). Notice how the results of ML are quite noisy.

row) how the results of ML are quite noisy.

Figure 6.12 presents the result of applying our method to a dual-ISO video sequence taken with a *Canon EOS 5D Mark II* camera using ISO values 100 and 1600. The top row shows some frames from the interlaced dual-ISO input and the bottom row the corresponding LDR images obtained with our approach. This video has several moving objects, but none of the resulting frames shows noticeable visual artifacts of any kind, and the output LDR video doesn’t suffer

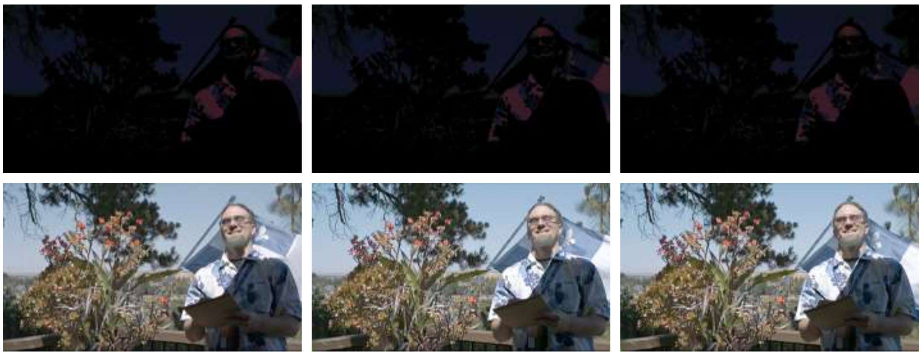


Figure 6.12: Top row: frames from a dual-ISO video. Bottom row: results obtained with proposed method. The video does not show temporal artifacts.

Figure 6.13: Tone-mapped results from two video sequences. First column: results of ML. Second column: results of our approach.

from temporal artifacts either.

Figure 6.13 compares ML and our method for two video sequences captured with a *Canon EOS 5D Mark II* camera using dual-ISO values 100 and 1600. Again, as it was the case in the static scenes of Figure 6.11, we can see that ML produces results that are rather noisy, as it can be appreciated on the shadowed parts of the images.

Finally, the top row of Figure 6.14 shows that while the default choice of parameter values produces good overall results, in some cases there can be some mismatches on the interpolated missing lines. The bottom row shows that these problems can be significantly reduced with a better, image-dependent choice of parameters (in this case: $\sigma = 20$, $B = 100$).

Analysis and optimisation of the proposed method

The Bachelor's Degree Final Project presented by [Sintes, 2017] extended the proposed method for HDR video reconstruction. The main contributions of the project are i) to analyse the influence of the parameters involved in the method, regarding execution time and image quality, and ii) to improve the execution time by optimising the code and implementing parallel programming when possible.

In the analysis of the method, three parameters are considered. These parameters have an impact in the final HDR reconstruction. Let us detail the selected parameters and their influence.

- Band size (B) determines the size of the neighbour pixels considered when computing the correlation matrix. Usually the value of this parameter will not be too large, since we assume that corresponding pixels in two consecutive rows are not far from each other. This parameter has a deep impact on the execution time. It also influences image quality at the edges of the objects. Notice that for horizontal edges, we need to set B with a large value, whereas for vertical and obliques edges the value can be reduced.

- Patch size (φ) defines the neighbourhood considered for each pixel to compute the correlation coefficient between the possible matchings in the next row. This parameter also influences the execution time. It is important to note that images with a lot of texture would require small values of φ, in contrast with homogeneous regions, where we may need to select larger values.

- Noise standard deviation (σ) is defined in order to consider noise at sensor level, and account for it. As mentioned in previous chapters, high ISO values increase the noise in the acquired image. This parameter works as

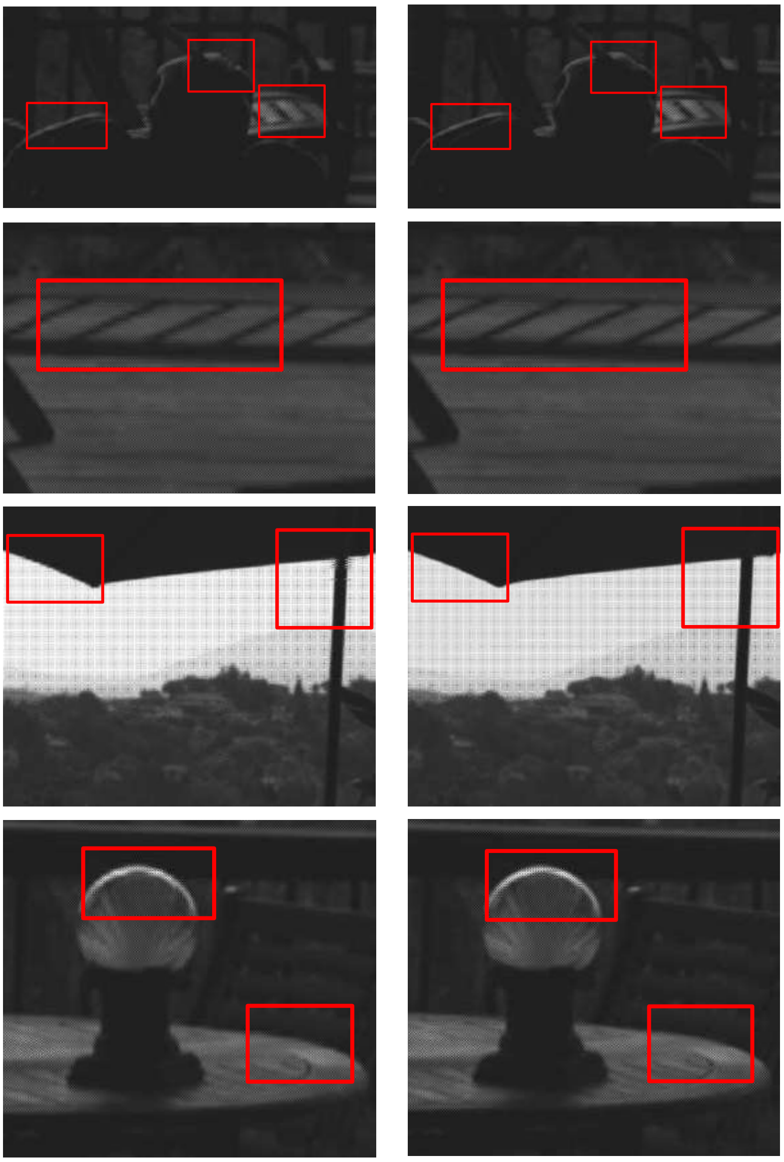


Figure 6.14: Top row: results of our method using default choice of parameter values, notice some artifacts. Bottom row: results with better, image-dependent parameter values.

Table 6.2: Comparison of the total execution time of the original version, and the two versions proposed in [Sintes, 2017].

	Original version	Dual-ISO RGB v.1	Dual-ISO RGB v.2
Execution time	1979 s	53 s	62 s

a threshold, and small values mean a restriction to the presence of noise. This parameter does not influence the execution time.

The code was initially developed in C/C++ using ImageMagick [Cristy and ImageMagick Studio LLC, 1990] and OpenCV [Bradski, 2000] libraries. Therefore, the first decision was to select only one library, in this case OpenCV, for the whole code. Different optimisation techniques were used in order to avoid memory leakage. Finally, the OpenMP [Dagum and Menon, 1998] library was used for parallel programming. In Table 6.2, the execution times of the original method, and the two proposed new implementations Dual-ISO RGB v.1 and v.2 are shown. On both implementations, the refinement step is removed, which results in a compromise between image quality and execution time. For more details on those implementations, see [Sintes, 2017].

Conclusion

We have presented a practical approach for creating LDR video of HDR scenes from single-camera interlaced dual-ISO footage. Our method is affordable, since the input is obtained from a DSLR camera; it does not suffer from ghosting artefacts due to the dual-ISO information in each frame, instead of alternating exposures in consecutive frames; it minimises spatial artefacts, thanks to the application of a high quality deinterlacing method; and it provides results with visible details in dark and bright areas, consistent with human perception, thanks to the tone mapping method used.

It would be interesting to adapt the parameters of the algorithm depending on the input image. For example, in high ISO values we might be more flexible in terms of noise. As mentioned in earlier chapters, we know that large values of ISO increase the noise in the image. Although small values of band size (e.g. 10) might be enough, we could adapt the parameter based on the direction of the edges in the image. This means that for horizontal edges we would choose a bigger value, rather than for vertical and oblique edges small values would work as good. The pixel neighbourhood for patches comparison can vary depending on the contrast measured in each pixel. Thus, for high contrast we would consider a small neighbourhood, and for homogeneous regions with low contrast values, we would rather consider larger neighbourhoods.

CHAPTER 7

Colour matching: relevance and methods

Motivation

Colour matching techniques aim to map the colours of one image, defined as *source*, to those of a second image, defined as *reference*. These colour differences are due to the use of different camera models, or even when shooting with the same camera model under different camera settings: white balance, exposure time, aperture, etc. For example, Figure 7.1 shows two pictures, the reference and the source, taken from the same scene and acquired using the same camera. In this case, we only varied the white balance from the camera in manual mode. These two images show large colour differences. The image on the right side corresponds to the colour matched source, which shares the same colour palette as the reference one.


Reference
Source
Colour matched

Figure 7.1: Colour matching a pair of images taken from the same scene. Images were acquired using the same camera model and varying the white balance setting.

Colour matching is used in many professional situations. For example in TV: when broadcasting from different cameras, it is necessary to guarantee that there are no colour changes between cameras. In cinema: a stereo rig set-up for stereoscopic 3D requires colour correction of the two cameras, since the two images would be projected together at the same time. Amateur situations also demand colour matching techniques: in mosaicing methods for panoramic images, whose objective is to concatenate, in a single image, a set of pictures taken by just rotating the camera.

Current solution on cinema and TV

In cinema and TV it is quite usual the need for working with footage coming from several cameras, which show noticeable colour differences among them even if they are all the same model. In TV broadcasts, technicians work in camera control units (CCU) [MediaCollege, 2012] so as to ensure colour consistency when cutting from one camera to another. In this case, the technicians have control of the white balance and exposure time of the set of cameras. In cinema productions, several cameras record the same scene from different points of view. Recording with the same camera model and setting the same parameters (like white balance) does not assure colour stability, since different points of view might require the change of shooting conditions. This issue becomes more noticeable when using different cameras, due to the fact that setting same camera parameters in all of them is not enough for eliminating colour differences.

In cinema post-production, colourists need to manually colour-match images coming from different sources. Aiming to help to perform this task automatically, the Academy Color Encoding System (ACES) introduced a colour management framework [Postma and Chorley, 2015]. The idea is to allow colourists to work within the same colour space and be able to use different cameras and displays; however, the ACES pipeline requires to have the cameras characterised previously, and therefore does not allow to work ‘in the wild’, a situation which is very common.

Colour charts can be used at the shooting time. In this way, colourists might colour match the footage either manually or using automatic softwares like DaVinci Resolve [Design, 2018], or Imatest Multicharts [Imatest, 2018]. These softwares estimate look up tables (LUT) in order to find the colour correction matrix. The matrix is computed from the corresponding squares of the colour charts that appear in the source and reference image.

Figure 7.2: On the left, a camera control unit; on the middle, the DaVinci Resolve tool; and on the right, the Imatest software for colour correction based on colour charts. Images from [Design, 2018] and [Imatest, 2018].

Figure 7.2 presents from left to right: a camera control unit used in TV broadcast, the DaVinci Resolve tool for colourists, and the Imatest software used for colour correction by utilising colour charts.

Colour matching in academia

In image processing and computer vision research, it is a challenge to colour match a pair of pictures (reference and source). We differentiate between the case where image pairs do not necessarily share any content (colour transfer), and those taken from the same scene (colour stabilisation). The latter can be understood as a constrained colour transfer problem. Some applications, as image stitching, require colour uniformity across a set of images (more than two images) from the same scene.

Colour transfer

The aim of colour transfer methods is to obtain an image that has the structure of the source image and the colours of the reference image. In this case, images might belong to completely different scenarios, and thus no content is shared between images. In Figure 7.3 we present a colour transfer situation. The reference image has vivid orange and blue colours, and the source shows pale colours. The image on the right corresponds to the result of colour transferring the reference palette into the source image.

Figure 7.3: Colour transfer example: on the left the reference image, on the middle the source image, and on the right the source image with the same look/appearance as the reference. Image from [Reinhard et al., 2001].

A seminal work in colour transfer was proposed by Reinhard *et al.* [Reinhard et al., 2001], where the pair of RGB images are first converted to a decorrelated colour space, and then the mean and variance from the reference are transferred to the source. Xiao and Ma [Xiao and Ma, 2006] also worked with colour statistics, but unlike Reinhard method, the colour space treated is not relevant. Same authors extended their work in [Xiao and Ma, 2010], where they proposed a preserving-gradient colour transfer technique, and an evaluation metric for colour transfer methods. The method presented by Kotera [Kotera, 2005] proposed to compute the principal components of the colour clusters, in order to match the principal axes of the source to the reference image by a matrix multiplication (rotation and scaling). Pouli and Reinhard [Pouli and Reinhard, 2011] performed histogram matching along different scales given images of different dynamic ranges. The authors worked in CIELAB colour space.

The method of Pitié *et al.* [Pitié et al., 2005] defines the images as probability density functions, and then matches them through an iterative non linear process, involving only one dimensional marginals. In [Pitié et al., 2007], they extended their work by considering the gradient, in order to remove the artefacts that might appear in the colour matched image. Some methods use optimal transport, in order to minimise the cost of transferring probability density distributions of the source image, into the reference one, [Rabin et al., 2014] and [Ferradans et al., 2014]. On the latter, the authors proposed a new formulation for discrete optimal transport, and then they applied it to colour transfer problem.

Nguyen *et al.* [Nguyen et al., 2014a] presented a colour transfer method that first applies colour constancy to the input images, then it performs luminance matching, and finally the colour gamuts are aligned by a linear transformation. The algorithm of [Nguyen et al., 2014b] works for RAW input images in order to colour match the linear RGB colour spaces of different cameras (source and reference). In addition, it needs several images taken under different illuminant from both source and reference cameras. Gong *et al.* [Gong et al., 2016,Gong et al., 2017] proposed a colour transfer method based on a homography transformation and a mean intensity mapping.

All the above algorithms are global methods, although we find local approaches like [Tai et al., 2005, Xiang et al., 2009, Fecker et al., 2008] to name a few.

Related applications to colour transfer have appeared recently. The method of Lindner and Süsstrunk [Lindner and Süsstrunk, 2015] uses semantics to determine a specific colour style or look for a given image, what they called *semantic colour transfer*. In this case, the semantics are equivalent to the reference image, and the input image would be the source. The authors apply tone-mapping curves to each colour channel depending on the semantic concept applied. Also the method of [Vazquez-Corral and Bertalmío, 2018] shows how to modify a gamut mapping method so it can be used for this problem, but also for colour coherence, where the goal is to modify a source image to make its colours consistent with the colours of an unrelated reference one, but without changing the look of the source.

Colour stabilisation

Colour stabilisation tackles the situation where some regions or objects appear in both the reference and the source images. These approaches focus on those corresponding regions to estimate the colour matching transformation. Figure 7.4 presents an example of colour stabilisation. On the left, the reference image, and on the middle the source image, which was taken in the same scene using zoom out. Finally on the right, the colour matched image where we can appreciate the same colour appearance as shown in the frame of the reference.

Figure 7.4: The reference image on the left, the source image on the middle, and on the right the colour stabilised image on the right. Image from [Vazquez-Corral and Bertalmío, 2014a].

In their seminal work HaCohen *et al.* [HaCohen et al., 2011] presented a method to obtain dense correspondences between the images by computing corresponding patches [Barnes et al., 2010]. The estimation of the correspondences is combined with a global colour mapping model.

Vazquez-Corral and Bertalmío [Vazquez-Corral and Bertalmío, 2014a] proposed a colour stabilisation algorithm that consists of estimating a power law (γ value) for each of the images, and a single 3×3 matrix, to colour match the source image to the reference. It is built on the assumption that in digital cameras the colour encoding can be expressed as a matrix multiplication followed by a power law (gamma correction). In a similar way, the method of Frigo *et al.* [Frigo et al., 2016] presented colour stabilisation for video sequences. The colour correction model used is based on channel-based non-linearity (gamma correction) and channel-based scaling estimation (3×3 diagonal matrix).

The authors of Hwang *et al.* [Hwang et al., 2014] proposed to use moving least squares for colour stabilisation, by incorporating a probabilistic measure to ensure robustness against noise and outliers. The colour transformation is defined as affine, and they used extrapolation on those parts of the image where there are no common regions.

Colour consistency among multiple views

Colour consitency is an extension of colour stabilisation, and it is used when a set of images from the same scene need to be colour matched. For example for image mosaicing, 3D reconstruction, etc. Figure 7.5 presents a set of images taken from the same location, and focused on different objects of the scene. Some of the pictures share the same content, which will be used for estimating colour consistency in the whole image sequence.

Figure 7.5: Colour consistency example: a set of images from the same scene, with shared content among them. Image from [HaCohen et al., 2013].

HaCohen *et al.* [HaCohen et al., 2013] extended their previous approach in colour stabilisation problem [HaCohen et al., 2011], to the case of more than two images. They computed colour transformation between image pairs with shared content, and then those mappings can be composed in order to obtain stability in the whole image stack.

In a recent work, Park *et al.* [Park et al., 2016] proposed a method that uses rank minimisation in all the set of correspondences between the image stack. They use a model for colour correction based on a channel-based power law (gamma correction), and a scale factor (white balance). Therefore, in the optimisation they use the 2-rank constraint.

Xia *et al.* [Xia et al., 2017] presented a method to achieve colour consistency in image stitching. On the overlapping regions among the shots, it computes parametric curves for each channel under colour, gradient and contrast constraints.

In Tian and Cohen [Tian and Cohen, 2017] approach colour consistency is achieved by first categorising images in low and high quality. Then, the set of high quality images are colour matched, and used for colour transferring with the best match among the low quality set.

A review of the performance of colour consistency methods is presented in Xu and Milligan *et al.* [Xu and Mulligan, 2010] for imaging stitching.

It is important to mention that in general, the source and reference images are assumed to be gamma corrected. Vazquez-Corral and Bertalmío [Vazquez-Corral and Bertalmío, 2016] proposed the first work on colour stabilisation dealing with logarithmic encoded images. The method relies on finding a sufficiently large number of achromatic matches among source and reference, which in some situations may be a challenging limitation. In the next chapter, we will present our proposed method on colour stabilisation, for image pairs encoded with unknown non-linearities.

CHAPTER 8

Colour stabilisation for cinema content

Colour matching techniques aim to map the colours of one image, defined as source, to those of a second image, defined as reference. A particular case is colour stabilisation, where the two pictures are taken from the same scene and differ in terms of colour. These colour differences are due to the use of different camera models, or even when the same camera model is used under different settings (white balance, exposure time, aperture, etc.). The final output from most digital cameras is a non-linear image that has been encoded with a power law function, known as gamma correction. Although this encoding technique is still used for LDR content, currently cinema cameras use other non-linear encodings that are more suitable for HDR scenes, like logarithmic-encodings. In this context, we present in this work a colour stabilisation technique that adapts to any non-linear encoded input.

This chapter is based on the work presented in SMPTE 2017 ‘Color-matching Shots from Different Cameras Having Unknown Gamma or Logarithmic Encoding Curves’ [Gil Rodríguez et al., 2017], and its journal extension, in preparation, with the title ‘Color matching images with unknown non-linear encodings’.

Motivation: encoding techniques for cinema content

Digital cameras perform, in general terms, the following in-camera steps: demosaicing, white balance, colour correction (from RGB camera sensor to device

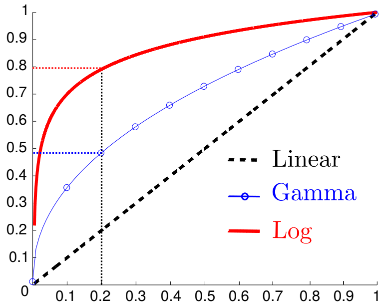


Figure 8.1: Linear response versus gamma-corrected and logarithmic response (perceptual). The gamma correction was defined as power $1/2.2$, and the logarithmic curve an ARRI Log C.

independent colour space), encoding standard (usually a gamma correction), and compression. Bianco *et al.* [Bianco et al., 2012] proposed a generic colour processing pipeline for digital cameras

$$I_{out} = (A \cdot I_{lin})^{1/\gamma}\,, \tag{8.1}$$

where I_{out} is the output image, A is a 3×3 matrix which carries colour information and white balance, γ value defines a power law function, and I_{lin} is the linear image read by the camera sensor after demosaicing. This is a simplified version of the pipeline, since other preprocessing techniques, like denoising, contrast enhancement, etc. might be applied.

Although gamma correction has been the most used encoding technique, it fails when working with high dynamic range (HDR) imaging, since in low luminance areas quantisation artifacts might appear. Current professional cinema cameras are able to capture a wide range of luminance, and therefore, a compression of the data is needed for storage, while preserving all the details and appearance. At present, cinema cameras allow to encode the final coloured image by replacing the gamma correction with a logarithmic function. The general form (common to the most popular log-encoding methods [Brendel, 2011], [Corporation, 2009]) can be expressed as:

$$I_{out} = c\log_{10}(a \cdot A \cdot I_{lin} + b) + d, \tag{8.2}$$

where I_{out} and I_{lin} are defined as above, and the parameters $a, b, c,$ and d are constant real values (varying for different camera manufacturers and camera settings). Figure 8.1 shows the plot of linear (dashed), gamma-corrected (circle) and logarithmic (continuous) responses to linear values. Notice that gamma correction and logarithmic curve assign respectively 50% and 80% of the output range to the 20% of the linear intensity values, respectively.

Our main contribution in this work is a method that i) converts images encoded following the type of Equation (8.2) to the type of Equation (8.1), and ii) that colour stabilises pair of images which were taken with unknown camera settings and/or unknown encoding non-linearities. The proposed method first modify the initial image, and then estimates a gamma correction, and a matrix that transforms the colours of a source image to a given reference. We compare our method against state-of-the-art colour matching methods, and we show that our results outperform the rest of the algorithms both quantitative and qualitatively.

The proposed framework

We present a colour stabilisation framework that takes as input an image pair encoded with gamma or logarithmic curves. The main steps of our method can be outlined as follows:

1. Given two input images, we correct the log-encoded image(s) in order to transfer the inputs to adopt the form of a regular gamma-corrected ones.
2. We colour stabilised the images by estimating a 4×4 matrix, and two power law values.
3. Finally, we undo the correction made in the first step if necessary (in case the original reference image is log-encoded).

We refer the reader to the flowchart of the proposed model in Figure 8.2.

From log-encoded to gamma-corrected images

Let us consider a log-encoded image as in Equation (8.2),

$$I_{\log} = c \log_{10}(a \cdot A \cdot I_{lin} + b) + d, \tag{8.3}$$

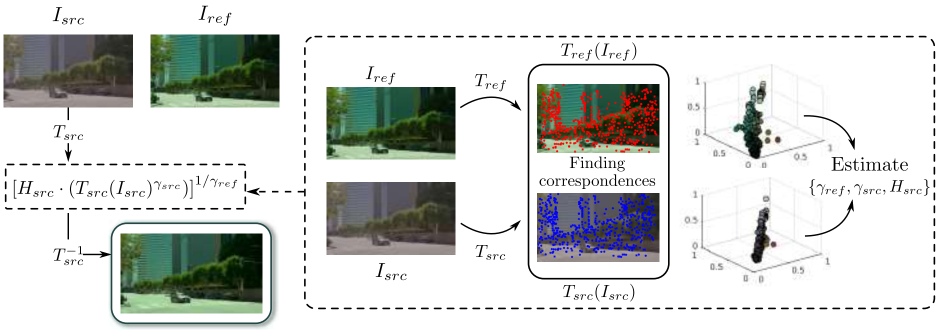


Figure 8.2: Flowchart of the proposed colour stabilisation method. For simplicity, the diagram shows only the colour matching from source image to reference one. Given two non-linear encoded images, reference (I_{ref}) and source (I_{src}), we apply the transformations T_{ref} and T_{src} to the image pair. These transformations are defined as the power 10 function $10^{\times}$, in case of a given log-encoded image; and as the identity Id, in case of gamma corrected input. Then, we compute a set of correspondences pts_{ref} and pts_{src}, using standard feature descriptor (e.g. in this chapter SIFT [Lowe, 1999]). From this set of corresponding pixel locations, we estimate the parameters $\{\gamma_{ref}, \gamma_{src}, H_{src}\}$ in the pixel values correspondences. The computed values are applied to the $T_{src}(I_{src})$ image. Finally, T_{src}^{-1} function is applied to the colour matched image.

If we apply a power 10 function to Equation (8.3), we obtain the following expression

$$\begin{aligned} 10^{I_{\log}} &= 10^{\log_{10}(a \cdot A \cdot I_{lin}+b)^{c}} \cdot 10^{d} \\ &= (a \cdot A \cdot I_{lin} + b)^{c} \cdot 10^{d}. \end{aligned} \tag{8.4}$$

In most of the definitions of the logarithmic curves, the value of parameter b is usually small. As Figure 8.3 shows, setting $b = 0$ does not change significantly the logarithmic curve. Notice that for the three different logarithmic curves (continuous lines), their equivalent curves fixing $b = 0$ (dashed lines) lie on top. Therefore, we can simplify Equation (8.3) by neglecting b,

$$10^{I_{\log}} = (a \cdot A \cdot I_{lin})^{c} \cdot 10^{d} = (K \cdot I_{lin})^{c}, \tag{8.5}$$

where $K = a \cdot A \cdot 10^{d/c}$ is a matrix with same size as A. We start from Equation (8.3), and by applying power 10 function to it, we end by getting an expression like in Equation (8.1). Therefore, given two log-encoded images I_{ref} and I_{src}, the first stage in our algorithm is to apply power 10 to obtain $10^{I_{ref}}$, and $10^{I_{src}}$. These two images behave as gamma corrected images.

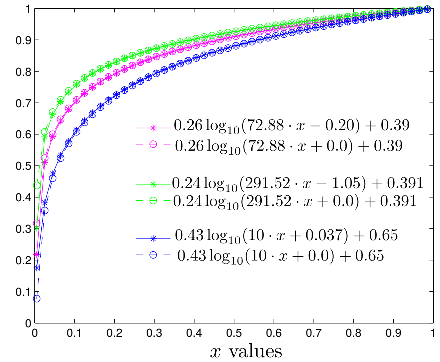


Figure 8.3: Graph of 3 logarithmic encoded curves in three different colours: Log C ARRI of EI 320 (green), Log C ARRI of EI 1280 (magenta) [Brendel, 2011], and S-Log [Corporation, 2009] (blue) plotted in continuous lines. In addition, the same logarithmic curves by setting $b = 0$ in their definitions (dashed lines). Note that the distance between the dashed and continuous lines from the same colour is small.

Colour stabilisation

We start by transforming, if needed, input images I_{ref} and I_{src} into gamma-corrected images I'_{ref} and I'_{src}. Then we compute a set of correspondences pts_{ref} and pts_{src}. Notice that we use SIFT [Lowe, 1999] for finding corresponding pixels, although it can be replaced by any other method. It is important to note that we compute the correspondences between $I'_{ref} \leftrightarrow I'_{src}$, and $I'_{src} \leftrightarrow I'_{ref}$, and select those that appear in both directions. This allows us to discard some potentially incorrect correspondences. Let us now define the pixel values in the corresponding locations of I'_{ref} and I'_{src} as

$$\{(R^{'}_{ref}, G^{'}_{ref}, B^{'}_{ref})^t\}_i, \text{ and } \{(R^{'}_{src}, G^{'}_{src}, B^{'}_{src})^t\}_i, \tag{8.6}$$

where $i = 1, \ldots, N$ denotes the number of correspondences. We follow the idea from the colour stabilisation model proposed in [Vazquez-Corral and Bertalmío, 2014a],

$$H_{src} \cdot {I'}_{src}^{\gamma_{src}} \sim {I'}_{ref}^{\gamma_{ref}}, \tag{8.7}$$

where H_{src} was a 3×3 matrix that transforms colours from the source to match the ones of the reference, and $\gamma_{ref}, \gamma_{src}$ are the gamma correction values. We extend H as a projective transformation with size 4×4 (inspired by the colour homography by [Gong et al., 2016], [Gong et al., 2017]). In this way, the model can deal not only with pixels in the core of the colour gamut, but also with those values that appear on the border, which are the most affected by gamut mapping and tone mapping. Then, from the set of correspondences, we can build a system of equations considering matrix size 4×4 and homogeneous coordinates,

$$\begin{aligned} H_{src} \cdot \begin{bmatrix} R'_{src} \\ G'_{src} \\ B'_{src} \\ 1 \end{bmatrix}^{\gamma_{src}} - \begin{bmatrix} R'_{ref} \\ G'_{ref} \\ B'_{ref} \\ 1 \end{bmatrix}^{\gamma_{ref}} &= 0, \qquad (8.8) \\ H_{ref} \cdot \begin{bmatrix} R'_{ref} \\ G'_{ref} \\ B'_{ref} \\ 1 \end{bmatrix}^{\gamma_{ref}} - \begin{bmatrix} R'_{src} \\ G'_{src} \\ B'_{src} \\ 1 \end{bmatrix}^{\gamma_{src}} &= 0, \end{aligned}$$

where $\{\gamma_{src}, \gamma_{ref}, H_{src}, H_{ref}\}$ are the unknowns. We propose a new single optimisation process, where the only constraint is $H_{ref} \cdot H_{src} \sim I$. The objective function considers the l_2-norm (in 3×1 non-homogeneous coordinates) on RGB, as well as the same norm in Lab colour space. Finally, the matrices and non-linearities are applied to the entire images as in Equation (8.7), and we obtain the colour matched images:

$$I''_{src} = \left[H_{src} \cdot I'^{\gamma_{src}}_{src} \right]^{1/\gamma_{ref}}. \qquad (8.9)$$

Undo power 10 function

If I_{src} was log-encoded, we apply a $\log_{10}$ function to the result of the previous step so as to undo the power 10 transform we applied at the beginning.

Results and Discussions

This section is divided into 4 different parts. First, we describe how we have created an image dataset for evaluation. Second, we compare our approach with seven popular colour matching methods. Third, we evaluate the performance of the rest of methods by applying the proposed preprocessing step in case of log-encoded images. Finally, we show results considering other non-linear encoding techniques coming from real cinema cameras.

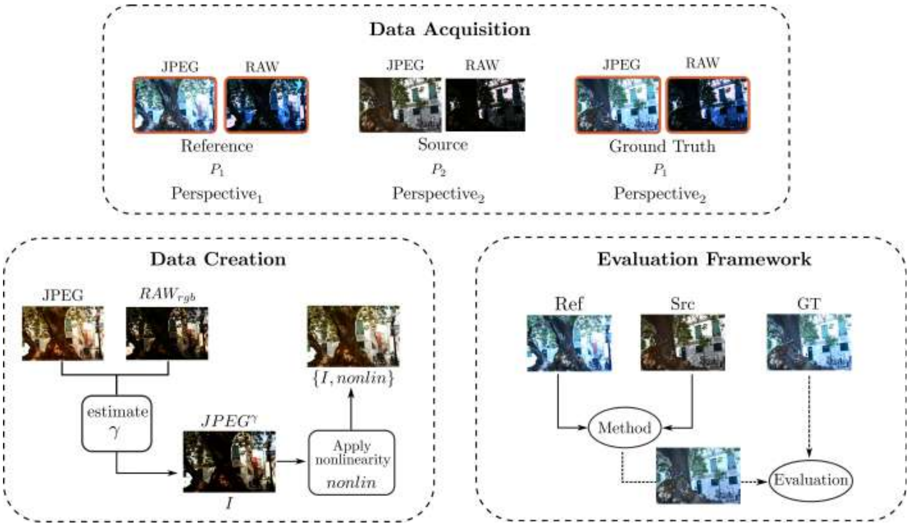


Figure 8.4: Evaluation framework. On the left, data acquisition is described. Pictures are taken from the same scene, and from two different points of view Perspective$_1$ and Perspective$_2$. From the first one, the reference image is taken, and from the second, the source and the ground truth. Images are stored in RAW and JPEG format, and we chose different camera settings P_1, for reference and ground truth, and P_2 parameters for the source. On the middle, data is created by linearising the JPEG image, i.e undoing the gamma correction I. Once linearised, a random non-linearity is applied $nonlin$, and the new image and the non-linearity are stored $\{I, nonlin\}$. Finally, the reference Ref and source Src become the input images for the colour matching methods, and the corresponding output is evaluated against the GT.

Dataset

Our data is composed of different scenes, where each of them contains a reference image Ref, a source image Src and a ground truth image GT. In order to acquire our data, we work in camera manual mode to have full control over exposure time, white balance, ISO value, and aperture. We stored RAW and JPEG formats for each image. In that way, we have the linear information read by the camera sensor (RAW), as well as the final compressed image (JPEG). Images were taken using two camera models, Nikon D3100 (12-bits) and Canon EOS80D (14-bits). Let us explain the steps we follow from acquisition to the final triplet Ref, Src, and GT images for each scene:

- Set the parameters of the camera (exposure time, white balance, ISO value, and aperture) P_1, the camera position Perspective$_1$, and acquire the Ref set RAW and JPEG.

- Change the camera position Perspective$_2$, and vary the camera settings to a different configuration P_2 to acquire the Src set.
- Use the same camera parameters as the Ref P_1, and the same camera position as the Src Perspective$_2$, then we acquire the GT set RAW and JPEG.
- For each pair (RAW, JPEG) define $\{I, nonlin\}$:
    1. Preprocess the RAW input to obtain an RGB linear image using *DCRAW* [Coffin, 2010] open source code, we refer to this image as RAW_{rgb}.
    2. Estimate the γ correction curve, between the preprocessed RAW_{rgb} and the JPEG using [Vazquez-Corral and Bertalmío, 2014a].
    3. Undo γ from the JPEG image in order to obtain a linear image called I with the camera colour processing still in.
    4. Apply a random generated non-linearity to I. In case of gamma correction, we set the range of values to be selected $[1.7, 2.7]$, and for logarithmic curves, we select the definitions from Log C ARRI (a total of 11 curves) [Brendel, 2011], and S-Log from Sony [Corporation, 2009]. We name the applied non-linearity $nonlin$.
    5. In case of GT, the same $nonlin$ as the one selected for the reference is applied.

Analysis of our colour stabilisation model

In this section, we explore the performance of our colour stabilisation model (Equation (8.9)), by modifying the size of the colour correction matrix, originally a 3×3 matrix. We redefine the linear transformation as an affine 3×4, and the projective 4×4 matrix (inspired by the colour homography of [Gong et al., 2016], [Gong et al., 2017]). We consider the affine transformation as particular case of a projective one, by adding a last row of zeros except the last value, which will be one. In this way, we redefine the RGB values as 4×1 vectors $[R, G, B, 1]^t$ (homogeneous coordinates). Thus, after performing the respective transformations, we go back to 3×1 size vectors, dividing all the elements by the last element of the resulting 4×1 vector, and discarding the last one.

In the quantitative evaluation we select the following colour metrics: ΔE^*_{00} [Sharma et al., 2005], PSNR of the luminance, colour *PSNR* defined as *CPSNR*, *CID* [Lissner et al., 2013], and root mean squared error (RMSE). In Table 8.1 the results of for the three matrix sizes are shown. We can observe that the performance of 3×4 and 4×4 matrices across different metrics is significantly better than the 3×3 transformation matrix. The most noticeable improvement is in the PSNR of the red colour channel, boosting almost 0.9 in decibels. This affects directly the performance of the CPSNR metric.

Table 8.1: Study of our approach considering different matrix size (3×3, 3×4, and 4×4). Results show mean (μ) and median ($\hat{\mu}$) among 35 pairs, where both reference and source are γ-corrected images.

		$\Delta \mathbf{E}^*_{00}$	**PSNR L**	**CPSNR**	**CID**	**RMSE**
3x3	μ	3.726	27.420	26.116	0.164	0.054
	$\hat{\mu}$	3.554	27.228	24.965	0.149	0.056
3x4	μ	3.312	27.578	26.745	0.153	0.050
	$\hat{\mu}$	3.197	27.205	26.318	0.134	0.048
4x4	μ	**3.263**	**27.650**	**26.907**	**0.145**	**0.049**
	$\hat{\mu}$	**3.092**	**27.271**	**26.576**	**0.125**	**0.047**

In Figure 8.5 we show the reference image (first column and first row), the source (second column and first row). The GT is presented in first column and second row. Then, from the second till the fourth column we show the colour stabilised images after applying our method, by varying the matrix sizes. For each result of the different matrix sizes 3×3, 3×4 and 4×4, we show some crops in the second row. Notice the differences in colour within the colourchecker. In case of the 3×3 matrix (the second column), we can appreciate that the colour of the orange in the tree does not match the colour of the GT. It can also be noticed some bluish colour cast in the whole image. However, the results obtained with affine and projective transform show better matching between colourcheckers, as well as the colour of the oranges. The difference between these two, although very subtle, appear in the door, in which 4×4 result recovers in more detail the bars of the front door. From now on, we will report our results for the 4×4 case.

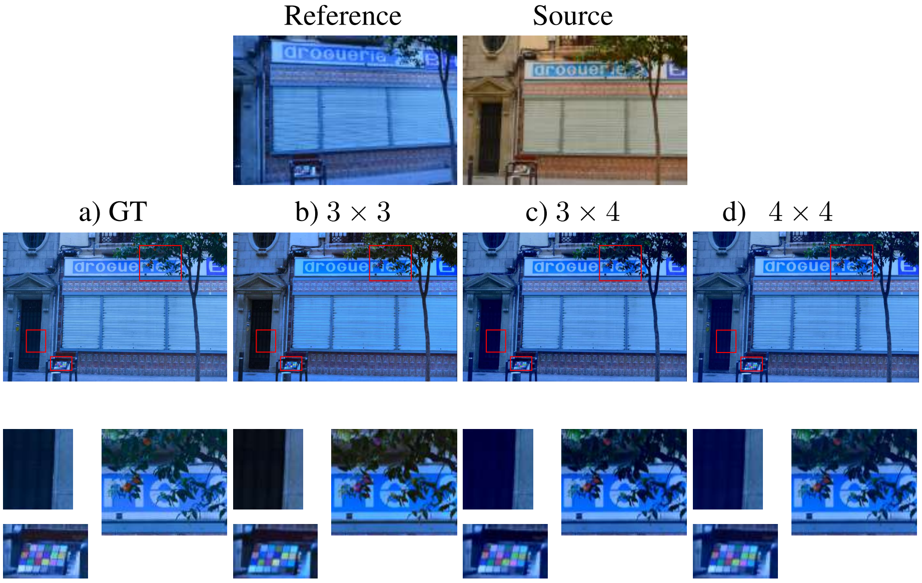


Figure 8.5: Results of varying matrix size in our colour stabilisation method. On the first row, the reference and source images. On the second row, we present the GT, and the rest of the columns the outputs of considering a 3×3, 3×4, and 4×4 matrix, repectively. On the last row, we show crops of the GT and our 3 different results.

Experiments versus state-of-the-art

We evaluate our approach against seven state-of-the-art methods: Reinhard *et al.* [Reinhard et al., 2001] (Reinhard), Kotera [Kotera, 2005] (Kotera), Xiao and Ma [Xiao and Ma, 2006] (Xiao), Pitié *et al.* [Pitié et al., 2007] (Pitie), Ferradans *et al.* [Ferradans et al., 2014] (Ferradans), Park *et al.* [Park et al., 2016] (Park), HaCohen *et al.* [HaCohen et al., 2011] (HaCohen), and Gil Rodríguez *et al.* [Gil Rodríguez et al., 2017] (Gil). We want to emphasise that for Pitié *et al.* [Pitié et al., 2007], we focus only on the global part of the method. We studied all possible combinations of applied non-linearities to the reference and source image: i) two gamma-corrected images, ii) two log-encoded images, iii) one gamma-corrected as reference and one log-encoded as source, and iv) one log-encoded as reference and one gamma-corrected as source. In the quantitative evaluation we select the following colour metrics: ΔE^*_{00} [Sharma et al., 2005], *PSNR* of luminance channel (PSNR L), colour *PSNR* defined as *CPSNR*, *CID* [Lissner

Table 8.2: Results from the comparison among 35 image pairs for: 1) two γ-encoded images, 2) two Log-encoded images, 3) reference Log-encoded and source γ-corrected, and 4) reference γ-corrected and source Log-encoded.

		ΔE^*_{00}		**PSNR L**		**CPSNR**		**CID**		**RMSE**	
		μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$
Ref γ-Src γ	**Kotera**	11.111	7.686	21.122	23.877	19.786	21.040	0.458	0.394	0.145	0.089
	Pitie	3.567	3.394	26.162	25.946	25.696	25.769	0.174	0.157	0.055	0.051
	Reinhard	4.777	4.652	25.525	25.162	23.904	23.571	0.205	0.191	0.068	0.066
	Xiao	4.377	4.232	25.940	26.077	25.183	25.270	0.196	0.160	0.059	0.055
	Ferradans	5.522	5.308	23.715	23.874	23.028	22.560	0.260	0.237	0.078	0.074
	Park	3.428	**3.020**	27.604	**27.381**	26.595	26.384	0.157	0.134	0.051	0.048
	Gil	3.726	3.554	27.420	27.228	26.116	24.965	0.164	0.149	0.054	0.056
	Ours	**3.263**	3.092	**27.650**	27.271	**26.907**	**26.576**	**0.145**	**0.125**	**0.049**	**0.047**
Ref log-Src log	**Kotera**	14.234	8.381	18.586	21.081	17.615	19.676	0.551	0.481	0.179	0.104
	Pitie	3.978	4.044	25.797	25.369	25.119	25.099	0.207	0.201	0.059	0.056
	Reinhard	7.878	7.916	22.656	22.512	19.899	19.369	0.364	0.368	0.107	0.108
	Xiao	5.632	5.599	24.330	23.910	23.199	23.190	0.272	0.264	0.072	0.069
	Ferradans	8.587	7.047	19.351	20.831	18.925	20.250	0.395	0.325	0.128	0.097
	Park	6.768	4.548	26.217	26.162	23.961	24.196	0.296	0.210	0.083	0.062
	Gil	4.057	3.644	27.027	26.689	25.665	25.379	0.193	0.155	0.057	0.054
	Ours	**3.400**	**3.022**	**27.446**	**27.158**	**26.587**	**26.479**	**0.161**	**0.135**	**0.050**	**0.047**
Ref log-Src γ	**Kotera**	15.704	12.405	17.017	18.864	15.970	16.625	0.631	0.586	0.199	0.148
	Pitie	3.909	3.830	25.796	25.498	25.225	25.020	0.200	0.201	0.059	0.056
	Reinhard	7.928	7.516	21.260	21.056	18.883	18.687	0.393	0.392	0.117	0.116
	Xiao	7.926	7.554	21.446	20.539	20.438	20.059	0.403	0.416	0.100	0.099
	Ferradans	8.578	7.954	19.654	19.163	19.172	18.518	0.381	0.369	0.122	0.119
	Park	5.895	5.242	24.038	23.352	22.972	22.294	0.305	0.290	0.078	0.077
	Gil	4.066	3.667	27.102	26.847	25.741	25.627	0.188	0.178	0.058	0.052
	Ours	**3.377**	**3.140**	**27.571**	**27.606**	**26.712**	**26.632**	**0.157**	**0.129**	**0.050**	**0.047**
Ref γ-Src log	**Kotera**	12.658	9.202	18.629	20.748	17.893	20.089	0.538	0.430	0.162	0.099
	Pitie	3.752	3.903	25.957	25.538	25.378	25.217	0.184	0.173	0.057	0.055
	Reinhard	6.438	6.246	22.861	22.642	21.776	21.666	0.291	0.291	0.084	0.083
	Xiao	6.794	5.734	23.023	22.770	22.097	22.215	0.322	0.314	0.081	0.077
	Ferradans	6.317	6.165	22.222	21.826	21.577	21.357	0.318	0.298	0.089	0.086
	Park	12.808	9.620	20.746	22.593	18.779	19.351	0.510	0.454	0.147	0.108
	Gil	3.863	3.476	27.197	26.672	25.889	25.341	0.173	0.162	0.054	0.054
	Ours	**3.444**	**3.313**	**27.395**	**26.922**	**26.563**	**25.684**	**0.152**	**0.144**	**0.050**	**0.052**

et al., 2013], and *root mean squared error* (RMSE). For each metric we show the mean (μ) and the median ($\hat{\mu}$). Notice that in order to compare the colour stabilised and the GT in case of log-encoded images, we first undo the non-linearity (since it is known) from the result and GT, and then we apply a gamma-correction of value $1/2.2$. We use the data computed as described in Section 8.3.1, which consists of 35 image pairs for each comparison.

In the quantitative comparisons we show in green the best results, and then in blue and orange, the second and third rated, respectively. In Figures 8.6 and 8.7, the log-encoded images are shown in BT.709.

Gamma-corrected inputs

The first block in Table 8.2 presents the results of pairs encoded both by gamma correction in the first block. In most of the metrics, our method outperforms the rest of algorithms. Except for median values of ΔE^*_{00} and PSNR L, in which [Park et al., 2016] obtains better results. The method of Park is ranked second, and [Pitié et al., 2007] and [Gil Rodríguez et al., 2017] dispute the third best ranked.

In the first two rows of Figure 8.6, we show the results for two different scenes. The first column shows the reference, the second the source, and the third the GT. In this example, we compare our algorithm (last column, first and second row) against the method of [Ferradans et al., 2014] (fourth column, first row), and [Gil Rodríguez et al., 2017] (fourth column, second row). In the first scene, [Ferradans et al., 2014] introduces gray colours in the output of the floor. From last scene, [Gil Rodríguez et al., 2017] method cannot recover the blueish intensity of the wall.

Log-encoded inputs

We present the results of 35 image pairs, this time the inputs were encoded with logarithmic curves. In the second block of Table 8.2, it is shown that our method outperforms the rest of algorithms in all the metrics.

From Figure 8.6 (third and fourth rows), [HaCohen et al., 2011] washes out the green colour. In the second scene, the output from [Pitié et al., 2007] cannot recover the red colour of the garage in the background, and it presents a purplish colour in one of the doors, and it makes appear some clouds on the sky.

Log-encoded reference and gamma-corrected source

In this comparison, the reference image is log-encoded, and the source is a gamma-corrected image. In the third block in Table 8.2, our proposed method outperforms the rest of the algorithms in all metrics. The method of [Gil Rodríguez et al., 2017] and the method of [Pitié et al., 2007] ranked second and third, respectively.

Figure 8.6: Results of all the methods for the four comparisons. Each block represents: 1) gamma-corrected image pair, 2) log-encoded input images, 3) log-encoded reference and gamma-corrected source and 4) gamma-corrected reference and log-encoded source. The first column presents the reference, the second shows the source, the third the GT, the fourth the methods result, and the last our result. We present for 1) [Ferradans et al., 2014] and [Gil Rodríguez et al., 2017] methods, 2) [HaCohen et al., 2011] and [Pitié et al., 2007] methods and 3) [Xiao and Ma, 2006] and [Park et al., 2016] methods, and 4) [Reinhard et al., 2001] and [Kotera, 2005] methods.

Figure 8.6 shows the results from [Xiao and Ma, 2006] (fifth row), and [Park et al., 2016] (sixth row). In the first scene, notice that [Xiao and Ma, 2006] enhanced yellow and red colours, and it saturates the upper right corner of the wall. [Park et al., 2016] method presented colour shift in the floor, in addition it intensifies purple colour on the right side.

Gamma-corrected reference and log-encoded source

The last block in Table 8.2 presents the results where the reference is a gamma-corrected image, and the source is log-encoded one. For all the metrics, our method outperforms the rest of algorithms. [Gil Rodríguez et al., 2017] method ranks second, and results from ΔE^*_{00}, CPSNR, CID and RMSE metrics are very close to [Pitié et al., 2007].

Figure 8.6 shows the results from [Reinhard et al., 2001] and [Kotera, 2005] (last two rows). The result from [Reinhard et al., 2001] shows a yellowish cast on the wall. The method of [Kotera, 2005] presents washed out colours, see the chair and the wall behind it.

Table 8.3: Results show mean (μ) and median ($\hat{\mu}$) among 22 pairs, where reference and source images are encoded using logarithmic curves.

		ΔE^*_{00}	PSNR L	CPSNR	CID	RMSE
HaCohen	μ	3.119	**27.700**	**27.243**	0.149	**0.046**
	$\hat{\mu}$	**2.836**	26.619	26.548	0.139	0.047
Ours	μ	**3.089**	27.695	27.093	**0.145**	0.047
	$\hat{\mu}$	2.989	**27.219**	**26.553**	**0.134**	**0.047**

Let us note that the method of [HaCohen et al., 2011] is not presented in Table 8.2, since it did not compute solutions for all the image pairs. In case 1), it did not perform results for 10 of the images, which represents almost the 30% of total image pairs. In case 2), 13 of the image pairs were not computed. In cases 3) and 4), [HaCohen et al., 2011] method does not provide result for 15 and 20 image pairs, respectively. In Table 8.3 we present [HaCohen et al., 2011] and our results, only on those image pairs computed by [HaCohen et al., 2011] method in case 2). In that reduced set of images HaCohen's results are comparable with ours, but its lack of results makes it unsuitable for real-like scenarios.

Experiments with power 10

In this section we want to check what happens if we perform the proposed preprocessing in order to adapt the log-encoded images as gamma-corrected, in all the methods we compare with.

Table 8.4: Results from the comparison among 35 image pairs for: 1)) two Log-encoded images, 2) reference Log-encoded and source γ-corrected, and 3) reference γ-corrected and source Log-encoded. In this case, we applied power 10 to the inputs (if necessary) for the rest of algorithms, except Gil.

		ΔE^*_{00}		PSNR L		CPSNR		CID		RMSE	
		μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$
Ref log-Src log	**Kotera 10**	11.919	8.285	20.771	22.503	19.286	20.719	0.487	0.403	0.148	0.092
	Pitie 10	3.790	4.074	25.961	25.871	25.377	25.332	0.197	0.188	0.058	0.054
	Reinhard 10	5.238	5.216	24.824	23.975	23.190	23.178	0.231	0.211	0.072	0.069
	Xiao 10	4.758	4.507	25.623	25.002	24.646	24.542	0.219	0.216	0.063	0.059
	Ferradans 10	9.885	7.208	18.695	19.915	18.318	19.464	0.419	0.316	0.145	0.106
	Park 10	4.137	3.843	26.858	26.733	25.626	24.901	0.204	0.170	0.057	0.057
	Gil	4.057	3.644	27.027	26.689	25.665	25.379	0.193	0.155	0.057	0.054
	Ours	**3.400**	**3.022**	**27.446**	**27.158**	**26.587**	**26.479**	**0.161**	**0.135**	**0.050**	**0.047**
Ref log-Src γ	**Kotera 10**	11.775	8.609	20.918	22.022	19.308	20.385	0.485	0.407	0.147	0.096
	Pitie 10	3.697	3.681	25.991	25.487	25.487	25.258	0.189	0.184	0.057	0.055
	Reinhard 10	5.379	5.082	25.034	25.086	23.259	23.384	0.235	0.216	0.072	0.068
	Xiao 10	4.905	4.565	25.556	25.031	24.599	24.649	0.227	0.205	0.064	0.059
	Ferradans 10	7.649	6.155	20.637	21.916	20.167	21.338	0.353	0.318	0.113	0.086
	Park 10	3.773	3.250	27.076	**27.734**	26.201	**26.897**	0.179	0.154	0.054	**0.045**
	Gil	4.066	3.667	27.102	26.847	25.741	25.627	0.188	0.178	0.058	0.052
	Ours	**3.377**	**3.140**	**27.571**	27.606	**26.712**	26.632	**0.157**	**0.129**	**0.050**	0.047
Ref γ-Src log	**Kotera 10**	11.564	8.588	20.466	22.328	19.164	20.831	0.480	0.379	0.148	0.091
	Pitie 10	3.721	3.895	26.064	25.783	25.477	25.508	0.183	0.165	0.056	0.053
	Reinhard 10	5.579	5.300	24.691	24.479	23.073	22.759	0.250	0.256	0.074	0.073
	Xiao 10	4.847	4.699	25.100	25.252	24.345	24.667	0.219	0.206	0.063	0.058
	Ferradans 10	6.280	6.010	22.433	22.802	21.903	22.115	0.311	0.312	0.084	0.078
	Park 10	6.262	4.023	26.168	26.592	24.278	24.857	0.269	0.192	0.080	0.057
	Gil	3.863	3.476	27.197	26.672	25.889	25.341	0.173	0.162	0.054	0.054
	Ours	**3.444**	**3.313**	**27.395**	**26.922**	**26.563**	**25.684**	**0.152**	**0.144**	**0.050**	**0.052**

Log-encoded inputs

Results show a considerable improvement between the original methods and after applying power 10, see first block in Table 8.4. The only exception is the algorithms of [Pitié et al., 2007] and [Ferradans et al., 2014], which have a performance similar with and without power 10. In this case, the method of [Park et al., 2016] is ranked second and/or third in all of the metrics, whereas in the previous performance, it was ranked third by only two metrics.

Figure 8.7 presents a comparison of the results of the methods with and without the power 10 preprocessing, in first and second rows. The first column corresponds to the GT, the second the output of the methods, the third the output of the methods after applying power 10, and last column shows our result. Notice that [Pitié et al., 2007] and [HaCohen et al., 2011] perform very similar with and without power 10. Nonetheless, [Pitié et al., 2007] after applying power 10 looks closer to the GT on the red doors of the fire fighters garage.

Log-encoded reference and gamma-corrected source

Results show a considerable difference between the original method and after applying the power 10 preprocessing. Notice the boosting of [Park et al., 2016] , which improves significantly versus its original version. It is ranked second after our approach, and in median PSNR L and CPSNR it gets the best results, see Table 8.7 (second block). In Figure 8.7, the result from [Xiao and Ma, 2006] after applying power 10 shows no saturation on the upper right corner of the wall, and in addition the red cast appearing in the previous result is removed. [Park et al., 2016] method present no colour shift on the floor, although it cannot completely recover the yellow colour of the truck. We emphasise that power 10 function boosts [HaCohen et al., 2011], from computing 57% of the cases, to reaching almost 80%. Once again, the method of [HaCohen et al., 2011] is not presented in this Table 8.4, since it did not compute a solution for 8 of the images.

Gamma-corrected reference and log-encoded source

In this last comparison, although [Park et al., 2016] improves their previous results, it is not as noticeable as in the previous comparison. In this case it ranks third only in PSNR L metric, see Table 8.4 last block. The method of [Pitié et al., 2007] outperforms [Park et al., 2016] as opposed to previous case. [Park et al., 2016] in this context shows a more constant performance in both cases. Notice that in Figure 8.7, [Reinhard et al., 2001] method presents noticeable difference between their original versions, and after applying power 10 function to the inputs (fifth row). In the fifth row and fourth column, we observe how the colour of the wall is brighter and closer to the greyish tone of the GT. The result of [Kotera,

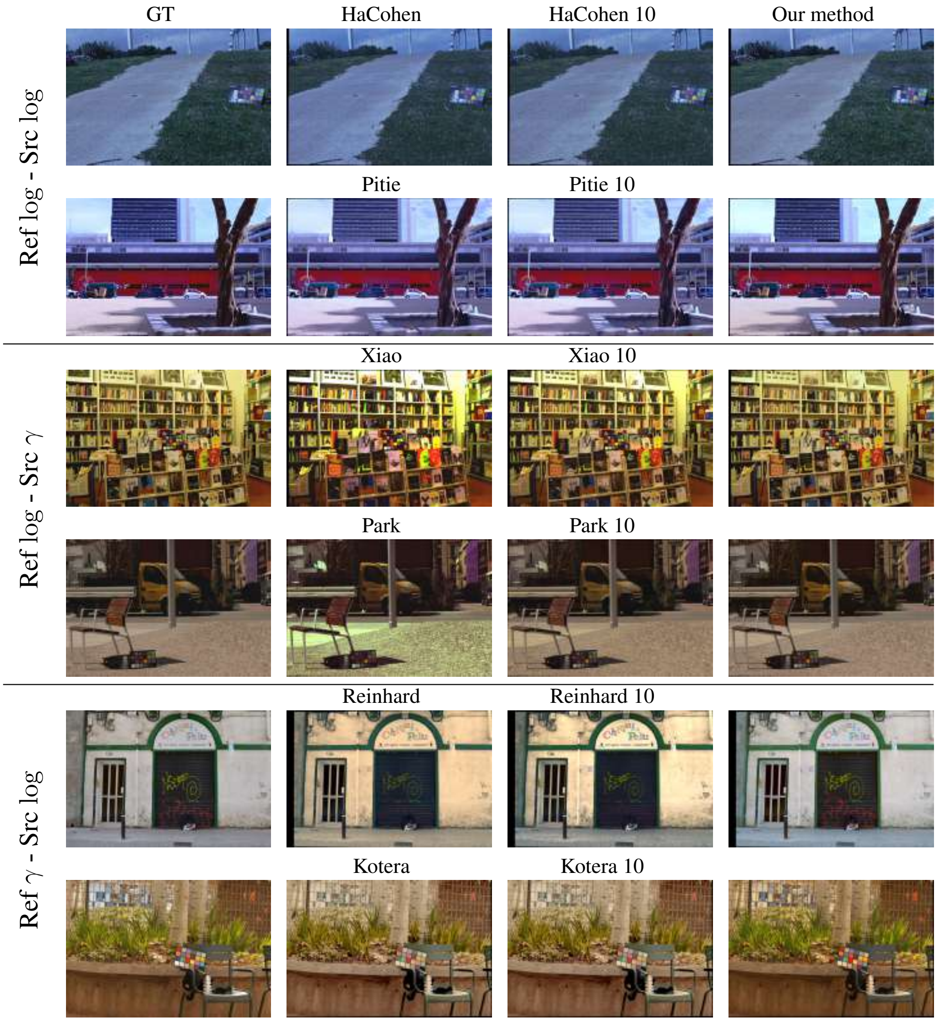


Figure 8.7: Results from applying power 10 to log-encoded inputs in all the methods. Each block represents: 1) log-encoded input images, 2) log-encoded reference and gamma-corrected source and 3) gamma-corrected reference and log-encoded source. The first column presents the GT, the second shows the output of the original method, the third shows the output of the method applying power 10, and the fourth our result. We present for 1) HaCohen and Pitie's methods, 2) Xiao and Park's methods and 3) Reinhard and Kotera's methods.

2005] when applying power 10 is similar to the result of the original method, nonetheless we can appreciate a subtle enhanced of contrast and more intense green colour. The method of [HaCohen et al., 2011] is not presented in this Table 8.4, since it did not compute a solution for 12 of the images. Nonetheless, it recovers 8 more solutions when power 10 function is applied.

The results of our experiments show that the proposed framework (applying a power 10 function to log-encoded images) boosts the performance of majority of the methods we compare with. This is true both in terms of quantitative metrics and image quality. With the exception of the algorithms of Pitié *et al.* [Pitié et al., 2007] and Ferradans *et al.* [Ferradans et al., 2014], since those are based on the probabilities of the images, and those are transferred even when power 10 function is applied.

Beyond gamma and log-encoded images: HLG and PQ

Current digital cinema cameras can encode images by using a logarithmic function. Nonetheless, there exist other decoding/encoding curves such as Perceptual Quantizer (PQ) [SMPTE, 2014] and Hybrid Log-Gamma (HLG) [Borer and Cotton, 2016]. These two non-linear encodings are well-known and defined. HLG is a scene-referred encoding,

$$E' = \begin{cases} r\sqrt{E} & \text{if } 0 \leq E \leq 1 \\ a\log(E-b)+c & \text{if } 1 < E \text{ is odd,} \end{cases} \tag{8.10}$$

where E is the signal, and r, a and b are real numbers. As opposite of HLG encoding, PQ is a display-referred decoding curve, and it needs to set the absolute luminance of the display. PQ is defined as,

$$Y = L\left(\frac{V^{\frac{1}{n}} - C_1}{C_2 - C_3 V^{\frac{1}{m}}}\right)^{\frac{1}{n}}, \tag{8.11}$$

where V is the input signal, Y is the display luminance, and $L,\ m,\ n,\ C_1,\ C_2,\ C_3$ are real values.

In this section, we colour match pair of images encoded using different transfer functions. The dataset we use for experiments is the one provided by ARRI in [Andriani et al., 2013]. This data contains high dynamic range images. We select 10 image pairs from the dataset, and apply PQ, HLG and Log C ARRI curves to the linear RAW data. In the case of PQ curve, we set up the absolute luminance of the display to 1000 cd/m^2. We colour stabilise all the possible combinations of these figures.

Table 8.5: Results show mean (μ) and median ($\hat{\mu}$) among 10 pairs, where reference and source images are encoded using HLG, PQ and logarithmic curves.

		ΔE^*_{00}	**PSNR L**	**CPSNR**	**CID**	**RMSE**
Kotera	μ	3.344	32.505	30.567	0.110	0.045
	$\hat{\mu}$	4.117	30.379	29.061	0.114	0.052
Pitie	μ	1.022	40.047	40.134	0.035	0.021
	$\hat{\mu}$	0.582	43.069	42.568	0.004	0.006
Reinhard	μ	1.861	35.311	35.020	0.062	0.040
	$\hat{\mu}$	2.058	32.023	31.408	0.042	0.037
Xiao	μ	1.891	32.965	32.789	0.061	0.032
	$\hat{\mu}$	2.066	30.379	30.244	0.054	0.033
Ferradans	μ	4.820	24.692	24.624	0.183	0.073
	$\hat{\mu}$	3.865	25.364	25.485	0.145	0.052
Park	μ	1.624	38.250	36.795	0.044	0.029
	$\hat{\mu}$	1.418	37.730	37.142	0.018	0.017
HaCohen	μ	0.657	43.933	43.076	0.006	0.007
	$\hat{\mu}$	0.494	44.474	43.809	0.002	**0.004**
Gil	μ	1.775	36.086	35.636	0.068	0.029
	$\hat{\mu}$	1.659	31.674	31.295	0.046	0.022
Ours	μ	**0.310**	**48.324**	**47.649**	**0.002**	**0.005**
	$\hat{\mu}$	**0.239**	**49.435**	**49.131**	**0.001**	**0.004**

We compare our method, described in Section 8.2, with the algorithms presented in the previous experiments: Reinhard *et al.* [Reinhard et al., 2001] (Reinhard), Kotera [Kotera, 2005] (Kotera), Xiao and Ma [Xiao and Ma, 2006] (Xiao), Pitié *et al.* [Pitié et al., 2007] (Pitie), Ferradans *et al.* [Ferradans et al., 2014] (Ferradans), Park *et al.* [Park et al., 2016] (Park), HaCohen *et al.* [HaCohen et al., 2011] (HaCohen), and Gil Rodríguez *et al.* [Gil Rodríguez et al., 2017] (Gil). In order to compute the quantitative results, we undo the non-linearity (since it is known) of the resulting colour matched image and the GT, and then apply a γ correction of $1/2.2$, as performed in previous experiments. From the data in Table 8.5, it is apparent that our method is accurate when working with real data

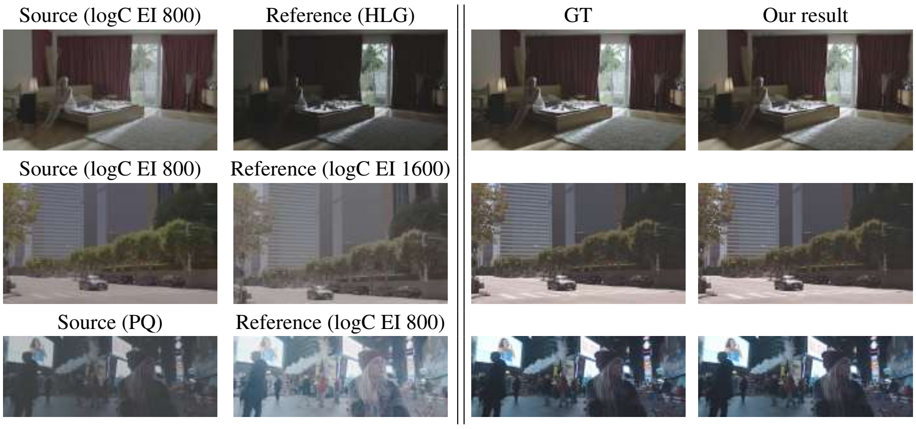


Figure 8.8: ARRI dataset. From left to right, source, reference, GT and our result. Each row represents a different comparison and scenario. The GTs and our results are tone mapped using [Cyriac et al., 2016].

and common situations. We compare our results against the methods considered above, using their original versions. These HDR images contain a high percentage of all the pixel values within very low levels, boosting the metric results for all the methods.

Figure 8.8 presents the image results, where we show the GTs and our results after applying tone mapping operator (TMO) from [Cyriac et al., 2016], instead of the gamma correction. The reference and the source are displayed directly, in order to appreciate the differences between applying the different curves. As it can be seen from last column in Figure 8.8, our method recovers the colours and appearance of the reference image, in different input situations. We show for 3 different scenes (rows), and for each scene: the reference (first column), source (second column), GT (third column) and our result (last column). Notice that on the last row, where reference is PQ and source HLG encoded, our result (last column) is not able to completely recover the blue on the t-shirt on the left upper corner. In our output, the blue appears brighter than in the GT. This is due to the fact that no correspondences are able in this particular blue, thus the recovery is not perfect.

Conclusion

We have presented a method for colour matching any combination of non-linear encoded images: gamma-corrected to gamma-corrected, logarithmic-encoded to gamma-corrected or logarithmic-encoded to logarithmic-encoded. The method is based on the modification of logarithmic-encoded images so as they behave as gamma-corrected ones. In this way, we can colour stabilise the images by estimating a 4×4 matrix and a power law value. Our results show that our method outperforms state-of-the-art algorithms quantitatively and qualitatively, considering all four different cases. In addition, we showed that our algorithm works as well for other non-linear enconding curves, such as HLG and PQ. We believe this work would encourage to consider further non-linear encodings in current camera colour processing models. In a future work, we would like to explore the more general case, when no content is shared among the input images.

CHAPTER 9

Conclusions and Future Work

The main objectives of this thesis were to propose HDR methods for still images and video, and to present a colour stabilisation method for images encoded with unknown non-linearities. With this purpose, we looked into the camera inner processes to be able to understand and redefine these problems in terms of digital camera knowledge.

We studied and analysed HDR reconstruction methods from multiple exposure images. Most of HDR methods reconstruct the final image by first estimating the camera response function (CRF). In Chapter 5, our implementation follows a different path. First, we showed that the building assumptions of CRF estimation are not correct. These assumptions are the uniqueness of the CRF, the independence of colour channels, and the full recovery of the dynamic range of the scene. We prove the first assumption is wrong by estimating the γ non-linearity applied when varying the exposure time between the RAW and its corresponding JPEG image. Our results indicates that gamma correction can be adapted to different exposure times, and thus it breaks the uniqueness assumption of the CRF. It is well known that colour channels in $sRGB$ space are correlated, and consequently not independent. We showed this fact by computing the colour transformation between the RAW and JPEG image, and highlighting that the matrix was not just diagonal, but diagonal dominant with non-zero values outside the diagonal. Finally, although the multiple exposure techniques aim to recover the whole DR of the scene, we showed that the presence of glare due to the optics of the camera reduces the DR that the camera can recover, and this is smaller than the actual

DR of the scene.

These observations make us to present a method to create HDR images from a stack of non-linear images. We show that by setting one of the images as the reference (the best exposed one) we can transform the rest of the stack by mapping them into the ‘space’ of the reference one. It is based on the definition of the colour processing pipeline, and in this way we can relate images from the same scene from a theoretical point of view. This transformation step let us work with a ‘linearised’ set of images that can be combined linearly to produce the final image containing details in dark and bright areas. We introduced also the possibility of choosing several images as reference, since it will help to preserve details in very dark/bright areas, and more robustly and independently than from the selection of just a single reference.

In Chapter 6, we proposed a method for HDR video generation from common DSLR cameras. One of the main issues of HDR video is the presence of movement in the scene, since it might cause ghosting artefacts on the final results. For this reason, we used a particular software available for many of the current DSLR cameras. This software gives the option of capturing video in high definition by alternating rows of two different ISO values. In this way, all the rows are captured at the same time, regardless of the ISO value used, and then the appearance of ghosting artefacts can be diminished. We first reconstruct two images of full resolution with these two different ISO values. For that, we modified a deinterlacing algorithm proposed for video. Then, these two images are fused linearly to create HDR video frames, and finally a TMO used for video is applied to obtain the final HDR video ready to be displayed.

The development of HDR technology is changing the encoding techniques that were the standard for LDR content. Many professional cameras record in logarithmic instead of γ power law function. In Chapter 8, we presented a technique that modifies logarithmic encoded images in order to make them behave as γ corrected images. In this way, we can colour match image pairs encoded with logarithmic curves in the same way as the colour stabilisation problem given two gamma corrected images. We showed that this modification could also improve existing colour matching techniques to handle log-encoded inputs. Moreover, we studied the performance of the colour processing pipeline model considering a projective transformation. This approximation is more accurate when dealing with pixels that are close to the border of the colour gamut.

Future work

Our current implementation of the method in Chapter 5, an HDR reconstruction from a set of multiple exposed images, sets one image as the reference, and the rest of images within the stack are transformed to the space of the 'reference one'. In order to transform them, we set as inputs the reference image, and the image in the stack that we want to map. In this context, a different implementation was also done, in which the images of the stack are transformed to the space of the 'reference one', by composing all the intermediate mappings till reaching the reference image. For example, in a stack of 5 images, if we select image 3 as the reference, in order to transform the first image into the reference, we will compute the mapping from image 1 to image 2, and then compose it with the mapping from image 2 to 3. It will be interesting, within this new implementation, to perform an exhaustive analysis of the improvements, and/or advantages, and limitations of it. On the one hand, we believe this new implementation will estimate better correspondences in case there exist large differences between images exposure times; and on the other hand, we should take into account that errors in the parameters estimation would be propagated. It will also be relevant to investigate weighting functions that can be used in order to combine the 'linearised' images in a more efficient manner.

Our implementation of HDR video and [Sintes, 2017] are not yet real-time. It will be interesting to code those parts that have been already parallelised in the GPU using CUDA programming. We would like to automatically obtain the values of the algorithm parameters depending on the image content, since these are image dependent.

Regarding colour stabilisation approaches, we would like to extend our work to tackle the more general case in which no content is shared between images. This will change completely the current definition we made of the problem, but nonetheless we could still use the fact that colour transformations can be expressed as matrix multiplications. It will also be interesting the study of different possibilities that will better approximate colours at the boundary of the gamut. For example, by considering a different γ function for each channel.

Bibliography

[Adams et al., 2016] Adams, W. J., Elder, J. H., Graf, E. W., Leyland, J., Lugtigheid, A. J., and Muryy, A. (2016). The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports*, 6.

[Aguerrebere et al., 2014] Aguerrebere, C., Delon, J., Gousseau, Y., and Musé, P. (2014). Best Algorithms for HDR Image Generation. A Study of Performance Bounds. *SIAM Journal on Imaging Sciences*, 7(1):1–34.

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

[Akbarinia and Párraga, 2018] Akbarinia, A. and Párraga, C. A. (2018). Colour constancy beyond the classical receptive field. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(9):2081–2094.

[Akiyama et al., 2015] Akiyama, H., Tanaka, M., and Okutomi, M. (2015). Pseudo Four-Channel Image Denoising for Noisy CFA Raw Data. *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 1–5.

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

[Aydin et al., 2008a] Aydin, T., Mantiuk, R., and Seidel, H.-P. (2008a). Extending Quality Metrics to Full Luminance Range Images. *Human Vision and Electronic Imaging XIII*, 6806.

[Aydin et al., 2008b] Aydin, T. O., Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2008b). Dynamic Range Independent Image Quality Assessment. *ACM Transactions on Graphics (TOG)*, 27(3):69.1–69.10.

[Aydin et al., 2014] Aydin, T. O., Stefanoski, N., Croci, S., Gross, M., and Smolic, A. (2014). Temporally Coherent Local Tone Mapping of HDR Video. *ACM Transactions on Graphics (TOG)*, 33(6):196.1–196.13.

[Backhaus et al., 1998] Backhaus, W., Kliegl, R., and Werner, J. (1998). *Color Vision: Perspectives from Different Disciplines*. Walter de Gruyter.

[Ballester et al., 2007] Ballester, C., Bertalmío, M., Caselles, V., Garrido, L., Marques, A., and Ranchin, F. (2007). An Inpainting-Based Deinterlacing Method. *IEEE Transactions on Image Processing (TIP)*, 16(10):2476–2491.

[Banterle et al., 2011] Banterle, F., Artusi, A., Debattista, K., and Chalmers, A. (2011). *Advanced High Dynamic Range Imaging: Theory and Practice*. A. K. Peters, Ltd.

[Banterle et al., 2006] Banterle, F., Ledda, P., Debattista, K., and Chalmers, A. (2006). Inverse Tone Mapping. In *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques*, GRAPHITE, pages 349–356, New York, NY, USA. ACM.

[Barnes et al., 2010] Barnes, C., Shechtman, E., Goldman, D. B., and Finkelstein, A. (2010). The Generalized PatchMatch Correspondence Algorithm. In *European Conference on Computer Vision (ECCV)*, pages 29–43.

[Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vison and Image Understanding*, 110(3):346–359.

[Bayer, 1975] Bayer, B. E. (1975). Color Imaging Array. US Patent 3,971,065.

[Berns, 1993] Berns, R. S. (1993). The mathematical development of CIE TC 1-29 proposed colour difference equation: CIELCH. In *Proceedings of the 7th Congress of International Colour Association*, volume B, pages C19.1 – C19.4.

[Bertalmío, 2014a] Bertalmío, M. (2014a). From Image Processing to Computational Neuroscience: A Neural Model Based on Histogram Equalization. *Frontiers in Computational Neuroscience*, 8:71.

[Bertalmío, 2014b] Bertalmío, M. (2014b). *Image Processing for Cinema*. Chapman & Hall/CRC Mathematical and Computational Imaging Sciences Series. CRC Press - Taylor & Francis, 1st edition.

[Bianco et al., 2012] Bianco, S., Bruna, A., Naccari, F., and Schettini, R. (2012). Color space transformations for digital photography exploiting information about the illuminant estimation process. *Journal of the Optical Society of America (JOSA)*, 29(3):374–384.

[Bianco et al., 2007] Bianco, S., Gasparini, F., Russo, A., and Schettini, R. (2007). A new method for RGB to XYZ transformation based on pattern search optimization. *IEEE Transactions on Consumer Electronics*, 53(3):1020–1028.

[Bist et al., 2016] Bist, C., Cozot, R., Madec, G., and Ducloux, X. (2016). Style Aware Tone Expansion for HDR Displays. In *Proceedings of Graphics Interface (GI)*, pages 57–63. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine.

[Borer and Cotton, 2016] Borer, T. and Cotton, A. (2016). A Display-Independent High Dynamic Range Television System. *SMPTE Motion Imaging Journal*, 125(4):50–56.

[Borer and Cotton, 2017] Borer, T. and Cotton, A. (2017). HLG vs PQ Systems for HDR Television. *SMPTE Webinar*.

[Bradski, 2000] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*.

[Brendel, 2011] Brendel, H. (2011). ALEXA Log C Curve- Usage in VFX. Technical report, ARRI.

[Buchsbaum, 1980] Buchsbaum, G. (1980). A Spatial Processor Model for Object Colour Perception. *Journal of the Franklin Institute*, 310(1):1–26.

[Cambridge, 2005] Cambridge (2005). Cambridge in colour. `https://www.cambridgeincolour.com/tutorials/cameras-vs-human-eye.htm`.

[Chang et al., 1999] Chang, E., Cheung, S., and Pan, D. (1999). Color filter array recovery using a threshold-based variable number of gradients. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3650, pages 36–43.

[Coffin, 2010] Coffin, D. (2010). Decoding raw digital photos in Linux. `http://www.cybercom.net/~dcoffin/dcraw/`. Accessed: 2010.

[Corporation, 2009] Corporation, S. (2009). S-Log White Paper, S-Log within Digital Intermediate workflow designed for cinema release. Technical report, SONY.

[Cox et al., 1996] Cox, I. J., Hingorani, S. L., Rao, S. B., and Maggs, B. M. (1996). A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63:542–567.

[Cristy and ImageMagick Studio LLC, 1990] Cristy, J. and ImageMagick Studio LLC, A. (1990). Imagemagick. `https://github.com/ImageMagick/ImageMagick`.

[Cyriac et al., 2015a] Cyriac, P., Bertalmío, M., Kane, D., and Vazquez-Corral, J. (2015a). A Tone Mapping Operator Based on Neural and Psychophysical Models of Visual Perception. In *Proceedings of SPIE - Human Vision and Electronic Imaging XX*, volume 9394, pages 1–10.

[Cyriac et al., 2015b] Cyriac, P., Kane, D., and Bertalmío, M. (2015b). Perceptual Dynamic Range for In-Camera Image Processing. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 19.1–19.11. BMVA Press.

[Cyriac et al., 2016] Cyriac, P., Kane, D., and Bertalmío, M. (2016). Optimized Tone Curve for In-Camera Image Processing. In *IS&T Electronic Imaging Conference*, volume 2016, pages 1–7.

[Dagum and Menon, 1998] Dagum, L. and Menon, R. (1998). OpenMP: An Industry-Standard API for Shared-Memory Programming. *IEEE Computational Science and Engineering*, 5(1):46–55.

[Dartnall et al., 1983] Dartnall, H. J. A., Bowmaker, J. K., and Mollon, J. D. (1983). Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal Society of London B: Biological Sciences*, 220(1218):115–130.

[Debevec and Malik, 1997] Debevec, P. E. and Malik, J. (1997). Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 369–378, New York, NY, USA.

[Design, 2018] Design, B. (2018). Explore Blackmagic Design Products. `https://www.blackmagicdesign.com/`.

[Dolby, 2016] Dolby (2016). What is ICPCT? Technical report, Dolby.

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

[Dufaux et al., 2016] Dufaux, F., Le Callet, P., Mantiuk, R., and Mrak, M. (2016). *High Dynamic Range Video From Acquisition to Display and Applications*. Elsevier.

[Ebner, 1998] Ebner, F. (1998). *Derivation and modelling hue uniformity and development of the IPT color space*. PhD thesis, Rochester Institute of Technology.

[Eilertsen et al., 2015] Eilertsen, G., Mantiuk, R. K., and Unger, J. (2015). Real-time Noise-aware Tone Mapping. *ACM Transactions on Graphics (TOG)*, 34(6):198.1–198.15.

[Eilertsen et al., 2017] Eilertsen, G., Mantiuk, R. K., and Unger, J. (2017). A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum*, 36(2):565–592.

[Fairchild, 2013] Fairchild, M. (2013). *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley.

[Fairchild and Wyble, 2010] Fairchild, M. and Wyble, D. (2010). HDR-CIELAB and HDR-IPT: Simple models for describing the color of high-dynamic-range and wide-color-gamut images. *Proceedings - IS and T/SID Color Imaging Conference (CIC)*, pages 322–326.

[Fairchild, 2007] Fairchild, M. D. (2007). The HDR photographic survey. *Color and Imaging Conference*, 2007(1):233–238.

[Fecker et al., 2008] Fecker, U., Barkowsky, M., and Kaup, A. (2008). Histogram-based prefiltering for luminance and chrominance compensation of multiview video. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(9):1258–1267.

[Ferradans et al., 2011] Ferradans, S., Bertalmío, M., Provenzi, E., and Caselles, V. (2011). An Analysis of Visual Adaptation and Contrast Perception for Tone Mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(10):2002–2012.

[Ferradans et al., 2014] Ferradans, S., Papadakis, N., Peyr, G., and Aujol, J. (2014). Regularized Discrete Optimal Transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882.

[Ferwerda et al., 1996] Ferwerda, J. A., Pattanaik, S. N., Shirley, P., and Greenberg, D. P. (1996). A Model of Visual Adaptation for Realistic Image Synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 249–258, New York, NY, USA. ACM.

[Finlayson et al., 1994] Finlayson, G. D., Drew, M. S., and Funt, B. V. (1994). Spectral sharpening: sensor transformations for improved color constancy. *Journal of the Optical Society of America (JOSA)*, 11(5):1553–1563.

[Finlayson et al., 2001] Finlayson, G. D., Hordley, S. D., and HubeL, P. M. (2001). Color by correlation: a simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(11):1209–1221.

[Finlayson et al., 2015] Finlayson, G. D., Mackiewicz, M., and Hurlbert, A. (2015). Color correction using root-polynomial regression. *IEEE Transactions on Image Processing (TIP)*, 24(5):1460–1470.

[Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395.

[Frigo et al., 2016] Frigo, O., Sabater, N., Delon, J., and Hellier, P. (2016). Motion Driven Tonal Stabilization. *IEEE Transactions on Image Processing (TIP)*, 25(11):5455–5468.

[Froehlich et al., 2014] Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A., and Brendel, H. (2014). Creating Cinematic Wide Gamut HDR-Video for the Evaluation of Tone Mapping Operators and HDR-Displays. In *Proceedings of the SPIE*, volume 9023, pages 1–11.

[Gallo et al., 2009] Gallo, O., Gelfand, N., Chen, W., Tico, M., and Pulli, K. (2009). Artifact-free High Dynamic Range Imaging. *IEEE International Conference on Computational Photography (ICCP)*, pages 1–7.

[Gil Rodríguez and Bertalmío, 2016] Gil Rodríguez, R. and Bertalmío, M. (2016). High quality video in high dynamic range scenes from interlaced dual-ISO footage. In *IS&T Electronic Imaging Conference*, pages 1–7.

[Gil Rodríguez et al., 2015] Gil Rodríguez, R., Vazquez-Corral, J., and Bertalmío, M. (2015). The intrinsic error of exposure fusion for hdr imaging, and a way to reduce it. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 126.1–126.12. BMVA Press.

[Gil Rodríguez et al., 2017] Gil Rodríguez, R., Vazquez-Corral, J., and Bertalmío, M. (2017). Color-matching Shots from Different Cameras Having Unknown Gamma or Logarithmic Encoding Curves. In *SMPTE Annual Technical Conference & Exhibition*, pages 1–15.

[Gong et al., 2016] Gong, H., Finlayson, G., and Fisher, R. (2016). Recoding Color Transfer as A Color Homography. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 17.1–17.11. BMVA Press.

[Gong et al., 2017] Gong, H., Finlayson, G. D., Fisher, R. B., and Fang, F. (2017). 3D color homography model for photo-realistic color transfer re-coding. *The Visual Computer*, pages 1–11.

[Granados et al., 2010] Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H.-P., and Lensch, H. (2010). Optimal HDR reconstruction with linear digital cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 215–222.

[Granados et al., 2013] Granados, M., Kim, K. I., Tompkin, J., and Theobalt, C. (2013). Automatic Noise Modeling for Ghost-free HDR Reconstruction. *ACM Transactions on Graphics (TOG)*, 32(6):201.1–201.10.

[Gu et al., 2010] Gu, J., Hitomi, Y., Mitsunaga, T., and Nayar, S. (2010). Coded Rolling Shutter Photography: Flexible Space-Time Sampling. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8.

[Guild, 1931] Guild, J. (1931). The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 230(681-693):149–187.

[Guthier et al., 2013] Guthier, B., Kopf, S., and Effelsberg, W. (2013). Algorithms for a real-time HDR video system. *Pattern Recognition Letters*, 34:25 – 33.

[Guthier et al., 2014] Guthier, B., Kopf, S., Wichtlhuber, M., and Effelsberg, W. (2014). Parallel Implementation of a Real-time High Dynamic Range Video System. *Integrated Computer-Aided Engineering*, 21(2):189–202.

[HaCohen et al., 2011] HaCohen, Y., Shechtman, E., Goldman, D. B., and Lischinski, D. (2011). Non-rigid Dense Correspondence with Applications for Image Enhancement. *ACM Transactions on Graphics (TOG)*, 30(4):70.1–70.10.

[HaCohen et al., 2013] HaCohen, Y., Shechtman, E., Goldman, D. B., and Lischinski, D. (2013). Optimizing Color Consistency in Photo Collections. *ACM Transactions on Graphics (TOG)*, 32(4):85.1 – 85.9.

[Hajisharif et al., 2014] Hajisharif, S., Kronander, J., and Unger, J. (2014). HDR reconstruction for alternating gain (ISO) sensor readout. In *Eurographics*, pages 25–28.

[Hanhart et al., 2015a] Hanhart, P., Bernardo, M., Pereira, M., Pinheiro, A., and Ebrahimi, T. (2015a). Benchmarking of objective quality metrics for HDR image quality assessment. *EURASIP Journal on Image and Video Processing*, 2015(39).

[Hanhart et al., 2015b] Hanhart, P., Rerabek, M., and Ebrahimi, T. (2015b). owards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies. *Applications of Digital Image Processing XXXVIII*, page 95990G.

[Heide et al., 2014] Heide, F., Steinberger, M., Tsai, Y.-T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., Kautz, J., and Pulli, K. (2014). FlexISP: A Flexible Camera Image Processing Framework. *ACM Transactions on Graphics (TOG)*, 33(6):231.1–231.13.

[Heo et al., 2011] Heo, Y. S., Lee, K. M., Lee, S. U., Moon, Y., and Cha, J. (2011). Ghost-Free High Dynamic Range Imaging. In *Asian Conference on Computer Vision (ACCV)*, pages 486–500, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Hirakawa et al., 2005] Hirakawa, K., Member, S., and Parks, T. W. (2005). Adaptive Homogeneity-Directed Demosaicing Algorithm. *IEEE Transactions on Image Processing (TIP)*, 14:360–369.

[Hoefflinger, 2007] Hoefflinger, B. (2007). *High-Dynamic-Range (HDR) Vision*, volume 26 of *Springer Series in Advanced Microelectronics*. Springer.

[Hu et al., 2012] Hu, J., Gallo, O., and Pulli, K. (2012). Exposure Stacks of Live Scenes with Hand-held Cameras. In *European Conference on Computer Vision (ECCV)*, pages 499–512.

[Hu et al., 2013] Hu, J., Gallo, O., Pulli, K., and Sun, X. (2013). HDR Deghosting: How to Deal with Saturation? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1163–1170.

[Hubel et al., 1997] Hubel, P., Holm, J., Finlayson, G., and Drew, M. (1997). Matrix Calculations for Digital Photography. In *Color Imaging Conference (CIC)*.

[Hunt and Pointer, 2011] Hunt, R. W. G. and Pointer, M. R. (2011). *Colour Vision*, chapter 1, pages 1–17. Wiley-Blackwell.

[Hwang et al., 2014] Hwang, Y., Lee, J. Y., Kweon, I. S., and Kim, S. J. (2014). Color Transfer Using Probabilistic Moving Least Squares. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3342–3349.

[Imatest, 2018] Imatest (2018). Documentation, color correction matrix. `http://www.imatest.com/docs/colormatrix/`.

[Jameson and Hurvich, 1955] Jameson, D. and Hurvich, L. M. (1955). Some Quantitative Aspects of an Opponent-Colors Theory. I. Chromatic Responses and Spectral Saturation. *Journal of the Optical Society of America (JOSA)*, 45(7):546–552.

[Janesick, 2001] Janesick, J. (2001). *Scientific Charge-coupled Devices*. Press Monograph Series. Society of Photo Optical.

[Judd, 1979] Judd, Deane Brewster; MacAdam, D. L. (1979). *Contributions to color science*, volume 545. NIST Research Library.

[Kang et al., 2003] Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2003). High Dynamic Range Video. *ACM Transactions on Graphics (TOG)*, 22(3):319–325.

[Kindlmann, 2002] Kindlmann, G. (2002). Face-based Luminance Matching for Perceptual Colormap Generation. `https://www.cs.utah.edu/~gk/papers/vis02/talk/index.html`.

[Kirk and Andersen, 2006] Kirk, K. and Andersen, H. J. (2006). Noise characterization of weighting schemes for combination of multiple exposures. In *Proceedings of the British Machine Vision Conference*, pages 115.1–115.10.

[Kodak, 2018] Kodak (2018). BASIC SENSITOMETRY AND CHARACTERISTICS OF FILM. `https://www.kodak.com/uploadedfiles/motion/US_plugins_acrobat_en_motion_newsletters_filmEss_06_Characteristics_of_Film.pdf`.

[König and Dieterici, 1892] König, A. and Dieterici, C. (1892). Die Grundempfindungen in normalen und anomalen Farbsystemen und ihre Intensittsvertheilung im Spectrum. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, pages 241–347.

[Kotera, 2005] Kotera, H. (2005). A scene-referred color transfer for pleasant imaging on display. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 5–8.

[Kronander et al., 2012] Kronander, J., Gustavson, S., and Unger, J. (2012). Real-time HDR Video Reconstruction for Multi-sensor Systems. In *ACM SIGGRAPH Posters*, pages 65.1–65.1, New York, NY, USA.

[Kundu et al., 2017a] Kundu, D., Ghadiyaram, D., Bovik, A. C., and Evans, B. L. (2017a). Large-Scale Crowdsourced Study for Tone-Mapped HDR Pictures. *IEEE Transactions on Image Processing*, 26(10):4725–4740.

[Kundu et al., 2017b] Kundu, D., Ghadiyaram, D., Bovik, A. C., and Evans, B. L. (2017b). No-Reference Quality Assessment of Tone-Mapped HDR Pictures. *IEEE Transactions on Image Processing (TIP)*, 26(6):2957–2971.

[Kunkel et al., 2016] Kunkel, T., Daly, S., Miller, S., and Froehlich, J. (2016). Chapter 15 - Perceptual Design for High Dynamic Range Systems. In Dufaux, F., Le Callet, P., Mantiuk, R. K., and Mrak, M., editors, *High Dynamic Range Video*, pages 391 – 430. Academic Press.

[Kunkel and Reinhard, 2010] Kunkel, T. and Reinhard, E. (2010). A Reassessment of the Simultaneous Dynamic Range of the Human Visual System. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV, pages 17–24, New York, NY, USA. ACM.

[Lam and Fung, 2008] Lam, E. and Fung, G. S. K. (2008). Automatic white balancing in digital photography. In *Single-sensor imaging: Methods and applications for digital cameras*, pages 265–294.

[Lantern, 2013] Lantern, M. (2013). Dynamic range improvement for some Canon dSLRs by alternating ISO during sensor readout. `http://www.magiclantern.fm/index.html`.

[Le Gray, 1856] Le Gray, G. (1856). The Brig. `http://www.getty.edu/art/collection/objects/62016/gustave-le-gray-the-brig-french-1856/`.

[Lee et al., 2014] Lee, C., Li, Y., and Monga, V. (2014). Ghost-Free High Dynamic Range Imaging via Rank Minimization. *IEEE Signal Processing Letters*, 21(9):1045–1049.

[Lee et al., 2013] Lee, J.-Y., Matsushita, Y., Shi, B., Kweon, I. S., and Ikeuchi, K. (2013). Radiometric Calibration by Rank Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(1):144–156.

[Li et al., 2016] Li, J., Skorka, O., Ranaweera, K., and Joseph, D. (2016). Novel Real-Time Tone-Mapping Operator for Noisy Logarithmic CMOS Image Sensors. *Electronic Imaging*, 2016(12):1–13.

[Li et al., 2008] Li, X., Gunturk, B., and Zhang, L. (2008). Image demosaicing: A systematic survey. *Proceedings of SPIE*, 6822.

[Lindner and Süsstrunk, 2015] Lindner, A. J. and Süsstrunk, S. (2015). Semantic-improved color imaging applications: It is all about context. *IEEE Transactions on Multimedia*, 17:700–710.

[Lissner et al., 2013] Lissner, I., Preiss, J., Urban, P., Lichtenauer, M. S., and Zolliker, P. (2013). Image-Difference Prediction: From Grayscale to Color. *IEEE Transactions on Image Processing (TIP)*, 22(2):435–446.

[Lissner and Urban, 2012] Lissner, I. and Urban, P. (2012). Toward a Unified Color Space for Perception-Based Image Processing. *IEEE Transactions on Image Processing (TIP)*, 21(3):1153–1168.

[Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157.

[Lulé et al., 1999] Lulé, T., Keller, H., Wagner, M., and B ohm, M. (1999). LARS II - a high dynamic range - image sensor with a-Si:H photo conversion layer. *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*.

[Luo et al., 2001] Luo, M. R., Cui, G., and Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, 26(5):340–350.

[Ma et al., 2014] Ma, K., Yeganeh, H., Zeng, K., and Wang, Z. (2014). High dynamic range image tone mapping by optimizing tone mapped image quality index. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

[Mai et al., 2011] Mai, Z., Mansour, H., Mantiuk, R., Nasiopoulos, P., Ward, R. K., and Heidrich, W. (2011). Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image and Video Compression. *IEEE Transactions on Image Processing (TIP)*, 20(6):1558–1571.

[Malvar et al., 2004] Malvar, R., He, L.-w., and Cutler, R. (2004). High-Quality Linear Interpolation for Demosaicing of Bayer-Patterned Color Images. In *International Conference of Acoustic, Speech and Signal Processing*, volume 3, page 485. Institute of Electrical and Electronics Engineers, Inc.

[Mann and Picard, 1995] Mann, S. and Picard, R. W. (1995). On Being ‘undigital’ With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures. In *Proceedings of IS&T*, pages 442–448.

[Mantiuk et al., 2008] Mantiuk, R., Daly, S., and Kerofsky, L. (2008). Display adaptive tone mapping. In *ACM Transactions on Graphics (TOG)*, volume 27, page 68. ACM.

[Mantiuk et al., 2011] Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W. (2011). HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions. *ACM Transactions on Graphics (TOG)*, 30(4):40.1–40.14.

[Masia et al., 2009] Masia, B., Agustin, S., Fleming, R. W., Sorkine, O., and Gutierrez, D. (2009). Evaluation of Reverse Tone Mapping Through Varying Exposure Conditions. In *ACM SIGGRAPH Asia*, SIGGRAPH Asia, pages 160.1–160.8, New York, NY, USA. ACM.

[Masia et al., 2017] Masia, B., Serrano, A., and Gutierrez, D. (2017). Dynamic range expansion based on image statistics. *Multimedia Tools and Applications*, 76(1):631–648.

[MathWorks, 2010] MathWorks (2010). pspectro: Photometric and colorimetric calculations. `https://es.mathworks.com/matlabcentral/fileexchange/28185-pspectro--photometric-and-colorimetric-calculations?focused=5168084&tab=example`.

[McCann and Rizzi, 2011] McCann, J. J. and Rizzi, A. (2011). *The art and science of HDR imaging*, volume 26. John Wiley & Sons.

[MediaCollege, 2012] MediaCollege (2012). *CCU (Camera Control Unit) Operations*.

[Menon and Calvagno, 2011] Menon, D. and Calvagno, G. (2011). Color image demosaicking: An overview. *Signal Processing: Image Communication*, 26(8):518 – 533.

[Mertens et al., 2007] Mertens, T., Kautz, J., and Reeth, F. V. (2007). Exposure Fusion. In *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, pages 382–390, Washington, DC, USA. IEEE Computer Society.

[Meylan et al., 2007] Meylan, L., Daly, S., and Süsstrunk, S. (2007). Tone Mapping For High Dynamic Range Displays. *PT/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII*, 6492.

[Miller et al., 2012] Miller, S., Nezamabadi, M., and Daly, S. (2012). Perceptual Signal Coding for More Efficient Usage of Bit Codes. In *The 2012 Annual Technical Conference Exhibition*, pages 1–9.

[Mitsunaga and Nayar, 1999] Mitsunaga, T. and Nayar, S. (1999). Radiometric self calibration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 374–380.

[Myszkowski et al., 2008] Myszkowski, K., Mantiuk, R., and Krawczyk, G. (2008). *High Dynamic Range Video*. Synthesis lectures on computer graphics and animation. Morgan & Claypool Publishers.

[Nakamura, 2005] Nakamura, J. (2005). *Image Sensors and Signal Processing for Digital Still Cameras*. Optical Science and Engineering. CRC Press.

[Narwaria et al., 2015a] Narwaria, M., Mantiuk, R. K., Silva, M. P. D., and Callet, P. L. (2015a). HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal on Electronic Imaging*, 24(1).

[Narwaria et al., 2015b] Narwaria, M., Silva, M. P. D., and Callet, P. L. (2015b). HDR-VQM: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46 – 60.

[Narwaria et al., 2015c] Narwaria, M., Silva, M. P. D., and Callet, P. L. (2015c). Study of high dynamic range video quality assessment. In *Proceedings on SPIE*, volume 9599, pages 1 – 13.

[Nemoto et al., 2015] Nemoto, H., Korshunov, P., Hanhart, P., and Ebrahimi, T. (2015). Visual attention in LDR and HDR images. *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*.

[Nguyen et al., 2014a] Nguyen, R., Prasad, D. K., and Brown, M. S. (2014a). Raw-to-Raw: Mapping between Image Sensor Color Responses. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3398–3405.

[Nguyen et al., 2014b] Nguyen, R. M. H., Kim, S. J., and Brown, M. S. (2014b). Illuminant Aware Gamut-Based Color Transfer. *Computer Graphics Forum*, 33(7):319–328.

[Oh et al., 2015] Oh, T.-H., Lee, J.-Y., Tai, Y.-W., and Kweon, I. S. (2015). Robust High Dynamic Range Imaging by Rank Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(6):1219–1232.

[Oskarsson, 2017] Oskarsson, M. (2017). Temporally Consistent Tone Mapping of Images and Video Using Optimal K-means Clustering. *Journal of Mathematical Imaging and Vision*, 57(2):225–238.

[Packer and Williams, 2003] Packer, O. and Williams, D. (2003). Light, the Retinal Image, and Photoreceptors. *The Science of Color, Optical Socieity of America*, 41102.

[Park et al., 2016] Park, J., Tai, Y. W., Sinha, S. N., and Kweon, I. S. (2016). Efficient and robust color consistency for community photo collections. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 430–438.

[Photography, 2018] Photography, D. (2018). Digital Photography Review. `https://www.dxomark.com/itext/insights_nikon_canon/image019.png`.

[PhotographyLife, 2018] PhotographyLife (2018). The Exposure Triangle - A Beginner's Guide. `https://photographylife.com/what-is-exposure-triangle`.

[Pilz et al., 2008] Pilz, M., Honold, S., and Kienle, A. (2008). Determination of the optical properties of turbid media by measurements of the spatially resolved reflectance considering the point-spread function of the camera system. *Journal of Biomedical Optics*, 13(5):1–6.

[Pitié et al., 2005] Pitié, F., Dahyot, R., and Kokaram, A. C. (2005). N-Dimensional Probablility Density Function Transfer and its Application to Colour Transfer. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1434–1439.

[Pitié et al., 2007] Pitié, F., Kokaram, A. C., and Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1):123 – 137.

[Popov et al., 2018] Popov, V., Ostarek, M., and Tenison, C. (2018). Practices and pitfalls in inferring neural representations. *NeuroImage*, 174.

[Postma and Chorley, 2015] Postma, P. and Chorley, B. (2015). Colour Grading with Colour Management. In *SMPTE15: Persistence of Vision - Defining the Future*, pages 1–8.

[Pouli and Reinhard, 2011] Pouli, T. and Reinhard, E. (2011). Progressive color transfer for images of arbitrary dynamic range. *Computers & Graphics*, 35:67 – 80.

[Poynton, 2003] Poynton, C. (2003). *Digital Video and HD: Algorithms and Interfaces*. The Morgan Kaufmann Series in Computer Graphics. Elsevier Science.

[Rabin et al., 2014] Rabin, J., Ferradans, S., and Papadakis, N. (2014). Adaptive color transfer with relaxed optimal transport. In *IEEE International Conference on Image Processing (ICIP)*, pages 4852–4856.

[Ramanath et al., 2005] Ramanath, R., Snyder, W., Yoo, Y., and Drew, M. (2005). Color image processing pipeline. *IEEE Signal Processing Magazine*, 22(1):34–43.

[RED, 2018] RED (2018). HIGH DYNAMIC RANGE VIDEO WITH HDRX. `http://www.red.com/learn/red-101/hdrx-high-dynamic-range-video`.

[Reinhard et al., 2001] Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P. (2001). Color Transfer Between Images. *IEEE Computer Graphics and Applications*, 21(5):34–41.

[Reinhard et al., 2010] Reinhard, E., Heidrich, W., Pattanaik, S., and Debevec, P. (2010). *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*. Morgan Kaufmann series in computer graphics. Morgan Kaufmann/Elsevier.

[Reinhard et al., 2002] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic Tone Reproduction for Digital Images. *ACM Transanctions on Graphics (TOG)*, 21(3):267–276.

[Robertson et al., 2003] Robertson, M. A., Borman, S., and Stevenson, R. L. (2003). Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12(2):219–228.

[Sen et al., 2012] Sen, P., Kalantari, N. K., Yaesoubi, M., Darabi, S., Goldman, D. B., and Shechtman, E. (2012). Robust Patch-Based HDR Reconstruction of Dynamic Scenes. *ACM Transactions on Graphics (TOG)*, 31(6):203.1–203.11.

[Shapley and Hawken, 2011] Shapley, R. and Hawken, M. (2011). Color in the Cortexsingle- and double-opponent cells. *Vision research*, 51(7):701–717.

[Sharma, 2002] Sharma, G. (2002). *Digital Color Imaging Handbook*. CRC Press, Inc., Boca Raton, FL, USA.

[Sharma et al., 2005] Sharma, G., Wu, W., and Dalal, E. N. (2005). The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color research and application*, 30(1):21–30.

[Sintes, 2017] Sintes, J. (2017). Analysis and Optimization of a High Dynamic Range Video Method. `https://repositori.upf.edu/handle/10230/32923`.

[SMPTE, 2014] SMPTE (2014). ST 2084:2014 - SMPTE Standard - High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays. *SMPTE ST 2084:2014*, pages 1–14.

[Solis et al., 2010] Solis, D., López-Luke, T., Rosa, E., Espinoza, O., and Anderson, S. (2010). Syntonized white up-converted emission by Tm3+Yb3+Er3+Ho3+ doped ZrO2 nanocrystals. *Proceedings of SPIE*, 7617.

[Srikantha and Sidibé, 2012] Srikantha, A. and Sidibé, D. (2012). Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 27(6):650 – 662.

[Stevens and Stevens, 1963] Stevens, J. C. and Stevens, S. S. (1963). Brightness Function: Effects of Adaptation. *Journal of Optical Society America*, 53(3):375–385.

[Stockman and Sharpe, 2001] Stockman, A. and Sharpe, L. T. (2001). Cone spectral sensitivities and color matching. In Gegenfurtner, K., Sharpe, L., and Boycott, B., editors, *Color Vision: From genes to perception*, chapter 2, pages 53–88.

[Sugito et al., 2017] Sugito, Y., Cyriac, P., Kane, D., and Bertalmío, M. (2017). Improved High Dynamic Range Video Coding with a Nonlinearity based on Natural Image Statistics. In *3rd International Conference on Signal Processing (ICOSP)*.

[Tai et al., 2005] Tai, Y.-W., Jia, J., and Tang, C.-K. (2005). Local Color Transfer via Probabilistic Segmentation by Expectation-Maximization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 747–754.

[Tian and Cohen, 2017] Tian, Q.-C. and Cohen, L. D. (2017). *Color Consistency for Photo Collections Without Gamut Problems*, pages 90–101. Springer International Publishing, Cham.

[Tocci et al., 2011] Tocci, M. D., Kiser, C., Tocci, N., and Sen, P. (2011). A Versatile HDR Video Production System. *ACM Transactions on Graphics (TOG)*, 30(4):41.1–41.10.

[Tomić et al., 2014] Tomić, I., Karlović, I., and Jurič, I. (2014). Practical assessment of veiling glare in camera lens system. *Journal of Graphic Engineering and Design*, 5(2):23.

[Valeton and Norren, 1983] Valeton, J. and Norren, D. v. (1983). Light adaptation of primate cones: An analysis based on extracellular data. *Vision Research*, 23(12):1539 – 1547.

[Vazquez-Corral and Bertalmío, 2014a] Vazquez-Corral, J. and Bertalmío, M. (2014a). Color Stabilization Along Time and Across Shots of the Same Scene, for One or Several Cameras of Unknown Specifications. *IEEE Transactions on Image Processing (TIP)*, 23(10):4564–4575.

[Vazquez-Corral and Bertalmío, 2014b] Vazquez-Corral, J. and Bertalmío, M. (2014b). Spectral Sharpening of Color Sensors: Diagonal Color Constancy and Beyond. *Sensors*, 14(3):3965–3985.

[Vazquez-Corral and Bertalmío, 2015] Vazquez-Corral, J. and Bertalmío, M. (2015). Simultaneous blind gamma estimation. *IEEE Signal Processing Letters*, 22(9):1316–1320.

[Vazquez-Corral and Bertalmío, 2016] Vazquez-Corral, J. and Bertalmío, M. (2016). Log-encoding Estimation for Color Stabilization of Cinematic Footage. In *IEEE International Conference on Image Processing (ICIP)*, pages 3349–3353.

[Vazquez-Corral and Bertalmío, 2018] Vazquez-Corral, J. and Bertalmío, M. (2018). Spatial gamut mapping among non-inclusive gamuts. *Journal of Visual Communication and Image Representation*, 54:204–212.

[Vazquez-Corral et al., 2014] Vazquez-Corral, J., Connah, D., and Bertalmío, M. (2014). Perceptual Color Characterization of Cameras. *Sensors*, 14(12):23205–23229.

[Vazquez-Corral et al., 2012] Vazquez-Corral, J., Vanrell, M., Baldrich, R., and Tous, F. (2012). Color Constancy by Category Correlation. *IEEE Transactions on Image Processing (TIP)*, 21(4):1997–2007.

[Vigier et al., 2016] Vigier, T., Krasula, L., Milliat, A., Silva, M. P. D., and Callet, P. L. (2016). Performance and robustness of HDR objective quality metrics in the context of recent compression scenarios. In *Digital Media Industry Academic Forum (DMIAF)*, pages 59–64.

[Wandell, 1995a] Wandell, B. A. (1995a). *Foundations of Vision*. Sinauer Associates.

[Wandell, 1995b] Wandell, B. A. (1995b). Foundations of Vision: Chapter 4 Wavelength Encoding. `https://foundationsofvision.stanford.edu/chapter-4-wavelength-encoding/`. Accessed 2018.

[Wang et al., 2004] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612.

[Wang et al., 2003] Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers*, volume 2, pages 1398–1402.

[Ward,] Ward, G. Anyhere software. `http://www.anyhere.com/gward/hdrenc/Encodings.pdf`.

[Wikipedia, 2017] Wikipedia (2017). Spectral Sensitivity. `https://en.wikipedia.org/wiki/Spectral_sensitivity`.

[Wikipedia, 2018a] Wikipedia (2018a). CIE 1931 Color Space. `https://en.wikipedia.org/wiki/CIE_1931_color_space`.

[Wikipedia, 2018b] Wikipedia (2018b). CIELUV. `https://en.wikipedia.org/wiki/CIELUV`.

[Wikipedia, 2018c] Wikipedia (2018c). HSV. `https://en.wikipedia.org/wiki/HSL_and_HSV`.

[Wikipedia, 2018d] Wikipedia (2018d). Impossible color. `https://en.wikipedia.org/wiki/Impossible_color`.

[Wikipedia, 2018e] Wikipedia (2018e). Luminosity Function. `https://en.wikipedia.org/wiki/Luminosity_function`.

[Wikipedia, 2018f] Wikipedia (2018f). Optical Biology. `https://en.wikipedia.org/wiki/CIE_1931_color_space`.

[Wong, 2018] Wong, J. (2018). Optical biology. `http://jamie-wong.com/post/color/#optical-biology`.

[Wright, 1929] Wright, W. D. (1929). A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141.

[Wüller, 2016] Wüller, D. (2016). Image flare measurement according to iso 18844. In *Digital Photography and Mobile Imaging*.

[Wyszecki and Stiles, 1982] Wyszecki, G. and Stiles, W. (1982). *Color science: concepts and methods, quantitative data and formulae*. Wiley.

[Xia et al., 2017] Xia, M., Renping, J. Y., Zhang, X. M., and Xiao, J. (2017). Color Consistency Correction Based on Remapping Optimization for Image Stitching. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2977–2984.

[Xiang et al., 2009] Xiang, Y., Zou, B., and Li, H. (2009). Selective color transfer with multi-source images. *Pattern Recognition Letters*, 30(7):682–689.

[Xiao and Ma, 2006] Xiao, X. and Ma, L. (2006). Color Transfer in Correlated Color Space. In *PACM International Conference on Virtual Reality Continuum and Its Applications*, VRCIA, pages 305–309, New York, NY, USA. ACM.

[Xiao and Ma, 2010] Xiao, X. and Ma, L. (2010). Gradient-Preserving Color Transfer. *Computer Graphics Forum*, 28(7):1879–1886.

[Xu and Mulligan, 2010] Xu, W. and Mulligan, J. (2010). Performance evaluation of color correction approaches for automatic multi-view image and video stitching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 263–270.

[Yeganeh and Wang, 2013] Yeganeh, H. and Wang, Z. (2013). Objective Quality Assessment of Tone-Mapped Images. *IEEE Transactions on Image Processing (TIP)*, 22(2):657–667.

[Young, 1802] Young, T. (1802). The Bakerian Lecture: On the Theory of Light and Colours. *Philosophical Transactions of the Royal Society of London*, 92:12–48.

[Zhang and Wu, 2005] Zhang, D. and Wu, X. (2005). Color demosaicking via directional linear minimum mean square-error estimation. *IEEE Transactions on Image Processing (TIP)*, 14(12):2167–2178.

[Zhao et al., 2015] Zhao, H., Shi, B., Fernandez-Cull, C., Yeung, S.-K., and Raskar., R. (2015). Unbounded High Dynamic Range Photography using a Modulo Camera. *IEEE International Conference on Computational Photography*, pages 1–10.